Method of moments

Method of moments estimation is widely applicable and particularly attractive for addressing instrumental variable estimation in nonlinear models and as an alternative to maximum likelihood estimation where the likelihood function presents analytic challenges. The idea is simple: equate the sample moments with the theoretical moments. The method requires at least as many moment conditions as parameters to be estimated (when we have more moment conditions than parameters the method is described as generalized method of moments or GMM; discussed below).

Consider some simple examples.

Example 1. Suppose we have a random sample of size n from a **normal distribution** $f(X;\mu,\sigma)$. The moment conditions are

$$\mu = E[X] = \frac{1}{n}\sum_i x_i = \overline{X}$$

and

$$\sigma^2 = Var[X] = E\left[(X-\mu)^2\right] = \frac{1}{n}\sum_i \left(x_i - \overline{X}\right)^2$$

In this simple case, the sample moments are the moment estimators for the parameters.

$$\mu_{MM} = \overline{X}$$

$$\sigma^2_{MM} = \frac{1}{n}\sum_i \left(x_i - \overline{X}\right)^2$$

Example 2. Suppose we have a random sample of size n from a **gamma distribution** $g(X;\alpha,\theta)$. The density function for the gamma distribution is $g(X;\alpha,\theta) = \frac{1}{\Gamma(\alpha)\theta^\alpha} x^{\alpha-1} e^{-\frac{x}{\theta}}$. The moment conditions are

$$\mu = E[X] = \alpha\theta = \frac{1}{n}\sum_i x_i = \overline{X}$$

and

$$\sigma^2 = Var[X] = E\left[(X-\mu)^2\right] = \alpha\theta^2 = \frac{1}{n}\sum_i \left(x_i - \overline{X}\right)^2$$

Solve the moment conditions to recover the parameter estimators. Let

$$\alpha = \frac{\overline{X}}{\theta}$$

then

$$\sigma^2 = \frac{\overline{X}}{\theta}\theta^2 = \overline{X}\theta = \frac{1}{n}\sum_i \left(x_i - \overline{X}\right)^2$$

$$\theta_{MM} = \frac{\frac{1}{n}\sum_i \left(x_i - \overline{X}\right)^2}{\overline{X}} = \frac{1}{\overline{X}n}\sum_i \left(x_i - \overline{X}\right)^2$$

and

$$\alpha_{MM} = \frac{\overline{X}}{\frac{1}{\overline{X}n}\sum_i \left(x_i - \overline{X}\right)^2} = \frac{n\overline{X}^2}{\sum_i \left(x_i - \overline{X}\right)^2}$$

Example 3. Suppose the data generating process (DGP) follows a **linear regression model** $y_i = x_i^T\beta + \varepsilon_i$ where x_i is a k element vector, $E\left[\varepsilon_i \mid x_i\right] = 0$, and we have a sample of size n. The moment (or orthogonality) condition is

$$E\left[x_i\left(y_i - x_i^T\beta\right)\right] = 0$$

and corresponding sample moment condition is

$$\frac{1}{n}\sum_i x_i\left(y_i - x_i^T\beta\right) = 0$$

Then, the method of moments estimator for β is

$$\beta_{MM} = \left(\sum_i x_i x_i^T\right)^{-1}\sum_i x_i y_i$$

or, in matrix form,

$$\beta_{MM} = \left(X^T X\right)^{-1} X^T Y$$

Example 4. Continue with example 3 except that the error condition $E\left[\varepsilon_i \mid x_i\right] = 0$ is violated but we have k exogenous or **instrumental variables** Z related to X that satisfy $E\left[\varepsilon_i \mid z_i\right] = 0$. The exactly-identified moment condition is

$$E\left[z_i\left(y_i - x_i^T\beta\right)\right] = 0$$

and corresponding sample moment condition is

$$\frac{1}{n}\sum_i z_i\left(y_i - x_i^T\beta\right) = 0$$

Then, the method of moments estimator for β is

$$\beta_{MM} = \left(\sum_i z_i x_i^T\right)^{-1}\sum_i z_i y_i$$

or, in matrix form,

$$\beta_{MM} = \left(Z^T X\right)^{-1} Z^T Y$$

Example 5. Suppose the DGP follows a **nonlinear regression model** $u_i = g(y_i, x_i, \theta)$ with non-additive error where $E[u_i \mid x_i] = 0$ and we have a random sample of size n. The moment condition is

$$E[h(x_i)\, g(y_i, x_i, \theta)] = 0$$

and the sample moment condition is

$$\frac{1}{n}\sum_i h(x_i)\, g(y_i, x_i, \theta) = 0$$

where $h(x_i) = \frac{\partial g(y_i, x_i, \theta)}{\partial \theta}$ and θ_{MM} is determined by numerical methods.

Generalized method of moments (GMM).

Example 6. Continue with example 4 except we have $j > k$ instruments Z so that the moment condition is **over-identified**. In this case we attempt to minimize the quadratic loss.

$$Q(\beta) = \left[\frac{1}{n}(Y - X\beta)^T Z\right] W \left[Z^T (Y - X\beta)\right]$$

where W is a weight matrix (discussed below). The first order condition provides the moment condition

$$-2\left[\frac{1}{n}X^T Z\right] W \left[Z^T (Y - X\beta)\right] = 0$$

and the GMM estimator.

$$\beta_{GMM} = \left[X^T Z W Z^T X\right]^{-1} X^T Z W Z^T Y$$

Example 7. Suppose we have $j > k$ **general (possibly, nonlinear) moment functions** $h_i(\theta)$ where $E[h(w, \theta_0)] = 0$ for θ_0 the true parameters and the average h_i evaluated at θ_0 converges in distribution to a mean zero normal random vector with variance $S_o = p\lim \frac{1}{n}\sum_i \sum_j \left[h_i h_j^T \mid_{\theta_0}\right]$. GMM quadratic to be minimized is

$$\frac{1}{2}\left[\frac{1}{n}\sum_i h(w_i, \theta)\right]^T W \left[\frac{1}{n}\sum_i h(w_i, \theta)\right]$$

where $h_i(\theta) = h_i(w_i, \theta)$. First order conditions are

$$\left[\frac{1}{n}\sum_i \frac{\partial h_i\left(\widehat{\theta}\right)^T}{\partial \theta} \mid_{\widehat{\theta}}\right] W \left[\frac{1}{n}\sum_i h_i\left(\widehat{\theta}\right)\right] = 0$$

Example 8. GMM nonlinear IV estimation takes various forms. Suppose we have $j \geq k$ instruments Z, $E[u_i \mid z_i] = 0$, where $u_i = y_i - g(x_i)$ for additive errors or $u_i = r(x_i, y_i, \theta)$ for non-additive errors. The GMM estimator for the exactly-identified case solves $Z^T u = 0$ for θ_{GMM}.

The general case minimizes

$$u^T Z W Z^T u$$

where the optimal GMM estimator employs $W = S^{-1}$ (discussed below) and NL2SLS-IV employs $W = \left(Z^T Z\right)^{-1}$. In other words, NL2SLS-IV minimizes $u^T P_z u$ where P_z is the projection matrix into the columns of Z, $P_z = Z\left(Z^T Z\right)^{-1} Z^T$.

Optimal weighting matrix.

If S_0 is known, the optimal weighting matrix is $W = S_0^{-1}$. In practice, S_0 is typically unknown and estimated via some variant of the Newey-West estimator.[1]

$$\widehat{S} = \Omega_0 + \sum_{l=1}^{j} \left(1 - \frac{l}{j+1}\right) \left(\Omega_l + \Omega_l^T\right)$$

where $\Omega_l = \frac{1}{T} \sum_{t=l+1}^{T} h_t h_{t-l}^T$.

Method of simulated moments (MSM).

GMM may be infeasible if the moment conditions are intractable. For example, the conditions may involve latent variables (variables unobserved by the analyst). In such instances, simulated moments can be substituted in place of theoretical moments where the expected value of the simulated moment equals the theoretical moment. This is the method of simulated moments (MSM).

Example 9.[2] Suppose we have a random sample from a **log-normal distribution** $g\left(y; \mu, \sigma^2\right)$. Then, $z = \log y$ has a normal distribution $f\left(z; \mu, \sigma^2\right)$. This setting is amenable to GMM so that MSM is not necessitated but rather allows for comparison of the relative efficacy of MSM. The elementary zero functions are

$$h_{1i}\left(\mu, \sigma^2\right) = z_i - \mu$$

and

$$h_{2i}\left(\mu, \sigma^2\right) = y_i - \exp\left[\mu + \frac{1}{2}\sigma^2\right]$$

[1] The Newey-West estimator is designed to assure positive semi-definiteness. In this instance, we require positive definiteness.

[2] This example is adapted from Davidson and MacKinnon, 2004, *Econometric Theory and Methods*. They provide a more detailed discussion of the properties of GMM and MSM estimators.

Derivatives of these functions with respect to the parameters are

$$\frac{\partial h_{1i}}{\partial \mu} = -1; \frac{\partial h_{2i}}{\partial \mu} = -\exp\left[\mu + \frac{1}{2}\sigma^2\right]$$

$$\frac{\partial h_{1i}}{\partial \sigma^2} = 0; \frac{\partial h_{2i}}{\partial \sigma^2} = -\frac{1}{2}\exp\left[\mu + \frac{1}{2}\sigma^2\right]$$

The weight matrix follows from

$$S_0 = E\left(\begin{bmatrix} h_{10} \\ h_{20} \end{bmatrix}\begin{bmatrix} h_{10}^T & h_{20}^T \end{bmatrix}\right) = \begin{bmatrix} \sigma_z^2 I & \sigma_{zy} I \\ \sigma_{yz} I & \sigma_y^2 I \end{bmatrix}$$

where h_{j0}, $j = 1, 2$, is h_j evaluated at the true values μ_0 and σ_0^2. Then, the efficient moment functions (in matrix form) are

$$H^T\left(\mu, \sigma^2\right) W h\left(\mu, \sigma^2\right) = 0$$

where $H\left(\mu, \sigma^2\right) = -\begin{bmatrix} \iota & 0 \\ \exp\left[\mu + \frac{1}{2}\sigma^2\right]\iota & \frac{1}{2}\exp\left[\mu + \frac{1}{2}\sigma^2\right]\iota \end{bmatrix}$, ι is a vector of ones, $W = S_0^{-1}$, and $h\left(\mu, \sigma^2\right) = \begin{bmatrix} h_1(\mu, \sigma^2) \\ h_2(\mu, \sigma^2) \end{bmatrix}$. While $H\left(\mu, \sigma^2\right)$ and W affect the variance or efficiency of the over-identified (GMM) estimator, they don't appear in the simplified (exactly-identified) estimating equations.

$$\iota^T h_1\left(\mu, \sigma^2\right) = 0; \iota^T h_2\left(\mu, \sigma^2\right) = 0$$

These two equations yield $\mu_{GMM} = \overline{z}$ and $\sigma_{GMM}^2 = 2\left(\log \overline{y} - \overline{z}\right)$.[3]

MSM

Next, we re-evaluate this setting via simulation, that is, employing MSM. Let u^* represent random draws from a standard normal distribution. The key to simulated moments is to generate simulated expected values of $y = \exp\left[\mu + \sigma u^*\right]$ and $z = \log\left[y\right]$. Then, simulated moments of z and y, respectively, involve averaging

$$m_1^*\left(u^*, \mu, \sigma^2\right) = \mu + \sigma u^*$$

and

$$m_2^*\left(u^*, \mu, \sigma^2\right) = \exp\left[\mu + \sigma u^*\right]$$

If we employ R simulations, the zero functions for MSM are

$$h_{t1}^*\left(z_t, \mu, \sigma^2\right) = z_t - \frac{1}{R}\sum_{r=1}^{R} m_1^*\left(u_{tr}^*, \mu, \sigma^2\right)$$

[3]Of course, we could add over-identifying moments such as $E\left[y^2\right]$ but we'll continue along this notationally simpler path with simulated moments.

and

$$h_{t2}^{*}\left(y_t,\mu,\sigma^2\right)=y_t-\frac{1}{R}\sum_{r=1}^{R}m_2^{*}\left(u_{tr}^{*},\mu,\sigma^2\right)$$

where u_{tr}^{*} are drawn once (a vector of length nR) from a standard normal distribution.

The simulated moment conditions average over the zero functions equating the sample average with the simulated sample average.

$$\frac{1}{n}\sum_{t=1}^{n}h_{t1}^{*}\left(z_t,\mu,\sigma^2\right)=\overline{z}-\frac{1}{n}\sum_{t=1}^{n}\frac{1}{R}\sum_{r=1}^{R}m_1^{*}(u_{tr}^{*},\mu,\sigma^2)=0$$

and

$$\frac{1}{n}\sum_{t=1}^{n}h_{t2}^{*}\left(y_t,\mu,\sigma^2\right)=\overline{y}-\frac{1}{n}\sum_{t=1}^{n}\frac{1}{R}\sum_{r=1}^{R}m_2^{*}(u_{tr}^{*},\mu,\sigma^2)=0$$

MSM estimation minimizes the quadratic form solving for the parameters.

$$\left[\begin{array}{cc}\frac{1}{n}\sum_{t=1}^{n}h_{t1}^{*}\left(z_t,\mu,\sigma^2\right) & \frac{1}{n}\sum_{t=1}^{n}h_{t2}^{*}\left(y_t,\mu,\sigma^2\right)\end{array}\right]\left[\begin{array}{c}\frac{1}{n}\sum_{t=1}^{n}h_{t1}^{*}\left(z_t,\mu,\sigma^2\right)\\ \frac{1}{n}\sum_{t=1}^{n}h_{t2}^{*}\left(y_t,\mu,\sigma^2\right)\end{array}\right]$$

Based on the moment conditions we know the MSM estimator is asymptotically consistent, we next explore MSM efficiency relative to GMM via simulation.

Simulation.

Let $n=R=1,000$, then we ran 100 simulations of GMM and MSM estimation of μ and σ for the lognormal distribution ($\mu=10,\sigma=5$) discussed above. Results are tabulated below.

	μ	μ	σ	σ
	GMM	MSM	GMM	MSM
mean	10.01450	10.319229	4.46997	4.453509
std. dev.	0.1380157	0.4346337	0.2930238	0.3678285
1% quantile	9.714255	9.482923	4.038108	3.850237
5% quantile	9.807114	9.651765	4.072466	3.885505
10% quantile	9.839347	9.771796	4.110955	3.948107
25% quantile	9.921804	9.980254	4.266044	4.209186
50% quantile	10.010791	10.320465	4.448000	4.451024
75% quantile	10.104375	10.558103	4.631000	4.723155
90% quantile	10.197654	10.924584	4.800516	4.900083
95% quantile	10.240818	11.087159	4.961245	5.021340
99% quantile	10.301261	11.254266	5.373312	5.331810

As expected, GMM is somewhat less variable than MSM, nonetheless, in this instance MSM performs quite satisfactorily.