Econometrics, Lecture 10A
ARIMA and unit roots

Ron Smith
EMS, Birkbeck, University of London

Autumn 2020

Last time

- Defined a covariance stationary stochastic process: constant mean, variances and covariances after removal of deterministic elements
- Defined the lag operator, $Ly_t = y_{t-1}$
- Introduced AR(p)

$$y_t = \sum_{i=1}^{p} \rho_i y_{t-i} + \varepsilon_t$$

  and MA(q) processes

$$y_t = \sum_{i=1}^{q} \mu_i \varepsilon_{t-i} + \varepsilon_t$$

- Defined the order of intergration the number of times a series needed to be differenced to make it stationary.

ARIMA

- Combining AR and MA processes, gives ARMA, ARMA(1,1) is

$$y_t = \alpha + \rho y_{t-1} + \varepsilon_t + \mu \varepsilon_{t-1}.$$

- Difference the data d times to make them stationary then model as an ARMA process of order p and q, giving an Autoregressive Integrated Moving Average, ARIMA(p,d,q) process:

$$A^p(L)\Delta^d y_t = \alpha + B^q(L)\varepsilon_t.$$

- ARIMA(1,1,1) and ARIMA(2,2,2) processes are

$$\begin{aligned}
\Delta y_t &= \alpha + \rho \Delta y_{t-1} + \varepsilon_t + \mu \varepsilon_{t-1,}. \\
\Delta^2 y_t &= \alpha + \rho_1 \Delta^2 y_{t-1} + \rho_2 \Delta^2 y_{t-2} + \varepsilon_t + \mu_1 \varepsilon_{t-1} + \mu_2 \varepsilon_{t-2}.
\end{aligned}$$

Common Factors

- ARIMA models often describe the univariate dynamics of a single economic time-series quite well and are widely used for forecasting.
- A potential problem is effects cancelling out. Suppose the true model is a random walk and we multiply both sides by $(1-\rho L)$

$$\begin{aligned}
\Delta y_t &= \alpha + \varepsilon_t, \\
(1-\rho L)\Delta y_t &= (1-\rho L)\alpha + (1-\rho L)\varepsilon_t, \\
\Delta y_t &= (1-\rho)\alpha + \rho \Delta y_{t-1} + \varepsilon_t - \rho \varepsilon_{t-1}.
\end{aligned}$$

- You can estimate an ARIMA(1,1,1) model with significant AR and MA coefficients of opposite signs, that are individually but not jointly significant, cancelling out.

Frisch-Waugh theorem

- Most economic time-series, e.g. log GDP, are non-stationary, trended, so can include a trend in the regression

$$y_t = \alpha + \beta x_t + \gamma t + \varepsilon_t,$$

- Alternatively you can detrend the data by regressing y_t and x_t on t and using the residuals

$$\begin{aligned} y_t &= a_y + b_y t + \widetilde{y}_t \\ x_t &= a_x + b_x t + \widetilde{x}_t \\ \widetilde{y}_t &= \beta \widetilde{x}_t + \varepsilon_t. \end{aligned}$$

- The two estimates of β will be identical.
- Can do this for any variable and it is a useful trick for looking at the relationship between two variables controlling for others.

Trend and difference stationary processes

- The trend in economic time-series can be generated in two ways.
- As stationary around a deterministic trend:

$$y_t = \alpha + \rho y_{t-1} + \gamma t + \varepsilon_t \tag{1}$$

  with $\mid \rho \mid < 1$. The effects of the shocks ε_t are transitory and die away through time, since ε_{t-i} is multiplied by ρ^i when you substitute back. If the variables are in logs, the long run growth rate is $g = \gamma/(1-\rho)$.
- As a random walk with drift, difference stationary:

$$\begin{aligned} \Delta y_t &= \alpha + \varepsilon_t \\ y_t &= \alpha + y_{t-1} + \varepsilon_t \end{aligned}$$

  The long run growth rate is α.

Testing

- We want to test the null of a difference stationary process (one with a unit root) against the alternative of a trend stationary process.
- Substitute $\gamma = g(1-\rho)$ then subtract y_{t-1} from both sides, so we can write the trend stationary process as:

$$\begin{aligned} y_t &= \alpha + \rho y_{t-1} + \gamma t + \varepsilon_t && (2) \\ \Delta y_t &= \alpha + (\rho - 1)(y_{t-1} - gt) + \varepsilon_t && (3) \\ \Delta y_t &= \alpha + \beta(y_{t-1} - gt) + \varepsilon_t && (4) \end{aligned}$$

  where $\beta = \rho - 1$.
- If $\rho = 1$ or equivalently $\beta = 0$, we get the random walk with drift: with growth rate α.
- We use the restricted trend form (4) since making $\rho = 1$ in (2) would make $\Delta y_t = \alpha + \gamma t + \varepsilon_t$. But we do not want a trend in the change.

Random Walks and stochastic trends

- Substituting back in the random walk we get

$$\begin{aligned} y_t &= \alpha + y_{t-1} + \varepsilon_t \\ y_t &= \alpha + (\alpha + y_{t-2} + \varepsilon_{t-1}) + \varepsilon_t \\ y_t &= y_{t-2} + 2\alpha + \varepsilon_t + \varepsilon_{t-1} \end{aligned}$$

- Continuing the process to period zero, we get:

$$y_t = y_0 + \alpha t + \sum_{i=0}^{t-1} \varepsilon_{t-i}. \tag{5}$$

- In a difference stationary process, the series is determined by an initial value, y_0, a deterministic trend αt, and a 'stochastic trend', $\sum_{i=0}^{t-1} \varepsilon_{t-i}$, the sum of past errors. the efects of the shocks are permanent or persistent, they last for ever, and

Trend in change is quadratic trend in level

If we had not restricted (4) so that the trend term dropped out when $\beta = 0$, there would be a quadratic trend in y_t. Show this by substituting back in

$$
\begin{aligned}
y_t &= \alpha + y_{t-1} + \gamma t + \varepsilon_t \\
y_t &= \alpha + (\alpha + y_{t-2} + \gamma(t-1) + \varepsilon_{t-1}) + \gamma t + \varepsilon_t
\end{aligned}
\tag{6}
$$

etc.

Testing for unit roots

- Choosing between the trend stationary and difference stationary model is a matter of testing $H_0 : \beta = 0$ or equivalently $\rho = 1$; whether there is a 'unit root' in y_t.
- To do this we can estimate (4), a regression of Δy_t on a constant, y_{t-1} and a linear trend to estimate $\widehat{\beta}$ the coefficient on y_{t-1};
- Construct the 't statistic' $\tau_\beta = \widehat{\beta}/SE(\widehat{\beta})$ to test $H_0 : \beta = 0$; against $H_1 : \beta < 0$.
- If we do not reject H_0 we conclude that there is a unit root in y_t, it is I(1), stationary after being differenced once.
- If we reject the null we conclude that y_t is trend stationary I(0). This is a one-sided test and if $\widehat{\beta} > 0$, we certainly do not reject the null of a unit root.

Dickey Fuller and Augmented Dickey Fuller tests

- The test statistic τ_β does not have a standard t distribution, but a Dickey Fuller distribution. This is because under H_0 the dependent variable is $I(0)$ but the regressor, y_{t-1} is $I(1)$, so usual Gauss-Markov assumptions do not hold.
- With trend the 5% critical value, CV, is about -3.5 (-2.9 no trend), programs give you CVs or p values.
- To get good estimates of (4) we require ε_t white noise. To remove any serial correlation, lags of the dependent variable are added to give the 'Augmented Dickey Fuller' (ADF) regression:

$$\Delta y_t = \alpha + \beta y_{t-1} + \gamma t + \sum_{i=1}^{p} \delta_i \Delta y_{t_{-i}} + \varepsilon_t \tag{7}$$

where p is chosen to try to make the residual white noise. Again the procedure is to use the t ratio on β with the non standard critical values to test $H_0 : \beta = 0$ against $H_1 : \beta < 0$.

Reparameterisations

- (7) is a reparameterisation of a AR(p+1) with trend, e.g. AR3.

$$
\begin{aligned}
y_t &= \rho_1 y_{t-1} + \rho_2 y_{t-2} + \rho_3 y_{t-3} + \varepsilon_t, \\
y_t &= \rho_1 y_{t-1} + (\rho_2 + \rho_3) y_{t-2} - \rho_3 (y_{t-2} - y_{t-3}) + \varepsilon_t, \\
y_t &= (\rho_1 + \rho_2 + \rho_3) y_{t-1} - (\rho_2 + \rho_3)(y_{t-1} - y_{t-2}) \\
&\quad -\rho_3 (y_{t-2} - y_{t-3}) + \varepsilon_t, \\
y_t - y_{t-1} &= (\rho_1 + \rho_2 + \rho_3 - 1) y_{t-1} - (\rho_2 + \rho_3)(y_{t-1} - y_{t-2}) \\
&\quad -\rho_3 (y_{t-2} - y_{t-3}) + \varepsilon_t, \\
\Delta y_t &= \beta y_{t-1} + \delta_1 \Delta y_{t-1} + +\delta_2 \Delta y_{t-2} + \varepsilon_t.
\end{aligned}
$$

Testing for I(2)

- To test for a unit root in levels, $H_0 : I(1)$ against $H_1 : I(0)$, in the intercept only version, we estimate

$$\Delta y_t = \alpha + \beta y_{t-1} + \sum_{i=1}^{p} \delta_i \Delta y_{t_{-i}} + \varepsilon_t$$

  and use the t statistic on $\widehat{\beta}$ to test $H_0 : \beta = 0$; against $H_1 : \beta < 0$.

- To test for a unit root in differences, $H_0 : I(2)$ against $H_1 : I(1)$, you just take a further difference:

$$\Delta^2 y_t = \alpha + \beta \Delta y_{t-1} + \sum_{i=1}^{p} \delta_i \Delta^2 y_{t_{-i}} + \varepsilon_t$$

  Trend in the change are not common for economic series. Again $H_0 : \beta = 0$; against $H_1 : \beta < 0$.

Other types of test

There are many procedures for determining whether there is a unit root. They differ, for instance, in

- whether they use the null of a unit root like the ADF or the null of stationarity, like KPSS;
- how they correct for serial correlation (in a parametric way like the ADF adding lags or in a non-parametric way like Phillips Peron where you use a robust variance estimator);
- whether they include other variables; whether they use GLS detrending; and whether they use both forward and backward regressions.
- Many programs give you a lot of choices.

It is hard to decide on the order of integration

- Tests have low power, it is hard to distinguish $\rho = 1$ from $\rho = 0.97$.
- Sensitive to choice of deterministics (intercept and trend), and treatment of serial correlation
- Power depends on the span of the data not the number of observations. Can seem I(1) on short samples, I(0) on long.
- An I(0) process with a step change will appear I(1), since the shock (change in level) is permanent.
- The order of integration is a univariate statistical summary of how the time series moves over the sample, it is not an inherent structural property of the series. Whether you treat a variable as $I(0)$ or $I(1)$ depends on the purpose of the exercise, for estimation it is often safer to treat it as $I(1)$.
- YOU DO NOT ALWAYS HAVE TO DECIDE. Unrestricted level ARDL or VARs are robust to not knowing the order of integration.

Next time

- Will combine the univariate ARIMA model with the linear regression model.
- Rather than having unobserved lagged shocks ε_{t-i}, will have observed lagged shocks x_{t-i}.
- Autoregessive Distributed Lag, ARDL(2,2)

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \varepsilon_t$$

- Very flexible framework which can handle diferent orders of integration and cointegration.
- Can be given a theoretical interpretation in terms of a long-run equilibrium relationship and an adjustment process.