Efficient GMM examples

Robert Kirkby*

May 24, 2024

Abstract

In some situations where the moment conditions are 'separable' efficient GMM can be done in a single step. But more commonly it must be implemented as two-iteration efficient GMM. This document just illustrates a few examples of each. We provide two examples using cross-sectional data. The first example estimates a simple linear regression and is standard, so we use two-iteration efficient GMM. The second example estimates a normal distribution, and since the moment conditions are separable we can simply estimate efficient GMM directly. We then provide two further examples using time-series data. The first example estimates a stochastic volatility model of interest rates, and we use two-iteration efficient GMM. The second example estimates an AR(1) process, and since the moment conditions are separable we can simply estimate efficient GMM directly. We then revisit the AR(1) process, and show that a different choice of moment conditions leads to non-separable moment conditions, and so we use two-iteration efficient GMM.

Keywords: Generalized Method of Moments, Efficient GMM, Two-iteration Efficient GMM.

Sometimes the moment conditions for GMM estimation are 'separable' into a first term which depends on the data (and not the parameters to be estimated) and a second terms which depends only on the parameters to be estimated. Because the first term is a random variable, while the second term is a constant. When moment conditions are separable the covariance matrix of the moment conditions can be estimated as the covariance matrix first term, which depends on the data (and not the parameters to be estimated). So efficient GMM, which uses the inverse of the covariance matrix of the moment conditions can be implemented in a single step when the moment conditions are 'separable'. In the more common situation that moment conditions are not separable, efficient GMM can be implemented via two-iteration efficient GMM, but doing it directly in a single step is not possible.

This document first provides a formal description of 'separable' moment conditions, and how they permit efficient GMM to be directly implemented. It does so for cross-sectional data, and then again for time-series data.

Four examples are provided to illustate how separability is sometimes possible, and is more a property of the kinds of moment conditions being used, rather than a property of models themselves. The first two examples use cross-sectional data, and the last two examples use time-series data.

*Please address all correspondence about this article to Robert Kirkby at <robertdkirkby@gmail.com>, robertdkirkby.com.

The first example is GMM estimation of a linear regression, and the moment conditions are non-separable. The second example is GMM estimation of a normal distribution, and the moment conditions are separable. The third example is GMM estimation of a stochastic volatility model, and the moment conditions are non-separable. The fourth example is GMM estimation of an AR(1) model, this estimation is performed twice for different moment conditions, and the first set of moment conditions are separable, but the second set of moment conditions are non-separable.

Codes implementing all the examples can be found at: github

1 GMM in cross-sectional data

GMM estimation involves estimating a parameter vector θ to satisfy moment conditions $G(\{x_i\}_{i=1}^N, \theta) \equiv \frac{1}{N}\sum_{i=1}^N g(x_i, \theta) = 0$, where $\{x_i\}_{i=1}^N$ is the observed cross-sectional data. That G can be written as the average $(\frac{1}{N}\sum_{i=1}^N)$ of a function of the individual observations is what it means to say that G is a moment. The GMM estimation is given by $\theta^* = \arg\min_\theta G'WG$, where W is a weighting matrix.

Under standard assumptions the GMM estimator θ^* is consistent and asymptotically normal, with asymptotic variance given by $(J'WJ)^{-1}(J'W\Omega WJ)(J'WJ)^{-1}$, where $J \equiv \frac{\partial G(X,\theta)}{\partial\theta}\big|_{\theta=\theta_0}$, and Ω is the variance-covariance matrix of G (at $\theta = \theta_0$). Efficient GMM uses $W = \Omega^{-1}$ which delivers the minimum asymptotic variance, namely $J'\Omega^{-1}J$. So to implement efficient GMM requires us to first estimate $\hat{\Omega}$.

It is always impossible to evaluate the moment conditions, $G(\{x_i\}_{i=1}^N, \theta) = 0$, without first somehow choosing a value for the parameter θ.

The purpose of these examples however, is that sometimes we can estimate the variance-covariance matrix of G, denoted $\hat{\Omega}$ without needing to choose a value for the parameter θ. In particular, this will *not* be possible for the linear regression example, where a value for θ is required to estimate Ω and so efficient GMM has to be implemented as two-iteration efficient GMM. But for our example of the normal distribution, the moment condition can be separated into a random variable term which depends on the data, and a constant term which is a function of the parameter. Since $Cov(Z+c) = Cov(Z)$ for a random variable Z and a constant c, it follows that the covariance matrix of our moment condition is equal to a covariance matrix that can be estimated from the data without needing a value for θ. Efficient GMM for the normal distribution example can therefore simply be implemented as a single step.

This 'separete terms' arises whenever our moment conditions $\frac{1}{N}\sum_{i=1}^N g(x_i, \theta) = 0$ can be expressed as $\frac{1}{N}\sum_{i=1}^N g(x_i, \theta) = \frac{1}{N}\sum_{i=1}^N [f(x_i) - h(\theta)] = \left[\frac{1}{N}\sum_{i=1}^N f(x_i)\right] - h(\theta)$. As now we have $\left[\frac{1}{N}\sum_{i=1}^N f(x_i)\right]$ which is a random variable that depends on the observed data but not the parameter vector θ, and $h(\theta)$ which is just a constant (given θ). So the covariance matrix of $\frac{1}{N}\sum_{i=1}^N g(x_i, \theta)$ is just equal to the covariance matrix of $\left[\frac{1}{N}\sum_{i=1}^N f(x_i)\right]$, which can be estimated from the observed data without any need for us to first choose a value for the parameter vector θ.

When the moment conditions are not separable in the way just described, we can still perform two-iteration efficient GMM. Because GMM is consistent regardless of what weighting matrix is used (as long as it is postive semi-definite) we can first use any weighting matrix, such as the identity matrix, to estimate a first-iteration $\hat{\theta}_1 = \arg\min_\theta G'WG$. Once we have θ_1 we can use this to evaluate the moment conditions, and so we are able to estimate the covariance matrix of the moment conditions, $\hat{\Omega}(\hat{\theta}_1)$. Our second-iteration estimates $\hat{\theta}_2 = \arg\min_\theta G'\hat{\Omega}(\hat{\theta}_1)^{-1}G$. Two-iteration efficient GMM is a standard technique in applications, and for most GMM setups (including two of our three examples) is necessary to be able to implement efficient GMM.

1.1 GMM estimation of a simple linear regression

Our DGP is

$$y_i = \alpha + \beta x_i + u_i,\ u \sim N(0, \sigma^2)$$

in pripnle x_i is any exogenously given data, in our code we will generate it as uniformly distributed.

We can estimate the parameter vector $\theta = [\alpha; \beta]$ using the following two moment conditions,

$$\begin{aligned} E(u_i) &= 0 \\ E(x_i u_i) &= 0 \end{aligned}$$

the first of which is just that the errors should be zero-mean, while the second is the assumption that x is predetermined (a standard assumption for OLS estimation). To implement the GMM estimator we use the sample analogues,

$$\begin{aligned} \frac{1}{N}\sum_{i=1}^{N} u_i &= 0 \\ \frac{1}{N}\sum_{i=1}^{N} x_i u_i &= 0 \end{aligned}$$

Efficient GMM requires us to first estimate Ω, so that we can then use $W = \Omega^{-1}$ as the weighting matrix. By definition $\Omega = E([u_i; x_i u_i][u_i; x_i u_i]')$, which can be estimated as,

$$\hat{\Omega} = \begin{bmatrix} \frac{1}{N}\sum_{i=1}^{N} u_i^2 & \frac{1}{N}\sum_{i=1}^{N} x_i u_i^2 \\ \frac{1}{N}\sum_{i=1}^{N} x_i u_i^2 & \frac{1}{N}\sum_{i=1}^{N} x_i^2 u_i^2 \end{bmatrix}$$

The catch is that we cannot observe u_i. We could easily calculate u_i using the equation $u_i = y_i - \alpha - \beta x_i$, but this requires us to first have α and β. So we cannot just directly estimate $\hat{\Omega}$ and implement efficient GMM as we do not have α and β. We can easily do this as two-iteration efficient GMM, where the first-iteration will give us estimates $[\hat{\alpha}_1; \hat{\beta}_1]$, and then we can easily estimate $\hat{\Omega}([\hat{\alpha}_1; \hat{\beta}_1])$ using the estimator described above. The second-iteration then using the inverse of this as a weighting matrix to estimate $[\hat{\alpha}_2; \hat{\beta}_2]$.

The code implement this two-iteration efficient GMM estimation. The first part of the code simulates data from the DGP, and the second part of the code then performs the estimation based on this data.

For our purposes it is important to observe that there is no way to simplify $\hat{\Omega}$ so that it can be estimated without knowledge of θ.

This is implemented as *GMM_LinearRegression.m*.

1.2 GMM estimation of a normal distribution

Our DGP is

$$x_i \sim N(\mu, \sigma^2)$$

we want to estimate the parameters $\theta = [\mu; \sigma]$.

We estimate this targeting the mean and variance, so the moment conditions are

$$E(x_i - \mu) = 0$$
$$E((x_i - E[x])^2 - \sigma^2) = 0$$

where $E[x]$ is the population mean of x. To implement the GMM estimator we use the sample analogues,

$$\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu) = 0$$
$$\frac{1}{N}\sum_{i=1}^{N}((x_i - \bar{x})^2 - \sigma^2) = 0$$

where $\bar{x}$ is the sample mean of x, $\bar{x} \equiv \frac{1}{N}\sum_{i=1}^{N} x_i$.

The key observation for our purposes is that we can rewrite these moment conditions as,

$$\left[\frac{1}{N}\sum_{i=1}^{N} x_i\right] - \mu = 0$$
$$\left[\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2\right] - \sigma^2 = 0$$

so now there is a term which is a random variable that depends on the data (but not on θ) and a term that is a constant (μ and σ^2).

The covariance matrix of these moments is defined as $\Omega = E([x_i - \mu; (x_i - E[x])^2 - \sigma^2][x_i - \mu; (x_i - E[x])^2 - \sigma^2]')$. But because the moments are separable with μ and σ^2 being constants (for given parameter values), this simplifies to just $\Omega = E([x_i; (x_i - E[x])^2][x_i; (x_i - E[x])^2]')$. And now this can be estimated as

$$\hat{\Omega} = \begin{bmatrix} \frac{1}{N}\sum_{i=1}^{N} x_i^2 & \frac{1}{N}\sum_{i=1}^{N} x_i(x_i - \bar{x})^2 \\ \frac{1}{N}\sum_{i=1}^{N} x_i(x_i - \bar{x})^2 & \frac{1}{N}\sum_{i=1}^{N}((x_i - \bar{x})^2)^2 \end{bmatrix}$$

where $\bar{x}$ is the sample mean of x, $\bar{x} \equiv \frac{1}{N}\sum_{i=1}^{N} x_i$.

Note that because we can estimate $\hat{\Omega}$ without a value for the parameter vector $[\mu, \sigma]$, two-iteration efficient GMM would here just give the exact same solution we get from our efficient GMM anyway (we demonstrate this in the codes).

This is implemented as *GMM_normaldist.m*.

2 GMM in time-series data

GMM estimation involves estimating a parameter vector θ to satisfy moment conditions $G(\{x_t\}_{t=1}^T, \theta) \equiv \frac{1}{T}\sum_{t=1}^T g(x_t, \theta) = 0$, where $\{x_t\}_{t=1}^T$ is the observed time-series data. That G can be written as the average ($\frac{1}{T}\sum_{i=T}^T$) of a function of the individual observations is what it means to say that G is a moment. The GMM estimation is given by $\theta^* = \arg\min_\theta G'WG$, where W is a weighting matrix.

Under standard assumptions the GMM estimator θ^* is consistent and asymptotically normal, with asymptotic variance given by $(J'WJ)^{-1}(J'W\Omega WJ)(J'WJ)^{-1}$, where $J \equiv \frac{\partial G(X,\theta)}{\partial \theta}\Big|_{\theta=\theta_0}$, and Ω is the variance-covariance matrix of G (at $\theta = \theta_0$). Efficient GMM uses $W = \Omega^{-1}$ which delivers the minimum asymptotic variance, namely $J'\Omega^{-1}J$. So to implement efficient GMM requires us to first estimate $\hat{\Omega}$.

It is always impossible to evaluate the moment conditions, $G(\{x_t\}_{t=1}^T, \theta) = 0$, without first somehow choosing a value for the parameter θ.

The purpose of these examples however, is that sometimes we can estimate the variance-covariance matrix of G, denoted $\hat{\Omega}$ without needing to choose a value for the parameter θ. In particular, this will *not* be possible for the stochastic volatility example, where a value for θ is required to estimate Ω and so efficient GMM has to be implemented as two-iteration efficient GMM. But for our example of the normal distribution, the moment condition can be separated into a random variable term which depends on the data, and a constant term which is a function of the parameter. For time series Ω is no longer just the covariance matrix of G, instead it is $\Omega \equiv \sum_{\tau=-\infty}^{\infty} E[g(x_t;\theta)g(x_{t-\tau};\theta)']$.

Since $Cov(Z + c) = Cov(Z)$ for a random variable Z and a constant c, it follows that the covariance matrix of our moment condition is equal to a covariance matrix that can be estimated from the data without needing a value for θ. Note that we can apply this logic for each lead/lag (each τ) in the formula for Ω, and thus it applies to Ω itself. Efficient GMM for the AR(1) example can therefore simply be implemented as a single step.

This 'separete terms' arises whenever our moment conditions $\frac{1}{T}\sum_{t=1}^T g(x_t, \theta) = 0$ can be expressed as $\frac{1}{T}\sum_{t=1}^T g(x_t, \theta) = \frac{1}{T}\sum_{t=1}^T [f(x_t) - h(\theta)] = \left[\frac{1}{T}\sum_{t=1}^T f(x_t)\right] - h(\theta)$. As now we have $\left[\frac{1}{T}\sum_{t=1}^T f(x_t)\right]$ which is a random variable that depends on the observed data but not the parameter vector θ, and $h(\theta)$ which is just a constant (given θ). So the covariance matrix of $\frac{1}{T}\sum_{t=1}^T g(x_t, \theta)$ is just equal to the covariance matrix of $\left[\frac{1}{T}\sum_{t=1}^T f(x_t)\right]$, which can be estimated from the observed data without any need for us to first choose a value for the parameter vector θ.

When the moment conditions are not separable in the way just described, we can still perform two-iteration efficient GMM. Because GMM is consistent regardless of what weighting matrix is used (as long as it is postive semi-definite) we can first use any weighting matrix, such as the identity matrix, to estimate a first-iteration $\hat{\theta_1} = \arg\min_\theta G'WG$. Once we have θ_1 we can use this to evaluate the moment conditions, and so we are able to estimate the covariance matrix of the moment conditions, $\hat{\Omega}(\hat{\theta_1})$. Our second-iteration estimates $\hat{\theta_2} = \arg\min_\theta G'\hat{\Omega}(\hat{\theta_1})^{-1}G$. Two-iteration efficient GMM is a standard technique in applications, and for most GMM setups (including two of our three examples) is necessary to be able to implement efficient GMM.

The time-series case is importantly different from the cross-sectional case. Partly as the underlying theory for the consistency and asymptotically normal results is different. But there is also an important difference in Ω, which now has to be estimated with the Newey-West estimator because of the autocorrelation.

2.1 GMM estimation of a stochastic volatility model

We are interested in a stochastic volatility model of interest rates,

$$\begin{aligned} \Delta r_t &= \alpha + \beta r_{t-1} + u_t \\ u_t &= \sigma r_{t-1}^{\gamma} z_t, \quad z_t \sim N(0,1) \end{aligned}$$

where r_t is the interest rate, and u_t is a shock whose variance depends on r_{t-1}. $\Delta r_t \equiv r_t - r_{t-1}$.

We can estimate the parameters $\theta = [\alpha, \beta, \sigma, \gamma]$ using GMM with the moment conditions,

$$\begin{aligned} E[u_t] &= 0 \\ E[u_t^2] &= \sigma^2 r_{t-1}^{2\gamma} \\ E[u_t r_{t-1}] &= 0 \\ E[(u_t^2 - \sigma^2 r_{t-1}^{2\gamma}) r_{t-1}] &= 0 \end{aligned}$$

note that the these are: first, that errors are mean zero; second, the variance of the innovations; third, exogenous regressors (shocks uncorrelated with regressor); fourth, variance of shocks is uncorrelated with regressor.

To implement the GMM estimator we use the sample analogues,

$$\begin{aligned} \frac{1}{T}\sum_{t=1}^{T} u_t &= 0 \\ \frac{1}{T}\sum_{t=1}^{T} u_t^2 &= \sigma^2 r_{t-1}^{2\gamma} \\ \frac{1}{T}\sum_{t=1}^{T} (u_t r_{t-1}) &= 0 \\ \frac{1}{T}\sum_{t=1}^{T} ((u_t^2 - \sigma^2 r_{t-1}^{2\gamma}) r_{t-1}) &= 0 \end{aligned}$$

Efficient GMM requires us to first estimate Ω, so that we can use $W = \Omega^{-1}$ as the weighting matrix. By definiion $\Omega = \sum_{\tau=-\infty}^{\infty} S_{t,\tau}$, where $S_{t,\tau} = E([u_t; u_t^2; u_t r_{t-1}; (u_t^2 - \sigma^2 r_{t-1}^{2\gamma}) r_{t-1}][u_{t-\tau}; u_{t-\tau}^2; u_{t-\tau} r_{t-\tau-1}; (u_{t-\tau}^2 - \sigma^2 r_{t-\tau-1}^{2\gamma}) r_{t-\tau-1}]')$, which can be estimated as,

$$\hat{S}_{t,\tau} = \begin{bmatrix} \frac{1}{T}\sum_{t=1}^{T} u_t^2 & \frac{1}{T}\sum_{t=1}^{T} u_t^3 & \frac{1}{T}\sum_{t=1}^{T} u_t(u_t r_{t-1}) & \cdots \\ \frac{1}{T}\sum_{t=1}^{T} u_t^3 & \frac{1}{T}\sum_{t=1}^{T} u_t^4 & \frac{1}{T}\sum_{t=1}^{T} u_t^2(u_t r_{t-1}) & \cdots \\ \frac{1}{T}\sum_{t=1}^{T} u_t(u_t r_{t-1}) & \frac{1}{T}\sum_{t=1}^{T} u_t^2(u_t r_{t-1}) & \frac{1}{T}\sum_{t=1}^{T}(u_t r_{t-1})^2 & \cdots \\ \frac{1}{T}\sum_{t=1}^{T} u_t((u_t^2 - \sigma^2 r_{t-1}^{2\gamma}) r_{t-1}) & \frac{1}{T}\sum_{t=1}^{T} u_t^2((u_t^2 - \sigma^2 r_{t-1}^{2\gamma}) r_{t-1}) & \frac{1}{T}\sum_{t=1}^{T}(u_t r_{t-1})((u_t^2 - \sigma^2 r_{t-1}^{2\gamma}) r_{t-1}) & \cdots \end{bmatrix}$$

and then $\hat{\Omega} = \sum_{\tau=-L}^{L} k(\tau, L)\hat{S}_{t,\tau}$, where $k(\tau, L)$ is a weighting kernel, and we use the Newey-West kernel $k(\tau, L) = (1 - \frac{\tau}{L+1})$ and set $L = floor(4 * ((T/100)^(2/9)))$ following the Newey-West plug-in procedure.

The catch is that we cannot observe u_t. We could easily calculate u_t using the equation $u_t = r_t - r_{t-1} - \alpha - \beta r_{t-1}$, but this requires us to first have α and β. So we cannot just directly estimate $\hat{\Omega}$ and implement efficient GMM as we do not have α and β. We can easily do this as two-iteration efficient GMM, where the first-iteration will give us estimates $[\hat{\alpha}_1; \hat{\beta}_1; \hat{\sigma}_1; \hat{\gamma}_1]$, and then we can easily estimate $\hat{\Omega}([\hat{\alpha}_1; \hat{\beta}_1])$ using the estimator described above. The second-iteration then using the inverse of this as a weighting matrix to estimate $[\hat{\alpha}_2; \hat{\beta}_2; \hat{\sigma}_2; \hat{\gamma}_2]$.

The code implement this two-iteration efficient GMM estimation. The first part of uses the interest rates on US 3-month Treasury's (and requires getFredData), and the second part of the code then performs the estimation based on this data.

For our purposes it is important to observe that there is no way to simplify $\hat{\Omega}$ so that it can be estimated without knowledge of θ.

This is implemented as *GMM_stochvolatility.m*.

2.2 GMM estimation of an AR(1) model

We are interested an AR(1) process,

$$y_t = \rho y_{t-1} + e_t, \quad e_t \sim N(0, \sigma^2)$$

We will GMM estimate the parameters $\theta = [\rho, \sigma]$ of this AR(1) model twice, the first time with moments that are separable, the second time using non-separable moments. This helps illustrate that whether or not the moments will be separable is mostly a question of the kinds of moments that are being estimated.

2.2.1 AR(1) with separable moments

We can estimate the parameters $\theta = [\rho, \sigma]$ using GMM with the moment conditions,

$$\begin{aligned} E[(y_t - \bar{Y})^2 - \frac{\sigma^2}{1-\rho^2}] = 0 \\ E[(y_t - \bar{Y})(y_{t-1} - \bar{Y}) - \rho\frac{\sigma^2}{1-\rho^2}] = 0 \\ E[(y_t - \bar{Y})(y_{t-2} - \bar{Y}) - \rho^2\frac{\sigma^2}{1-\rho^2}] = 0 \end{aligned}$$

note that the these are: first, the variance; second, the first-order auto-covariance; third, the second-order auto-covariance (the analytical formulae for these moments of an AR(1) process are standard and can be found on wikipedia or in any textbook). $\bar{Y}$ is the population mean of Y, that is $\bar{Y} = E[Y]$.

To implement the GMM estimator we use the sample analogues,

$$\frac{1}{T}\sum_{t=1}^{T}((y_t-\bar{y})^2-\frac{\sigma^2}{1-\rho^2})=0$$

$$\frac{1}{T}\sum_{t=1}^{T}((y_t-\bar{y})(y_{t-1}-\bar{y})-\rho\frac{\sigma^2}{1-\rho^2})=0$$

$$\frac{1}{T}\sum_{t=1}^{T}((y_t-\bar{y})(y_{t-2}-\bar{y})-\rho^2\frac{\sigma^2}{1-\rho^2})=0$$

where $\bar{y}$ is the sample mean, that is $\bar{y}=\frac{1}{T}\sum_{t=1}^{T}y_t$.

The key observation for our purposes is that we can rewrite these moment conditions as,

$$[\frac{1}{T}\sum_{t=1}^{T}(y_t-\bar{y})^2]-\frac{\sigma^2}{1-\rho^2}=0$$

$$[\frac{1}{T}\sum_{t=1}^{T}(y_t-\bar{y})(y_{t-1}-\bar{y})]-\rho\frac{\sigma^2}{1-\rho^2}=0$$

$$[\frac{1}{T}\sum_{t=1}^{T}(y_t-\bar{y})(y_{t-2}-\bar{y})]-\rho^2\frac{\sigma^2}{1-\rho^2}=0$$

so now there is a term which is a random variable that depends on the data (but not on θ) and a term that is a constant ($\frac{\sigma^2}{1-\rho^2}$, $\rho\frac{\sigma^2}{1-\rho^2}$, and $\rho^2\frac{\sigma^2}{1-\rho^2}$).

The covariance matrix of these moments is defined as $\Omega=E([(y_t-\bar{Y})^2-\frac{\sigma^2}{1-\rho^2};(y_t-\bar{Y})(y_{t-1}-\bar{Y})-\rho\frac{\sigma^2}{1-\rho^2};(y_t-\bar{Y})(y_{t-2}-\bar{Y})-\rho^2\frac{\sigma^2}{1-\rho^2}][(y_t-\bar{Y})^2-\frac{\sigma^2}{1-\rho^2};(y_t-\bar{Y})(y_{t-1}-\bar{Y})-\rho\frac{\sigma^2}{1-\rho^2};(y_t-\bar{Y})(y_{t-2}-\bar{Y})-\rho^2\frac{\sigma^2}{1-\rho^2}]')$. But because the moments are separable with the second terms being constants (for given parameter values), this simplifies to just $\Omega=E([(y_t-\bar{Y})^2;(y_t-\bar{Y})(y_{t-1}-\bar{Y});(y_t-\bar{Y})(y_{t-2}-\bar{Y})][(y_t-\bar{Y})^2;(y_t-\bar{Y})(y_{t-1}-\bar{Y});(y_t-\bar{Y})(y_{t-2}-\bar{Y})]')$. And now this can be estimated as

$$\hat{\Omega}=\begin{bmatrix} \frac{1}{T}\sum_{t=1}^{T}(y_t-\bar{y})^4 & \frac{1}{T}\sum_{t=1}^{T}(y_t-\bar{y})^2(y_t-\bar{y})(y_{t-1}-\bar{y}) & \frac{1}{T}\sum_{t=1}^{T}(y_t-\bar{y})^2(y_t- \\ \frac{1}{T}\sum_{t=1}^{T}(y_t-\bar{y})^2(y_t-\bar{y})(y_{t-1}-\bar{y}) & \frac{1}{T}\sum_{t=1}^{T}(y_t-\bar{y})^2(y_{t-1}-\bar{y})^2 & \frac{1}{T}\sum_{t=1}^{T}(y_t-\bar{y})(y_{t-1}-\bar{y})(\\ \frac{1}{T}\sum_{t=1}^{T}(y_t-\bar{y})^2(y_t-\bar{y})(y_{t-2}-\bar{y}) & \frac{1}{T}\sum_{t=1}^{T}(y_t-\bar{y})(y_{t-1}-\bar{y})(y_t-\bar{y})(y_{t-2}-\bar{y}) & \frac{1}{T}\sum_{t=1}^{T}(y_t-\bar{y})^2(y \end{bmatrix}$$

where $\bar{y}$ is the sample mean of y, $\bar{y}\equiv\frac{1}{T}\sum_{t=1}^{T}y_t$.

Note that because we can estimate $\hat{\Omega}$ without a value for the parameter vector $[\rho,\sigma]$, two-iteration efficient GMM would here just give the exact same solution we get from our efficient GMM anyway (we demonstrate this in the codes).

This is implemented as $GMM\_AR1.m$.

2.2.2 AR(1) with non-separable moments

We can estimate the parameters $\theta = [\rho, \sigma]$ using GMM with the moment conditions,

$$
\begin{aligned}
E[e_t] &= 0 \\
E[e_t y_{t-1}] &= 0
\end{aligned}
$$

note that the these are: first, the innovations are mean zero; second, exogenous regressors (shocks uncorrelated with regressor, which is the lag).

To implement the GMM estimator we use the sample analogues,

$$
\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T} e_t &= 0 \\
\frac{1}{T}\sum_{t=1}^{T} e_t y_{t-1} &= 0
\end{aligned}
$$

Efficient GMM requires us to first estimate Ω, so that we can use $W = \Omega^{-1}$ as the weighting matrix. By definition $\Omega = \sum_{\tau=-\infty}^{\infty} S_{t,\tau}$, where $S_{t,\tau} = E([e_t; e_t y_{t-1}][e_t; e_t y_{t-1}]')$, which can be estimated as,

$$
\hat{S}_{t,\tau} = \begin{bmatrix} \frac{1}{T}\sum_{t=1}^{T} e_t^2 & \frac{1}{T}\sum_{t=1}^{T} e_t^2 y_{t-1} \\ \frac{1}{T}\sum_{t=1}^{T} e_t^2 y_{t-1} & \frac{1}{T}\sum_{t=1}^{T} e_t^2 y_{t-1}^2 \end{bmatrix}
$$

and then $\hat{\Omega} = \sum_{\tau=-L}^{L} k(\tau, L)\hat{S}_{t,\tau}$, where $k(\tau, L)$ is a weighting kernel, and we use the Newey-West kernel $k(\tau, L) = (1 - \frac{\tau}{L+1})$ and set $L = floor(4 * ((T/100)^(2/9)))$ following the Newey-West plug-in procedure.

The catch is that we cannot observe e_t. We could easily calculate e_t using the equation $e_t = y_t - \rho y_{t-1}$, but this requires us to first have ρ. So we cannot just directly estimate $\hat{\Omega}$ and implement efficient GMM as we do not have ρ. We can easily do this as two-iteration efficient GMM, where the first-iteration will give us estimates $[\hat{\rho}_1; \hat{\sigma}_1]$, and then we can easily estimate $\hat{\Omega}(\hat{\rho}_1)$ using the estimator described above. The second-iteration then using the inverse of this as a weighting matrix to estimate $[\hat{\rho}_2; \hat{\sigma}_2]$.

For our purposes it is important to observe that there is no way to simplify $\hat{\Omega}$ so that it can be estimated without knowledge of θ.

This is implemented as $GMM\_AR1\_B.m$.

3 Conclusion

Some models have 'separable' GMM moment conditions, and when this occurs we can implement efficient GMM in a single step. More commonly, efficient GMM must be implemented as two-iteration efficient GMM (more than two iterations can be used, but using two is standard practice as gains from, e.g., seven-iteration efficient GMM are considered small in practice).

As can be seen from the examples here, it is the kind of moment conditions used to estimate the model that is key to whether or not the moments are 'seperable'. It is much less about what kind of model is being estimated, and whether we have cross-sectional, time-series, or panel data is unimportant. It is also unimportant if the model is just-identified or over-identified; we saw examples of separable moments in both situations. In particular, moment conditions that can be described as 'matching the mean/variance/etc.' of the data are typically separable, while moment conditions that are about the properties of residuals are typically non-separable.

That the moments are separable is fairly unimportant for the models being estimated here, in the sense that the runtime costs of using two-iteration efficient GMM when the moment conditions are separable is negligible. However there are situations, like GMM estimation of life-cycle models, where the runtime gains of not having to do two iterations are substantial. In GMM estimation of life-cycle model the use of separable moments is wide-spread in applications and so GMM estimation of life-cycle models can benefit from taking advantage of when moment conditions are separable.

References