Hypothesis testing/significance for time-series and accuracy in forecasting

Overview

Dr. Cristian Rodriguez Rivero

Friday, 29^{th}, 2020

Agenda

- Overview of classification in time-series analysis
- Notions of stationarity and Violations of stationarity
- General tests (Unit Root)
- Specific Tests (Autocorrelation or serial correlation)
- Stationary tests
- Forecast errors and Test

Time-series Analysis – Statistic Tests

General

- Decomposition
- Trend
- Stationarity
- Seasonal adjustment
- Exponential smoothing
- Cointegration
- Structural break
- Granger causality

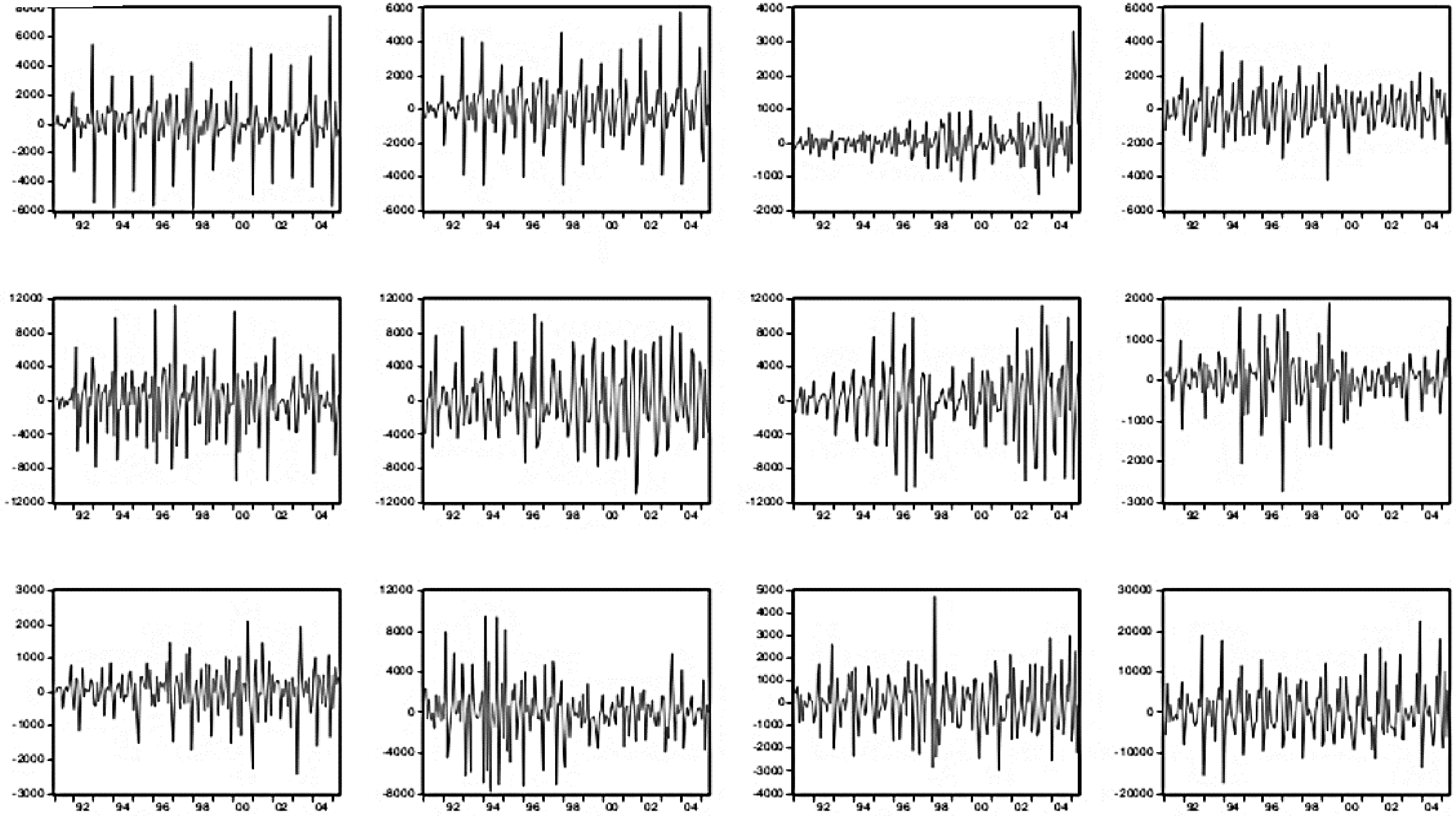
Specific tests

- Johansen (Cointegration)
- Q-statistic (Ljung–Box**)** Box-Pierce and Ljung-Box
- Durbin–Watson
- Breusch–Godfrey

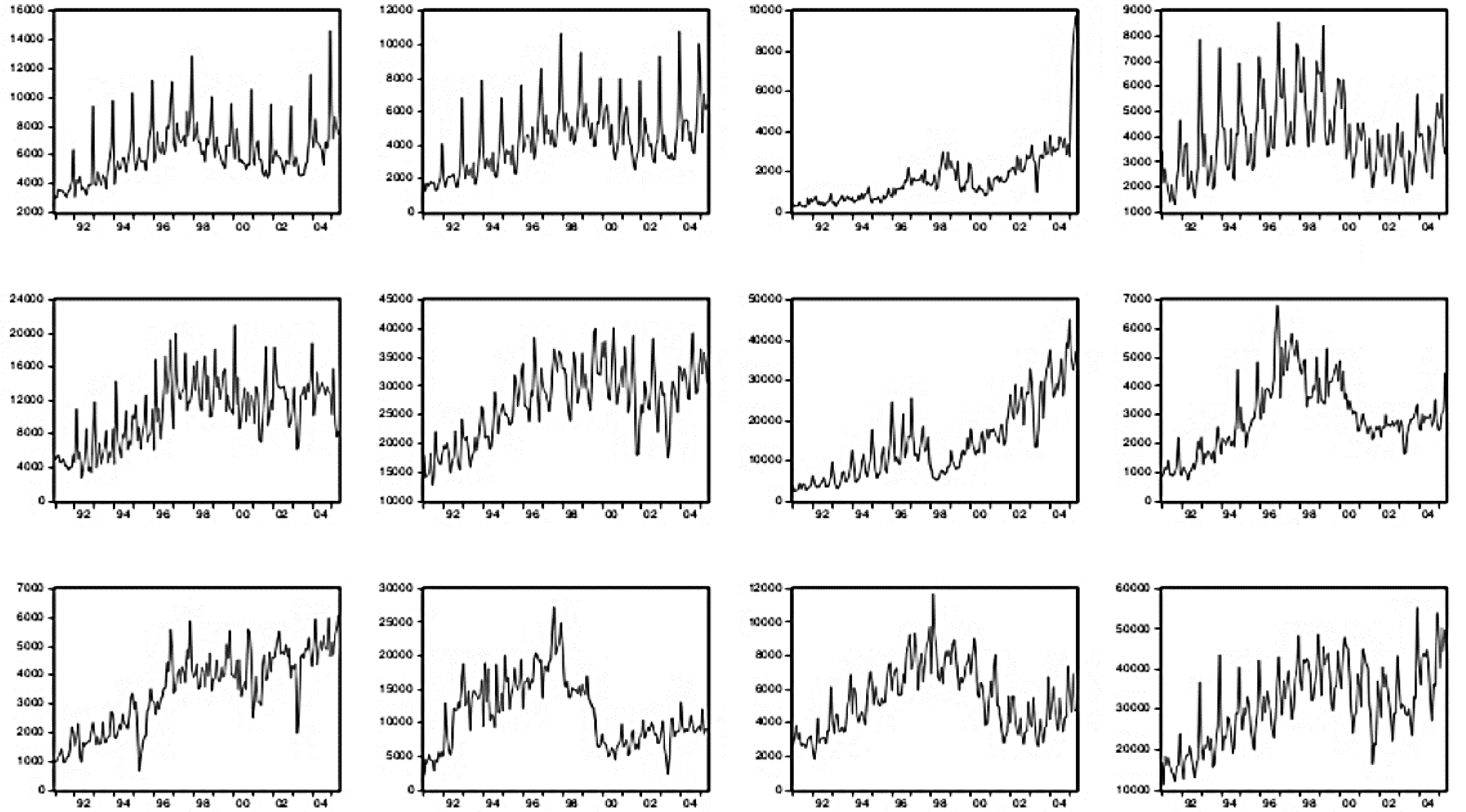
Time-series Analysis in Time and Frequency domain

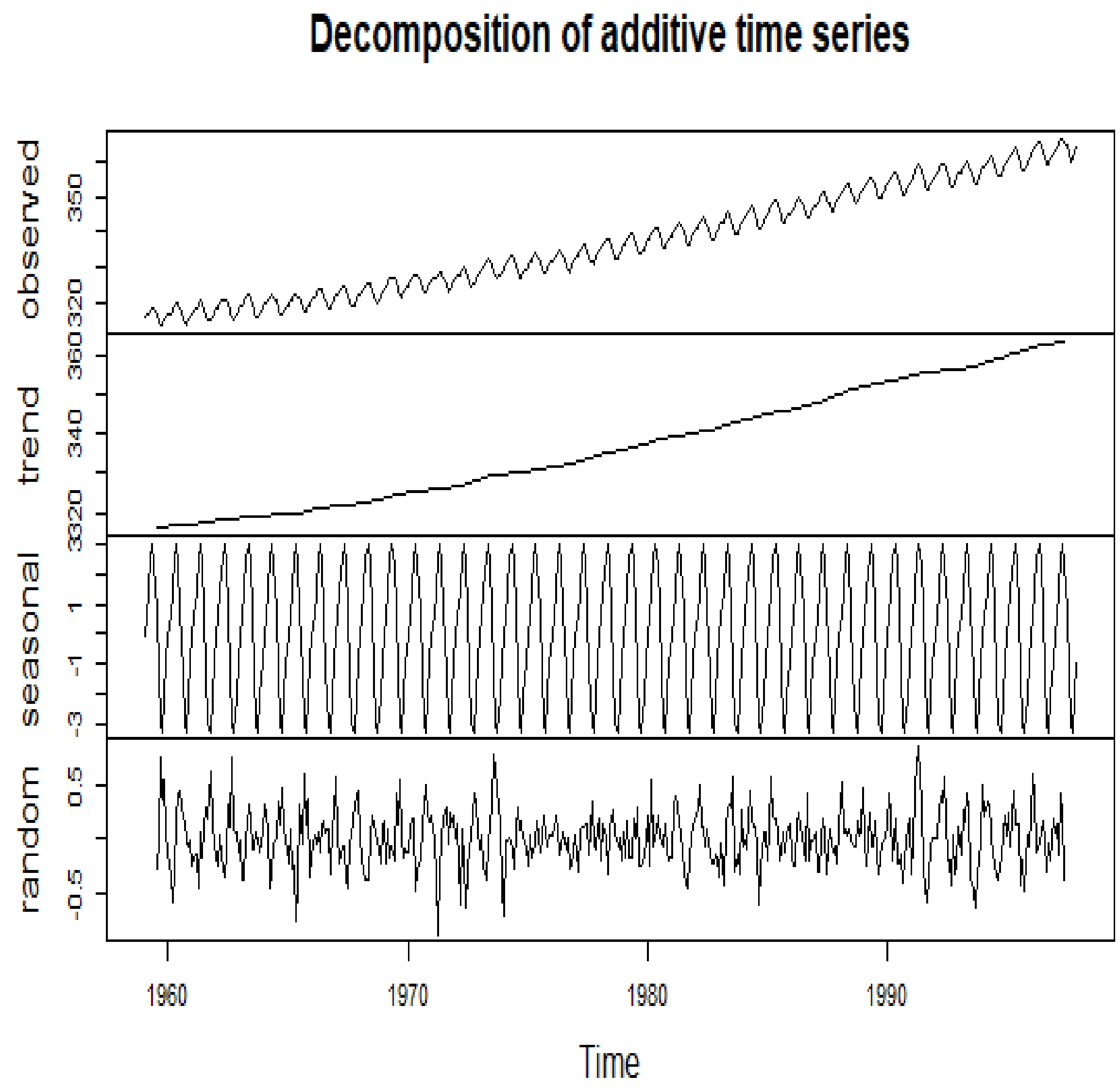
Time domain	Autocorrelation (ACF) partial (PACF) Cross-correlation (XCF) ARMA model ARIMA model (Box–Jenkins) Autoregressive conditional heteroskedasticity (ARCH) Vector autoregression (VAR)
Frequency domain	Spectral density estimation Fourier analysis Wavelet Whittle likelihood

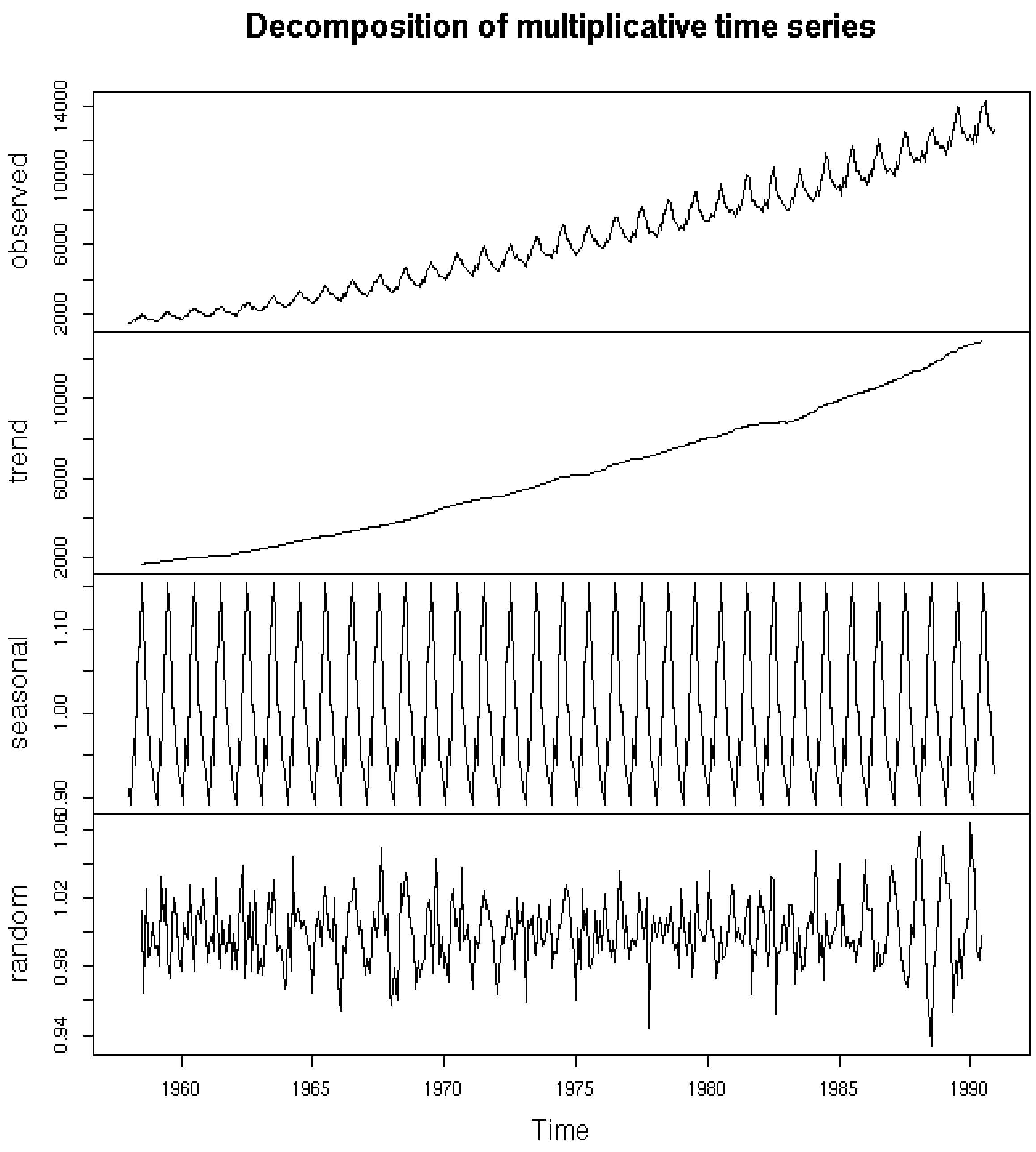
Example of stationary time series



Example of non-stationary time series







https://cran.r-project.org/web/packages/stR/vignettes/stRvignette.html

Stationarity

- The statistical basis for our estimation and forecasting depends on series being covariance stationary. (constant mean and variance across time and, the roots of the characteristic polynomial of the autoregressive process are outside the unit circle in the complex plane.)

- Covariance stationary (random variables) = if all the terms of the sequence have the same mean, and if the covariance between any two terms depends only on the relative positions of the two terms (how far apart they are located from each other, not on their absolute position located in the sequence).

Types of stationarity

- **Strict stationarity** means that the joint distribution of any moments of any degree (e.g. expected values, variances, third order and higher moments) within the process is never dependent on time.
- **First-order stationarity** series have means that never changes with time. Any other statistics (like variance) can change.
- **Second-order stationarity** (weak stationarity) time series have a constant mean, variance and an autocovariance that doesn’t change with time. Other statistics in the system are free to change over time. This constrained version of strict stationarity is very common.
- **Trend-stationary** models fluctuate around a deterministic trend (the series mean). These deterministic trends can be linear or quadratic, but the amplitude (height of one oscillation) of the fluctuations neither increases nor decreases across the series.
- **Difference-stationary** models are models that need one or more differencings to become stationary.

Non-Stationary Time Series?

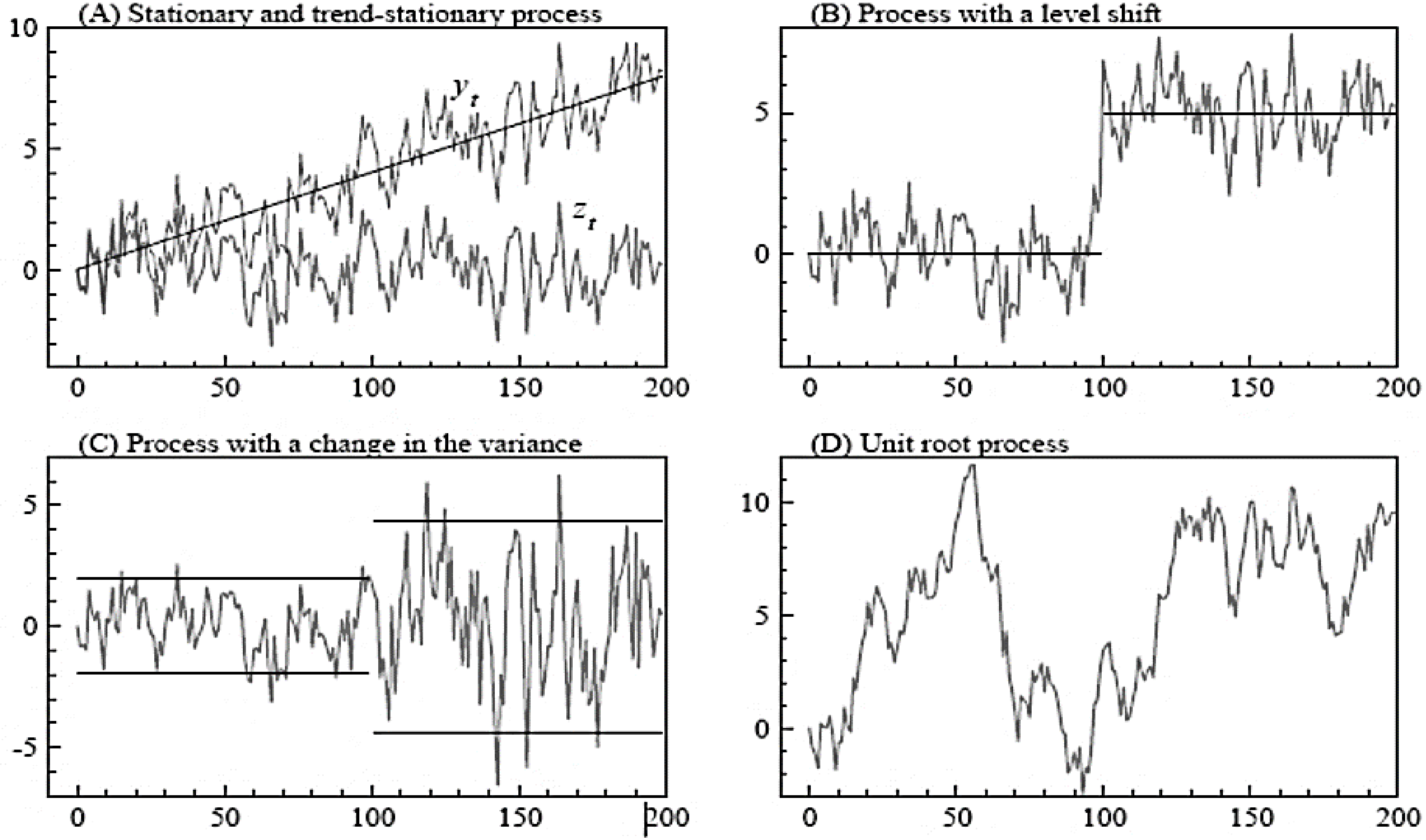
- Many time series are nonstationary.
- Need tests for stationarity, unit roots, cointegration (will depend on the nature of the time series)

- Stationarity can be violated in different ways. Examples of non-stationarity:

- 1) Deterministic trends (trend stationarity).
- 2) Level shifts.
- 3) Variance changes (volatility).
- 4) Unit roots (stochastic trends).

https://bookdown.org/ccolonescu/RPoE4/time-series-nonstationarity.html

Types of non-stationary time series



Unit root tests

- DF (Dickey-Fuller) test
- ADF (Augmented Dickey Fuller) test

Other tests

- PP (Phillipps-Perron) test
- ZA (Zivot-Andrews) test
- ADF-GLS (Augmented Dickey-Fuller Generalized Least Square) test

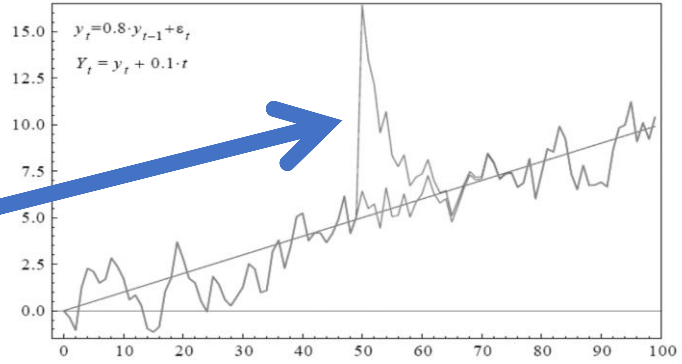
Why are unit roots important?

- Interesting to know if shocks have permanent or transitory effects.
- It is important for forecasting to know if the process has an attractor.

This is caused by:

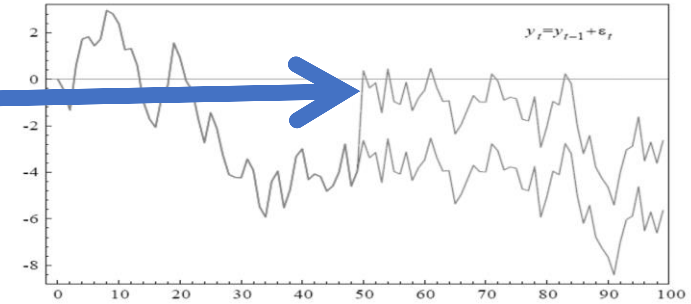
A stationary process with a deterministic trend:

- Shocks have transitory effects.



A process with a stochastic trend or a unit root:

- Shocks have permanent effects.



Problems with Unit Roots?

If time-series are not covariance stationary unit roots require some special treatment.

Drawbacks

- Statistically, the existence of unit roots can be problematic because OLS estimate of the AR(1) coef. φ is biased.
  - In multivariate frameworks, one can get spurious regression results

So to identify the correct underlying time series model, we must test whether a unit root exists or not.

When to use Unit Roots?

Bear in mind!

Unit roots is a test for a specific type of non-stationarity for a specific type of time series models.

Other tests can test for other forms of nonstationarity and deal with more general forms of time series.

The Dickey-Fuller (DF) Test

The Dickey-Fuller (DF) test is simply the t− test of H0 :

null hypothesis :rho = 0

alternative hypothesis: rho < 0.

- If the series is stationary (or trend stationary), then it has a tendency to return to a constant (or deterministically trending) mean.

- This test is asymmetrical, we are only concerned with negative values of our test statistic.

Dickey, D. A.; Fuller, W. A. (1979).

Augmented Dickey-Fuller (ADF) test

- Null Hypothesis: the process has a unit-root ("difference stationary")

- Alternative Hypothesis: the process has no unit root. It can mean either that the process is stationary, or trend stationary, depending on which version of the ADF test is used.

Specific Tests for autocorrelation

Q-statistic (Ljung–Box) as an extension of Box-Pierce

Durbin–Watson

Breusch–Godfrey

Durbin-Watson test

Durbin-Watson test detect the presence of autocorrelation at lag 1 in the residuals (prediction errors) from a regression analysis.

Null hypothesis: rho=0 ,

Alternative hypothesis rho≠0

$$DW = \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2} \quad , \ 0<DW<4,$$

Interpretation:

if DW < 1.0, alarm!

Reject Null. Accept that there is +ve A/C.	Test inconclusive	Accept Null (No autocorrelation)	Test inconclusive	Reject Null Accept there is –ve A/C.
0	DW_{low}	DW_{upper} 2	$4\text{-}DW_{upper}$	$4\text{-}DW_{low}$

Q-statistics

The **Ljung–Box (also called Portmanteau)** test may be defined as:

- **H0:** the autocorrelations up to lag k are all 0:
- **HA:** the autocorrelations of one or more lags differ from 0.
- Test statistics:

$$Q_k = n(n+2)\sum_{j=1}^{k} \frac{r_j^2}{n-j},$$

which is approximately χ2k-distributed.

- Commonly used in autoregressive integrated moving average (ARIMA)

Breusch–Godfrey test

It's a test for autocorrelation in the errors in a regression model.

The test is more general than that using the Durbin–Watson statistic, for autocorrelation of any order

Null hypothesis: rho1= rho2= rho3=… rhop=0, there is no autocorrelation of any order up to p.

Either compute the F test for the joint significance of the residuals and if F > Fcritical reject null of no p order autocorrelation

or, compute (N-p)*R2 auxillary ~ χ2 (p)

Stationary tests

KPSS (Kwiatkowski, Phillips, Schmidt, Shin,1992)

These tests are used for testing a null hypothesis that an observable time series is stationary around a deterministic trend.

- Null Hypothesis: the process is trend-stationary
- Alternative Hypothesis: the process has a unit root

Leybourne-McCabe (Leybourne and McCabe1994, 1999)

- Null Hypothesis: stationarity
- Alternative Hypothesis: unit root

Has been shown more powerful than the KPSS test.

Models Evaluation based on Mean Forecast Error (or Forecast Bias)

- **Scale-dependent errors**

Mean absolute error: MAE

Root mean squared error: RMSE

- **Percentage errors**

Mean absolute percentage error: MAPE

Symmetric Mean absolute percentage error: sMAPE

- **Scaled errors**

Mean Absolute Scaled Error: MASE

Hyndman & Koehler (2006), Armstrong (1978, p. 348)

Models Evaluation based on Mean Forecast Error (or Forecast Bias)

- **Scale-dependent errors**

Mean absolute error: MAE

Root mean squared error: RMSE

- **Percentage errors**

Mean absolute percentage error: MAPE

Symmetric Mean absolute percentage error: sMAPE

- **Scaled errors**

Mean Absolute Scaled Error: MASE

Hyndman & Koehler (2006), Armstrong (1978, p. 348)

The Diebold-Mariano Test

- Null hypothesis H0 : E(dt) = 0 ∀t
- Alternative hypothesis H1 : E(dt) ≠ 0

Under the null hypothesis, the test statistics DM is asymptotically N(0, 1) distributed and rejected at significance level.

Diebold, F.X. and R.S. Mariano. (1995)

Diebold, F.X. and Kilian, L. (2000)

The Diebold-Mariano Test

When not use!

The Diebold-Mariano test should not be applied to situations where the competing forecasts are obtained using two nested models.

Harvey, Leybourne, and Newbold (1997) (HLN) suggest to improved for small-sample properties,

1) making a bias correction to the DM test statistic,

2) comparing the corrected statistic with a Student-t distribution with (T-1) degrees of freedom, rather than the standard normal.

Harvey, Leybourne, and Newbold (1997)

References

- Enders, Walter (2010). Applied Econometric Time Series.
- J. Durbin and G. S. Watson, 1951, "Testing for Serial Correlation in Least Squares Regression: II", Biometrika, Vol. 38, No. 1/2, pp. 159-177.
- Durbin and G. S. Watson, 1950, "Testing for Serial Correlation in Least Squares Regression: I", Biometrika, Vol. 37, No. 3/4, pp. 409-428.
- Bagnall, A., Lines, J., Bostrom, A., Large, J., & Keogh, E. (2017). The great time series classification bake off: a review and experimental evaluation of recent algorithmic advances. Data Mining and Knowledge Discovery, 31(3), 606–660.
- Hyndman, R.J., & Athanasopoulos, G. (2018) Forecasting: principles and practice, 2nd edition, OTexts: Melbourne, Australia.
- Time Series Analysis: Forecasting and Control, 5th Edition, by George E. P. Box, Gwilym M. Jenkins, Gregory C. Reinsel and Greta M. Ljung, 2015.
- Kwiatkowski, D.; Phillips, P. C. B.; Schmidt, P.; Shin, Y. (1992). "Testing the null hypothesis of stationarity against the alternative of a unit root". *Journal of Econometrics*. **54** (1–3): 159–178
- Box, G. E. P. and Cox, D. R. (1964). An analysis of transformations, Journal of the Royal Statistical Society, Series B, 26, 211-252.

References

- Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. International Journal of Forecasting, 22, 679–688.
- Armstrong, J. S. (1978). Long-range forecasting: From crystal ball to computer. John Wiley & Sons.
- Ord, J. K., Fildes, R., & Kourentzes, N. (2017). Principles of business forecasting (2nd ed.). Wessex Press Publishing Co.
- Francis X. Diebold (2015) Comparing Predictive Accuracy, Twenty Years Later: A Personal Perspective on the Use and Abuse of Diebold–Mariano Tests, Journal of Business & Economic Statistics, 33:1, 1.
- Dickey, D. A.; Fuller, W. A. (1979). "Distribution of the Estimators for Autoregressive Time Series with a Unit Root". Journal of the American Statistical Association. 74 (366): 427–431.
- Makridakis, Spyros (1993-12-01). "Accuracy measures: theoretical and practical concerns". International Journal of Forecasting. 9 (4): 527–529.
- Harvey, D., S. Leybourne, and P. Newbold. (1997). Testing the Equality of Prediction Mean Squared Errors. International Journal of Forecasting, 13: 281-91.

References

- Diebold, Francis X.; Mariano, Roberto S. (2002). "Comparing predictive accuracy". Journal of Business and Economic Statistics. 20 (1): 134–144.
- Diebold, F.X. and Kilian, L. (2000), "Unit Root Tests are Useful for Selecting Forecasting Models," Journal of Business and Economic Statistics, 18, 265-273.
- Cheung, Yin-Wong; Kon S. Lai (1995). "Practitioner's Corner: Lag Order and Critical Values of a Modified Dickey-Fuller Test". Oxford Bulletin of Economics and Statistics. 57 (3): 411–419.
- Godfrey, L. G. (1988). Misspecification Tests in Econometrics. Cambridge, UK: Cambridge.
- Breusch, T. S. (1978). "Testing for Autocorrelation in Dynamic Linear Models". Australian Economic Papers. 17: 334–355.
- Godfrey, L. G. (1978). "Testing Against General Autoregressive and Moving Average Error Models when the Regressors Include Lagged Dependent Variables". Econometrica. 46: 1293–1301