Methods of Economic Investigation II

Lecture 10

Oliver Linton

December 5, 2002

1 Dynamic Regression Models

- We have looked at pure time series models with dynamic response and at static regression models. In practice, we may want to consider models that have both features.

- Distributed lag

$$y_t = \alpha + \sum_{j=0}^{q} \beta_j X_{t-j} + u_t,$$

[could have $q = \infty$], where for now

$$u_t \overset{iid}{\sim} 0, \quad \sigma^2.$$

Captures the idea of dynamic response: affect on y of change in x may take several periods to work through.

- Temporary change. Suppose that

$$x_t \longrightarrow x_t + \Delta$$

but that future x_s are unaffected, then

$$\begin{aligned} y_t &\longrightarrow y_t + \beta_0 \Delta \\ y_{t+1} &\longrightarrow y_t + \beta_1 \Delta \text{ etc.} \end{aligned}$$

- Permanent change. Suppose that

$$x_s \longrightarrow x_s + \Delta, \quad \forall s \geq t.$$

Then

$$\begin{aligned} y_t &\longrightarrow y_t + \beta_0 \Delta \\ y_{t+1} &\longrightarrow y_t + (\beta_0 + \beta_1) \Delta \text{ etc.} \end{aligned}$$

  - The impact effect is $\beta_0 \Delta$.
  - Long run effect is $\Delta \sum_{s=0}^{\infty} \beta_s$.

- When q is large (infinite) there are too many free parameters β_j, which makes estimation difficult and imprecise. To reduce the dimensionality it is appropriate to make restrictions on β_j.

  – The polynomial lag

$$\beta_j = \begin{cases} a_0 + a_1 j + \ldots + a_p j^p & \text{if } j \leq p \\ 0 & \text{else.} \end{cases}$$

  – The Geometric lag

$$\beta_j = \beta\lambda^j, \quad j = 0, 1, \ldots$$

  for some $0 < \lambda < 1$.

- Goemetric lag. This implies that

$$
\begin{aligned}
y_t &= \alpha + \beta \sum_{j=0}^{\infty} \lambda^j x_{t-j} + u_t \\
&= \alpha + \beta \left[\sum_{j=0}^{\infty} (\lambda^j L^j) \right] x_t + u_t \\
&= \alpha + \beta \frac{1}{1 - \lambda L} x_t + u_t.
\end{aligned}
$$

- Therefore,

$$(1 - \lambda L) y_t = \alpha(1 - \lambda L) + \beta x_t + (1 - \lambda L) u_t,$$

which is the same as

$$y_t = \alpha(1 - \lambda) + \lambda y_{t-1} + \beta x_t + u_t - \lambda u_{t-1}.$$

The last equation is called the lagged dependent variable representation.

- More generally [ADL model]

$$A(L)y_t = B(L)x_t + u_t,$$

where A, B are polynomials of order p, q, while

$$C(L)u_t = D(L)\varepsilon_t, \quad \varepsilon_t \text{ i.i.d. } 0, \sigma^2.$$

- This is a very general class of models; estimation, forecasting, and testing have all been worked out at this generality, and one can find accounts of this in advanced time series texts.

- Can write

$$D(L)^{-1}C(L)A(L)y_t = D(L)^{-1}C(L)B(L)x_t + \varepsilon_t,$$

$$y_t = A(L)^{-1}B(L)x_t + A(L)^{-1}C(L)^{-1}D(L)\varepsilon_t,$$

1.1 Adaptive expectations

- Suppose that

$$\underbrace{y_t}_{\text{demand}} = \alpha + \beta \underbrace{x^*_{t+1}}_{\text{expected price}} + \varepsilon_t,$$

but that the expected price is made at time t and is unobserved by the econometrician.

- We observe x_t, where

$$\underbrace{x^*_{t+1} - x^*_t}_{\text{revised expectations}} = (1 - \lambda) \underbrace{(x_t - x^*_t)}_{\text{forecast error}},$$

i.e.,

$$x^*_{t+1} = \underbrace{\lambda x^*_t}_{\text{old forecast}} + \underbrace{(1 - \lambda)x_t}_{\text{news}}.$$

- Write

$$(1 - \lambda L) x_t^* = (1 - \lambda) x_t,$$

which implies that

$$\begin{aligned} x_t^* &= \frac{(1-\lambda)}{1-\lambda L} x_t \\ &= (1-\lambda)\left[x_t + \lambda x_{t-1} + \lambda^2 x_{t-2} + \ldots\right]. \end{aligned}$$

- Therefore,

$$y_t = \alpha + \frac{\beta(1-\lambda)}{1-\lambda L} x_t + \varepsilon_t,$$

which implies that

$$y_t = \lambda y_{t-1} + \alpha(1-\lambda) + \beta(1-\lambda) x_t + \varepsilon_t - \lambda \varepsilon_{t-1}.$$

This is an ADL with an $MA(1)$ error term.

1.2 Partial adjustment

- Suppose that

$$y_t^* = \alpha + \beta x_t,$$

where y_t^* is the desired level.

- However, because of costs of adjustment

$$\underbrace{y_t - y_{t-1}}_{\text{actual change}} = (1-\lambda)(y_t^* - y_{t-1}) + \varepsilon_t.$$

- Substituting we get

$$\begin{aligned} y_t &= (1-\lambda)y_t^* + \lambda y_{t-1} + \varepsilon_t \\ &= \alpha(1-\lambda) + \lambda y_{t-1} + \beta(1-\lambda)x_t + \varepsilon_t. \end{aligned}$$

This is an ADL with an i.i.d. error term - assuming that the original error term was i.i.d.

1.3 Error Correction

- Suppose long run equilibrium is

$$y = \lambda x.$$

- Disequilibria are corrected according to

$$\Delta y_t = \beta \left(y_{t-1} - \lambda x_{t-1} \right) + \lambda \Delta x_{t-1} + \varepsilon_t,$$

where $\beta < 0$.

- This implies that

$$y_t = y_{t-1}(1 + \beta) + \lambda(1 - \beta) x_{t-1} - \lambda x_{t-2} + \varepsilon_t.$$

1.4 Estimation of ADL Models

- Suppose that

$$y_t = \theta_1 + \theta_2 y_{t-1} + \theta_3 x_t + \varepsilon_t,$$

where we have two general cases regarding the error term:

(1) ε_t is i.i.d. 0, σ^2

(2) ε_t is autocorrelated.

- In case (1), we can use OLS regression to get consistent estimates of θ_1, θ_2 and θ_3. The original parameters are related to the θ_j in some way, for example

$$\left.\begin{array}{l} \theta_1 = \alpha(1-\lambda) \\ \theta_2 = \lambda \\ \theta_3 = \beta(1-\lambda) \end{array}\right\}.$$

- In this case, we would estimate the original parameters by indirect least squares

$$\begin{aligned} \widehat{\lambda} &= \widehat{\theta}_2 \\ \widehat{\alpha} &= \frac{\widehat{\theta}_1}{1-\widehat{\theta}_2} \\ \widehat{\beta} &= \frac{\widehat{\theta}_3}{1-\widehat{\theta}_2}. \end{aligned}$$

- In case (2), we must use instrumental variables or some other procedure because OLS will be inconsistent.

  – For example, if
  $$\varepsilon_t = \eta_t - \theta\eta_{t-1},$$
  then y_{t-1} is correlated with ε_t through η_{t-1}. In this case there are many instruments: (1) All lagged x_t, (2) $y_{t-2}, \ldots$.

  – However, when
  $$\varepsilon_t = \rho\varepsilon_{t-1} + \eta_t,$$
  η_t i.i.d. lagged y are no longer valid instruments and we must rely on lagged x.

- There are many instruments; efficiency considerations require that one has a good way of combining them such as in our GMM discussion.

- IV are not generally as efficient as ML when the error terms are normally distributed.

2 GARCH Models

- Engle (1982) introduced the following class of models

$$r_t = \varepsilon_t \sigma_t,$$

where ε_t is i.i.d. $(0,1)$, while

$$\sigma_t^2 = var\left(r_t \left| \mathcal{F}_{t-1}\right.\right)$$

is the (time-varying) conditional variance.

- For example,

$$\sigma_t^2 = \alpha + \gamma r_{t-1}^2,$$

which is the $ARCH(1)$ model.

- Provided $\gamma < 1$, the process r_t is weakly stationary and has finite unconditional variance σ^2 given by

$$\sigma^2 = E(\sigma_t^2) < \infty,$$

where

$$\sigma^2 = \alpha + \gamma\sigma^2 = \frac{\alpha}{1-\gamma}.$$

- This uses the law of iterated expectations $E(Y) = E\left(E(Y\,|I)\right)$ to argue

$$\begin{aligned} E\left(r_{t-1}^2\right) &= E\left(E\left(\varepsilon_{t-1}^2\,|I_{t-1}\right)\sigma_{t-1}^2\right) \\ &= E\left(\sigma_{t-1}^2\right) = \sigma^2. \end{aligned}$$

- The unconditional distribution of r_t is thick-tailed; that is, even if ε_t is normally distributed, r_t is going to have an unconditional distribution that is a mixture of normals and is more leptokurtic. Suppose ε_t is standard normal, then

$$E(\varepsilon_t^4) = 3E^2(\varepsilon_t^2) = 3$$

but (assuming it exists)

$$\mu_4 = E(r_t^4) = E\left(\varepsilon_t^4 \sigma_t^4\right) = 3E(\sigma_t^4),$$

where

$$\begin{aligned} E(\sigma_t^4) &= E\left[\left(\alpha^2 + \gamma^2 r_{t-1}^4 + 2\alpha\gamma r_{t-1}^2\right)\right] \\ &= \alpha^2 + \gamma^2 \mu_4 + 2\alpha\gamma\sigma^2. \end{aligned}$$

- Therefore,

$$\begin{aligned} \mu_4 &= 3\left(\alpha^2 + \gamma^2 \mu_4 + 2\alpha\gamma\sigma^2\right) \\ &= \frac{3\left(\alpha^2 + 2\alpha\gamma\sigma^2\right)}{1 - 3\gamma^2} \\ &\geq 3\sigma^4 = \frac{3\alpha^2}{(1-\gamma)^2}. \end{aligned}$$

- The process r_t is uncorrelated, i.e.,

$$\mathsf{cov}(r_t, r_{t-s}) = 0$$

for all $s \neq 0$. However, the process r_t is dependent so that

$$E\left(g(r_t)g(r_{t-s})\right) \neq E\left(g(r_t)\right)E(h(r_{t-s}))$$

for arbitrary functions g, h, certainly for $g(r) = h(r) = r^2$ this is not true.

- Can write the process as an $AR(1)$ process in u_t^2, i.e.,

$$r_t^2 = \alpha + \gamma r_{t-1}^2 + \eta_t,$$

where $\eta_t = r_t^2 - \sigma_t^2$ is a mean zero innovation that is uncorrelated with its past.

- Therefore, since $\gamma > 0$, the volatility process is positively autocorrelated, i.e.,

$$\mathsf{cov}\left(\sigma_t^2, \sigma_{t-j}^2\right) > 0.$$

Hence we get volatility clustering.

- We can rewrite the process as

$$\sigma_t^2 - \sigma^2 = \gamma \left(r_{t-1}^2 - \sigma^2 \right).$$

  Suppose that $\sigma_{t-1}^2 = \sigma^2$. When we get a large shock, i.e., $\varepsilon_{t-1}^2 > 1$, we get $\sigma_t^2 > \sigma^2$ but the process decays rapidly to σ^2 unless we get a sequence of large shocks $\varepsilon_{t-1+s}^2 > 1$, $s = 0, 1, 2, \ldots$. In fact, for a normal distribution the probability of having $\varepsilon^2 > 1$ is only about 0.32 so we generally see little persistence.

- Although the ARCH model implies volatility clustering, it does not in practice generate enough.

- Generalize to $ARCH(p)$, write

$$\sigma_t^2 = \alpha + \sum_{j=1}^{p} \gamma_j r_{t-j}^2,$$

where p is some positive integer and γ_j are positive coefficients.

- This model is fine, but estimation is difficult. When p is large one finds that the coefficients are imprecisely estimated and can be negative. Have to impose some restrictions on the coefficients.

- Instead $GARCH(1,1)$

$$\sigma_t^2 = \alpha + \beta\sigma_{t-1}^2 + \gamma r_{t-1}^2,$$

where α, β, γ are positive.

  – We have

  $$\sigma_t^2 = \frac{\alpha}{1-\beta} + \gamma \sum_{j=1}^{\infty} \beta^{j-1} r_{t-j}^2,$$

  so that it is an infinite order ARCH model with geometric decline in the coefficients.

- If $\gamma + \beta < 1$, then the process r_t is weakly stationary, i.e., the unconditional variance exists, and

$$\sigma^2 = E(\sigma_t^2) < \infty,$$

where

$$\sigma^2 = \alpha + \beta\sigma^2 + \gamma\sigma^2 = \frac{\alpha}{1 - (\beta + \gamma)}.$$

- Surprisingly, even for some values of β, γ with $\gamma + \beta \geq 1$, the process σ_t^2 is strongly stationary although the unconditional variance does not exist in this case.

- More general class of models $GARCH(p, q)$

$$B(L)\sigma_t^2 = \alpha + C(L)r_{t-1}^2,$$

where A and B are lag polynomials. Usually assume that the parameters in α, B, $C > 0$ to ensure that the variance is positive.

- Other models. For example, one can write the model for log of variance, i.e.,

$$\log \sigma_t^2 + \alpha + \beta \log \sigma_{t-1}^2 + \gamma r_{t-1}^2.$$

This automatically imposes the restriction that $\sigma_t^2 \geq 0$ so there is no need to impose restrictions on the parameters.

- Nelsons $EGARCH$

$$\log \sigma_t^2 = \alpha + \beta \log \sigma_{t-1}^2 + \gamma \varepsilon_t + \delta \left(|\varepsilon_t| - E\left(|\varepsilon_t|\right)\right).$$

- $TARCH$, $SGARCH$, $CGARCH$ etc.

2.1 Estimation

- More general model

$$\begin{aligned} y_t &= b' x_t + \varepsilon_t \sigma_t \\ B(L)\sigma_t^2 &= \alpha + C(L)\left(y_{t-1} - b' x_{t-1}\right)^2 . \end{aligned}$$

- If ARCH effects are present, then we need to use robust estimates of the standard errors for the parameters b of the mean model.

- Also, the variance process itself is of interest. Want to estimate the parameters of σ_t^2 too.

- Let
$$\theta = \left(b, \alpha, \beta_1, \ldots, \beta_p, \gamma_1, \ldots, \gamma_q\right).$$
Estimation by ML suggested by ε_t being standard normal. In this case
$$\ell_T(\theta) = -\frac{1}{2}\sum_{t=1}^{T} \log \sigma_t^2(\theta) - \frac{1}{2}\sum_{t=1}^{T} \frac{\left(y_t - b' x_t\right)^2}{\sigma_t^2(\theta)}.$$
The ML estimator of b, θ can be obtained from this criterion.

  - This involves nonlinear optimization.

  - Have to impose the inequality restrictions on the parameters which can be tricky.