14.385 Recitation 1

Paul Schrimpf

September 15, 2008

Types of Estimators

- Moment methods
  - GMM
  - IV
- Extremum methods
  - MLE
  - M-estimators
  - Quantile regression
  - **Minimum Distance**

Figure 1: Relationship among estimators

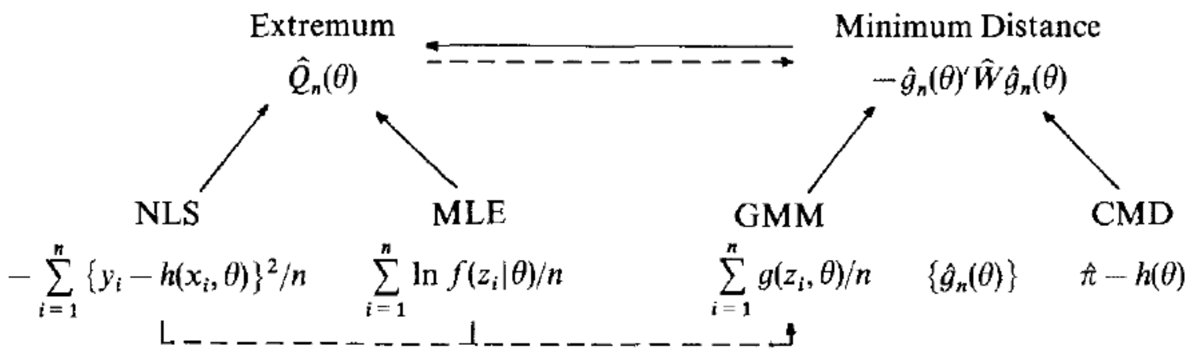

Figure 1.

[a]

[a]Source: Newey and McFadden (1994)

1 Minimum Distance

$$\hat{\theta} = \arg\min \hat{f}_n(\theta)'\hat{W}\hat{f}_n(\theta) \tag{1}$$

where $\text{plim}\, \hat{f}_n(\theta_0) = 0$.

Includes:

- GMM with $\hat{f}_n(\theta) = \frac{1}{n}\sum g(z_i, \theta)$ and $\hat{W} = \hat{A}$
- MLE with $\hat{f}_n(\theta) = \frac{1}{n}\sum \frac{\partial \ln f(z_i|\theta)}{\partial\theta}$ and $\hat{W} = I$
- Classical Minimum Distance (CMD): $\hat{f}_n(\theta) = \hat{\pi} - h(\theta)$ where $\hat{\pi} \xrightarrow{p} \pi_0 = h(\theta_0)$. Usually, π are reduced form parameters, θ are structural parameters, and $h(\theta)$ is a mapping from the structural parameters to the reduced form.
  - *Example:* Chamberlain (1982, HoE 1984) approach to panel data. Model:

    $$y_{it} = x_{it}\beta + c_i + e_{it}\ ,\ E[e_{it}|x_i, c_i] = 0$$

    Reduced form: regress y_{it} on all $x_{i\cdot}$ to get π_t.
    $h(\theta)$: we know that $x_{i\cdot}\pi_t$ is the best linear predictor of y_{it} given $x_{i\cdot}$. We also know that

    $$\begin{aligned} BLP(y_{it}|x_{i\cdot}) =& BLP(x_{it}\beta + e_{it}|x_{i\cdot}) + BLP(c_i|x_{i\cdot}) \\ =& x_{it}\beta + x_{i\cdot}\lambda \end{aligned}$$

    So if we stack the π_t into a $t \times tk$ matrix π, we know that

    $$\pi = h(\beta, \lambda) = I_T \otimes \beta' + \imath_T\lambda'$$

    where β is $k \times 1$ and λ is $tk \times 1$.
- Indirect Inference: is mathematically the same as CMD, $\hat{f}_n(\theta) = \hat{\pi} - h(\theta)$ where $\hat{\pi} \xrightarrow{p} \pi_0 = h(\theta_0)$, but the justification is slightly different. We have an economic model, which we are not entirely certain is the true DGP (or perhaps is just difficult to compute the likelihood for), but we do believe can capture some important features of the data. These features of the data are summarized by the parameters of an easy to estimate auxillary model, π. $h(\theta)$ gives the estimates of the auxillary model that we would expect if our economic model were the true DGP and had parameters θ. $h(\theta)$ is often calculated through simulation.
  - *Example: DSGE* (taken from 14.384 notes, for a real application see e.g. Del Negro,Schorfheide, Smets and Wouters (2007)) Consider a simple RBC:

    $$\begin{aligned} &\max E_0 \sum \omega^t \frac{c^{-\gamma} - 1}{\gamma} \\ &\text{s.t. } c_t + i_t = A\lambda_t k_t^\alpha \\ &\quad k_{t+1} = (1-\delta)k_t + i_t \\ &\quad \lambda_t = \rho\lambda_{t-1} + \epsilon_t \ \ \epsilon_t \sim N(0, \sigma^2) \end{aligned}$$

    This model has many parameters ($\theta = (\omega, \gamma, A, \alpha, \delta, \rho, \sigma^2)$) and it would be difficult to write down a likelihood or moment functions. Moreover, we don't really believe that this model is the true DGP and we don't want to use it to explain all aspects of the data. Instead we just want the model to explain some feature of the data, say the dynamics as captured by VAR coefficients. Also, although it is hard to write the likelihood function for this model, it is fairly easy to simulate the model. The we can use indirect inference as follows:

    1. Estimate (possibly misspecified) VAR from data. A VAR is simply OLS on:

       $$Y_t = \pi_1 Y_{t-1} + ... + \pi_p Y_{t-p} + u_t$$

       where Y_t is the vector of observed variables at time t. In this example, Y_t might be c_t and i_t.
    2. Given β, simulate model, estimate VAR from simulations, repeat until minimize objective function

1.1 Consistency

Recall the general theorem on consistency from lecture 2:

Theorem 1. *If (i) $Q(\theta)$ is uniquely minimized at the true parameter value θ_0, (ii) Θ is compact, (iii) $Q(\cdot)$ is continuous, and (iv) $\sup_{\theta\in\Theta}|\hat{Q}(\theta)-Q(\theta)| \xrightarrow{p} 0$, then $\hat{\theta} \xrightarrow{p} \theta_0$.*

We will now discuss applying this theorem to minimum distance. To do that, we need to verify each of the conditions:

1. (Identification): suppose $\hat{f}_n(\theta) \xrightarrow{p} f(\theta)$ and $\hat{W} \xrightarrow{p} W$, so that $Q(\theta) = f(\theta)'Wf(\theta)$. As with GMM, showing that this function has a unique minimum is difficult. A local identification condition is that $rank\frac{\partial f}{\partial \theta} = p$, where θ is $p \times 1$. Global identification is typically just assumed.

2. (Compactness): assume it.

3. (Continuity): depends on the particular application. For CMD and indirect inference, $f(\theta) = \pi - h(\theta)$ is continuous as long as $h(\theta)$ is continuous. Since $h(\theta)$ does not depend on the data at all, this condition is easily checked. In the panel data example, $h(\theta)$ is obviously continuous.

4. (Uniform Convergence): depends on the particular application. Recall lemma 3 from lecture 2. It was:

   Lemma 2. *Suppose $\hat{Q}(\theta) \xrightarrow{p} Q(\theta)$ for each $\theta \in \Theta$. Then uniform convergence holds if for some $h > 0$, we have uniformly for $\theta, \theta' \in \Theta$*

   $$|\hat{Q}(\theta) - \hat{Q}(\theta')| \leq B_T||\theta - \theta'||^h \ , \ B_T = O_p(1)$$

   For CMD and indirect inference,

   $$\begin{aligned}\hat{Q}(\theta) - \hat{Q}(\theta') =&|(\hat{\pi} - h(\theta))'\hat{W}(\hat{\pi} - h(\theta)) - (\hat{\pi} - h(\theta'))'\hat{W}(\hat{\pi} - h(\theta'))'| \\ =&|2(h(\theta) - h(\theta')\hat{W}\hat{\pi} + h(\theta)'\hat{W}h(\theta) - h(\theta')'\hat{W}h(\theta')| \\ \leq&|2(h(\theta) - h(\theta')\hat{W}\hat{\pi}| + |h(\theta)'\hat{W}h(\theta) - h(\theta')'\hat{W}h(\theta')| \\ \leq&|2(h(\theta) - h(\theta')\hat{W}\hat{\pi}| + |(h(\theta) - h(\theta'))'\hat{W}(h(\theta) - h(\theta'))|\end{aligned}$$

   so a sufficient condition is that $h(\theta)$ is Hölder continuous on Θ, i.e.

   $$|h(\theta) - h(\theta')| \leq K||\theta - \theta'||^h$$

   for some $h > 0$ and all $\theta, \theta' \in \Theta$. A sufficient condition for Hölder continuity is that $h()$ is differentiable with a bounded derivative because then

   $$|h(\theta) - h(\theta')| \leq \sup_{\Theta} ||h'||||\theta - \theta'||$$

   Clearly, this condition holds for the panel data example. It could also be checked in other applications.

   - If $h(\theta)$ is computed through simulation, then some additional steps need to be taken to show consistency. Let $h_S(\theta)$ denote the value of $h(\theta)$ computed from S simulations. Typically, $h_S(\theta)$ will be some standard estimator and we will know that $h_S(\theta) \xrightarrow{p} h(\theta)$ as $S \to \infty$. For $\hat{Q}$ to converge uniformly, we need to promise that $S \to \infty$ as $T \to \infty$, and we will need the convergence of h_S to h to be uniform in addition to the conditions above.

2 Checking Consistency

When verifying consistency, always start by checking whether $E\nabla Q(\theta_0) = 0$

Example: Probit[1]

$$Q_n(\beta) = \frac{1}{n}\sum y_i \log(\Phi(x_i\beta)) + (1-y_i)\log(1-\Phi(x_i\beta))$$

The derivative is:[2]

$$\begin{aligned}\nabla Q_n(\beta) =& \frac{1}{n}\sum\left(\frac{y_i}{\Phi(x_i\beta)} - \frac{1-y_i}{1-\Phi(x_i\beta)}\right)\phi(x_i\beta)x_i \\ =& \frac{1}{n}\sum\left(\frac{y_i-\Phi(x_i\beta)}{\Phi(x_i\beta)(1-\Phi(x_i\beta))}\right)\phi(x_i\beta)x_i\end{aligned}$$

which has plim

$$\text{plim}\,\nabla Q_n(\beta) = E\left[\left(\frac{E[y_i|x_i]-\Phi(x_i\beta)}{\Phi(x_i\beta)(1-\Phi(x_i\beta))}\right)\phi(x_i\beta)x_i\right]$$

which is 0 at β_0.

[1] Taken from Konrad's notes from last year.

[2] We could just as well calculate $Q = \text{plim}\, Q_n$ and then differentiate.