Athens University of Economics and Business
Department of Economics

Postgraduate Program - Master's in Economic Theory
Course: Econometrics II
Prof: Stelios Arvanitis
TA: Alecos Papadopoulos

Semester: Spring 2017-2018

May 19, 2018 (corrected)

Generalized Method of Moments (GMM) - why use it?

This document is to be studied after, and as a complement to,
"Tutorial 5: A Linear Model with Instrumental Variables in a Time Series Framework."
where one will find the proper technical treatment of GMM.
Consult also Hayashi (2000), ch. 4 and 6.

Why use the GMM estimator? What is its fundamental value-added that justifies its existence and application? The issue matters because among other things the GMM estimator is considered one of the most important novel contributions of Econometrics to the science of Statistics (and it is what essentially got an Economics Nobel to its creator, Lars Peter Hansen).

We will show that the *unique* contribution of the GMM estimator is ... the weighting matrix: the proof that the weighting matrix is not the identity matrix, and the determination of an estimable optimal weighting matrix. It is in this aspect that GMM surpasses the traditional Method of Moments (MM) estimator: efficiency and lower variance.

A. Estimating an overidentified model using Method-of-Moments

A usual framework in which GMM is introduced is a model with regressor endogeneity and more instrumental variables than regressors. This educational approach ended up being misleading, because we are left with the impression that "when the system is overidentified, we *have* to use GMM". This is not correct, and we will show it by first obtaining the "traditional" Method-of-Moments estimator for an over-identified model. In matrix notation, assume that the model is

$$\mathrm{y} = \mathrm{X}\beta + \mathrm{u}, \quad E(\mathrm{u}) = \mathbf{0}, \quad E(\mathrm{X}'\mathrm{u}) \neq \mathbf{0}$$

where X is an $n \times L$ matrix, and assume that the Instruments matrix is Z of dimension $n \times K,\ K > L$, including any deterministic and exogenous stochastic regressors, as well as the instruments for the endogenous regressors. The additional assumption here is $E(\mathrm{Z}'\mathrm{u}) = \mathbf{0}$.

In traditional Method of Moments fashion, let's use these orthogonality conditions to obtain an estimator. Applying the Analogy principle, we have

$$\mathrm{Z}'\mathrm{u} = \mathbf{0} \Rightarrow \mathrm{Z}'(\mathrm{y} - \mathrm{X}\beta) = \mathbf{0} \Rightarrow \mathrm{Z}'\mathrm{X}\beta = \mathrm{Z}'\mathrm{y}$$

The matrix $\mathrm{Z}'\mathrm{X}$ is not square, and cannot be inverted. Writing $\mathrm{S}_{XZ} = \mathrm{X}'\mathrm{Z},\ \ \mathrm{S}_{ZX} = \mathrm{Z}'\mathrm{X},\ \mathrm{S}'_{ZX} = \mathrm{S}_{XZ}$ we have

$$\mathrm{S}_{ZX}\beta = \mathrm{Z}'\mathrm{y} \Rightarrow \mathrm{S}_{XZ}\mathrm{S}_{ZX}\beta = \mathrm{S}_{XZ}\mathrm{Z}'\mathrm{y} \Rightarrow \hat{\beta}_{MM} = \left(\mathrm{S}_{XZ}\mathrm{S}_{ZX}\right)^{-1}\mathrm{S}_{XZ}\mathrm{Z}'\mathrm{y}$$

This is the MM estimator that uses all the available instruments and does not discard information embodied in the orthogonality conditions. By analyzing y we get

$$\hat{\beta}_{MM} - \beta = \left(\mathrm{S}_{XZ}\mathrm{S}_{ZX}\right)^{-1}\mathrm{S}_{XZ}\mathrm{Z}'\mathrm{u}$$

and under usual regularity assumptions together with $E(\mathrm{Z}'\mathrm{u}) = \mathbf{0}$ we get consistency of $\hat{\beta}_{MM}$.

Turning to the limiting distribution of the MM estimator we look at

$$\sqrt{n}\left(\hat{\beta}_{MM} - \beta\right) = \left(\tfrac{1}{n}\mathrm{S}_{XZ}\,\tfrac{1}{n}\mathrm{S}_{ZX}\right)^{-1}\left(\tfrac{1}{n}\mathrm{S}_{XZ}\right)\left(\tfrac{1}{\sqrt{n}}\mathrm{Z}'\mathrm{u}\right)$$

Under the assumption $E\left(\mathrm{Z'u}\right)=\mathbf{0}$ the *finite-sample* variance of the above statistic is by construction,

$$\mathrm{Var}\left[\sqrt{n}\left(\hat{\beta}_{MM}-\beta\right)\right]=E\left[\left(\tfrac{1}{n}\mathrm{S}_{XZ}\,\tfrac{1}{n}\mathrm{S}_{ZX}\right)^{-1}\left(\tfrac{1}{n}\mathrm{S}_{XZ}\right)\left(\tfrac{1}{n}\mathrm{Z'uu'Z}\right)\left(\tfrac{1}{n}\mathrm{S}_{ZX}\right)\left(\tfrac{1}{n}\mathrm{S}_{XZ}\,\tfrac{1}{n}\mathrm{S}_{ZX}\right)^{-1}\right]$$

Under regularity conditions

$$\tfrac{1}{n}\mathrm{S}_{XZ}\xrightarrow{p}E\left(\mathbf{x}_t\mathbf{z}_t'\right)\equiv\Sigma_{XZ},\quad \tfrac{1}{n}\mathrm{S}_{ZX}\xrightarrow{p}E\left(\mathbf{z}_t\mathbf{x}_t'\right)\equiv\Sigma_{ZX}$$

Assuming that $\left\{\mathbf{z}_t u_t\right\}$ **is a vector martingale difference** process, we can get our CLT. As $n\rightarrow\infty$ we have

$$\sqrt{n}\left(\hat{\beta}_{MM}-\beta\right)\xrightarrow{d}N\left(0,\mathrm{AV}_{MM}\right),\quad \mathrm{AV}_{MM}=\left(\Sigma_{XZ}\Sigma_{ZX}\right)^{-1}\Sigma_{XZ}\left(\mathrm{plim}\tfrac{1}{n}\mathrm{Z'uu'Z}\right)\Sigma_{ZX}\left(\Sigma_{XZ}\Sigma_{ZX}\right)^{-1}$$

We still need to obtain an estimable expression for $\left(\mathrm{plim}\tfrac{1}{n}\mathrm{Z'uu'Z}\right)$. It is tedious, but instructive, to carry out the matrix multiplications here, to see what kind of additional assumptions, if any, we need to impose. Write Z' as a vector-bloc matrix,

$\mathrm{Z}'=\begin{bmatrix}\mathbf{z}_1'\\ \vdots\\ \mathbf{z}_K'\end{bmatrix}$ where each row-vector contains the whole series of the realizations of a regressor. Then

$$\mathrm{Z'uu'Z}=\begin{bmatrix}\mathbf{z}_1'\\ \vdots\\ \mathbf{z}_K'\end{bmatrix}\mathrm{uu}'\left[\mathbf{z}_1\quad\cdots\quad\mathbf{z}_K\right]=\begin{bmatrix}\mathbf{z}_1'\mathrm{uu}'\mathbf{z}_1 & \mathbf{z}_1'\mathrm{uu}'\mathbf{z}_2 & \cdots & \mathbf{z}_1'\mathrm{uu}'\mathbf{z}_K\\ \mathbf{z}_2'\mathrm{uu}'\mathbf{z}_1 & \mathbf{z}_2'\mathrm{uu}'\mathbf{z}_2 & \cdots & \mathbf{z}_2'\mathrm{uu}'\mathbf{z}_K\\ \vdots & \cdots & \ddots & \vdots\\ \mathbf{z}_K'\mathrm{uu}'\mathbf{z}_1 & \cdots & \mathbf{z}_K'\mathrm{uu}'\mathbf{z}_{K-1} & \mathbf{z}_K'\mathrm{uu}'\mathbf{z}_K\end{bmatrix}$$

Let's analyze a main-diagonal element:

$$[1,1]:\ \mathbf{z}_1'\mathrm{uu}'\mathbf{z}_1 = \begin{bmatrix} z_{11} & \cdots & z_{1n} \end{bmatrix} \begin{bmatrix} u_1^2 & u_1 u_2 & \cdots & u_1 u_n \\ u_2 u_1 & u_2^2 & \cdots & u_2 u_n \\ \vdots & \cdots & \ddots & \vdots \\ u_n u_1 & u_n u_2 & u_n u_{n-1} & u_n^2 \end{bmatrix} \begin{bmatrix} z_{11} \\ \vdots \\ z_{1n} \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{t=1}^{n} z_{1t} u_t u_1 & \sum_{t=1}^{n} z_{1t} u_t u_2 & \cdots & \sum_{t=1}^{n} z_{1t} u_t u_n \end{bmatrix} \begin{bmatrix} z_{11} \\ \vdots \\ z_{1n} \end{bmatrix}$$

$$\Rightarrow \mathbf{z}_1'\mathrm{uu}'\mathbf{z}_1 = z_{11} \sum_{t=1}^{n} z_{1t} u_t u_1 + z_{12} \sum_{t=1}^{n} z_{1t} u_t u_2 + \ldots + z_{1n} \sum_{t=1}^{n} z_{1t} u_t u_n$$

$$\begin{aligned} &= z_{11}^2 u_1^2 + z_{11} z_{12} u_2 u_1 + \ldots + z_{11} z_{1n} u_n u_1 \\ &+ z_{12} z_{11} u_1 u_2 + z_{12}^2 u_2^2 + \ldots + z_{12} z_{1n} u_n u_2 \\ &+ \ldots \\ &+ z_{1n} z_{11} u_1 u_n + z_{1n} z_{12} u_2 u_n + \ldots + z_{1n}^2 u_2^2 \end{aligned}$$

$$\Rightarrow \mathbf{z}_1'\mathrm{uu}'\mathbf{z}_1 = \sum_{t=1}^{n} z_{1t}^2 u_t^2 + \sum \sum_{t \neq \ell}^{n} z_{1t} z_{1\ell} u_t u_\ell$$

The double sum contains $n(n-1)$ terms. All elements of the matrix are divided by *n*, so

$$\frac{1}{n}\mathbf{z}_1'\mathrm{uu}'\mathbf{z}_1 = \frac{1}{n}\sum_{t=1}^{n} z_{1t}^2 u_t^2 + \frac{1}{n}\sum \sum_{t \neq \ell}^{n} z_{1t} z_{1\ell} u_t u_\ell = \frac{1}{n}\sum_{t=1}^{n} z_{1t}^2 u_t^2 + (n-1)\frac{1}{n(n-1)}\sum \sum_{t \neq \ell}^{n} z_{1t} z_{1\ell} u_t u_\ell$$

With an ergodic-stationary sample, letting n go to infinity we will get

$$\text{plim}\frac{1}{n}\mathbf{z}_1'\mathrm{uu}'\mathbf{z}_1 = E\left(z_{1t}^2 u_t^2\right) + \lim(n-1)E\left(z_{1t} z_{1\ell} u_t u_\ell\right)\Big|_{t \neq \ell}$$

Keep that, and let's turn to a typical off-diagonal element,

$$[1,2]:\ \mathbf{z}_1'\mathrm{uu}'\mathbf{z}_2 = \begin{bmatrix} z_{11} & \cdots & z_{1n} \end{bmatrix} \begin{bmatrix} u_1^2 & u_1 u_2 & \cdots & u_1 u_n \\ u_2 u_1 & u_2^2 & \cdots & u_2 u_n \\ \vdots & \cdots & \ddots & \vdots \\ u_n u_1 & u_n u_2 & u_n u_{n-1} & u_n^2 \end{bmatrix} \begin{bmatrix} z_{21} \\ \vdots \\ z_{2n} \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{t=1}^n z_{1t} u_t u_1 & \sum_{t=1}^n z_{1t} u_t u_2 & \cdots & \sum_{t=1}^n z_{1t} u_t u_n \end{bmatrix} \begin{bmatrix} z_{21} \\ \vdots \\ z_{2n} \end{bmatrix}$$

$$\Rightarrow \mathbf{z}_1'\mathrm{uu}'\mathbf{z}_2 = z_{21}\sum_{t=1}^n z_{1t} u_t u_1 + z_{22}\sum_{t=1}^n z_{1t} u_t u_2 + \ldots + z_{2n}\sum_{t=1}^n z_{1t} u_t u_n$$

$$\begin{aligned} &= z_{21} z_{11} u_1^2 + z_{21} z_{12} u_2 u_1 + \ldots + z_{21} z_{1n} u_n u_1 \\ &+ z_{22} z_{11} u_1 u_2 + z_{22} z_{12} u_2^2 + \ldots + z_{22} z_{1n} u_n u_2 \\ &+ \ldots \\ &+ z_{2n} z_{11} u_1 u_n + z_{2n} z_{12} u_2 u_n + \ldots + z_{2n} z_{1n} u_2^2 \end{aligned}$$

$$\Rightarrow \mathbf{z}_1'\mathrm{uu}'\mathbf{z}_2 = \sum_{t=1}^n z_{2t} z_{1t} u_t^2 + \sum_{t\neq\ell}\sum^n z_{2t} z_{1\ell} u_t u_\ell$$

Proceeding as before,

$$\frac{1}{n}\mathbf{z}_1'\mathrm{uu}'\mathbf{z}_2 = \frac{1}{n}\sum_{t=1}^n z_{2t} z_{1t} u_t^2 + \frac{1}{n}\sum_{t\neq\ell}\sum^n z_{2t} z_{1\ell} u_t u_\ell = \frac{1}{n}\sum_{t=1}^n z_{2t} z_{1t} u_t^2 + (n-1)\frac{1}{n(n-1)}\sum_{t\neq\ell}\sum^n z_{2t} z_{1\ell} u_t u_\ell$$

$$\Rightarrow \operatorname{plim}\frac{1}{n}\mathbf{z}_1'\mathrm{uu}'\mathbf{z}_2 = E\left(z_{2t} z_{1t} u_t^2\right) + \lim (n-1) E\left(z_{2t} z_{1\ell} u_t u_\ell\right)\Big|_{t\neq\ell}$$

So the matrix $\operatorname{plim}\frac{1}{n}\mathrm{Z}'\mathrm{uu}'\mathrm{Z}$ will contain elements of the form

$$\operatorname{plim}\frac{1}{n}\mathbf{z}_1'\mathrm{uu}'\mathbf{z}_1 = E\left(z_{1t}^2 u_t^2\right) + \lim (n-1) E\left(z_{1t} z_{1\ell} u_t u_\ell\right)\Big|_{t\neq\ell} \text{ in the main diagonal}$$

$$\operatorname{plim}\frac{1}{n}\mathbf{z}_1'\mathrm{uu}'\mathbf{z}_2 = E\left(z_{2t} z_{1t} u_t^2\right) + \lim (n-1) E\left(z_{2t} z_{1\ell} u_t u_\ell\right)\Big|_{t\neq\ell} \text{ off diagonal}$$

The martingale difference property of the product process $\{\mathbf{z}_t \mathrm{u}_t\}$ says that

$$E\left(\mathbf{z}_t u_t \left|\mathbf{z}_{t-1}u_{t-1}, \mathbf{z}_{t-2}u_{t-2}, \ldots\right.\right) \equiv E\left(\mathbf{z}_t \mathrm{u}_t \mid f_{t-i}; i=1,2,\ldots\right) = 0$$

Applying the law of iterated expectations

$$\begin{aligned} t > \ell,\ \ E\left(z_{1t}z_{1\ell}u_t u_\ell\right) &= E\left[E\left(z_{1t}z_{1\ell}u_t u_\ell \mid f_{t-i}; i=1,2,\ldots\right)\right] = E\left[z_{1\ell}u_\ell E\left(z_{1t}u_t \mid f_{t-i}; i=1,2,\ldots\right)\right] \\ &= E\left[z_{1\ell}u_\ell \cdot 0\right] = 0 \end{aligned}$$

$$\begin{aligned} \ell > t,\ \ E\left(z_{1t}z_{1\ell}\mathrm{u}_t \mathrm{u}_\ell\right) &= E\left[E\left(z_{1t}z_{1\ell}u_t u_\ell \mid f_{\ell-i}; i=1,2,\ldots\right)\right] = E\left[z_{1t}u_t E\left(z_{1\ell}u_\ell \mid f_{\ell-i}; i=1,2,\ldots\right)\right] \\ &= E\left[z_{1t}u_t \cdot 0\right] = 0 \end{aligned}$$

... and analogously for $E\left(z_{2t}z_{1\ell}u_t u_\ell\right)$. So under the m.d. property, these terms vanish and we are left with

$$\operatorname{plim}\frac{1}{n}\mathbf{z}_1'\mathrm{uu}'\mathbf{z}_1 = E\left(z_{1t}^2 u_t^2\right), \quad \operatorname{plim}\frac{1}{n}\mathbf{z}_1'\mathrm{uu}'\mathbf{z}_2 = E\left(z_{2t}z_{1t}u_t^2\right)$$

We see that we need to make an assumption about the above products and their expected values. We make the simplest one, that of Conditional homoskedasticity:

$$E\left(u_t^2 \left|\mathbf{z}_t\right.\right) = \sigma^2$$

which implies by application of the LIE,

$$\operatorname{plim}\frac{1}{n}\mathbf{z}_1'\mathrm{uu}'\mathbf{z}_1 = E\left(z_{1t}^2 u_t^2\right) = \sigma^2 E\left(z_{1t}^2\right), \quad \operatorname{plim}\frac{1}{n}\mathbf{z}_1'\mathrm{uu}'\mathbf{z}_2 = E\left(z_{2t}z_{1t}u_t^2\right) = \sigma^2 E\left(z_{2t}z_{1t}\right)$$

Given these results, and going back to the full matrix, we obtain

$$E\left(\text{plim}\tfrac{1}{n}\,\text{Z}'\text{uu}'\text{Z}\right) = \sigma^2 E\left(\mathbf{z}_t\mathbf{z}_t'\right) \equiv \sigma^2\Sigma$$

So

$$\sqrt{n}\left(\hat{\beta}_{MM} - \beta\right) \xrightarrow{d} N\left(0, \text{AV}_{MM}\right), \quad \text{AV}_{MM} = \sigma^2\left(\Sigma_{XZ}\Sigma_{ZX}\right)^{-1}\Sigma_{XZ}\Sigma\Sigma_{ZX}\left(\Sigma_{XZ}\Sigma_{ZX}\right)^{-1}$$

To recapitulate, we obtained the above using:

- The Instruments orthogonality condition $E\left(\text{Z}'\text{u}\right) = \mathbf{0}$
- The martingale-difference property $E\left(\mathbf{z}_t u_t \mid f_{t-i}; i = 1, 2, \ldots\right) = 0$
- The Conditional Homoskedasticity property $E\left(u_t^2 \mid \mathbf{z}_t\right) = \sigma^2$

We see that there is no need to assume the stronger condition of mean-indepedence, $E\left(\text{u} \mid Z\right) = \mathbf{0}$ to obtain the asymptotic normality result and the estimable form of the asymptotic variance.

So this is the MM estimator of an overidentified system. We have used all the instruments and have not wasted any information. The question is "can we do better in terms of asymptotic efficiency?". The answer is yes.

B. Estimating an overidentified model using GMM

We can obtain the GMM estimator by going back to the roots: think "minimize a sum of squares", as we do in OLS. But here, minimizing the sum of squared residuals won't do, since we will get nothing else than OLS itself, since in the end, the residuals will be determined by $\hat{\text{u}} = \text{y} - \text{X}\hat{\beta}$, whatever estimator we use.

So we think "minimize the sum of squares of the orthogonality conditions, $\text{u}'\text{ZZ}'\text{u}$ "... Well, if we do that *we will obtain the MM estimator:* the same relation between MM and OLS, holds also in the overidentified case.

Then**, the idea**: why not insert a (square and p.d.) matrix of "weights", W, currently unspecified, and see where it leads us? So let's

$$\min_{\beta} \mathrm{u}'\mathrm{ZWZ}'\mathrm{u}$$

If we do that and apply matrix operations including differentiation by a vector, we will get the GMM estimator (called here a "weighted minimum distance estimator")

$$\hat{\beta}_{GMM} = \left(\mathrm{S}_{XZ}\mathrm{W}\mathrm{S}_{ZX}\right)^{-1}\mathrm{S}_{XZ}\mathrm{WZ}'\mathrm{y} \quad \text{or} \qquad \hat{\beta}_{GMM} - \beta = \left(\mathrm{S}_{XZ}\mathrm{W}\mathrm{S}_{ZX}\right)^{-1}\mathrm{S}_{XZ}\mathrm{WZ}'\mathrm{u}$$

and

$$\sqrt{n}\left(\hat{\beta}_{GMM} - \beta\right) \xrightarrow{d} N\left(0, \mathrm{AV}_{GMM}\right), \quad \mathrm{AV}_{GMM} = \sigma^2\left(\Sigma_{XZ}\mathrm{W}\Sigma_{ZX}\right)^{-1}\Sigma_{XZ}\mathrm{W}\Sigma\mathrm{W}\Sigma_{ZX}\left(\Sigma_{XZ}\mathrm{W}\Sigma_{ZX}\right)^{-1}$$

Comparing this expression with the one obtained for the MM estimator we see now in what sense the GMM estimator "generalizes" the MM estimator: it does not generalize it in the sense of estimating an overidentified model because supposedly the MM estimator can be applied only to the estimation of an exactly identified system, since we have seen that this is not true. The MM estimator is defined as the estimator arising directly from the "moment equations/orthogonality restrictions" of the model and the application of the Analogy principle, it is nowhere said that we *have* to have an exactly overidentfied system (although it is true from a historical perspective that Karl Pearson introduced and developed the MM estimator using the rule "for each unknown parameter one moment equation").

But the GMM *does* generalize the MM estimator, in the sense that the latter is a special case of the GMM *where we have set the weighting matrix* W *equal to the Identity matrix.*

This is what really new does GMM bring into the picture: the matrix W. The question now becomes "**why should we select W to be anything else than the Identity**

matrix?" And the answer is that there are other matrices for which the asymptotic variance of the GMM estimator becomes not larger than the MM estimator, and so better in efficiency terms.

One, and estimable, such matrix is $\mathrm{W} = \Sigma^{-1}$. Inserting this into AV_{GMM} a lot of things simplify and we get

$$\mathrm{AV}^{*}_{GMM} = \sigma^2 \left(\Sigma_{XZ}\Sigma^{-1}\Sigma_{ZX}\right)^{-1}$$

In order to prove the optimality of this and the superiority of GMM we have to prove that the difference

$$\mathrm{AV}_{GMM} - \mathrm{AV}^{*}_{GMM} \Rightarrow \sigma^2 \left(\Sigma_{XZ}\mathrm{W}\Sigma_{ZX}\right)^{-1} \Sigma_{XZ}\mathrm{W}\Sigma\mathrm{W}\Sigma_{ZX} \left(\Sigma_{XZ}\mathrm{W}\Sigma_{ZX}\right)^{-1} - \sigma^2 \left(\Sigma_{XZ}\Sigma^{-1}\Sigma_{ZX}\right)^{-1} \geq \mathbf{0}$$

i.e that it is a positive semidefinite matrix, for *any* W, and so also for $\mathrm{W} = I$ which is what holds for the MM estimator.

The proof of this is a good example of the power of matrix algebra.

Matrix Algebra Fact 1: if matrices A and B are p.d. then A-B is p.s.d. if and only if $\mathrm{B}^{-1} - \mathrm{A}^{-1}$ is p.s.d. This is useful because it allows us to reduce the number of matrix inverses we have to work with. Ignoring the common scaling term σ^2, in order to prove

$$\mathrm{V}_{GMM} - \mathrm{V}^{*}_{GMM} \geq \mathbf{0}$$

we can therefore equivalently prove

$$Q = \Sigma_{XZ}\Sigma^{-1}\Sigma_{ZX} - \Sigma_{XZ}\mathrm{W}\Sigma_{ZX} \left(\Sigma_{XZ}\mathrm{W}\Sigma\mathrm{W}\Sigma_{ZX}\right)^{-1} \Sigma_{XZ}\mathrm{W}\Sigma_{ZX} \geq \mathbf{0}$$

Matrix Algebra Fact 2: Since Σ is p.d. there exist an invertible matrix C such that

$$\mathrm{C}'\mathrm{C} = \Sigma^{-1}, \quad \Sigma = \mathrm{C}^{-1}\left(\mathrm{C}'\right)^{-1}$$

Inserting this, we examine

$$Q = \Sigma_{XZ}\mathrm{C}'\mathrm{C}\Sigma_{ZX} \; - \; \Sigma_{XZ}\mathrm{W}\Sigma_{ZX}\left(\Sigma_{XZ}\mathrm{W}\mathrm{C}^{-1}\left(\mathrm{C}'\right)^{-1}\mathrm{W}\Sigma_{ZX}\right)^{-1}\Sigma_{XZ}\mathrm{W}\Sigma_{ZX}$$

Using also the fact that $\Sigma_{ZX} = \Sigma'_{XZ}$ we get

$$Q = \Sigma_{XZ}\mathrm{C}'\mathrm{C}\Sigma'_{XZ} \; - \; \Sigma_{XZ}\mathrm{W}\Sigma'_{XZ}\left(\Sigma_{XZ}\mathrm{W}\mathrm{C}^{-1}\left(\mathrm{C}'\right)^{-1}\mathrm{W}\Sigma'_{XZ}\right)^{-1}\Sigma_{XZ}\mathrm{W}\Sigma'_{XZ}$$

Take common factor from the left $\Sigma_{XZ}\mathrm{C}'$ and from the right $\mathrm{C}^{-1}\Sigma'_{XZ}$:

$$Q = \Sigma_{XZ}\mathrm{C}'\left[I - \left(\mathrm{C}'\right)^{-1}\mathrm{W}\Sigma'_{XZ}\left(\Sigma_{XZ}\mathrm{W}\mathrm{C}^{-1}\left(\mathrm{C}'\right)^{-1}\mathrm{W}\Sigma'_{XZ}\right)^{-1}\Sigma_{XZ}\mathrm{W}\mathrm{C}^{-1}\right]\mathrm{C}\Sigma'_{XZ}$$

Define for clarity $\mathrm{H} \equiv \mathrm{C}\Sigma'_{XZ} \Rightarrow \mathrm{H}' = \Sigma_{XZ}\mathrm{C}'$, and also

$$\mathrm{G} \equiv \left(\mathrm{C}'\right)^{-1}\mathrm{W}\Sigma'_{XZ} \;\Rightarrow\; \mathrm{G}' = \Sigma_{XZ}\mathrm{W}\left[\left(\mathrm{C}'\right)^{-1}\right]' = \Sigma_{XZ}\mathrm{W}\left[\left(\mathrm{C}^{-1}\right)'\right]' = \Sigma_{XZ}\mathrm{W}\mathrm{C}^{-1}$$

With these we obtain

$$Q = \mathrm{H}'\left[I - \mathrm{G}\left(\mathrm{G}'\mathrm{G}\right)^{-1}\mathrm{G}'\right]\mathrm{H}$$

The middle matrix is a projection matrix, idempotent and symmetric

$I - \mathrm{G}\left(\mathrm{G}'\mathrm{G}\right)^{-1}\mathrm{G}' = \mathrm{M}, \;\; \mathrm{M}' = \mathrm{M}, \;\; \mathrm{MM} = \mathrm{M}$. Then

$$Q = \mathrm{H}'\mathrm{M}\mathrm{H} \; = \; \mathrm{H}'\mathrm{M}\mathrm{M}\mathrm{H} \; = \; \mathrm{H}'\mathrm{M}'\mathrm{M}\mathrm{H} \; = \left(\mathrm{M}\mathrm{H}\right)'\mathrm{M}\mathrm{H}$$

But this is the Gram matrix of the matrix product MH, and a Gram matrix is always positive semi-definite.

We have proved that $\mathrm{AV}_{GMM} - \mathrm{AV}^{*}_{GMM} \geq \mathbf{0}$, and so that we optimize efficiency (minimize variance) if

a) We use GMM instead of MM

b) We set the weighting matrix equal to $\mathrm{W} = \Sigma^{-1}$.

C. GMM with autocorrelation

The martingale-difference assumption does not allow for the existence of autocorrelation in the $\{\mathbf{z}_t \mathrm{u}_t\}$ process. But in a time-series setting this is difficult to defend in many cases.

If we want to allow for autocorrelation, we need to assume that the process $\{\mathbf{z}_t \mathrm{u}_t\}$ satisfies Gordin's condition, and so that the related CLT is applicable. In such a case, we obtain that

$$\tfrac{1}{\sqrt{n}} \mathrm{Z}'\mathrm{u} \xrightarrow{d} N(0,\Sigma), \quad \Sigma = \Gamma_0 + 2\sum_{j=1}^{\infty}\Gamma_j, \quad \Gamma_0 = E\left(\mathbf{z}_t\mathbf{z}_t' u_t^2\right), \quad \Gamma_j = E\left(\mathbf{z}_t u_t u_{t-j} \mathbf{z}_{t-j}'\right)$$

while the expressions for the distribution of the estimator remains unchanged, with the only difference being the expression for Σ.

A special case.

Assume that autocorrelation exists only in $\{u_t\}$.

Assume conditional homoskedasticity as before.

Assume that the Instruments are centered on their mean (implying that there is no constant term in the regression)

Assume that the sample, as regards the Instruments, is i.i.d.

Then you should be able to show that $\Gamma_j = 0 \quad \forall\, j \neq 0,\ \ \Gamma_0 = \mathrm{Var}(u) E\left(\mathbf{z}_t \mathbf{z}_t'\right)$

with the only difference from the non-autocorrelation case being that, here, $\mathrm{Var}(u)$ is expected to be larger than before, due to the autocorrelation.

--