Single Equation Linear GMM: Hypothesis Testing

Eric Zivot

Winter 2013

Hypothesis Testing for Linear GMM Models

The main types of hypothesis tests are

- Overidentification restrictions
- Coefficient restrictions (linear and nonlinear)
- Subsets of orthogonality restrictions
- Instrument relevance.

Remark:

One should always first test the overidentifying restrictions before conducting the other tests. If the model specification is rejected, it does not make sense to do the remaining tests.

Specification Tests in Overidentified Models

An advantage of the GMM estimation in overidentified models is the ability to test the specification of the model

$$\begin{aligned} y_t &= \mathbf{z}_t'\boldsymbol{\delta}_0 + \varepsilon_t \\ E[\mathbf{g}_t] &= E[\mathbf{x}_t\varepsilon_t] = \mathbf{0},\ E[\mathbf{g}_t\mathbf{g}_t'] = E[\mathbf{x}_t\mathbf{x}_t'\varepsilon_t^2] = \mathbf{S} \end{aligned}$$

The J-*statistic*, introduced in Hansen (1982), refers to the value of the GMM objective function evaluated using *an efficient* GMM estimator:

$$J = J(\hat{\boldsymbol{\delta}}(\hat{\mathbf{S}}^{-1}), \hat{\mathbf{S}}^{-1}) = n\mathbf{g}_n(\hat{\boldsymbol{\delta}}(\hat{\mathbf{S}}^{-1}))'\hat{\mathbf{S}}^{-1}\mathbf{g}_n(\hat{\boldsymbol{\delta}}(\hat{\mathbf{S}}^{-1}))$$

$$\mathbf{g}_n(\boldsymbol{\delta}) = \mathbf{S}_{xy} - \mathbf{S}_{xz}\boldsymbol{\delta}$$

$$\hat{\boldsymbol{\delta}}(\hat{\mathbf{S}}^{-1}) = \text{any efficient GMM estimator},\ \hat{\mathbf{S}} \xrightarrow{p} \mathbf{S}$$

Recall, If $K = L$, then $J = 0$; if $K > L$, then $J > 0$.

Under regularity conditions (see Hayashi, 2000, Chap. 3) and if the moment conditions are valid, then as $n \longrightarrow \infty$

$$J \xrightarrow{d} \chi^2(K-L)$$

Remarks

1. In a well-specified overidentified model with valid moment conditions the J-statistic behaves like a chi-square random variable with degrees of freedom equal to the number of overidentifying restrictions.

2. If the model is misspecified and/or some of the moment conditions do not hold (e.g., $E[x_{it}\varepsilon_t] = E[x_{it}(y_t - \mathbf{z}_t'\boldsymbol{\delta}_0)] \neq 0$ for some i), then the J-statistic will be large relative to a chi-square random variable with $K-L$ degrees of freedom.

3. The J-statistic acts as an omnibus test statistic for model misspecification. A large J-statistic indicates a misspecified model.

4. Unfortunately, the J-statistic does not, by itself, give any information about how the model is misspecified.

The J-statistic and TSLS

When $E[\mathbf{x}_i\mathbf{x}_i'\varepsilon_i^2] = \mathbf{S} = \sigma^2\mathbf{\Sigma}_{xx}$, efficient GMM reduces to TSLS. The J-statistic then takes the form

$$\begin{aligned} & J(\hat{\boldsymbol{\delta}}_{\mathsf{TSLS}}, \hat{\sigma}^{-2}_{\mathsf{TSLS}}\mathbf{S}^{-1}_{xx}) \\ = \; & n\frac{(\mathbf{s}_{xy} - \mathbf{S}_{xz}\hat{\boldsymbol{\delta}}_{\mathsf{TSLS}})'\mathbf{S}^{-1}_{xx}(\mathbf{s}_{xy} - \mathbf{S}_{xz}\hat{\boldsymbol{\delta}}_{\mathsf{TSLS}})}{\hat{\sigma}^2_{\mathsf{TSLS}}} \end{aligned}$$

The TSLS J-statistic is also known as *Sargan's statistic* (see Sargan, 1958).

Remark: Most statistical software that computes TSLS does not report Sargan's statistic.

Asymptotic Distribution of Sample Moments and J-statistic

$$\begin{aligned} y_t &= \mathbf{z}_t'\boldsymbol{\delta}_0 + \varepsilon_t \\ E[\mathbf{g}_t] &= E[\mathbf{x}_t\varepsilon_t] = \mathbf{0},\ E[\mathbf{g}_t\mathbf{g}_t'] = E[\mathbf{x}_t\mathbf{x}_t'\varepsilon_t^2] = \underset{K\times K}{\mathbf{S}} \end{aligned}$$

Result 1: The normalized sample moment evaluated at $\boldsymbol{\delta}_0$ is asymptotically normally distributed

$$\sqrt{n}\mathbf{g}_n(\boldsymbol{\delta}_0) = \sqrt{n}\mathbf{S}_{x\varepsilon} = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\mathbf{x}_t\varepsilon_t \xrightarrow{d} N(\mathbf{0}, \mathbf{S})$$

This follows directly by the CLT for ergodic-stationary MDS.

Result 2: The J-statistic evaluated at $\boldsymbol{\delta}_0$ and $\hat{\mathbf{S}}^{-1}$ is asymptotically chi-square distributed with K degrees of freedom

$$J = J(\boldsymbol{\delta}_0, \hat{\mathbf{S}}^{-1}) = n\mathbf{g}_n(\boldsymbol{\delta}_0)'\hat{\mathbf{S}}^{-1}\mathbf{g}_n(\boldsymbol{\delta}_0) \xrightarrow{d} \chi^2(K)$$

provided $\hat{\mathbf{S}} \xrightarrow{p} \mathbf{S}$.

This follows directly from Result 1, Slutsky's theorem and the CMT:

$$\begin{aligned} n\mathbf{g}_n(\boldsymbol{\delta}_0)'\hat{\mathbf{S}}^{-1}\mathbf{g}_n(\boldsymbol{\delta}_0) &= \left(\sqrt{n}\hat{\mathbf{S}}^{-1/2}\mathbf{g}_n(\boldsymbol{\delta}_0)\right)'\left(\sqrt{n}\hat{\mathbf{S}}^{-1/2}\mathbf{g}_n(\boldsymbol{\delta}_0)\right) \\ \xrightarrow{d} N(\mathbf{0}, \mathbf{I}_k)'N(\mathbf{0}, \mathbf{I}_k) &\equiv \mathbf{z}'\mathbf{z} \sim \chi^2(K) \end{aligned}$$

where $\mathbf{z} \sim N(\mathbf{0}, \mathbf{I}_k)$ and $\hat{\mathbf{S}}^{-1} = \hat{\mathbf{S}}^{-1/2\prime}\hat{\mathbf{S}}^{-1/2}$.

Result 3: The J-statistic evaluated at $\boldsymbol{\delta}_0$ and $\hat{\mathbf{W}} \xrightarrow{p} \mathbf{W} \neq \mathbf{S}^{-1}$ *is not* asymptotically chi-square distributed with K degrees of freedom

$$J = J(\boldsymbol{\delta}_0, \hat{\mathbf{W}}) = n\mathbf{g}_n(\boldsymbol{\delta}_0)'\hat{\mathbf{W}}\mathbf{g}_n(\boldsymbol{\delta}_0) \overset{d}{\nrightarrow} \chi^2(K)$$

This follows directly from Result 1, Slutsky's theorem and the CMT:

$$\begin{gathered} n\mathbf{g}_n(\boldsymbol{\delta}_0)'\hat{\mathbf{W}}\mathbf{g}_n(\boldsymbol{\delta}_0) = \left(\sqrt{n}\hat{\mathbf{W}}^{1/2}\mathbf{g}_n(\boldsymbol{\delta}_0)\right)' \left(\sqrt{n}\hat{\mathbf{W}}^{1/2}\mathbf{g}_n(\boldsymbol{\delta}_0)\right) \\ \xrightarrow{d} \left(\mathbf{W}^{1/2}N(\mathbf{0},\mathbf{S})\right)' \left(\mathbf{W}^{1/2}N(\mathbf{0},\mathbf{S})\right) \equiv \left(\mathbf{W}^{1/2}\mathbf{S}^{1/2}\mathbf{z}\right)' (\mathbf{W}^{1/2}\mathbf{S}^{1/2}\mathbf{z}) \\ \neq \mathbf{z}'\mathbf{z} \sim \chi^2(K) \end{gathered}$$

where $\mathbf{z} \sim N(\mathbf{0}, \mathbf{I}_k)$ and $\mathbf{W} = \mathbf{W}^{1/2\prime}\mathbf{W}^{1/2}$.

Note: $\left(\mathbf{W}^{1/2}\mathbf{S}^{1/2}\mathbf{z}\right)' (\mathbf{W}^{1/2}\mathbf{S}^{1/2}\mathbf{z}) \neq \mathbf{z}'\mathbf{z} \sim \chi^2(K)$ unless $\mathbf{W}^{1/2} = \mathbf{S}^{-1/2}$

Q1: What is the distribution of normalized sample moment evaluated at the GMM estimates $\hat{\boldsymbol{\delta}}(\hat{\mathbf{W}})$ and $\hat{\boldsymbol{\delta}}(\hat{\mathbf{S}}^{-1})$?

Q2: What is the distribution of the J-statistic evaluated at the GMM estimate $\hat{\boldsymbol{\delta}}(\hat{\mathbf{S}}^{-1})$?

Intuition: Estimation of the $L \times 1$ vector $\boldsymbol{\delta}_0$ uses up L degrees of freedom. Therefore, the asymptotic normal distribution of $\mathbf{g}_n(\hat{\boldsymbol{\delta}}(\hat{\mathbf{W}}))$ and $\mathbf{g}_n(\hat{\boldsymbol{\delta}}(\hat{\mathbf{S}}^{-1}))$ will be normal but with a $K-L$ dimensional covariance matrix. The asymptotic distribution of the J-statistic evaluated at $\hat{\boldsymbol{\delta}}(\hat{\mathbf{S}}^{-1})$ will be chi-square with $K-L$ degrees of freedom.

Properties of Quadratic Forms in Symmetric Idempotent Matrices

Result: Let $\mathbf{Q}$ be an $n \times n$ symmetric idempotent (i.e. projection) matrix with $rank(\mathbf{Q}) = K < n$ such that $\mathbf{Q} = \mathbf{Q}'$ and $\mathbf{Q} \times \mathbf{Q} = \mathbf{Q}$, and let $\mathbf{z} \sim N(0, \mathbf{I}_n)$. Then

$$\mathbf{z}'\mathbf{Q}\mathbf{z} \sim \chi^2(K)$$

Sketch of proof.

Consider the spectral (eigenvalue) decomposition of $\mathbf{Q}$

$$\begin{aligned}
\mathbf{Q} &= \mathbf{P}\mathbf{\Lambda}\mathbf{P}', \\
\mathbf{P} &= \text{orthonormal matrix of eigenvectors; } \mathbf{P}' = \mathbf{P}^{-1} \\
\mathbf{\Lambda} &= \text{diagonal matrix of eigenvalues}
\end{aligned}$$

Now, because $\mathbf{Q}$ is a rank K projection matrix it has K eigenvalues equal to 1 and $n-K$ eigenvalues equal to 0. Hence, we can write

$$\boldsymbol{\Lambda} = \begin{pmatrix} \mathbf{I}_K & \\ & \mathbf{O}_{n-K} \end{pmatrix}$$

In addition, because $\mathbf{P}$ is orthonormal

$$\mathbf{Pz} \equiv \mathbf{z}$$

That is,

$$\begin{aligned} E[\mathbf{Pz}] &= \mathbf{P}E[\mathbf{z}] = \mathbf{0}, \\ var(\mathbf{Pz}) &= E[\mathbf{PzzP}'] = \mathbf{P}E[\mathbf{zz}']\mathbf{P}' = \mathbf{PP}^{-1} = \mathbf{I}_n \end{aligned}$$

Then

$$\mathbf{z}'\mathbf{Qz} = \mathbf{z}'\mathbf{P}\boldsymbol{\Lambda}\mathbf{P}'\mathbf{z} = \mathbf{z}'\boldsymbol{\Lambda}\mathbf{z} = \sum_{i=1}^{K} z_i^2 \sim \chi^2(K)$$

Proposition 1: The normalized sample moment evaluated at the GMM estimate $\hat{\boldsymbol{\delta}}(\hat{\mathbf{W}})$ is asymptotically normally distributed

$$\hat{\mathbf{W}}^{1/2}\sqrt{n}\mathbf{g}_n(\hat{\boldsymbol{\delta}}(\hat{\mathbf{W}})) = \hat{\mathbf{W}}^{1/2}\sqrt{n}(\mathbf{s}_{xy} - \mathbf{S}_{xz}\hat{\boldsymbol{\delta}}(\hat{\mathbf{W}}))$$
$$\xrightarrow{d} N(\mathbf{0}, \mathbf{NSN}')$$
$$\mathbf{N} = (\mathbf{I}_K - \mathbf{P}_F)\mathbf{W}^{1/2},\ \mathbf{P}_F = \mathbf{F}(\mathbf{F}'\mathbf{F})^{-1}\mathbf{F}', \mathbf{F} = \mathbf{W}^{1/2}\boldsymbol{\Sigma}_{xz}$$
$$\text{rank}(\mathbf{F}) = L,\ \text{rank}(\mathbf{N}) = K - L,\ \text{rank}(\mathbf{NSN}') = K - L$$

Corollary: The normalized sample moment evaluated at the efficient GMM estimate $\hat{\boldsymbol{\delta}}(\hat{\mathbf{S}}^{-1})$ is asymptotically normally distributed

$$\hat{\mathbf{S}}^{-1/2}\sqrt{n}\mathbf{g}_n(\hat{\boldsymbol{\delta}}(\hat{\mathbf{S}}^{-1})) \xrightarrow{d} N(\mathbf{0}, (\mathbf{I}_K - \mathbf{P}_F))$$

This follows directly from Proposition 1 by setting $\hat{\mathbf{W}} = \hat{\mathbf{S}}^{-1}$.

Proposition 2: The asymptotic distribution of the J-statistic evaluated at the efficient GMM estimate $\hat{\boldsymbol{\delta}}(\hat{\mathbf{S}}^{-1})$ is chi-square with $K-L$ degrees of freedom

$$\begin{aligned} J(\hat{\boldsymbol{\delta}}(\hat{\mathbf{S}}^{-1}), \hat{\mathbf{S}}^{-1}) &= n\mathbf{g}_n(\hat{\boldsymbol{\delta}}(\hat{\mathbf{S}}^{-1}))'\hat{\mathbf{S}}^{-1}\mathbf{g}_n(\hat{\boldsymbol{\delta}}(\hat{\mathbf{S}}^{-1})) \\ &\xrightarrow{d} \chi^2(K-L) \end{aligned}$$

This follows directly from the Corollary to Proposition 1:

$$\begin{aligned} &n\mathbf{g}_n(\hat{\boldsymbol{\delta}}(\hat{\mathbf{S}}^{-1}))'\hat{\mathbf{S}}^{-1}\mathbf{g}_n(\hat{\boldsymbol{\delta}}(\hat{\mathbf{S}}^{-1})) = \\ &\left(\hat{\mathbf{S}}^{-1/2}\sqrt{n}\mathbf{g}_n(\hat{\boldsymbol{\delta}}(\hat{\mathbf{S}}^{-1}))\right)'\left(\hat{\mathbf{S}}^{-1/2}\sqrt{n}\mathbf{g}_n(\hat{\boldsymbol{\delta}}(\hat{\mathbf{S}}^{-1}))\right) \\ &\xrightarrow{d} N(\mathbf{0}, (\mathbf{I}_K - \mathbf{P}_F))'N(\mathbf{0}, (\mathbf{I}_K - \mathbf{P}_F)) \equiv \mathbf{z}'(\mathbf{I}_K - \mathbf{P}_F)\mathbf{z} \sim \chi^2(K-L) \end{aligned}$$

where

$$\mathbf{z} \sim N(\mathbf{0}, \mathbf{I}_k),\ \text{rank}(\mathbf{I}_K - \mathbf{P}_F) = K - L$$

Sketch of Proof to Proposition 1.

First note that

$$\begin{aligned}\hat{\boldsymbol{\delta}}(\hat{\mathbf{W}}) &= (\mathbf{S}'_{xz}\hat{\mathbf{W}}\mathbf{S}_{xz})^{-1}\mathbf{S}'_{xz}\hat{\mathbf{W}}\mathbf{s}_{xy} \\ \mathbf{g}_n(\hat{\boldsymbol{\delta}}(\hat{\mathbf{W}})) &= \mathbf{s}_{xy} - \mathbf{S}_{xz}\hat{\boldsymbol{\delta}}(\hat{\mathbf{W}}) \\ &= \mathbf{s}_{xy} - \mathbf{S}_{xz}(\mathbf{S}'_{xz}\hat{\mathbf{W}}\mathbf{S}_{xz})^{-1}\mathbf{S}'_{xz}\hat{\mathbf{W}}\mathbf{s}_{xy}\end{aligned}$$

Define (e.g. Cholesky factorization)

$$\hat{\mathbf{W}} = \hat{\mathbf{W}}^{1/2\prime}\hat{\mathbf{W}}^{1/2}$$

Then

$$\begin{aligned}&\hat{\mathbf{W}}^{1/2}\mathbf{g}_n(\hat{\boldsymbol{\delta}}(\hat{\mathbf{W}})) \\ =\ &\hat{\mathbf{W}}^{1/2}\mathbf{s}_{xy} - \hat{\mathbf{W}}^{1/2}\mathbf{S}_{xz}(\mathbf{S}'_{xz}\hat{\mathbf{W}}\mathbf{S}_{xz})^{-1}\mathbf{S}'_{xz}\hat{\mathbf{W}}^{1/2\prime}\hat{\mathbf{W}}^{1/2}\mathbf{s}_{xy} \\ =\ &(\mathbf{I}_K - \hat{\mathbf{W}}^{1/2}\mathbf{S}_{xz}(\mathbf{S}'_{xz}\hat{\mathbf{W}}\mathbf{S}_{xz})^{-1}\mathbf{S}'_{xz}\hat{\mathbf{W}}^{1/2\prime})\hat{\mathbf{W}}^{1/2}\mathbf{s}_{xy}\end{aligned}$$

Define

$$\begin{aligned}
\hat{\mathbf{F}} &= \hat{\mathbf{W}}^{1/2}\mathbf{S}_{xz} \\
\mathbf{P}_{\hat{F}} &= \hat{\mathbf{F}}(\hat{\mathbf{F}}'\hat{\mathbf{F}})^{-1}\hat{\mathbf{F}}' \\
&= \hat{\mathbf{W}}^{1/2}\mathbf{S}_{xz}(\mathbf{S}_{xz}'\hat{\mathbf{W}}\mathbf{S}_{xz})^{-1}\mathbf{S}_{xz}'\hat{\mathbf{W}}^{1/2\prime} \\
\text{rank}(\mathbf{P}_{\hat{F}}) &= L \text{ as } n \to \infty
\end{aligned}$$

Then

$$\begin{aligned}
\hat{\mathbf{W}}^{1/2}\mathbf{g}_n(\hat{\boldsymbol{\delta}}(\hat{\mathbf{W}})) &= (\mathbf{I}_K - \mathbf{P}_{\hat{F}})\hat{\mathbf{W}}^{1/2}\mathbf{s}_{xy} \\
\text{rank}(\mathbf{I}_K - \mathbf{P}_{\hat{F}}) &= K - L \text{ as } n \to \infty
\end{aligned}$$

Now

$$\begin{aligned}
\mathbf{s}_{xy} &= \mathbf{S}_{xz}\boldsymbol{\delta}_0 + \mathbf{S}_{x\varepsilon} \\
\hat{\mathbf{W}}^{1/2}\mathbf{s}_{xy} &= \hat{\mathbf{W}}^{1/2}\mathbf{S}_{xz}\boldsymbol{\delta}_0 + \hat{\mathbf{W}}^{1/2}\mathbf{S}_{x\varepsilon} \\
&= \hat{\mathbf{F}}\boldsymbol{\delta}_0 + \hat{\mathbf{W}}^{1/2}\mathbf{S}_{x\varepsilon}
\end{aligned}$$

Therefore,

$$\begin{aligned}
\hat{\mathbf{W}}^{1/2}\mathbf{g}_n(\hat{\boldsymbol{\delta}}(\hat{\mathbf{W}})) &= (\mathbf{I}_K - \mathbf{P}_{\hat{F}})\hat{\mathbf{W}}^{1/2}\mathbf{s}_{xy} \\
&= (\mathbf{I}_K - \mathbf{P}_{\hat{F}})\left(\hat{\mathbf{F}}\boldsymbol{\delta}_0 + \hat{\mathbf{W}}^{1/2}\mathbf{S}_{x\varepsilon}\right) \\
&= (\mathbf{I}_K - \mathbf{P}_{\hat{F}})\hat{\mathbf{W}}^{1/2}\mathbf{S}_{x\varepsilon}
\end{aligned}$$

since $(\mathbf{I}_K - \mathbf{P}_{\hat{F}})\hat{\mathbf{F}} = \mathbf{0}$.

Consider the normalized sample moment

$$\hat{\mathbf{W}}^{1/2}\sqrt{n}\mathbf{g}_n(\hat{\boldsymbol{\delta}}(\hat{\mathbf{W}})) = (\mathbf{I}_K - \mathbf{P}_{\hat{F}})\hat{\mathbf{W}}^{1/2}\sqrt{n}\mathbf{S}_{x\varepsilon}$$

By the ergodic theorem and Slutsky's theorem

$$\begin{aligned}
\hat{\mathbf{F}} &= \hat{\mathbf{W}}^{1/2}\mathbf{S}_{xz} \xrightarrow{p} \mathbf{W}^{1/2}\boldsymbol{\Sigma}_{xz} = \mathbf{F} \\
& \mathbf{P}_{\hat{F}} \xrightarrow{p} \mathbf{P}_F \\
& (\mathbf{I}_K - \mathbf{P}_{\hat{F}})\hat{\mathbf{W}}^{1/2} \xrightarrow{p} (\mathbf{I}_K - \mathbf{P}_F)\mathbf{W}^{1/2}
\end{aligned}$$

By the CLT for ergodic-stationary MDS and the CMT

$$\sqrt{n}\mathbf{S}_{x\varepsilon} \xrightarrow{d} N(\mathbf{0}, \mathbf{S})$$

Therefore, by the CMT

$$\begin{aligned}
\hat{\mathbf{W}}^{1/2}\sqrt{n}\mathbf{g}_n(\hat{\boldsymbol{\delta}}(\hat{\mathbf{W}})) &= (\mathbf{I}_K - \mathbf{P}_{\hat{F}})\hat{\mathbf{W}}^{1/2}\sqrt{n}\mathbf{S}_{x\varepsilon} \\
&\xrightarrow{d} (\mathbf{I}_K - \mathbf{P}_F)\mathbf{W}^{1/2}N(\mathbf{0}, \mathbf{S}) \equiv N(\mathbf{0}, \mathbf{NSN}')
\end{aligned}$$

where $\mathbf{N} = (\mathbf{I}_K - \mathbf{P}_F)\mathbf{W}^{1/2}$. Note

$$\begin{aligned} \text{rank}(\mathbf{N}) &= K - L \\ \text{rank}(\mathbf{NSN}') &= K - L \end{aligned}$$

Testing Restrictions on Coefficients: **Wald Statistics**

Wald-type statistics are based on the asymptotic normality of the GMM estimator $\hat{\boldsymbol{\delta}}(\hat{\mathbf{W}})$ for an arbitrary weight matrix $\hat{\mathbf{W}}$. Simple tests on individual coefficients of the form

$$H_0 : \delta_k = \delta_k^0$$

may be conducted using the asymptotic t-ratio

$$t_k = \frac{\hat{\delta}_k(\hat{\mathbf{W}}) - \delta_k^0}{\widehat{\mathsf{SE}}(\hat{\delta}_k(\hat{\mathbf{W}}))}$$

where $\widehat{\mathsf{SE}}(\hat{\delta}_k(\hat{\mathbf{W}}))$ is the square root of the kth diagonal element of

$$\frac{1}{n}\widehat{\mathsf{avar}}(\hat{\boldsymbol{\delta}}(\hat{\mathbf{W}})) = \frac{1}{n}(\mathbf{S}_{xz}'\hat{\mathbf{W}}\mathbf{S}_{xz})^{-1}\mathbf{S}_{xz}'\hat{\mathbf{W}}\hat{\mathbf{S}}\hat{\mathbf{W}}\mathbf{S}_{xz}(\mathbf{S}_{xz}'\hat{\mathbf{W}}\mathbf{S}_{xz})^{-1}$$

Under the null hypothesis , the t-ratio has an asymptotic standard normal distribution.

Linear hypotheses of the form

$$
\begin{aligned}
H_0 &: \mathbf{R}\boldsymbol{\delta}_0 = \mathbf{r} \\
\mathbf{R} &= \text{fixed } q \times L \text{ matrix of rank } q \\
\mathbf{r} &= \text{fixed } q \times 1 \text{ vector}
\end{aligned}
$$

may be tested using the Wald statistic

$$
\begin{aligned}
\text{Wald} &= n(\mathbf{R}\hat{\boldsymbol{\delta}}(\hat{\mathbf{W}}) - \mathbf{r})' \left[\mathbf{R} \cdot \widehat{\text{avar}}(\hat{\boldsymbol{\delta}}(\hat{\mathbf{W}})) \cdot \mathbf{R}'\right]^{-1} \\
&\quad \times (\mathbf{R}\hat{\boldsymbol{\delta}}(\hat{\mathbf{W}}) - \mathbf{r}) \\
\widehat{\text{avar}}(\hat{\boldsymbol{\delta}}(\hat{\mathbf{W}})) &= (\mathbf{S}_{xz}'\hat{\mathbf{W}}\mathbf{S}_{xz})^{-1}\mathbf{S}_{xz}'\hat{\mathbf{W}}\hat{\mathbf{S}}\hat{\mathbf{W}}\mathbf{S}_{xz}(\mathbf{S}_{xz}'\hat{\mathbf{W}}\mathbf{S}_{xz})^{-1}
\end{aligned}
$$

Under the null, the Wald statistic has a limiting chi-square distribution with q degrees of freedom.

Remark

The Wald statistic is valid for any consistent and asymptotically normal GMM estimator $\hat{\boldsymbol{\delta}}(\hat{\mathbf{W}})$ based on an arbitrary symmetric and positive definite weight matrix $\hat{\mathbf{W}} \xrightarrow{p} \mathbf{W}$. Usually, Wald statistics are computed using $\hat{\mathbf{W}} = \hat{\mathbf{S}}^{-1}$.

Sketch of Proof

Under $H_0 : \mathbf{R}\boldsymbol{\delta}_0 = \mathbf{r}$

$$\sqrt{n}(\mathbf{R}\hat{\boldsymbol{\delta}}(\hat{\mathbf{W}}) - \mathbf{r}) = \sqrt{\mathbf{n}}(\mathbf{R}\hat{\boldsymbol{\delta}}(\hat{\mathbf{W}}) - \mathbf{R}\boldsymbol{\delta}_{\mathbf{0}}) = \mathbf{R}\sqrt{n}(\hat{\boldsymbol{\delta}}(\hat{\mathbf{W}}) - \boldsymbol{\delta}_0)$$

By the asymptotic normality of $\hat{\boldsymbol{\delta}}(\hat{\mathbf{W}})$ and Slutsky's theorem

$$\begin{aligned} & \mathbf{R}\sqrt{n}(\hat{\boldsymbol{\delta}}(\hat{\mathbf{W}}) - \boldsymbol{\delta}_0) \xrightarrow{d} \mathbf{R}\cdot N(\mathbf{0}, \text{avar}(\hat{\boldsymbol{\delta}}(\hat{\mathbf{W}}))) \\ \equiv \; & N(\mathbf{0}, \mathbf{R}\cdot\text{avar}(\hat{\boldsymbol{\delta}}(\hat{\mathbf{W}})) \cdot \mathbf{R}') \end{aligned}$$

Therefore, by the CMT

$$\begin{aligned}
\mathsf{Wald} &= \sqrt{n}(\mathbf{R}\hat{\boldsymbol{\delta}}(\hat{\mathbf{W}}) - \mathbf{r})' \left[\mathbf{R}\cdot\widehat{\mathsf{avar}}(\hat{\boldsymbol{\delta}}(\hat{\mathbf{W}})) \cdot \mathbf{R}'\right]^{-1} \\
&\times \sqrt{n}(\mathbf{R}\hat{\boldsymbol{\delta}}(\hat{\mathbf{W}}) - \mathbf{r}) \\
&\xrightarrow{d} N(\mathbf{0}, \mathbf{R}\cdot\mathsf{avar}(\hat{\boldsymbol{\delta}}(\hat{\mathbf{W}})) \cdot \mathbf{R}')' \left[\mathbf{R}\cdot\widehat{\mathsf{avar}}(\hat{\boldsymbol{\delta}}(\hat{\mathbf{W}})) \cdot \mathbf{R}'\right]^{-1} \\
&\times N(\mathbf{0}, \mathbf{R}\cdot\mathsf{avar}(\hat{\boldsymbol{\delta}}(\hat{\mathbf{W}})) \cdot \mathbf{R}') \\
&\sim \chi^2(q)
\end{aligned}$$

Nonlinear hypotheses of the form

$$H_0: \underset{q\times 1}{\mathbf{a}(\boldsymbol{\delta}_0)} = \mathbf{0}$$

$$\text{rank}\left(\mathbf{A}(\boldsymbol{\delta}_0)\right) = q,\ \underset{q\times L}{\mathbf{A}(\boldsymbol{\delta}_0)} = \frac{\partial \mathbf{a}(\boldsymbol{\delta}_0)}{\partial \boldsymbol{\delta}'}$$

may be tested using the Wald statistic

$$\begin{aligned}\text{Wald} &= n\mathbf{a}(\hat{\boldsymbol{\delta}}(\hat{\mathbf{W}}))'\left[\mathbf{A}(\hat{\boldsymbol{\delta}}(\hat{\mathbf{W}}))\widehat{\text{avar}}(\hat{\boldsymbol{\delta}}(\hat{\mathbf{W}}))\mathbf{A}(\hat{\boldsymbol{\delta}}(\hat{\mathbf{W}}))'\right]^{-1} \\ &\quad \times \mathbf{a}(\hat{\boldsymbol{\delta}}(\hat{\mathbf{W}}))\end{aligned}$$

Under the null, the Wald statistic has a limiting chi-square distribution with q degrees of freedom.

Sketch of Proof

The proof follows from the asymptotic normality of $\hat{\boldsymbol{\delta}}(\hat{\mathbf{W}})$ and the delta method:

$$\sqrt{n}(\hat{\boldsymbol{\delta}}(\hat{\mathbf{W}}) - \boldsymbol{\delta}_0) \xrightarrow{d} N(\mathbf{0}, \text{avar}(\hat{\boldsymbol{\delta}}(\hat{\mathbf{W}})))$$
$$\sqrt{n}(\mathbf{a}(\hat{\boldsymbol{\delta}}(\hat{\mathbf{W}})) - \mathbf{a}(\boldsymbol{\delta}_0)) = \sqrt{n}\mathbf{a}(\hat{\boldsymbol{\delta}}(\hat{\mathbf{W}})$$
$$\xrightarrow{d} N\left(\mathbf{0}, \mathbf{A}(\boldsymbol{\delta}_0)\text{avar}(\hat{\boldsymbol{\delta}}(\hat{\mathbf{W}}))\mathbf{A}(\boldsymbol{\delta}_0)'\right)$$

since $\mathbf{a}(\boldsymbol{\delta}_0) = 0$ under H_0. Therefore, by Slutsky's theorem and the CMT

$$\text{Wald} = \sqrt{n}\mathbf{a}(\hat{\boldsymbol{\delta}}(\hat{\mathbf{W}}))' \left[\mathbf{A}(\hat{\boldsymbol{\delta}}(\hat{\mathbf{W}}))\widehat{\text{avar}}(\hat{\boldsymbol{\delta}}(\hat{\mathbf{W}}))\mathbf{A}(\hat{\boldsymbol{\delta}}(\hat{\mathbf{W}}))'\right]^{-1}$$
$$\times \sqrt{n}\mathbf{a}(\hat{\boldsymbol{\delta}}(\hat{\mathbf{W}}))$$
$$\xrightarrow{d} \chi^2(q)$$

Testing Restrictions on Coefficients: GMM LR-Type Statistics

Motivation: F-statistic in linear regression

Consider the linear regression model

$$y_t = \underset{(1\times K)}{\mathbf{x}_t'}\underset{(K\times 1)}{\boldsymbol{\beta}} + \varepsilon_t,\ t = 1,\ldots,n$$

and consider testing the simple hypothesis

$$H_0 : \beta = \beta_0$$

A standard test statistic is the F-statistic

$$\begin{aligned} F &= \frac{\left(RSS(\boldsymbol{\beta}_0) - RSS(\hat{\boldsymbol{\beta}}_{OLS})\right)/K}{RSS(\hat{\boldsymbol{\beta}}_{OLS})/(n-K)} \\ &= \frac{\left(RSS(\boldsymbol{\beta}_0) - RSS(\hat{\boldsymbol{\beta}}_{OLS})\right)/K}{\hat{\sigma}^2} \\ RSS(\boldsymbol{\beta}) &= (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \\ &= \text{OLS objective function} \end{aligned}$$

If the regressors are fixed and the errors are normally distributed, then $F \sim F_{K,n-K}$

The F-statistic may be re-written as

$$K \cdot F = \frac{RSS(\boldsymbol{\beta}_0)}{\hat{\sigma}^2} - \frac{RSS(\hat{\boldsymbol{\beta}}_{OLS})}{\hat{\sigma}^2}$$

which is the difference between the scaled restricted OLS objective function and the scaled unrestricted OLS objective function.

Under more general conditions on the regressors and the errors,

$$K \cdot F \xrightarrow{d} \chi^2(K)$$

Note: It can be shown that $K \cdot F$ is numerically identical to the usual Wald statistic for testing $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$. (e.g. typical first year econometrics calculation)

Linear and nonlinear restrictions of the form

$$H_0 : \underset{q\times 1}{\mathbf{R}\boldsymbol{\delta}_0} = \mathbf{r}$$
$$H_0 : \underset{q\times 1}{\mathbf{a}(\boldsymbol{\delta}_0)} = \mathbf{0}$$

can be tested using a statistic based on the difference between the GMM objective functions (J-statistics) evaluated under the restricted and unrestricted models.

In efficient GMM estimation, the unrestricted objective function is

$$J(\hat{\boldsymbol{\delta}}(\hat{\mathrm{S}}^{-1}), \hat{\mathrm{S}}^{-1})$$

where $\hat{\boldsymbol{\delta}}(\hat{\mathrm{S}}^{-1})$ is computed without restrictions.

The restricted efficient GMM estimator solves

$$\tilde{\delta}_{\mathsf{R}}(\hat{\mathrm{S}}^{-1}) = \arg\min_{\delta}\ J(\boldsymbol{\delta}, \hat{\mathrm{S}}^{-1}) \text{ subject to } H_0$$

The GMM LR-type statistic is the difference between the restricted and unrestricted J-statistics:

$$\mathsf{LR}_{\mathsf{GMM}} = J(\tilde{\boldsymbol{\delta}}_{\mathsf{R}}(\hat{\mathbf{S}}^{-1}), \hat{\mathbf{S}}^{-1}) - J(\hat{\boldsymbol{\delta}}(\hat{\mathbf{S}}^{-1}), \hat{\mathbf{S}}^{-1})$$

Under the null hypotheses it can be shown (homework assignment)

$$\mathsf{LR}_{\mathsf{GMM}} \xrightarrow{d} \chi^2(q)$$

Remarks

1. As $n \to \infty$, it can be shown that Wald $-$ $\mathsf{LR}_{\mathsf{GMM}} \xrightarrow{p} \mathbf{0}$, although the two statistics may differ in finite samples for nonlinear hypotheses.

2. For linear restrictions, Wald and $\mathsf{LR}_{\mathsf{GMM}}$ are numerically equivalent provided that the same value of $\hat{\mathbf{S}}$ is used to compute the restricted and unrestricted efficient GMM estimators. Typically $\hat{\mathbf{S}}$ computed under the unrestricted model is used in constructing $\mathsf{LR}_{\mathsf{GMM}}$. In this case, when the restricted efficient GMM estimator is computed the weight matrix $\hat{\mathbf{S}}_{\mathsf{UR}}^{-1}$ is held fixed during the estimation (no iteration is performed on the weight matrix).

3. If $\mathsf{LR}_{\mathsf{GMM}}$ is computed using two different consistent estimates of $\mathbf{S}$, say $\hat{\mathbf{S}}$ and $\tilde{\mathbf{S}}$, then it is not guaranteed to be positive in finite samples but is asymptotically valid.

4. The LR_{GMM} statistic has the advantage over the Wald statistic for nonlinear hypotheses in that it is invariant to how the nonlinear restrictions are represented. Additionally, Monte Carlo studies have shown that LR_{GMM} often performs better than Wald in finite samples. In particular, Wald tends to over reject the null hypothesis when it is true.

Testing Subsets of Orthogonality Conditions

Consider the linear GMM model

$$y_t = \mathbf{z}_t'\boldsymbol{\delta}_0 + \varepsilon_t$$

with instruments $\mathbf{x}_t = (\mathbf{x}_{1t}', \mathbf{x}_{2t}')'$ such that

$$\begin{aligned} \underset{K\times 1}{\mathbf{x}_t} &= \begin{pmatrix} \mathbf{x}_{1t} \\ \mathbf{x}_{2t} \end{pmatrix} \begin{matrix} K_1 \times 1 \\ K_2 \times 1 \end{matrix} \\ K_1 &\geq L \text{ and } K_1 + K_2 = K \end{aligned}$$

The instruments $\mathbf{x}_{1t}$ are assumed to be valid (i.e., $E[\mathbf{x}_{1t}\varepsilon_t] = 0$), whereas the instruments $\mathbf{x}_{2t}$ are suspected not to be valid (i.e., $E[\mathbf{x}_{2t}\varepsilon_t] \neq 0$). That is, the hypotheses to be tested are

$$\begin{aligned} H_0 &: E[\mathbf{x}_t\varepsilon_t] = \mathbf{0} \\ H_1 &: E[\mathbf{x}_{2t}\varepsilon_t] \neq \mathbf{0} \text{ given } E[\mathbf{x}_{1t}\varepsilon_t] = \mathbf{0} \end{aligned}$$

A procedure to test for the validity of $\mathbf{x}_{2t}$ due to Newey (1985), and Eichenbaum, Hansen, and Singleton (1988) is as follows.

First, estimate the model by efficient GMM using the full set of instruments $\mathbf{x}_t$, giving

$$\hat{\boldsymbol{\delta}}(\hat{\mathbf{S}}_{\mathsf{Full}}^{-1}) = (\mathbf{S}_{xz}'\hat{\mathbf{S}}_{\mathsf{Full}}^{-1}\mathbf{S}_{xz})^{-1}\mathbf{S}_{xz}'\hat{\mathbf{S}}_{\mathsf{Full}}^{-1}\mathbf{S}_{xy}$$

where

$$\hat{\mathbf{S}}_{\mathsf{Full}} = \begin{bmatrix} \hat{\mathbf{S}}_{11,\mathsf{Full}} & \hat{\mathbf{S}}_{12,\mathsf{Full}} \\ \hat{\mathbf{S}}_{21,\mathsf{Full}} & \hat{\mathbf{S}}_{22,\mathsf{Full}} \end{bmatrix}$$

$$\hat{\mathbf{S}}_{11,\mathsf{Full}} \text{ is } K_1 \times K_1$$

Second, estimate the model by efficient GMM using only the instruments $\mathbf{x}_{1t}$ (which are valid under H_1) and using the weight matrix $\hat{\mathbf{S}}_{11,\mathsf{Full}}^{-1}$ giving

$$\tilde{\boldsymbol{\delta}}(\hat{\mathbf{S}}_{11,\mathsf{Full}}^{-1}) = (\mathbf{S}_{x_1z}'\hat{\mathbf{S}}_{11,\mathsf{Full}}^{-1}\mathbf{S}_{x_1z})^{-1}\mathbf{S}_{x_1z}'\hat{\mathbf{S}}_{11,\mathsf{Full}}^{-1}\mathbf{S}_{x_1y}$$

Third, form the statistic

$$C = J(\hat{\boldsymbol{\delta}}(\hat{\mathbf{S}}^{-1}_{\mathsf{Full}}), \hat{\mathbf{S}}^{-1}_{\mathsf{Full}}) - J(\tilde{\boldsymbol{\delta}}(\hat{\mathbf{S}}^{-1}_{11,\mathsf{Full}}), \hat{\mathbf{S}}^{-1}_{11,\mathsf{Full}})$$

Under the null hypothesis that $E[\mathbf{x}_t \varepsilon_t] = \mathbf{0}$, the statistic C has a limiting chi-square distribution with $K - K_1$ degrees of freedom.

Note: The use of $\hat{\mathbf{S}}^{-1}_{11,\mathsf{Full}}$ guarantees that the C statistic is non-negative.

Testing Instrument Relevance

In order to obtain consistent GMM estimates the instruments $\mathbf{x}_t$ must satisfy

$$\begin{aligned} E[\mathbf{x}_t \varepsilon_t] &= 0 \text{ (orthogonality or validity)} \\ \text{rank}(E[\mathbf{x}_t \mathbf{z}_t']) &= L \text{ (relevancy)} \end{aligned}$$

For the rank condition to hold, the endogenous variables $\mathbf{z}_t$ must be correlated with the instruments.

To see this, consider the simple GMM regression involving a single endogenous variable and a single instrument (all variables demeaned)

$$\begin{aligned} y_t &= z_t \delta + \varepsilon_t \\ E[x_t \varepsilon_t] &= 0 \end{aligned}$$

The rank condition is $\text{rank}(\mathbf{\Sigma}_{zx}) = 1$, which implies that $\mathbf{\Sigma}_{zx} = \text{cov}(x, z) \neq 0$.

Remarks

1. If there are K instruments $x_{1t}, \ldots, x_{Kt}$ but only one endogenous variable z_t, then the rank condition holds as long as cov$(x_k, z) \neq 0$ *for some* k.

2. If cov$(x_k, z) \approx 0$ for all k, then the instruments are called *weak*.

3. Testing instrument relevance is important in practice because recent research (e.g., Nelson and Startz (1989), Dufour (1997), Staiger and Stock (1997), Zivot, Startz, Nelson (1998), Stock and Wright (2000)) has shown that standard GMM procedures for estimation and inference can be highly misleading if instruments are weak.

4. Stock, Wright, and Yogo (2002) and Kleibergen and Mavorides (2009) gave nice surveys of the issues associated with using GMM in the presence of weak instruments and discussed the nonstandard inference procedures that should be used.

Simple procedure for testing instrument relevance

Consider the general linear GMM regression

$$
\begin{aligned}
y_t &= \mathbf{z}_t' \boldsymbol{\delta} + \varepsilon_t \\
&= \underset{(1\times L_1)}{\mathbf{z}_{1t}'} \underset{(L_1\times 1)}{\boldsymbol{\delta}_1} + \underset{(1\times K_1)}{\mathbf{x}_{1t}'} \underset{(K_1\times 1)}{\boldsymbol{\delta}_2} + \varepsilon_t \\
L &= L_1 + K_1
\end{aligned}
$$

such that

$$
\begin{aligned}
E[\mathbf{z}_{1t}\varepsilon_t] &\neq 0 \text{ (L_1 endogenous regressors)} \\
E[\mathbf{x}_{1t}\varepsilon_t] &= 0 \text{ (K_1 included exogenous regressors)}
\end{aligned}
$$

Assume there are K_2 additional instruments (excluded exogenous variables) $\mathbf{x}_{2t}$ satisfying

$$
E[\mathbf{x}_{2t}\varepsilon_t] = 0
$$

What is important for the rank condition are the correlations between the endogenous variables in $\mathbf{z}_{1t}$ and the instruments in $\mathbf{x}_{2t}$.

To measure the correlations between the elements of $\mathbf{z}_{1t}$ and $\mathbf{x}_{2t}$ and to test for instrument relevance, estimate by least squares the so-called first-stage regression

$$z_{1lt} = \mathbf{x}_{1t}'\boldsymbol{\pi}_{1l} + \mathbf{x}_{2t}'\boldsymbol{\pi}_{2l} + v_{lt}, \;\; l = 1, \ldots, L_1$$

for each endogenous variable in $\mathbf{z}_{1t}$.

The t-ratios on the variables in $\mathbf{x}_{2t}$ can be used to assess the strength of the correlation between z_{1lt} and the variables in $\mathbf{x}_{2t}$. The F-statistic for testing $\boldsymbol{\pi}_{2l} = \mathbf{0}$ can be used to assess the joint relevance of $\mathbf{x}_{2t}$ for z_{1lt}.