Consistency and Asymptotic Normality of Instrumental Variables Estimators

So far we have analyzed, under a variety of settings, the limiting distribution of $T^{1/2}\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^{\dagger}\right)$ as well as Wald, Lagrange Multiplier and Likelihood Ratio test for $H_0 : \mathbf{R}\boldsymbol{\beta}^{\dagger}=\mathbf{r}$ versus $H_A : \mathbf{R}\boldsymbol{\beta}^{\dagger}\neq\mathbf{r}$, under the asumption that $E\left(\mathbf{X}_t\left(y_t-\mathbf{X}_t'\boldsymbol{\beta}^{\dagger}\right)\right)=0$. Indeed, we can always find a $\boldsymbol{\beta}^{\dagger}$ such that $E\left(\mathbf{X}_t\left(y_t-\mathbf{X}_t'\boldsymbol{\beta}^{\dagger}\right)\right)=0$. The issue is: are we really interested in conducting inference on $\boldsymbol{\beta}^{\dagger}$?

Suppose we want to learn about the effect on wages of an extra year of education. Also, suppose that the "true" model is:

$$\ln w_i=\beta_{0,0}+\beta_{0,1}educ_i+\beta_{0,2}abil_i+\epsilon_i$$

where w_i denote the wage of individual i, and $educ_i$ and $abil_i$ the years of education and the ability of individual i, $i=1,...,n$ (i.e. we have a cross section of individuals). We are interested in conductinfg inference on $\beta_{0,1}$. The problem is that individual ability is unobserved and is likely to be (positively) correlated with education.

Now, we run a OLS regression of $\ln w_i$ on a constant and $educ_i$,

$$\begin{aligned}&\widehat{\beta}_{ols}\\ =\ &\frac{\frac{1}{n}\sum_{i=1}^{n}\left(\ln w_i-\frac{1}{n}\sum_{i=1}^{n}\ln w_i\right)\left(educ_i-\frac{1}{n}\sum_{i=1}^{n}educ_i\right)}{\frac{1}{n}\sum_{i=1}^{n}\left(educ_i-\frac{1}{n}\sum_{i=1}^{n}educ_i\right)^2}\end{aligned}$$

It is easy to see (do it!) that

$$\widehat{\beta}_{ols}\overset{a.s.}{\to}\beta^{\dagger}=\beta_{0,1}+\rho_{ab,ed}\beta_{0,2},$$

where $\rho_{ab,ed}$ denotes the correlation between education and ability, and that

$$E\left((y_i-\left(\beta_{0,1}+\rho_{ab,ed}\beta_{0,2}\right)educ_i)educ_i\right)=0.$$

Though, we are interested in making inference on $\beta_{0,1}$ and not on $\beta_{0,1}+\rho_{ab,ed}\beta_{0,2}$!!!!

In order to do that, we need to use a different estimator, known as *Instrumental Variable (IV)* estimators. Suppose the $E(z_i educ_i)\neq 0$ and $E(z_i\left(\epsilon\right)_i)=0$, then

$$\widehat{\beta}_{iv}=\frac{\frac{1}{n}\sum_{i=1}^{n}\left(\ln w_i-\frac{1}{n}\sum_{i=1}^{n}\ln w_i\right)\left(z_i-\frac{1}{n}\sum_{i=1}^{n}z_i\right)}{\frac{1}{n}\sum_{i=1}^{n}\left(educ_i-\frac{1}{n}\sum_{i=1}^{n}educ_i\right)\left(z_i-\frac{1}{n}\sum_{i=1}^{n}z_i\right)}$$

and $\widehat{\beta}_{ols}\overset{a.s.}{\to}\beta_{0,1}$.

Let's generalize. Suppose $y_t=\mathbf{X}_t'\boldsymbol{\beta}^{\ddagger}+\epsilon_t$, where $\mathbf{X}_t$ is $k\times 1$, and we want an estimator consistent for $\boldsymbol{\beta}^{\ddagger}$. Now, $E\left(\mathbf{X}_t\left(y_t-\mathbf{X}_t'\boldsymbol{\beta}^{\ddagger}\right)\right)\neq 0$, but there are p

instruments $\mathbf{Z}_t$, $p \geq k$, such that $E\left(\mathbf{Z}_t\left(y_t - \mathbf{X}_t'\boldsymbol{\beta}^{\ddagger}\right)\right) = 0$ and $E\left(\mathbf{Z}_t\mathbf{X}_t'\right)$ is of rank k uniformly in t.

Typically, this scenario appears in simultaneous equations systems, i.e.

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta}^{\ddagger} + \boldsymbol{\epsilon}$$

$$\mathbf{X} = \mathbf{Z}\boldsymbol{\Pi} + \mathbf{u}$$

where $\mathbf{y}$ is $T \times 1$, $\mathbf{X}$ is $T \times k$, $\boldsymbol{\beta}$ is $k \times 1$, $\mathbf{Z}$ is $T \times p$, $p \geq k$, $\boldsymbol{\Pi}$ $p \times k$. If $E\left(\mathbf{u}_t\epsilon_t\right) \neq 0$, then $E\left(\mathbf{Z}_t\epsilon_t\right) \neq 0$.

In this case, we want to use the IV estimator

$$\begin{aligned}
&\widehat{\boldsymbol{\beta}}_{T,IV} \\
= \quad &\left(\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{X}_t\mathbf{Z}_t'\right)\widehat{\mathbf{P}}_T^{zz}\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{Z}_t\mathbf{X}_t'\right)\right)^{-1} \\
&\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{X}_t\mathbf{Z}_t'\right)\widehat{\mathbf{P}}_T^{zz}\frac{1}{T}\sum_{t=1}^{T}\mathbf{Z}_t y_t,
\end{aligned}$$

where $\mathbf{P}_T^{zz}$ is $p \times p$.[1]

Assumption IV-1:

(i) $y_t = \mathbf{X}_t'\boldsymbol{\beta}^{\ddagger} + \epsilon_t$ and $E\left(\mathbf{Z}_t\left(y_t - \mathbf{X}_t'\boldsymbol{\beta}^{\ddagger}\right)\right) = 0$, with $\mathbf{X}_t$ $k \times 1$ and $\mathbf{Z}_t$ $p \times 1$, $p \geq k$.

(ii) $\frac{1}{T}\sum_{t=1}^{T}\mathbf{Z}_t\mathbf{X}_t' - \mathbf{Q}_T = o_p(1)$, where $\mathbf{Q}_T$ is $O(1)$ and uniformly of full rank k,

(iii) $\widehat{\mathbf{P}}_T^{zz} - \mathbf{P}_T^{zz} = o_p(1)$ and $\mathbf{P}_T^{zz}$ is $O(1)$ and uniformly positive definite.

(iv) $V_T^{-1/2}\frac{1}{T^{1/2}}\sum_{t=1}^{T}\mathbf{Z}_t\epsilon_t \xrightarrow{d} N(0, I_p)$, with $V_T = var\left(\frac{1}{T^{1/2}}\sum_{t=1}^{T}\mathbf{Z}_t\epsilon_t\right)$ is uniformly positive definite.

Theorem IV-1

(a) Let Assumption IV-1(i)-(iii) hold, then

$$\widehat{\boldsymbol{\beta}}_{T,IV} - \boldsymbol{\beta}^{\ddagger} \xrightarrow{p} 0$$

(b) Let Assumption IV-1(i)-(iv) hold, then

$$\mathbf{D}_T^{-1/2}T^{1/2}\left(\widehat{\boldsymbol{\beta}}_{T,IV} - \boldsymbol{\beta}^{\ddagger}\right) \xrightarrow{d} N(0, I_k),$$

where

$$\mathbf{D}_T = \left(\mathbf{Q}_T'\mathbf{P}_T^{zz}\mathbf{Q}_T\right)^{-1}\mathbf{Q}_T'\mathbf{P}_T^{zz}\mathbf{V}_T\mathbf{P}_T^{zz}\mathbf{Q}_T\left(\mathbf{Q}_T'\mathbf{P}_T^{zz}\mathbf{Q}_T\right)^{-1}$$

(c) Let Assumption IV-1(i)-(iv) hold, and there exists $\widehat{\mathbf{V}}_T$ such that $\widehat{\mathbf{V}}_T - \mathbf{V}_T = o_p(1)$, then

$$\widehat{\mathbf{D}}_T^{-1/2}T^{1/2}\left(\widehat{\boldsymbol{\beta}}_{T,IV} - \boldsymbol{\beta}^{\ddagger}\right) \xrightarrow{d} N(0, I_k),$$

[1] When $\widehat{\mathbf{P}}_T^{zz} = \left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{Z}_t\mathbf{Z}_t'\right)^{-1}$, then $\widehat{\boldsymbol{\beta}}_{T,IV}$ is known as 2SLS (Two-Stage Least Squares).

where

$$\begin{aligned}
& \widehat{\mathbf{D}}_T \\
= \quad & \left(\left(\frac{1}{T} \sum_{t=1}^{T} \mathbf{X}_t \mathbf{Z}_t' \right) \widehat{\mathbf{P}}_T^{zz} \left(\frac{1}{T} \sum_{t=1}^{T} \mathbf{Z}_t \mathbf{X}_t' \right) \right)^{-1} \\
& \times \left(\frac{1}{T} \sum_{t=1}^{T} \mathbf{X}_t \mathbf{Z}_t' \right) \widehat{\mathbf{P}}_T^{zz} \widehat{\mathbf{V}}_T \widehat{\mathbf{P}}_T^{zz} \left(\frac{1}{T} \sum_{t=1}^{T} \mathbf{Z}_t \mathbf{X}_t' \right) \\
& \times \left(\left(\frac{1}{T} \sum_{t=1}^{T} \mathbf{X}_t \mathbf{Z}_t' \right) \widehat{\mathbf{P}}_T^{zz} \left(\frac{1}{T} \sum_{t=1}^{T} \mathbf{Z}_t \mathbf{X}_t' \right) \right)^{-1}
\end{aligned}$$

Corollary IV-1

Let Assumption IV-1(i)-(iv) hold and assume there exists $\widehat{\mathbf{V}}_T$ such that $\widehat{\mathbf{V}}_T - \mathbf{V}_T = o_p(1)$. If $\mathbf{P}_T^{zz} = \mathbf{V}_T^{-1}$, then

$$\widehat{\mathbf{D}}_{0,T}^{-1/2} T^{1/2} \left(\widehat{\boldsymbol{\beta}}_{T,IV} - \boldsymbol{\beta}^{\ddagger} \right) \xrightarrow{d} N(0, I_k),$$

where

$$\widehat{\mathbf{D}}_{0,T} = \left(\left(\frac{1}{T} \sum_{t=1}^{T} \mathbf{X}_t \mathbf{Z}_t' \right) \widehat{\mathbf{P}}_T^{zz} \left(\frac{1}{T} \sum_{t=1}^{T} \mathbf{Z}_t \mathbf{X}_t' \right) \right)^{-1}.$$

Proof of Theorem IV-1:

(a) Given A-IV1(i)-(iii) and recalling that the inversion of a uniformly positive definite matrix is a continuous operation, and recallling Property PR1,

$$\begin{aligned}
& \widehat{\boldsymbol{\beta}}_{T,IV} - \boldsymbol{\beta}^{\ddagger} \\
= \quad & \left(\left(\frac{1}{T} \sum_{t=1}^{T} \mathbf{X}_t \mathbf{Z}_t' \right) \widehat{\mathbf{P}}_T^{zz} \left(\frac{1}{T} \sum_{t=1}^{T} \mathbf{Z}_t \mathbf{X}_t' \right) \right)^{-1} \\
& \left(\frac{1}{T} \sum_{t=1}^{T} \mathbf{X}_t \mathbf{Z}_t' \right) \widehat{\mathbf{P}}_T^{zz} \frac{1}{T} \sum_{t=1}^{T} \mathbf{Z}_t \epsilon_t \\
& \xrightarrow{p} \left(\mathbf{Q}_T' \mathbf{P}_T^{zz} \mathbf{Q}_T \right)^{-1} \mathbf{Q}_T' \mathbf{P}_T^{zz} \times \mathbf{0}
\end{aligned}$$

(b) Let

$$\begin{aligned}
& \mathbf{D}_T^{-1/2} T^{1/2} \left(\widehat{\boldsymbol{\beta}}_{T,IV} - \boldsymbol{\beta}^{\ddagger} \right) \\
= \quad & \mathbf{D}_T^{-1/2} \left(\mathbf{Q}_T' \mathbf{P}_T^{zz} \mathbf{Q}_T \right)^{-1} \mathbf{Q}_T' \mathbf{P}_T^{zz} \mathbf{V}_T^{1/2} \mathbf{V}_T^{-1/2} \frac{1}{T^{1/2}} \sum_{t=1}^{T} \mathbf{Z}_t \epsilon_t \\
& + o_P(1) \qquad (1)
\end{aligned}$$

where the $o_p(1)$ terms comes from the product rule, as $\frac{1}{T^{1/2}}\sum_{t=1}^{T}\mathbf{Z}_t\epsilon_t = O_p(1)$, as it converges in distribution, and $\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{Z}_t\mathbf{X}_t' - \mathbf{Q}_T\right) = o_P(1)$, $\left(\widehat{\mathbf{P}}_T^{zz} - \mathbf{P}_T^{zz}\right) = o_p(1)$ becasuse of A-IV1(ii)(iii). Now, given the definition of $\mathbf{D}_T$, the first term on the RHS of (1) converges in distribution to $N(0, I_k)$. The result then follows from the asymptotic equivalence lemma.

(c) By the same argument as in Part (c) of Theorem OLS-1, given that $\widehat{\mathbf{D}}_T - \mathbf{D}_T = o_p(1)$.

Hypothesis testing can be performed along the same line outline for the OLS case, simply replacing OLS estimators with IV estimators and using the appropriate estimator for the covariance matrix.

Sufficient conditions for A-IV1(ii)-(iv) and the consistent estimation of $var\left(T^{-1/2}\sum_{t=1}^{T}\mathbf{Z}_t\epsilon_t\right)$ are obtained simply estending the moment and memory conditions to Z_t too.

We have seen that instrument should satisfy two properties, i.e. they should be uncorrelated with the error and they should be correlated with the regressors.

Provided $p > k$, we can test the null,

$$H_0 : E\left(\mathbf{Z}_t\left(y_t - \mathbf{X}_t'\boldsymbol{\beta}^{\ddagger}\right)\right) = 0$$

versus

$$H_A : E\left(\mathbf{Z}_t\left(y_t - \mathbf{X}_t'\boldsymbol{\beta}^{\ddagger}\right)\right) \neq 0$$

Theorem IV-2 (Tests for overidentifying restrictions).

Let Assumption IV-1(i)-(iv) hold, and there exists $\widehat{\mathbf{V}}_T$ such that $\widehat{\mathbf{V}}_T - \mathbf{V}_T = o_p(1)$. If $\mathbf{P}_T^{zz} = \mathbf{V}_T^{-1}$, then

$$J_{IV,T} \overset{d}{\to} \chi^2_{p-k}$$

where

$$\begin{aligned} & J_{IV,T} \\ = \; & T\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{Z}_t\left(y_t - \mathbf{X}_t'\widehat{\boldsymbol{\beta}}_{T,IV}\right)\right)' \\ & \widehat{\mathbf{P}}_T^{zz}\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{Z}_t\left(y_t - \mathbf{X}_t'\widehat{\boldsymbol{\beta}}_{T,IV}\right)\right) \end{aligned}$$

Proof: We shall derive it as a special case when doing GMM testing.

Hausman Test

$$H_0 : E\left(\mathbf{X}_t\left(y_t - \mathbf{X}_t'\boldsymbol{\beta}^{\ddagger}\right)\right) = 0$$

versus

$$H_A : E\left(\mathbf{X}_t\left(y_t - \mathbf{X}_t'\boldsymbol{\beta}^{\ddagger}\right)\right) \neq 0$$

Define,

$$
\begin{aligned}
\mathbf{H}_T &= T\left(\widehat{\boldsymbol{\beta}}_{T,IV}-\widehat{\boldsymbol{\beta}}_{T,ols}\right)'\widehat{Var}\left(T^{1/2}\left(\widehat{\boldsymbol{\beta}}_{T,IV}-\widehat{\boldsymbol{\beta}}_{T,ols}\right)\right)\\
&\quad\left(\widehat{\boldsymbol{\beta}}_{T,IV}-\widehat{\boldsymbol{\beta}}_{T,ols}\right) \qquad (2)
\end{aligned}
$$

where $\widehat{Var}\left(T^{1/2}\left(\widehat{\boldsymbol{\beta}}_{T,IV}-\widehat{\boldsymbol{\beta}}_{T,ols}\right)\right)$ is an estimator of $var\left(T^{1/2}\left(\widehat{\boldsymbol{\beta}}_{T,IV}-\widehat{\boldsymbol{\beta}}_{T,ols}\right)\right)$.

$$
\begin{aligned}
& var\left(T^{1/2}\left(\widehat{\boldsymbol{\beta}}_{T,IV}-\widehat{\boldsymbol{\beta}}_{T,ols}\right)\right)\\
=\ & var\left(T^{1/2}\left(\widehat{\boldsymbol{\beta}}_{T,IV}-\boldsymbol{\beta}^{\dagger}\right)\right)
\end{aligned}
$$

$$
\begin{aligned}
&+var\left(T^{1/2}\left(\widehat{\boldsymbol{\beta}}_{T,ols}-\boldsymbol{\beta}^{\dagger}\right)\right)\\
&-2cov\left(T^{1/2}\left(\widehat{\boldsymbol{\beta}}_{T,IV}-\boldsymbol{\beta}^{\dagger}\right),T^{1/2}\left(\widehat{\boldsymbol{\beta}}_{T,ols}-\boldsymbol{\beta}^{\dagger}\right)\right) \qquad (3)
\end{aligned}
$$

The difficult part is to provide an estimator for the last term of (3). We have already see the two variance terms, it remains to see the covariance term. Consider the case in which $\mathbf{P}_T^{zz}=\mathbf{V}_T^{-1}$. Recall from (1) that,

$$
\begin{aligned}
& T^{1/2}\left(\widehat{\boldsymbol{\beta}}_{T,IV}-\boldsymbol{\beta}^{\ddagger}\right)\\
=\ & \left(\mathbf{Q}_T'\mathbf{P}_T^{zz}\mathbf{Q}_T\right)^{-1}\mathbf{Q}_T'\mathbf{P}_T^{zz}\frac{1}{T^{1/2}}\sum_{t=1}^{T}\mathbf{Z}_t\epsilon_t+o_P(1)
\end{aligned}
$$

and from (**??**) that

$$
\begin{aligned}
& T^{1/2}\left(\widehat{\boldsymbol{\beta}}_{T,ols}-\boldsymbol{\beta}^{\dagger}\right)\\
=\ & \mathbf{M}_T^{-1}\frac{1}{T^{1/2}}\sum_{t=1}^{T}\mathbf{X}_t\epsilon_t+o_P(1)
\end{aligned}
$$

$\mathbf{M}_T=T^{-1}\sum E\left(\mathbf{X}_t\mathbf{X}_t'\right)$. Under the null of $E(\mathbf{X}_t\epsilon_t)=0$, note that $\boldsymbol{\beta}^{\ddagger}=\boldsymbol{\beta}^{\dagger}$. Thus,

$$
\begin{aligned}
& cov\left(T^{1/2}\left(\widehat{\boldsymbol{\beta}}_{T,IV}-\boldsymbol{\beta}^{\dagger}\right),T^{1/2}\left(\widehat{\boldsymbol{\beta}}_{T,ols}-\boldsymbol{\beta}^{\dagger}\right)\right)\\
=\ & \left(\mathbf{Q}_T'\mathbf{P}_T^{zz}\mathbf{Q}_T\right)^{-1}\mathbf{Q}_T'\mathbf{P}_T^{zz}\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}E\left(\mathbf{Z}_t\epsilon_t\epsilon_s\mathbf{X}'s\right)\mathbf{M}_T^{-1}
\end{aligned}
$$

If $E\left(\mathbf{Z}_t\mathbf{Z}_t'\epsilon_t^2\right)=\sigma_\epsilon^2E\left(\mathbf{Z}_t\mathbf{Z}_t'\right),E\left(\mathbf{X}_t\mathbf{X}_t'\epsilon_t^2\right)=\sigma_\epsilon^2E\left(\mathbf{X}_t\mathbf{X}_t'\right)$, and if $E\left(\mathbf{Z}_t\epsilon_t\epsilon_s\mathbf{X}'s\right)=0$ for all $t\neq s$, and if $E\left(\mathbf{Z}_t\mathbf{X}_t'\epsilon_t^2\right)=\mathbf{Q}_T$, then

$$
\begin{aligned}
& cov\left(T^{1/2}\left(\widehat{\boldsymbol{\beta}}_{T,IV}-\boldsymbol{\beta}^{\dagger}\right),T^{1/2}\left(\widehat{\boldsymbol{\beta}}_{T,ols}-\boldsymbol{\beta}^{\dagger}\right)\right)\\
=\ & var\left(T^{1/2}\left(\widehat{\boldsymbol{\beta}}_{T,ols}-\boldsymbol{\beta}^{\dagger}\right)\right)=\sigma_\epsilon^2E\left(\mathbf{X}_t\mathbf{X}_t'\right)
\end{aligned}
$$

Thus, in this case

$$\begin{aligned}
& var\left(T^{1/2}\left(\widehat{\boldsymbol{\beta}}_{T,IV}-\widehat{\boldsymbol{\beta}}_{T,ols}\right)\right)\\
= \;& var\left(T^{1/2}\left(\widehat{\boldsymbol{\beta}}_{T,IV}-\boldsymbol{\beta}^{\ddagger}\right)\right)\\
& -var\left(T^{1/2}\left(\widehat{\boldsymbol{\beta}}_{T,ols}-\boldsymbol{\beta}^{\ddagger}\right)\right)
\end{aligned}$$

and nothing that $\mathbf{P}_T^{zz}=\left(\frac{1}{T}\sum_{t=1}^{T}E(\mathbf{Z}_t'\mathbf{Z}_t)\sigma_\epsilon^2\right)^{-1}$

$$\begin{aligned}
& var\left(T^{1/2}\widehat{\left(\widehat{\boldsymbol{\beta}}_{T,IV}}-\widehat{\boldsymbol{\beta}}_{T,ols}\right)\right)\\
= \;& \left(\widehat{\mathbf{Q}}_T'\widehat{\mathbf{P}}_T^{zz}\widehat{\mathbf{Q}}_T\right)^{-1}-\widehat{\sigma}_\epsilon^2 T^{-1}\sum E\left(\mathbf{X}_t\mathbf{X}_t'\right)
\end{aligned}$$

Theorem IV3 (Hausman Test)

Let Assumption IV-1(i)-(iv) hold, and assume that

$$\begin{aligned}
& \widehat{Var}\left(T^{1/2}\left(\widehat{\boldsymbol{\beta}}_{T,IV}-\widehat{\boldsymbol{\beta}}_{T,ols}\right)\right)\\
= \;& Var\left(T^{1/2}\left(\widehat{\boldsymbol{\beta}}_{T,IV}-\widehat{\boldsymbol{\beta}}_{T,ols}\right)\right)+o_p(1)
\end{aligned}$$

Then, under H_0, $\mathbf{H}_T\xrightarrow{d}\chi_k^2$ and under H_A, $\mathbf{H}_T$ diverges to infinity.

Proof: Under H_0,

$$\begin{aligned}
& T^{1/2}\left(\widehat{\boldsymbol{\beta}}_{T,IV}-\widehat{\boldsymbol{\beta}}_{T,ols}\right)\\
= \;& T^{1/2}\left(\widehat{\boldsymbol{\beta}}_{T,IV}-\boldsymbol{\beta}\dagger\right)-T^{-1/2}\left(\widehat{\boldsymbol{\beta}}_{T,ols}-\boldsymbol{\beta}^{\dagger}\right)
\end{aligned}$$

The statement then follows by the same type of argument used in the proof of the Wald test.

Under H_A,

$$\begin{aligned}
& T^{1/2}\left(\widehat{\boldsymbol{\beta}}_{T,IV}-\widehat{\boldsymbol{\beta}}_{T,ols}\right)\\
= \;& T^{1/2}\left(\widehat{\boldsymbol{\beta}}_{T,IV}-\boldsymbol{\beta}^{\ddagger}\right)-T^{-1/2}\left(\widehat{\boldsymbol{\beta}}_{T,ols}-\boldsymbol{\beta}^{\dagger}\right)\\
& +T^{1/2}\left(\boldsymbol{\beta}^{\ddagger}-\boldsymbol{\beta}^{\dagger}\right)
\end{aligned}$$

and note that $T^{1/2}\left(\boldsymbol{\beta}^{\ddagger}-\boldsymbol{\beta}^{\dagger}\right)$ diverges.

Weak Instruments

In this section we assume that all data are *iid*, and we have conditional homoskedasticity. There is almost nothing in the literature about weak instruments dealing with dependence and heterogeneity or even allowing for conditional heteroskedasticity.

We have seen how to test for instruments exogeneity in the overidentified case. Though, good instruments has to be correlated with the regressors. Consider again,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta}^{\ddagger} + \boldsymbol{\epsilon} \tag{4}$$

$$\mathbf{X} = \mathbf{Z}\boldsymbol{\Pi} + \mathbf{u} \tag{5}$$

we need that Π be of full rank k. Note that, as $E(\mathbf{Z}_t\mathbf{u}_t)$ is zero,

$$\Pi = \left(E\left(\mathbf{X}'\mathbf{X}\right)\right)^{-1} E\left(\mathbf{X}'\mathbf{Z}\right)$$

If Π is of rank less than k, then we have what we the so called *weak instruments problem.*[2]

We proceed in two stages. First, we analyze how to test for the null of NO weak instruments versus the alternative of weak instruments. For the case of $k = 1$, can just do an F test for the null $\Pi = 0$. For the case of $k > 1$, think are more complex.

Hall, Rudebusch and Wilcox (International Economic Review 1996) suggest the following approach. Note that they assume that observations are not only *iid* but also jointly normal.

Compute the k *canonical correlation* $r_1, ..., r_k$, between $\mathbf{X}$ and $\mathbf{Z}$ as the k non-negative solutions to the determinant equations:

$$\det\left(\mathbf{X}'(\mathbf{r}^2\mathbf{I}_T - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}')\mathbf{X}\right) = 0$$

Note that r_i^2 are *estimated* canonical correlations. Order them so that $r_i^2 \geq r_{i+1}^2$

As the canonical correlations are strictly linked with the k eigenvalues of $(\mathbf{X}'\mathbf{X})^1(\mathbf{X}'\mathbf{Z})(\mathbf{Z}'\mathbf{Z})^{-1}(\mathbf{Z}'\mathbf{X})$, the full rank condition is equivalent to the fact that the smallest canonical correlation is strictly positive. Let $\rho_1 \geq ...\rho_k \geq 0$, be the "population" canonical correlations.

Construct the statistic:

$$-T\ln\left(1 - r_k^2\right)$$

Under the null $H_0 : \rho_k = 0$, $-T\ln\left(1 - r_k^2\right) \overset{d}{\rightarrow} \chi_{(p-(k-1))}$, where p is the number of instruments.

[2] Sometime the case of Π of rank smaller than k is termed *irrelevant instruments,* while the case of $\mathbf{Z}'\mathbf{X}/T \simeq CT^{-1/2}$ is indeed called weak instruments (e.g. Staiger and Stock Econometrica 1997).

The statistic above assume iid normal (note that is very similar to Johansenn test for rank of cointegrating vector...).

If the full rank condition fails, then $\widehat{\boldsymbol{\beta}}_{IV,T}$ is no longer consistent for $\boldsymbol{\beta}^{\ddagger}$. Thus, means that we can no longer performs inference on $\boldsymbol{\beta}^{\ddagger}$, based on Wald, LM and LR tests. Inference with chi-squared limiting distribution can be performed using the Anderson and Rubin (AR) statistic (Annals of Mathematical Statistics 1949), which does not use the IV estimators. Suppose, we want to test

$$H_0 : \boldsymbol{\beta}^{\dagger} = \boldsymbol{\beta}_0 \text{ vs } H_A : \boldsymbol{\beta}^{\dagger} \neq \boldsymbol{\beta}_0$$

where $\boldsymbol{\beta}_0$ is a $k \times 1$ vector of given "numbers". In terms of the system (4) and (5),

$$\begin{aligned} & AR_T\left(\boldsymbol{\beta}_0\right) \\ = & \frac{\frac{1}{p}\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}_0\right)'\left(\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}\right)\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}_0\right)'}{\frac{1}{T-p}\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}_0\right)'\left(\mathbf{I}_T-\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}\right)\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}_0\right)} \end{aligned}$$

If errors are iid normal then $AR_T\left(\boldsymbol{\beta}_0\right)$ is distributed as $F(p, T-p)$, otherwise if we drop normality $p^{-1}AR_T \xrightarrow{d} \chi^2_p$, regardless failure of full rank conditions. Again, this is because the AR statistics does not make use of the IV estimators! Thus, we can test hypotheses even if in the presence of weak instruments.

Though, we now want to see what are the confidence set in the presence of weak identification. Let's $c_{0.05}$ be the 5% percent critical values of a $\chi^2_{(p-(k-1))}$. We take a grid of values to test under the null, $...\beta_{01}, \beta_{02}, ..., \beta_{0,m}...$ We say that $\beta_{0,i}$ belongs to the 95%−confidence interval if the P-value associated to $AR_T\left(\boldsymbol{\beta}_0\right)$ is larger than 0.05. In the case of failure of the rank conditions the confidence set may be unbounded. For example, suppose $k = p = 1$ and $E(X_t Z_t) = 0$, in this case the confidence set is all the real line.

More Recent Developments

(1) The Anderson and Rubin statistic is characterized by a limiting distribution with degree of freedom equal to the number of instruments. This creates a problem whenever we have a large number of instruments. Kleibergen (Econometrica 2002), proposes a modification of the AR statistics which has a limiting distribution equal to the number of parameters to be estimated, provided that the number of instruments grows at a rate slower than T, i.e. $p = p_T = o(T)$.

(2) If the number of instruments grows with the sample size, but not too fast, then it is possible to obtain consistent IV estimator even in the case of weak instrments (Chao and Swanson, Econometrica 2005).

Generalized Method of Moment Estimator

Let $g_t(\boldsymbol{\beta}) = g(y_t, \mathbf{X}_t, \mathbf{Z}_t; \boldsymbol{\beta})$ where $\boldsymbol{\beta} \in R^k$, and g is R^p-valued, with $p \geq k$. Define, the GMM estimators as:

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}_{T,GMM} &= \arg\min_{\beta \in B} \left(\frac{1}{T} \sum_{t=1}^{T} g_t(\beta) \right)' \widehat{\Omega}_T \left(\frac{1}{T} \sum_{t=1}^{T} g_t(\beta) \right) \\
&= \arg\min_{\beta \in B} G_T(\boldsymbol{\beta})' \widehat{\Omega}_T G_T(\boldsymbol{\beta})
\end{aligned}$$

with Ω_T be $p \times p$. When $p = k$ we have *exactly identified* GMM, when $p > k$ we have *overidentified* GMM. $\widehat{\Omega}_T$ is called *estimated weighting matrix.*

Note that OLS is GMM with $G_T(\boldsymbol{\beta}) = \frac{1}{T}\sum_{t=1}^{T} \mathbf{X}_t (y_t - \mathbf{X}_t' \boldsymbol{\beta})$, and $\widehat{\Omega}_T = I_k$.

Also, note that IV is GMM with $G_T(\boldsymbol{\beta}) = \frac{1}{T}\sum_{t=1}^{T} \mathbf{Z}_t (y_t - \mathbf{X}_t' \boldsymbol{\beta})$, and $\widehat{\Omega}_T = \widehat{\mathbf{P}}_T^{zz}$.

Example 1: Nonlinear IV

$$y_t = \phi(\mathbf{X}_t, \boldsymbol{\beta}) + \epsilon_t$$

and $\mathbf{X}_t$ is endogeneous, and we use $\mathbf{Z}_t$ as instruments for $\mathbf{X}_t$. I want to estimate $\boldsymbol{\beta}$. In this case,

$$G_T(\boldsymbol{\beta}) = \frac{1}{T} \sum_{t=1}^{T} \mathbf{Z}_t (y_t - \phi(\mathbf{X}_t, \boldsymbol{\beta}))$$

and one choice of weighting matrix can be $\widehat{\Omega}_T = \frac{1}{T}\sum_{t=1}^{T} \mathbf{Z}_t \mathbf{Z}_t'$, which is $p \times p$, with $p \geq k$.

Example 2: Estimation of Stochastic Differential Equations (Square Root Model)

The model below is often used for describing the dynamic of short term rates.

$$dX(t) = \kappa(\mu - X(t))\, dt + \eta X(t)^{1/2} dW(t)$$

where $X(t)$ denotes the process in continuous time, with $t \in \mathcal{R}^+$, and $W(t)$ denotes a standard Brownian motion (i.e. a process whose increments $W(t) - W(s)$, $s < t$, are independent normal with variance $t - s$). We want to estimate κ, μ and η from discretely sampled observations $X_1, ..., X_T$ (i.e. X_t is the process $X(t)$ sampled at time $t = 1, 2, ..., T$). The explicit form of the first two moments of the first two autocorrelation is known, in fact

$$E(X_t) = \mu$$

$$Var(X_t) = \frac{\mu\eta^2}{\kappa^3} (\exp(-\kappa) + \kappa - 1)$$

$$Cov(X_t, X_{t-1}) = \frac{\mu\eta^2}{2\kappa} \frac{(1 - \exp(\kappa))^2}{k^2}$$

$$Cov(X_t, X_{t-2}) = \frac{\mu\eta^2}{2\kappa}\frac{\exp(-\kappa)\left(1-\exp(\kappa)\right)^2}{k^2}$$

Thus, in this case can define $G_T(\boldsymbol{\beta})$ as the vector containing the difference between sample moments and model implied moments.

$$\begin{aligned} & G_T(\boldsymbol{\beta}) \\ = & \begin{pmatrix} \frac{1}{T}\sum_{t=1}^{T} X_t \\ \frac{1}{T}\sum_{t=1}^{T}(X_t - \frac{1}{T}\sum_{t=1}^{T} X_t)^2 \\ \frac{1}{T}\sum_{t=2}^{T}(X_t - \frac{1}{T}\sum_{t=1}^{T} X_t)(X_{t-1} - \frac{1}{T}\sum_{t=1}^{T} X_t) \\ \frac{1}{T}\sum_{t=3}^{T}(X_t - \frac{1}{T}\sum_{t=1}^{T} X_t)(X_{t-2} - \frac{1}{T}\sum_{t=1}^{T} X_t) \end{pmatrix} \\ & - \begin{pmatrix} \mu \\ \frac{\mu\eta^2}{\kappa^3}\left(\exp(-\kappa)+\kappa-1\right) \\ \frac{\mu\eta^2}{2\kappa}\frac{(1-\exp(\kappa))^2}{k^2} \\ \frac{\mu\eta^2}{2\kappa}\frac{\exp(-\kappa)(1-\exp(\kappa))^2}{k^2} \end{pmatrix} \end{aligned}$$

When the moment conditions are nonlinear in β, typically we can no longer define $\widehat{\boldsymbol{\beta}}_{T,GMM}$ in a closed form. In the nonlinear case, we need also two additional conditions, known as *unique identifiability* and *uniform law of large numbers.*

In the GMM case, we constrain our attention to the stationary case.

Define,

$$\boldsymbol{\beta}^{\dagger}_{GMM} = \arg\min_{\beta\in B} G_{\infty}(\boldsymbol{\beta})'\Omega_{\infty}G_{\infty}(\boldsymbol{\beta}),$$

where $G_{\infty}(\boldsymbol{\beta})$ is the almost sure limit of $G_T(\boldsymbol{\beta})$, i.e. $G_T(\boldsymbol{\beta}) \overset{a.s.}{\to} G_{\infty}(\boldsymbol{\beta})$, and $\widehat{\Omega}_T \overset{pr}{\to} \Omega_{\infty}$. $\boldsymbol{\beta}^{\dagger}_{GMM}$ is said to be *uniquely identifiable* if

$$G_{\infty}(\boldsymbol{\beta}^{\dagger}_{GMM})'\Omega_{\infty}G_{\infty}(\boldsymbol{\beta}^{\dagger}_{GMM}) < G_{\infty}(\boldsymbol{\beta})'\Omega_{\infty}G_{\infty}(\boldsymbol{\beta})$$

for all $\boldsymbol{\beta} \neq \boldsymbol{\beta}^{\dagger}_{GMM}$.

Note that in the OLS and IV case $\boldsymbol{\beta}^{\dagger}_{ols}$ and $\boldsymbol{\beta}^{\dagger}_{IV}$ are always uniquely identified as in that case $G_{\infty}(\boldsymbol{\beta})'\Omega_{\infty}G_{\infty}(\boldsymbol{\beta})$ is convex, and thus it has a unique minimum.

Note, that by the first order conditions:

$$\nabla_{\beta}G_{\infty}(\boldsymbol{\beta}^{\dagger}_{GMM})'\Omega_{\infty}G_{\infty}(\boldsymbol{\beta}^{\dagger}_{GMM}) = 0$$

Consider the *overidentifed case, $p > k$.* If the moment conditions are true, then $G_{\infty}(\boldsymbol{\beta}^{\dagger}_{GMM}) = 0$; otherwise $\nabla_{\beta}G_{\infty}(\boldsymbol{\beta}^{\dagger}_{GMM})'\Omega_{\infty}G_{\infty}(\boldsymbol{\beta}^{\dagger}_{GMM}) = 0$ but $G_{\infty}(\boldsymbol{\beta}^{\dagger}_{GMM}) \neq 0$. On the other hand, in the *exactly identified* case, $\nabla_{\beta}G_{\infty}(\boldsymbol{\beta}^{\dagger}_{GMM})'\Omega_{\infty}$ is an invertible matrix, and thus $\nabla_{\beta}G_{\infty}(\boldsymbol{\beta}^{\dagger}_{GMM})'\Omega_{\infty}G_{\infty}(\boldsymbol{\beta}^{\dagger}_{GMM}) = 0$ is equivalent to $G_{\infty}(\boldsymbol{\beta}^{\dagger}_{GMM}) = 0$. As we shall see below, the limiting distribution of GMM is driven by $\sqrt{T}G_T(\boldsymbol{\beta}^{\dagger}_{GMM})$. Now, in the case of mispecified overidentified models, $G_{\infty}(\boldsymbol{\beta}^{\dagger}_{GMM}) \neq 0$ and so $\sqrt{T}G_T(\boldsymbol{\beta}^{\dagger}_{GMM})$ cannot satisfy a CLT as it's a non-zero mean, and it will diverge to either plus or minus infinity. Threrefore, exactly identified GMM can be used to estimate misspecified models, but overidentified GMM cannot (see Hall and Inoue, Journal of Econometrics 2003). The issue is that typically exactly identified GMM do not perform well. For overidentified misspecified case, a possibility is Generalized Empirical Likelihood (Schennach, 2008, Annals of Statistics).

As we'll see below, before using overidentified GMM it is better to run a so called J-test to test for the validity of the overidentifying restrictions.

Uniform (Strong) Law of Large Numbers.

We say that $G_T(\boldsymbol{\beta})$ satisfies a uniform strong law of large numbers, if

$$\sup_{\beta \in B} |G_T(\boldsymbol{\beta}) - G_\infty(\boldsymbol{\beta})| \overset{a.s.}{\to} 0$$

Instead, we say that $G_T(\boldsymbol{\beta})$ satisfies a uniform weak law of large numbers, if

$$\sup_{\beta \in B} |G_T(\boldsymbol{\beta}) - G_\infty(\boldsymbol{\beta})| \overset{pr}{\to} 0.$$

Uniform convergence means that

$$G_T(\boldsymbol{\beta}) - G_\infty(\boldsymbol{\beta}) = o_p(1)$$

and the $o_p(1)$ term **does not** depend on $\boldsymbol{\beta}$.

Assumption GMM-1:

A-GMM-1(i): $\sup_{\beta \in B} |G_T(\boldsymbol{\beta}) - G_\infty(\boldsymbol{\beta})| \overset{pr}{\to} 0$ and $\widehat{\Omega}_T \overset{pr}{\to} \Omega_\infty$, with $\mathbf{B}$ be a compact set in $\mathcal{R}^k$ (uniform LLN)

A-GMM-1(ii): $G_\infty(\boldsymbol{\beta}^\dagger_{GMM})'\Omega_\infty G_\infty(\boldsymbol{\beta}^\dagger_{GMM}) < G_\infty(\boldsymbol{\beta})'\Omega_\infty G_\infty(\boldsymbol{\beta})$ Unique identifiability

A-GMM-1(iii): $G_\infty(\boldsymbol{\beta}^\dagger_{GMM}) = 0$

A-GMM-1(iv): $G_T(\boldsymbol{\beta})$ is differentiable in the interior of $\mathbf{B}$, $\widehat{\boldsymbol{\beta}}_{T,GMM}$ and $\boldsymbol{\beta}^\dagger_{GMM}$ are in the interior of $\mathbf{B}$.

A-GMM-1(v): $\nabla_\beta G_T(\boldsymbol{\beta}) - \mathbf{D}_\infty(\boldsymbol{\beta}) \overset{pr}{\to} 0$ uniformly for all $\boldsymbol{\beta}$ in a neighborhood of $\boldsymbol{\beta}^\dagger_{GMM}$, and $\mathbf{D}_\infty(\boldsymbol{\beta})$ has full rank k and is uniformly continuous in $\boldsymbol{\beta}$ all $\boldsymbol{\beta}$ in a neighborhood of $\boldsymbol{\beta}^\dagger_{GMM}$ (uniform LLN in a neighborhood of $\boldsymbol{\beta}^\dagger_{GMM}$)

A-GMM-1(vi): $\sqrt{T}V_T^{-1/2}G_T\left(\boldsymbol{\beta}^\dagger_{GMM}\right) \overset{d}{\to} N(0, I_p)$, with $V_T = var\left(\sqrt{T}G_T\left(\boldsymbol{\beta}^\dagger_{GMM}\right)\right)$ and $trV_T < \infty$ and V_T is positive definite.

A-GMM-1(i) is a uniform law of large numbers, as well as **A-GMM-1(v),** the full rank conditions is the counterpart of the "relevance of instruments" condition in IV. We'll see which additional conditions we need to pass from a pointwise LLN to a uniform LLN.

As we mentioned already, **A-GMM-1(iii)** is trivially satisfied when $p = k$, but for $p > k$ is equivalent to correctness of the moment conditions. As we'll see it can be tested.

A-GMM-1(ii) is a primitive assumptions. In certain case, when $G_\infty(\boldsymbol{\beta}^\dagger_{GMM})'\Omega_\infty G_\infty(\boldsymbol{\beta}^\dagger_{GMM})$ is convex, then it is trivially satisfied.

A-GMM-1(vi) requires that $\sqrt{T}G_T\left(\boldsymbol{\beta}^{\dagger}_{GMM}\right)$ satisfies a CLT. Note that in the OLS case, $\sqrt{T}G_T\left(\boldsymbol{\beta}^{\dagger}_{GMM}\right) = \mathbf{X}'\boldsymbol{\epsilon}/T^{1/2}$ and the IV case $\sqrt{T}G_T\left(\boldsymbol{\beta}^{\dagger}_{GMM}\right) = \mathbf{Z}'\boldsymbol{\epsilon}/T^{1/2}$. Thus, $\sqrt{T}G_T\left(\boldsymbol{\beta}^{\dagger}_{GMM}\right)$ satisfies a CLT if the data do not display too much memory and/or heterogeneity. We need nothing new here, just require that the observations satisfy one of the CLT we have seen.

Theorem GMM-1:
(a) Let A-GMM-1(i)-(ii) hold. Then,

$$\widehat{\boldsymbol{\beta}}_{T,GMM} \overset{pr}{\to} \boldsymbol{\beta}^{\dagger}_{GMM}$$

(b) Let A-GMM-1(i)-(vi) hold. Then,

$$\widehat{\Sigma}_T^{-1/2}\sqrt{T}\left(\widehat{\boldsymbol{\beta}}_{T,GMM} - \boldsymbol{\beta}^{\dagger}_{GMM}\right) \overset{d}{\to} N(0, I_k)$$

where

$$\begin{aligned}\widehat{\Sigma}_T &= \left(\nabla_\beta G_T\left(\widehat{\boldsymbol{\beta}}_{T,GMM}\right)'\widehat{\Omega}_T\nabla_\beta G_T\left(\widehat{\boldsymbol{\beta}}_{T,GMM}\right)\right)^{-1} \\ &\quad \times\nabla_\beta G_T\left(\widehat{\boldsymbol{\beta}}_{T,GMM}\right)'\widehat{\Omega}_T V_T\widehat{\Omega}_T\nabla_\beta G_T\left(\widehat{\boldsymbol{\beta}}_{T,GMM}\right) \\ &\quad \times\left(\nabla_\beta G_T\left(\widehat{\boldsymbol{\beta}}_{T,GMM}\right)'\widehat{\Omega}_T\nabla_\beta G_T\left(\widehat{\boldsymbol{\beta}}_{T,GMM}\right)\right)^{-1}\end{aligned}$$

(c) Let A-GMM-1(i)-(vi) hold. If $\Omega_\infty = V^{-1}$, where $V = \lim_{T\to\infty} V_T$, then,

$$\widetilde{\Sigma}_T^{-1/2}\sqrt{T}\left(\widehat{\boldsymbol{\beta}}_{T,GMM} - \boldsymbol{\beta}^{\dagger}_{GMM}\right) \overset{d}{\to} N(0, I_k)$$

where

$$\widetilde{\Sigma}_T = \left(\nabla_\beta G_T\left(\widehat{\boldsymbol{\beta}}_{T,GMM}\right)'\widehat{\Omega}_T\nabla_\beta G_T\left(\widehat{\boldsymbol{\beta}}_{T,GMM}\right)\right)^{-1}$$

Remark: When $\Omega_\infty = V^{-1}$, with $V = \lim Var\left(\sqrt{T}G_T\left(\boldsymbol{\beta}^{\dagger}_{GMM}\right)\right)$, we say that Ω_∞ is the *optimal weighting matrix.* In this case, primitive sufficient conditions for $\widehat{\Omega}_T \overset{pr}{\to} \Omega_\infty$ follow by the same argument used to show the consistency of the variance of the score in the OLS case.

Proof:

(a) Given A-GMM1(i), by the uniform law of large numbers,

$$G_T\left(\boldsymbol{\beta}\right)'\Omega_T G_T\left(\boldsymbol{\beta}\right) - G_\infty(\boldsymbol{\beta})'\Omega_\infty G_\infty(\boldsymbol{\beta}) = o_P(1)$$

with the $o_P(1)$ term independent of β. As the argmin is a continuous function, by Property PR1,

$$\arg\min_\beta G_T\left(\boldsymbol{\beta}\right)'\Omega_T G_T\left(\boldsymbol{\beta}\right) \overset{pr}{\to} \arg\min_\beta G_\infty(\boldsymbol{\beta})'\Omega_\infty G_\infty(\boldsymbol{\beta})$$

But, given the definition of $\widehat{\boldsymbol{\beta}}_{T,GMM}$ and $\boldsymbol{\beta}^{\dagger}_{GMM}$, and given A-GMM1(ii), unique identifiability, this means that

$$\widehat{\boldsymbol{\beta}}_{T,GMM} - \boldsymbol{\beta}^{\dagger}_{GMM} \overset{pr}{\to} 0.$$

(b) By the first order conditions,

$$\nabla_{\beta} G_T \left(\widehat{\boldsymbol{\beta}}_{T,GMM}\right)' \widehat{\Omega}_T G_T \left(\widehat{\boldsymbol{\beta}}_{T,GMM}\right) = 0.$$

Recalling the intermediate value theorem, via a expansion of the last term in the LHS above, around $\boldsymbol{\beta}^{\dagger}_{GMM}$, we have

$$\begin{aligned} 0 &= \nabla_{\beta} G_T \left(\widehat{\boldsymbol{\beta}}_{T,GMM}\right)' \widehat{\Omega}_T G_T \left(\boldsymbol{\beta}^{\dagger}_{GMM}\right) \\ &\quad + \left(\nabla_{\beta} G_T \left(\widehat{\boldsymbol{\beta}}_{T,GMM}\right)' \widehat{\Omega}_T \nabla_{\beta} G_T \left(\overline{\boldsymbol{\beta}}_{T,GMM}\right)\right) \left(\widehat{\boldsymbol{\beta}}_{T,GMM} - \boldsymbol{\beta}^{\dagger}_{GMM}\right) \end{aligned}$$

where $\overline{\boldsymbol{\beta}}_{T,GMM} \in \left(\widehat{\boldsymbol{\beta}}_{T,GMM}, \boldsymbol{\beta}^{\dagger}_{GMM}\right)$. Thus,

$$\begin{aligned} & T^{1/2} \left(\widehat{\boldsymbol{\beta}}_{T,GMM} - \boldsymbol{\beta}^{\dagger}_{GMM}\right) \\ = & \left(\nabla_{\beta} G_T \left(\widehat{\boldsymbol{\beta}}_{T,GMM}\right)' \widehat{\Omega}_T \nabla_{\beta} G_T \left(\overline{\boldsymbol{\beta}}_{T,GMM}\right)\right)^{-1} \\ & \times \nabla_{\beta} G_T \left(\widehat{\boldsymbol{\beta}}_{T,GMM}\right)' \widehat{\Omega}_T T^{1/2} G_T \left(\boldsymbol{\beta}^{\dagger}_{GMM}\right) \end{aligned}$$

We now need to show that,

$$\nabla_{\beta} G_T \left(\widehat{\boldsymbol{\beta}}_{T,GMM}\right) - \mathbf{D}_{\infty} \left(\boldsymbol{\beta}^{\dagger}_{GMM}\right) = o_p(1) \tag{6}$$

Now,

$$\begin{aligned} & \nabla_{\beta} G_T \left(\widehat{\boldsymbol{\beta}}_{T,GMM}\right) - \mathbf{D}_{\infty} \left(\boldsymbol{\beta}^{\dagger}_{GMM}\right) \\ = & \left(\nabla_{\beta} G_T \left(\widehat{\boldsymbol{\beta}}_{T,GMM}\right) - \mathbf{D}_{\infty} \left(\widehat{\boldsymbol{\beta}}_{T,GMM}\right)\right) \\ & + \left(\mathbf{D}_{\infty} \left(\widehat{\boldsymbol{\beta}}_{T,GMM}\right) - \mathbf{D}_{\infty} \left(\boldsymbol{\beta}^{\dagger}_{GMM}\right)\right) \end{aligned}$$

As $\widehat{\boldsymbol{\beta}}_{T,GMM}$ is in the interior of $\mathbf{B}$, given A-GMM1(v), $\nabla_{\beta} G_T \left(\widehat{\boldsymbol{\beta}}_{T,GMM}\right) - \mathbf{D}_{\infty} \left(\widehat{\boldsymbol{\beta}}_{T,GMM}\right) = o_p(1)$, also given the uniform continuity $\mathbf{D}_{\infty}(\boldsymbol{\beta})$ in $\boldsymbol{\beta}$ in a neighborhood of $\boldsymbol{\beta}^{\dagger}$, given that $\widehat{\boldsymbol{\beta}}_{T,GMM} - \boldsymbol{\beta}^{\dagger}_{GMM} \overset{pr}{\to} 0$, and recalling Property PR1, it follows that $\mathbf{D}_{\infty} \left(\widehat{\boldsymbol{\beta}}_{T,GMM}\right) - \mathbf{D}_{\infty} \left(\boldsymbol{\beta}^{\dagger}_{GMM}\right) = o_p(1)$. Thus, (6) follows. As, $\overline{\boldsymbol{\beta}}_{T,GMM} \in \left(\widehat{\boldsymbol{\beta}}_{T,GMM}, \boldsymbol{\beta}^{\dagger}_{GMM}\right)$,

$$\nabla_{\beta} G_T \left(\overline{\boldsymbol{\beta}}_{T,GMM}\right) - \mathbf{D}_{\infty} \left(\boldsymbol{\beta}^{\dagger}_{GMM}\right) = o_p(1)$$

by the same argument. Recalling, A-GMM-1(i), because of the product rule, by a similar argument as that used in the proof of Theorem IV-1(b),

$$\begin{aligned}
& T^{1/2}\left(\widehat{\boldsymbol{\beta}}_{T,GMM}-\boldsymbol{\beta}^{\dagger}_{GMM}\right) \\
= & \left(\mathbf{D}_{\infty}\left(\boldsymbol{\beta}^{\dagger}_{GMM}\right)'\Omega_{\infty}\mathbf{D}_{\infty}\left(\boldsymbol{\beta}^{\dagger}_{GMM}\right)\right)^{-1} \\
& \times\mathbf{D}_{\infty}\left(\boldsymbol{\beta}^{\dagger}_{GMM}\right)'\Omega_{\infty}T^{1/2}G_{T}\left(\boldsymbol{\beta}^{\dagger}_{GMM}\right)+o_{p}(1)
\end{aligned}$$

Thus, by the asymptotic equivalence lemma,

$$\Sigma_{\infty}^{-1/2}T^{1/2}\left(\widehat{\boldsymbol{\beta}}_{T,GMM}-\boldsymbol{\beta}^{\dagger}_{GMM}\right)\xrightarrow{d}N(0,I_{k}),$$

where

$$\begin{aligned}
\Sigma_{\infty} = & \left(\mathbf{D}_{\infty}\left(\boldsymbol{\beta}^{\dagger}_{GMM}\right)'\Omega_{\infty}\mathbf{D}_{\infty}\left(\boldsymbol{\beta}^{\dagger}_{GMM}\right)\right)^{-1}\mathbf{D}_{\infty}\left(\boldsymbol{\beta}^{\dagger}_{GMM}\right)'\Omega_{\infty} \\
& \mathbf{V}\boldsymbol{\Omega}_{\infty}\mathbf{D}_{\infty}\left(\boldsymbol{\beta}^{\dagger}_{GMM}\right)\left(\mathbf{D}_{\infty}\left(\boldsymbol{\beta}^{\dagger}_{GMM}\right)'\Omega_{\infty}\mathbf{D}_{\infty}\left(\boldsymbol{\beta}^{\dagger}_{GMM}\right)\right)^{-1}.
\end{aligned}$$

Now, given (6) and given A-GMM1(i),

$$\widehat{\Sigma}_{T}-\Sigma_{\infty}=o_{P}(1).$$

Thus, because of the product rule,

$$\begin{aligned}
& \widehat{\Sigma}_{T}^{-1/2}T^{1/2}\left(\widehat{\boldsymbol{\beta}}_{T,GMM}-\boldsymbol{\beta}^{\dagger}_{GMM}\right) \\
= & \Sigma_{\infty}^{-1/2}T^{1/2}\left(\widehat{\boldsymbol{\beta}}_{T,GMM}-\boldsymbol{\beta}^{\dagger}_{GMM}\right) \\
& +\left(\widehat{\Sigma}_{T}^{-1/2}-\Sigma_{\infty}^{-1/2}\right)T^{1/2}\left(\widehat{\boldsymbol{\beta}}_{T,GMM}-\boldsymbol{\beta}^{\dagger}_{GMM}\right) \\
= & \Sigma_{\infty}^{-1/2}T^{1/2}\left(\widehat{\boldsymbol{\beta}}_{T,GMM}-\boldsymbol{\beta}^{\dagger}_{GMM}\right)+o_{P}(1)
\end{aligned}$$

and the statement follows by the asymptotic equivalence lemma.

(c) Immediate, by noting that when $\Omega_{\infty}=V^{-1}$, then

$$\Sigma_{\infty}=\left(\mathbf{D}_{\infty}\left(\boldsymbol{\beta}^{\dagger}_{GMM}\right)'\Omega_{\infty}\mathbf{D}_{\infty}\left(\boldsymbol{\beta}^{\dagger}_{GMM}\right)\right)^{-1}.$$

Testing for Overidentifying Restrictions-J-test

One of the main advantage of GMM in the overidentified case, is that it can lead to a test for the validity of the moment conditions. For example, suppose we have a sample of *iid* observations X_t. We want to test whether X_t is normally distributed $N(\mu, \sigma^2)$. In this case we know that $E(X_t) = \mu$, $Var(X_t) = \sigma^2$, $E((X_t - \mu))^3 = 0$, and $E((X_t - \mu))^4 = 3\left(\sigma^2\right)^2$. Thus,

$$\begin{aligned} & G_T(\boldsymbol{\beta}) \\ = & \begin{pmatrix} \frac{1}{T}\sum_{t=1}^{T} X_t - \mu \\ \frac{1}{T}\sum_{t=1}^{T}(X_t - \frac{1}{T}\sum_{t=1}^{T} X_t)^2 - \sigma^2 \\ \frac{1}{T}\sum_{t=2}^{T}(X_t - \frac{1}{T}\sum_{t=1}^{T} X_t)^3 - 0 \\ \frac{1}{T}\sum_{t=3}^{T}(X_t - \frac{1}{T}\sum_{t=1}^{T} X_t)^4 - 3\left(\sigma^2\right)^2 \end{pmatrix} \end{aligned}$$

We estimate $\boldsymbol{\beta}$ by GMM, and then we want to use these estimate for testing the null hypothesis $H_0 : G_\infty(\boldsymbol{\beta}^\dagger_{GMM}) = \mathbf{0}$ versus $H_A : G_\infty(\boldsymbol{\beta}^\dagger_{GMM}) \neq \mathbf{0}$. Now, in the case of $p = k$ (exact identification), by FOC $G_T\left(\widehat{\boldsymbol{\beta}}_{T,GMM}\right) = 0$. On the other hand, in the overidentified case, by the FOC $G_T\left(\widehat{\boldsymbol{\beta}}_{T,GMM}\right)' \widehat{\Omega}_T G_T\left(\widehat{\boldsymbol{\beta}}_{T,GMM}\right) = 0$, but $G_T\left(\widehat{\boldsymbol{\beta}}_{T,GMM}\right)$ is NOT identically zero. Though if null is true $G_T\left(\widehat{\boldsymbol{\beta}}_{T,GMM}\right)$ will approach zero in probability, while if alternative is true it will approach a probability limit different from zero. Thus, we want to construct a statistics based on $T^{1/2}G_T\left(\widehat{\boldsymbol{\beta}}_{T,GMM}\right)$; intuitively under H_0 it will converge in distribution, under the alternative, it will diverge.

Hereafter, let

$$H_0 : G_\infty(\boldsymbol{\beta}^\dagger_{GMM}) = 0$$

$$H_A : G_\infty(\boldsymbol{\beta}^\dagger_{GMM}) \neq \mathbf{0}$$

Now, construct the following statistic (often known as J-test),

$$J_T = TG_T\left(\widehat{\boldsymbol{\beta}}_{T,GMM}\right)' \widehat{V}_T^{-1} G_T\left(\widehat{\boldsymbol{\beta}}_{T,GMM}\right)$$

where $V_T = var\left(T^{1/2}G_T(\boldsymbol{\beta}^\dagger_{GMM})\right)$ and

$$\widehat{\boldsymbol{\beta}}_{T,GMM} = \arg\min_{\beta \in B} G_T(\boldsymbol{\beta})' \widehat{V}_T^{-1} G_T(\boldsymbol{\beta})$$

The test has been computed using a GMM estimator based on the optimal weighting matrix, i.e. using as weighting matrix a consistent estimator of the inverse of the variance of the scaled moment conditions.

Note that the test we have performed for the exogeneity of the instruments was indeed a J-test.

Theorem J-test

Let Assumption A-GMM1(i)-(ii) and A-GMM1(iv)-(v) hold. Also, assume that A-GMM1(vi) hold with $\Omega_\infty = V_\infty^{-1}$, where $V_\infty = \lim_{T\to\infty} Var\left(T^{1/2}G_T\left(\boldsymbol{\beta}^\dagger_{GMM}\right)\right)$.

(a) If $\widehat{V}_T - V_\infty = o_p(1)$ and if $p > k$, then under H_0,

$$J_T \xrightarrow{d} \chi^2_{p-k}$$

(b) Under H_A, J_T diverges to infinity at rate T.

Proof:

(a) Under the assumptions above, $\widehat{\boldsymbol{\beta}}_{T,GMM} - \boldsymbol{\beta}^\dagger_{GMM}$, by part (a) of Theorem GMM-1. Via a mean value expansion around $\boldsymbol{\beta}^\dagger_{GMM}$,

$$\begin{aligned} & T^{1/2}G_T\left(\widehat{\boldsymbol{\beta}}_{T,GMM}\right) \\ = \ & T^{1/2}G_T\left(\boldsymbol{\beta}^\dagger_{GMM}\right) + \nabla_\beta G_T\left(\overline{\boldsymbol{\beta}}_{T,GMM}\right)T^{1/2}\left(\widehat{\boldsymbol{\beta}}_{T,GMM} - \boldsymbol{\beta}^\dagger_{GMM}\right) \end{aligned}$$

where $\overline{\boldsymbol{\beta}}_{T,GMM} \in \left(\widehat{\boldsymbol{\beta}}_{T,GMM}, \boldsymbol{\beta}^\dagger_{GMM}\right)$. By the same argument used in the proof of Theorem GMM-1 part (b), we have that

$$\begin{aligned} & T^{1/2}\left(\widehat{\boldsymbol{\beta}}_{T,GMM} - \boldsymbol{\beta}^\dagger_{GMM}\right) \\ = \ & -\left(\mathbf{D}_\infty\left(\boldsymbol{\beta}^\dagger_{GMM}\right)' V_\infty^{-1}\mathbf{D}_\infty\left(\boldsymbol{\beta}^\dagger_{GMM}\right)\right)^{-1} \\ & \mathbf{D}_\infty\left(\boldsymbol{\beta}^\dagger_{GMM}\right)' V_\infty^{-1}T^{1/2}G_T\left(\boldsymbol{\beta}^\dagger_{GMM}\right) \\ & + o_p(1) \end{aligned}$$

Furthermore, recall that $\nabla_\beta G_T\left(\widehat{\boldsymbol{\beta}}_{T,GMM}\right) - \mathbf{D}_\infty\left(\boldsymbol{\beta}^\dagger_{GMM}\right) = o_p(1)$. For notation brevity, let $\mathbf{D}_\infty\left(\boldsymbol{\beta}^\dagger_{GMM}\right) = \mathbf{D}_\infty$.Thus,

$$\begin{aligned} & T^{1/2}G_T\left(\widehat{\boldsymbol{\beta}}_{T,GMM}\right) \\ = \ & \Bigg(I_p - \mathbf{D}_\infty\left(\boldsymbol{\beta}^\dagger_{GMM}\right)\left(\mathbf{D}_\infty\left(\boldsymbol{\beta}^\dagger_{GMM}\right)' \mathbf{V}_\infty^{-1}\mathbf{D}_\infty\left(\boldsymbol{\beta}^\dagger_{GMM}\right)\right)^{-1} \\ & \mathbf{D}_\infty\left(\boldsymbol{\beta}^\dagger_{GMM}\right)' \mathbf{V}_\infty^{-1}\Bigg)T^{1/2}G_T\left(\boldsymbol{\beta}^\dagger_{GMM}\right) + o_p(1) \end{aligned}$$

For notation brevity, hereafter, let $\mathbf{D}_\infty\left(\boldsymbol{\beta}^\dagger_{GMM}\right) = \mathbf{D}_\infty$.Thus,

$$\lim T^{1/2}G_T\left(\widehat{\boldsymbol{\beta}}_{T,GMM}\right) \xrightarrow{d} N(0, \Sigma_\infty)$$

where

$$\begin{aligned}
& \Sigma_{\infty} \\
= & \left(I_p - \mathbf{D}_{\infty}\left(\mathbf{D}_{\infty}{}'\mathbf{V}_{\infty}^{-1}\mathbf{D}_{\infty}\right)^{-1}\mathbf{D}'_{\infty}\mathbf{V}_{\infty}^{-1}\right) \\
& \mathbf{V}_{\infty}\left(I_p - \mathbf{D}_{\infty}\left(\mathbf{D}_{\infty}\left(\boldsymbol{\beta}^{\dagger}_{GMM}\right)'\mathbf{V}_{\infty}^{-1}\mathbf{D}_{\infty}\right)^{-1}\mathbf{D}'_{\infty}\mathbf{V}_{\infty}^{-1}\right) \\
= & \left(V_{\infty}^{1/2} - \mathbf{D}_{\infty}\left(\mathbf{D}'_{\infty}V_{\infty}^{-1}\mathbf{D}_{\infty}\right)^{-1}\mathbf{D}'_{\infty}V_{\infty}^{-1/2}\right) \\
& \times\left(V_{\infty}^{1/2} - \mathbf{D}_{\infty}\left(\mathbf{D}'_{\infty}V_{\infty}^{-1}\mathbf{D}_{\infty}\right)^{-1}\mathbf{D}'_{\infty}V_{\infty}^{-1/2}\right)
\end{aligned}$$

and thus, $T^{1/2}\mathbf{V}_{\infty}^{-1/2}G_T\left(\widehat{\boldsymbol{\beta}}_{T,GMM}\right) \xrightarrow{d} N(0, \mathbf{V}_{\infty}^{-1/2}\Sigma_{\infty}\mathbf{V}_{\infty}^{-1/2})$. Now,

$$\begin{aligned}
& \mathbf{V}_{\infty}^{-1/2}\Sigma_{\infty}\mathbf{V}_{\infty}^{-1/2} \\
= & \left(I_P - \mathbf{V}_{\infty}^{-1/2}\mathbf{D}_{\infty}\left(\mathbf{D}'_{\infty}V_{\infty}^{-1}\mathbf{D}_{\infty}\right)^{-1}\mathbf{D}'_{\infty}V_{\infty}^{-1/2}\right) \\
& \times\left(I_P - \mathbf{V}_{\infty}^{-1/2}\mathbf{D}_{\infty}\left(\mathbf{D}'_{\infty}V_{\infty}^{-1}\mathbf{D}_{\infty}\right)^{-1}\mathbf{D}'_{\infty}V_{\infty}^{-1/2}\right) \\
= & \left(I_P - \mathbf{V}_{\infty}^{-1/2}\mathbf{D}_{\infty}\left(\mathbf{D}'_{\infty}V_{\infty}^{-1}\mathbf{D}_{\infty}\right)^{-1}\mathbf{D}'_{\infty}V_{\infty}^{-1/2}\right)
\end{aligned}$$

as $\left(I_P - \mathbf{V}_{\infty}^{-1/2}\mathbf{D}_{\infty}\left(\mathbf{D}'_{\infty}V_{\infty}^{-1}\mathbf{D}_{\infty}\right)^{-1}\mathbf{D}'_{\infty}V_{\infty}^{-1/2}\right)$ is idempotent. Given that $\widehat{\mathbf{V}}_T^{-1}$ is consistent for $\mathbf{V}_{\infty}^{-1}$, $T^{1/2}\widehat{\mathbf{V}}_T^{-1/2}G_T\left(\widehat{\boldsymbol{\beta}}_{T,GMM}\right) \xrightarrow{d} N(0, \mathbf{V}_{\infty}^{-1/2}\Sigma_{\infty}\mathbf{V}_{\infty}^{-1/2})$.As $\left(I_P - \mathbf{V}_{\infty}^{-1/2}\mathbf{D}_{\infty}\left(\mathbf{D}'_{\infty}V_{\infty}^{-1}\mathbf{D}_{\infty}\right)^{-1}\mathbf{D}'_{\infty}V_{\infty}^{-1/2}\right)$ has rank $p-k$, it follows that $J_T \xrightarrow{d} \chi^2_{p-k}$.

There are cases in which the optimal weighting matrix depend on the parameters, that is $\Omega_{\infty} = \Omega_{\infty}\left(\boldsymbol{\beta}\right)$. In this case, we proceed in three steps.

Step 1: We use an arbitrary positive definite weighting matrix, e.g. an identity, ans we get a first estimator $\widetilde{\boldsymbol{\beta}}_{T,GMM}$. Now, $\widetilde{\boldsymbol{\beta}}_{T,GMM}$ is inefficient but consistent.

Step 2: We use $\widetilde{\boldsymbol{\beta}}_{T,GMM}$ to construct a consistent estimator of the optimal weighting matrix, say $\widehat{\boldsymbol{\Omega}}_T\left(\widetilde{\boldsymbol{\beta}}_{T,GMM}\right)$.

Step 3: We use $\widehat{\boldsymbol{\Omega}}_T\left(\widetilde{\boldsymbol{\beta}}_{T,GMM}\right)$ in order to find an efficient estimator $\widehat{\boldsymbol{\beta}}_{T,GMM}$.

An alternative is to keep iterating, using as weighting matrix the covariance estimator obtained at the first step. Then, sto when the difference between estimators is below a given tolerance level. GMM, 2-Step Iterative GMM can be estimated by optimization procedures build in GAUSS, Matlab.

GMM tends to have a substantial small sample bias, specially when there are many monent conditions. An alternative estimator it the Countinuous Updating

Estimator (CUE) of Hansen, Eaton and Yaron (Journal of Business Economics and Statistics 1996), where

$$\begin{aligned}\widehat{\boldsymbol{\beta}}_{T,CUE} &= \arg\min_{\beta\in B}\left(\frac{1}{T}\sum_{t=1}^{T}g_t\left(\beta\right)\right)' V_T^{-1}\left(\beta\right)\left(\frac{1}{T}\sum_{t=1}^{T}g_t\left(\beta\right)\right)\\ &= \arg\min_{\beta\in B}G_T'\left(\beta\right)V_T^{-1}\left(\beta\right)G_T\left(\beta\right)\end{aligned}$$

where $V_T\left(\beta\right)=var\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}g_t\left(\beta\right)\right)$. The first order conditions now are

$$\begin{aligned}&\nabla_\beta G_T'\left(\widehat{\beta}_{CUE}\right)V_T^{-1}\left(\widehat{\beta}_{CUE}\right)\nabla_\beta G_T\left(\widehat{\beta}_{CUE}\right)\\ &-\widehat{\Lambda}'\left(\widehat{\beta}_{CUE}\right)V_T^{-1}\left(\widehat{\beta}_{CUE}\right)G_T\left(\widehat{\beta}_{CUE}\right)V_T^{-1}\left(\widehat{\beta}_{CUE}\right)G_T\left(\widehat{\beta}_{CUE}\right)\\ =\;& 0\end{aligned}$$

where $\widehat{\Lambda}\left(\widehat{\beta}_{CUE}\right)=\frac{1}{T}\sum_{t=1}^{T}g_t\nabla_\beta G_T'\left(\widehat{\beta}_{CUE}\right)\nabla_\beta g_t\left(\widehat{\beta}_{CUE}\right)$.

The difference with 2-step GMM is the second term in the first order condition, which is set to zero in usual case. The effect of this extra term is to recenter the first order condition, and reduce the bias. Big inconvenience, computationally quite cumbersone, cannot do with standard Newton-Raphson etc algorithm.

Uniform Law of Large Numbers

Assumptions A-GMM1-(i) require that $G_T(\beta) = \frac{1}{T}\sum_{t=1}^{T} g_t(\beta)$ converge to $G_\infty(\beta)$ uniformly in $\boldsymbol{\beta}$ (A-GMM-(v) requires the weaker condition of uniform convergence only in a neighborhood of $\boldsymbol{\beta}^\dagger_{GMM.}$ Uniform law of large numbers are generally required when performing nonlinear estimation, when the closed form of the estimator is no longer available.

We shall see that a (Strong) *Weak Uniform of Large Numbers* requires a pointwise (strong) weak law of large numbers (i.e. a law of large numbers for any $\boldsymbol{\beta} \in \mathbf{B}$, plus a condition known as *(strong) stochastic equicontinuity.*

We outline the case of Weak Uniform Law of Large Numbers (as if we are interested in hypothesis testing suffices). Though, keep in mind that the "strong" counterpart exists.

What below is taken from Andrews, Econometric Theory, 1992.

Stochastic Equicontinuity. $\{G_T(\boldsymbol{\beta}) : T \geq 1\}$ is stochastically equicontinuous on $\mathbf{B}$ if, $\forall \varepsilon > 0$, $\exists \delta > 0$, such that[3]

$$\limsup_{T\to\infty} P\left(\sup_{\beta\in B}\sup_{\beta'\in S(\beta,\delta)} \left|G_T\left(\boldsymbol{\beta}'\right) - G_T\left(\boldsymbol{\beta}\right)\right|\right) < \varepsilon$$

where $S(\beta, \delta)$ denote a ball of radius δ around β.[4]

In word, for any $\beta \in B$, we take a ball of radius δ, now for the worst possible β' in the $\delta-$ball around the worst possible β, the convergence has to occur.

Uniform Law of Large Numbers. If

(i) B is a totally bounded set (i.e. it can be covered by a finite numbers of balls).[5]

(ii) $\forall \boldsymbol{\beta} \in \mathbf{B}$, $G_T\left(\boldsymbol{\beta}\right) - E(G_T\left(\boldsymbol{\beta}\right)) \overset{pr}{\to} 0$ (pointwise weak law of large numbers)

(iii) $\{G_T(\boldsymbol{\beta}) : T \geq 1\}$ is stochastically equicontinuous.

Then:

$$\sup_{\beta\in B}\left|G_T\left(\boldsymbol{\beta}\right) - E(G_T\left(\boldsymbol{\beta}\right))\right| = o_P(1).$$

[3] The "prime" has nothing to do with transpose, β, β' are two elements of B.

[4] Given a sequence b_T,

$$\limsup_T b_T = \inf_T \sup_{m\geq T} b_m$$

while

$$\liminf_T b_T = \sup_T \inf_{m\geq T} b_m$$

Thus,

$$\liminf b_T \leq \lim_T b_T \leq \limsup_T b_T$$

[5] Recall a compact set is totally bounded and closed. Sometime we are interesting in thesting say $H_0 : \beta^\dagger \leq 1$ versus $H_A : \beta^\dagger > 1$. In this case we want to have a open parameter space.

We know already how to get a pointwise law of large numbers for any fixed β, the parameters space assumption is very mild, thus it remains to find some primitive conditions for stochastic equicontinuity.

Here we outline the *Lipschitz Conditions* of Andrews (there are several variant and other approaches, but this is one of the more comprehensible and most used). Hereafter: $G_T(\beta) = \frac{1}{T}\sum_{t=1}^{T} g_t\left(w_t, \beta\right)$. where w_t simply denotes the dependence of g_t on the data, e.g. in the nonlinear-IV GMM case, $g_t\left(w_t, \beta\right) = (y_t - \boldsymbol{\phi}(\mathbf{X}_t, \beta))\mathbf{Z}_t$.

Assumption Weak Lipschitz

WL1: $\left|g_t(w_t, \beta') - g_t(w_t, \beta)\right| \leq C_t(w_t) h\left(d\left(\beta, \beta'\right)\right)$ for all $\beta, \beta' \in B$ where $d\left(\beta, \beta'\right)$ is a metric, e.g. the Euclidean norm, h a deterministic function such that $h \to 0$ as $d\left(\beta, \beta'\right) \to 0$, and $C_t(w_t)$ is a measurable function.

WL2: $\frac{1}{T}\sum_{t=1}^{T}\left(C_t(w_t) - E(C_t(w_t))\right) \overset{pr}{\to} 0$.

If Assumption Weak Lipschitz hold, then $G_T(\beta)$ is stochastic equicontinuous and $E(g_t(\beta))$ is continuous on B.

Consider the example above, $g_t\left(w_t, \beta\right) = (y_t - \boldsymbol{\phi}(\mathbf{X}_t, \beta))\mathbf{Z}_t$ and suppose that ϕ is a bounded function with bounded first derivative. Also, for simplicity of notation, suppose that X_t, Z_t and so β are scalar. Then, for say $\beta < \beta'$

$$\begin{aligned}
\left|g_t(w_t, \beta') - g_t(w_t, \beta)\right| &\leq \left|\sup_{\beta} \nabla_{\beta} \boldsymbol{\phi}(\mathbf{X}_t, \beta)\right| \left(\beta - \beta'\right) |\mathbf{Z}_t| \\
&\leq \Delta\left(\beta - \beta'\right) |\mathbf{Z}_t|
\end{aligned}$$

Now, under very mild conditions $\frac{1}{T}\sum(|Z_t| - E(|Z_t|)) \overset{pr}{\to} 0$.

Some simpler conditions apply for the iid case and stationary ergodic (see attached leaflet).

Introduction to the Bootstrap

So far we have studied asymptotic normality of various estimators, OLS, IV, GMM and several related hypothesis testing. Inference on parameters is based on asymptotic critical values. But, how good is the normal approximations? Can we get some improvement over that? We shall see that *bootstrap critical values* can provide refinements over asymptotic critical value under various circumstances.

First, we want to outline the logic underlying the bootstrap, and then we see how the use of bootstrap can lead to more accurate inference.

We begin by consider a very simple situation. We have a sample of T iid observations, $X_1, ..., X_T$ and we want to test the null hypothesis:

$$H_0 : E(X_1) = \mu \text{ versus } H_A : E(X_1) \neq \mu$$

note that given the identical distribution assumption, $E(X_1) = E(X_2) = ... = E(X_T)$.

Consider the t-statistic

$$t_{\mu,T} = \frac{\frac{1}{T^{1/2}} \sum_{t=1}^{T} (X_t - \mu)}{\widehat{\sigma}_X},$$

where $\widehat{\sigma}_X^2 = \frac{1}{T} \sum_{t=1}^{T} \left(X_t - \frac{1}{T} \sum_{t=1}^{T} X_t \right)^2$. Provided, $var(X_1) < \infty$, we know that under H_0, $t_\mu \overset{d}{\to} N(0,1)$. Thus, we compare t_μ with 2.5% and 97.5% critical values of a standard normal, and we reject at 5% if we $t_{\mu,T} < -1.96$ or $t_{\mu,T} > 1.96$.

The idea underlying the bootstrap is to pretend that the sample is the population, and so we can draw from the sample as many (bootstrap) samples as we want and we construct many bootstrap statistic.

The simplest form of bootstrap is the *iid nonparametric bootstrap,* which is suitable for iid observations.

Imagine we put all our T observations in an urn, and then we make T draws with replacement (i.e. we make one draw, get one observation, put it back in the urn, get another one, put it back in the urn, and so on, for T times). Let $X_1^*, X_2^*, ..., X_T^*$ be the resampled observations, and note that $X_1^* = X_t$, $t = 1, ..., T$ with probability $1/T$. In order words, $X_1^*, X_2^*, ..., X_T^*$ is equal to $X_{I_1}, X_{I_2}, ..., X_{I_T}$, where for $i = 1, ..., T$ I_i is a random variable taking values $1, 2, ..., T$ with equal probability $1/T$. $X_1^*, X_2^*, ..., X_T^*$ form a boostrap sample. Needless to say, we can repeat the same operation and get a second bootstrap sample, and so on. Note that, given the original sample, the probability law governing the resample is nothing else that the probability law of I_i, $i = 1, ..., T$. As I_i are iid discrete uniform on $[1, T]$, X_i^* are also *iid*, conditional on the sample. Now, let E^* and Var^* denotes the mean and the variance of the resampled series, conditional on sample (note that E^* and Var^* are mean and variance operators in terms of the law governing the bootstrap, i.e. in terms of I_i, $i = 1, ..., T$).

Now, given the identical distribution, $E^*(X_1) = E^*(X_2^*) = ... = E^*(X_T^*)$, and

$$\begin{aligned} E^*(X_1^*) &= X_1\frac{1}{T} + X_2\frac{1}{T} + ... + X_T\frac{1}{T} \\ &= \frac{1}{T}\sum_{t=1}^{T} X_t \end{aligned}$$

Also,

$$E^*\left(\frac{1}{T}\sum_{t=1}^{T} X_t^*\right) = E^*(X_1^*) = \frac{1}{T}\sum_{t=1}^{T} X_t$$

Thus, the boostrap mean is equal to the sample mean.

Given that $X_1^*, ..., X_T^*$ are independent,

$$\begin{aligned} & Var^*\left(\frac{1}{T^{1/2}}\sum_{t=1}^{T} X_t^*\right) \\ = & \frac{1}{T}\sum_{t=1}^{T} Var(X_t^*) = Var^*(X_1^*) \\ \\ = & E^*(X_1^{*2}) - (E^*(X_1^*))^2 \\ = & \frac{1}{T}\sum_{t=1}^{T} X_t^2 - \left(\frac{1}{T}\sum_{t=1}^{T} X_t\right)^2 \\ = & \frac{1}{T}\sum_{t=1}^{T}\left(X_t^2 - \frac{1}{T}\sum_{t=1}^{T} X_t\right)^2. \end{aligned}$$

Thus, the boostrap variance is equal to the sample variance.

Let $\widehat{\sigma}_X^{*2} = \frac{1}{T}\sum_{t=1}^{T}\left(X_t^* - \frac{1}{T}\sum_{t=1}^{T} X_t^*\right)^2$.Given that $X_1^*, ..., X_T^*$ are *iid* with mean and variance equal to the sample mean and sample variance,

$$\begin{aligned} t_{\mu,T}^* &= \frac{\frac{1}{T^{1/2}}\sum_{t=1}^{T}\left(X_t^* - \frac{1}{T}\sum_{t=1}^{T} X_t\right)}{\widehat{\sigma}_X^*} \\ &\overset{d^*}{\to} N(0,1), \end{aligned}$$

where d^* denotes convergence in distribution according to the bootstrap probability measure, conditional on the sample. IMPORTANT: $t_{\mu,T}^* \overset{d^*}{\to} N(0,1)$, regardless whether the null hypothesis is true or not. Thus, under the null $t_{\mu,T}$ and $t_{\mu,T}^*$ have the same limiting distribution; under the alternative $t_{\mu,T}^* \overset{d^*}{\to} N(0,1)$ while $t_{\mu,T}$ diverges (to $\mp\infty$).

This suggest to proceed in the following manner. We construct B (B large) bootstrap statistics, say $t_{\mu,T}^{*(1)},...,t_{\mu,T}^{*(B)}$. We sort from the smallest to the largest. Suppose $B = 1000$, then the 25th bootstrap statistic gives the 2.5% critical values, say $z_{T,2.5\%}^{*}$ and the 975-th boot statistics the 97.5% critical values, say $z_{T,97.5\%}^{*}$. If B is large enough, then to reject H_0 if $t_{\mu,T} < z_{2.5\%}^{*}$ or $t_{\mu,T} > z_{T,97.5\%}^{*}$ and do not reject if $z_{T,2.5\%}^{*} < t_{\mu,T} < z_{T,97.5\%}^{*}$ gives a test with asymptotic (as $T \to \infty$) size equal to 5% and asymptotic unit power.

It is important to note, that the boostrap higher moments also are equal to the sample moment. In fact, given independence,

$$\begin{aligned}
& E^{*}\left(\frac{1}{T^{1/2}}\sum_{t=1}^{T}X_t^{*}\right)^{3} \\
= \quad & \frac{1}{T^{3/2}}E^{*}(X_1^{*3}) = \frac{1}{T^{1/2}}\frac{1}{T}\sum_{t=1}^{T}X_t^{3}
\end{aligned}$$

and so on for the fourth etc.

Question: is inference based on $z_{T,2.5\%}^{*}$ and $z_{T,97.5\%}^{*}$ more accurate than inference based on standard normal approximation (i.e. on ±1.96)?

Answer YES. Why?

Edgeworth Expansion

(nothing to do with Edgeworth box!)

Under mild assumptions (satisfied for the sample mean in the iid case provided there are enough finite moments), we can express the distribution of the t-statistic as a leading term, which is the CDF of a standard normal, plus other terms capturing deviation from normality. We have,

$$P\left(t_{\mu,T} \leq x\right) = \Phi(x) + T^{-1/2}p_1(x)\phi(x) + T^{-1}p_2(x)\phi(x) + T^{-3/2}p_3(x)\phi(x) ... \quad (7)$$

where $\Phi(x)$ and $\phi(x)$ are the cumulative distribution function and the density of a standard normal evaluated at x, $p_1(x)$ is a polynomial in x depending on the central third moment, $p_2(x)$ is a polynomial in x depending on the fourth moment minus 3, etc. Therefore, $p_1(x)$ captures deviation from normality in the form of skewness, $p_2(x)$ captures deviation from normality in the sense of excess kurtosis. The successive terms captures more complex deviations and higher order effect. From (7) we see that the order of approximation of the normal distribution is $T^{-1/2}$.

Analogously, we can write the Edgeworth expansion for $t_{\mu,T}^{*}$, i.e.[6]

$$P^{*}\left(t_{\mu,T}^{*} \leq x\right) = \Phi(x) + T^{-1/2}\widehat{p}_1(x)\phi(x) + T^{-1}\widehat{p}_2(x)\phi(x) + T^{-3/2}\widehat{p}_3(x)\phi(x) .. \quad (8)$$

where $\widehat{p}_1(x)$ is a polynomial in x depending on the sample central third moment, $\widehat{p}_2(x)$ is a polynomial in x depending on the sample fourth moment minus 3, etc. Therefore, as sample moments converge to population moments, and under

[6] Recall that the boostrap moments are the sample moments.

mild assumption the convergence is at rate $T^{-1/2}$, we have that $\widehat{p}_1(x) - p_1(x) = O_p(T^{-1/2})$, $\widehat{p}_2(x) - p_2(x) = O_p(T^{-1/2})$, etc. Recall that $\Pr\left(t^*_{\mu,T} \leq x\right)$ depends on the sample, and so it's a random variable, while $\Pr\left(t_{\mu,T} \leq x\right)$ it's a number (between 0 and 1!) depending on T.

Thus,

$$\Pr\left(t_{\mu,T} \leq x\right) - \Pr\left(t^*_{\mu,T} \leq x\right) = O_P(T^{-1}),$$

while

$$\Pr\left(t_{\mu,T} \leq x\right) - \Phi(x) = O(T^{-1/2}).$$

Thus, if we approximate $P\left(t_{\mu,T} \leq x\right)$ with a standard normal CDF we have an error of order $O(T^{-1/2})$, while if we approximate $\Pr\left(t_{\mu,T} \leq x\right)$ with the bootstrap distribution $P^*\left(t^*_{\mu,T} \leq x\right)$ we have an error of order $O_P(T^{-1})$. Thus, the bootstrap distribution provides a more accurate approximation than the normal CDF.

In practice, we do not compare $P\left(t_{\mu,T} \leq x\right)$ with $\Phi(x)$, but instead we compare $t_{\mu,T}$ with z_α. Let $z_{T,\alpha}$ be defined as below,

$$P\left(t_{\mu,T} \leq z_{T,\alpha}\right) = \alpha$$

and analogously, define $z^*_{T,\alpha}$ as

$$P^*\left(t^*_{\mu,T} \leq z^*_{T,\alpha}\right) = \alpha$$

Cornish Expansion

Whenever we have an Edgeworth expansion, we can always obtain a Cornish expansion by inversion.

$$z_{T,\alpha} = z_\alpha + T^{-1/2} q_1(\alpha) + T^{-1} q_2(\alpha) + T^{-3/2} q_3(\alpha) ... \tag{9}$$

where $q_1(\alpha), q_2(\alpha)$ are again polynomial in α capturing skewness and kurtosis, and

$$z^*_{T,\alpha} = z_\alpha + T^{-1/2} \widehat{q}_1(\alpha) + T^{-1} \widehat{q}_2(\alpha) + T^{-3/2} \widehat{q}_3(\alpha) ...$$

where $\widehat{q}_1(\alpha), \widehat{q}_2(\alpha)$ are again polynomial in α capturing sample skewness and sample kurtosis. Now,

$$q_1(\alpha) - \widehat{q}_1(\alpha) = O_P(T^{-1/2})$$

$$q_2(\alpha) - \widehat{q}_2(\alpha) = O_P(T^{-1/2})$$

Thus,

$$z_{T,\alpha} - z^*_{T,\alpha} = O(T^{-1})$$

while

$$z_{T,\alpha} - z_\alpha = O_P(T^{-1/2}).$$

Therefore, we say that inference based on bootstrap critical values is more accurate than that based on asymptotic normal critical values.

Bootstrap for Time Series

The iid nonparametric bootstrap does not work with dependent observation. The reason is that the resampled observations are iid, while the actual observations are not.

In the case of dependent observations things are more complicated. On one side we want to draw "blocks" of data long enough to preserve the dependence structure present in the original sample, on the other side we want to have a large enough number of blocks independent each other. The most used resampling methods for time series data is the block bootstrap (Kunsch, Annals of Statistics 1989).

Block Bootstrap: Let $T = bl$, where b denotes the number of blocks and l denotes the length of each block.

We first draw a discrete uniform random variable I_1, that can take value $0, 1, ..., T{-}l$ with probability $1/(T{-}l{+}1)$, the first block is given by $X_{I_{1+1}}, ..., X_{I_1+l}$, we then draw another discrete uniform say I_2, and the second block of length l is $X_{I_2+1}, ..., X_{I_2+l}$, and we go ahead in the same manner, until we draw the last discrete uniform say I_b, and so the last block is $X_{I_b+1}, ..., X_{I_b+l}$. Let's call the X_t^* the resampled series, and note that $X_1^*, X_2^*, ..., X_T^*$ correspond to $X_{I_1+1}, X_{I_1+2}, ..., X_{I_b+l}$, thus conditionally on the sample, the only random element is the beginning of each block. In particular $X_1^*, ..., X_l^*$, $X_{l+1}^*, ..., X_{2l}^*$, $X_{T-l+1}^*, ..., X_T^*$, conditionally on the sample, can be treated as b iid block of discrete uniform. It can be shown that conditional on the sample and for all sample but a set of measure approaching zero,

$$E^* \left(\frac{1}{T} \sum_{t=1}^{T} X_t^* \right) = \frac{1}{T} \sum_{t=1}^{T} X_t + O_P^*(l/T) \tag{10}$$

$$Var^* \left(\frac{1}{T^{1/2}} \sum_{t=1}^{T} X_t^* \right)$$

$$\begin{aligned} = \quad & \frac{1}{T} \sum_{t=l}^{T-l} \sum_{i=-l}^{l} (X_t - \frac{1}{T} \sum_{t=1}^{T} X_t)(X_{t+i} - \frac{1}{T} \sum_{t=1}^{T} X_t) \\ & + O_P(l^2/T) \end{aligned} \tag{11}$$

where E^* and Var^* denotes the expectation and the variance operator with respect to P^* (the probability law governing the resampled series, i.e. the probability law governing the iid uniform, conditional on the sample). $O_{P^*}(l/T)$ $(O_{P^*}(l^2/T))$ denotes a term converging in probability P^* to zero if $l/T \to 0$ $(l^2/T \to 0)$.

Sketch of proof of (10) and (11).

$$\begin{aligned} E^*\left(\frac{1}{T}\sum_{t=1}^{T}X_t^*\right) &= E^*\left(\frac{1}{bl}\sum_{i=1}^{b}\sum_{j=1}^{l}X_{I_i+j}\right) \\ &= E^*\left(\frac{1}{l}\sum_{j=1}^{l}X_{I_1+j}\right), \qquad (12) \end{aligned}$$

as I_i, $i=1,...,b$ are independent uniform and so, conditionally on the sample, blocks are independent and identically distributed (note that conditionally on sample that only randomness is due to $I_1,...,I_b$ that are *iid* uniform). Thus (12) can be rewritten as:

$$\begin{aligned} &\frac{1}{l}(X_1+X_2+...+X_l)\Pr(I_1=0) \\ &+\frac{1}{l}(X_2+X_3+...+X_{l+1})\Pr(I_1=1) \\ &+...+\frac{1}{l}(X_{l+1}+X_{l+2}+...+X_{2l})\Pr(I_1=l) \\ &+...+\frac{1}{l}(X_{bl-l+1}+X_{bl-l+2}+...+X_{bl})\Pr(I_1=T-l+1) \qquad (13) \end{aligned}$$

Now $\Pr(I_1=0)=\Pr(I_1=1)=...\Pr(I_1=T-l)=\frac{1}{T-l+1}$.

Note that for $l+1\leq t\leq T-l$ we have lX_t summands, while we have only 1 X_1 and X_{bl}, 2 X_2 and X_{bl-1}, and $l-1$ X_l and X_{bl-l}. Thus summing up the terms in (13) we have that $E^*\left(\frac{1}{T}\sum_{t=1}^{T}X_t^*\right)$ is equal to

$$\begin{aligned} &\frac{1}{T-l+1}\sum_{t=l+1}^{T-l}X_t+O_P(l/T) \qquad (14) \\ =\ &\frac{1}{T}\sum_{t=1}^{T}X_t+O_P(l/T) \end{aligned}$$

Now we want to sketch the proof of (11). As I_i, $i=1,2,...,b$ are *iid*, and given (14),

$$\begin{aligned} &Var^*\left(\frac{1}{T^{1/2}}\sum_{t=1}^{T}X_t^*\right) \\ =\ &Var^*\left(\frac{1}{b^{1/2}l^{1/2}}\sum_{i=1}^{b}\sum_{j=1}^{l}X_{I_i+j}\right) \\ =\ &Var^*\left(\frac{1}{l^{1/2}}\sum_{j=1}^{l}X_{I_1+j}\right) \end{aligned}$$

$$
\begin{aligned}
&= E^* \left(\frac{1}{l} \sum_{k=1}^{l} \sum_{j=1}^{l} (X_{I_1+k} - X^a)(X_{I_1+j} - X^a) \right) \\
&\quad + O_P(l^2/T),
\end{aligned}
$$

where $X^a = \frac{1}{T} \sum_{t=1}^{T} X_t$. The first term on the RHS above is in turn equal to

$$
\begin{aligned}
&\frac{1}{l} \sum_{k=1}^{l} \sum_{j=1}^{l} (X_{k+1} - X^a)(X_{j+1} - X^a) \Pr(I_1 = 0) \\
&+ \frac{1}{l} \sum_{k=1}^{l} \sum_{j=1}^{l} (X_{k+2} - X^a)(X_{j+2} - X^a) \Pr(I_1 = 1) \\
&+ ... + \frac{1}{l} \sum_{k=1}^{l} \sum_{j=1}^{l} (X_{k+(T-l)} - X^a) \\
&(X_{j+(T-l)} - X^a) \Pr(I_1 = T - l - 1)
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{1}{T-l-1} \sum_{t=l}^{T-l} \sum_{j=-l}^{l} (X_t - X^a) \\
&\quad (X_{t+j} - X^a) + O_P(l^2/T)
\end{aligned}
$$

Now Kunsch has shown that conditional on the sample, and for all sample but a set of probability measure approaching zero, as $l \to \infty$,

$$t_{\mu,T}^{*b} =$$

$$t_{\mu,T}^{*b} = \frac{\frac{1}{T^{1/2}} \sum_{t=1}^{T} (X_t^* - E^*(X_t^*))}{\widehat{\sigma}_T^{*HAC}} \xrightarrow{d^*} N(0, I). \tag{15}$$

where

$$\widehat{\sigma}_T^{*HAC} = \frac{1}{T} \sum_{k=1}^{b} \sum_{j=1}^{l} \sum_{i=1}^{l} \left(X_{I_k+i} - \overline{X}^* \right) \left(X_{I_k+i} - \overline{X}^* \right),$$

with $\overline{X}^* = T^{-1} \sum_{t=1}^{T} X_t^*$ and $I_1, ..., I_b$ are the draws from the discrete uniform on $[0, T-l-1]$, which we observe after resampling the date. Let

$$t_{\mu,T}^{HAC} = \frac{\frac{1}{T^{1/2}} \sum_{t=1}^{T} (X_t - \mu)}{\widehat{\sigma}_T^{HAC}},$$

where $\widehat{\sigma}_T^{2,HAC}$ is an HAC covariance estimator.Thus, if we use the block bootstrap, we know that $t_{\mu,T}^{HAC}$ and $t_{\mu,T}^{*b}$ have the same limiting distribution and so bootstrap critical values are asymptotically valid.

Though, while in the iid case (iid observations and iid bootstrap),

$$E^* \left(\frac{1}{T} \sum_{t=1}^{T} X_t^* \right) = \frac{1}{T} \sum_{t=1}^{T} X_t$$

and

$$Var^* \left(\frac{1}{T^{1/2}} \sum_{t=1}^{T} X_t^* \right) = \frac{1}{T} \sum_{t=1}^{T} \left(X_t - \overline{X} \right)^2 .$$

In the case of the block boostrap and dependent observations (but the same will be true if we use the block bootstrap and we have iid observations),

$$E^* \left(\frac{1}{T} \sum_{t=1}^{T} X_t^* \right) = \frac{1}{T} \sum_{t=1}^{T} X_t + O_P \left(\frac{l}{T} \right)$$

$$Var^* \left(\frac{1}{T^{1/2}} \sum_{t=1}^{T} X_t^* \right) = \frac{1}{T-l-1} \sum_{t=l}^{T-l} \sum_{j=-l}^{l} (X_t - \overline{X})(X_{t+j} - \overline{X}) + O_P \left(\frac{l^2}{T} \right)$$

As a consequence, it is no longer true that $P\left(t_{\mu,T}^{HAC} \leq x\right) - P^*\left(t_{\mu,T}^{*b} \leq x\right) = O_P(T^{-1})$.

Gotze and Hipp (Annals of Statistics 1996), for the case of stationary mixing observations, have shown that if we choose the block length l equal to the lag truncation parameter used in the construction of the HAC variance estimator (i.e. $l = m_T$), then

$$\begin{aligned} & P\left(t_{\mu,T}^{HAC} \leq x\right) - P^*\left(t_{\mu,T}^{*b} \leq x\right) \\ = \quad & O_P(lT^{-1}) + O\left(l^{-1}T^{-1/2}\right) \end{aligned}$$

Thus, for $l = T^{1/4}$,[7]

$$P\left(t_{\mu,T}^{HAC} \leq x\right) - P^*\left(t_{\mu,T}^{*b} \leq x\right) = O_P(T^{-3/4}).$$

[7] Note that while we need $m_T/T^{1/4} \to 0$ for the case of possibly heterogeneous observations, in the strict stationary case we can allow for $m_T = T^{1/4}$.

Bootstrap Refinements for GMM estimators

(based on Andrews, Econometrica 2002).

Now, we outline how to bootstrapping GMM estimators, and we see how bootstrap critical value can provide an improvement over asymptotic (normal) critical values. Improvement over standard asymptotics are called *higher order refinements.*

In the sequel, we need that $E\left(g_t\left(\beta^{\dagger}_{GMM}\right)g_{t-k}\left(\beta^{\dagger}_{GMM}\right)\right)=0$ for all $k>\kappa$, where κ is finite, that is the correlation between the moment condition is zero after the $\kappa-$th term. Currently, for the case of general nonlinear GMM estimators, there are no results about bootstrap higher order refinements for the general case, in which $\kappa=\kappa_T$ with $\kappa_T\rightarrow\infty$ as $T\rightarrow\infty$.[8]

For generality, we consider the case in which the variance of the moment conditions depend on the parameters, and therefore we use a two-step GMM approach. In the first step, we use an arbitrary $p\times p$ weigthing matrix, say Ω, and we compute,

$$\begin{aligned}&\widehat{\boldsymbol{\beta}}_{T,GMM}\\ =\;&\arg\min_{\beta\in B}\left(\frac{1}{T}\sum_{t=1}^{T}g_t\left(\beta\right)\right)'\Omega\left(\frac{1}{T}\sum_{t=1}^{T}g_t\left(\beta\right)\right)\\ =\;&\arg\min_{\beta\in B}G_T\left(\boldsymbol{\beta}\right)'\Omega G_T\left(\boldsymbol{\beta}\right).\end{aligned}\tag{16}$$

Given $\widehat{\boldsymbol{\beta}}_{T,GMM}$, we compute the second step estimator

$$\begin{aligned}&\widetilde{\boldsymbol{\beta}}_{T,GMM}\\ =\;&\arg\min_{\beta\in B}G_T\left(\boldsymbol{\beta}\right)'\widehat{\Omega}_T\left(\widehat{\boldsymbol{\beta}}_{T,GMM}\right)G_T\left(\boldsymbol{\beta}\right),\end{aligned}\tag{17}$$

where

$$\begin{aligned}&\widehat{\Omega}_T\left(\widehat{\boldsymbol{\beta}}_{T,GMM}\right)^{-1}\\ =\;&\frac{1}{T}\sum_{t=1}^{T}g_t\left(\widehat{\boldsymbol{\beta}}_{T,GMM}\right)g_t\left(\widehat{\boldsymbol{\beta}}_{T,GMM}\right)'\\ &+\frac{2}{T}\sum_{t=1}^{T}\sum_{j=1}^{\kappa}g_t\left(\widehat{\boldsymbol{\beta}}_{T,GMM}\right)g_{t-j}\left(\widehat{\boldsymbol{\beta}}_{T,GMM}\right)'\end{aligned}$$

The two-step GMM covariance matrix estimator is given by:

$$\widetilde{\sigma}^2_T=\left(D_T\left(\widetilde{\boldsymbol{\beta}}_{T,GMM}\right)\widehat{\Omega}_T\left(\widetilde{\boldsymbol{\beta}}_{T,GMM}\right)D_T\left(\widetilde{\boldsymbol{\beta}}_{T,GMM}\right)\right)^{-1},$$

[8] Inoue and Shintani (Journal of Econometrics 2006) provide GMM refinements in the case of $\kappa=\kappa_T$ for linear IV overidentified estimators).

where

$$D_T\left(\widetilde{\boldsymbol{\beta}}_{T,GMM}\right)=\frac{1}{T}\sum_{t=1}^{T}\frac{\partial}{\partial\beta}g_t\left(\boldsymbol{\beta}\right)|_{\boldsymbol{\beta}=\widetilde{\boldsymbol{\beta}}_{T,GMM}}.$$

Let $\widetilde{\sigma}_{ii,T}^2$ be the $ii-$th element of $\widetilde{\sigma}_T^2$.

Suppose, $g_t\left(\beta\right)=g\left(y_t,X_t,Z_t,\beta\right)$, we resample b blocks of length l of $\left(y_t,X_t,Z_t\right)$, in order to obtain $\left(y_t^*,X_t^*,Z_t^*\right)$.

Let

$$\begin{aligned}&g_t^*\left(\beta\right)\\=\;&g\left(y_t^*,X_t^*,Z_t^*,\beta\right)-E^*\left(g\left(y_t^*,X_t^*,Z_t^*,\widehat{\boldsymbol{\beta}}_{T,GMM}\right)\right),\end{aligned}\tag{18}$$

where

$$\begin{aligned}&\frac{1}{T}\sum_{t=1}^{T}E^*\left(g\left(y_t^*,X_t^*,Z_t^*,\widehat{\boldsymbol{\beta}}_{T,GMM}\right)\right)\\=\;&\frac{1}{T-l+1}\sum_{t=1}^{T}w_tg\left(y_t,X_t,Z_t,\widehat{\boldsymbol{\beta}}_{T,GMM}\right)\end{aligned}$$

with

$$w_t=t/l\ t=1,...,l-1$$
$$w_t=1\ t=l,...,T-l+1$$
$$w_t=\frac{T-t+1}{l},\ t=T-l+2,...,T$$

The weigth w_t is smaller than one for the first and last l observations, as they have less chances of being drawn.

Note, that in general $g\left(y_t^*,X_t^*,Z_t^*,\widehat{\beta}_{T,GMM}\right)$ has non-zero mean even if $g\left(y_t,X_t,Z_t,\beta_{GMM}^{\dagger}\right)$ has zero mean; hence the need of recentering the bootstrap moment conditions. In fact, $E^*\left(g_t^*\left(\widehat{\beta}_{T,GMM}\right)\right)=0$.

Now, we define the bootstrap counterpart of $\widehat{\boldsymbol{\beta}}_{T,GMM}$, $\widehat{\boldsymbol{\beta}}_{T,GMM}^*$

$$\begin{aligned}&\widehat{\boldsymbol{\beta}}_{T,GMM}^*\\=\;&\arg\min_{\beta\in B}\left(\frac{1}{T}\sum_{t=1}^{T}g_t^*\left(\beta\right)\right)'\Omega\left(\frac{1}{T}\sum_{t=1}^{T}g_t^*\left(\beta\right)\right)\\=\;&\arg\min_{\beta\in B}G_T^*\left(\beta\right)'\Omega G_T^*\left(\beta\right),\end{aligned}$$

where $g_t^*\left(\beta\right)$ is defined as in (18). Also, define the bootstrap counterpart of

$\widetilde{\boldsymbol{\beta}}_{T,GMM}$, $\widetilde{\boldsymbol{\beta}}^*_{T,GMM}$ as

$$\begin{aligned}
& \widetilde{\boldsymbol{\beta}}^*_{T,GMM} \\
= \; & \arg\min_{\beta\in B}\left(\frac{1}{T}\sum_{t=1}^{T} g_t^{**}\left(\beta\right)\right)' \widehat{\Omega}_T^*\left(\widehat{\boldsymbol{\beta}}^*_{T,GMM}\right)\left(\frac{1}{T}\sum_{t=1}^{T} g_t^{**}\left(\beta\right)\right) \\
= \; & \arg\min_{\beta\in B} G_T^{**}\left(\beta\right)' \widehat{\Omega}_T^*\left(\widehat{\boldsymbol{\beta}}^*_{T,GMM}\right) G_T^{**}\left(\beta\right),
\end{aligned}$$

where

$$\begin{aligned}
& g_t^{**}\left(\beta\right) \\
= \; & g\left(y_t^*, X_t^*, Z_t^*, \beta\right) - E^*\left(g\left(y_t^*, X_t^*, Z_t^*, \widetilde{\boldsymbol{\beta}}_{T,GMM}\right)\right),
\end{aligned} \tag{19}$$

and

$$\begin{aligned}
& \widehat{\Omega}_T^*\left(\widehat{\boldsymbol{\beta}}^*_{T,GMM}\right)^{-1} \\
= \; & \frac{1}{T}\sum_{t=1}^{T} g_t^{**}\left(\widehat{\boldsymbol{\beta}}^*_{T,GMM}\right) g_t^{**}\left(\widehat{\boldsymbol{\beta}}^*_{T,GMM}\right)' \\
& + \frac{2}{T}\sum_{t=1}^{T}\sum_{j=1}^{\kappa} g_t^{**}\left(\widehat{\boldsymbol{\beta}}^*_{T,GMM}\right) g_{t-j}^{**}\left(\widehat{\boldsymbol{\beta}}^*_{T,GMM}\right)'
\end{aligned}$$

Thus, $\widehat{\Omega}_T^*\left(\widehat{\boldsymbol{\beta}}^*_{T,GMM}\right)$ is the bootstrap analog of $\widehat{\Omega}_T\left(\widehat{\boldsymbol{\beta}}_{T,GMM}\right)$.

The bootstrap covariance matrix, is given by

$$\widetilde{\sigma}_T^{2*} = \left(D_T^*\left(\widetilde{\boldsymbol{\beta}}^*_{T,GMM}\right)\widehat{\Omega}_T^*\left(\widetilde{\boldsymbol{\beta}}^*_{T,GMM}\right) D_T^*\left(\widetilde{\boldsymbol{\beta}}^*_{T,GMM}\right)\right)^{-1},$$

where

$$D_T^*\left(\widetilde{\boldsymbol{\beta}}^*_{T,GMM}\right) = \frac{1}{T}\sum_{t=1}^{T}\frac{\partial}{\partial\beta} g_t^{**}\left(\boldsymbol{\beta}\right)|_{\boldsymbol{\beta}=\widetilde{\boldsymbol{\beta}}^*_{T,GMM}}.$$

Now, let $\widetilde{\sigma}_{ii,T}^{2*}$ be the $ii-th$ element of $\widetilde{\sigma}_T^{2*}$.

We are interested in testing $H_0 : \beta_i = \beta^\dagger_{i,GMM}$ vs $H_A : \beta_i \neq \beta^\dagger_{i,GMM}$.

Define the t-stastic as:

$$t_{\beta_i,T} = \frac{T^{1/2}\left(\widetilde{\boldsymbol{\beta}}_{i,T,GMM} - \boldsymbol{\beta}^\dagger_{i,GMM}\right)}{\widetilde{\sigma}_{ii,T}}$$

The bootstrap analog of $t_{\beta_i,T}$ is:

$$t^*_{\beta_i,T} = \frac{T^{1/2}\left(\widetilde{\boldsymbol{\beta}}^*_{i,T,GMM} - \widetilde{\boldsymbol{\beta}}_{i,T,GMM}\right)}{\widetilde{\sigma}^*_{ii,T}}.$$

Now, $\widetilde{\sigma}_{ii,T}^{2*}$ is the bootstrap counterpart of $\widetilde{\sigma}_{ii,T}^{2}$, but it does not coincide with $var^{*}T^{1/2}\left(\widetilde{\boldsymbol{\beta}}_{i,T,GMM}^{*}-\widetilde{\boldsymbol{\beta}}_{i,T,GMM}\right)$.
Why? The dependence in the sample moment conditions and in the bootstrap moment conditions is not the same. This is due to the so called "joint problem". Blocks are independent, conditional on the sample. So, the last observation of a block and the first of the next block are uncorrelated. Though, this is not true in the original sample. As there are b joint points (as many as the blocks), has to be taken into account.
Summaring the issue is: $\widetilde{\sigma}_{ii,T}^{2*}$ properly mimics $\widetilde{\sigma}_{ii,T}^{2}$ (i.e. $E^{*}\left(\widetilde{\sigma}_{ii,T}^{2*}\right)=\widetilde{\sigma}_{ii,T}^{2}$), but $\widetilde{\sigma}_{ii,T}^{2*}$ is NOT $var^{*}T^{1/2}\left(\widetilde{\boldsymbol{\beta}}_{i,T,GMM}^{*}-\widetilde{\boldsymbol{\beta}}_{i,T,GMM}\right)$.

We thus need a correction factor. Define

$$\begin{aligned}
&\widetilde{\widetilde{\sigma}}_{ii,T}^{2}\\
= &\left(D_{T}\left(\widetilde{\boldsymbol{\beta}}_{T,GMM}\right)\widehat{\Omega}_{T}\left(\widetilde{\boldsymbol{\beta}}_{T,GMM}\right)D_{T}\left(\widetilde{\boldsymbol{\beta}}_{T,GMM}\right)\right)^{-1}\\
&\times D_{T}\left(\widetilde{\boldsymbol{\beta}}_{T,GMM}\right)\widehat{\Omega}_{T}\left(\widetilde{\boldsymbol{\beta}}_{T,GMM}\right)\widetilde{\Omega}_{T}\left(\widetilde{\boldsymbol{\beta}}_{T,GMM}\right)^{-1}\widehat{\Omega}_{T}\left(\widetilde{\boldsymbol{\beta}}_{T,GMM}\right)\\
&\left(D_{T}\left(\widetilde{\boldsymbol{\beta}}_{T,GMM}\right)\widehat{\Omega}_{T}\left(\widetilde{\boldsymbol{\beta}}_{T,GMM}\right)D_{T}\left(\widetilde{\boldsymbol{\beta}}_{T,GMM}\right)\right)^{-1},
\end{aligned}$$

where

$$\begin{aligned}
&\widetilde{\Omega}_{T}\left(\widetilde{\boldsymbol{\beta}}_{T,GMM}\right)^{-1}\\
= &E^{*}\left(\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}g_{t}^{**}\left(\widetilde{\boldsymbol{\beta}}_{T,GMM}\right)g_{s}^{**}\left(\widetilde{\boldsymbol{\beta}}_{T,GMM}\right)'\right)\\
= &\frac{1}{l(T-l-1)}\sum_{t=0}^{T-l}\sum_{j=1}^{l}\sum_{i=1}^{l}g_{t+j}^{**}\left(\widetilde{\boldsymbol{\beta}}_{T,GMM}\right)g_{t+i}^{**}\left(\widetilde{\boldsymbol{\beta}}_{T,GMM}\right)'.
\end{aligned}$$

Note that $\widetilde{\widetilde{\sigma}}_{ii,T}^{2}=var^{*}T^{1/2}\left(\widetilde{\boldsymbol{\beta}}_{i,T,GMM}^{*}-\widetilde{\boldsymbol{\beta}}_{i,T,GMM}\right)$. The correction factor is given by

$$\tau_{ii,T}=\frac{\widetilde{\sigma}_{ii,T}}{\widetilde{\widetilde{\sigma}}_{ii,T}}.$$

Now, consider the adjusted bootstrap statistic,

$$\widetilde{t}_{\beta_{i},T}^{*}=\frac{T^{1/2}\left(\widetilde{\boldsymbol{\beta}}_{i,T,GMM}^{*}-\widetilde{\boldsymbol{\beta}}_{i,T,GMM}\right)}{\widetilde{\sigma}_{ii,T}^{*}}\tau_{ii,T}$$

which is given by the product of the bootstrap analog of the t-statistic time the correction term.
Note that in the case of *iid* bootstrap, there is no join points issue, and therefore there is no need for the adjustment factor.

Assumption A1 does not suffice for bootstrap refinements. While a complete set of sufficient conditions is provided by Assumptions 1-5 in Andrews (2002), below we just sketch which the type of assumptions we need in addition to A1 above.

Assumption A2

A2(i): $E\left(g_t\left(\beta^{\dagger}\right)g_{t+j}\left(\beta^{\dagger}\right)\right)=0$ for all $j>\kappa$

A2(ii): (y_t, X_t) is stationary and strong mixing with exponentially decaying coefficient (see e.g. Assumption 1 in Andrews (2002) or in Hall and Horowitz (1996)).

A2(iii): The t-statistic and its bootstrap counterpart admit an Edgeworth expansion

A2(iv): Let

$$\begin{aligned} f_t(\beta) &= \Big(g_t(\beta), g_t(\beta)g_{t-j}(\beta), \frac{\partial^i}{\partial\beta^i}g_t(\beta), \\ &\qquad \frac{\partial^i}{\partial\beta^i}g_t(\beta)g_{t-j}(\beta), \ j\leq\kappa \text{ and } i\geq d_1\Big) \end{aligned}$$

. The derivatives, up of order d_2 of $f_t(\beta)$ have all moments finite, and satisfy a Lipschitz condition.

We then have:

Theorem 2 (from Theorem 2 in Andrews (2002))

(a) Let A1 and A2 hold, with $d_1\geq 5$ and $d_2\geq 4$. Let $l\approx T^{\gamma}$, suppose $0\leq\xi\leq 1/2-\gamma$ and $\xi<\gamma$, then

$$P\left(\left|t_{\beta_i,T}\right|<\widetilde{z}^*_{T,\alpha/2}\right)=\alpha+O\left(T^{-(1+\xi)}\right),$$

where $\widetilde{z}^*_{T,\alpha/2}$ is such that $\Pr\left(\widetilde{t}^*_{\beta_i,T}\leq\widetilde{z}^*_{T,\alpha/2}\right)=\alpha/2$, where $t_{\beta_i,T}$ and $\widetilde{t}^*_{\beta_i,T}$ are defined as in (**??**) and (**??**).

(b) Let A1 and A2 hold, with $d_1\geq 4$ and $d_2\geq 3$. Let $l=T^{\gamma}$, suppose $0\leq\xi\leq 1/2-\gamma$ and $\xi<\gamma$, then

$$\begin{aligned} & P\left(t_{\beta_i,T}<-\widetilde{z}^*_{T,\alpha/2} \text{ or } t_{\beta_i,T}>\widetilde{z}^*_{T,1-\alpha/2}\right) \\ =\ & \alpha+O\left(T^{-1/2+\xi}\right). \end{aligned}$$

The proof of Theorem 2 is based on the following steps. First, $t_{\beta_i,T}$ can be approximated by a smooth function, say G, of $f_t(\beta^{\dagger})$, as defined in A2(iv), and the bootstrap statistic without correction term $t^*_{\beta_i,T}$ can be approximated by $G\left(f^*_t(\widehat{\beta}_T)\right)$, where $f^*_t(\widehat{\beta}_T)$ is defined as $f_t(\widehat{\beta}_T)$ but with the sample moment conditions replaced by the bootstrap ones. Then, given A2(iii), $G\left(f_t(\beta^{\dagger})\right)$ and $G\left(f^*_t(\widehat{\beta}_T)\right)$ admit an Edgeworth expansion, and, given the Lipschitz and moment condition in A2(iv), the difference between the first two terms in the

Edgeworth expansion of $G\left(f_t(\beta^{\dagger})\right)$ and $G\left(f_t^{*}(\widehat{\beta}_T)\right)$ approach zero sufficiently fast. Finally, if $\xi < \gamma$, the correction term approaches one fast enough, thus ensuring that also the Edgeworth expansions of the corrected bootstrap statistic $\widetilde{t}^{*}_{\beta_i,T}$ and that of $t_{\beta_i,T}$ get closer and closer.

From Theorem 2 it is immediate to see that, if set $\gamma = 1/4$, i.e. $l = T^{1/4}$, then ξ can be made arbitrarily close to 1/4. Thus, the bootstrap improvement in the error probability is of order $T^{-\xi}$, with $0 < \xi < 1/4$. The condition $\xi < \gamma$ ensures that as $T \to \infty$, the correction factor $\tau_{ii,T} \to 1$. As mentioned already, in the case of *iid* observation there is no need for correction factor and so we do no longer require $\xi < \gamma$. Thus, one can set $\gamma = 0$ (i.e. $l = 1$), so that $\xi = 1/2$, thus leading to a improvement in the error in the rejection probability of order $T^{-1/2}$.

If the moment conditions are a martingale difference sequence, as in the case of dynamic correct specification, then $\kappa = 0$. Though, we still need to use a block size l, with $l \to \infty$. This in order to capture dependence in the higher (higher than second) moments.

When computing higher moments there is substantial difference between mds and *iid*. Example: If ϵ_t is *iid*, then

$$\mathrm{E}\left(\epsilon_t\epsilon_s^2\right) = \mathrm{E}\left(\epsilon_t\right)\mathrm{E}\left(\epsilon_s^2\right) = 0, \text{ for all } t \neq s$$

Suppose ϵ_t is mds but not independent. Now, in the *iid* case, let $s > t$, $\mathcal{F}_t = \sigma\left(\epsilon_1, ..., \epsilon_t\right)$

$$\mathrm{E}\left(\epsilon_t\epsilon_s^2\right) = \mathrm{E}\left(\mathrm{E}\left(\epsilon_t\epsilon_s^2|\mathcal{F}_t\right)\right) = \mathrm{E}\left(\mathrm{E}\left(\epsilon_t\epsilon_s^2|\mathcal{F}_t\right)\right) = \mathrm{E}\left(\epsilon_t\mathrm{E}\left(\epsilon_s^2|\mathcal{F}_t\right)\right).$$

Now, as $\mathrm{E}\left(\epsilon_s^2|\mathcal{F}_t\right)$ can be a measurable function of$\mathcal{F}_t$, $\mathrm{E}\left(\epsilon_t\mathrm{E}\left(\epsilon_s^2|\mathcal{F}_t\right)\right)$ can be different from zero.

How to Construct Bootstrap Critical Values.

(a) In practice, we do not know the bootstrap critical value $\widetilde{z}^{*}_{T,\alpha/2}$. The standard approach is to construct B bootstrap statistics, say $\widetilde{t}^{*(j)}_{\beta_i,T}$, $j = 1, ..., B$ and obtain $\widetilde{z}^{*}_{T,B,\alpha/2}$ as the $(1-\alpha/2)$ percentile of the empirical distribution of the $(\widetilde{t}^{*(1)}_{\beta_i,T}, ..., \widetilde{t}^{*(B)}_{\beta_i,T})$. The problem is how to choose B large enough, in order to ensure that the inference based on $\widetilde{z}^{*}_{T,\alpha/2}$ and on $\widetilde{z}^{*}_{T,B,\alpha/2}$ lead to the same higher order improvements. The issue of the optimal selection the number of bootstrap replications B has been addressed by e.g. Davidson and MacKinnon (2000) and Andrews and Buchinski (2000).

(b) The construction of the bootstrap statistic requires the choice of the block length parameter l. An adaptive procedure for choosing l has been suggested by Hall, Horowitz and Jing (1995).

(c) The computation of the bootstrap estimator $\widehat{\beta}^{*}_T$ can be quite demanding, as it involves the solution of B nonlinear optimization problems. Davidson and

MacKinnon (1999) have suggested an alternative $k-$step estimator. Basically, one can set $\widehat{\beta}_T^{*(0)} = \widehat{\beta}_T$ and take k step towards $\widehat{\beta}_T^*$, via a Newtwon-Raphson algorithm for example. Andrews (2002, Theorem 1) has shown that inference based on $z^*_{T,k,\alpha/2}$, i.e. on the critical values based on $\widehat{\beta}_T^{*(k)}$ leads to the same order of refinements as inference based on $\widehat{\beta}_T^*$, for $k \geq 3$ or $k \geq 4$, depending whether we consider symmetrical or equally tailed tests.

Improved Refinements

The block-block bootstrap

As stated in Theorem 2, the block bootstrap provide refinements in the error in rejection probability up to order $T^{-\xi}$, with $\xi < 1/4$, while the *iid* bootstrap provide refinements of order $T^{-1/2}$. One of the reason is the *join points problem* mentioned above. Andrews (2004) suggests to construct block statistics, so that the same join problem occurs in both the bootstrap and the actual sample. In other words, the statistic is computed by deleting the πl observations immediately preceding the join points $l+1, 2l+1, ..., (b-1)l+1$, where as $T \to \infty$, $\pi \to 0$ and $\pi l \to \infty$. As the underlying sample is strong mixing, the $l(1-\pi)$-th and the $l+1$-th observations become independent as $\pi l \to \infty$. Given that, there is no longer need for the correction term and then we do no longer require $\gamma > \xi$. Thus, we can choose $\gamma < 1/4$, thus allowing for $\xi > 1/4$. Nevertheless, we still need to choose a large enough block length, to capture the dependence in the data.

The Markov Bootstrap

If the underlying generating process is Markov, or it can be well approximated by a Markov process, then one could rely on the Markov Bootstrap proposed by Horowitz (2003). Basically, sample observations are used to construct a kernel estimator of the conditional density. Then, bootstrap samples are drawn from the estimated conditional density. Under mild regularity conditions, the Markov bootstrap leads to refinements in the error in rejection probability of order $T^{1/2-\varepsilon}$, with ε arbitrarily small.