Method of Moments (Classical Method of Moments in Hamilton):

Suppose $\{y_t\}_{t=1}^T$ is an i.i.d. sample of random variable $Y$ from density:

$f_Y(y|\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is a ($k$ x 1) dimensional vector of parameters.

Suppose $k$ population moments can be written as functions of $\boldsymbol{\theta}$:

$E[Y_t^i] = \mu_i(\boldsymbol{\theta}), i = i_1, i_2, ..., i_k$

The method of moments estimator, $\hat{\boldsymbol{\theta}}_{mm}$, of $\boldsymbol{\theta}$ is the value:

$$\mu_i(\hat{\theta}_{mm}) = \frac{1}{T}\sum_{t=1}^{T} y_t^i, i = i_1, i_2, ..., i_k$$

Note that if you need to estimate $k$ parameters, you must specify exactly $k$ moments

Example: Normal

$\boldsymbol{\theta} = (\mu, \sigma^2)'$

$k = 2$

$E[Y^1] = \mu$

$E[Y^2] = \text{Var}(Y) + E[Y]^2 = \sigma^2 + \mu$

Example: Beta Distribution

Suppose $Y \sim \text{Beta}(\alpha, \beta)$:

$$f_Y(y|\alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} y^{\alpha-1}(1-y)^{\beta-1}$$

In this case:

$\boldsymbol{\theta} = (\alpha, \beta)'$ (or $k = 2$)

$$\mu_1 = E[Y^1] = \frac{\alpha}{\alpha+\beta}$$

$$\sigma^2 = \text{Var}(Y) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

$$\mu_2 = E[Y^2] = \text{Var}(Y) + E[Y^1]^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} + \frac{\alpha^2}{(\alpha+\beta)^2} = \frac{\alpha\beta + \alpha^2(\alpha+\beta+1)}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

Solve for $\beta$ using $\mu_1$:

$\alpha = \mu_1(\alpha + \beta)$

$\implies \alpha = \mu_1\alpha + \mu_1\beta$

$\implies \alpha(1 - \mu_1) = \mu_1\beta$

$$\implies \beta = \frac{\alpha(1-\mu_1)}{\mu_1}$$

Useful equations:

$$(\alpha+\beta) = \alpha + \alpha\frac{1-\mu_1}{\mu_1} = \frac{\cancel{\mu_1\alpha} + \alpha(1 - \cancel{\mu_1})}{\mu_1} = \frac{\alpha}{\mu_1}$$

$$(\alpha+\beta+1) = \frac{\alpha}{\mu_1} + 1 = \frac{\alpha+\mu_1}{\mu_1}$$

Now substitute for $\beta$ in $\mu_2$:

$$\mu_2 = \frac{\alpha^2\left(\dfrac{1-\mu_1}{\mu_1}\right) + \alpha^2\left(\dfrac{\alpha+\mu_1}{\mu_1}\right)}{\dfrac{\alpha^2}{\mu_1}\cdot\dfrac{\alpha+\mu_1}{\mu_1}} = \frac{1-\mu_1+\alpha+\mu_1}{\dfrac{\alpha+\mu_1}{\mu_1^2}} = \frac{(1-\alpha)\mu_1^2}{\alpha+\mu_1}$$

$$\implies \alpha\mu_2 + \mu_1\mu_2 = \mu_1^2 + \alpha\mu_1^2$$

$$\alpha(\mu_2 - \mu_1^2) = \mu 1^2 - \mu_1\mu_2$$

$$\implies \alpha = \frac{\mu_1^2 - \mu_1\mu_2}{\underbrace{\mu_2 - \mu_1^2}_{\sigma^2}} = \frac{\mu_1^2 - \mu_1\mu_2 + \mu_1^3 - \mu_1^3}{\sigma^2} = \frac{\mu_1^2(1-\mu_1) - \mu_1\overbrace{(\mu_2 - \mu_1^2)}^{\sigma^2}}{\sigma^2} = \frac{\mu_1^2(1-\mu_1)}{\sigma^2} - \mu_1$$

$$\beta = \frac{\alpha(1-\mu_1)}{\mu_1} = \frac{\mu_1(1-\mu_1)^2}{\sigma^2} - (1-\mu_1)$$

$$\hat{\alpha}_{mm} = \frac{\hat{\mu}_1^2(1-\hat{\mu}_1)}{\hat{\sigma^2}} - \hat{\mu}_1$$

$$\hat{\beta}_{mm} = \frac{\hat{\mu}_1(1-\hat{\mu}_1)^2}{\hat{\sigma^2}} - (1-\hat{\mu}_1)$$

where:

$$\hat{\mu}_1 = \frac{1}{T}\sum_{t=1}^{T} y_t$$

$$\hat{\mu}_2 = \frac{1}{T}\sum_{t=1}^{T} y_t^2$$

$$\hat{\sigma}^2 = \hat{\mu}_2 - \hat{\mu}_1^2$$

<u>Example: t distribution</u>

Suppose $Y \sim t(\nu)$:

$$f_Y(y|\nu) = \frac{\Gamma\left(\dfrac{\nu+1}{2}\right)}{(\pi\nu)^{1/2}\Gamma\left(\dfrac{\nu}{2}\right)}\left(1+\frac{y^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

In this case:

$\boldsymbol{\theta} = \nu$ (or $k = 1$)

If $\nu > 2, \mu_2 = \dfrac{\nu}{\nu-2}$

Thus,

$$\nu = \nu\mu_2 - 2\mu_2$$

$$\implies \nu = \frac{2\mu_2}{\mu_2 - 1}$$

$$\implies \hat{\nu}_{mm} = \frac{2\hat{\mu}_2}{\hat{\mu}_2 - 1}$$

where $\hat{\mu}_2 = \frac{1}{T}\sum_{t=1}^{T} y_t^2$

Generalized Method of Moments

Let $\boldsymbol{Y}$ be an (n x 1) vector of random variables and $\boldsymbol{\theta}$ a (k x 1) vector of parameters governing the process $\boldsymbol{y}_t$.

Denote the true parameter vector as $\boldsymbol{\theta}_0$

Suppose we can specify an (r x 1) vector valued function $\boldsymbol{h}(\boldsymbol{\theta}, \boldsymbol{y}_t) : (\mathbb{R}^k \text{x} \mathbb{R}^n) \to \mathbb{R}_1^r$ such that:

$E[\boldsymbol{h}(\boldsymbol{\theta}_0, \boldsymbol{y}_t)] = 0$, where $r \geq k$

Define:

$$\boldsymbol{g}_T(\boldsymbol{\theta}) = \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{h}(\boldsymbol{\theta}, \boldsymbol{y}_t)$$

$\boldsymbol{g}_T(\boldsymbol{\theta}) : \mathbb{R}^k \to \mathbb{R}^r$

We want to choose $\hat{\theta}_{gmm}$ such that the sample moments $\boldsymbol{g}_T(\hat{\theta}_{gmm})$ are close to zero.

If $r = k$, we can choose $\hat{\theta}_{gmm}$ such that $\boldsymbol{g}_T(\hat{\theta}_{gmm}) = 0$ because we have $k$ equations and $k$ unknowns

If $r > k$ we have more equations than unknowns; in general there is no $\hat{\theta}_{gmm}$ such that $\boldsymbol{g}_T(\hat{\theta}_{gmm}) = 0$

Instead we minimize a quadratic form:

$$\underbrace{Q_T(\boldsymbol{\theta})}_{\text{(1 x 1)}} = \underbrace{\boldsymbol{g}_T(\boldsymbol{\theta})'}_{(1xr)} \underbrace{W_T}_{\text{(r x r)}} \underbrace{\boldsymbol{g}_T(\boldsymbol{\theta})}_{\text{(r x 1)}}$$

The matrix $W_T$ places where weight on some moment conditions and less on others.

We might have to use numerical optimization to minimze $Q_T(\boldsymbol{\theta})$

Example: t-distribution

The method of moments estimator of the t-distribution is a special case of the GMM estimator

$\boldsymbol{y}_t$

$\boldsymbol{\theta} = \nu$

$w_T = 1$

$$\boldsymbol{h}(\boldsymbol{\theta}, \boldsymbol{y}_t) = y_t^2 - \frac{\nu}{\nu - 2}$$

$$E[y_t^2] = \frac{\nu}{\nu - 2}$$

$$E[y_t^2 - \frac{\nu}{\nu - 2}] = 0$$

$$E[\boldsymbol{h}(\boldsymbol{\theta}, \boldsymbol{y}_t)] = 0$$

$$\boldsymbol{g}_T(\boldsymbol{\theta}) = \frac{1}{T}\sum_{t=1}^{T}(y_t^2 - \frac{\nu}{\nu - 2})$$

In this case, $r = k = 1$, and

$$Q_T(\boldsymbol{\theta}) = [\frac{1}{T}\sum_{t=1}^{T}(y_t^2 - \frac{\nu}{\nu-2})]^2$$

Since $r = k = 1$, $\hat{\nu}_{gmm}$ can be chosen such that $Q_T(\boldsymbol{\theta}) = 0$

Example: t-distribution with r = 2

Suppose we add a moment condition for the t-distribution

If $\nu > 4 \rightarrow \mu_T = E[Y_t^4] = \dfrac{3\nu^2}{(\nu-2)(\nu-4)}$

Now $r = 2 > 1 = k$

That is, we now have more moment conditions than parameters.

We map this problem into the GMM form in the following way:

$\boldsymbol{y}_t$

$\boldsymbol{\theta} = \nu$

$$w_T = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\boldsymbol{h}(\boldsymbol{\theta}, \boldsymbol{y}_t) = \begin{bmatrix} y_t^2 - \dfrac{\nu}{\nu-2} \\ y_t^4 - \dfrac{3\nu^2}{(\nu-2)(\nu-4)} \end{bmatrix}$$

$$\boldsymbol{g}_T(\boldsymbol{\theta}) = \frac{1}{T}\sum_{t=1}^{T} h(\boldsymbol{\theta}, \boldsymbol{y}_t)$$

The weighting matrix $w_T = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ places equal weight on the two moment conditions.

We could alter this matrix to emphasize one condition more than another.

If $\boldsymbol{y}_t$ is strictly stationary and $\boldsymbol{h}$ continuous, a law of large numbers will hold:

$\boldsymbol{g}_T(\boldsymbol{\theta}) \xrightarrow{p} E[\boldsymbol{h}(\boldsymbol{\theta}, \boldsymbol{y}_t)]$

$E[X] = Bob$

$x_1, ..., x_n \sim f_X(x)$

$\dfrac{1}{n}\displaystyle\sum_{i=1}^{n} xi \xrightarrow{p} Bob$

Under certain regularity conditions, it can be shown that

$\boldsymbol{\theta}_{gmm} \xrightarrow{p} \theta_0$