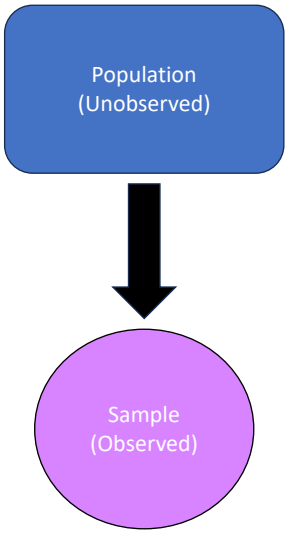
ECOM90024 – FORECASTING IN ECONOMICS & BUSINESS

LECTURE 11: REVIEW LECTURE

1

THE FORECASTING MINDSET

- All forecasting models need to be trained (or estimated) using data.
- A good forecasting model should reflect the features of the underlying data generating process of the variable/phenomena of interest. In the language of statistics, this is the population of interest.
- The data that we train or estimate forecasting models on is a sample drawn from the population of interest.
- Does a predictive model that fits the sample well generate good predictions?
- Does prediction = truth?



2

MODEL SELECTION FOR FORECASTING

- In this course, we have primarily evaluated models based on how well they fit the data that they have been trained on . This is because we want to learn how to correctly identify the various forms of dependency that may be present in the data.
- Statistics such as the AIC, BIC and R^2 are measures of fit.
- However, since we can only ever observe a sample drawn from the underlying population, we can never be sure which features of the sample are idiosyncratic (i.e., unique to the sample) or truly representative of the data generating process.
- When we try to build a model that tries to account for all the features that are present in the data, we run the risk of ***overfitting***. Overfitted models will produce poor forecasts since they will incorporate the idiosyncratic features of the training data.
- Measures of fit are not the be all and end all of model building! Especially for the purpose of forecasting!

3

MODEL SELECTION FOR FORECASTING

- To evaluate the ***forecasting performance*** of models, we need to use measures such as:
- Root mean square error (RMSE)

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} e_i^2} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}$$

- Mean absolute error (MAE)

$$MAE = \frac{1}{N}\sum_{i=1}^{N} |e_i| = \frac{1}{N}\sum_{i=1}^{N} |y_i - \hat{y}_i|$$

- https://www.r-bloggers.com/2021/07/how-to-calculate-root-mean-square-error-rmse-in-r/

4

MODEL SELECTION FOR FORECASTING

- Note that these measures are statistical metrics. They are simply measures of distances (or scaled/relative distances). In optimization and decision theory, these are known as ***loss functions***.
- In many cases, these metrics will favor different forecasting models.
- The impact of prediction errors on the business may not be reflected in certain metrics.
- There may be some business contexts which may call for a ***user-specific loss function***.

5

MODEL SELECTION FOR FORECASTING

- For instance, you may be faced with a business context in which an overprediction might have a greater impact on the business compared to an underprediction. So, you may need to incorporate that asymmetry into your measure.
- For instance, you might define the following loss function:

$$\sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i-\hat{y}_i)^2+\gamma(y_i-\hat{y}_i)^2 I(y_i-\hat{y}_i<0)}$$

- **Example:** Predicting demand has asymmetric implications on inventory management.
- **Example:** Predicting an asset price for the purpose of assessing loan collateral.

6

MODEL SELECTION FOR FORECASTING

- All models are wrong, but some are useful! – George Box (1987)
- When building forecasting and prediction models, there is always an element of human judgement involved. We need to be clear about the assumptions we are making and what we are basing our decisions on.
- You need to be able to convince people that your reasoning and assumptions are reasonable.

7

REVIEW LECTURE

- We began by conceptualizing all time series as comprising of a set of fluctuations occurring at different time scales:

$$Y_t = T_t + S_t + C_t + \varepsilon_t$$

- Where,

$T_t =$ Trend component (long horizon fluctuations)

$S_t =$ Seasonal component (regular fluctuations that occur at fixed intervals within a year)

$C_t =$ Cyclical component (fluctuations that occur an non-fixed intervals over the course of several years)

8

CLASSICAL DECOMPOSITION

- The decompose function computes the additive decomposition via the following steps:

1. Setting the seasonal period m equal to the frequency in which the data is observed (e.g., $m = 4$ for quarterly data, $m = 12$ for monthly data, etc.)
2. Compute the trend-cycle component $\hat{T}_t + \hat{C}_t$ using an $MA(m)$ if m is odd and $\overline{MA(m)}$ if m is even.
3. Calculate the detrended series $y_t - \hat{T}_t - \hat{C}_t$
4. Calculate the seasonal component for each season as the average of the detrended values for that season (e.g., the average of all the January observations etc.) These seasonal component values are then adjusted to ensure that they add to zero. The seasonal component is obtained by stringing together these monthly values, and then replicating the sequence for each year of data. This gives $\hat{S}_t$
5. The remainder component is calculated by subtracting the estimated seasonal and trend-cycle components: $\hat{R}_t = y_t - \hat{T}_t - \hat{C}_t - \hat{S}_t$

9

CLASSICAL DECOMPOSITION

- A multiplicative decomposition is specified as:

$$Y_t = T_t \times S_t \times C_t \times \varepsilon_t$$

- The decompose function computes the multiplicative decomposition via the following steps:

1. Setting the seasonal period m equal to the frequency in which the data is observed (e.g., $m = 4$ for quarterly data, $m = 12$ for monthly data, etc.)
2. Compute the trend-cycle component $\hat{T}_t \times \hat{C}_t$ using an $MA(m)$ if m is odd and $\overline{MA(m)}$ if m is even.
3. Calculate the detrended series $y_t / (\hat{T}_t \hat{C}_t)$

10

CLASSICAL DECOMPOSITION

- Calculate the seasonal component for each season as the average of the detrended values for that season (e.g., the average of all the January observations etc.) These seasonal component values are then adjusted to ensure that they add to m. The seasonal component is obtained by stringing together these monthly values, and then replicating the sequence for each year of data. This gives $\hat{S}_t$

- The remainder component is calculated by dividing out the estimated seasonal and trend-cycle components

$$\hat{R}_t = \frac{y_t}{\hat{T}_t \times \hat{S}_t \times \hat{C}_t}$$

11

SIMPLE EXPONENTIAL SMOOTHING

- We can compute a trend as simple exponentially smoothed series in the following way:

- Start with the initial values of the time series y_0 and y_1 then, we set $l_0 = y_0$ and compute:

$$l_1 = \alpha y_1 + (1-\alpha) l_0$$

- Then, proceed to compute:

$$l_2 = \alpha y_2 + (1-\alpha) l_1$$

- And proceed iteratively until

$$l_T = \alpha y_T + (1-\alpha) l_{T-1}$$

Through recursive substitution we can see that this becomes the original geometric series that we presented initially a couple of slides ago!

12

EXPONENTIAL SMOOTHING: HOLT'S LINEAR TREND

- In order to accommodate a time trend, we will have to augment the updating equations to incorporate a trend component b_t:

$$\text{Level Equation:} \quad l_t = \alpha y_t + (1-\alpha)(l_{t-1} + b_{t-1})$$

$$\text{Trend Equation:} \quad b_t = \beta(l_t - l_{t-1}) + (1-\beta)b_{t-1}$$

$$Forecasting\ Equation: \quad \hat{y}_{t+h|t} = l_t + hb_t$$

- Where again, h denotes the forecast horizon.
- b_t is a weighted average of the change in the level $l_t - l_{t-1}$ and the estimated trend for time $t-1$

13

EXPONENTIAL SMOOTHING: EXPONENTIAL TREND

- We can also specify the model to have a multiplicative form if we want to model an exponential trend:

$$\text{Level Equation:} \quad l_t = \alpha y_t + (1-\alpha)(l_{t-1} b_{t-1})$$

$$\text{Trend Equation:} \quad b_t = \beta \frac{l_t}{l_{t-1}} + (1-\beta)b_{t-1}$$

$$Forecasting\ Equation: \quad \hat{y}_{t+h|t} = l_t b_t^h$$

- Where b_t is now the growth rate of the trend.

14

EXPONENTIAL SMOOTHING: DAMPED TREND

- The additive damped trend model is given by:

$$\text{Level Equation:} \quad l_t = \alpha y_t + (1-\alpha)(l_{t-1} + \phi b_{t-1})$$

$$\text{Trend Equation:} \quad b_t = \beta(l_t - l_{t-1}) + (1-\beta)\phi b_{t-1}$$

$$Forecasting\ Equation: \quad \hat{y}_{t+h|t} = l_t + \left(\phi + \phi^2 + \cdots + \phi^h\right) b_t$$

- Where $0 < \phi < 1$

15

EXPONENTIAL SMOOTHING: DAMPED TREND

- The multiplicative damped trend is given by:

$$\text{Level Equation:} \quad l_t = \alpha y_t + (1-\alpha)(l_{t-1} b_{t-1}^{\phi})$$

$$\text{Trend Equation:} \quad b_t = \beta \frac{l_t}{l_{t-1}} + (1-\beta) b_{t-1}^{\phi}$$

$$Forecasting\ Equation: \quad \hat{y}_{t+h|t} = l_t b_t^{\left(\phi+\phi^2+\cdots+\phi^h\right)}$$

16

DETERMINISTIC TREND ESTIMATION

- The trend component of a time series can also be modeled as a deterministic function of time. Letting $t = 1,2, \dots T$ we have that

$$T_t = \delta_0 + \delta_1 t \text{ (linear trend)}$$

$$T_t = \delta_0 + \delta_1 t + \delta_2 t^2 + \cdots + \delta_k t^k \text{ (polynomial trend)}$$

$$T_t = \delta_0 e^{\delta_1 t} \text{ (exponential trend)}$$

- How we choose to model the deterministic trend will depend on the underlying data generating process.
- As a general rule, simplicity should be preferred to avoid overfitting. A rich or complex deterministic trend specification needs to be well justified.

17

DETERMINISTIC SEASONALITY

- The seasonal component of a time series can be modeled through the use of dummy variables. For instance, if the data is observed in quarterly intervals, we could define

$$D_1 = \{1,0,0,0,1,0,0,0,1,0,0,0, \dots\}$$
$$D_2 = \{0,1,0,0,0,1,0,0,0,1,0,0, \dots\}$$
$$D_3 = \{0,0,1,0,0,0,1,0,0,0,1,0, \dots\}$$
$$D_4 = \{0,0,0,1,0,0,0,1,0,0,0,1, \dots\}$$

- So that,

$$S_t = \xi_1 D_1 + \xi_2 D_2 + \xi_3 D_3 + \xi_4 D_4$$

- How we choose to define our seasonality dummies will ultimately depend on the manner in which the seasonality manifests. For instance, if the data generating process for quarterly data has a seasonality component that occurs at fixed half year intervals, we could define,

$$H_1 = D_1 + D_2 = \{1,1,0,0,1,1,0,0,1,1,0,0, \dots\}$$
$$H_2 = D_3 + D_4 = \{0,0,1,1,0,0,1,1,0,0,1,1, \dots\}$$

18

OLS ESTIMATION OF DETERMINISIC COMPONENTS

- The deterministic trend and seasonality components can be easily estimated using OLS. This simply involves specifying a multiple regression model such as

$$Y_t = \delta_1 t + \xi_1 D_1 + \xi_2 D_2 + \xi_3 D_3 + \xi_4 D_4 + \varepsilon_t$$

$$\varepsilon_t \sim i.i.d.(0, \sigma^2)$$

- When specified in this way, the coefficients ξ_i, $i = 1,2,3,4$ have the interpretation as the season specific means, otherwise known as ***seasonal factors***.
- We can test for the presence of seasonality by testing the joint hypothesis,

$$H_0: \xi_1 = \xi_2 = \xi_3 = \xi_4$$

- It is important to note that when the full set of seasonal dummies are included in the regression, the constant term is omitted. This is so as to avoid the problem of ***multicollinearity***.

19

MULTICOLLINEARITY

- Multicollinearity in a multiple regression occurs when two or more explanatory variables (i.e. regressors) are highly correlated and results in imprecise estimates of the regression coefficients.
- This will typically manifest as large standard errors which may lead us to incorrectly fail to reject the null hypothesis in a significance test (i.e. a Type II error). Doing so may result in incorrect modeling decisions!
- To illustrate, consider the following regression,

$$Y_i = \beta_1 X_i + \beta_2 Z_i + \varepsilon_i$$

- Suppose that both X_i and Z_i are each individually significantly correlated with Y_i and also highly correlated with one another. This will produce large standard errors for the estimates of β_1 and β_2 which may cause us to conclude that both are not statistically different from zero which would in turn lead us to drop both variables from the regression.

20

MULTICOLLINEARITY

- To see how multicollinearity can manifest in a seasonal dummy specification, let's consider the following model for quarterly data,

$$Y_t = \delta_0 + \delta_1 t + \xi_1 D_1 + \xi_2 D_2 + \xi_3 D_3 + \xi_4 D_4 + \varepsilon_t$$

$$\varepsilon_t \sim i.i.d.(0, \sigma^2)$$

- The constant term δ_0 can be thought of as the slope coefficient for a variable that is simply a sequence of ones, $\iota = 1,1,1, \dots, 1$
- We can also recognize that the set of dummy variables are related in the following way,

$$D_1 = \iota - D_2 - D_3 - D_4$$

- Since, any single seasonal dummy can be expressed as an exact linear function of ι and the the other seasonal dummies, we have perfect multicollinearity!

21

MULTICOLLINEARITY

- The problem of multicollinearity in this setting can be solved by dropping one of the seasonal dummies,

$$Y_t = \delta_0 + \delta_1 t + \xi_2 D_2 + \xi_3 D_3 + \xi_4 D_4 + \varepsilon_t$$

- In such a specification, the constant δ_0 now represents the mean associated with the season represented by D_1 and the coefficients ξ_2, ξ_3 and ξ_4 represent the seasonal deviations in the mean in the seasons represented by D_2, D_3 and D_4 respectively.

- Therefore, when specified in this way, we can test for the presence of seasonality by testing the joint hypothesis,

$$H_0: \xi_2 = \xi_3 = \xi_4 = 0$$

22

JOINT TEST OF SIGNIFICANCE

- To test a joint hypothesis we have to estimate the unrestricted model and the restricted model. For instance, if we wish to test

$$H_0: \xi_2 = \xi_3 = \xi_4 = 0$$

Then we have that

$$Y_t = \delta_0 + \delta_1 t + \xi_2 D_2 + \xi_3 D_3 + \xi_4 D_4 + \varepsilon_t \text{ (unrestricted model)}$$

$$Y_t = \delta_0 + \delta_1 t + \varepsilon_t \text{ (restricted model)}$$

- For a multiple regression with a sample size of T and K explanatory variables to perform a joint test with J restrictions, the test statistic is computed as

$$F_{test} = \frac{\left(\frac{SSR_U - SSR_R}{J}\right)}{\frac{SSR_U}{(T-K-1)}} \sim F_{(J,(T-K-1))}$$

23

CYCLICAL DYNAMICS

- Having accounted for deterministic trend and seasonal components, we can then proceed to account for the cyclical fluctuations which can be thought of broadly as the stable, mean-reverting dynamics that are not captured by trends or seasonal factors.
- We characterize cyclical dynamics as covariance-stationary time series.
- A covariance-stationary time series must have the following stochastic properties:

1. A constant mean: $E[Y_t] = \mu$
2. A finite variance: $E[(Y_t - \mu)^2] < \infty$
3. The covariance between any two terms in the series depends only on the relative position of the two terms.

$$E\left[(Y_t - \mu)(Y_{t-j} - \mu)\right] = E\left[(Y_{t+k} - \mu)(Y_{t+k-j} - \mu)\right] = \gamma_j$$

$$\forall j, k$$

24

DEPENDENCE STRUCTURE OF A TIME SERIES

- The better we are able to account for the dependence structure in the time series, the more precise our forecasts will be. To see this, let's consider the following two processes,

Process 1: $Y_t = \mu_t + \upsilon_t$

Process 2: $Y_t = \mu_t + \varepsilon_t$

$\varepsilon_t = \phi\varepsilon_{t-1} + \upsilon_t$

- Where $|\phi| < 1 \; and \; \upsilon_t \sim i.i.d(0, \sigma^2)$
- We can see clearly that the variance of the serially correlated innovations ε_t will be greater than the white noise innovations υ_t. Since we know from our analysis of covariance stationary processes that,

$$\sigma_\varepsilon^2 = \frac{\sigma^2}{1-\phi^2}$$

- This means that the forecast intervals generated from the second process will be wider than the forecast intervals generated from the first process.

25

DEPENDENCE STRUCTURE OF A TIME SERIES

- The dependence structure of a covariance stationary time series process is succinctly summarized by its autocorrelation and partial autocorrelation functions.
- Let Y_t be a covariance stationary time series. It's j-th autocovariance is given by,

$$\gamma_j = E\big[(Y_t - E[Y_t])(Y_{t-j} - E[Y_{t-j}])\big]$$

- It follows that it's j-th autocorrelation is given by

$$\rho_j = \frac{\gamma_j}{\gamma_0}$$

- We can see from this that a time series that is serially uncorrelated (i.e. white noise) will have $\rho_j = 0$ for all $j \geq 1$.

26

DEPENDENCE STRUCTURE OF A TIME SERIES

- Given a set of observations $\{y_1, y_2, \ldots, y_T\}$, the sample j-th autocovariance can be computed as

$$\hat{\gamma}_j = \frac{1}{T-j-1}\sum_{t=j+1}^{T}(y_t - y)(y_{t-j} - y)$$

- While the sample j-th autocorrelation can be computed as

$$\hat{\rho}_j = \frac{\hat{\gamma}_j}{\hat{\gamma}_0}$$

- The sample autocovariance and autocorrelations are estimates of the population autocovariance and autocorrelations.
- Given a sample size of T, the sample autocorrelations from a white noise process will be distributed as

$$\hat{\rho}_j \sim N\left(0, \frac{1}{T}\right)$$

27

PORTMANTEAU TESTS

- The Box-Pierce and Ljung-Box statistics allow us to test the null hypothesis that the data we observe are generated from a white noise process.

$$Q_{BP} = T\sum_{\tau=1}^{m}\hat{\rho}_j^2 \sim \chi_m^2 \quad \text{(Box-Pierce Q-Statistic)}$$

$$Q_{LB} = T(T+2)\sum_{\tau=1}^{m}\left(\frac{1}{T-\tau}\right)\hat{\rho}^2(\tau) \sim \chi_m^2 \text{ (Ljung-Box Q-Statistic)}$$

- Intuitively, if the data we observe are indeed generated from a white noise process, the sample autocorrelations will be small (i.e. close to zero) and thus these test statistics will be small relative to the critical value and will therefore have large p-values.
- Thus, if we perform these tests with respect to the residuals of our time series model and reject the null, it means that we have not accounted for all of the serial dependence and our model can be improved!

28

PARTIAL AUTOCORRELATIONS

- Given a time series Y_t, the j-th partial autocorrelation is the autocorrelation between Y_t and Y_{t-j} ***that has not been accounted for*** by the autocorrelations between Y_t and Y_{t-k} for $k = 1, \dots, j-1$.
- To illustrate, let's consider an AR(1) model,

$$Y_t = \phi Y_{t-1} + \varepsilon_t$$

- The j-th autocorrelation is given by

$$\rho_j = \phi^j$$

- All of the autocorrelations for $j \geq 2$ are due to the first autocorrelation. If we account for the first autocorrelation ϕ, there are no other higher order dynamics in play!

29

PARTIAL AUTOCORRELATIONS

- Consider an AR(2)

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t$$

- It can be shown that the first two autocorrelations are given by

$$\rho_1 = \frac{\phi_1}{1-\phi_2}$$

$$\rho_2 = \phi_1 \rho_1 + \phi_2$$

- And the subsequent autocorrelations are given by

$$\rho_j = \phi_1 \rho_{j-1} + \phi_2 \rho_{j-2}.$$

- This means that all the higher order autocorrelations depend only on the first two autocorrelations. Therefore, if we account for the dependence between Y_t and Y_{t-1} and Y_t and Y_{t-2}, there are no other higher order dynamics at play!

30

PARTIAL AUTOCORRELATIONS

- In order to understand how we can compute partial autocorrelations, let's consider the following multiple regression:

$$Z_i = \beta_1 X_{1,i} + \beta_2 X_{2,i} + \varepsilon_i$$

- The coefficient β_1 is interpreted as the average change in Z per unit change in X_1 with X_2 held constant. That is, it represents the partial effect of X_1 on Z or the unique contribution of X_1 towards predicting Z. Applying this to a time series model, we can see that the first partial autocorrelation is computed as π_1 in the regression

$$Y_t = \pi_1 Y_{t-1} + \varepsilon_t$$

- Then the second partial autocorrelation is computed as π_2 in the regression

$$Y_t = \pi_1 Y_{t-1} + \pi_2 Y_{t-2}\, \varepsilon_t$$

- The j-th partial autocorrelation as the j-th autoregressive coefficient π_j in j-th order autoregression,

$$Y_t = \pi_1 Y_{t-1} + \pi_2 Y_{t-2} + \cdots + \pi_j Y_{t-j} + \varepsilon_t$$

31

AUTOREGRESSIVE SERIES

- Therefore, when we have a set of observations that have been generated from an AR(p)

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \varepsilon_t$$

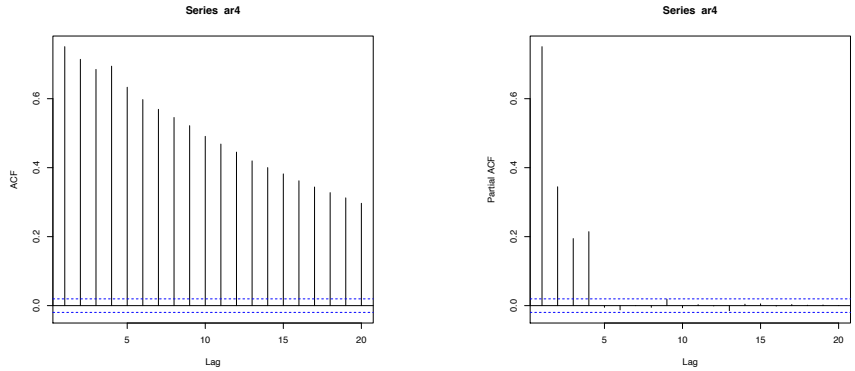
- Estimating a sequence of autoregressions of order $1,2, \ldots, p$ will yield significant coefficient estimates. Higher order autoregressions however will have coefficient estimates that will not be statistically different from zero for coefficients multiplying terms of order greater than p.

- Therefore, AR(p) processes will have partial autocorrelation functions and sample partial autocorrelation functions that cut off at lags greater than p.

32

AUTOREGRESSIVE SERIES

- Consider the following sample ACF and PACF from data generated from an AR(4),



33

MA SERIES

- In the case of a MA model, we showed previously that an invertible MA(1) may be represented as an infinite order AR,

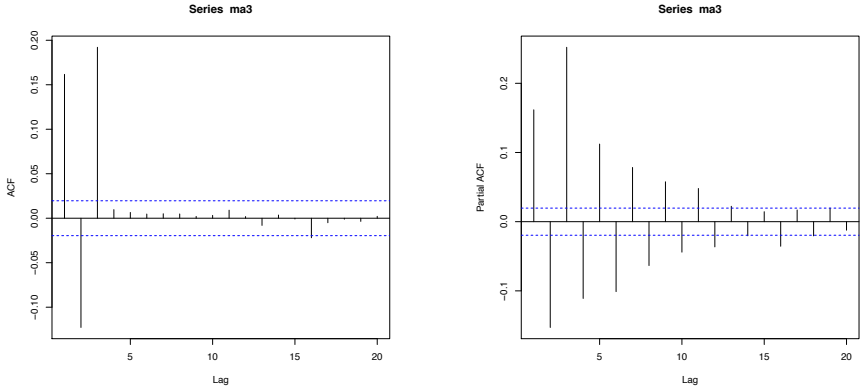
$$Y_t = \varepsilon_t + \theta Y_{t-1} - \theta^2 Y_{t-2} + \theta^3 Y_{t-3} - \cdots$$

- Then the partial autocorrelation function for an MA model will have partial autocorrelations that gradually decay as $j \to \infty$

34

MA SERIES

- Here we have the sample ACF and PACF computed from data generated from an MA(3)



35

YULE-WALKER EQUATIONS

- It can be shown that the autocorrelation function for an AR(p) has a recursive structure. When expressed in this way, the autocorrelation functions are known as the Yule-Walker equations.

$$\rho_j = \phi_1\rho_{j-1} + \phi_2\rho_{j-2} + \cdots + \phi_p\rho_{j-p}$$

- These equations tells us that we can can express the p autoregressive coefficients in terms of the first p autocorrelations.
- Hence we can compute estimates for the autoregressive coefficients either via OLS or solving the Yule-Walker equations for the sample autocorrelations.

36

COVARIANCE STATIONARITY

- An AR(p) expressed in terms of lag operators is given by

$$(1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p) Y_t = \varepsilon_t$$

- It is covariance stationary if the roots of

$$1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p = 0$$

- Lie ***outside the unit circle*** or equivalently if the roots of:

  - $\lambda^p - \phi_1 \lambda^{p-1} - \phi_2 \lambda^{p-2} - \cdots - \phi_{p-1} \lambda - \phi_p = 0$

- Lie ***inside the unit circle.***

37

INVERTIBILITY

- An MA(q) expressed in terms of lag operators is given by

$$Y_t = (1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q) \varepsilon_t$$

- All moving average models are covariance stationary. However, not all moving average models are invertible. A moving average model is invertible if the roots of:

$$1 + \theta_1 z + \theta_2 z^2 + \cdots + \theta_q z^q = 0$$

- Lie ***outside the unit circle*** or equivalently if the roots of:

$$\lambda^q + \theta_1 \lambda^{q-1} + \theta_2 \lambda^{q-2} + \cdots + \theta_{q-1} \lambda + \theta_q = 0$$

- Lie ***inside the unit circle.***

38

WOLD REPRESENTATION THEOREM

- Let Y_t be ***ANY*** zero mean, covariance-stationary process. Wold's representation theorem states that:

$$Y_t = \Psi(L)\varepsilon_t = \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i}$$

$$\varepsilon_t \sim_{iid} (0, \sigma^2)$$

$$\psi_0 = 1$$

$$\sum_{i=0}^{\infty} \psi_i^2 < \infty$$

- That is to say, any zero mean, covariance stationary process can be expressed as an infinite order moving average. In other words, the correct model (or data generating process) for any covariance stationary series is ***some infinite distributed lag of white noise.***

39

THE GENERAL LINEAR MODEL

- If the invertibility and stationarity conditions are satisfied, we can write an ARMA(p, q) as:

$$\frac{\Theta(L)}{\Phi(L)}\varepsilon_t$$

- Recall that the general linear model is written as:

$$Y_t = \Psi(L)\varepsilon_t = \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i}$$

- Therefore the ARMA(p, q) model approximates an infinite distributed lag or current and past innovations using the ratio of two finite-order lag polynomials:

$$\Psi(L) \approx \frac{\Theta(L)}{\Phi(L)}$$

40

TIME VARYING CONDITIONAL VOLATILITY

- Once we have an appropriate ARMA specification for the conditional mean (i.e. we no longer have any dependence in the residuals). We can proceed to look at whether there exists any dependence in the squared residuals (i.e. volatility clustering!).
- If so, we can specify a GARCH type of model. Recall that a standard GARCH(1,1) process is specified as:

$$\varepsilon_t = \sigma_t \upsilon_t$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

$$\upsilon_t \sim_{i.i.d.} N(0,1)$$

41

A GENERAL APPROACH TO ESTIMATING A TIME SERIES MODEL

1. Plot the time series data that you have collected.
2. Specify and estimate using OLS any deterministic components (i.e. trend and seasonality) that you believe to exist, along with the mean of the series.
3. Generate the residuals from your deterministic regression. This will be the demeaned, detrended and seasonally adjusted series.
4. Perform an ADF test on your demeaned, detrended and seasonally adjusted series to determine the order of integration.
5. If the null-hypothesis of a unit root is not rejected, take the first difference of the data and perform the unit root test again. Once you reject the null-hypothesis of a unit root, proceed to the next step with your differenced data. The number of times you have difference the data to achieve stationarity is your order of integration d.

42

A GENERAL APPROACH TO ESTIMATING A TIME SERIES MODEL

6. Generate the sample ACF and PACF of the demeaned, detrended and seasonally adjusted series. These will give you a sense of the dependence structure of the time series.

7. Estimate a range of ARIMA(p, d, q) and choose your preferred model using the AIC and BIC.

8. Extract residuals from Step 7 and test for ARCH effects. If ARCH effects are detected, specify an ARCH or GARCH model for the innovations,

$$\varepsilon_t = \sigma_t \upsilon_t$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

$$\upsilon_t \sim_{i.i.d.} N(0,1)$$

9. Check that the ARCH/GARCH specification is adequate by generating the squared standardized residuals from the ARCH/GARCH and testing whether any dependence remains (i.e. are there any ARCH effects that are unaccounted for).

43

APPROACH TO ESTIMATING & FORECASTING AN ARMA MODEL WITH ARCH EFFECTS

4. If the ARCH/GARCH model is adequate, proceed to generate the h-step ahead point forecasts from the ARIMA(p, d, q) model.

5. To generate the h-step ahead forecast intervals, first derive the expression for the h-step ahead forecast error and associated conditional variance. This will be a function of the ARMA coefficients and the conditional variances from the ARCH/GARCH model, $\boldsymbol{\theta} = \{c, \phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q, \sigma_{t+1|t}^2, \dots \sigma_{t+h|t}^2\}$

6. Using your estimates of the ARMA coefficients and the forecasts of the conditional variances, compute the 95% forecast interval as

$$E\left[Y_{t+h} \middle| \Omega_t\right] \pm 1.96\sigma_h$$

44

FINAL EXAM – HURDLE REQUIREMENT

- 15 MINS READING TIME
- 120 MINUTES WRITING TIME
- 4 QUESTIONS

- MIX OF NUMERICAL AND ANALYTICAL QUESTIONS

- YOU MAY BE ASKED TO INTERPRET OR USE *R* OUTPUT (i.e., CHARTS, ESTIMATION RESULTS etc.)

- YOU MAY BE ASKED TO PERFORM NUMERICAL CALCULATIONS USING A SMALL NUMBER OF DATA POINTS. USEFUL TO PRACTICE DOING THIS BEFORE THE EXAM. IT'S NOT JUST ABOUT WHAT YOU KNOW, IT'S ABOUT WHAT YOU CAN DO!

- GO THROUGH THE TUTORIAL QUESTIONS AND ASSIGNMENTS CAREFULLY.

45