Macro III. 8/26

Generalized Method of Moments.

1. The idea of GMM is simple: fit your model by choosing parameters to minimize the distance between model moments (function of parameters) and data moments.

   - "Model" can be a linear relation (without economic theory), a probability function (say, $N(\mu, \sigma^2)$), a likelihood function, first order condition in a model (Euler equations, which were the first major application), a full-blown DSGE model.
   - For a given model, the researcher chooses the moments. (Sometimes badly.)
   - Historically, some people associated GMM with Minnesota economics, Rational Expectations, and similar. Useful in such contexts, but also in many more.

2. Formally a model takes the form $m_t = m(x_t, \theta)$ such that $E_{\theta_0}(m_t) = 0$ and (identification) the expectation is not zero for other values of θ (θ_0

is the "true value.") We estimate θ by choosing the value that minimize

$$M_T(\theta) = \Sigma_{t=1}^T m_t(x_t, \theta) \,.$$

M_T is often a non-linear function of θ. OLS and IV are examples of GMM, but people do not think of it like that. Usually, we divide M_T by T but the solution is the same. (If you want to prove consistency, you need to divide by T to use the LLN.)

3. Let us do a bunch of examples.

   - Model: $x \sim N(\mu, 1)$. $M_T = \Sigma_{t=1}^T (x_t - \mu)$.
   - Model: $x \sim N(\mu, \sigma^2)$. $M_T = \Sigma_{t=1}^T (x_t - \mu, x_t^2 - \sigma^2 - \mu^2)$.
   - Model (any i.i.d likelihood function): $M_T = \Sigma_{t=1}^T d\, l(x_t, \theta)$ (call the score, where l is the log of the likelihood function (the density)). ML is (at least asymptotically) typicall the most efficient estimator (satisfies the Cramer-Rao lower bound asymptotically under some conditions).
   - Model: OLS. $M_T = \Sigma_{t=1}^T (x_t^1 (y_t - \beta_1 x_t^1 - \beta_2 x_t^2, x_t^2 (y_t - \beta_1 x_t^1 - \beta_2 x_t^2)'$. Verify that this is the OLS FOC $X'(Y - X\beta) = 0$.

- Model: IV. $M_T = \Sigma_{t=1}^T (z_t^1 (y_t - \beta_1 x_t^1 - \beta_2 x_t^2, z_t^2 (y_t - \beta_1 x_t^1 - \beta_2 x_t^2))'$. Usually, GMM theory is written with instrumentsm but the instrument can be just unity (as in the ML case) or x_t itself as in OLS. Remember, OLS is clearly the most efficient IV estimator if it is valid.

- Model: Nonlinear least squares. Say, minimize the sum of squares of $y_t - \alpha x_t^\rho$. The moments are the FOCs, $M_T = \Sigma_{t=1}^T (x_t^\rho (y_t - \alpha x_t^\rho, \rho x_t^{\rho-1} (y_t - \alpha x_t^\rho)'$.

- Model: Nonlinear least squares with endogenous regressors. Most people would do $M_T = \Sigma_{t=1}^T (z_t^1 (y_t - \alpha x_t^\rho, z_t^2 (y_t - \alpha x_t^\rho)'$. and call it GMM, not non-linear regression.

- Examples so far are exactly identified, very commonly more instruments than variables, e.g. $M_T = \Sigma_{t=1}^T (z_t^1 (y_t - \alpha x_t^\rho, ..., z_t^K (y_t - \alpha x_t^\rho)'$. in which case people chose a weighting matrix W (more on that later). Then you minimize $M_T' W M_T$. Logic as for GLS.