Generalized Method of Moments

Gabriel Marin[1]

Inter-American Development Bank[1]

Macroeconomic Policy: Fiscal, Monetary and Finance

Outline

Table of Contents

Antecedents of the Generalized Method of Moments (GMM)

Let us begin our lecture by analyzing three antecedents of the Generalized Method of Moments (GMM):

1. Method of Moments
2. Minimum Chi-Square
3. IV Estimation

We will make evaluate each of the three and derive the GMM estimator. PDF with algebra should be uploaded in Brightspace. Our lecture is based on Hall (2004).

Method of Moments

Recall from your Introduction to Econometrics class the Method of Moments. The general idea is to estimate the **population** moments using the **sample** moment conditions. In general, we would have moment conditions of the form:

$$E[v_t] - \mu_0 = 0 \tag{1}$$

$$E[v_t^2] - (\sigma_0^2 + \mu_0^2) = 0 \tag{2}$$

Substituting with the sample moments and rearranging yields:

$$T^{-1} \sum_{t=1}^{T} v_t = \hat{\mu}_T \tag{3}$$

$$T^{-1} \sum_{t=1}^{T} (v_t - \hat{\mu}_T)^2 = \hat{\sigma}_T^2 \tag{4}$$

One caveat of this methodology is that the values of $(\hat{\mu}_T, \hat{\sigma}_T^2)$ depend on the moments chosen.

Minimum Chi-Square Estimation

The second antecedent comes from the Minimum Chi-Square estimation. First, let us suppose we want to model the probability that the outcome of an experiment lies in one of k mutually exclusive groups. Then our null hypothesis would be:

$$p_i = h(i, \theta_0), \tag{5}$$

where $h(i, \theta_0)$ is a functional form with parameter θ_0. Now, we can perform inference on this model by using the goodness-of-fit statistic.

$$GF_t(\theta_0) = \sum_{i=1}^{k} \frac{[T_i - T_h(i, \theta_0)]^2}{T_i}, \tag{6}$$

where T_i is the frequency of the outcomes in the i^{th} group in a sample size of T. Pearson (1900) shows that GF_t is distributed as $\chi^2(k - 1 - p)$ under the null.

Minimum Chi-Square Estimation - Continued

Neyman and Pearson (1928) suggested that a reasonable approach for the solution of this model is to find the value of $\hat{\theta}_T$ that minimizes the goodness-of-fit statistic. They are the ones who created the *Minimum Chi-Square Estimation*. Now, how is it that the GF_t statistic relates with the GMM?

From GF_t to GMM...

We can rexpress Equation (6) as follows,

$$GF_t(\theta_0) = \sum_{i=1}^{k} \frac{[\hat{p}_i - h(i;\theta_0)]^2}{\hat{p}_i}, \tag{7}$$

where $\hat{p}_i = \frac{T_i}{T}$, the relative frequency in the sample of the outcomes in the i^{th} group. With this, we will make first a connection with the method of moments. Let us consider a set of indicator variables $\{D_t(i); i = 1, 2, ...k; t = 1, 2, ...T\}$, if we have that Equation (5) holds, $P(D_t(i) = 1) = E[D_t(i)] = h(i;\theta_0)$. This implies the following vector of k population moment conditions.

$$E \begin{bmatrix} D_t(1) - h(1;\theta_0) \\ D_t(2) - h(2;\theta_0) \\ . \\ . \\ D_t(k) - h(k;\theta_0) \end{bmatrix} = 0$$

From GF_t to GMM - Continued

Since we have that our population moments add up to zero, we only need $k-1$ of the population moment conditions actually provide information of θ_0. Since we assumed that $p \leq k-1$, then we can use these population moment conditions to estimate θ_0. In other words, we have:

$$\begin{bmatrix} \hat{p}_1 - h(1;\theta_0) \\ \hat{p}_2 - h(2;\theta_0) \\ . \\ . \\ \hat{p}_k - h(k;\theta_0) \end{bmatrix} = 0 \tag{8}$$

Which are the same elements of the numerator of Equation (7)! Yet, we do not have the GF_t per se. This is where we connect the Minimum Chi-Square estimation and the population moment conditions.

From GF_t to GMM - Continued

Assuming $k-1=p$ (same moment conditions as unknown parameters, we have that $\hat{\theta}_T$ satisfies $GF_t(\hat{\theta}_T)=0$. Since GF_t is a chi-squared, the minimum is zero! Now, if we consider the situation where $k-1>p$, the method of moments is not valid as we have more moments than parameters, but we can still use the Minimum Chi-Square estimation. In this case, let us use the GF_t in matrix form.

$$GF_t(\theta)=T\begin{bmatrix}\hat{p}_1-h(1;\theta_0)\\ \hat{p}_2-h(2;\theta_0)\\ .\\ .\\ \hat{p}_k-h(k;\theta_0)\end{bmatrix}'\begin{bmatrix}\hat{p}_1^{-1} & 0 & . & . & 0\\ 0 & \hat{p}_2^{-1} & . & . & 0\\ . & . & . & . & .\\ . & . & . & . & .\\ 0 & 0 & . & . & \hat{p}_k^{-1}\end{bmatrix}\begin{bmatrix}\hat{p}_1-h(1;\theta_0)\\ \hat{p}_2-h(2;\theta_0)\\ .\\ .\\ \hat{p}_k-h(k;\theta_0)\end{bmatrix} \quad (9)$$

In this case, we want to find the value of θ_0 which is closest to solving the sample moment conditions.

IV Estimation - Hall (2005)

Let us consider the following system of equations

$$q_t^D = \alpha_0 p_t + u_t^D \tag{10}$$

$$q_t^S = \beta_1' n_t + \beta_2 p_t + u_t^S \tag{11}$$

$$q_t^D = q_t^S = q_t \tag{12}$$

If we wanted to regress q_t on p_t we see that both are determined at the same time, this results in a problem of endogeneity. To resolve this issue, we make use of a methodology called instrumental variable (IV) regression. The idea is to find an instrument z_t^D that is not correlated with the error term ϵ_t, and correlated with the endogenous regressor p_t.

IV Assumptions

Formally, an instrumental variable must satisfy three conditions:

- Independence: Endogenous variable orthogonal to the instrument.
- Exclusion
$$Cov(Z_i, \varepsilon_i) = 0$$
- Relevance
$$Cov(Z_i, x_i) \neq 0$$
- Monotonicity

If our IV satisfies such conditions, our previous example we would have:

$$Cov[z_t^D, q_t] - \alpha_0 Cov[z_t^D, p_t] = 0, \tag{13}$$

which is equivalent to

$$E[z_t^D q_t] - \alpha_0 E[z_t^D p_t] = 0 \tag{14}$$

Table of Contents

Generalized Method of Moments - Hall (2005)

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

$$E[f(v_t, \theta_0)] = 0, \forall t. \tag{15}$$

Definition 2. Generalized Method of Moments Estimator: The GMM estimator is the value of θ which minimizes:

$$Q_T(\theta) = T^{-1} \sum_{t=1}^{T} f(v_t, \theta)' W_T T^{-1} \sum_{t=1}^{T} f(v_t, \theta), \tag{16}$$

where W_T is a positive semi-definite matrix which may depend on the data but converges in probability to a positive definite of matrix of constants.

GMM- Assumptions

- Strict Stationarity (recall VAR slides)
- Orthogonality condition - z_t satisfies $E[z_t ut] = 0$
- Identification condition - $rank(E[z_t x_t']) = p$. You will probably encounter problems with this one. Two cases: i) either $E[z_t x_t']$ contains a row of zeros, or ii) $E[z_t x_t]$ is a multiple of other.

GMM and IV - Revisited

Let y be a $(T \times 1)$ vector whose t^{th} element is y_t; X be the $(T \times p)$ matrix whose t^{th} row is x_t'; Z be the $(T \times p)$ matrix whose t^{th} row is z_t' u be the $(T \times 1)$ vector whose t_{th} element is u_t; and $u(\theta) = y - X\theta$.
Substituting with Equation (16), we obtain:

$$Q_T(\theta) = \{T^{-1}u(\theta)'Z\}W_t\{T^{-1}Z'u(\theta)\} \tag{17}$$

Recall we want to find the value of θ that minimizes Equation (17). Rearranging the associated first order condition yields

$$\hat{\theta}_T = \{(T^{-1}X'Z)W_T(T^{-1}Z'X)\}^{-1}\{(T^{-1}X'Z)W_T(T^{-1}Z'y)\}. \tag{18}$$

We can also rearrange the first order conditions as

$$(T^{-1}X'Z)W_T T^{-1}Z'u(\hat{\theta}_T) = 0, \tag{19}$$

which is identical to the method of moments estimator of

$$E[x_t z_t']WE[z_t u_t(\theta_0)] = 0 \tag{20}$$

GMM and IV - Revisited

In the case that we have an exactly identified model (same number of parameters as moment conditions), we have that the weighting matrix plays no role in our estimation, reducing Equation (20) to

$$\hat{\theta} = (T^{-1}Z'X)^{-1}(T^{-1}Z'y). \tag{21}$$

This is exactly the same as the standard IV estimator! Therefore, in the case of an exactly identified model, the GMM estimator and the IV estimator are equivalent. If we were to have more moments than parameters (overidentified model), then W_T would determine the information we use to minimize the Equation (20).

GMM - Optimal Weighting Matrix

Recall the weighting matrix W_t will help us determine how the information will be used for our minimization problem of $Q_T(\theta)$. In that sense, Hansen (1982) shows the GMM optimal estimator θ_T would be obtained in two steps. First, estimate θ_0 using a sub-optimal weighting matrix like $W_t = I$, then obtain the variance covariance matrix from the first estimation. The second step uses the inverse of that first-step matrix and estimates the GMM model. Hansen shows this is sufficient to obtain an asymptotic covariance matrix

Table of Contents

Delgadillo, Marin & Von Der Meden (2020) - Households

We developed a New-Keynesian DSGE model that included a fiscal feedback to the Dynamic IS equation. The model is as follows: Households maximize their utility function by selecting a combination of consumption C_t and working hours H_t subject to a budget constraint.

$$\max_{C_t,H_t} \sum_{t=0}^{\infty} \beta^t \left(\frac{C_t^{1-\sigma}}{1-\sigma} - \frac{H_t^{1-\gamma}}{1-\gamma}\right) \tag{22}$$

s.t.

$$(1+\tau^c)(P_tC_t)+V_tb_t+K_{t+1}-(1-\delta)K_t = w_tH_t+(1-\tau^k)r_tK_t+b_{t-1}+T_t, \tag{23}$$

where C_t, H_t, b_t, K_t stand for consumption, working hours, government bonds and capital, respectively. Also, P_t, w_t, r_t and V_t corresponds to consumption price index, wages, real interest rate and bond price, respectively.

Delgadillo, Marin & Von Der Meden (2020) - Households

In the model, households divide their expenditure into consumption goods and services (C_t), investment on capital goods (K_t), and government bonds (b_t). They receive labor income through wages (w_t) and capital income via (r_t). In present time they receive the nominal value and interest gain on past any purchase of government bonds (b_{t-1}) done in $t-1$. Finally, all government revenues are fully transferred to households (T_t).

The government obtains resources through taxes on capital income and final consumption.[1] A second source of revenues is bond financing. At any period, it satisfies a balanced budget condition:

$$r_t \tau^k + P_t C_t \tau^c + V_t b_t = T_t + b_{t-1}. \tag{24}$$

[1] This is done to approximate the reality of Mexico, where the main taxation instruments are value added taxes, income taxes, and excise taxes.

Delgadillo, Marin & Von Der Meden (2020) - Households

One of the features that differentiates this model from the NK canonical model (see for instance Gali (2008)), is that the latter uses only consumption to explain the production, that is $Y_t = C_t$. Whereas this thesis proposes a variation of this model with investment and government expenditure. Therefore, gross domestic product is defines as $Y_t = C_t + X_t + G_t$. Where X_t, and G_t are the investment and government expenditure in time t, respectively. This influences the households' consumption by distorting the consumption-saving decisions. Such result is reflected on the following set of equations:

$$1 = E_t[\beta(\frac{P_t}{P_{t+1}})(\frac{C_t}{C_{t+1}})^{\sigma}(1 - \delta + (1 - \tau^k)r_{t+1})], \tag{25}$$

$$\frac{w_t}{P_t} = (1 + \tau^c)C_t^{\sigma} H_t^{\gamma}, \tag{26}$$

$$V_t = E_t[\beta(\frac{P_t}{P_{t+1}})(\frac{C_t}{C_{t+1}})^{\sigma}]. \tag{27}$$

Delgadillo, Marin & Von Der Meden (2020) - Firms

As in the canonical NK monetary framework, we assume a price setting behavior governed by an inter temporal profit maximization

$$\max_{P_t} \sum_{k=0}^{\infty} \theta^k E_t[Q_{t,t+k}(P_t^* Y_{t+k|t} - M\Psi_{t+k}(Y_{t+k|t}))] \tag{28}$$

s.t

$$Y_{t+k|t} = (\frac{P_t^*}{P_{t+k}})^{-\epsilon} C_{t+k}, \tag{29}$$

where $Q_{t,t+k}$, Ψ_{t+k}, are the stochastic discount factor on returns and the cost function, respectively, and $t, t+k$ indicates from time t to time $t+k$, and $t+k|t$ indicates the value $t+k$ given the value in t.

Delgadillo, Marin & Von Der Meden (2020) - Firms

The price structure follows a Calvo price setting dynamic (Calvo (1983)) where firms have a probability θ of maintaining their prices unchanged this period and $1-\theta$ to reset their prices. This dynamic allows θ to become the price stickiness index and introduces inflation through monopolistic competition. Also, the profit optimization problem faced by firms determines the price P^* while it remains effective to market value profits. This problem is subject to the firms' costs function, profit margin, market stability, and θ. The cost function may take different forms. As in the standard model, it includes a marginal cost markup M. The associated first order condition takes the form:

$$\sum_{k=0}^{\infty} \theta^k E_t[Q_{t,t+k} Y_{t+k|t}(P_t^* - M\Psi_{t+k|t})], \tag{30}$$

As in Gali (2008). We make use of a Cobb-Douglas production function, where the capital's share in production is represented by α, while the labor share is represented by $1-\alpha$

$$Y_t = A_t K_t^{\alpha} H_t^{1-\alpha}. \tag{31}$$

Delgadillo, Marin & Von Der Meden (2020) - Firms

In terms of the aggregate price index, following the traditional NK framework's Calvo price setting (Calvo (1983)), the index takes the form:

$$\Pi_t^{1-\epsilon} = \theta + (1-\theta)\frac{P_t^*}{P_{t-1}}^{1-\epsilon}. \tag{32}$$

Then, we log-linearized[2] (32) around the steady state,[3] which yields the following:

$$\pi_t = (1-\theta)(p_t^* - p_{t-1}). \tag{33}$$

The marginal cost in logs is defined as the difference of the required real wage by the household, and the marginal product of labor from the firms:

$$mc_t = (w_t - p_t) - mph_t. \tag{34}$$

[2]Variables in lower-case are to be understood as log-linearized values.

[3]It is assumed as steady state zero inflation.

Delgadillo, Marin & Von Der Meden (2020) - Firms

One of the innovations to the canonical model, is that the marginal product of labor now contains a component of capital. From the first order conditions from firms, we know that:

$$w_t = (1-\alpha)A_t(\frac{K_t}{H_t})^{\alpha}, \tag{35}$$

that after log-linearizing:

$$mph_t = log(1-\alpha) + a_t + \alpha k_t - \alpha h_t. \tag{36}$$

We will replace in the marginal cost (mc_t), with the real wage required by households and the mph_t from firms. This combination leads to:

$$mc_t = log(1-\tau^c) + \sigma c_t + (\gamma+\alpha)h_t - log(1-\alpha) - a_t - \alpha k_t. \tag{37}$$

Delgadillo, Marin & Von Der Meden (2020) - Equilibrium

Following Gali (2008), log-linearizing (30) around the steady state and combining it with the deviation from steady state of Equation (37) - mc_t - yields the inflation equation

$$\pi_t = \beta E_t[\pi_{t+1}] + \lambda mc_t. \tag{38}$$

Then, log-linearizing (25) around the steady state yields the Dynamic IS.

$$\tilde{y}_t = \frac{1}{w_1}(E_t[\tilde{y}_{t+1}] - E_t[g_{t+1}] - E_t[x_{t+1}] + \Delta(E_t[r_{t+1}] - \delta - E_t[\pi_{t+1}]). \tag{39}$$

Next, knowing that $Y_t = C_t + X_t + G_t$, we can find an equation that expresses consumption as a share of income. Then, by combining such expression with (38), we can further obtain the New Keynesian Phillips Curve (NKPC).

$$\pi_t = \beta E_t[\pi_{t+1}] + \kappa^* \tilde{y}_t. \tag{40}$$

Finally, we use a standard Taylor rule whenever is needed

$$i_t = \phi_0 + \phi_\pi \pi_t + \phi_{\tilde{y}} \tilde{y}_t. \tag{41}$$

GMM Estimation in R

$$E_t[w_t - p_t - log(1+\tau^c) - c_t - \gamma h_t] = 0, \tag{42}$$

$$E_t[v_t - log(\beta) - p_t + p_{t+1} - c_t + c_{t+1}] = 0, \tag{43}$$

$$E_t[1 - log(\beta) - p_t + p_{t+1} - c_t + c_{t+1} - log(1 - \delta + (1-\tau^k)r_{t+1})] = 0, \tag{44}$$

$$E_t[r_t - log(\alpha) + log(1-\tau^k) - \alpha(h_t - k_t)] = 0, \tag{45}$$

References

- Calvo, G. A. (1983). Staggered prices in a utility-maximizing framework. Journal of monetary Economics, 12(3), 383-398.
- Marin, Gabriel and Delgadillo, Julio and Von der Meden, Jürgen, Enhancing Central Bank Decision Making with Machine Learning: An Application of Random Forest Regressions to Mexico (April 4, 2020).
- Galí, J. (2008). Monetary policy design in the basic new Keynesian model. Monetary Policy, Inflation, and the Business Cycle.
- Hall, A. R. (2004). Generalized method of moments. OUP Oxford.
- Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. Econometrica: Journal of the econometric society, 1029-1054.
- Neyman, J., Pearson, E. S. (1928). On the use and interpretation of certain test criteria for purposes of statistical inference: Part I. Biometrika, 175-240.

References

- Pearson, K. (1900). X. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 50(302), 157-175.

Thank you!

E-mail: gmarinmunoz@iadb.org