ECON 7130 - Microeconomics III

Spring 2016

Notes for Lecture #9

Today:

- GMM

Generalized Method of Moments

- General idea:
  - Recall that the likelihood analysis is based on a full specification of the distributional form of the data, and the DGP is assumed to be known apart from a finite number of parameters to be estimated
  - The main condition for the asymptotic efficiency of the ML estimator is that the likelihood function is correctly specified
  - If there is much uncertainty on the distributional form, it may be preferable to apply an estimation technique that assumes less structure on the DGP
  - GMM is an alternative principle, where the estimator is derived from a set of minimal assumptions, the so-called moment conditions that the model should satisfy.
  - Key difference from ML and other methods:
    * The likelihood analysis begins with a statistical description of the data, and the econometrician should ensure that the likelihood function accounts for the main characteristics of the data
    * Based on the likelihood function we can test hypotheses implied by economic theory (i.e., the conditional probability should say something about the parameters)
    * A GMM estimation, on the other hand, typically begins with an economic theory and the data are used to produce estimates of the model parameters
    * Thus it's very much a structural approach - what are important are the model parameters and there is an emphasis on econ theory (e.g. does demand slope down??)
    * Estimation is done under minimal statistical assumptions, and often less attention is given to the fit of the model (in theory, it must be right if the moments hold)
  - Note the elegance in this - economic theory drives the model estimation
- Moment conditions:
  - A moment condition is a statement involving the data and the parameters of interest.
  - Generally: $g(\theta_0) = E[f(w_t, z_t, \theta_0)] = 0$, where
    * θ is a $K \times 1$ dimensional vector
    * $f(\cdot)$ is an R dimensional vector of potentially non-linear functions
    * w_t is a vector of variables appearing in the model
    * z_t is a vector of instruments
  - Expectation of moment condition is zero when evaluated at the true parameters values, θ_0
  - For a given set of observations, w_t and z_t $(t = 1, 2, ..., T)$, we cannot calculate the expectation, and it is natural to rely on sample averages
  - Define the sample analogue to the moment conditions above as: $g_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} f(w_t, z_t, \theta)$
  - We then define the estimator $\hat{\theta}$ as the solution to $g_T(\theta) = 0$.

    * Need at least as many equations as we have parameters, $R \geq K$ (the "order condition" for identification).
    * If R = K we say that the system is exactly identified
    * The estimator is referred to as the method of moments (MM) estimator.
  - Example: MM estimator of the mean
    * Let μ_0 be the population expectation for y_t
    * Let $f(y_t, \mu_0) = y_t - \mu_0$
    * Then $g(\mu_0) = E[f(y_t, \mu_0)] = E[y_t - \hat{\mu}] = 0$
    * Based on the observed samples, y_t $(T = 1, 2, ..., T)$ we can construct the the sample moment conditions:
    * $g_T(\hat{\mu}) = \frac{1}{T} \sum_{t=1}^{T} (y_t - \hat{\mu}) = 0$
    * The MM mean estimator is the solution to this, or $\hat{\mu} = \frac{1}{T} \sum_{t=1}^{T} y_t$ - the sample average

- Identification of MM estimators:
  - The moment conditions are said to identify the parameters in θ_0 if there is a unique solution, so that $E[f(w_t, z_t, \theta)] = 0$ if and only if $\theta = \theta_0$

- OLS as MM:
  - Consider the linear regression model, $y_t = x_t^{'} \beta_0 + \epsilon_t$, $t = 1, 2, ..., T$,
    * x_t is a $K \times 1$ vector of regressors
    * Assume it represents the conditional expectation: $E[y_t | x_t] = x_t^{'} \beta_0$ so that $E[\epsilon_t | x_t] = 0$
  - This implies K unconditional moments:
  - $g(\beta_0) = E[x_t \epsilon_t] = E[x_t (y_t - x_t^{'} \beta_0)]$ (NOTE - these are OLS assumptions that the X's are uncorrelated with the errors)
  - Corresponding sample moments:
  - $g_T(\hat{\beta}) = \frac{1}{T} \sum_{t=1}^{T} x_t (y_t - x_t' \hat{\beta}) = \frac{1}{T} \sum_{t=1}^{T} x_t y_t - \frac{1}{T} \sum_{t=1}^{T} x_t x_t^{'} \hat{\beta} = 0$
  - The MM estimator can then be derived as the unique sol'n to: $\hat{\beta}_{MM} = \left(\sum_{t=1}^{T} x_t x_t^{'} \right)^{-1} \sum_{t=1}^{T} x_t y_t$
    * Provided that $\sum_{t=1}^{T} x_t x_t^{'}$ is non-singular, so the inverse exists
  - Note: $\hat{\beta}_{MM} = \hat{\beta}_{OLS}$

- IV as MM:
  - Partition the regression model above so that we have: $y_t = x_{1t}^{'} \gamma_0 + x_{2t}^{'} \delta_0 + \epsilon_t$
  - The K_1 variables in x_{1t} are pre-determined
  - The $K_2 = K - K_1$ variables in x_{2t} are endogenous
  - This means:
    * $E[x_{1t} \epsilon_t] = 0$ $(K_1 \times 1)$
    * $E[x_{2t} \epsilon_t] \neq 0$ $(K_2 \times 1)$
  - There is no unique sol'n to the model like this since there are K parameters, but only $K_1 < K$ moment conditions
  - Non consider K_2 new variables, correlated with x_{2t}, but not with the errors: $E[z_{2t} \epsilon_t] = 0$
  - These K_2 new moment conditions can be added to the K_1 above so that the model is now identified
  - Note that we now have two $K \times 1$ vectors: $x_t = \begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix}$ and $z_t = \begin{pmatrix} x_{1t} \\ z_{2t} \end{pmatrix}$
  - z_t is called the vector of instruments

    * As we discussed with IV estimation...
    * x_{1t} are instruments for themselves, since they are predetermined
    * z_{2t} are instruments for x_{2t}
  - Now we have K moment conditions: $g(\beta_0) = E[z_t\epsilon_t] = E[z_t(y_t - x_t^{'}\beta_0)] = 0$
  - The corresponding sample moment conditions are given by: $g_T(\hat{\beta}) = \frac{1}{T}\sum_{t=1}^{T} z_t(y_t - x_t^{'}\hat{\beta}) = 0$
  - The MM estimator is the solution to: $\hat{\beta}_{MM} = \left(\sum_{t=1}^{T} z_t x_t^{'}\right)^{-1} \sum_{t=1}^{T} z_t y_t$
    * Provided that the $K \times K$ matrix $\sum_{t=1}^{T} z_t x_t^{'}$ is non-singular, so the inverse exists
  - This MM estimator coincides with the simple IV estimator

- ML as MM:
  - Can do MM instead of ML
  - Moment conditions are that the first derivative of the likelihood function w.r.t. the parameters must equal zero at θ_0 (this is the likelihood score function)
  - MM gives consistent estimate of θ even if likelihood function is misspecified

- Over-identified models and GMM:
  - If $R > K$, then model is over-identified and, in general, no solution to $g_T(\theta) = 0$ exists
  - In this case, use the Generalized Method of Moments (GMM) estimator, $\hat{\theta}_{GMM}$
  - $\hat{\theta}_{GMM}$ is chosen to minimize the distance between the $g_T(\hat{\theta})$ and 0
  - Usually use a distance corresponding to the sum of squares, $g_T(\theta)^{'}g_T(\theta)$
  - Weighting matrices:
    * A disadvantage of the simple sum of squared errors is that you weight moments depending upon the units used
    * Note that $\hat{\theta}_{GMM}$ depends upon the weighting matrix used:

$$\hat{\theta}_{GMM}(W_T) = \underset{\theta}{\operatorname{argmin}}\{g_T(\theta)'W_T g_T(\theta)\} \tag{1}$$

    * Where W_T is the weighting matrix chosen
    * Need W_T to be positive definite so that put some weight on all moments (don't throw info away!)
    * Note that the weight matrix is redundant when exactly identified (and in this case, estimator does not depend on weight matrix)
  - Why use an over identified model?
    * Practical - may have trouble making exactly identified hit exactly
    * Test model with over identifying restrictions

- Identification of GMM:
  - Need unique solution to $\hat{\theta}_{GMM}(W_T) = argmin_\theta\{g_T(\theta)'W_T g_T(\theta)\}$
  - Also need the Law of Large Numbers to apply to the moment conditions (i.e, $\frac{1}{T}\sum_{t=1}^{T} f(w_t, z_t, \theta) \rightarrow E[f(w_t, z_t, \theta]$ for $T \rightarrow \infty$)
    * If data are IID, these are fulfilled
    * If data are time series, need stationarity

- Std Errors
  - As we noted before, the GMM estimator depends on the weight matrix

- Some weight matrices produce precise estimators while others produce poor estimators with large variances
- We want to choose the optimal weight matrix to produce estimates with smallest possible asymptotic variance
- This is an efficient or optimal GMM estimator
- Intuition: moments with a small variance are very informative on the parameters and should have a large weight
- Moments with a high variance should have a smaller weight
- 2-Step
  * 1st, estimate θ with an identity matrix: $\hat{\theta}_{[1]} = \text{argmin}_{\theta}\, g_T(\theta)' W_{[1]} g_T(\theta)$
  * 2nd, use $\hat{\theta}_{[1]}$ to find the optimal weight matrix, $W_{[2]}^{opt} = \hat{\Omega}^{-1} = \left(g_T(\hat{\theta}_{[1]}) g_T(\hat{\theta}_{[1]})' \right)^{-1}$
  * Ω is the VCV matrix of the moments
  * Now you can use the optimal weight matrix to find the parameter estimates:

$$\hat{\theta}_{[2]} = \underset{\theta}{\text{argmin}}\, g_T(\theta)' W_{[2]}^{opt} g_T(\theta) \tag{2}$$

  * Note that estimator is not unique, but depends upon initial weight matrix
  * This is the *two-step* GMM estimator
  * If continue this until convergence (i.e. until the weight matrix doesn't change between iterations), get the *iterated* GMM estimator
  * This estimator does not depend upon the initial weight matrix
  * BUT, 2 steps is usually enough

- Overidentification tests
  - Hansen J-test for over identifying restrictions (same as Sargan test for linear models - as we talked about with IV)
  - $\xi_J = T * g_T(\hat{\theta}_{GMM})' W_T^{opt} g_T(\hat{\theta}_{GMM}) = T * Q_T(\theta) \sim \chi^2(R - K)$
  - Where, $Q_T = g_T(\theta)' W_T g_T(\theta)$
  - $R - K$ degrees of freedom (because if want, can have K moments equal zero, so they don't contribute to test)
  - The intuition is that if we can set K moments to zero, but if all R are valid, then the remaining $R - K$ should also be close to zero
  - So a lower J-stat means model more likely valid. Large means violated by data.
  - NOTES:
    * J-stat does not test validity of model per se
    * It's not a test of underlying economic theory
    * It tests the whether over identifying restrictions are correct given identification using K moments
    * Often structural models fail this - usually just used to compare one model's fit to others'

How to code estimation:

- Stata
  * `gmm` command
  * Allows you to specify the moment conditions as substitutable expressions
    · You'll enclose model parameters in braces, {}
  * Think about moment conditions taking the form of $E[ze(\beta)] = 0$, where z is a vector of instruments, and β are the model parameters

    * Will write out $e(\beta)$ and then list instruments in the gmm command
    * e.g., OLS by GMM
        · Regression equation is: $y_i = \beta_0 + \beta_1 * x_{1i} + \beta_2 * x_{2i} + \varepsilon_i$
        · Moments are $E\left[\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \varepsilon\right] = 0$
        · To estimate via gmm do:
        · `gmm(y-x1*{b1}-x2*{b2}-{b3}), instruments(x1 x2), [options]`
        · Where the first expression is the expectation of the error (i.e. the expectation that'll be zero)
        · The options available include using a one-step or iterated estimator (default is two-step) and a number of std error corrects (robust, cluster, etc)
        · After run `gmm`, can do `estat overid` to get the Hansen J-test statistic
- Matlab
    * Did you all do this in Adam's class?
    * Write down the moments function to get Q_T
    * Use `fminsearch` or other optimization routine (e.g. simulated annealing) to minimize the distance
    * Find optimal weight matrix by iterating on the above steps - use step 1 results to calculate W2, the W2 to get θ_2
    * With these code, easy to make iteration until convergence

GMM: Example 1, Hansen and Singleton, "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models" (*Econometrica*, 1982):

- Mostly a methods paper about GMM with dynamic models
- Question: What are the coefficient of relative risk aversion and the time preference parameter for the representative consumer?
- Problem: Need a structural model to estimate this, most methods for estimating a structural parameters from dynamic models need to find a numerical solution to the model many times during estimation. This is computationally intensive.
- Solution: GMM using the Euler equations as moments
- Basic Model:
    - Consumption Capital Asset Pricing Model (Consumption CAPM)
        * Model will give equilibrium prices in terms of consumption
        * e.g., return on assets r_t
    - consumer solves: $\max_{\{c_t\}_{t=0}^{\infty}} E_t \sum_{t=0}^{\infty} \beta^t u(c_t)$ s.t. $c_t + s_t \leq w_t + R_t s_{t-1}$
    - Euler equation is: $\beta E_t[u'(c_{t+1})R_{t+1}|\mathcal{I}_t] = u'(c_t)$
        * Where β is the time discount factor
        * R is the gross, real interest rate
        * c is consumption
    - If we have Constant Relative Risk Aversion (CRRA) utility function, then $u(c_t) = \frac{c_t^{1-\gamma}}{1-\gamma}$
    - $\gamma < 1$ is the Coeff of Relative Risk Aversion
    - Euler is now: $\beta E_t\left(\left[\frac{c_{t+1}}{c_t}\right]^{-\gamma} R_{t+1}|\mathcal{I}_t\right) - 1 = 0$

- Can then write this as: $\beta \left[\frac{c_{t+1}}{c_t}\right]^{-\gamma} R_{t+1} - 1 = \varepsilon_{t+1}$
- Where ε_{t+1} is expectational error - uncorrelated with any variable in the time t information set, $\mathcal{I}_t$
- The economic interpretation is that under rational expectations a variable in the information set must be uncorrelated to the expectation error.
- These give the orthogonality conditions of the moments - i.e. the model gives you the instrument set you need (anything in the info set)
- Note that this model is nonlinear. This is not a problem for GMM

• Identification:

- We have two parameters to estimate, β and γ.
    * Note that these are the deep parameters of the model- they are policy invariant
- Thus we'll need $R \geq 2$ instruments to identify the model
- Any relevant variable in the information set, $\mathcal{I}_t$ may be an instrument
- e.g., a constant, c_t, R_t, $\frac{c_t}{c_{t-1}}$, or lags of these
- Thus, can construct a vector of moment conditions:

$$\begin{aligned} E\left[\left(\beta\left(\frac{c_{t+1}}{c_t}\right)^{-\gamma} R_{t+1} - 1\right)\right] &= 0 \\ E\left[\left(\beta\left(\frac{c_{t+1}}{c_t}\right)^{-\gamma} R_{t+1} - 1\right)\left(\frac{c_t}{c_{t-1}}\right)\right] &= 0 \\ E\left[\left(\beta\left(\frac{c_{t+1}}{c_t}\right)^{-\gamma} R_{t+1} - 1\right) R_t\right] &= 0 \end{aligned} \quad (3)$$

- These moments will need to hold for $t = 1, 2, ...T$.
- Thus can construct the sample analogue to the expectation, which his the sample mean of each moment
    * e.g. $\frac{1}{T}\sum_{t=1}^{T} \beta \left(\frac{c_{t+1}}{c_t}\right)^{-\gamma} R_{t+1} = 1$ (this is the first moment condition above)
- Of course, also need economic model to be correct for identification

• Data:

- Stock return data
- Aggregate consumption data

• Results:

- Note how you don't have to solve the dynamic model - the FOC (euler equations) are all you need to estimate the parameters
- Estimates of $\beta = 0.99$ or so, $\gamma = 0.68 - 0.97$

BLP GMM....