**AN6902 Time Series Analysis Date: 11/12/2022**

|  |  |
| --- | --- |
| **Week 1: Linear Regression Models** | |
| * **Gauss Markov Theorem:** OLS method to estimate beta parameters providing **Unbiasedness, Efficiency, Consistency, Estimation** * Best Linear Unbiased Estimator (BLUE): lowest sampling variance and which is also unbiased. * **Structured Time Series (STM) = Multivariate Time Series** * **Pure Time Series (PTM) = Univariate Time Series** * AR(p) = using Y only   historical Yt numbers as input   * MA(q)  historical error terms (i.e. residues) to interpret Y * **Time Series:** stock index v.s. macroeconomic fundamentals * **Cross-Sectional**: country’s GDP v.s. probability default on debt * **Panel:** Time Series + Cross-sectional * **Jarque-Bera test:** Goodness-of-fit test determining if sample data have **skewness and kurtosis matching normal distribution** | **EViews**  **Stationary**: constant mean, variance, autocovariance **s =** lag **c =** constant  **PACF (MA)** - invertibility **ACF (AR)** - stationarity |
| **Week 2 and 3: Univariate Time Series Modelling and Forecasting** | |
| 1. Autoregressive Model (AR) yt = μ + φ1 yt−1 + φ2 yt−2 +···+ φp yt−p + ut 2. Wold’s Decomposition Theorem (MA)   Any stationary series = deterministic + stochastic (i.e., random probability distribution)  *Definitions*   * **Strictly stationary:** joint distribution unchanged * **Weakly stationary:** covariance stationary * **Covariance:** pair of random variables defined on the same probability space * **Autocovariance:** pair of values in a discrete-time stochastic process * **Autocovariances:** covariances of Ys with its own previous values * **Autocorrelation:** correlations X | **EViews**   * **Residues Analysis:** if series is random = good model * H0: white noise (i.e., stationary) * **Rejecting null hypothesis (P-value < 0.05) = non-stationary**   ***Detecting Serial Correlation:*** *If a variable is serially correlated, it is not random.* If there is no serial correlation (i.e. random):   1. ACF and PACF at all lags should be near zero 2. Ljung-Box Q-statistics should be insignificant   **Building ARMA models: Box-Jenkins Approach**   1. ***Identification (Information Criteria)***  * Akaike Information Criteria (AIC) * Bayesian Information Criteria (SBIC) * Hannah Quinn Criterion (HQIC)   (Less Strict Penalty) **AIC > HQIC > SBIC** (Strict Penalty)   1. ***Estimation*** 2. ***Diagnostic Checking***  * Overfitting/Residual Diagnostics * Examining if residuals are free from autocorrelation |
| **Week 4: Modelling Long-Run Relationships in Finance** | |
| * **Durbin-Watson test:** autocorrelation of 1st order * **Breusch-Godfrey test**: autocorrelation of higher orders * **Random walks:** independent and non-stationary * Non-stationary data can lead to spurious regressions * I(0) = stationary * I(1) = 1 unit root, requires 1 differencing to induce stationarity * I(2) = 2 unit root, requires 2 differencing to induce stationarity | **EViews**   * **Dickey-Fuller Test** * H0: series contains unit root * H1: series does not contain unit root (i.e., stationary) * **Philips-Perron (PP) Test** * Considers unit root in presence of **known structural breaks** (e.g. changes in monetary policy/removal of exchange rate controls) * **KPSS Test** * Test for stationarity * H0: series is stationary * H1: series is not stationary   **Graphical user interface, text, application  Description automatically generated**  **Best outcomes:**   1. Reject H0 and Do not reject H0 (ADF/PP and KPSS) 2. Do not reject H0 and Reject H0 (ADF/PP and KPSS)   ***Unit Root Test (ADF = left-tailed test, KPSS = right-tailed test)*** *View > Unit Root Test > Standard Unit Root Test > ADF* Level = Yt 1st difference = Delta Yt 2nd difference = Delta square Yt |
| **Week 5 and 6: Modelling Volatility and Correlation** | |
| **autoregressive conditional heteroscedasticity (ARCH)**:  a time series model for volatilities.  Limitations of ARCH:  1. Value of the no. of lags of squared residual has no clear best approach 2. Value of the no. of lags of squared error might be very large, resulting in large conditional variance that is not parsimonious  **generalised autoregressive conditional heteroscedasticity (GARCH)**:  a common specification of dynamic model for volatility.  GARCH (1, 1) = ARCH (infinity) Limitations of GARCH:  1. GARCH enforces a symmetric response of volatility to positive/negative shocks  ***Alternative GARCH:*** 1. (Threshold) T-ARCH (GJR) Model – to overcome leverage effects 2. EGARCH Model – to overcome negativity constraints 3. GARCH-in-mean – to incorporate “higher risk, higher return” concept |  |
| **EViews** | |
| **1. Check Stationarity: Unit Root Test** If I(1) i.e. not stationary, use differencing to make the model stationary If I(0) all good, no need to use differencing  **2. Residues Analysis: Check if proposed model is a good approximation -  if residues is white noise**  **3. Check cointegration: Engle-Granger cointegration test** (working out residues) Part 1 > All variables are I(1) Part 2 > Use the step 1 residuals as one variable in the error correction model  **4. Check variance of the residuals** If variance is unchanged, just use simple time series If variance is different, use ARCH/GARCH | ***Questions for Exam:***  What is ARMA Unit Root Test What is ARCH Extension of ARCH What is GARCH ARCH-GARCH Condition Model Why do we use these models Why cointegration is important Purpose of ECM |

**EViews Summary/Writing**

**1. Generate Series > r = 100\*d(logprice)**The function d() takes the first difference of a series, and the first difference of a log is approximately the percentage change.

**2. Estimate Equation > r c ar(1)  
Residual Diagnostics > Correlogram – Q Statistics > Lag 36 (Default)**

**Prob (Correlogram)**  
**Null hypothesis (h0) = not autocorrelated (white noise)**  
**Prob. > 0.05, do not reject h0 (Good)**  
Errors are independent and are not autocorrelated, one of the linear regression assumptions   
If the data fits well, the residual gets smaller.

**Null hypothesis (h0) = white noise**  **T-statistic < 0.05 = reject h0 (Bad)**  
Thus, the model is not good as the error term is NOT white noise.   
White noise = not predictable, just random variation. If residuals look like white noise, you did everything to make the model as good as possible.

Probability < 0.05, reject null hypothesis of white noise

***3. Unit Root Test (ADF Test)***e.g. Augmented Dickey-Fuller test statistic Prob\* = 0.001  
Null hypothesis (h0) = stationary  
**Probability < 0.05, reject null hypothesis**Since Prob = 0.001, strongly reject null hypothesis, cannot be white noise.   
The data is non-stationary and require differencing.

**4. Use Engle-Granger test for residuals (H0 = no cointegration)  
Estimate Equation > d(ls) c d(lf) res1(-1)**

**5. ARCH (H0 = no Arch effect)  
View > Residual Diagnostics > Heteroskedasticity   
T-statistic < 0.05 = reject (Good) = no Arch effect  
  
View > Correlogram of Standardised Residuals (H0 = ARCH)  
Probability (Correlogram) > 0.05, do not reject (Good) = not autocorrelated (i.e. all squared residuals have no linear relations and ARCH effect is insignificant) = no ARCH**

**Probability < 0.05 = ARCH**If significant ARCH effects, model proposed is not good. Will need improvement with ARCH/GARCH.  
Square residuals correlogram = ARCH effect test

**Essay**

**ARMA Model**

The ARMA model includes the combination of the autoregressive (AR) and moving average (MA) components. With ARMA, there is both a geometrically decaying ACF and PACF. The ARMA model must be stationary and invertible, for the AR and MA components, respectively.

***Why Stationarity***

Stationarity affirms that the statistical properties of a time series do not change over time. The ARMA process requires that the autocovariance function cannot change with time.

***Unit Root Test***

A series following a unit root process is non-stationary but becomes stationary by taking first differences. The null hypothesis of a unit root (i.e. H0 = unit root) is rejected if the test statistic < 0.05.

An I(0) series is a stationary series, while an I(1) series contains 1 unit root. An I(2) series contains 2 unit roots and would require differencing twice to induce stationarity.

The Augmented Dickey-Fuller (ADF) test is widely used to test if a series contains a unit root, based on a regression of the change in that variable on the lag of the level of that variable. Meanwhile, the KPSS test checks if the series is stationary (H0 = stationary).

ADF test is used when the errors are homoscedastic and PP test is preferred for heteroscedastic errors.

A great **advantage** of [Philips-Perron](http://en.wikipedia.org/wiki/Phillips%E2%80%93Perron_test) test is that it is non-parametric, i.e. it does not require to select the level of serial correlation as in ADF. It rather takes the same estimation scheme as in DF test, but corrects the statistic to conduct for autocorrelations and heteroscedasticity (HAC type corrections).

The main **disadvantage** of the PP test is that it is based on asymptotic theory. Therefore it works well only in large samples that are indeed luxury if not it comes for financial time series data. And it also shares disadvantages of ADF tests: sensitivity to structural breaks, poor small sample power too often resulting in unit root conclusions.

It is **advisable** to make several tests and see if the results match, if not check the properties of the time series (PP is more robust to deviations from "gentleman's" set of properties!). You may also consider Zivot-Andrew test if you believe the data has structural breaks.

**Why Cointegration is important**

Cointegration aims to uncover causal relationships among variables by determining if the stochastic trends in a group of variables (e.g. spot vs. futures price) are shared by the series. Cointegration essentially means two variables have a long-run relationship.

Cointegration tests identify scenarios where two or more non-stationary time series are integrated together in a way that they cannot deviate from equilibrium in the long term. The tests are used to identify the degree of sensitivity of two variables to the same *average price* over a specified period of time. Essentially, cointegration is a technique used to find a possible correlation between time series processes in the long term.

**The Engle-Granger approach** is used to determine cointegration between one variables (Johansen’s test for multiple cointegrating vectors). The Engle-Granger test is a unit root test (with the Augmented Dickey-Fuller (ADF) test) applied to the **residuals of a potentially cointegrating regression.**

If the residuals have a unit root (i.e. I(1) form), there is no cointegration.

**Engle-Granger test (H0 = no cointegration)**

1. Estimate the cointegration regression.
2. Test the residuals from the cointegration regression for unit roots.

The Engle-Granger test statistic for cointegration reduces to an ADF unit root test of the residuals of the cointegration regression:

* **If the residuals contain a unit root, then there is no cointegration.**
* The null hypothesis of the ADF test is that the residuals have a unit root. Therefore, the Engle-Granger test considers the null hypothesis that there is no cointegration.
* As the Engle-Granger test statistic decreases:
  + We are more likely to reject the null hypothesis of no cointegration.
  + We have stronger evidence that the variables are cointegrated.

***Purpose of ECM***

If two series are cointegrated, an error correction model (ECM) could be appropriate rather than a model in pure first difference (i.e. I(1) form) because it would enable us to capture the long-run relationship between series as well as the short-run. The error-correction infers that the last-period’s deviation from a long-run equilibrium, called the error, influences its short-run dynamics. The ECM makes it possible to deal with non-stationary data series, separating the short-run and long-run.

**White Noise**

White noise can be used to construct predictable AR or MA processes. A time series model is adequate if the residuals are white noise. The goodness-of-fit of a time series model can be identified by the white noise. From the EViews correlogram, f the Prob\* column > 0.05, we can safely not reject null hypothesis of white noise. This highlights that the model is white noise.

**Serial Correlation (H0 = no serial correlation)**

If untreated, serial correlation can lead to a number of issues:

* 1. Reported standard errors and t-statistics are invalid (even asymptotically)
  2. Coefficients may be biased
  3. In the presence of lagged dependent variables, OLS estimates are biased and inconsistent

Overall, serial correlation has a large impact on standard errors and efficiency of estimators.

According to the correlogram of residuals, if there are no serial correlation, the AC and PACF at all lags should be near zero and all Ljung-Box Q-statistics should be insignificant.   
Clearly, this is not the case in the correlogram which shows substantial and persistent autocorrelation in residuals.

There are two other test statistic methods to test for serial correlation – the Durbin Watson and Breusch-Godfrey tests.

To test the hypothesis of no serial correlation, compare the reported ***Durbin-Watson stat*** to a table of critical values. E.g. DW = 0.027, we soundly reject the null hypothesis of no serial correlation. For **Breusch-Godfrey test,** the null hypothesis of no serial correlation is easily rejected.

To resolve serial correlation, the data can be differenced. Taking the first-order differences addresses a number of issues that arises in the time series data:

1. It eliminates most serial correlation
2. It de-trends the data
3. It transforms an I(1) process to an I(0)

After differencing,

1. The time trend now is not significant (Prob > 0.05). Taking the first-differences has detrended the data.
2. The R-squared value is much lower now reflecting the fact that it is harder to fit differences data (compared to levels).

**Heteroskedasticity and Autocorrelation (H0 = no heteroskedasticity)**

In many financial time series, the conditional variance of the error term depends on past values of the error term. This is also known as autoregressive conditional heteroskedasticity (ARCH).

When testing for heteroskedasticity, **residuals should not be serially correlated.** Any serial correlation will generate invalidate tests for heteroskedasticity. Since we have identified that there is no serial correlation, we can then test for heteroskedasticity.

You can correct both heteroskedasticity and autocorrelation of unknown form using the HAC Consistent Covariance (Newey-West).

We can perform the White test (i.e. test of H0 = no heteroskedasticity, against heteroskedasticity of unknown, general form).

Based on the test statistics, we reject the null hypothesis of no heteroskedasticity, which means that the error term is heteroskedastic and standard errors should be adjusted.

**ARCH effects (H0 = no ARCH)**

Testing for ARCH effects, also known as a test for autocorrelation in the squared residuals, allows us to determine of the assumption of constant variance is valid.

We can test for ARCH terms with the ARCH LM test. We reject the null hypothesis of no ARCH, which means that residuals suffer from heteroskedasticity.

**What is ARCH**

ARCH provides a framework for time series models of volatility.

**Extension of ARCH**

Owing to the limitations of ARCH, namely

1. Value of the no. of lags of squared residual has no clear best approach
2. Value of the no. of lags of squared error might be very large, resulting in large conditional variance that is not parsimonious

**What is GARCH**

GARCH is more parsimonious and avoid overfitting compared with ARCH. Maximum likelihood is used to find the most likely values of parameters. GARCH is the “ARMA Equivalent” of ARCH, and GARCH is a better fit for data exhibiting heteroskedasticity and volatility clustering.

The GARCH(1, 1) model will be sufficient to capture the volatility clustering in the data.

IGARCH = Integrated GARCH, ‘unit root in variance’

Owing to the limitations of GARCH, namely

1. GARCH enforces a symmetric response of volatility to positive/negative shocks

Other models arose:

1. (Threshold) T-GARCH (GJR) Model – to overcome leverage effects
2. EGARCH Model – to overcome negativity constraints
3. GARCH-in-mean – to incorporate “higher risk, higher return” concept

**ARCH-GARCH Condition Model**

The ARCH and GARCH models are useful when the goal is to **analyse and forecast volatility.**

**Glossary**

**autocorrelation**: a standardised measure, which must lie between −1 and +1, of the extent to which the current value of a series is related to its own previous values.

**autocorrelation function (ACF)**: a set of estimated values showing the strength of association between a variable and its previous values as the lag length increases.

**autocovariance**: an unstandardised measure of the extent to which the current value of a series is related to its own previous values.

**autoregressive conditional heteroscedasticity (ARCH) model**: a time series model for volatilities.

**autoregressive (AR) model**: a time series model where the current value of a series is fitted with its previous values.

**autoregressive moving average (ARMA) model**: a time series model where the current value of a series is fitted with its previous values (the autoregressive part) and the current and previous values of an error term (the moving average part).

**Bera–Jarque test**: a widely employed test for determining whether a series closely approximates a normal distribution.

**best linear unbiased estimator (BLUE)**: is one that provides the lowest sampling variance and which is also unbiased.

**Box–Jenkins approach**: a methodology for estimating ARMA models.

**Breusch–Godfrey test**: a test for autocorrelation of any order in the residuals from an estimated regression model, based on an auxiliary regression of the residuals on the original explanatory variables plus lags of the residuals.

**cointegration**: a concept whereby time series have a fixed relationship in the long run.

**cointegrating vector**: the set of parameters that describes the long-run relationship between two or more time series.

**conditional expectation**: the value of a random variable that is expected for time *t* + *s* (*s* = 1, 2, . . .) given information avail- able until time *t*.

**Dickey–Fuller test**: an approach to determining whether a series contains a unit root, based on a regression of the change in that variable on the lag of the level of that variable.

**differencing**: a technique used to remove a (stochastic) trend from a series that involves forming a new series by taking the lagged value of the original series away from the cur- rent one.

**Durbin–Watson statistic**: a test for first order autocorrelation, i.e. a test for whether a (residual) series is related to its immediately preceding values.

**Engle–Granger test**: a unit root test applied to the residuals of a potentially cointegrating regression.

**error correction model (ECM)**: a model constructed using variables that are employed in stationary, first-differenced forms together with a term that captures movements back towards long run equilibrium.

**exponential (EGARCH)**: a model where volatility is modelled in an exponential form so that no non-negativity conditions need to be applied to the parameters. This specification also that allows for asymmetries in the relationship between volatility and returns of different signs.

**first differences**: are new series constructed by taking the immediately previous value of a series from its current value.

**generalised autoregressive conditional heteroscedasticity (GARCH) models**: a common specification of dynamic model for volatility.

**GARCH-in-mean (GARCH-M)**: a dynamic model for volatility where the standard deviation (or variance) enters into the generating process for returns.

**Gauss–Markov theorem**: a derivation using algebra showing that, providing a certain set of assumptions hold, the OLS estimator is the best linear unbiased estimator (BLUE).

**Hannan–Quinn information criterion**: a metric that can be used to select the best fitting from a set of competing models and that incorporates a moderate penalty term for including additional parameters.

**heteroscedasticity**: where the variance of a series is not constant throughout the sample.

**information criteria**: a family of methods for selecting between competing models that incorporate automatic correction penalties when larger numbers of parameters are included.

**integrated GARCH (IGARCH)**: a model where the variance process is non-stationary so that the impact of shocks on volatility persists indefinitely.

**invertibility**: a condition for a moving average (MA) model to be representable as a valid infinite-order autoregressive model.

**Johansen test**: an approach to determining whether a set of variables is cointegrated – i.e. if they have a long-run equilibrium relationship.

**KPSS test**: a test for stationarity – in other words, a test where the null hypothesis is that a series is stationary against an alternative hypothesis that it is not.

**kurtosis**: the standardised fourth moment of a series; a measure of whether a series has ‘fat tails’.

**Ljung–Box test**: a general test for autocorrelation in a variable or residual series.

**moving average process**: a model where the dependent variable depends upon the cur- rent and past values of a white noise (error) process.

**multicollinearity**: a phenomenon where the two or more of the explanatory variables used in a regression model are highly related to one another.

**multivariate generalised autoregressive conditionally heteroscedastic (GARCH) models**: a family of dynamic models for time-varying variances and covariances.

**partial autocorrelation function (pacf)**: measures the correlation of a variable with its value *k* periods ago (*k* = 1,2,...) after removing the effects of observations at all intermediate lags.

**random walk**: a simple model where the current value of a series is simply the previous value perturbed by a white noise (error) term. Therefore the optimal forecast for a variable that follows a random walk is simply the most recently observed value of that series.

**random walk with drift**: a random walk model that also includes an intercept, so that changes in the variable are not required to average zero.

Chart, line chart

Description automatically generated

**residual sum of squares (RSS)**: the addition of all of the squared values of the differences between the actual data points and the corresponding model fitted values.

**residual terms**: the differences between the actual values of the dependent variable and the values that the model estimated for them – in other words, the parts of the dependent variable that the model could not explain.

**Schwarz’s Bayesian information criterion (SBIC)**: a metric that can be used to select the best fitting from a set of competing models and that incorporates a strict penalty term for including additional parameters.

**spurious regressions**: if a regression involves two or more independent non-stationary variables, the slope estimate(s) may appear highly significant to standard statistical tests and may have highly significant *t*-ratios even though in reality there is no relationship between the variables.

**stochastic trend**: some levels time series possess a stochastic trend, meaning that they can be characterised as unit root processes, which are non-stationary.

**Theil’s** *U***-statistic**: a metric to evaluate forecasts, where the mean squared error of the forecasts from the model under study is divided by the mean squared error of the fore- casts from a benchmark model. A *U*-statistic of less than one implies that the model is superior to the benchmark.

**unit root process**: a series follows a unit root process if it is non-stationary but becomes stationary by taking first differences.

**VECH model**: a relatively simple multivariate approach that allows for the estimation of time-varying volatilities and covariances that are stacked into a vector.

**vector autoregressive (VAR) model**: a multivariate time series specification where lagged values of (all) the variables appear on the right-hand side in (all) the equations of the (unrestricted) model.

**vector autoregressive moving average (VARMA) model**: a VAR model where there are also lagged values of the error terms appearing in each equation.

**vector error correction model (VECM)**: an error correction model that is embedded into a VAR framework so that the short- and long-run relationships between a set of variables can be modelled simultaneously.

**Wald test**: an approach to testing hypotheses where estimation is undertaken only under the alternative hypothesis; most common forms of hypothesis tests (e.g. *t*- and *F*-tests) are Wald tests.