Masters Level, Time Series Notes

Anastasia Zervou

February 18, 2019

1 Lecture 1

We talked about:
-Nature of time series data
-One equation models
-Univariate vs. multivariate one-equation models
-Static vs. dynamic model
-Distributed lag models and temporary vs. permanent shocks

Readings:
-Slides and notes of lecture 1
-Wooldridge Chapter 10

1.1 Additional Notes on Distributed Lag Model

Clarification on Distributed lag models: The effect of a change in an explanatory variable on the dependent variable, depends on the nature of the change and on the modeling choice we have made. I will use a second order distributed lag model, a temporary and a permanent change in explanatory variable, for exposition.

Distributed lag model:

$$y_t = \alpha_0 + \delta_0 x_t + \delta_1 x_{t-1} + \delta_2 x_{t-2} + u_t$$

Temporary shock: We are interested in the effect of x_t on current and future values of y. Let's assume that $x_{t-j} = x_{t+s} = c$ for all $j, s \neq 0$, and $x_t = c + 1$. Then, ignoring for a second the errors we have:

$$y_{t-1} = \alpha_0 + \delta_0 c + \delta_1 c + \delta_2 c,$$

$$y_t = \alpha_0 + \delta_0 (c+1) + \delta_1 c + \delta_2 c,$$

$$y_{t+1} = \alpha_0 + \delta_0 c + \delta_1 (c+1) + \delta_2 c,$$

$$y_{t+2} = \alpha_0 + \delta_0 c + \delta_1 c + \delta_2 (c+1),$$

$$y_{t+3} = \alpha_0 + \delta_0 c + \delta_1 c + \delta_2 c.$$

Here, the change in x_t affects y until y_{t+2}, as $y_t - y_{t-1} = \delta_0$, $y_{t+1} - y_{t-1} = \delta_1$, $y_{t+2} - y_{t-1} = \delta_2$, but $y_{t+3} - y_{t-1} = 0 = y_{t+s} - y_{t-1}$ for all $s > 2$.

Permanent shock: We are interested in the effect of x_t on current and future values of y. Let's assume that $x_{t-j} = c$ for all $j > 0$, and $x_t = x_{t+s} = c + 1$ for all $s > 0$. Then, ignoring for a second the errors we have:

$$y_{t-1} = \alpha_0 + \delta_0 c + \delta_1 c + \delta_2 c,$$
$$y_t = \alpha_0 + \delta_0(c+1) + \delta_1 c + \delta_2 c,$$
$$y_{t+1} = \alpha_0 + \delta_0(c+1) + \delta_1(c+1) + \delta_2 c,$$
$$y_{t+2} = \alpha_0 + \delta_0(c+1) + \delta_1(c+1) + \delta_2(c+1),$$
$$y_{t+3} = \alpha_0 + \delta_0(c+1) + \delta_1(c+1) + \delta_2(c+1).$$

Here, $y_t - y_{t-1} = \delta_0$, $y_{t+1} - y_{t-1} = \delta_0 + \delta_1$, $y_{t+2} - y_{t-1} = \delta_0 + \delta_1 + \delta_2 = y_{t+s} - y_{t-1}$ for all $s > 2$.

Notice that the effect of x on current and future values of y, depends on the number of lags we decided to use in the distributed lag model, and on the nature of the change in x, i.e., temporary or permanent shock.

2 Lecture 2

We talked about:
-TS Assumptions (vs CS Assumptions)
-Trends and Seasonality -Distributed Lag model and bands

Readings:
-Slides and notes of lecture 2
-Wooldridge Chapter 10

2.1 Differences between CS and TS

-CS sample: collection of observations from population; TS sample: collection of observations specifically ordered
-CS sample size: number of observations in the sample; TS sample size: number of time periods
-CS random variable vs TS stochastic process (= collection of random variables with time subscript
-CS population vs TS set of all possible realizations of the stochastic process

So in time series, we are forced to think about joint distribution of x_t, x_{t+1}, x_{t+2} etc. In CS, we did not have to, as the random sampling assumption is reassuring that x_i is independent of x_j for two different i, j individuals.

2.2 Small sample assumptions for time series

1. Linear model in coefficients (same as CS)
2. No perfect collinearity and some variation on the X's (same as CS)
3. Strict exogeneity: this combines the random sampling assumption of CS, and the (simple) exogeneity assumption of CS.
Those 3 assumptions give us unbiased OLS estimators
4. Homoskedasticity (same as CS)
5. No serial correlation: we did not need to make this assumption in CS, as we assume random sampling.
Assumptions 1-5 imply, apart form unbiased OLS estimators, OLS unbiased variance estimator, and also Best Linear Unbiased Estimators (same as CS).

2.3 Inference

We usually want to make inference, and thus we need to know the distribution of the estimators. Conditional on the right hand side variables, we then need to assume a distribution for the errors (or the left-hand-side variable). In CS, we make the assumption that conditional on the right hand side variables, the errors are identically and independently distributed as normal with mean μ and variance σ^2. This assumption is strong and implies also exogeneity and homoskedasticity.

In TS we make the normality assumption as well, but conditional on all right hand side variables, for all time subscripts. The normality assumption in TS is strong and implies the assumptions of exogeneity, homoskedasticity and no serial correlation.

2.4 Distributed Lag Model and bands calculation

In its general form:

$$y_t = \alpha_0 + \delta_0 x_t + \delta_1 x_{t-1} + \delta_2 x_{t-2} + .. + +\delta_T x_{t-T} + \rho Z_t + u_t \qquad (1)$$

Let's name $\theta = \sum_{k=0}^{T} \delta_k$. Then we can rewrite equation (1) as:

$$y_t = \alpha_0 + (\theta - \sum_{k=1}^{T} \delta_k) x_t + \sum_{k=1}^{T} \delta_k x_{t-k} + \rho Z_t + u_t$$

$$= \alpha_0 + \theta x_t + \sum_{k=1}^{T} \delta_k (x_{t-k} - x_t) + \rho Z_t + u_t.$$

and the estimate of θ is the cumulative effect, for which we also obtain standard errors. Specifically, we run:

$$y_t = \alpha_0 + \delta_0 x_t + \sum_{k=1}^{T} \delta_k x_{t-k} + \rho Z_t + u_t,$$

to find the effect of the $\theta_0 = \delta_0$. We run:

$$y_t = \alpha_0 + (\delta_0 + \delta_1) x_t + \delta_1 (x_{t-1} - x_t) + \sum_{k=2}^{T} \delta_k x_{t-k} + \rho Z_t + u_t,$$

to find the effect of the the sum of $\theta_1 = \delta_0 + \delta_1$. We run:

$$y_t = \alpha_0 + (\delta_0 + \delta_1 + \delta_2) x_t + \delta_1 (x_{t-1} - x_t) + \delta_2 (x_{t-2} - x_t) + \sum_{k=3}^{T} \delta_k x_{t-k} + \rho Z_t + u_t,$$

to find the effect of the the sum of $\theta_2 = \delta_0 + \delta_1 + \delta_2$, and so on. Then, we can also graph the estimates $\hat{\theta_0}, \hat{\theta_1}, \hat{\theta_2}$, etc, together with their error bands. Note: Distributed lags models (together with static models) have hopes of satisfying the small sample TS assumptions, and thus producing unbiased estimators, as well as using the $t - test$ for inference (after assuming normality of errors).

2.5 Creating a trend variable

A trend variable is an array: $1, 2, 3, ..T$, depending on the size of the time series sample. The trend variable does not need to start at 1, but needs to be increasing by 1 unit each time. A quadratic trend is similarly, an array with a quadratic trend $1, 4, 9..T^2$.

2.6 Trend and Seasonality

Correcting for trend or seasonality can be done in two ways:
a. Include a trend or seasonality dummies on the right-hand side.
b. De-trend or seasonally adjust the data and run regressions on the de-trended/seasonally adjusted data.
The coefficients should be the same in either case.
However, trends and seasonality overstate the variance of the left hand side variable, and thus R^2 and adjusted R^2 are higher in an a-type regression.

3 Lecture 3

We talked about:
-Asymptotic/ Large sample assumptions (Wooldridge Chapter 11)
-AR(1), MA(1) models
-Autocorrelation Functions

Readings:
-Slides and notes of lecture 3
-Wooldridge Chapter 11

In TS, under the (very restrictive) small sample assumptions, we could fit either static, or distributed lag models. However, we often think in terms of ARMA (Autoregressive Moving Average) models. Any ARMA representation would violate strict exogeneity. Thus, in TS we rely heavily on asymptotic properties.

3.1 Asymptotic properties for TS

1. Linear model in coefficients **and** right and left hand side variables are stationary and weakly dependent.
2. No perfect collinearity and some variation on the X's
3. Contemporaneous exogeneity
Those 3 assumptions give consistency for the OLS estimators
4. Homoskedasticity
5. No serial correlation
Assumptions 1-5 imply asymptotic normality, so we can make inference without extra assumptions. And, very importantly, we can use ARMA models as the strict exogeneity assumption has been relaxed.

3.2 Autoregressive Model of order 1: AR(1) model

Does an AR(1) model satisfy covariance stationarity? Let the error be serially uncorrelated, with mean zero and constant variance σ^2. An AR(1) model is a first order autoregressive model:

$$y_t = \theta y_{t-1} + \varepsilon_t.$$

It has mean:

$$E(y_t) = \theta E(y_{t-1}) = 0,$$

for $\theta \neq 1$.

Its variance is:

$$Var(y_t) = Var(\theta y_{t-1} + \varepsilon_t) = \theta^2 Var(\theta y_{t-1}) + \sigma^2;$$

thus, for $Var(y_t) = Var(y_{t-1})$ (required for covariance stationary series) we have:

$$Var(y_t) = \frac{\sigma^2}{1-\theta^2},$$

which is bounded for $|\theta| < 1$.

Also,

$$Cov(y_t, y_{t-k}) = \theta^k \frac{\sigma^2}{1-\theta},$$

where only the distance between y_t and y_{t-k} matters, and goes to zero for $|\theta| < 1$ as that distance goes to infinity.[1]

Thus, the process is covariance stationary if $|\theta| < 1$.

We can aslo expand the model $y_t = \theta y_{t-1} + \varepsilon_t = \theta^2 y_{t-2} + \theta\varepsilon_{t-1} + \varepsilon_t = ...$, so we have:

$$y_t = \theta^n y_{t-n} + \Sigma_{j=1}^{n-1} \theta^j \varepsilon_{t-j}.$$

From the above equation we can see: $a)$ for n going to infinity, the AR(1) process translates into an infinite MA process for $|\theta| < 1$. $b)$ the effect of a shock, decays over time for $|\theta| < 1$, as $\frac{\partial y_t}{\partial \varepsilon_t} = \theta$, $\frac{\partial y_{t+1}}{\partial \varepsilon_t} = \theta^2 < \theta$ and so on.

[1]For example: $Cov(y_t, y_{t-1}) = E(y_t y_{t-1}) - 0 = E((\theta y_{t-1} + \varepsilon_t) y_{t-1}) = \theta E(y_{t-1})^2 + \theta E(\varepsilon_t y_{t-1}) = \theta E(y_{t-1})^2 = \theta Var(y_t) = \theta \frac{\sigma^2}{1-\theta}$. Remember that the mean of y_t is zero; also, the expectations are conditional on the observables. Similarly, $Cov(y_t, y_{t-2}) = E(y_t y_{t-2}) - 0 = E((\theta y_{t-1} + \varepsilon_t) y_{t-2}) = E((\theta^2 y_{t-2} + \theta\varepsilon_{t-1} + \varepsilon_t) y_{t-2}) = \theta^2 E(y_{t-2})^2 + \theta E(\varepsilon_{t-1} y_{t-2}) + E(\varepsilon_t y_{t-2}) = \theta^2 E(y_{t-2})^2 = \theta^2 Var(y_t) = \theta^2 \frac{\sigma^2}{1-\theta}$.

3.3 Moving Average Model of order 1: MA(1) model

Does an MA(1) model satisfy covariance stationarity? Let the error be serially uncorrelated, with mean zero and constant variance σ^2. An MA(1) model is:

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

It has mean $E(y_t) = 0$. Its variance is:

$$Var(y_t) = E(\varepsilon_t + \alpha\varepsilon_{t-1})^2 = \sigma^2 + \alpha^2\sigma^2 = (1+\alpha^2)\sigma^2.$$

Also,

$$Cov(y_t, y_{t-1}) = E[(\varepsilon_t + \alpha\varepsilon_{t-1})(\varepsilon_{t-1} + \alpha\varepsilon_{t-2})] = E(\varepsilon_t\varepsilon_{t-1} + \alpha\varepsilon_t\varepsilon_{t-2} + \alpha\varepsilon_{t-1}^2 + \alpha^2\varepsilon_{t-1}\varepsilon_{t-2})$$

$$= \alpha E(\varepsilon_{t-1}^2) = \alpha\sigma^2.$$

$Cov(y_t, y_{t-k}) = 0$ for $k > 1$.[2] So a finite MA process is covariance stationary and weakly dependent. Also, the effect of a shock is short-lived, as $\frac{\partial y_t}{\partial \varepsilon_t} = 1$, $\frac{\partial y_{t+1}}{\partial \varepsilon_t} = \alpha$, but $\frac{\partial y_{t+2}}{\partial \varepsilon_t} = 0$ and $\frac{\partial y_{t+k}}{\partial \varepsilon_t} = 0$ for $k > 1$.

Finally, we can expand the model, as: $y_t = \varepsilon_t + \alpha\varepsilon_{t-1} = \varepsilon_t + \alpha y_{t-1} - \alpha^2\varepsilon_{t-2} = \varepsilon_t + \alpha y_{t-1} - \alpha^2 y_{t-2} + \alpha^3\varepsilon_{t-3} = .. = \varepsilon_t + \alpha\Sigma_{j=0}^{n-1}(-\alpha)^j y_{t-j-1} + \alpha(-\alpha)^n\varepsilon_{t-n-1}$. Thus, we can write the MA(1) model as an infinite AR model for $|\alpha| < 1$:

$$y_t = \alpha\Sigma_{j=0}^{\infty}(-\alpha)^j y_{t-j-1} + \varepsilon_t.$$

3.4 ARMA models

We will use a lot ARMA models, combining AR and MA terms:

$$y_t = \theta_1 y_{t-1} + \theta_2 y_{t-2} + ... + \theta_p y_{t-p} + \varepsilon_t + \alpha_1\varepsilon_{t-1} + \alpha_2\varepsilon_{t-2} + ... + \alpha_q\varepsilon_{t-q}$$

3.5 ACF

Autocorrelation: Evolution of autocorrelation coefficients, $\rho_{t,t-k}$ along k. Autocorrelation coefficient $\rho_{t,t-k}$ is:

$$\rho_{t,t-k} = \frac{Cov(y_t, y_{t-k})}{\text{Var}\, y_t}$$

[2] For example, $Cov(y_t, y_{t-2}) = E[(\varepsilon_t + \alpha\varepsilon_{t-1})(\varepsilon_{t-2} + \alpha\varepsilon_{t-3})] = E(\varepsilon_t\varepsilon_{t-2} + \alpha\varepsilon_t\varepsilon_{t-3} + \alpha\varepsilon_{t-1}\varepsilon_{t-2} + \alpha^2\varepsilon_{t-1}\varepsilon_{t-3}) = 0$.

4 Lecture 4

We talked about:
-ARMA models
-PACF

Readings:
-Slides and notes of lecture 3
-Wooldridge Chapter 11
-Further readings: Verbeek Ch 8

4.1 ARMA models

$$y_t = \theta_1 y_{t-1} + \theta_2 y_{t-2} + ... + \theta_p y_{t-p} + \varepsilon_t + \alpha_1 \varepsilon_{t-1} + \alpha_2 \varepsilon_{t-2} + ... + \alpha_q \varepsilon_{t-q}$$

4.2 ACF

Autocorrelation: Evolution of autocorrelation coefficients, $\rho_{t,t-k}$ along k.
Autocorrelation coefficient $\rho_{t,t-k}$ is:

$$\rho_{t,t-k} = \frac{Cov(y_t, y_{t-k})}{\text{Var}\, y_t}$$

For example, for an AR(1) model, $y_t = \theta y_{t-1} + \varepsilon_t$: $Cov(y_t, y_{t-1}) = E(y_t y_{t-1}) - 0 = E((\theta y_{t-1} + \varepsilon_t) y_{t-1}) = \theta E(y_{t-1})^2 + \theta E(\varepsilon_t y_{t-1}) = \theta E(y_{t-1})^2 = \theta Var(y_t) = \theta \frac{\sigma^2}{1-\theta}$. Remember that the mean of y_t is zero; also, the expectations are conditional on the observables. Then, $\rho_{t,t-1} = \frac{Cov(y_t, y_{t-1})}{\text{Var}\, y_t} = \theta$.

Similarly, $Cov(y_t, y_{t-2}) = E(y_t y_{t-2}) - 0 = E((\theta y_{t-1} + \varepsilon_t) y_{t-2}) = E((\theta^2 y_{t-2} + \theta \varepsilon_{t-1} + \varepsilon_t) y_{t-2}) = \theta^2 E(y_{t-2})^2 + \theta E(\varepsilon_{t-1} y_{t-2}) + E(\varepsilon_t y_{t-2}) = \theta^2 E(y_{t-2})^2 = \theta^2 Var(y_t) = \theta^2 \frac{\sigma^2}{1-\theta}$. Thus, $\rho_{t,t-2} = \frac{Cov(y_t, y_{t-2})}{\text{Var}\, y_t} = \theta^2$.

In general, for the AR(1) model with coefficient θ, $\rho_{t,t-k} = \theta^k$, and thus tails off as k increases, for $|\theta| < 1$.

For an MA(1) model, we found in section 3.3 above that $Cov(y_t, y_{t-1}) = E[(\varepsilon_t + \alpha \varepsilon_{t-1})(\varepsilon_{t-1} + \alpha \varepsilon_{t-2})] = E(\varepsilon_t \varepsilon_{t-1} + \alpha \varepsilon_t \varepsilon_{t-2} + \alpha \varepsilon_{t-1}^2 + \alpha^2 \varepsilon_{t-1} \varepsilon_{t-2}) = \alpha E(\varepsilon_{t-1}^2) = \alpha \sigma^2$, and $Var(y_t) = E(\varepsilon_t + \alpha \varepsilon_{t-1})^2 = \sigma^2 + \alpha^2 \sigma^2 = (1 + \alpha^2)\sigma^2$. Thus $\rho_{t,t-1} = \frac{\alpha \sigma^2}{(1+\alpha^2)\sigma^2} = \frac{\alpha}{1+\alpha^2}$. In addition, $Cov(y_t, y_{t-k}) = 0$ for $k > 1$, so $\rho_{t,t-k} = 0$ for $k > 1$. That means that the ACF goes to zero for lags larger than 1 for the MA(1) model.

4.3 PACF

Partial autocorrelation function shows the evolution of the partial autocorrelation coefficient.

The kth-order sample partial autocorrelation coefficient is the estimate for θ_p in an $AR(p)$ model, which we denote by $\hat{\theta}_{kk}$. It measures the extra correlation between y_t and y_{t-k}, after adjusting for the intermediate values $y_{t-1}, .., y_{t-k+1}$.

For example, estimating:

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

gives us the partial autocorrelation coefficient $\hat{\theta}_{11}$, and estimating

$$y_t = \theta_1 y_{t-1} + \theta_2 y_{t-2} + \varepsilon_t,$$

gives us the partial autocorrelation coefficient $\hat{\theta}_{22}$ and so on.

We can use this information for model selection. Given the large sample assumptions (including stationarity of the process), the OLS estimators of θs are consistent; also, asymptotic normality is implied and we can make inference. We can test the $AR(k-1)$ vs. the $AR(k)$ model, testing the H_0: $\theta_{kk} = 0$. So, the partial autocorrelation coefficient is zero for $k > p$ in a $AR(p)$ model. Thus, the PACF for an AR(p) model goes to zero for lags higher than p.

Given that a (covariance stationary) MA model could be written as an infinite AR process, we realize that for an MA process, the PACF would decay slowly and have infinite length.

4.4 Choice of model

We can get some useful information for modeling time series data from looking the ACF and PACF. Specifically:

An AR(p) model has:
- ACF that is infinite in extent (it tails off).
-PACF that goes to zero for lags larger than p.

An MA(q) process has:
-ACF that goes to zero for lags larger than q.
-PACF that is infinite in extent (it tails off).

In the absence of any of these two situations, a combined ARMA model may provide a good representation of the data.

5 Lecture 5

We talked about:
What is autocorrelation
What OLS properties fail under autocorrelation
How to test for autocorrelation
How to correct for autocorrelation

Readings:
-Slides and notes of lecture 5
-Wooldridge Chapter 12

5.1 Autocorrelation: What is it, what does it imply

One of the TS assumptions is that of no autocorrelation, i.e., $E(u_t, u_s) = 0$ for $t \neq s$. That means, that if for example there is an omitted variable in the model, that omitted variable does not exhibit persistence. However, it is very usual that macroeconomic variables exhibit persistence; this is true for the omitted variables as well. Thus, positive autocorrelation is common in TS data.

The most usual form of autocorrelation is the $AR(1)$ type: $u_t = \rho u_{t-1} + e_t$, where e_t is identically and independently distributed with mean zero and constant variance.

-Small Sample: If we have $AR(1)$ type of autocorrelation, and we don't have lagged dependent variable as explanatory variable in our model, then the OLS estimates are unbiased, under strict exogeneity. However, our OLS estimates are in general not efficient, and thus are not BLUE.
-Large sample-Asymptotic: If we have $AR(1)$ type of autocorrelation, and we don't have lagged dependent variable as explanatory variable in our model, then the OLS estimates are consistent, under contemporaneous exogeneity, for covariance stationary and weak dependent process. However, our OLS estimates are in general not efficient.

If we have a lagged dependent variable in our model, then the OLS estimators could be biased and inconsistent, when there is $AR(1)$ type of autocorrelation. For example if our model is:

$$Y_t = \beta_1 + \beta_2 X_t + \beta_3 Y_{t-1} + u_t,$$

then lagging the model by a period we get:

$$Y_{t-1} = \beta_1 + \beta_2 X_{t-1} + \beta_3 Y_{t-2} + u_{t-1}.$$

If there is autocorrelation of the usual AR(1) form: $u_t = \rho u_{t_1} + \varepsilon_t$, then Y_{t-1} that correlates with u_{t-1}, also correlates with u_t. In that case, contemporaneous exogenity does not hold, and as a consequence, OLS will yield inconsistent estimates.

5.2 Autocorrelation Tests

1) Assume that TS 1-3 assumptions hold, and also assume asymptotic normality for the errors of your model. One way to detect autocorrelation is to run your model, save residuals, and test H_0: $\rho = 0$ in the regression $\hat{u_t} = \rho \hat{u}_{t-1} + e_t$. Note that we have assumed that $\hat{u_t}$ follows a normal distribution as u_t; how valid is this assumption depends on how large is your sample (see slides of lecture 5 for a graphical representation).

2) Durbin-Watson test (exact):

$$DW = \Sigma_{t=2}^{n} (\hat{u}_t - \hat{u}_{t-1})^2 / \Sigma_{t=2}^{n} (\hat{u}_t)^2 \approx 2(1 - \hat{\rho}).$$

Under the assumptions TS.1-TS.6, the Durbin-Watson test is an exact test. We need to test the null $H_0 : \rho = 0$ vs. $H_0 : \rho > 0$. The Durbin-Watson test works with a lower and an upper bound for the critical value. In the area between the bounds the test result is inconclusive. Then, we reject if $DW < d_L$, and fail to reject if $DW > d_U$. Example: Sstatic Phillips Curve: $DW = 0.80 < d_L = 1.32$, thus, reject null hypothesis of no serial correlation.

3) Assume that TS 1-3 assumptions hold, and also assume asymptotic normality for the errors of your model. We can generalize the first test for a more general model:

$$y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + ... + \beta_k x_{tk} + u_t.$$

The t-test for autocorrelation can be easily generalized to allow for the possibility that the explanatory variables are not strictly exogenous. Now the residual could be $u_t = a_0 + a_1 x_{t1} + ... + a_k x_{tk} + \rho u_{t-1} + e_t$:, where we have an AR(1) model for serial correlation and exogeneity not holding (with an i.i.d. series e_t). Then we run:

$$\hat{u}_t = a_0 + a_1 x_{t1} + ... + a_k x_{tk} + \rho \hat{u}_{t-1} + error,$$

and test the null $H_0 : \rho = 0$. The test now allows for the possibility that the strict exogeneity assumption is violated.

4) What we use often in practice: LM test (asymptotic):

$$T \cdot R^2 \sim \chi^2(z),$$

with R^2 computed from the regression

$$\hat{u}_t = c + \rho_1 \hat{u}_{t-1} + \ldots + \rho_z \hat{u}_{t-z} + v_t.$$

Here we test the null is $H_0 : \rho_1 = .. = \rho_z = 0$.

Or use the more general Breusch-Godfrey test:

$$\hat{u}_t = c + \rho_1 \hat{u}_{t-1} + \ldots + \rho_z \hat{u}_{t-z} + a_1 x_{t1} + .. + a_k x_{tk} + \beta_2 v_t,$$

which is also an LM test, i.e., it uses the statistic $T \cdot R^2 \sim \chi^2(z)$. We test again the null $H_0 : \rho_1 = .. = \rho_z = 0$.

5.3 Correcting/Modeling Autocorrelation

If we have the usual AR(1) autocorrelation type, we could transform our model appropriately. E.g., if our model is:

$$y_t = \beta_0 + \beta_1 x_t + u_t,$$

lagging and multiplying by ρ gives:

$$\rho y_{t-1} = \rho\beta_0 + \rho\beta_1 x_{t-1} + \rho u_{t-1}.$$

Then subtracting the second from the first equation gives:

$$y_t - \rho y_{t-1} = \beta_0(1-\rho) + \beta_1(x_t - \rho x_{t-1}) + (u_t - \rho u_{t-1}).$$

Given AR(1) autocorrelation, i.e., $u_t = \rho u_{t-1} + e_t$, we can estimate the transformed model free from autocorrelation:

$$y_t - \rho y_{t-1} = \beta_0(1-\rho) + \beta_1(x_t - \rho x_{t-1}) + e_t.$$

However, we do not know ρ and need to estimate it.

In practice, we would estimate:

$$y_t = \beta_0(1-\rho) + \rho y_{t-1} + \beta_1 x_t - \beta_1 \rho x_{t-1} + e_t.$$

The revised specification involves a nonlinear restriction: The coefficient of x_{t-1} is minus the product of the coefficients of x_t and y_{t-1}. Thus we need non-linear least squares or (better) MLE.

Autocorrelation is problematic also for inference. In the presence of serial correlation, OLS standard errors overstate statistical significance because there is less independent variation. One can compute serial correlation-robust standard errors after OLS. Serial correlation-robust standard errors are:

$$correctse(\hat{\beta}_j) = \left[\frac{se(\hat{\beta}_j)}{\hat{\sigma}}\right]^2 \sqrt{\hat{v}},$$

where $\hat{v} = \Sigma_{t=1}^{z}\hat{\alpha}_t^2 + 2\Sigma_{h=1}^{g}\left[1 - h/(g+1)\right]\left(\Sigma_{t=h+1}^{g}\hat{\alpha}_t\hat{\alpha}_{t-h}\right)$ and the term $\hat{\alpha}_t = \hat{u}_t\hat{r}_t$, i.e., it is the product of the residuals and the residuals of a regression of x_{tj} on all other explanatory variables. The integer g controls how much serial correlation is allowed. For example, for $g = 1$ we have: $\hat{v} = \Sigma_{t=1}^{z}\hat{\alpha}_t^2 + \Sigma_{h=2}^{z}\hat{\alpha}_t\hat{\alpha}_{t-1}$, for $g = 2$ we have $\hat{v} = \Sigma_{t=1}^{z}\hat{\alpha}_t^2 + 4/3\Sigma_{t=2}^{z}\hat{\alpha}_t\hat{\alpha}_{t-1} + 2/3\Sigma_{t=3}^{z}\hat{\alpha}_t\hat{\alpha}_{t-2}$.

This is the Newly-West correction for autocorrelation. The formulas are also robust to heteroskedasticity; they are therefore called "heteroskedasticity and autocorrelation consistent?"(=HAC) For the integer g, values such as $g = 2$ or $g = 3$ are normally sufficient. Also, serial correlation-robust standard errors are only valid asymptotically; they may be severely biased if the sample size is not large enough. The bias is the higher the more autocorrelation there is; if the series are highly correlated, it might be a good idea to difference them first. Serial correlation-robust errors should be used if there is serial correlation and strict exogeneity fails (e.g. in the presence of lagged dependent variable).

5.4 Fitting a model with Autocorrelation in practice

Test for autocorrelation
-Check the model's specification: possibly add further lags and test again for autocorrelation
-Take first differences of your dependent variable
-Use corrected standard errors.

What is autocorrelation
What OLS properties fail under autocorrelation
How to test for autocorrelation
How to correct for autocorrelation

Readings:
-Slides and notes of lecture 5

-Wooldridge Chapter 12

6 Lecture 6

We talked about:
What is heteroskedasticity
Common forms of heteroskedasticity
How to test for heteroskedasticity
What is Unit root
How to test for Unit root, DF and ADF tests

Readings:
-Slides and notes of lecture 6
Wooldridge Chapter 12

6.1 Heteroskedasticity: What is it, what does it imply, how to model it

Heteroskedastic errors are errors with no constant variance. E.g., when large shocks hit then variance is high and the other way around. We discuss two common models of conditional heteroskedasticity in TS: ARCH (Autoregressive Conditional Heteroskedasticity) and GARC (Generalized Autoregressive Conditional Heteroskedasticity).

6.1.1 ARCH

ARCH(1) is a model that takes int account the strength of the shock in determining the error's variance. Note, that we model conditional variance, conditioned on the information set until time $t-1$, denoted by I_{t-1} (that usually includes the right hand side variables in an ARMA model):

$$\sigma_t^2 = E\{u_t^2|I_{t-1}\} = a_0 + a_1 u_{t-1}^2.$$

Note that variance should be positive, so we need to constrain the set of values that the parameters can take, so $a_0 > 0$ and $a_1 > 0$.
Also, note that having ARCH error, does not mean that the series you model is not covariance stationary. That's because the unconditional variance, $\sigma_t^2 = a_0 + a_1\sigma_{t-1}^2 = a_0 + a_1 a_0 + a_1^2\sigma_{t-2}^2 = ... = a_0(1 + a_1 + a_1^2 +$

$.. + a_1^{t+M-1}) + a_1^{t+M}\sigma_{t-M}^2$. Then, for M going to ∞, $\sigma^2 = \frac{a_0}{1-a_1}$, given that $|a_1| < 1$. Thus, $|a_1| < 1$ would be another constraint we impose when using ARCH estimation.
We can have a more general ARCH(q) representation, as follows:

$$\sigma_t^2 = a_0 + a_1 u_{t-1}^2 + a_2 u_{t-2}^2 + ... + a_q \varepsilon_{t-q}^2.$$

6.1.2 GARCH

An alternative representation of conditional heteroskedasticity is that of GARCH(p,q):

$$\sigma_t^2 = a_0 + a_1 u_{t-1}^2 + a_2 u_{t-2}^2 + ... + a_q u_{t-q}^2 + \phi_1 \sigma_{t-1}^2 + \phi_2 \sigma_{t-2}^2 + ... + \phi_p \sigma_{t-p}^2.$$

Often, a GARCH(1,1) model is good representation of TS data, so include it in the list of models under consideration:

$$\sigma_t^2 = a_0 + a_1 u_{t-1}^2 + \phi_1 \sigma_{t-1}^2,$$

with $a_0, a_1, \phi > 0, 0 < a_1 + \phi < 1$.

6.2 Breusch-Pagan Test for Heteroskedasticity

It is a Lagrange Multiplier (LM) test, that uses the R^2 from the regression of the saved residuals squared, $\hat{u}_t^2$, as $\hat{u}_t^2 = a_0 + a_1 \hat{u}_{t-1}^2 + ... + a_q \hat{u}_{t-q}^2$. For sample size T, the test uses the statistic TR^2, which under the null, H_0: homoskedasticity ($a_1 = .. = a_q = 0$), the test statistic has χ^2 distribution with q degrees of freedom.
Note that this is a test for ARCH. In order to detect the order of heteroskedasticity, you might want to start with a high order ARCH test; if the high order lags are significant, then possibly a GARCH model is appropriate. This is because observing e.g., the GARCH(p,q) mode above, we see that it looks like an ARMA(p,q) model of the variance; and we know that the AR part of a covariance stationary AR model can be written as an infinite order MA model. Thus, our GARCH(p,q) model could be written as an infinite order ARCH model.

7 Lecture 7

We talked about:
Visualized heteroskedasticity and did an example from slides lecture 6

Model Selection in stationary time series
Example on model selection: putting what we learned together on practical applications

Readings:
-Slides and notes of lecture 6
-Slides of lecture 7
Wooldridge Chapter 12

8 Lecture 8: Forecasting

We talked about:
Forecasting definitions
Forecasting with MA(1), MA(2), MA(q)
Computing optimal predictor, forecast error, and forecast error variance
Model Selection and forecasting
Example on model selection and forecasting

Readings:
-Slides and notes of lecture 8

Call the information set available $I_t = \{y_T, y_{T-1}, y_{T-2}, ..., \varepsilon_T, \varepsilon_{T-1}, \varepsilon_{T-2}, , ...\}$ in order to make predictions for the future evolution of the variable of interest y. It is usually true that the optimal predictor of y_{T+H} is it's expected value conditional on the information set available at T, i.e., $E(y_{T+h}|T)$, which we will denote as $y_{T+h|T}$.
Also, we define the forecast error, as the difference between the true value of y and the predicted one, i.e., the forecast error is:

$$e_{T+h|T} = y_{T+h} - y_{T+h|T}. \tag{2}$$

8.1 Predicting using MA(1) processes

Assume an MA(1) process:

$$y_t = a\varepsilon_{t-1} + \varepsilon_t,$$

with ε_t white noise with zero mean and σ^2 variance.[3] [Note: We keep the WN assumption of errors for all subsequent models.]

[3]Note that how good an assumption is the one for the errors, depends on the model selection we made earlier.

We are at period T and want to predict y_{T+1}. First, we write down what y_{T+1} is, given the model of choice (here MA(1)):

$$y_{T+1} = a\varepsilon_T + \varepsilon_{T+1}.$$

Then, we project on the current information set T, computing $E(y_{T+h}|I_T)$:

$$E(y_{T+1}|I_T) = y_{T+1|T} = a\varepsilon_T.$$

Also, $y_{T+2} = a\varepsilon_{T+1} + \varepsilon_{T+2}$, so $y_{T+2|T} = 0$. Thus, for an $MA(1)$ process, $y_{T+h|T} = 0$, for $h > 1$. The statement will be general, and the optimal predictor for an $MA(q)$ process will be zero for $h > q$.

Following equation (2) the forecast errors for the model will be:

$$e_{T+1|T} = y_{T+1} - y_{T+1|T} = \varepsilon_{T+1},$$

$$e_{T+2|T} = y_{T+2} - y_{T+2|T} = a\varepsilon_{T+1} + \varepsilon_{T+2},$$

and so on. So, the forecast errors for an $MA(1)$ model is ε_{T+1} for $h = 1$, and for $h > 1$ it is $e_{T+h|T} = y_{T+h} - 0 = a\varepsilon_{T+h-1} + \varepsilon_{T+h}$ an $MA(1)$ model for $h > 1$ (given that $y_{T+h|T} = 0$).

Also, the forecast error variance for $h = 1$ is $\sigma_1^2 = \sigma^2$, and for $h > 1$ it is $\sigma_h^2 = (1 + a^2)\sigma^2$. This is because for $h = 1$, $e_{T+1|T} = \varepsilon_{T+1}$ and $Var(\varepsilon_{T+1}) = \sigma_1^2 = \sigma^2$. For $h > 1$, $e_{T+h|T} = y_{T+h} - 0 = a\varepsilon_{T+h-1} + \varepsilon_{T+h}$, so $Var(e_{T+h|T}) = a^2(\sigma^2 + 1)$.

8.2 Predicting using MA(2) processes

Assume an MA(2) process:

$$y_t = \varepsilon_t + a_1\varepsilon_{t-1} + a_2\varepsilon_{t-2},$$

with ε_t white noise with zero mean and σ^2 variance.
We are at period T and want to predict y_{T+1}. First, we write down what y_{T+1} is, given the model of choice (here MA(2)):

$$y_{T+1} = \varepsilon_{T+1} + a_1\varepsilon_T + a_2\varepsilon_{T-1}.$$

Then, we project on the current information set T, computing $E(y_{T+h}|I_T)$:

$$E(y_{T+1}|I_T) = y_{T+1|T} = a_1\varepsilon_T + a_2\varepsilon_{T-1}.$$

Also, $y_{T+2} = \varepsilon_{T+2} + a_1\varepsilon_{T+1} + a_2\varepsilon_T$, so $y_{T+2|T} = a_2\varepsilon_T$. Finally, $y_{T+3} = \varepsilon_{T+3} + a_1\varepsilon_{T+2} + a_2\varepsilon_{T+1}$, so $y_{T+3|T} = 0$. Thus, for an $MA(2)$ process, $y_{T+h|T} = 0$, for $h > 2$. The statement will be general, and the optimal predictor for an $MA(q)$ process will be zero for $h > q$.

Following equation (2) the forecast errors for the model will be:

$$e_{T+1|T} = y_{T+1} - y_{T+1|T} = \varepsilon_{T+1},$$

$$e_{T+2|T} = y_{T+2} - y_{T+2|T} = \varepsilon_{T+2} + a_1\varepsilon_{T+1},$$

$$e_{T+3|T} = y_{T+3} - y_{T+3|T} = \varepsilon_{T+3} + a_1\varepsilon_{T+2} + a_2\varepsilon_{T+1},$$

and so on. So, the forecast errors for an $MA(2)$ model is $e_{T+h|T} = \varepsilon_{T+1}$ for $h = 1$, $e_{T+h|T} = \varepsilon_{T+2} + a_1\varepsilon_{T+1}$, an $MA(1)$ process, for $h = 2$, and for $h > 2$ it is $e_{T+h|T} = y_{T+h} - 0 = \varepsilon_{T+3} + a_1\varepsilon_{T+2} + a_2\varepsilon_{T+1}$ an $MA(2)$ process.

Also, the forecast error variance for $h = 1$ is $\sigma_1^2 = \sigma^2$, for $h > 1$ is $\sigma_h^2 = (1 + a_1^2)\sigma^2 < Var(y_t)$, and for $h > 2$ it is $\sigma_h^2 = (1 + a_1^2 + a_2^2)\sigma^2 = Var(y_t)$. Thus, making a prediction for one periods ahead gives us less variance (and thus more accuracy) than making a prediction for two periods ahead; also, making a prediction for two periods ahead gives us less variance (and thus more accuracy) than making a prediction for three periods ahead. However, going further ahead, the prediction accuracy does not further increase.

8.3 Predicting using MA(q) processes

Assume an MA(q) process:

$$y_t = \varepsilon_t + a_1\varepsilon_{t-1} + a_2\varepsilon_{t-2} + a_q\varepsilon_{t-q}$$

with ε_t white noise with zero mean and σ^2 variance.

The optimal predictor for an $MA(q)$ process will be zero for $h > q$.

The forecast errors are:
$e_{T+h|T}$: an $MA(h-1)$ process for $h \leq q$, and
$e_{T+h|T}$: an $MA(q)$ process for $h > q$.

Then, the forecast error variance is:

$\sigma_1^2 = \sigma^2$

$\sigma_h^2 < Var(y_t)$ for $h \leq q$

$\sigma_h^2 = Var(y_t)$ for all $h > q$.

8.4 Predicting using AR(1) processes: HW

8.5 Prediction accuracy

We will use the forecast errors variance to evaluate prediction accuracy. Specifically we can:

-Compute error bands: $y_{T+h|T} - 1.96\sigma_h$ and $y_{T+h|T} + 1.96\sigma_h$.

-Compute the Root Mean Squared Error (RMSE):

$$RMSE = \sqrt{\frac{1}{H}\sum_{h=1}^{H} e_{T+h}^2},$$

where H is the maximum period of forecast. The lower value of the statistic, the better the prediction accuracy.

–Compute the Diebold-Mariano test:

$$d = \frac{1}{H}\sum_{h=1}^{H}(e_{1,T+h}^2 - e_{2,T+h}^2),$$

where we compare two competing models, $1, 2$.