Lecture 1 Toolbox/review: GMM, ML, SML, SMM, indirect inference

Ciprian Domnisoru

Aalto University

Generalized Method of Moments

- Introduced by Hansen, L. (1982), Large Sample Properties of Generalized Method of Moments Estimators, Econometrica 50, 1029-1054.

  In this paper we study the large sample properties of a class of generalized method of moments (GMM) estimators which subsumes many standard econometric estimators. To motivate this class, consider an econometric model whose parameter vector we wish to estimate. The model implies a family of ***orthogonality conditions that embed any economic theoretical restrictions*** *that we wish to impose or test.*

Portfolio Choices (Hansen and Singleton,1982)

The consumer optimization problem

- Suppose there are J financial securities.
- Let p_{tj} denote the price of the j^{th} security in period t consumption units, and $q_{t-1,j}$ the amount a consumer owns at the beginning of the period.
- Let r_{tj} denote the real return on assets purchased in period $t-1$.
- The investor's budget constraint is:

$$c_t + \sum_{j=1}^{J} p_{tj} q_{tj} \leq \sum_{j=1}^{J} r_{tj} p_{t-1,j} q_{t-1,j}$$

- At t the consumer maximizes a concave objective function with linear constraints, choosing $(q_{s1}, ..., q_{sJ})$ to maximize:

$$u(c_t) + E_t \left[\sum_{s=t+1}^{T} \beta^{s-t} u(c_s) \right]$$

subject to the sequence of all the future budget constraints.

Portfolio Choices

First order conditions

- Nonsatiation guarantees:

$$c_t = \sum_{j=1}^{J} (r_{tj} p_{t-1,j} q_{t-1,j} - p_{tj} q_{tj})$$

- The interior first order condition (FOC) for each $k \in \{1, \ldots, J\}$ requires:

$$\begin{aligned} & p_{tk} u' \left(\sum_{j=1}^{J} (r_{tj} p_{t-1,j} q_{t-1,j} - p_{tj} q_{tj}) \right) \\ \geq \quad & E_t \left[p_{tk} r_{t+1,k} \beta u' \left(\sum_{j=1}^{J} (r_{t+1,j} p_{tj} q_{tj} - p_{t+1,j} q_{t+1,j}) \right) \right] \end{aligned}$$

with equality holding if $q_{tj} > 0$.

Portfolio Choices

Estimation and testing

- For any $r \times 1$ vector x_t belonging to the information set at t and all k:

$$0 = E_t \left[r_{t+1,k} \beta \frac{u'(c_{t+1})}{u'(c_t)} - 1 \right] = E \left[r_{t+1,k} \beta \frac{u'(c_{t+1})}{u'(c_t)} - 1 | x_t \right]$$

and hence:

$$0 = E \left\{ x_t \left[r_{t+1,k} \beta \frac{u'(c_{t+1})}{u'(c_t)} - 1 \right] \right\}$$

- Given a sample of length T we can estimate the $1 \times l$ vector (β, α) for a parametrically defined utility function $u(c_t; \alpha)$ by solving:

$$0 = A_T \sum_{t=1}^{T} x_t \left[r_{t+1,k} \beta \frac{u'(c_{t+1}; \alpha)}{u'(c_t; \alpha)} - 1 \right]$$

where A_T is an $l \times r$ weighting matrix.

Summary

- Basic model of consumption and portfolio choice
- From the Euler equations, we derive a nonlinear moment condition model used to estimate the preference parameters of the agent: no need to explicitly solve for the stochastic equilibrium.
- GMM: assumptions about population moment conditions that come from economic models -> solve for the parameters that make those assumptions hold in our data
- Moment conditions: first-order conditions from economic models; econometric model: instruments must be uncorrelated with errors;
- Specification test for the validity of the moment conditions $E[x_i \ \varepsilon_i]=0$ (J-test of overidentifying[1] restrictions). Derive a statistic with a known distribution when this assumption holds - deviations from this distribution seen as evidence of our conditions not holding. For the optimal matrix, χ^2 with K-L d.f.
- GMM is limited-information estimation method, does not need a specification of likelihood, ML is a special case!

[1] L of the K sample moment conditions can be set exactly equal to zero: we are measuring how close the remaining K-L overidentifying restrictions are.

TABLE III

INSTRUMENTAL VARIABLES ESTIMATION WITH MULTIPLE RETURNS

Equally- and Value-Weighted Aggregate Returns 1959:2–1978:12

Cons	*NLAG*	$\hat{\alpha}$	$\widehat{SE}(\hat{\alpha})$	$\hat{\beta}$	$\widehat{SE}(\hat{\beta})$	χ^2	DF	Prob.
NDS	1	− 0.5901	1.7331	.9989	.0041	18.309	6	.9945
NDS	2	1.0945	1.4907	.9961	.0040	24.412	12	.9821
NDS	4	0.3835	1.4208	.9975	.0039	40.234	24	.9798
ND	1	− 0.6494	0.6838	.9982	.0025	19.976	6	.9972
ND	2	− 0.0200	0.6071	.9982	.0025	27.089	12	.9925
ND	4	− 0.1793	0.5928	.9986	.0025	42.005	24	.9871

Model of consumption and portfolio choice

The optimum of the agent is characterized by the first order conditions, which can be expressed as:

$$1 = \mathbf{E}_t \left[r_{t+1,k} \beta \frac{u'(c_{t+1})}{u'(c_t)} \right]$$

where:

- $\mathbf{E}_t$ is the expectation conditional on the information available at time t.
- $r_{t+1,k}$ is the gross real return on asset k
- $\beta \frac{u'(c_{t+1})}{u'(c_t)}$ is the marginal rate of substitution (MRS)

If you think of $r_{t+1,k}$ as the inverse of relative prices, this model is simply equating MRS with relative prices in expectation.

Determine moment equations from the model

Subtract 1 from each side and bring it inside the expectation, so we have an expression equal to zero:

$$\mathbf{E}_t \left[r_{t+1,k} \beta \frac{u'(c_{t+1})}{u'(c_t)} - 1 \right] = 0$$

This is a population moment equation.

Under the assumption this model is true, we want to use data (r_{t+1}, c_t, c_{t+1}) and estimate preference parameters, i.e., (α, β) or (γ, β).

Parameterizing the utility function

We will parameterize the utility function, first using the same as in HS.

$$u(c_t) = (1+\alpha)^{-1} c_t^{1+\alpha}$$

Notice this utility function exhibits constant relative risk aversion (CRRA) if $\alpha < 0$. [2]

The population moment becomes a specific nonlinear function of data and parameters:

$$\mathbf{E}_t \left[r_{t+1,k} \beta \frac{u'(c_{t+1})}{u'(c_t)} - 1 \right] = 0 \Rightarrow \mathbf{E}_t \left[r_{t+1,k} \beta \left(\frac{c_{t+1}}{c_t} \right)^{\alpha} - 1 \right] = 0$$

[2]This utility function is concave with respect to wealth/consumption for $c \in \mathbb{R}_+$ if $\alpha < 0$. The coefficient of relative risk aversion, by definition, is $-\frac{cu''(c)}{u'(c)}$. You can show that for this utility function, the coefficient is constant (it equals $1+\alpha$).

Dataset

Following HS, we use aggregate real consumption data per capita from FRED.[3] We need:

- aggregate consumption data, nondurables & services
- consumption prices indices (CPIs) for both nondurables and services
- US population

Date-Range	Frequency	FRED Code	Series Description
1/1959-9/2020	Monthly	PCEND	Nondurables (Billions USD)
1/1959-9/2020	Monthly	DNDGRG3M086SBEA	Nondurables CPI (Index: 2012 = 100)
1/1959-9/2020	Monthly	PCES	Services (Billions USD)
1/1959-9/2020	Monthly	DSERRG3M086SBEA	Services CPI (Index: 2012 = 100)
1/1959-9/2020	Monthly	POPTHM	US Population (Thousands)

Then we calculate (for example) real nondurables per capita via:

$$\text{real nondurables per capita}_t = \frac{(\text{nondurables}_t * 1e9)/(\text{nondurables CPI}_t/100)}{(\text{population}_t * 1e3)}$$

[3] St. Louis Federal Reserve Bank Data, available here: https://fred.stlouisfred.org/

Dataset

HS uses NYSE data, but we use S&P500 indices from WRDS/CRSP.[4] We need:

- monthly equally-weighted returns including dividends
- monthly value-weighted returns including dividends
- monthly risk-free rate (we used 1yr T-Bill rate from FRED)

FRED/WRDS	Date Range	Frequency	FRED Code	Series Description
WRDS	1/1946-12/2019	Monthly	-	Equally-Weighted Returns (including dividends)
WRDS	1/1946-12/2019	Monthly	-	Value-Weighted Returns (including dividends)
FRED	4/1953-10/2020	Monthly	GS1	1yr Treasury Constants Maturity Rate (%)

Then we can calculate (for example) real value-weighted gross returns via:

$$\text{real value-weighted return}_t = \frac{(\text{value-weighted return}_t) + 1}{\frac{CPI_t}{CPI_{t-1}}}$$

[4] Only available with a license (e.g. through your university library).

Determine orthogonality conditions

Notice that we can label the expression inside the expectation as error.

$$\mathbf{E}_t \left[\underbrace{r_{t+1,k}\beta \left(\frac{c_{t+1}}{c_t} \right)^{\alpha} - 1}_{\varepsilon_t} \right] = 0$$

Let $x_t \in I_t$, with I_t information known by the agent and x_t observable by econometrician at time t. By definition, $\mathbf{E}_t[\varepsilon_t] := \mathbf{E}[\varepsilon_t | I_t]$. So $\mathbf{E}_t[\varepsilon_t] = 0$ implies:

$$\begin{aligned} \mathbf{E}[x_t \varepsilon_t] &= \mathbf{E}[\mathbf{E}[x_t \varepsilon_t | x_t, I_t]] \\ &= \mathbf{E}[x_t \mathbf{E}[\varepsilon_t | I_t]] \\ &= 0 \end{aligned}$$

We therefore have the orthogonality conditions (plural for vector $\mathbf{x}_t$):

$$\mathbf{E}\left[\mathbf{x}_t \cdot \left(r_{t+1,k}\beta \left(\frac{c_{t+1}}{c_t} \right)^{\alpha} - 1 \right) \right] = 0$$

Unconditional expectations based on conditional expectations

- Very common in GMM to use the law of iterated expectations to derive unconditional expectations based on conditional expectations.
- Take the OLS model, $y = x'\beta + u$, we assume $E(u \mid x) = 0$.
- Implies E(xu)=0, because, by law of iterated expectations,

$$E(xu) = E_x[E(xu|x)] = E_x[xE(u \mid x)] = 0$$

- $E(xu) = E[x(y - x'\beta)] = 0$

Orthogonality conditions

The first line estimated of Table III in HS uses equally- and value-weighted returns, with one lag each of the consumption ratio and equally- and value-weighted returns. This is a set of 8 population moment conditions in 2 parameters (α, β):

$$\mathbf{E}\left[\begin{pmatrix} 1 \\ \frac{c_t}{c_{t-1}} \\ r_{t,ew} \\ t_{t,vw} \end{pmatrix} \cdot \left(r_{t+1,ew}\beta \left(\frac{c_{t+1}}{c_t}\right)^{\alpha} - 1 \right)\right] = 0$$

$$\mathbf{E}\left[\begin{pmatrix} 1 \\ \frac{c_t}{c_{t-1}} \\ r_{t,ew} \\ t_{t,vw} \end{pmatrix} \cdot \left(r_{t+1,vw}\beta \left(\frac{c_{t+1}}{c_t}\right)^{\alpha} - 1 \right)\right] = 0$$

Because in each specification the number of orthogonality conditions exceeds the number of parameters to be estimated, the system is **overidentified**. This has two implications:

- How to weight each equation? (Answer: invertible weighting matrix $\mathbf{S}$!)
- No parameters exactly satisfy all equations.

This is why we use a GMM estimator, which selects parameters that minimize the sample moment error:

$$\theta_{GMM}^{(N)} = \mathbf{S}_N^* \left[\frac{1}{N} \sum_{n=1}^{N} \left(\mathbf{x}_t^{(n)} \cdot \left(r_{t+1}^{(n)} \beta \left(\frac{c_{t+1}^{(n)}}{c_t^{(n)}} \right)^{\alpha} - 1 \right) \right) \right]$$

The estimator is efficient if $\mathbf{S}_N^*$ is a consistent estimator of the covariance matrix of the population orthogonality conditions. It can be obtained through the two-step efficient GMM estimation procedure.

Optimal weighting matrix

- Let $S = Cov(zu)$, covariance of the moment conditions. Using a weight matrix which is the inverse of the moment covariance matrix. $W = S^{-1}$ produces the most efficient asymptotically normal estimator.
- Difficulty: you need W for estimation of u, but you don't have it.
- Cov (zu)=$E(u^2zz')$=σ^2 E (zz') for i.i.d. errors
- β_{GMM} minimizes the objective $Q(\beta) = \{\frac{1}{N}\sum_i z_i u_i(\beta)\}' W \{\frac{1}{N}\sum_i z_i u_i(\beta)\}$
- Because σ^2 positive scalar, ignore when minimizing objective
- $\widehat{W_1} = (\frac{1}{N}\sum_i z_i z_i')^{-1}$
- Given $\widehat{W_1}$, solve for $\hat{\beta}_1$
- Use residuals at the value β_1, estimate $S = \frac{1}{N}\sum_i u_i^2 z_i z_i'$
- re-solve minimization , using $\widehat{W_2} = \hat{S}^{-1}$

```
. gmm (real_vw*{beta=0.9}*(ratio_nds^{alpha})-1) ///
>         (rate_rf_f1*{beta=0.9}*(ratio_nds^{alpha})-1) ///
>         if time_hs==1, ///
>         two winitial(unadjusted, independent) ///
>         instruments(ratio_nds_l1 real_vw_l1 ratio_rf ratio_rf_l1)

Step 1
Iteration 0:   GMM criterion Q(b) =  .01227509
Iteration 1:   GMM criterion Q(b) =  .00143988
Iteration 2:   GMM criterion Q(b) =  .00143983

Step 2
Iteration 0:   GMM criterion Q(b) =  .68669564
Iteration 1:   GMM criterion Q(b) =  .57333493
Iteration 2:   GMM criterion Q(b) =   .5733348

GMM estimation

Number of parameters =   2
Number of moments    =  10
Initial weight matrix: Unadjusted                 Number of obs   =        239
GMM weight matrix:     Robust

------------------------------------------------------------------------------
             |               Robust
             | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
       /beta |   .9671057   .0026614   363.38   0.000     .9618894     .972322
      /alpha |  -.2804721   1.190222    -0.24   0.814    -2.613264    2.052319
------------------------------------------------------------------------------
Instruments for equation 1: ratio_nds_l1 real_vw_l1 ratio_rf ratio_rf_l1 _cons
Instruments for equation 2: ratio_nds_l1 real_vw_l1 ratio_rf ratio_rf_l1 _cons
```

Tests of overidentifying restrictions

Take the Q from objective (slide 13). Then, under the null hypothesis that your moment conditions hold, and using the optimal weighting matrix, $J = N \times Q \sim \chi^2$ with K-L d.f.

```
.
. estat overid

  Test of overidentifying restriction:

  Hansen's J chi2(8) = 137.027 (p = 0.0000)
. estadd scalar j_stat=r(J), replace

added scalar:
             e(j_stat) =  137.02702
. estadd scalar j_df=r(J_df), replace

added scalar:
               e(j_df) =  8
. estadd scalar j_prob=1-r(J_p), replace

added scalar:
             e(j_prob) =  1
.
```

Testing equality of parameters before and after Hansen sample

```
gmm ((real_ew*{beta1=0.9}*(ratio_nds^{alpha1})-1)*time_hs + ///
                (real_ew*{beta2=0.9}*(ratio_nds^{alpha2})-1)*time_after) ///
        ((real_vw*{beta1=0.9}*(ratio_nds^{alpha1})-1)*time_hs + ///
                (real_vw*{beta2=0.9}*(ratio_nds^{alpha2})-1)*time_after) ///
        two winitial(unadjusted, independent) ///
        instruments(ratio_nds_l1 real_ew_l1 real_vw_l1)
```

Other hints for Assignment 1

- Unit root indicates time dependence
- If we can't reject the null, we would worry about the sensitivity of the estimates to the time period selected. A stationary time series' properties don't depend on the time at which you observe it.
- dfuller ratio nds , lags(6) trend
- Alternate utility function parameterization: Think about function's properties and economic implications. Calculate marginal utility; adjust orthogonality conditions; run estimation; discuss.