Short Introduction to the Generalized Method of Moments*

Peter Zsohar
PhD Student
Central European University
E-mail:
zsohar_peter@ceu-budapest.edu

The generalized method of moments (GMM) is the centrepiece of semiparametric estimation frameworks. After putting GMM into context and familiarizing the reader with the main principles behind the method, we discuss the estimation procedure and the properties of the GMM estimator in details. We also provide a short survey of recent research areas in the field. To facilitate understanding, most concepts are illustrated by simple examples.

KEYWORDS:
GMM.
Semiparametric Estimation.

* Acknowledgements: The author wants to thank both *Laszlo Hunyadi*, Editor-In-Chief of the *Hungarian Statistical Review* and *Sergei Lychagin* referee, Assistant Professor of the Central European University for their helpful comments which have improved the study. All remaining errors are the author's.

Econometric analysis begins with some economic phenomenon that is of interest to us that we intend to analyse. First we turn to economic theory to see what insights it can offer. It postulates an explanation in some sort of conditions that describe the phenomena in terms of the key economic variables and model parameters. However, to answer specific questions, we have to quantify the parameters involved. We would like to adopt an estimation method whose implementation does not require the imposition of additional restrictions to the data generating process beyond those implied by the economic model. If it turns out that just for the purpose of getting these estimates we have to place further restrictions and make more assumptions and these are found to be unjustified by theory or inappropriate for the data then we run the risk that the invalidity will undermine all our subsequent inferences about the phenomenon of our interest. We would like to use a method of statistical estimation that fits well with exactly the kind of information we are getting out of our economic models. But what form does that information take? Very often restrictions implied by economic theory take the form what we will refer to as population moment conditions. The generalized method of moments (GMM) is a statistical method that combines observed economic data with the information in population moment conditions to produce estimates of the unknown parameters of this economic model. Once we have those parameters, we can go back to perform inference about the basic question that is of interest to us. Shortly we will see that GMM is very well tailored exactly to the kind of information we are getting out from our economic models.

The purpose of this article is to provide an introduction to the GMM framework and to give a rough picture of current on-going issues in the field. There are excellent textbooks and reference books available on the topic which are more precise and elaborate in all aspects like *Mátyás* [1999] or *Hall* [2005]. We will heavily rely on them and the interested reader is encouraged to study them. Our treatment misses many details but all simplifications were made to facilitate easy understanding.

After introducing the principle of the method of moments in Section 2, we show how to generalize the idea into GMM in Section 3. In Section 4 we discuss the properties of the GMM estimator. The estimation procedure is described in Section 5, while Section 6 provides a short description of testing in the GMM framework. We will also address briefly the question of moment selection in Section 7. After a short survey of the recent research in Section 8, Section 9 concludes.

1. The method of moments principle

The population moment conditions will play a crucial role in the discussion so it is worth going back to the primitives to understand the mechanics of GMM.

The raw uncentered moments are easy to compute and they reveal important aspects of a distribution. For example, the first four moments tell us about the population mean, variance, skewness and kurtosis. Using them we can immediately place restrictions according to our theory on the location, scale or shape of the distribution without specifying a full model or distribution.

Once we have some information on the population, the question remains how to use the sample to estimate the parameters of interest. In general, sample statistics each have a counterpart in the population, for example, the correspondence between the sample mean and the population expected value. The natural next step in the analysis is to use this analogy to justify using the sample moments as bases of estimators of the population parameters. This was the original idea in *Karl Pearson*'s work [1893], [1894], [1895] in the late 19^{th} century.

The Pearson family of distributions is a very flexible mathematical representation that has several important and frequently used distributions among its members depending on the parameterization you choose. Pearson's problem was to select an appropriate member of the family for a given dataset.

Example 1 – Simple method of moments estimator

To show a very simple example, assume that the population distribution has unknown mean μ and variance equal to one. In this case, the population moment condition states that $E[x_i] = \mu$. If $\{x_i : i = 1,2,\ldots,n\}$ is an independent and identically distributed sample from the distribution described formerly, then the sample average $\bar{x} = \frac{1}{n}\sum_{i=i}^{n} x_i$ is the sample analogue to the population mean $E[x_i]$. By utilizing this analogy principle, the method of moments (MM) estimator for $E[x_i] = \mu$ is simply given by $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \hat{\mu}_n$.

Basically we had to work out the first moment, then to replace it with the sample analogue and to solve the equation for the unknown parameter. What remains to be established is whether this approach is the best, or even a good way to use the sample data to infer the characteristics of the population.[1] Our intuition suggests that the bet-

[1] We will return to this subject in Section 4 discussing the properties of the GMM estimator.

ter the approximation is for the population quantity by the sample quantity, the better the estimates will be.

To make a step further, it is time to introduce some more general definitions.

Definition 1 – Method of moments estimator

Suppose that we have an observed sample $\{x_i : i = 1,2,\ldots,n\}$ from which we want to estimate an unknown parameter vector $\theta \in \mathrm{R}^p$ with true value θ_0. Let $f(x_i,\theta)$ be a continuous and continuously differentiable $\mathrm{R}^p \to \mathrm{R}^q$ function of θ, and let $E[f(x_i,\theta)]$ exist and be finite for all i and θ. Then the population moment conditions are that $E[f(x_i,\theta_0)] = 0$. The corresponding sample moments are given by

$$f_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} f(x_i,\theta).$$

The method of moments estimator of θ_0 based on the population moments $E[f(x_i,\theta)]$ is the solution to the system of equations $f_n(\theta) = 0$.

Note that if $q = p$, then for an unknown parameter vector θ the population moment conditions $E[f(x_i,\theta)] = 0$ represent a set of p equations for p unknowns. Solving these moment equations would give the value of θ which satisfies the population moment conditions and this would be the true value θ_0. Our intuition suggests that if the sample moments provide good estimates of the population moments, we might expect that the estimator $\hat{\theta}$ that solves the sample moment conditions $f_n(\hat{\theta}) = 0$ would provide a good estimate of the true value θ_0 that solves the population moment conditions $E[f(x_i,\theta_0)] = 0$.

Now we present some common models in terms of the MM terminology.

Example 2 – Ordinary least squares (OLS)

Consider the linear regression model

$$y_i = x_i'\beta_0 + u_i,$$

where x_i is the vector of p covariates, β_0 is the true value of the p unknown parameters in β, and u_i is an exogenous error term. In this case our population moment condition $E\left[f\left(x_i,\theta_0\right)\right]=0$ translates to $E\left[x_i u_i\right]=E\left[x_i\left(y_i-x_i'\beta_0\right)\right]=0$. Then the sample moment conditions are given by

$$\frac{1}{n}\sum_{i=1}^{n}x_i\hat{u}_i=\frac{1}{n}\sum_{i=1}^{n}x_i\left(y_i-x_i'\hat{\beta}\right)=0.$$

Thus the MM estimator of β_0 is given by $\hat{\beta}$ that solves this system of p linear equations and is equivalent to the standard OLS estimator.

Example 3 – Instrumental variables (IV)

If in Example 2 we allow u_i to be correlated with the covariates in x_i, we can state the population moment conditions in terms of the exogeneity assumption on the p instruments. Our population moment conditions are given by $E\left[z_i u_i\right]=E\left[z_i\left(y_i-x_i'\beta_0\right)\right]=0$ and the sample moment conditions are

$$\frac{1}{n}\sum_{i=1}^{n}z_i\left(y_i-x_i'\hat{\beta}\right)=0.$$

Just like previously, the MM estimator of β_0 is given by $\hat{\beta}$ that solves this system of p linear equations and this result shows that the standard IV estimator is also an MM type estimator.

Note that as long as the exogeneity of the error term and the instrument can be justified by economic reasoning, these examples do not impose any additional restrictions on the population that is not implied by some theory.

Example 4 – Maximum likelihood (ML)

In case we have a fully specified model, the sample log-likelihood is $\frac{1}{n}\sum_{i=1}^{n}l\left(\theta\middle|x_i\right)$. The first order conditions for the maximization of the log-likelihood function are then

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\partial l\left(\theta|x_i\right)}{\partial\theta}\Bigg|_{\theta=\hat{\theta}}=0\ .$$

These first order conditions can be regarded as a set of sample moment conditions so the maximum likelihood estimator can be given an MM interpretation as well.

So far we have considered cases where the number of moment conditions *q* was equal to the number of unknown parameters *p*. Assuming functionally independent moment equations, the resulting system of equations provided by the moment conditions can be solved to obtain the MM estimator. In the case of $q < p$ there is insufficient information and the model is not identified. If $q > p$, the model is over-identified, and in most cases, we are not able to solve the system of equations. However, estimation still can proceed and the next section will show the proper way to follow.

2. The GMM Estimator

We shall recall that population moment conditions represent information implied by some theory. It is quite natural that we want to use the most information available.[2] Unfortunately the MM estimator cannot incorporate more moments than parameters.[3]

Example 5 – Motivation for GMM

Consider again Example 1. Notice that our estimation was based solely on the first raw moment of the distribution. Now suppose that we believe to know that the sample at hand is a result of n independent draws from a Poisson distribution with parameter λ . Thus the new (additional) population moment condition based on the second raw moment is $E\left[x_i^2\right]-\lambda^2-\lambda=0$. The MM estimator of λ should satisfy the system of equations based on the sample moments

[2] Resisting the temptation to impose additional assumptions that might be unjustified by theory.

[3] However, there are still many possible actions one might think of like using all different sets of moments and then averaging the estimates, etc. but here this is not the road taken.

$$\begin{bmatrix} \frac{1}{n}\sum_{i=1}^{n} x_i - \hat{\lambda} \\ \frac{1}{n}\sum_{i=1}^{n}\left(x_i^2\right) - \hat{\lambda}^2 - \hat{\lambda} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Now we have two moment conditions and one unknown parameter which means that we do not have a general solution for $\hat{\lambda}$.

We could use only p number of moments to estimate the parameters but by dismissing the $q - p(>0)$ additional moments, we would loose the information contained in those conditions. The remedy for this situation was introduced to the econometrics literature by *Hansen* [1982] in his famous article and it is called GMM. The idea behind GMM estimation is that once it is impossible to solve the system of equations provided by the sample moment conditions, we can still have an estimate of θ that brings the sample moments as close to zero as possible.[4] Note that in the population still all moment conditions hold and the problem arises because we have a finite sample.

Definition 2 – Generalized method of moments estimator

Suppose that the conditions in Definition 1 are met and we have an observed sample $\{x_i : i = 1,2,\ldots,n\}$ from which we want to estimate an unknown parameter vector $\theta \in \Theta \subseteq \mathbb{R}^p$ with true value θ_0. Let $E\left[f\left(x_i,\theta\right)\right]$ be a set of q population moments and $f_n(\theta)$ the corresponding sample counterparts. Define the criterion function $Q_n(\theta)$ as

$$Q_n(\theta) = f_n(\theta)' W_n f_n(\theta),$$

where W_n, the weighting matrix, converges to a positive definite matrix W as n grows large. Then the GMM estimator of θ_0 is given by

$$\hat{\theta} = \underset{\theta \in \Theta}{arg\,min}\, Q_n(\theta).$$

[4] There are, of course, some statistical antecedents to GMM. The method of minimum Chi-square by *Neyman* and *Pearson, E.* [1928] deals with the general question how to estimate parameters when having more moment conditions than unknown parameters. However, they did not work with population moment conditions explicitly, the general idea was basically the same.

Basically the GMM estimator is a way of exploiting information from our general form of population moment conditions. When the number of moment conditions equals the number of unknown parameters GMM = MM. When $q > p$ then the GMM estimator is the value of θ closest to solving the sample moment conditions and $Q_n(\theta)$ is the measure of closeness to zero.

It might be useful to have a look at two practical applications from the literature that result in over-identifying moment conditions.

Example 6 by *Hansen* and *Singleton* [1982]

In their classical paper they analysed the movement of assets over time in a consumption-based capital asset pricing model. In a somewhat simpler version of their non-linear rational expectations model, the representative agent maximizes expected discounted lifetime utility

$$E\left[\sum_{\tau=0}^{\infty}\beta^{\tau}U\left(c_{t+\tau}\right)\middle|\Omega_t\right]$$

subject to the budget constraint

$$c_t + p_t q_t \le r_t q_{t-1} + w_t \qquad \forall t ,$$

where c_t is per period consumption, p_t, q_t, r_t are relative price, quantity and return on the asset with one period maturity, w_t is real wage and Ω_t is the information set of the agent in period *t*. Hansen and Singleton use a constant relative risk aversion utility function $U(c) = \left(c^{\gamma} - 1\right)/\gamma$ so the first order conditions to this optimization problem are

$$E\left[\beta\left(\frac{c_{t+1}}{c_t}\right)^{\gamma}\frac{r_{t+1}}{p_t} - 1\middle|\Omega_t\right] = 0 .$$

This looks pretty much like a population moment condition but the problem is that we have two parameters to estimate (β, γ) and only one moment condition. However, by an iterated conditional expectations argument for any vector $z_t \in \Omega_t$ the Euler-equation becomes

$$E\left[\left(\beta\left(\frac{c_{t+1}}{c_t}\right)^{\gamma}\frac{r_{t+1}}{p_t}-1\right)z_t\right]=0\,,$$

so in theory the model is identified by using any variables that are known to the agent in period *t*, such as lagged values of r_t/p_{t-1} or c_t/c_{t-1} and can be estimated consistently with GMM.[5] In contrast, maximum likelihood estimation of this model would involve exactly specifying conditional distributions of the variables and a lot of numerical integration which is computationally burdensome.

In the structural model from the previous example we were originating population moment conditions from what we were referring to as economic theory. However, sometimes "economic theory" means just some plausible assumptions based on intuition or other reasoning. Next we show an example which is based on a much less structural model, and moment conditions come from the exogeneity assumption on the instrumental variables.

Example 7 by *Angrist* and *Krueger* [1991]

The authors investigate the number of years spent in education and the subsequent earning potentials of individuals. They were interested in the impact of compulsory schooling laws in the US and estimated the following equation:

$$ln(w_i)=\beta_0+\beta_1 ed_i+controls+u_i\,.$$

The parameter of interest was β_1, the semi-elasticity of wage with respect to education. Estimating this linear equation by OLS could be biased and inconsistent as ed_i is probably correlated with individual factors in the regression error term u_i such as individual costs and potential benefits of schooling or other options outside the schooling system, most of which are unobserved by the researcher. Using the structure of compulsory school attendance laws at that time in the US they were able to argue that (in addition to the controls) dummy variables indicating the quarter of birth for each individual could be used to in-

[5] Note that the original variables in the model need not be stationary as taking consequent ratios makes the series stationary.

strument for the years spent in education. Their exogeneity assumption implies that the following population moment conditions hold:

$$E\left[z_i \left(ln\left(w_i \right) - \beta_0 - \beta_1 ed_i - controls \right) \right] = 0,$$

where the vector of instruments z_i contains the exogenous variables from the original model supplemented by the quarter of birth dummies. Note that there are more moment conditions then parameters and we could estimate the model by GMM.

3. Properties of the GMM Estimator

Under some sufficient conditions the GMM estimator as given in Definition 2 is consistent and asymptotically normally distributed. In the following we will discuss these properties and the sufficient conditions in somewhat more detail.

Population moment conditions provide information about the unknown parameters. The quality and the utilization method of this information are crucial in several aspects. First, a natural question arises about the sufficiency of the information contained in the moment conditions whether it is enough for the estimation to be "successful". This leads us to the issue of identification.

Assumption 1 – Identification

In the following, we present the necessary conditions for identification.

– *Order condition*: As we have already seen if $q < p$, the model is not identified and we are unable to estimate the parameters. So we need $q \geq p$.

– *Rank condition*: Once we have enough moment conditions, it is still crucial that among those moments should be at least *p* functionally independent ones which are satisfied if the expectation of the $q \times p$ Jacobian of the moment equations evaluated at θ_0 has rank (at least) *p*.

– *Uniqueness*: If we think of $E\left[f\left(x_i, \theta \right) \right]$ as a function of θ, then for successful estimation $E\left[f\left(x_i, \theta \right) \right] = 0$ has to be a unique property of θ_0. It means that θ_0 should be the only parameter vector which satisfies the population moment conditions.

We also need to establish a connection between the population moments and their sample counterparts. This will ensure that in the limit, the true parameter vector will be the one that solves the sample moment equations.

Assumption 2 – Convergence of sample moments

If the data generating process is assumed to meet the conditions for some kind of law of large numbers to apply, we may assume that the sample moments converge in probability to their expectation. That is $f_n(\theta_0)\left(=\frac{1}{n}\sum_{i=1}^{n} f(x_i,\theta_0)\right)$ converges to $E\left[f_n(\theta_0)\right]=E\left[f(x_i,\theta_0)\right]=0.$

Note that we have basic laws of large numbers only for independent observations. For a more general case, with dependent or correlated observations, we would assume that the sequence of observations $f(x_i,\theta)$ constitutes an ergodic q-dimensional process. Assumptions 1 and 2 together with the conditions from Definitions 1 and 2 establish that the parameter vector will be estimable.

Now we make a statistical assumption that allows us to establish the properties of asymptotic distribution of the GMM estimator.

Assumption 3 – Distribution of Sample Moments

We assume that the sample moments obey a central limit theorem. This assumes that the moments have a finite asymptotic covariance matrix, $(1/n)F$, so that

$$\sqrt{n} f_n(\theta_0) \xrightarrow{d} N(0,F).$$

Again, if the observations are not independent, it is necessary to make some assumptions about the data so that we could apply an appropriate central limit theorem.

Theorem 1 – GMM is consistent and asymptotically normal

Under the preceding assumptions, the GMM estimator is consistent and asymptotically normally distributed with asymptotic covariance matrix V_{GMM} defined as

$$V_{GMM}=\frac{1}{n}\left[G(\theta_0)' WG(\theta_0)\right]^{-1} G(\theta_0)' WFWG(\theta_0)\left[G(\theta_0)' WG(\theta_0)\right]^{-1},$$

where G is a $(q \times p)$ matrix defined as

$$G(\theta) = E\left[\frac{\partial f(x,\theta)}{\partial \theta}\right] = E\begin{bmatrix} \frac{\partial f_1(x,\theta)}{\partial \theta_1} & \cdots & \frac{\partial f_1(x,\theta)}{\partial \theta_p} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_q(x,\theta)}{\partial \theta_1} & \cdots & \frac{\partial f_q(x,\theta)}{\partial \theta_p} \end{bmatrix},$$

that is, $G(\theta_0)$ is the expected value of the Jacobian of the population moment functions evaluated at the true parameter value θ_0.

The point is that in general the variance of the GMM estimator depends on the choice of W_n, so we can extract the most information from the moment conditions by choosing an appropriate weighting matrix. By analysing the quadratic form in the GMM criterion function

$$Q_n(\theta) = f_n(\theta)' W_n f_n(\theta),$$

we see that setting $W_n = I$ gives us the sample moments' error sum of squares.

In fact, we could use a diagonal weighting matrix $W_n = \langle w \rangle$ to minimize the weighted sum of squared errors. This is a natural idea as some moments might be more volatile than others and, thus, it makes sense to normalize the errors in the moments by their variance.

However, the elements of $f_n(\theta)$ are freely correlated. Suppose the asymptotic covariance of the sample moments normalized by the root of the sample size is $Asy.Var\left[\sqrt{n} f_n(\theta_0)\right] = F$. Then the choice of $W_n = F^{-1}$ weights all elements of the criterion function appropriately so should be optimal based on the same idea that motivates generalized least squares.

Theorem 2 – Optimal weighting matrix

For a given set of moment conditions with the optimal choice of the weighting matrix $W_n = F^{-1}$, the GMM estimator is asymptotically efficient with covariance matrix

$$V_{GMM,optimal} = \frac{1}{n}\left[G(\theta_0)' F^{-1} G(\theta_0)\right]^{-1}.$$

It is important to emphasize that the efficiency result is valid only for a given set of moment conditions. That is, GMM is asymptotically efficient in the class of consistent and asymptotically normal estimators that do not use any additional information on top of that is contained in the moment conditions. The traditional ML utilizes a fully specified distribution so the two estimators are incomparable in a sense that they rely on different information sets. However, as we saw earlier in Example 4 if the moment conditions are the same as the first order conditions of ML estimation then the two estimators are numerically equal.

Especially, if the model is correctly specified and the underlying distribution is one from the exponential family, we can use the sufficient statistics as bases for moment conditions. In these cases GMM is efficient in a sense that it attains the Cramer–Rao lower bound asymptotically. The problem with this theoretical case is that it is unoperational as GMM's main strength is not specifying an exact distribution.

4. Estimation

After having discussed the properties of the GMM estimator, it is time to turn to some more practical issues like estimation. The question is how do we get those numbers when we have the data.

In the exactly identified case when $q = p$, GMM works the same as MM and there is no need for optimization as the system of moment conditions can be solved for the unknown parameters.

Example 8 – Exactly identified case

Consider the Poisson model from Example 5. Recall that we had two moment conditions for the single unknown parameter λ. Suppose we have a sample of $n = 20$ observations. Now we are going to estimate λ based on both moment conditions separately. The estimators relying on the first two raw moments are

$$\hat{\lambda}_{first} = \frac{1}{20}\sum_{i=1}^{20} x_i = 3.55,$$

$$\hat{\lambda}_{second} = \frac{-1 + \sqrt{1 + 4\frac{1}{20}\sum_{i=1}^{20} x_i^2}}{2} = 3.3859.$$

In order to utilize both moments at once, we need to compute the GMM estimator. Recall that in the over-identified case when $q > p$ the asymptotic variance of the GMM estimator, V_{GMM} depends on the weighting matrix. We want to get the most information out of our moment conditions thus we would like to use the optimal weighting matrix that minimizes V_{GMM}. As we discussed earlier, this would be F^{-1}. Logic suggests that first we should estimate the optimal weighting matrix so that we could use it in the criterion function to estimate θ_0 efficiently. The problem is that to get an estimator of F^{-1}, we already need an estimate of θ_0.

We can resolve this circularity by adopting a multi-step procedure.

1. We can choose a sub-optimal weighting matrix, say *I,* and minimize the simple sum of squared errors in the moments $Q_n(\theta) = f_n(\theta)' f_n(\theta)$. This will deliver a preliminary but consistent estimate of θ_0 which can be used then to estimate F and thus F^{-1} consistently.

2. With the optimal weighting matrix estimate at hand, we can minimize the new criterion function $Q_n(\theta) = f_n(\theta)' \hat{F}^{-1} f_n(\theta)$, and estimate θ_0 efficiently.

This is the so-called two-step GMM estimator which is consistent and efficient.

Example 9 – GMM estimation

We now continue with the Poisson example. In the first step we have to minimize the criterion function with using I as a weighting matrix:

$$Q_n(\theta) = f_n(\theta)' f_n(\theta) = \begin{bmatrix} \frac{1}{20}\sum_{i=1}^{20} x_i - \lambda \\ \frac{1}{20}\sum_{i=1}^{20}\left(x_i^2\right) - \lambda^2 - \lambda \end{bmatrix}' \begin{bmatrix} \frac{1}{20}\sum_{i=1}^{20} x_i - \lambda \\ \frac{1}{20}\sum_{i=1}^{20}\left(x_i^2\right) - \lambda^2 - \lambda \end{bmatrix}.$$

To facilitate computation, we started the optimization routine from the MM estimate based on the first raw moment $\hat{\lambda}_{first} = 3.55$. The first-step GMM estimate is

$$\hat{\lambda}_1 = 3.3885$$

which can be used to estimate F as follows:

$$\hat{F}_n = \frac{1}{n}\sum_{i=1}^{n} f\left(x_i, \hat{\theta}_1\right) f\left(x_i, \hat{\theta}_1\right)' = \frac{1}{20}\sum_{i=1}^{20}\begin{bmatrix} x_i - \hat{\lambda}_1 \\ x_i^2 - \hat{\lambda}_1^2 - \hat{\lambda}_1 \end{bmatrix}\begin{bmatrix} x_i - \hat{\lambda}_1 \\ x_i^2 - \hat{\lambda}_1^2 - \hat{\lambda}_1 \end{bmatrix}'.$$

Substituting in for $\hat{\lambda}_1$ and inverting $\hat{F}$ the estimated optimal weighting matrix is

$$\hat{W}_{optimal} = \hat{F}_n^{-1} = \begin{bmatrix} 10.5333 & -1.4504 \\ -1.4504 & 0.2085 \end{bmatrix}.$$

In the second step we have to minimize the new criterion

$$Q_n(\theta) = f_n(\theta)' \hat{F}^{-1} f_n(\theta).$$

The optimization routine was started from the first-step GMM estimate. Solving the minimization problem, for the second-step GMM estimator of λ we get

$$\hat{\lambda}_{GMM} = 3.2651.$$

We still have to estimate the variance of the estimator, V_{GMM}. First we recompute $\hat{F}_n^{-1}$ with $\hat{\lambda}_{GMM}$ at hand exactly as we did previously. Then we also have to estimate G, the matrix of the derivatives. G is the expected value of the Jacobian but notice that in our case the derivatives do not depend directly on the data so it can be estimated simply as

$$\hat{G} = \begin{bmatrix} -1 \\ -2\hat{\lambda}_{GMM} - 1 \end{bmatrix}.$$

Now we can compute the estimated variance of $\hat{\lambda}_{GMM}$ as

$$\hat{V}_{GMM} = \frac{1}{n}\left[\hat{G}'\hat{F}_n^{-1}\hat{G}\right]^{-1} = 0.0973.$$

Notice that the MM estimate based on the first raw moment equals the ML estimate for which we can estimate the asymptotic variance from the Cramer–Rao lower bound as $\hat{\lambda}/n$.

The Table compares the results.

Comparison of estimators

Results	ML	Two-step GMM
λ	3.55	3.2651
Standard error	0.4213	0.3119

In fact, we could continue this multi-step procedure to obtain the so-called iterated GMM estimator. *Hansen*, *Heaton* and *Yaron* [1996] suggest a method where the dependence of the weighting matrix on the unknown parameters is acknowledged and taken care of during the optimization procedure. Their approach became known as the continuously updated GMM. There is fairly compelling evidence to suggest that there are gains to iteration in terms of finite sample performance of the estimator but in most cases the two-step estimator is applied.

Given the mathematical form of the moment conditions, in some cases we can solve the optimization problem analytically and get a closed form representation of the estimates in terms of the data which will speed up the computations. However, in other cases such as with nonlinear models, we have to use numerical optimization routines. The problem with the widely used Newton–Raphson and other practical numerical optimization methods is that global optimization is not guaranteed. The GMM estimator is defined as a global minimizer of a GMM criterion function, and the proof of its asymptotic properties depends on this assumption. Therefore, the use of a local optimization method can result in an estimator that is not necessarily consistent or asymptotically normal.

Care should be taken with nonconvex problems where the existence of possibly multiple local minima may cause problems. With starting the optimization algorithm from several initial values spread out in the parameter space, one might be able to find the global minimum. However, it should be noted that the multi-start algorithm does not necessarily find he global optimum and is computationally intensive.

There are of course much more advanced numerical techniques and there is a freely available and fairly user friendly GMM toolbox for MATLAB by *Kyriakoulis* [2004].

An alternative solution in such cases is the use of Monte Carlo simulation methods to compute an otherwise intractable criterion function. The method of simulated moments approach is the simulated counterpart of the traditional GMM procedure and is applicable when the theoretical moments cannot be computed analytically. An extensive survey of recent (mostly theoretical) results in the subject can be found in *Liesenfeld–Breitung* [1999].

5. Testing in the GMM framework

Most of the times there are three broad inference questions that are of interest to us:

– Is the model correctly specified?
– Does the model satisfy certain particular restrictions?
– Which model appears to be more consistent with the data?

The first question is particularly important. Recall that the population moment conditions were deduced from an underlying economic model and all our inference is going to be based on them. As our estimate is relying on the information contained in the moment conditions, it is crucial whether the original model is consistent with the data or whether it appears to be a good representation of the data.

If the hypothesis of the model that led to the moment equations in the first place is incorrect, at least some of the sample moment restrictions will be systematically violated. This conclusion provides the basis for a test of the over-identifying restrictions and if we have more moments than parameters, we have scope for testing that. There is a very simple to compute statistic to use as an over-identifying restrictions test (the so-called J test) which is just the sample size times the value of the GMM criterion function evaluated at the second step GMM estimator

$$nQ_n\left(\hat{\theta}\right)=\left[\sqrt{n}\,f_n\left(\hat{\theta}\right)\right]'\left(Est.Asy.Var\left[\sqrt{n}\,f_n\left(\theta_0\right)\right]\right)^{-1}\left[\sqrt{n}\,f_n\left(\hat{\theta}\right)\right].$$

Notice that this is a Wald statistic and under the null

$$H_0{:}E\left[f\left(x_i,\theta_0\right)\right]=0,$$

and it has a large sample Chi-squared distribution with $q-p$ degrees of freedom. However, the over-identifying restrictions test can be computed only in case of $q>p$, as in the exactly identified model the criterion function is zero. The reason for the importance and the popularity of this test is that it really examines the heart, the crux of GMM, and it is easy to calculate, as it is an obvious by-product of the estimation procedure. The statistic is ubiquitously reported in all applications involving GMM estimation just as reporting the log of the likelihood function in ML estimation. We would like to stress that it is very important to do some kind of misspecification test as in misspecified models the properties of the GMM estimator are substantially different which is likely to make all subsequent inferences misleading.

Example 10 – Test for over-identifying restrictions
Consider the Poisson model from Example 9. Now we have

$$J_n = nQ_n\left(\hat{\lambda}_{GMM}\right) = 20 \times 0.2694 = 5.388 .$$

As $J_n \sim \chi^2[1]$, there is only very weak evidence in favour of the population moment conditions so the model can be rejected.[6]

The second inference question asks whether the model satisfies certain additional restrictions implied by economic and statistical theory that we could impose and what might tell us about economic behaviour. Fortunately all the well-known likelihood-based testing procedures have their GMM counterparts with very similar implementations. The GMM-based LR test is computed by using nQ_n instead of $ln L$ in the test statistic. The GMM-based Wald statistic is computed identically to the likelihood-based one by using the GMM estimates instead of the ML estimates. The LM test is derived by the same logic applied to the derivatives of the GMM criterion function.

The third question is model selection. The previously mentioned tests are applicable for nested models but selection from a set of non-nested models would require specifying the distribution of the data generating process.

Those interested in details should read the extensive discussion in Chapter 5 of *Hall* [2005].

6. Choice of moment conditions

So far we have covered how we can exploit information from our moment conditions in an efficient way but we haven't mentioned what is the best set of moment conditions to be used. It turns out that there is quite a straightforward answer to this, although it won't be very useful in terms of practical work.

Maximum likelihood is the asymptotically efficient estimator in the class of consistent and asymptotically normal estimators. Recall that we have already shown that ML is an MM type estimator based on the score function. Thus, if we use the derivatives of the log-likelihood function as moments, we will get an efficient estimator. Unfortunately this is not feasible as in most economic settings the population distribution is

[6] However, the small sample size and the discrete nature of the Poisson distribution should raise some concerns.

unknown. Making an additional assumption on the underlying distribution places restrictions on the economic variables involved that might be unjustified by economic theory and that is exactly the kind of thing that GMM was designed to help us avoid.

But what if we have more moment conditions than parameters? We might expect that more information never hurts but it turns out that sometimes in fact it doesn't help either. There are two main approaches in the literature to moment selection. One suggests optimal moment condition selection based on asymptotic theory among the class of generalized instrumental variables. Unfortunately in many settings it is infeasible just like with the score function. The other strand of literature emphasizes practical data based moment selection introducing different selection criteria. Some results suggest that they may help avoid situations where the asymptotic approximation of finite sample behaviour is poor. For a detailed summary of recent results please see *Hall* [2005].

7. Actively researched topics

All our reasoning and inferences so far were based on large sample theory. Two important questions arise:

– How well does this theory approximate finite sample behaviour in the kind of places where we want to apply GMM?
– Can we identify factors and aspects of model specification that appear to affect the quality of this approximation?

Numerous studies try to address these issues in the literature. There are analytical approaches based on higher order asymptotics and simulation supported studies applied to generated artificial data from structures to which we typically fit our economic theories. A detailed discussion and summary of the topic can be found in *Podivinsky* [1999] or in Chapter 6 of *Hall* [2005]. *Harris* and *Mátyás* [2004] provide an extensive comparative analysis of different IV and GMM estimators, focusing on their small sample properties. These studies assess how well the methods perform and the findings are perhaps not that surprising. Loosely speaking, sometimes GMM works well but sometimes it does not. The main factors that were found to affect the quality of asymptotic approximation are:

– form of moment conditions $f(x_i, \theta)$. Basically, the more nonlinearity is involved, the less good the approximation is;
– degree of over-identification $(q - p)$;

– interrelation between the elements of moment conditions;
– quality of identification.

How can we improve on the quality of inference then? There are three main strands of responses in the literature, two of which stay in the GMM framework and one suggesting another method.

– If we would like to stick to first order asymptotic theory, the method of moment selection has to be revised. There are procedures for selecting those moments that contribute to parameter estimation and retain the ones that help and discard those that don't add new information.

– There are alternative considerations to develop a large sample theory and try to use this alternative asymptotic framework to come up with inference procedures:

♦ weak identification – to tackle the case of uninformative moment conditions;

♦ artificial resampling techniques, especially bootstrap – to improve the accuracy of critical points used in tests;

♦ alternative theory of many moment conditions asymptotics for cases when $q >> p$.

– Step outside GMM. The problems arise because of the structure of GMM estimation so propose the generalized empirical likelihood class of estimators which contains the so-called continuously updated GMM and other empirical likelihood-based estimators.

8. Concluding Remarks

The econometrics literature offers the researcher a broad variety of estimation methods differing in the amount of information they use, ranging from fully parameterized likelihood-based techniques to pure nonparametric methods and a rich variety in between. Choosing one appropriately is a respectful task as a correctly specified parametric model provides much better quality estimates than methods that assume little more than mere association between variables at one another. However, this efficiency comes at a cost of possibly false restrictions. From another standpoint, semi- and non parametric methods are much more robust to variations in the underlying data generating process and still may provide consistent estimates without imposing additional assumptions. We have discussed that GMM is more robust to model specification than ML as it requires less information. This explains the increasing popularity of semi-parametric estimation frameworks like GMM, as they allow to incorporate only as much restriction as economic theory implies.

To state it differently, the GMM estimator is built on more general (assumed) characteristics of the population than in the classical likelihood-based framework, as it requires fewer and weaker assumptions.

References

ANGRIST, J. – KRUEGER, D. [1991]: Does Compulsory School Attendance Affect Schooling and Earnings? *Quarterly Journal of Economics*. Vol. 106. No. 4. pp. 979–1014.

HALL, A. R. [2005]: *Generalized Method of Moments*. Advanced Texts in Econometrics. Oxford University Press. Oxford.

HANSEN, L. P. [1982]: Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica*. Vol. 50. No. 4. pp. 1029–1054.

HANSEN, L. P. – HEATON, J. – YARON, A. [1996]: Finite-Sample Properties of Some Alternative GMM Estimators. *Journal of Business and Economic Statistics*. Vol. 14. No. 3. pp. 262–280.

HANSEN, L. P. – SINGLETON, K. J. [1982]: Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models. *Econometrica*. Vol. 50. No. 5. pp. 1269–1286.

HARRIS, M. N. – MÁTYÁS, L. [2004]: A Comparative Analysis of Different IV and GMM Estimators of Dynamic Panel Data Models. *International Statistical Review*. Rev. 72. No. 3. pp. 397–408.

KYRIAKOULIS, K. [2004]: *Gmm Toolbox for Matlab*. http://www.kostaskyriakoulis.com/gmmgui.html

LIESENFELD, R. – BREITUNG, J. [1999]: Simulation Based Method of Moments. In: *Mátyás, L*. (ed.): *Generalized Method of Moments Estimation*. Themes in Modern Econometrics. Cambridge University Press. Cambridge. pp. 275–300.

MÁTYÁS, L. (ed.) [1999]: *Generalized Method of Moments Estimation*. Themes in Modern Econometrics. Cambridge University Press. Cambridge.

NEYMAN, J. – PEARSON, E. [1928]: On the Use and Interpretation of Certain Test Criteria for Purposes of Statistical Inference. Part 2. *Biometrika*. Vol. 20. No. 3–4. pp. 263–294.

PEARSON, K. [1893]: Asymmetrical Frequency Curves. *Nature*. Vol. 49. Issue 1253. p. 6.

PEARSON, K. [1894]: *Contributions to the Mathematical Theory of Evolution*. Philosophical Transactions of the Royal Society of London. Royal Society of London. London.

PEARSON, K. [1895]: *Contributions to the Mathematical Theory of Evolution II*: *Skew Variation*. Philosophical Transactions of the Royal Society of London. Royal Society of London. London.

PODIVINSKY, J. M. [1999]: Finite Sample Properties of GMM Estimators and Tests. In: *Mátyás, L*. (ed.): *Generalized Method of Moments Estimation*. Themes in Modern Econometrics. Cambridge University Press. Cambridge.