Revista de Administração Contemporânea
Journal of Contemporary Administration

ASSOCIAÇÃO NACIONAL DE PÓS-GRADUAÇÃO E PESQUISA EM ADMINISTRAÇÃO
ANPAD

e-ISSN: 1982-7849

Methodological Article

A Tutorial on the Generalized Method of Moments (GMM) in Finance

Um Tutorial sobre o Método Generalizado dos Momentos (GMM) em Finanças

OPEN DATA

Open Access

Alan de Genaro*[1]
Paula Astorino[1]

ABSTRACT

Context: empirical problems in which the researcher is faced with a model that is partially specified. In these cases, the GMM method is the natural alternative for estimating the parameters of interest. **Objective:** the goal of this paper is to offer a tutorial that allows the researcher to understand both the theory and empirical aspects of the GMM method. **Methods:** we discuss the GMM concepts, forms of estimation, and limitations associated with the method. As a way of illustrating the method, we use two applications in the area of empirical finance. The first application is the estimation of the parameters of a consumption-based asset pricing models; the second is the estimation of the parameters of the evolution of the interest rate in continuous time. The data and codes in R are provided as online appendices. **Conclusion:** the GMM method can be used in problems where other methods such as maximum likelihood are not feasible, or even when the researcher wants to estimate a model partially specified.

Keywords: GMM; asset pricing; interest rate.

RESUMO

Contexto: problemas empíricos em que o pesquisador se depara com um modelo que seja parcialmente especificado. Nestes casos, o método GMM é a alternativa natural para estimação dos parâmetros de interesse. **Objetivo:** o propósito deste artigo é oferecer um tutorial que permita ao pesquisador compreender os aspectos conceituais e práticos do método GMM. **Métodos:** são apresentadas as características, formas de estimação, e algumas limitações associadas ao método em duas aplicações na área de finanças empíricas. A primeira aplicação é para a estimação dos parâmetros dos modelos de apreçamento de ativos baseados em consumo; o segundo é a estimação dos parâmetros do modelo para descrever a taxa de juros em tempo contínuo. Os dados e o código em R são fornecidos nos apêndices on-line. **Conclusão:** o método GMM pode ser utilizado em problemas onde outros métodos como máxima verossimilhança não são factíveis, ou ainda quando se deseja estimar um modelo parcialmente especificado.

Palavras-chave: GMM; apreçamento de ativos; taxa de juros.

* Corresponding Author.
1. Fundação Getulio Vargas, Escola de Administração de Empresas de São Paulo, São Paulo, SP, Brazil.

Cite as: Genaro, A., & Astorino, P. (2022). A tutorial on the generalized method of moments (GMM) in finance. *Revista de Administração Contemporânea, 26*(Sup. 1), e210287. https://doi.org/10.1590/1982-7849rac2022210287.en

Published as Early Access: January 24, 2022.
Assigned to this issue: August 08, 2022.

Note: This text is translated from the original Portuguese version, which can be accessed here.

JEL Code: C51, G12, C58.

Editor-in-chief: Marcelo de Souza Bispo (Universidade Federal da Paraíba, PPGA, Brazil)
Reviewers: Aureliano Angel Bressan (Universidade Federal de Minas Gerais, CEPEAD, Brazil)
André Alves Portela (Universidade Federal de Santa Catarina, Brazil)
Peer Review Report: The Peer Review Report is available at this external URL.

Received: November 07, 2021
Last version received: November 29, 2021
Accepted: January 06, 2022

# of invited reviewers until the decision:

	1	2	3	4	5	6	7	8	9
1st round	✓	✓							

INTRODUCTION

Hansen (1982) introduced[1] the generalized method of moments (GMM) and made significant contributions to empirical research in finance, in particular in asset pricing. This method has since been used to estimate parameters in models for which only moments are available.

Historically, as Jagannathan, Skoulakis and Wang (2002) recall, before the rise of GMM, the main method used to estimate asset pricing was the maximum likelihood estimator (MLE). However, asset pricing parameter estimation with MLE presented many difficulties, limiting its usefulness. First, it was necessary to derive a test to see whether the model was poorly specified. In addition, adopting some form of linearization of the model was common since asset pricing models are generally nonlinear. Finally, the researcher needed to define hypotheses about the joint distribution of the model's variables. When the hypothesized distribution is not corroborated by the data, the estimated parameters could be biased, even in large samples. As Cameron and Trivedi (2005) explain, we know that MLE is asymptotically efficient under certain conditions.

The main condition for asymptotic efficiency of MLE is that the likelihood function is correctly specified. This means that the true data generating process (DGP) is known, and the parameters are perfectly identified.

GMM, in turn, appears to be a viable alternative, since it allows the estimation of model parameters, linear or not, from a partially specified model. It thus circumvents the MLE method's difficulties. The convenience of not assuming a priori a joint distribution of variables, and the generality of GMM, which allows it to be used in many problems, are briefly the two main reasons why GMM has become so popular in empirical finance re-search.

Despite its advantages, GMM may not be as efficient as the MLE method. That is, the standard error of the estimates can be much larger than that found using MLE, for models that can also be estimated by MLE.

Given its broad applicability, this tutorial article aims to introduce and explain GMM to researchers, and to help them use it. Our first empirical example deals with asset pricing and its relationship with an individual's aggregate lifetime consumption. The second exam-ple describes short-term interest rates in continuous time. Although they may seem to be restricted to the field of finance, or even dated, these models are still used and adapted to reflect preferences of individuals, whose consumption and future expectations take into account values, such as social well-being, or reducing the environmental impacts of econom-ic activities.

FROM THE METHOD OF MOMENTS TO THE GMM

The modern treatment of GMM was formalized by Hansen (1982) based on the concepts and methods of Amemiya (1977) and Gallant (1977).

The starting point for introducing GMM is to understand how the method of moments works, and how GMM actually changes to handle the more general case. In both cases, the formulation begins with a moment condition defined by economic or statistical theory. For example, for the first moment of a random variable X, with population mean μ, we have the following moment condition:

$$\mathbb{E}[X] - \mu = 0 \tag{1}$$

In turn, given a sample $x_1, \dots, x_T$, we can rewrite Equation (1) through its empirical counterpart:

$$\sum_{t=1}^{T} \frac{x_t}{T} - \mu = 0 \tag{2}$$

Hence, the moment method estimator for μ is obtained by:

$$\hat{\mu}_{MM} = \sum_{t=1}^{T} \frac{x_t}{T} \tag{3}$$

The construction above can be extended to the linear regression model $y = \boldsymbol{x}'\beta + u$, where β is a vector of dimension K×1. Suppose the error u, conditioned to the regressors, has a mean of zero. This generates the following conditional momentum condition:

$$\mathbb{E}[u|\mathbf{x}] = 0 \tag{4}$$

Combining the foregoing with the law of iterated expectations,[2] we can write unconditional expectations:

$$\mathbb{E}[u\mathbf{x}] = 0 \tag{5}$$

$$\mathbb{E}[\mathbf{x}(y - \mathbf{x}'\beta)] = 0 \tag{6}$$

Using the empirical counterpart of the moment condition, we can write (6) as:

$$\frac{1}{N}\sum_{i=1}^{N} \mathbf{x}_i(y_i - \mathbf{x}_i'\beta) = 0 \tag{7}$$

Solving (7), we obtain the moment estimator:

$$\hat{\beta}_{MM} = \left(\sum_i \boldsymbol{x}_i \boldsymbol{x'}_i\right)^{-1} \sum_i \boldsymbol{x}_i y_i \quad (8)$$

Thus, the OLS method can be seen as a particular case of the method of moments. Further, the example above could be generalized to handle the case where x could be correlated with u. Then, the OLS estimator would be inconsistent. However, with the method of moments, it would be enough to include a new moment condition based on the existence of instruments z that are uncorrelated with u, that is, $\mathbb{E}[u|\mathbf{z}] = 0$.

To find the estimator, we simply set dim(z)=K, that is, that the number of instruments is exactly equal to the number of regressors. In this case $\hat{\beta}_{MM} = \left(\sum_i \boldsymbol{z}_i \boldsymbol{x'}_i\right)^{-1} \sum_i \boldsymbol{z}_i y_i$ which represents the linear method using instrumental variables.

The idea behind using instrumental variables is to find a variable that carries only the 'good' variation contained in x and helps explain y, albeit indirectly.

This instrument must pass the test of relevance. That is, when regressing x on the other variables, the coefficient of z must be statistically significant. In other words, z is significant in explaining x even considering the effects of other regressors.

The instrument must also satisfy the exclusion condition: $cov(z,u) = 0$. Since the error term u, by definition, is unobservable, no test can guarantee that the instrument meets this condition. Therefore, it is up to the researcher to theoretically justify using the chosen instrument to address the problem. In the end, given the moment conditions above, the instrumental variable method can also be seen as a specific case of GMM.

Hansen's (1982) contribution was to generalize the concept above to more general moment conditions than that expressed in Equation (1). Accordingly, the general case can be written using the set of R moment conditions:

$$\mathbb{E}\{f(x_t, z_t, \theta)\} = 0 \quad (9)$$

where $f(\cdot)$ is an R×1 vector of real functions, θ is a K-dimensional vector containing all parameters of interest, x_t is a vector with observable variables, and z_t is a vector of instruments.

The problem that GMM proposes to solve is how, from a sample $x_1, \dots, x_T$, of population X, the parameter θ can be estimated, having specified only the moment condition described in (4).

Again, the empirical counterpart of the R moment conditions defined in (9) is:

$$g_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} f(x_t, z_t, \theta) \quad (10)$$

To identify the parameters in this model, we must see how R and K are related. When $R = K$, the model is exactly identifiable. We can obtain the estimator of the moments method, $\hat{\theta}_{MM}$, by solving $g_T(\hat{\theta}_{MM}) = 0$. On the other hand, when $R < K$, the model is under-identified, and it is impossible to solve $g_T(\hat{\theta}_{MM}) = 0$. Finally, when $R > K$ the model is over-identified, in which case the estimator of the generalized method of moments, $\hat{\theta}_{GMM}$ becomes important.

When the model is over-identified ($R > K$), we have more equations R than unknown K. Therefore, it is impossible to find a vector $\hat{\theta}$ to make the entire set of R moments exactly equal to zero. The solution to this problem consists of finding the vector of parameters $\hat{\theta}$ that puts $g_T(\hat{\theta})$ as close to zero as possible (this distance concept depends on the norm adopted; here, the Euclidean norm is adopted, L^2). Additionally, since not all R moment conditions can be satisfied, we adopt a positive definite symmetric matrix W_T to weight these conditions properly. The method thus seeks to find the vector $\hat{\theta}$ such that $g_T(\hat{\theta}_{GMM}) \cong 0$. The GMM estimator for θ seeks to minimize the following quadratic form:

$$\hat{\theta}_{GMM} = \underset{\theta}{\operatorname{argmin}}\, g_T(\theta)' W_T g_T(\theta) \quad (11)$$

Usually, the solution to (11) is not available analytically, so it ends up being found using numerical methods[3].

Regarding desirable properties of the estimator, any positive symmetric W_T weighting matrix meets the conditions to guarantee that the estimator is consistent. Consequently, there are as many GMM estimators as there are choices for the W_T weighting matrix. However, estimator efficiency is not obtained for any arbitrary W_T matrix, only for the case where $W_T = S^{-1}$, when S^{-1} represents[4] the long-run covariance matrix of the moments. These, in turn, must also be estimated through $\hat{S}_T(\theta)$.

Estimating $\hat{\theta}_{GMM}$ presents the challenge of how to construct a consistent estimator for $S(\theta)$, as well as the dependent relationship between $\hat{S}_T(\theta)$ and $\hat{\theta}_{GMM}$. If we initially assume that θ is known, the optimal weighting matrix $\hat{S}_T^{-1}(\theta)$ can be estimated using the method of heteroskedasticity autocorrelation covariance (HAC), which is robust to the presence of heteroskedasticity and autocorrelation.

Solving Equation (11) depends on the method adopted to estimate the weight matrix $\hat{S}_T^{-1}(\theta)$. This, however, also depends on $\hat{\theta}_{GMM}$. This dependence is undesirable since inference on $\hat{\theta}_{GMM}$ relies on how we represent the weight matrix W_T. To address this, several methods have been developed to try to solve (11). The first alternative, the two-step method (two-step GMM) initially defines an arbitrary matrix W_T, often represented by the identity matrix, and then obtains the value of $\hat{\theta}^1$. Once we have this value, we use it in the functional form of $\widehat{W}_T = \hat{S}_T^{-1}(\hat{\theta}^1)$ to find the solution $\hat{\theta}_{GMM}$.

Although, in theory, the two-step GMM has good statistical properties, empirical studies show that its behavior in small samples tends to be unsatisfactory. To circumvent its limitations, Hansen, Heaton and Yaron (1996) propose two new methods for estimating $\hat{\theta}_{GMM}$.

The first alternative involves a natural evolution of the two-step method, in which the process is repeated n times until it reaches a convergence criterion. This is the iterative method (iterative GMM). A second alternative is the continuously updated estimator method (CUE GMM), in which the matrix $\hat{S}_T$ and $\hat{\theta}_{GMM}$ are estimated simultaneously. In addition to improved performance in small samples, another advantage is that this method is invariant to certain types of transformations and normalizations that can be performed with the data (Hall, 2005). On the other hand, the optimization problem to be solved is highly nonlinear, and therefore extremely sensitive to initial conditions. For this reason, the initial conditions of the CUE estimator usually use a result of earlier estimation via the two-step method or the iterative method.

Finally, GMM minimizes a quadratic function that contains the moment conditions properly weighted by W_T; if these moment conditions are correct, i.e., $\mathbb{E}\{f(x_t, z_t, \theta)\} = 0$, then $g_T(\hat{\theta}_{GMM}) \cong 0$. This result naturally leads to a model specification test, known as the J test, defined as:

$$J = \mathrm{T}^{-1}\, g_T(\hat{\theta})' \widehat{W}_T g_T(\hat{\theta}) \sim \chi^2_{r-q} \tag{12}$$

Rejecting the test indicates that the model is poorly specified, because some moments are not statistically equal to zero. That is, the model is rejected, since the moment condition is not valid.

The number of moments to include in the specification is not easily determined. Nevertheless, Hall (2005) recommends that for small samples, with up to 100 observations, we can include up to five moments more than the number of parameters to be estimated.

APPLICATIONS

Having introduced the main concepts of GMM, in this section we present two applications based on well-known studies.

For our data, we opt to make use of public data from previous research (Chan, Karolyi, Longstaff, & Sanders, 1992; Verbeek, 2004). This facilitates the reproduction of analysis and comparison of results. In addition, this tutorial aligns with the concepts and suggestions of Martins (2021), observing best practices in research and embracing the concept of open data.

We invite interested readers to use the *R* codes available in the online appendix to replicate all results presented in this article and adapt them in their own research.

Estimation of consumption-based asset pricing models (CCAPM)

In their groundbreaking article Hansen and Singleton (1982) develop the GMM model from the formal construction of Lucas (1978). This model aims to explain aggregate movements of consumption and asset returns. The framework to support using aggregated data is based on positing a single representative agent who wants to choose the optimal consumption path by maximizing the present value of the expected utility of consumption:

$$max\ \mathbb{E} \sum_{s=0}^{\infty} \left[\beta^s U(C_{t+s}) \middle| \mathcal{F}_t\right] \tag{13}$$

subject to budget constraint:

$$C_{t+s} + q_{t+s} = w_{t+s} + (1 + r_{t+s}) q_{t+s-1} \tag{14}$$

where F_t represents the set of information available at time t, q_{t+s} is the individual's wealth at the end of period $t + s$, r_{t+s} is the rate of return obtained from investing in a set of assets, and w_{t+s} represents labor income.

The agent can invest in N risky assets, which have a gross return of $r_{i,t}$, and in a risk-free asset that has a rate of return $r_{f,t}$. Solving the consumption and intertemporal asset allocation problem involves determining the first order condition of Equation (13), subject to constraint (14), producing the following Euler equation:

$$\mathbb{E}\left\{\beta \frac{U'(C_{t+1})}{U'(C_t)}(1 + r_{f,t+1}) - 1 \,\middle|\, \mathcal{F}_t\right\} = 0 \tag{15}$$

$$\mathbb{E}\left\{\beta \frac{U'(C_{t+1})}{U'(C_t)}(r_{i,t+1} + r_{f,t+1}) \mid \mathcal{F}_t\right\} = 0, \text{for } i = 1, \dots, N \quad (16)$$

where $U'(\cdot)$ is the consumer's marginal utility.

Euler equation imposes restrictions on the joint movement between consumption and the price of financial assets, helping us understand the behavior of these two variables.

In turn, the moment condition to estimate the parameters via GMM requires adopting a functional form to represent this agent's utility. The simplest case involving consumption-based asset pricing models is where the utility function belongs to the class of functions with constant relative risk aversion (CRRA):

$$U(C_t) = \begin{cases} \frac{c_t^{1-\gamma}}{1-\gamma} \; with\; \gamma \neq 1 \\ \ln(c_t) \; with\; \gamma = 1 \end{cases} \quad (17)$$

where the parameter γ represents the investor's risk aversion coefficient.

So, using Equation (17) in Expression (16) we obtain:

$$\mathbb{E}\left\{\beta\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma}(1 + r_{i,t+1}) - 1 \mid \mathcal{F}_t\right\} = 0 \quad (18)$$

Assuming that $z_t \in \mathcal{F}_t$, the law of iterated expectations allows us to write the unconditional moment expectation as:

$$\mathbb{E}\left\{\left[\beta\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma}(1 + r_{i,t+1}) - 1\right] z_t\right\} = 0 \text{ for } i = 1, \dots, \quad (19)$$

In general, the instrumental variable z_t may include a constant, equivalent to the unconditional expectation of the Euler equation, and macroeconomic variables used to describe the set of information available to the representative agent at the moment when he makes his decisions on optimal consumption and investment.

Initially using $z_t = \{1\}$ as an instrument, Equations (18) and (19) represent a set of N+1 moment conditions that allow us to identify the parameters $\theta = (\beta, \gamma)'$. We can write them in the general form of the GMM model:

$$\mathbb{E}[f(\theta, x_t, z_t)] \equiv \mathbb{E}\begin{bmatrix} \beta\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma}(1 + r_{f,t+1}) - 1 \\ \beta\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma}(r_{1,t+1} - r_{f,t+1}) \\ \vdots \\ \beta\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma}(r_{N,t+1} - r_{f,t+1}) \end{bmatrix} \quad (20)$$

To estimate the CCAPM model using GMM, we use a set of 10 portfolios. These portfolios are built based on the market value of shares (size-based) traded on the NYSE. Portfolio 1 contains the return of the smallest 10% of companies listed on the NYSE, while portfolio 10 is formed by the largest 10% of companies listed on the NYSE. The risk-free asset is represented by the three-month Treasury bill (T-bill) rate.

We use monthly data from February 1959 to November 1993. Portfolio data are obtained from the Center for Research in Security Prices (CRSP). Consumption data consist of the amounts spent on household consumption of non-durable goods and services in the US economy.

Results of the estimation by GMM (Tab. 1) are as follows:

Table 1 shows that the coefficients are statistically significant in all forms of estimation. While the estimate of the parameter β is reasonable, according to economic theory, the value of the risk aversion coefficient γ is excessively high. However, the J test does not reject that the model is poorly specified for any of the estimation methods.

Table 1. CCAPM Model.

	Two-step GMM		Iteratitve GMM		CUE GMM	
	Coefficient	Std. error	Coefficient	Std. error	Coefficient	Std. error
β	0.81***	0.12	0.83***	0.12	0.70***	0.12
γ	62.08*	33.79	57.40*	34.22	96.19***	31.62
J-stat	4.54	p-value 0.87	5.76	p-value 0.76	5.14	p-value 0.82

Note. Models estimated with the instrument $z_t = \{1\}$. Standard error is calculated using the HAC estimator of Andrews (1991). 'Two-step' is the two-step estimation method proposed by Hansen (1982). 'CUE' and 'Iterative' are, respectively, the continuous updating and the interactive methods proposed by Hansen et al. (1996). *** denotes significance at 1%. ** denotes significance at 5%. * denotes significance at 10%.

In an attempt to obtain a more reasonable parameter γ, a second exercise uses other variables as instruments — for example, $z_t = \{1, C_t/C_{t-1}\}$. We now have 2(N+1) moment conditions, as shown below:

$$\mathbb{E}[f(\theta, x_t, z_t)] \equiv \mathbb{E}\begin{bmatrix} \beta\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma}(1 + r_{f,t+1}) - 1 \\ \beta\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma}(r_{1,t+1} - r_{f,t+1}) \\ \vdots \\ \beta\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma}(r_{N,t+1} - r_{f,t+1}) \\ \left(\frac{C_t}{C_{t-1}}\right)\left[\beta\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma}(1 + r_{f,t+1}) - 1\right] \\ \left(\frac{C_t}{C_{t-1}}\right)\left[\beta\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma}(r_{1,t+1} - r_{f,t+1})\right] \\ \vdots \\ \left(\frac{C_t}{C_{t-1}}\right)\left[\beta\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma}(r_{N,t+1} - r_{f,t+1})\right] \end{bmatrix} \quad (21)$$

Simplifying the notation in (21) using the Kronecker product gives us another way to represent the moment conditions:

$$\mathbb{E}\left\{\left[\beta\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma}(1 + r_{i,t+1}) - 1\right] \bigotimes z_t\right\} = 0 \text{ for } i = 1, \dots, N + 1 \quad (22)$$

The GMM estimation with the same three criteria, using the robust error, gives the following estimation results (Tab. 2).

Again, the parameter β proves to be economically and statistically significant. However, the parameter γ ends up being estimated with a negative value, which is inconsistent. In this case, the J test correctly rejects the proposed model specification.

However, having observed that the risk aversion parameter found for the American economy was excessively high, Mehra and Prescott (1985) coined the expression 'equity premium puzzle' (EPP), and opened a fruitful new area of research. Since then, research has undertaken to modify the original problem and find the best value for the risk aversion parameter. Much of this research relies on proposing alternative forms for the utility function, as discussed in detail by Campbell (2018).

Table 2. CCAPM Model,

	Two-step GMM		Iterative GMM		CUE GMM	
	Coefficient	Std. error	Coefficient	Std. error	Coefficient	Std. error
β	0.91***	0.03	0.98***	8.83E-04	0.98***	8.74E-04
γ	33.9***	8.28	-1.25***	0.38	-1.23***	0.372
J-stat	133	p-value 0.00	30.2	p-value 0.067	30.2	p-value 0.067

Note. Models estimated with $z_t = \{1, C_t/C_{t-1}\}$. The standard error was calculated using the HAC estimator of Andrews (1991). 'Two-step' is the two-step estimation method proposed by Hansen (1982). 'CUE' and 'Iterative' are, respectively, the continuous updating and the interactive methods proposed by Hansen et al. (1996). *** denotes significance at 1%. ** denotes significance at 5%. * denotes significance at 10%.

For the Brazilian economy, some authors have sought to empirically evaluate the existence of the EPP, among them Issler and Piqueira (2002) and Cysne (2006). Unlike those from the US, Brazilian data on different functional forms for the utility function do not support the existence of the EPP, except as found by Cysne (2006).

Another way of using Equations (18) and (19) is to compare expected returns predicted by the model with the return realized for each portfolio. To do this, we use the concept of the stochastic discount factor (SDF) identified by this model:

$$m_{t+1}(\theta) = \beta\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma} \quad (23)$$

Advantages of using the SDF include: (a) the ability to describe a large class of asset pricing models with just one representation; and (b) a general and unified way to study and analyze variation in the expected return of different assets. Therefore, using the SDF representation, we can obtain the decomposition of the expected return for this model. Starting from:

$$\mathbb{E}\{m_{t+1}(\theta)\,(r_{i,t+1} + r_{f,t+1}) \mid \mathcal{F}_t\} = 0, \text{for } i = 1, \dots, N \quad (24)$$

we obtain:

$$\mathbb{E}[(r_{i,t+1} + r_{f,t+1})] = -\frac{\operatorname{cov}\left(m_{t+1}(\theta), (r_{i,t+1} + r_{f,t+1})\right)}{\mathbb{E}[m_{t+1}(\theta)]}, \text{for } i = 1, \dots, N \quad (25)$$

One implication of Equation (25) arising from consumption-based models is that assets that have positive covariance with the consumption growth rate must offer a higher rate of return so that, in equilibrium, the investor chooses these assets for his portfolio.

Additionally, we can investigate the model's quality by calculating the pricing error, which consists of comparing excess returns predicted by the model to those observed.

Figure 1 shows the excess realized return and the excess return predicted by the model for all 10 portfolios. For model-predicted returns, we follow Cochrane (1996) suggestion and use the parameters estimated in one-step procedure, having assumed that W_T is the identity matrix.

An analysis of Figure 1 confirms that the model's performance is not satisfactory, since it is unable to explain cross-sectional variation observed in excess return. In particular, the portfolio formed by the smallest 10% of firms is the one that most deviates from the 45° line. Therefore, the consumption-based asset pricing model is also unable to solve the size-effect pointed out by Banz (1981).

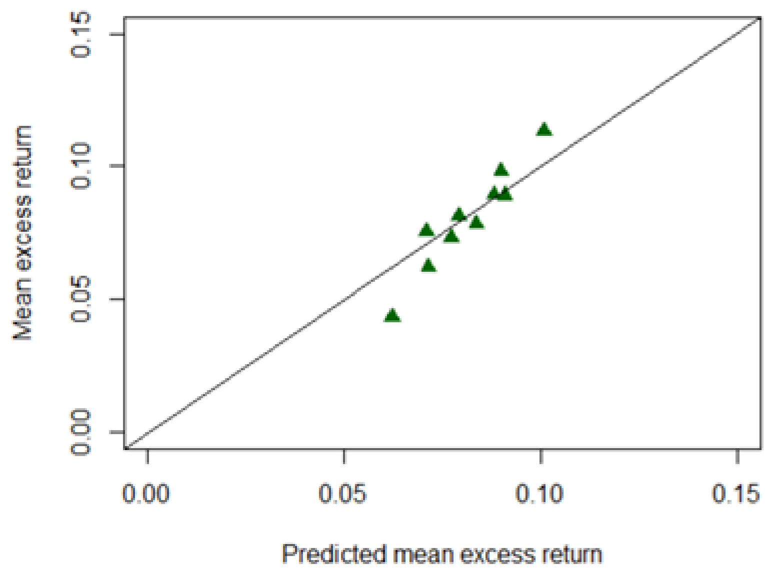

Figure 1. Dispersion between observed excess return and the excess return predicted by the CCAPM for 10 stock portfolios (size-based).

Interest rate model estimation

A second use of the GMM method is that of Chan, Karolyi, Longstaff and Sanders (1992), who seek to estimate the parameters of the stochastic differential equation (SDE) used to describe the dynamics of the short-term interest rate in continuous time:

$$dr_t = \kappa(\theta - r_t)dt + \sigma r_t^{\gamma} dW_t \quad (26)$$

where the parameter κ represents the speed with which the rate moves toward the long-term mean and θ the long-term mean.

Or equivalently,

$$dr_t = (\alpha - \beta r_t)dt + \sigma r_t^{\gamma} dW_t \quad (27)$$

In this representation, $\alpha = \kappa\theta$ and $\beta = -\kappa$. Specification (26) encompasses the following models, according to the restriction in the parameters: (a) Brownian motion with drift ($\beta = 0$ and $\gamma = 0$); (b) Ornstein-Uhlenbeck (or Vasicek) ($\gamma = 0$) ; and (c) Cox-Ingersoll-Ross ($\gamma = 1/2$).

The derivation of the moment conditions uses the Euler-Maruyama discretization to obtain the discrete-time version of Equation (27), thus:

$$r_{t+\Delta t} - r_t = (\alpha - \beta r_t)\Delta t + \sigma r_t^{\gamma} \sqrt{\Delta t}\epsilon_{t+\Delta t} \quad (28)$$

with

$$\mathbb{E}[\epsilon_{t+1}|\mathcal{F}_t] = 0 \text{ and } \mathbb{E}[\epsilon_{t+1}^2|\mathcal{F}_t] = \sigma^2 r_t^{2\gamma}$$

and term $\epsilon_{t+\Delta t}$ describes a standardized normal random variable.

The MLE method cannot be used to estimate the model, since the distribution of increments depends on the value of γ. When $\gamma = 0$, the increments have a normal distribution, and when $\gamma = 1/2$ they have a gamma distribution.

In turn, GMM provides the econometric conditions necessary to estimate the models, since (28) enables us to write the moment conditions. Chan et al. (1992) use $z_t = (1, r_t)$ as instruments. With this, we can write the following moment conditions:

$$\mathbb{E}[f(\theta, x_t, z_t)] \equiv \mathbb{E}\begin{bmatrix} \epsilon_{t+1} \\ \epsilon_{t+1} r_t \\ \epsilon_{t+1}^2 - \sigma^2 r_t^{2\gamma} \Delta t \\ (\epsilon_{t+1}^2 - \sigma^2 r_t^{2\gamma} \Delta t) r_t \end{bmatrix}$$

Once the model is exactly identifiable, we obtain the GMM estimator for $\theta = (\alpha, \beta, \sigma, \gamma)'$ by solving $g_T(\hat{\theta}_{GMM}) = 0$. The weight matrix W_T becomes redundant.

We use the same data as Chan et al. (1992): monthly data for the average bid-ask spread implicit in the 30-day T-bill for the period from June 1964 to November 1989, available from CRSP. Since the data are monthly, the interval is defined as $\Delta t = 1/12$.

Following are the results (Tab. 3).

Table 3. Unrestricted CKLS model.

	Coefficient	Standard error
α	0.042**	0.019
β	-0.607*	0.350
σ	1.324	0.956
γ	1.505***	0.279

Note. The standard error is calculated using the HAC estimator of Andrews (1991). *** denotes significance at 1%. ** denotes significance at 5%. * denotes significance at 10%.

These results are similar to those of Chan et al. (1992). In particular, the coefficient γ is approximately equal to 1.5 and significant at 1%. Since the model is exactly identified, the J test is equal to zero by construction.

A second exercise consists of estimating the restricted model CIR SR: we impose $\gamma = 1/2$, and estimate the model proposed by Cox, Ingersoll and Ross (1985). In this case, we have an over-identified model, and the weighting matrix W_T, becomes relevant (Tab. 4).

The J test is rejected at a significance level of 5%, indicating that the model is poorly specified. This is an intriguing result, since the unconstrained model estimates the parameter $\gamma \approx 1{,}5$.

Table 4. Restricted CKLS model — Two-step GMM.

	Coefficient	Standard error
α	0.023	0.016
β	-0.269	0.287
σ	0.082***	0.007
γ	0.5	-
J-stat	5.13	p-value: 0.023

Note. Model estimated with $\gamma = 0.5$. Standard error is calculated using the HAC estimator of Andrews (1991). 'Two-step' is the two-step estimation method proposed by Hansen (1982). *** denotes significance at 1%. ** denotes significance at 5%. * denotes significance at 10%.

In addition to these two examples, a classic reference for the GMM estimation is by Arellano and Bond (1991), whose estimator[5] is known as GMM in differences (GMM-Dif) or AB-GMM. The estimator was developed in a dynamic panel data context, that is, a fixed-effect model including the dependent variable in the lagged form as a regressor:

$$y_{it} = X_{it}\beta + \rho y_{it-1} + \alpha_i + u_{it} \text{ for } t = 1, \ldots, T \text{ and } i = 1, \ldots ,$$

The inclusion of the lagged term y_{it-1} generally implies that the regressor is not strictly exogenous, since the lagged variable is usually correlated with the error term u_{it}, giving rise to the so-called dynamic panel bias, as Nickell (1981) argues. This bias makes estimation by OLS unfeasible and justifies using GMM estimators.

Despite its merits, the AB-GMM estimate has known deficiencies. In particular, (a) if the dependent variable y_{it} has significant persistence, $\rho \approx 1$; (b) in the case of panel data in which the time dimension T has many observations, the number of instruments required grows substantially, as $T/(T - 1)/2$ instruments are required.

LIMITATIONS OF GMM

Although theory provides the conditions to claim that the GMM estimator is consistent, efficiency and bias fundamentally depend on the choice of moment conditions and the choice of instruments.

Asymptotic properties are well documented in Hansen's (1982) seminal paper, but the properties for finite samples are not fully known. The choice of right instruments is an ongoing research agenda. So-called 'weak instruments,' that is, variables that are weakly correlated with the variables of interest, represent some of the difficulties that may arise in finite samples. In GMM, weak instruments correspond to the weak identification of some or all the model parameters.

Therefore, the researcher must be careful when selecting instruments. Stock, Wright and Yogo (2002) review recent developments in instrument choice and impacts on the quality of estimation in finite samples.

In addition to questions about the choice of instrument, a second limitation of GMM concerns the moment condition, in particular when: (a) the model moment condition cannot be obtained in an analytical form or presents a functional form that is too complex to evaluate; (b) the moment condition depends on the behavior of an unobservable latent variable; or (c) the moment condition is derived from censored variables, and therefore, they are partially observed by the researcher.

To overcome the problems described above, a new class of estimators is proposed, called the simulated moments method (SMM), developed from the GMM model. Early contributors include McFaden (1989) and Duffie and Singleton (1993).

Conceptually, SMM uses simulations to approximate the moment condition that is unavailable in the analytical form. Thus, it allows parameters to be estimated in a similar way to GMM. The challenge associated with the SMM model, compared to GMM, is that the process of estimating parameters becomes computationally intensive. For a set of simulations, the parameters must be re-estimated until convergence occurs.

Using simulations to approximate an analytical result that is difficult to obtain also led to the development of the simulated maximum likelihood method (SML). Pedersen (1995) and Brandt and Santa-Clara (2002) independently propose that the likelihood function can be numerically approximated through simulations. The advantage of simulation methods is the possibility of including complex structures in a relatively straightforward way, something that is not trivial when devising an analytical solution. For example, Genaro and Avellaneda (2018) estimate via SML the parameters of a continuous-time model in which jumps occur endogenously, which cannot be estimated by maximum likelihood because the intensity of the jump process is a latent variable.

FINAL REMARKS

One difficulty for a reader encountering GMM for the first time is that it needs the moment conditions to be defined. These are different for each model, and can be confusing. The purpose of this tutorial is to present the method, step by step.

Precisely because each model has a specific moment condition, we present two relevant cases in the empirical finance literature that use GMM. These examples can help consolidate the reader's understanding of the method.

To complement this, we also consider using GMM in the presence of instrumental variables, as well as the theoretical limitations of the method. These limitations have motivated the development of alternatives, such as the simulated moments method (SMM).

It is worth noting, again, that pricing models using the technique presented here are still widely applied. In the current market context, especially, it is extremely important to measure the impacts of purportedly sustainable investments made by corporations worldwide.

Finally, the present tutorial article offers a view of the GMM method in line with the attitude expressed by Hansen (2013), allowing the researcher "to do something without having to do everything."

NOTES

1. In 2013, Lars Peter Hansen was awarded the Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel for his contributions to research in robust econometric methods, in particular for formally introducing the principle of "doing something without doing everything," which motivated the development of GMM. Lars Hansen shared that year's Nobel Prize with Eugene Fama and Robert Shiller.
2. In general, if $\mathbb{E}[x_1|x_2] = 0$, then $\mathbb{E}[x_1 g(x_2)] = 0$ for every function g.
3. In this tutorial, we adopt the *gmm* package available in *R*. Due to the popularity of the method, standard statistical software includes GMM packages. Currently, *R* 's *gmm* package is being replaced by the package *momentfit*.
4. The fact that the matrix S is inverted shows that moments of greater variance end up having a lower weight in the solution of (9).
5. *R*'s dynpanel package allows the implementation of the estimator, as in Arellano and Bond (1991).

REFERENCES

Amemiya, T. (1977). The maximum likelihood and nonlinear three-stage least squares estimator in the general nonlinear simultaneous equation model. *Econometrica*, *45*(4), 955-968. https://doi.org/10.2307/1912684

Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, *59*(3), 817-858. https://doi.org/10.2307/2938229

Arellano, M., & Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *The Review of Economic Studies*, *58*(2), 277-297. https://doi.org/10.2307/2297968

Banz, R. (1981). The relationship between return and market value of common stocks. *Journal of Financial Economics*, *9*, 3–18. https://doi.org/10.1016/0304-405X(81)90018-0

Brandt, M. W., & Santa-Clara, P. (2002). Simulated likelihood estimation of diffusions with an application to exchange rate dynamics in incomplete markets. *Journal of Financial Economics*, *63*, 161–212. Retrieved from https://ideas.repec.org/a/eee/jfinec/v63y2002i2p161-210.html

Campbell, J. Y. (2018). *Financial decisions and markets: A course in asset pricing*. Princeton, NJ: Princeton University Press.

Cameron, A. C., & Trivedi, P. (2005). *Microeconometrics: Methods and applications*. New York: Cambridge University Press.

Chan, K. C., Karolyi, G. A., Longstaff, F. A., & Sanders A. B. (1992). An empirical comparison of alternative models of the short-term interest rate. *Journal of Finance*, *47*(3), 1209-1227. https://doi.org/10.1111/j.1540-6261.1992.tb04011.x

Cochrane, J. (1996). A cross-sectional test of an investment-based asset pricing model. *Journal of Political Economy*, *104*(3), 572-621. https://doi.org/10.1086/262034

Cox, J. C., Ingersoll, J. E., & Ross, S. A. (1985). A theory of the term structure of interest rates. *Econometrica*, *53*(2), 385-407. https://doi.org/10.2307/1911242

Cysne, R. (2006). Equity-premium puzzle: Evidence from Brazilian data. *Economia Aplicada*, *10*(2), 161-180. https://doi.org/10.1590/S1413-80502006000200001

Duffie, D., & Singleton, K. (1993). Simulated moments estimation of Markov models of asset prices. *Econometrica*, *61*(4), 929-952. https://doi.org/10.2307/2951768

Gallant, A. R. (1977). Three-stage least-squares estimation for a system of simultaneous, nonlinear, implicit equations. *Journal of Econometrics*, *5*(1), 71-88. https://doi.org/10.1016/0304-4076(77)90035-5

Genaro, A., & Avellaneda, M. (2018). Does the lending rate impact ETF's prices? *Brazilian Review of Econometrics*, *38*(2), 287-319. https://doi.org/10.12660/bre.v38n22018.31732

Hall, A. R. (2005). *Generalized method of moments*. Oxford, UK: Oxford University Press.

Hansen, L. P. (2013). *Uncertainty outside and inside economic models*. Retrieved from https://www.nobelprize.org/uploads/2018/06/hansen-lecture.pdf

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, *50*(4), 1029-1054. https://doi.org/10.2307/1912775

Hansen, L. P., & Singleton, K. J. (1982). Generalized instrumental variables estimation of nonlinear rational expectations models. *Econometrica*, *50*(5), 1269-1286. https://doi.org/10.2307/1911873

Hansen, L. P., Heaton, J., & Yaron, A. (1996). Finite-sample properties of some alternative GMM estimators. *Journal of Business and Economic Statistics*, *14*(3), 262-280. https://doi.org/10.1080/07350015.1996.10524656

Issler, J. V., & Piqueira, N. S. (2000). Estimating relative risk aversion, the discount rate, and the intertemporal elasticity of substitution in consumption for Brazil using three types of utility function. *Brazilian Review of Econometrics*, *20*(2), 201-239. https://doi.org/10.12660/bre.v20n22000.2758

Jagannathan, R., Skoulakis, G., & Wang, Z. (2002). Generalized methods of moments: Applications in finance. *Journal of Business and Economic Statistics*, *20*(4), 470-481. https://doi.org/10.1198/073500102288618612

Lucas, R. E., Jr. (1978). Asset prices in an exchange economy. *Econometrica,* *46*(6), 1429-1445. https://doi.org/10.2307/1913837

Martins, H. C. (2021). Tutorial-articles: The importance of data and code sharing. *Revista de Administração Contemporânea*, *25*(1), e200212. https://doi.org/10.1590/1982-7849rac2021200212

McFadden, D. (1989). A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica*, *57*(5), 995-1026. https://doi.org/10.2307/1913621

Mehra, R., & Prescott, E. C. (1985). The equity premium: A puzzle. *Journal of Monetary Economics*, *15*(2), 145-161. https://doi.org/10.1016/0304-3932(85)90061-3

Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica*, *49*(6), 1417–1426. https://doi.org/10.2307/1911408

Pedersen, A. R. (1995). A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations. *Scandinavian Journal of Statistics*, *22*(1), 55–71. Retrieved from https://www.jstor.org/stable/4616340

Stock, J. H., Wright, J., & Yogo, M. (2002). A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business & Economic Statistics*, *20*(4), 518-529.

Verbeek, M. (2004). *A guide to modern econometrics*. Chichester, England: John Wiley & Sons.

Authorship

Alan de Genaro*

Fundação Getulio Vargas, Escola de Administração de Empresas de São Paulo
Rua Itapeva, n. 474, Bela Vista, 01332-000, São Paulo, SP, Brazil.
E-mail: alan.genaro@fgv.br
https://orcid.org/0000-0002-9839-6116

Paula Astorino

Fundação Getulio Vargas, Escola de Administração de Empresas de São Paulo
Rua Itapeva, n. 474, Bela Vista, 01332-000, São Paulo, SP, Brazil.
E-mail: Paula.astorino@fgv.br
https://orcid.org/0000-0001-9811-146X

* Corresponding Author

Copyrights

Conflict of Interests

The authors have stated that there is no conflict of interest.

Plagiarism Check

The RAC maintains the practice of submitting all documents approved for publication to the plagiarism check, using specific tools, e.g.: iThenticate.

Peer Review Method

This content was evaluated using the double-blind peer review process. The disclosure of the reviewers' information on the first page, as well as the Peer Review Report, is made only after concluding the evaluation process, and with the voluntary consent of the respective reviewers and authors.

Authors' Contributions

1st author: conceptualization (lead); data curation (equal); methodology (lead); project administration (equal); software (lead); supervision (lead); writing – original draft (lead).

2nd author: data curation (equal); project administration (equal); validation (lead); writing – review & editing (lead).

Funding

The authors reported that there is no financial support for the research in this article.

Data Availability

RAC informs that the videos and their contents mentioned in this article were an option by the author to offer extra material to readers. Thus, the videos were not part of the peer review process and are not an official RAC publication.

The authors claim that all data used in the research have been made publicly available through the Harvard Dataverse platform and can be accessed at:

Genaro, Alan de; Astorino, Paula, 2022, "Replication Data for "A Tutorial on the GMM Method" published by RAC - Revista de Administração Contemporânea", Harvard Dataverse, V1. https://doi.org/10.7910/DVN/ZQRNZA

RAC encourages data sharing but, in compliance with ethical principles, it does not demand the disclosure of any means of identifying research subjects, preserving the privacy of research subjects. The practice of open data is to enable the reproducibility of results, and to ensure the unrestricted transparency of the results of the published research, without requiring the identity of research subjects.