GMM and SMM notes and comments for Tyler Ransom

Richard W. Evans

October 2, 2020

1 GMM

1.0 GMM General Specification

My most general definition of GMM is the following. Let $\boldsymbol{x}$ be an $N \times M$ matrix of data with N observations (rows) and M variables (columns) and typical element $x_{n,m}$. Let $\boldsymbol{\theta} \in \mathbb{R}^K$ be the K-dimensional parameter vector to be estimated, and let $\hat{\boldsymbol{\theta}}_{GMM}$ be the GMM estimator of that parameter vector. The GMM estimator is the one that minimizes some general norm (distance measure) on an R-dimensional vector of model moments $\boldsymbol{m}\big(\boldsymbol{x}|\boldsymbol{\theta}\big)$ and a corresponding R-dimensional vector of data moments $\boldsymbol{m}\big(\boldsymbol{x}\big)$ such that $R \geq K$.

$$\hat{\boldsymbol{\theta}}_{GMM}: \quad \min_{\boldsymbol{\theta}} \big\| \boldsymbol{m}\big(\boldsymbol{x}|\boldsymbol{\theta}\big) - \boldsymbol{m}\big(\boldsymbol{x}\big) \big\| \tag{1}$$

The most common GMM criterion function or choice of norm is a weighted sum of squared moment errors. We can first define an $R \times 1$ error vector $\boldsymbol{e}\big(\boldsymbol{x}, \boldsymbol{\theta}\big)$ as the following function of the data $\boldsymbol{x}$ and the parameter vector $\boldsymbol{\theta}$ which can either be given in levels or in percentage differences (there are advantages to each).

$$\boldsymbol{e}\big(\boldsymbol{x}, \boldsymbol{\theta}\big) \equiv \begin{cases} \boldsymbol{m}\big(\boldsymbol{x}|\boldsymbol{\theta}\big) - \boldsymbol{m}\big(\boldsymbol{x}\big) \\ \quad \text{or} \\ \frac{\boldsymbol{m}\big(\boldsymbol{x}|\boldsymbol{\theta}\big) - \boldsymbol{m}\big(\boldsymbol{x}\big)}{\boldsymbol{m}\big(\boldsymbol{x}\big)} \end{cases} \tag{2}$$

The most common weighted sum of squared moment errors GMM specification takes the following form,

$$\hat{\boldsymbol{\theta}}_{GMM}: \quad \min_{\boldsymbol{\theta}} \, \boldsymbol{e}\big(\boldsymbol{x}, \boldsymbol{\theta}\big)^T \boldsymbol{W} \, \boldsymbol{e}\big(\boldsymbol{x}, \boldsymbol{\theta}\big) \tag{3}$$

where $\boldsymbol{W}$ is an $R \times R$ weighting matrix. The standard arguments regarding asymptotic biasedness and efficiency apply to the weighting matrix. An optimal weighting matrix is most efficient (inverse variance estimators, take care of heteroscedasticity and autocorrelation), but identity matrix is unbiased.

1.1 GMM Linear Regression: Method 1

Define a specific linear regression model as the following equation with the following assumptions on the error terms,

$$\begin{aligned} y_i =& \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + ...\beta_K x_{K,i} + \varepsilon_i \\ & \text{where} \quad \varepsilon_i \sim f(\varepsilon), \quad E[\varepsilon_i] = 0, \quad \text{and} \quad E[x_{k,i}\varepsilon_i] = 0 \quad \forall k \end{aligned} \tag{4}$$

where the data are the matrix of $i = 1, 2, ...N$ observations on $K + 1$ variables $\big(y_i, x_{1,i}, x_{2,i}, ...x_{K,i}\big)$, the parameter vector to be estimated is the $K + 1$ vector $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, ...\beta_K)$, and the error terms ε_i are i.i.d. according to the general pdf $f(\varepsilon)$ with the moment conditions $E[\varepsilon_i] = 0$ and $E[x_{k,i}\varepsilon_i] = 0$ for all k.

The most standard way of teaching GMM using linear regression is to use the $K+1$ moments implied by the unconditional expected value condition $E[\varepsilon_i] = 0$ and the K conditional expected value conditions $E[x_{k,i}\varepsilon_i] = 0$ for all k. To map this into the notation in Section 1.0, let the vector of data moments be the vector of $K + 1$ zeros assumed by the $K + 1$ zero-expected values on ε_i.

$$\boldsymbol{m}\big(\boldsymbol{y}, \boldsymbol{x}\big) = \begin{bmatrix} E[\varepsilon_i] \\ E[x_{1,i}\varepsilon_i] \\ E[x_{2,i}\varepsilon_i] \\ \vdots \\ E[x_{K,i}\varepsilon_i] \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{5}$$

These $K \times 1$ data moments are zero by assumption and do not actually use the data $\boldsymbol{y}$ or $\boldsymbol{x}$ in this case.

The model moments from Section 1.0 were specified as a function of data and parameters $\boldsymbol{m}\big(\boldsymbol{x}|\boldsymbol{\theta}\big)$. In this linear regression case, the $K \times 1$ model moments m_k are defined by solving the linear regression model (4) for ε_i and taking the sample averages that correspond to the assumed unconditional and conditional expectations on ε_i.

$$\boldsymbol{m}\big(\boldsymbol{y}, \boldsymbol{x}|\boldsymbol{\beta}\big) = \begin{bmatrix} \frac{1}{N}\sum_{i=1}^{N}\big(y_i - \beta_0 - \beta_1 x_{1,i} - \beta_2 x_{2,i} - ...\beta_K x_{K,i}\big) \\ \frac{1}{N}\sum_{i=1}^{N}\big(y_i - \beta_0 - \beta_1 x_{1,i} - \beta_2 x_{2,i} - ...\beta_K x_{K,i}\big)x_{1,i} \\ \frac{1}{N}\sum_{i=1}^{N}\big(y_i - \beta_0 - \beta_1 x_{1,i} - \beta_2 x_{2,i} - ...\beta_K x_{K,i}\big)x_{2,i} \\ \vdots \\ \frac{1}{N}\sum_{i=1}^{N}\big(y_i - \beta_0 - \beta_1 x_{1,i} - \beta_2 x_{2,i} - ...\beta_K x_{K,i}\big)x_{K,i} \end{bmatrix} \tag{6}$$

Define the moment error vector, analogous to (2) above, as the following function of data and parameters.

$$\boldsymbol{e}\big(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{\beta}\big) \equiv \boldsymbol{m}\big(\boldsymbol{y}, \boldsymbol{x}|\boldsymbol{\beta}\big) - \boldsymbol{m}\big(\boldsymbol{y}, \boldsymbol{x}\big) = \boldsymbol{m}\big(\boldsymbol{y}, \boldsymbol{x}|\boldsymbol{\beta}\big) \tag{7}$$

Note that we cannot use the percent deviation form of the moment errors shown in the second specification of (2) because the denominator is zero.

The GMM estimator in this case $\hat{\boldsymbol{\beta}}_{GMM1}$ is the following exactly identified problem of choosing $K+1$ parameters to minimize the weighted sum of squared errors in $K+1$ moment conditions,

$$\hat{\boldsymbol{\beta}}_{GMM1}: \quad \min_{\boldsymbol{\beta}} \boldsymbol{e}\big(\boldsymbol{y},\boldsymbol{x},\boldsymbol{\beta}\big)^T \boldsymbol{W}\, \boldsymbol{e}\big(\boldsymbol{y},\boldsymbol{x},\boldsymbol{\beta}\big) \tag{8}$$

where $\boldsymbol{W}$ is a $\big(K+1 \times K+1\big)$ weighting matrix.

1.2 GMM Linear Regression: Method 2

The second method for estimating the parameter vector $\boldsymbol{\beta}$ by GMM is just weighted linear least squares. For nonlinear model cases, this is just weighted nonlinear least squares. This ends up being an overidentified GMM estimation in which each error term is treated as a model moment. This approach does not really use any of the expected value of ε_i conditions in (4). It just uses the idea that we want to choose the parameter vector $\boldsymbol{\beta}$ to maximize the fit of the predicted values of the endogenous variable $\hat{y}_i$ to the data values of the endogenous variable y_i. This is equivalent to minimizing the absolute value of the error terms ε_i.

Similar to Method 1 in Section 1.1, we assume the data moments are a vector of zeros. However, in this case, the data moment vector is length N, a zero for each observation of ε_i.

$$\boldsymbol{m}\big(\boldsymbol{y},\boldsymbol{x}\big) = \begin{bmatrix} \min|\varepsilon_1| \\ \min|\varepsilon_2| \\ \vdots \\ \min|\varepsilon_N| \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{9}$$

Similar to (5), these $N \times 1$ data moments are zero by assumption and do not actually use the data $\boldsymbol{y}$ or $\boldsymbol{x}$ in this case.

The model moments are simply the model error terms found by solving linear regression equation (4) for ε_i for each of the $i = 1, 2, ...N$ observations. Note that each model error term is a function of data $\boldsymbol{y}$ and $\boldsymbol{x}$ and paramters $\boldsymbol{\beta}$.

$$\boldsymbol{m}\big(\boldsymbol{y},\boldsymbol{x}|\boldsymbol{\beta}\big) = \begin{bmatrix} y_1 - \beta_0 - \beta_1 x_{1,1} - \beta_2 x_{2,1} - ...\beta_K x_{K,1} \\ y_2 - \beta_0 - \beta_1 x_{1,2} - \beta_2 x_{2,2} - ...\beta_K x_{K,2} \\ \vdots \\ y_N - \beta_0 - \beta_1 x_{1,N} - \beta_2 x_{2,N} - ...\beta_K x_{K,N} \end{bmatrix} \tag{10}$$

Define the $N \times 1$ moment error vector, analogous to (2), as the following function of data and parameters.

$$\boldsymbol{e}\big(\boldsymbol{y},\boldsymbol{x},\boldsymbol{\beta}\big) \equiv \boldsymbol{m}\big(\boldsymbol{y},\boldsymbol{x}|\boldsymbol{\beta}\big) - \boldsymbol{m}\big(\boldsymbol{y},\boldsymbol{x}\big) = \boldsymbol{m}\big(\boldsymbol{y},\boldsymbol{x}|\boldsymbol{\beta}\big) \tag{11}$$

As with (7), we cannot use the percent deviation form of the moment errors shown in the second specification of (2) because the denominator is zero.

The GMM estimator in this case $\hat{\boldsymbol{\beta}}_{GMM2}$ is the following overidentified problem of choosing $K+1$ parameters to minimize the weighted sum of squared errors in N moment conditions,

$$\hat{\boldsymbol{\beta}}_{GMM2}: \quad \min_{\boldsymbol{\beta}} \boldsymbol{e}\big(\boldsymbol{y},\boldsymbol{x},\boldsymbol{\beta}\big)^T \boldsymbol{W}\, \boldsymbol{e}\big(\boldsymbol{y},\boldsymbol{x},\boldsymbol{\beta}\big) \tag{12}$$

where $\boldsymbol{W}$ is a $\big(N \times N\big)$ weighting matrix.

1.3 GMM Comparison of Methods

2 SMM

2.0 SMM General Specification

Simulated method of moments (SMM) is analogous to the generalized method of moments (GMM) estimator. SMM could really be thought of as a particular type of GMM estimator. The SMM estimator chooses model parameters $\boldsymbol{\theta}$ to make simulated model moments match data moments. Seminal papers developing SMM are McFadden (1989), Lee and Ingram (1991), and Duffie and Singleton (1993). Good textbook treatments of SMM are found in Adda and Cooper (2003, pp.87-100) and Davidson and MacKinnon (2004, pp. 383-394).

In the general specification of GMM estimation from Section 1.0, we use data $\boldsymbol{x}$ and model parameters $\boldsymbol{\theta}$ to minimize the distance between model moments $\boldsymbol{m}(\boldsymbol{x}|\boldsymbol{\theta})$ and data moments $\boldsymbol{m}(\boldsymbol{x})$.

$$\hat{\boldsymbol{\theta}}_{GMM}: \quad \min_{\boldsymbol{\theta}} \big\| \boldsymbol{m}\big(\boldsymbol{x}|\boldsymbol{\theta}\big) - \boldsymbol{m}\big(\boldsymbol{x}\big) \big\| \tag{1}$$

The following difficulties can arise with GMM making it not possible or very difficult.

- The model moment function $\boldsymbol{m}(\boldsymbol{x}|\boldsymbol{\theta})$ is not known analytically.
- The data moments you are trying to match come from another model (indirect inference).
- The model moments $\boldsymbol{m}(\boldsymbol{x}|\boldsymbol{\theta})$ are derived from latent variables that are not observed by the modeler. You only have moments, not the underlying data. See Laroque and Salanie (1993).
- The model moments $\boldsymbol{m}(\boldsymbol{x}|\boldsymbol{\theta})$ are derived from censored variables that are only partially observed by the modeler.
- The model moments $\boldsymbol{m}(\boldsymbol{x}|\boldsymbol{\theta})$ are just difficult to derive analytically. Examples include moments that include multiple integrals over nonlinear functions as in McFadden (1989).

Let $\boldsymbol{x}$ be the $N \times M$ vector of data as in Section 1.0, and let $\tilde{\boldsymbol{x}}_s$ be the sth simulation of the dataset $\boldsymbol{x}$. SMM estimation is simply to simulate the model data S times, and use the average values of the moments from the simulated data as the estimator for the model moments. Let $\tilde{\boldsymbol{x}} = \{\tilde{\boldsymbol{x}}_1, \tilde{\boldsymbol{x}}_2, ...\tilde{\boldsymbol{x}}_s, ...\tilde{\boldsymbol{x}}_S\}$ be the S simulations of the model data.

It is important to note a major difference between SMM and GMM. In the GMM examples in Section 1, the parameter vector $\boldsymbol{\theta}$ could be estimated without strong distributional assumptions about the random variable error term ε_i. In the first linear regression example in Section 1.1 with the $K+1$ moments, all that was needed was that the errors be distributed i.i.d. according to general pdf $\varepsilon_i \sim f(\varepsilon)$ and the two sets of moment conditions on the distribution of the errors $E[\varepsilon_i] = 0$ and $E[\varepsilon_i x_{k,i}] = 0$. In SMM, we need to know the exact distribution $f(\varepsilon)$ in order to simulate the data. For example, if we assume the errors are distributed normally, we need to know the parameters of the distribution (variance σ^2 in the case of the normal distribution). We need to be able

The $R \times 1$ data moment vector $\boldsymbol{m}\big(\boldsymbol{x}\big)$ is the same as in the GMM specification of Section 1.0. We can compute a set of R model moments $\boldsymbol{m}\big(\tilde{\boldsymbol{x}}_s|\boldsymbol{\theta}\big)$ from each of the simulated datasets $\tilde{\boldsymbol{x}}_s$. One conceptual advantage of SMM relative to GMM is that the model moments are computed in the same way as the data moments because the model moments are computed from simulated datasets $\tilde{\boldsymbol{x}}_s$ that are supposed to be representative of the real world dataset $\boldsymbol{x}$. The final model moment vector $\hat{\boldsymbol{m}}\big(\tilde{\boldsymbol{x}}|\boldsymbol{\theta}\big)$ to be used in SMM is an average across each of the moment vectors for each simulation $\boldsymbol{m}\big(\tilde{\boldsymbol{x}}_s|\boldsymbol{\theta}\big)$.

$$\hat{\boldsymbol{m}}\big(\tilde{\boldsymbol{x}}|\boldsymbol{\theta}\big) \equiv \frac{1}{S}\sum_{s=1}^{S} \boldsymbol{m}\big(\tilde{\boldsymbol{x}}_s|\boldsymbol{\theta}\big) \tag{13}$$

Once we have an estimate of the model moments $\hat{\boldsymbol{m}}\big(\tilde{\boldsymbol{x}}|\boldsymbol{\theta}\big)$ from our S simulations, SMM estimation is very similar to our presentation of GMM. The SMM approach of estimating the parameter vector $\hat{\boldsymbol{\theta}}_{SMM}$ is to choose $\boldsymbol{\theta}$ to minimize some distance measure of the simulated model moments $\hat{\boldsymbol{m}}\big(\tilde{\boldsymbol{x}}|\boldsymbol{\theta}\big)$ from the data moments $\boldsymbol{m}\big(\boldsymbol{x}\big)$.

$$\hat{\boldsymbol{\theta}}_{SMM}: \quad \min_{\boldsymbol{\theta}} \big\|\hat{\boldsymbol{m}}\big(\tilde{\boldsymbol{x}}|\boldsymbol{\theta}\big) - \boldsymbol{m}\big(\boldsymbol{x}\big)\big\| \tag{14}$$

The distance measure $||\hat{\boldsymbol{m}}(\tilde{\boldsymbol{x}}|\boldsymbol{\theta}) - \boldsymbol{m}(\boldsymbol{x})||$ can be any kind of norm. But it is important to recognize that your estimates $\hat{\boldsymbol{\theta}}_{SMM}$ will be dependent on what distance measure (norm) you choose. The most widely studied and used distance metric in GMM and SMM estimation is the L^2 norm or the weighted sum of squared errors in moments. Define the moment error function $\boldsymbol{e}(\tilde{\boldsymbol{x}}, \boldsymbol{x}|\boldsymbol{\theta})$ as the simple difference or percent difference in the vector of simulated model moments from the data moments.

$$\boldsymbol{e}\big(\tilde{\boldsymbol{x}}, \boldsymbol{x}, \boldsymbol{\theta}\big) \equiv \begin{cases} \hat{\boldsymbol{m}}\big(\tilde{\boldsymbol{x}}|\boldsymbol{\theta}\big) - \boldsymbol{m}\big(\boldsymbol{x}\big) \\ \quad \text{or} \\ \frac{\hat{\boldsymbol{m}}\big(\tilde{\boldsymbol{x}}|\boldsymbol{\theta}\big) - \boldsymbol{m}\big(\boldsymbol{x}\big)}{\boldsymbol{m}\big(\boldsymbol{x}\big)} \end{cases} \tag{15}$$

The weighted sum of squared simulated moment errors SMM specification takes the following form,

$$\hat{\boldsymbol{\theta}}_{SMM}: \quad \min_{\boldsymbol{\theta}} \; \boldsymbol{e}\big(\tilde{\boldsymbol{x}}, \boldsymbol{x}, \boldsymbol{\theta}\big)^T \, \boldsymbol{W} \, \boldsymbol{e}\big(\tilde{\boldsymbol{x}}, \boldsymbol{x}, \boldsymbol{\theta}\big) \tag{16}$$

where W is a $R \times R$ weighting matrix in the criterion function. We call the quadratic form expression $e(\tilde{x}, x|\theta)^T \, W \, e(\tilde{x}, x|\theta)$ the criterion function because it is a strictly positive scalar that is the object of the minimization in the SMM problem statement. The $R \times R$ weighting matrix W in the criterion function allows the econometrician to control how each moment is weighted in the minimization problem. For example, an $R \times R$ identity matrix for W would give each moment equal weighting, and the criterion function would be a simply sum of squared percent deviations (errors). Other weighting strategies can be dictated by the nature of the problem or model.

One last item to emphasize with SMM is that the errors that are drawn for the S simulations of the model must be drawn only once so that the minimization problem for $\hat{\boldsymbol{\theta}}_{SMM}$ does not have the underlying sampling changing for each guess of a value of $\boldsymbol{\theta}$. Put more simply, you want the random draws for all the simulations to be held constant so that the only thing changing in the minimization problem is the value of the vector of parameters $\boldsymbol{\theta}$.

2.1 SMM Linear Regression: Method 1

The SMM approach analogous to the GMM approach in Section 1.1 does not work. In most cases, the parameters are not identified. Loss of identification comes first because we have a maximum of $K + 1$ moments with potentially $K + 2$ parameters to estimate. A second characteristic that takes away a degree of freedom in this case is that one of the moment conditions $E[\varepsilon_i] = 0$ might be true for many parameter values in the assumed random variable distribution

In SMM, we need to modify the regression equation (4) to include a specific parameterized pdf for the i.i.d. random shocks ε_i. We assume here a normal distribution, which could add another parameter σ to our vector of parameters to be estimated $\boldsymbol{\theta}$.

$$\begin{aligned} y_i =& \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + ...\beta_K x_{K,i} + \varepsilon_i \\ & \text{where} \quad \varepsilon_i \sim N(0, \sigma^2), \quad E[\varepsilon_i] = 0, \quad \text{and} \quad E[x_{k,i}\varepsilon_i] = 0 \quad \forall k \end{aligned} \tag{17}$$

The $K + 1$ model moment conditions are either not functions of the parameters

2.2 SMM Linear Regression: Method 2

Only the SMM method analogous to the second GMM method in Section 1.2 will work. The data $\boldsymbol{y}$ and $\boldsymbol{x}$ in linear regression model (17) has N observations on one dependent variable y_i and K independent variables $x_{k,i}$. The parameter vector to be estimated is now $\boldsymbol{\theta} = \big(\beta_0, \beta_1, ...\beta_K, \sigma\big)$, which has length $K + 2$ and includes the standard deviation σ of the mean-zero, normally distributed random shocks ε_i.

In this case, let the vector of data moments be the vector of N dependent variable values y_i.

$$\boldsymbol{m}(\boldsymbol{y},\boldsymbol{x}) = \boldsymbol{m}(\boldsymbol{y}) = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} = \boldsymbol{y} \tag{18}$$

Then let the data moments for the sth simulation be the vector of predicted values including the simulated error terms $\varepsilon_{s,i}$,

$$\boldsymbol{m}(\tilde{\boldsymbol{y}}_s, \tilde{\boldsymbol{x}}_s|\boldsymbol{\theta}) = \boldsymbol{m}(\tilde{\boldsymbol{x}}_s|\boldsymbol{\theta}) = \begin{bmatrix} \beta_0 + \beta_1 x_{1,1} + \beta_2 x_{2,1} + ...\beta_K x_{K,1} + \varepsilon_{1,s} \\ \beta_0 + \beta_1 x_{1,2} + \beta_2 x_{2,2} + ...\beta_K x_{K,2} + \varepsilon_{2,s} \\ \vdots \\ \beta_0 + \beta_1 x_{1,N} + \beta_2 x_{2,N} + ...\beta_K x_{K,N} + \varepsilon_{N,s} \end{bmatrix} = \begin{bmatrix} \hat{y}_{1,s} \\ \hat{y}_{2,s} \\ \vdots \\ \hat{y}_{N,s} \end{bmatrix} = \hat{\boldsymbol{y}}_s \tag{19}$$

where $\varepsilon_{i,s}$ is the ith observation error term in the sth simulation of the errors, and $\hat{y}_{i,s}$ is the ith observation predicted value of the dependent variable in the sth simulation of the errors.

The final model moment vector $\hat{\boldsymbol{m}}(\tilde{\boldsymbol{x}}|\boldsymbol{\theta})$ to be used in SMM is an average across each of the moment vectors for each simulation $\boldsymbol{m}(\tilde{\boldsymbol{x}}_s|\boldsymbol{\theta})$.

$$\hat{\boldsymbol{m}}(\tilde{\boldsymbol{x}}|\boldsymbol{\theta}) \equiv \frac{1}{S}\sum_{s=1}^{S} \boldsymbol{m}(\tilde{\boldsymbol{x}}_s|\boldsymbol{\theta}) = \frac{1}{S}\sum_{s=1}^{S} \hat{\boldsymbol{y}}_{\boldsymbol{s}} = \hat{\boldsymbol{y}} \tag{20}$$

Define the moment error function as the following.

$$\boldsymbol{e}(\tilde{\boldsymbol{x}}, \boldsymbol{y}, \boldsymbol{\theta}) = \hat{\boldsymbol{m}}(\tilde{\boldsymbol{x}}|\boldsymbol{\theta}) - \boldsymbol{m}(\boldsymbol{y}) = \hat{\boldsymbol{y}} - \boldsymbol{y} \tag{21}$$

The SMM approach of estimating the parameter vector $\hat{\boldsymbol{\theta}}_{SMM}$ is to choose $\boldsymbol{\theta}$ to minimize some distance measure of the simulated model moments $\hat{\boldsymbol{m}}(\tilde{\boldsymbol{x}}|\boldsymbol{\theta})$ from the data moments $\boldsymbol{m}(\boldsymbol{x})$.

$$\hat{\boldsymbol{\theta}}_{SMM}: \quad \min_{\boldsymbol{\theta}} \ \boldsymbol{e}(\tilde{\boldsymbol{x}}, \boldsymbol{y}, \boldsymbol{\theta})^T \, \boldsymbol{W} \, \boldsymbol{e}(\tilde{\boldsymbol{x}}, \boldsymbol{y}, \boldsymbol{\theta}) \tag{22}$$

1. Draw S samples of N uniformly distributed random errors $u_{s,i} \sim U(0,1)$ such that the sample of errors for the sth simulation is $\boldsymbol{u}_s = (u_{s,1}, u_{s,2}, ...u_{s,i}, ...u_{s,N})$.

2. Guess values for the parameter vector to be estimated $\boldsymbol{\theta}_v = (\beta_{0,v}, \beta_{1,v}, ...\beta_{K,v}, \sigma_v)$, where v indexes the iteration of the guess.

3. Compute S samples of N normally distributed random shocks $\varepsilon_{i,s}$ that correspond to the uniformly distributed random errors $u_{s,i}$ in step (1) and the current guess of the standard deviation σ_v of the normally distributed errors.

4. Use the simulated values of $\varepsilon_{i,s}$ and the guessed values of the coefficients $(\beta_{0,v}, \beta_{1,v}, ...\beta_{K,v})$ to compute average predicted values $\hat{\boldsymbol{y}}$ using (19) and (20).

5. Compute the value of the criterion function from (22) given the guess for the parameter vector $\boldsymbol{\theta}_v$.

6. Update guess for parameter vector to $\boldsymbol{\theta}_{v+1}$ to minimize criterion function from (22). This should be done with a minimizer function.

3 Linking SMM to Cross Validation in Machine Learning

References

Adda, Jérôme and Russell Cooper, *Dynamic Economics: Quantitative Methods and Applications*, MIT Press, 2003.

Davidson, Russell and James G. MacKinnon, *Econometric Theory and Methods*, Oxford University Press, 2004.

Duffie, Darrell and Kenneth J. Singleton, "Simulated Moment Estimation of Markov Models of Asset Prices," *Econometrica*, July 1993, *61* (4), 929–952.

Laroque, G. and B. Salanie, "Simulation Based Estimation Models with Lagged Latent Variables," *Journal of Applied Econometrics*, December 1993, *8* (Supplement), 119–133.

Lee, Bong-Soo and Beth Fisher Ingram, "Simulation Based Estimation of Time-series Models," *Journal of Econometrics*, February 1991, *47* (2-3), 197–205.

McFadden, Daniel, "A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration," *Econometrica*, September 1989, *57* (5), 995–1026.