ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/227564036

Generalized Method of Moments

Chapter · November 2007

DOI: 10.1002/9780470996249.ch12 · Source: RePEc

CITATIONS
301

READS
5,410

1 author:

Alastair R. Hall
The University of Manchester

100 PUBLICATIONS **3,152** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project Bootstrapping Structural Change Tests View project

All content following this page was uploaded by Alastair R. Hall on 26 November 2018.

The user has requested enhancement of the downloaded file.

Generalized Method of Moments[1]

Alastair R. Hall

The University of Manchester

Manchester, UK[2]

March 10, 2009

[1]Manuscript prepared for inclusion in the section edited by Ole Barndorff-Nielsen and Eric Renault on financial econometrics that is to appear in the *Encyclopedia of Quantitative Finance*.

[2]contact address: Economics, School of Social Sciences, The University of Manchester, Oxford Road, Manchester M13 9PL, UK; phone: 0161 275 4875; fax: 0161 275 4812; email: Alastair.Hall@manchester.ac.uk

List of included elements:

- article title
- contributor's name
- contributor's affiliations
- keywords
- abstract
- main text
- references

Information on files:

- the file name is : `GMM_EQF_100309.tex` and it is created using PCTeX (LaTeX)

Keywords: population moment conditions, parameter estimation, test for model misspecification, time series.

Abstract: This entry describes the basic framework for statistical estimation and inference using Generalized Method of Moments and also illustrates the types of empirical models in finance to which these techniques have been applied.

Generalized Method of Moments (GMM) provides a computationally convenient method of obtaining consistent and asymptotically normally distributed estimators of the parameters of statistical models. The method has been applied in many areas of economics but has arguably been most frequently applied in finance. In fact, it was in empirical finance that the power of the method was first illustrated: for while Hansen [7] introduced GMM and presented its fundamental statistical theory, Hansen and Hodrick [9] and Hansen and Singleton [10] showed the potential of the GMM approach to estimation through their empirical analyses of, respectively, foreign exchange markets and asset pricing. This entry briefly describes the GMM framework for estimation and inference in the case where time series data are used. The reader is referred to [6] for a comprehensive treatment of GMM that presents both a rigorous statistical analysis of the method and empirical illustrations in finance. For applications of GMM to panel data, see [15].

The popularity of GMM can be understood by comparing the requirements for the

method to those for Maximum Likelihood (ML). While ML is the best available estimator within the Classical statistics paradigm, its optimality stems from its basis on the joint probability distribution of the data. However, in many scenarios of interest in finance, this dependence on the probability distribution can become a weakness for two main reasons. Firstly, in some models, the joint probability distribution is only implicitly specified and is not available in a convenient closed form with result that ML is computationally infeasible; *e.g.* in stochastic volatility models, see [11]. Secondly, in other cases, the underlying theoretical model places restrictions on the distribution of the data but does not completely specify its form with the result that ML is infeasible unless the researcher imposes an arbitrary assumption about the distribution. The latter is an unattractive strategy because if the assumed distribution is incorrect then the optimality of ML is lost, and the resulting estimator may even be inconsistent; *e.g.* in nonlinear Euler equation models, see [10]. In contrast, GMM estimation is based on population moment conditions. It turns out that in many cases where the financial-theoretic model does not specify the complete distribution, it does specify population moment conditions. Therefore, in these settings, GMM can be prefered to ML because it offers a way to estimate the parameters based solely on the information deduced from the underlying financial model.

The cornerstone of GMM estimation is the population moment condition:

Definition 1 Population Moment Condition

Let θ_0 be a vector of unknown parameters which are to be estimated, v_t be a vector of random

variables and $f(.)$ a vector of functions then a population moment condition takes the form

$$E[f(v_t, \theta_0)] = 0 \tag{1}$$

for all t.

To illustrate the types of population moment condition that can arise in finance models, we consider the specific examples in the seminal papers [9] and [10].

Example 1: the efficiency of foreign exchange markets

Consider the following regression model for the relationship between the forward and spot exchange rates:

$$s_{t+k} \;=\; \beta_{0,1} \;+\; \beta_{0,2} f_{t,k} \;+\; u_t \tag{2}$$

where $f_{t,k}$ is the log of the $k-$ period forward exchange rate at time t, s_{t+k} is the log of the spot exchange rate in period $t+k$, $\beta_{0,1}$ and $\beta_{0,2}$ are unknown parameters and u_t is an error term. Hansen and Hodrick [9] demonstrate that under the simple efficient-markets hypothesis, the following population moment condition holds

$$E[f_{t,k}(s_{t+k} \;-\; \beta_{0,1} \;-\; \beta_{0,2} f_{t,k})] = 0 \tag{3}$$

where $(\beta_{0,1}, \beta_{0,2}) = (0, 1)$. $\diamond$

Example 2: the consumption based asset pricing model

The consumption based asset pricing model involves a representative agent who makes investment and consumption decisions at time t to maximize his/her discounted lifetime utility subject to a budget constraint. Assuming that the agent's utility function is of the constant relative risk aversion form and that the only one asset available as a possible

investment yields a pay-off in the following period then, Hansen and Singleton [10] show that the model implies the following population moment condition

$$E[z_t\{\delta_0(c_{t+1}/c_t)^{\gamma_0-1}(g_{t+1}/p_t)-1\}]=0 \tag{4}$$

where c_t is consumption in period t, g_{t+1} is the gain (cash-flow) on the asset in period $t+1$, p_t is the price of the asset in period t, z_t is a vector of variables in the agent's information set at time t, and the unknown parameters $1-\gamma_0$ and δ_0 are respectively the agent's coefficient of relative risk aversion and discount factor. $\diamond$

The population moment condition states that $E[f(v_t,\theta)]$ equals zero when evaluated at θ_0. For the GMM estimation to be successful, this must be a unique property of θ_0, that is $E[f(v_t,\theta)]$ is not equal to zero when evaluated at any other value of θ. If that latter holds then θ_0 is said to be *identified* by $E[f(v_t,\theta_0)]=0$. A first order condition for identification (often refered to as a "local condition") is that $rank\{G(\theta_0)\}=p$, where $G(\theta)=E[\partial f(v_t,\theta)/\partial\theta']$, and this plays a crucial role in standard asymptotic distribution theory for GMM. By definition the moment condition involves q pieces of information about p unknowns, therefore identification can only hold if $q\geq p$. For reasons that emerge below it is convenient to split this scenario into two parts: $q=p$, in which case θ_0 is said to be *just-identified*, and $q>p$, in which case θ_0 is said to be *over-identified*.

Let $g_T(\theta)$ be the analogous sample moment to $E[f(v_t,\theta)]$, that is $g_T(\theta)=T^{-1}\sum_{t=1}^{T}f(v_t,\theta)$ where T is the sample size. The GMM estimator is then as follows.

Definition 2 Generalized Method of Moments Estimator

The Generalized Method of Moments estimator based on (1) is the value of θ which mini-

mizes:

$$Q_T(\theta) \;=\; g_T(\theta)' W_T g_T(\theta) \tag{5}$$

where W_T is known as the weighting matrix and is restricted to be a positive semi–definite matrix that converges in probability to W, some positive definite matrix of constants.

The restrictions on the weighting matrix are required to ensure that $Q_T(\theta)$ is a meaningful measure of the distance the sample moment is from zero at different values of θ. The choice of weighting matrix is discussed further below.

The first order conditions for this minimization are:

$$G_T(\hat{\theta}_T)' W_T g_T(\hat{\theta}_T) \;=\; 0 \tag{6}$$

where $G_T(\theta) \;=\; T^{-1} \sum_{t=1}^{T} \partial f(v_t, \theta)/\partial \theta'$. The structure of these conditions reveals some interesting insights into GMM estimation. Since $G_T(\theta)$ is $q \times p$, it follows that (6) involves calculating $\hat{\theta}_T$ as the value of θ that sets the p linear combinations of $g_T(.)$ to zero. Therefore, if $p = q$ (and $G_T(\hat{\theta}_T)' W_T$ is nonsingular) then $\hat{\theta}_T$ satisfies the analogous sample moment condition to (1), $g_T(\hat{\theta}_T) = 0$, and is, thus, the Method of Moments estimator based on the original moment condition. However, if $q > p$ then the first order conditions are not equivalent to solving the sample moment condition. Instead, $\hat{\theta}_T$ is equivalent to the Method of Moments estimator based on

$$G(\theta_0)' W E[f(v_t, \theta_0)] \;=\; 0 \tag{7}$$

Although (1) implies (7), the reverse does not hold because $q > p$; therefore, in this case, the estimation is actually based on only part of the original information. As a result, if

$q > p$ then GMM can be viewed as decomposing the original moment condition into two parts, the *identifying restrictions*, which contain the information actually used in the estimation, and the *overidentifying restrictions*, which represents a remainder. Furthermore, GMM estimation produces two fundamental statistics and each is associated with a particular component: the estimator $\hat{\theta}_T$ is a function of the information in the identifying restrictions, and the estimated sample moment, $g_T(\hat{\theta}_T)$, is a function of the information in the overidentifying restrictions. While unused in estimation, the overidentifying restrictions play a crucial role in inference about the validity of the model as is discussed below.

To establish the consistency and asymptotic normality of the GMM estimator, it is necessary to place restrictions on v_t and $f(.)$. In the standard asymptotic framework (see [7]), it is assumed that v_t is stationary and ergodic, $f(.)$ is continuous and differentiable and that its derivative is continuous in θ. Various expectations of $f(.)$ and its derivative must also exist. These conditions permit the applications of Laws of Large Numbers and Central Limit Theorems to the requisite sample averages. Chief among the restrictions on expectations of $f(.)$ are that the population moment condition is valid and identifies θ_0. Under the above conditions, it can be shown that the $\hat{\theta}_T \xrightarrow{p} \theta_0$ and

$$T^{1/2}(\hat{\theta}_T - \theta_0) \xrightarrow{d} N(0, V) \tag{8}$$

where

$$V = [G(\theta_0)'WG(\theta_0)]^{-1}G(\theta_0)'WS(\theta_0)WG(\theta_0)[G(\theta_0)'WG(\theta_0)]^{-1}$$

and $S(\theta) = lim_{T\to\infty}Var[T^{1/2}g_T(\theta)]$.

At this point, we return to the issue of the choice of weighting matrix. In the discussion

of the first order conditions above, it is noted that if $p = q$ then the GMM estimator can be found by solving $g_T(\hat{\theta}_T)$. As a result, the estimator does not depend on W_T. The asymptotic properties must also be invariant to W_T and it can be shown that V reduces to $\{G(\theta_0)'S(\theta_0)^{-1}G(\theta_0)\}^{-1}$ in this case. However, if $q > p$ then the first order conditions depend on W_T and therefore so does $\hat{\theta}_T$ in general. This dependence is unattractive because it raises the possibility that subsequent inferences can be affected by the choice of weighting matrix. However, in terms of asymptotic properties, the choice of weighting matrix only manifests itself in V, the asymptotic variance of the estimator. Since this is the case, it is natural to choose W_T such that V is minimized in a matrix sense. Hansen [7] shows that this can be achieved by setting $W_T = \hat{S}_T^{-1}$ where $\hat{S}_T$ is a consistent estimator of $S(\theta_0)$. The resulting asymptotic variance is $V = V^0 = \{G(\theta_0)'S(\theta_0)^{-1}G(\theta_0)\}^{-1}$; Chamberlain [4] shows V^0 is a semi-parametric efficiency bound for estimation of θ_0 based on (1).

In practical terms, two issues arise in the implementation of GMM with this choice of weighting matrix: (i) how to construct $\hat{S}_T$ so that it is a consistent estimator of $S(\theta_0)$; (ii) how to handle the dependence of $\hat{S}_T$ on $\hat{\theta}_T$. We treat each in turn.

(i) Estimation of $S(\theta_0)$: Under stationarity and ergodicity and certain other technical restrictions, it can be shown that $S(\theta_0) = \Gamma_0(\theta_0) + \sum_{i=1}^{\infty}\{\Gamma_i(\theta_0) + \Gamma_i(\theta_0)'\}$ where $\Gamma_i(\theta_0) = Cov[f(v_t, \theta_0), f(v_{t-i}, \theta_0)]$ is known as the i-lag autocovariance matrix of $f(v_t, \theta_0)$; see [2]. In some cases, the structure of the model implies $\Gamma_i(\theta_0) = 0$ for all $i > k$ for some k, and this simplifies the estimation problem; see [6][Section 3.5]. In the absence of such a restriction on the autocovariance matrices, the long run variance can be estimated by a member of

the class of *heteroscedasticity autocorrelation covariance (HAC)* estimators defined as

$$\hat{S}_{HAC} = \hat{\Gamma}_0 + \sum_{i=1}^{T-1} \omega(i; b_T)(\hat{\Gamma}_i + \hat{\Gamma}_i') \tag{9}$$

where $\hat{\Gamma}_j = T^{-1} \sum_{t=j+1}^{T} \hat{f}_t \hat{f}_{t-j}'$, $\hat{f}_t = f(v_t, \hat{\theta}_T)$, $\omega(.)$ is known as the *kernel*, and b_T is known as the *bandwidth*. The kernel and bandwidth must satisfy certain restrictions to ensure $\hat{S}_{HAC}$ is both consistent and positive semi–definite. As an illustration, Newey and West [13] propose the use of the kernel $\omega(i, b_T) = \{1 - i/(b_T + 1)\}\mathcal{I}\{i \leq b_T\}$ where $\mathcal{I}\{i \leq b_T\}$ is an indicator variable that takes the value of one if $i \leq b_T$ and zero otherwise. This choice is an example of a truncated kernel estimator because the number of included autocovariances is determined by b_T. For consistency, we require $b_T \to \infty$ with $T \to \infty$ but $b_T = o(T^{1/2})$. Various choices of kernel have been proposed and their properties analyzed: while theoretical rankings are possible, the evidence suggest that the choice of b_T is a far more important determinant of finite sample performance. Andrews [2] and Newey and West [14] propose data-based methods for the selection of b_T. See [6][Section 3.5] for further discussion.

(ii) Dependence of $\hat{S}_T$ on $\hat{\theta}_T$: As is apparent from the above discussion, the calculation of a consistent estimator for $S(\theta_0)$ requires knowledge of a (consistent) estimator of θ_0. Therefore, in order to calculate a GMM estimator that attains the efficiency bound, a multi-step procedure is used. On the first step, GMM is performed with an arbitrary weighting matrix; this preliminary estimator is then used in the calculation of $\hat{S}_T$. On the second step, GMM estimation is performed with $W_T = \hat{S}_T^{-1}$. For obvious reasons, the resulting estimator is commonly referred to as the *two-step GMM estimator*. Instead of stopping after just two steps, the procedure can be continued so that on the ith step the

GMM estimation is performed using $W_T = \hat{S}_T^{-1}$, where $\hat{S}_T$ is based on the estimator from the $(i-1)^{th}$ step. This yields the so-called *iterated GMM estimator*.

While two-steps are sufficient to attain the efficiency bound, simulation evidence suggests that there are often considerable gains to iteration in the sense of improvements in the quality of asymptotic theory as an approximation to finite sample behaviour; see [6][Chapter 6]. Notwithstanding these improvements, there is evidence that the quality of the asymptotic approximation may be poor in some circumstances of interest and this motivated the development of alternative methods for attaining the efficiency bound. One such method is the *continuous-updating GMM (CUGMM) estimator* proposed by Hansen, Heaton and Yaron [8]. The CUGMM estimator is the value of θ that minimizes $g_T(\theta)'S_T(\theta)^{-1}g_T(\theta)$ where $S_T(\theta)$ is a (matrix) function of θ such that $S_T(\theta) \xrightarrow{p} S(\theta)$ uniformly in θ. While the CUGMM estimator has the same limiting distribution as the iterated GMM estimator, Newey and Smith [12] and Anatolyev [1] demonstrate analytically that the continuous-updating estimator can be expected to exhibit lower finite sample bias than its two-step counterpart; see also [3].

The foregoing results are predicated on the assumption that (1) represents valid information about θ_0, and, in this sense, that the model is correctly specified. In practice, it is desirable to assess whether the data appear consistent with the restriction implied by the population moment condition. As noted above, if $p = q$ then the first order conditions force $g_T(\hat{\theta}_T) = 0$ irrespective of whether (1) holds and so the latter cannot be tested directly using the estimated sample moment, $g_T(\hat{\theta}_T)$. However, if $q > p$ then $g_T(\hat{\theta}_T) \neq 0$ because

GMM estimation only imposes the identifying restrictions and ignores the overidentifying restrictions. The latter represent $q-p$ restrictions which are true if (1) is itself true. To motivate the most commonly applied test statistic, it is useful to recall two aspects of our discussion above: (a) the GMM minimand measures the distance of $g_T(\theta)$ from zero; (b) the estimated sample moment contains information about overidentifying restrictions. Combining (a) and (b), it can be shown that GMM minimand evaluated at $\hat{\theta}_T$ is a measure of how for the sample is from satisfying the overidentifying restrictions. This leads to *overidentifying restrictions test* statistic,

$$J_T \;=\; T g_T(\hat{\theta}_T)' \hat{S}_T^{-1} g_T(\hat{\theta}_T)$$

which converges to a χ^2_{q-p} distribution under the null hypothesis that (1) holds. Notice that J_T is conveniently calculated as the sample size times the two-step (or iterated) GMM minimand evaluated at the associated estimator. The overidentifying restrictions test is the standard model diagnostic within the GMM framework. Nevertheless, J_T does not have power against all alternatives of interest: Ghysels and Hall [5] show that J_T has no (local) power against parameter variation and so it is prudent to complement the overidentifying restrictions test with tests of structural stability; see [6][Section 5.4].

As noted above, there is evidence that in some cases of interest, the aforementioned (standard) asymptotic theory can provide a poor approximation to the finite sample performance of the GMM estimator and its associated inference techniques. This has prompted researchers to develop ways of ameliorating these deficiencies. The proposed methods fall into three basic categories: (i) methods for *moment selection* which are designed to deter-

mine the "best" set of moments to use with a view to then basing inference on the standard asymptotic theory described above; (ii) resampling methods such as the *bootstrap* that are valid under similar assumptions to the standard theory but aim to provide more accurate approximations to finite sample behaviour; (iii) *alternative asymptotic theories* that are based on different assumptions to the standard framework and are designed to provide more accurate approximations in certain scenarios of interest, such as *weak identification*. For further discussion of (i)-(iii), see [6][Chapters 6,7 and 8].

References:

[1] Anatolyev, S. (2005) GMM, GEL, Serial correlation, and asymptotic bias. *Econometrica* **73**,983-1002.

[2] Andrews, D. (1991) Heteroscedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* **59**, 817–858.

[3] Antoine, B., Bonnal, H., Renault, E. (2007) On the efficient use of the informational content of estimating equations: implied probabilities and Euclidean Empirical Likelihood. *Journal of Econometrics* **138**, 461-487.

[4] Chamberlain, G. (1987) Asymptotic efficiency in estimation with conditional moment restrictions. *Journal of Econometrics* **34**, 305–334.

[5] Ghysels, E., Hall, A. (1990) Are consumption based intertemporal asset pricing models structural? *Journal of Econometrics* **45**, 121–139.

[6] Hall, A. (2005),*Generalized Method of Moments.* Oxford University Press, Oxford, UK.

[7] Hansen, L. (1982) Large sample properties of Generalized Method of Moments estimators. *Econometrica* **50**, 1029–1054.

[8] Hansen, L., Heaton, J., Yaron, A. (1996) Finite sample properties of some alternative GMM estimators obtained from financial market data. *Journal of Business and Economic Statistics* **14**, 262–280.

[9] Hansen, L., Hodrick, R. (1980) Forward exchange rates as optimal predictors of future spot rates. *Journal of Political Economy*, **87**, 829-853.

[10] Hansen, L., Singleton, K. (1982) Generalized instrumental variables estimation of nonlinear rational expectations models. *Econometrica* **50**, 1269–1286.

[11] Melino, A., Turnbull, S. (1990) Pricing foreign currency options with stochastic volatility. *Journal of Econometrics* **45**, 239–266.

[12] Newey, W., Smith, R. (2004) Higher order properties of GMM and generalized empirical likelihood estimators. *Econometrica* **72**, 219-255.

[13] Newey, W., West, K. (1987) A simple positive semi–definite heteroscedasticity and autocorrelation consistent covariance matrix. *Econometrica* **55**,703–708.

[14] Newey, W., West, K. (1994) Automatic lag selection in covariance matrix estimation. *Review of Economic Studies* **61**, 631–653.

[15] Wooldridge, J. (2002) *Econometric Analysis of Cross Section and Panel Data.* MIT Press, Cambridge, MA, USA.

View publication stats