Nonlinear Least Squares

Applications to MIDAS and Probit Models

Brandon Lee

15.450 Recitation 11

Nonlinear Least Sqares

- Consider the model

$$y_t = h(x_t, \theta) + \varepsilon_t$$

  Here we assume that we know the functional form of $h(x_t, \theta)$ and we need to estimate the unknown parameter θ. The linear regression specification is a special case where $h(x_t, \theta) = x_t' \cdot \theta$.

- The nonlinear least squares (NLS) estimator minimizes the squared residuals (exactly the same as in the OLS):

$$\hat{\theta}_{NLS} = \arg\min_{\theta} \sum_{t=1}^{T} (y_t - h(x_t, \theta))^2$$

  The first order condition is

$$0 = \sum_{t=1}^{T} \frac{\partial h\left(x_t, \hat{\theta}_{NLS}\right)}{\partial \theta} \left(y_t - h\left(x_t, \hat{\theta}_{NLS}\right)\right)$$

NLS through QMLE

- Pretend that the errors are normally distributed $\varepsilon_t \sim N\left(0, \sigma^2\right)$. Then the log-likelihood function is

$$\mathscr{L}(\theta) = \sum_{t=1}^{T} \left[-\ln\left(\sqrt{2\pi\sigma^2}\right) - \frac{\left(y_t - h(x_t, \theta)\right)^2}{2\sigma^2} \right]$$

- We can obtain a QMLE estimate of θ from the first order condition when maximizing the log-likelihood function:

$$\begin{aligned} 0 &= \frac{\partial \mathscr{L}}{\partial \theta} \\ &= \frac{1}{\sigma^2} \sum_{t=1}^{T} \frac{\partial h(x_t, \theta)}{\partial \theta} \cdot \left(y_t - h(x_t, \theta)\right) \end{aligned}$$

This is exactly the same condition as before.

NLS Viewed as GMM

- Note that our optimality condition

$$0 = \sum_{t=1}^{T} \frac{\partial h(x_t, \theta)}{\partial \theta} \cdot (y_t - h(x_t, \theta))$$

  can be rewritten as

$$0 = \hat{E}\left[\frac{\partial h(x_t, \theta)}{\partial \theta} \cdot (y_t - h(x_t, \theta))\right]$$

- This is a GMM moment condition and thus our NLS estimator can be interpreted as a GMM estimator. This is a simple, yet extremely powerful observation because now we can use the expressions for GMM standard errors (with Newey-West procedure that is robust to both heteroskedastic and correlated errors).

Continued

- This moment condition is not surprising because the identification assumption $E[\varepsilon_t | x_t] = 0$ leads to the moment condition

$$E[g(x_t)\varepsilon_t] = 0$$

or

$$E[g(x_t)(y_t - h(x_t, \theta))] = 0$$

for any function $g(\cdot)$. We are simply picking

$$g(x_t) = \frac{\partial h(x_t, \theta)}{\partial \theta}$$

This choice of $g(\cdot)$ is motivated by QMLE.

MIDAS

- Suppose we are interested in the behavior of a monthly volatility measure (say, sum of daily squared returns over a month).
- One possibility is to specify a GARCH structure to monthly data series, but this approach throws away lots of valuable data. Or we could use GARCH on daily data series, and try to infer the behavior of monthly volatility measure from the estimated daily volatility dynamics. A potential problem here, though, is that small specification errors that may be acceptable in the daily series may add up to large errors in the monthly volatility measure.
- The idea of Mixed Data Sampling (MIDAS) is to allow for a very flexible and direct relationship between current month's daily data and next month's monthly volatility measure.

- Specification:

$$V_{t+H,t}^{H} = a_H + \phi_H \sum_{k=0}^{K} b_H(k,\theta) X_{t-k,t-k-1} + \varepsilon_{Ht}$$

- We have flexibility in 1) predictive variables: we could use squared daily returns, absolute daily returns, etc and 2) the functional form $b_H(k,\theta)$: this facilitates easier curve fitting.
- Once specified (after picking a functional form $b_H(k,\theta)$), we can use NLS to estimate the unknown parameter θ. Use numerical optimization routines to minimize the squared residuals.
- A very practical way to forecast volatility over a relevant holding period.

Probit Model

- Suppose our dependent variable y_i is binary and takes on values 0 and 1. The running example will be models of corporate defaults. In this case, let's say $y_i = 1$ means that firm i defaults and $y_i = 0$ if otherwise. We also have a set of regressors X_i that influence y_i. In our example, they may be leverage ratios, profitability, or macroeconomic conditions. The Probit specification says that y_i and X_i are linked through

$$Prob(y_i = 1|X_i) = \Phi\left(X_i'\beta\right)$$

- We can also write this as

$$y_i = \Phi\left(X_i'\beta\right) + \varepsilon_i$$

where $E\left[\varepsilon_i|X_i\right] = 0$. We can see how the Probit model is a special case of nonlinear regression specification.

Continued

- Note that y_i can only be either 0 or 1. That means that the error term ε_i can only take on two values, so it is not valid to assume that ε_i follows a continuous distribution (for example, normal distribution).
- We can write the likelihood function directly quite easily. Note that

$$Prob(y_i = 1 | X_i, \beta) = \Phi\left(X_i'\beta\right)$$
$$Prob(y_i = 0 | X_i, \beta) = 1 - \left(X_i'\beta\right)$$

and therefore the log-likelihood function is given by

$$\mathscr{L}(\beta | y, X) = \sum_{i=1}^{N} \left[y_i \ln\left(\Phi\left(X_i'\beta\right) + (1 - y_i)\ln\left(1 - \Phi\left(X_i'\beta\right)\right)\right)\right]$$

- We can find $\hat{\beta}$ that maximizes the above log-likelihood function using very simple numerical algorithms.

15.450 Analytics of Finance
Fall 2010