AB1202
Statistics and Analysis

Lecture 12
Time Series Predictive Models

Chin Chee Kai
cheekai@ntu.edu.sg
Nanyang Business School
Nanyang Technological University

Time Series Predictive Models

- Regression Prediction in Time Series Data
- Selecting “Better” Models in Time Series
- Autocorrelation
- Durbin-Watson Test
- Auto-Regressive (AR) Predictive Models

Regression Prediction in Time Series

- Nobody can predict the future. Agreed!
- Then why do we talk about forecasting and prediction?
- We have to ask a totally different question first. How do house flies avoid getting hit so naturally?
- Flies see 10 times better than human, ie 300 frames/sec! This means things happening slower than 3.33 msec can be detected by the fly. (Source: http://www.berkeley.edu/news/media/releases/98legacy/04_09_98a.html)
- How fast can your hand slap the house fly? Say at an incredible 10m/s when your hand is about 20cm away from the fly? To complete traveling 20cm, your hand will take 20 msec. Within 3.33 msec, the fly already saw your hand moving towards it! Its forecast projecting that your hand will hit it in the next 16 msec gives it ample time to escape.
- To you, it's a split second. To the fly, you were stationary while it chooses its best escape route.

Image Source: https://www.istockphoto.com/sg/photo/slap-gm480624571-37127578

Image Source: http://www.doyourownpestcontrol.com/images/housefly-1.jpg

To forecast a fast-moving hand, you need to sample even faster like a fly

Regression Prediction in Time Series

- In linear regression, we obtain models from past data to predict response values.
- We normally try to stick to known range of input values which have been used in forming the regression model
  - substituting values outside the original data ranges is often frowned upon.
- In time series regression, however, we MUST substitute a time value outside of past data.
- We shall use "forecasting" to refer to such future value prediction where input variable involves time.

Selecting "Better" Models

- Need objective, quantifiable way to select model
  - Could be to justify using one model (possibly proposed by someone) over another model (possibly by somebody else)
  - Could be to automate through algorithmic trading, drone flight systems, robotics, e-commerce, etc.
- Error Measurement
  - At time $t = i$, forecast $\hat{y}_i$, then actual data y_i comes in.
  - So error = $\varepsilon_i = y_i - \hat{y}_i$
- Error Measurement Function
  - A formula that combines all the n errors $\varepsilon_1, \dots \varepsilon_n$, and possibly $y_1, \dots y_n$ and $\hat{y}_1, \dots \hat{y}_n$ and turn them into a single, non-negative decimal value (smaller better).

Selecting "Better" Models

- MAD – Mean Absolute Deviation
  - $\text{MAD} = \frac{1}{n}\sum_{i=1}^{n}|\varepsilon_i| = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$
  - Easy and fast to calculate on computers; numbers are small and manageable.
  - Not nice for analysis.
- MSE – Mean Squared Error
  - $\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}\varepsilon_i^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$
  - Formula is friendly to analysis.
  - But numbers typically huge; has squared units; hard to interpret.
- RMSE – Root-Mean Squared Error
  - $\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i^2} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$
  - Improvement of MSE – numbers are smaller; unit is same as data; easier to interpret.

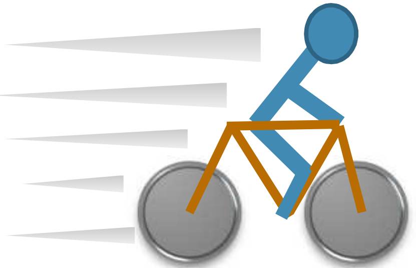
Autocorrelation

- Data series whose **errors** correlate with its lagged values are said to have *autocorrelation*.
- Lagged values mean delayed values. Eg:

Time t	1	2	3	4	5	…
Error ε_t	3	-2	-1	5	4	…
Lag-1 ε_{t-1}		3	-2	-1	5	…
Lag-2 ε_{t-2}			3	-2	-1	…

- Lag-1 means time t sees the data up to 1 period before (ie at $t-1$).
- Lag-2 means time t sees the data up to 2 periods before (ie at $t-2$).
- And so on.
- Just like error series is itself a random variable, lagged error series are also random variables – their values just happened to be mirrored values from the past.

It's mind boggling perhaps, but you just have to take it like how you ride a bicycle – it's mind boggling how you can ride so fast balancing on just two thin wheels!



Durbin-Watson Test

- Since correlation is never quite zero for practical data, autocorrelation always exists – the question is whether it is significant.
- Durbin-Watson Test is such a test for autocorrelation.
- Durbin-Watson test statistic $DW = \frac{\sum_{i=2}^{n}(\varepsilon_i - \varepsilon_{i-1})^2}{\sum_{i=1}^{n}\varepsilon_i^2}$
- Take note that the ε_i's are errors left over after a regression model's predicted values have been deducted from the observed data series.
  - Common mistake is to apply DW test on the data itself.
  - DW Test is about the autocorrelation of *errors* with its lagged-1 self, not data values with itself.
- Critical values are from look-up tables. Require (1) significance level α, (2) sample size n, (3) number of explanatory variables k
- We will mainly be concerned with only lag-1 analysis

Example

Eg, for the table of errors from time 1 to 5 ($n = 5$),

$$DW = \frac{\sum_{i=2}^{5}(\varepsilon_i - \varepsilon_{i-1})^2}{\sum_{i=1}^{5}\varepsilon_i^2} = \frac{(-2-3)^2 + \left(-1-(-2)\right)^2 + \left(5-(-1)\right)^2 + (4-5)^2}{3^2 + (-2)^2 + (-1)^2 + 5^2 + 4^2}$$

$$= \frac{63}{55} = 1.1455$$

Durbin-Watson Test

For illustration, we let $n = 30$, $k = 1$, and $\alpha = 0.05$. Test Statistic $DW = 2.6$

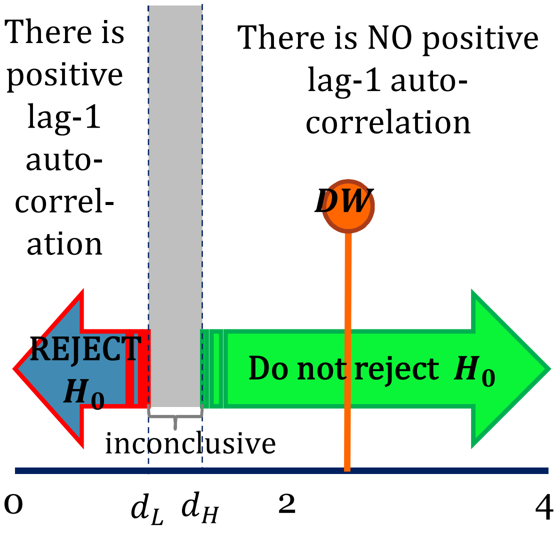
Case 1: Test for existence of POSITIVE autocorrelation

$$H_0\text{: } \rho \leq 0$$
$$H_1\text{: } \rho > 0$$

Get critical values:

$d_L = 1.35$
$d_H = 1.49$

Conclusion: Since $DW = 2.6 > d_H = 1.49$, we do not reject H_0 and conclude there is NO positive autocorrelation at 5%.


There is positive lag-1 auto-correl-ation
There is NO positive lag-1 auto-correlation
DW
REJECT H_0
Do not reject H_0
inconclusive
0
d_L
d_H
2
4

Case 2: Test for existence of NEGATIVE autocorrelation

$$H_0\text{: } \rho \geq 0$$
$$H_1\text{: } \rho < 0$$

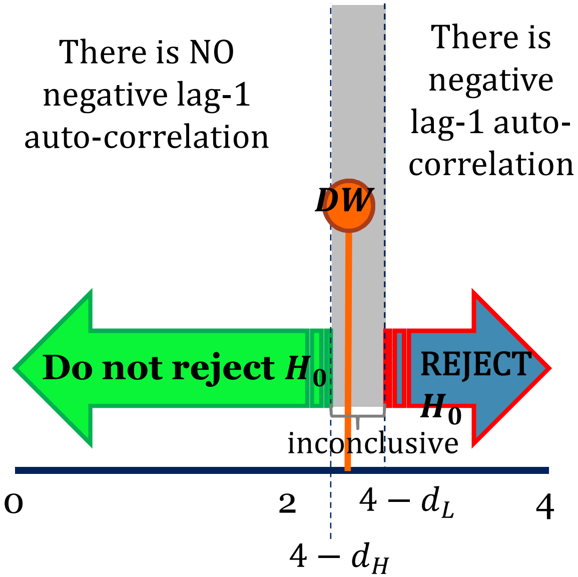
Get crit. vals.:

$d_L = 1.35$
$d_H = 1.49$

Calculate:

$4 - d_H = 2.51$
$4 - d_L = 2.65$

Conclusion: Since $DW = 2.6$ is between $4 - d_H = 2.51$ and $4 - d_L = 2.65$, the test is inconclusive at 5%.


There is NO negative lag-1 auto-correlation
There is negative lag-1 auto-correlation
DW
Do not reject H_0
REJECT H_0
inconclusive
0
2
$4 - d_L$
4
$4 - d_H$

Case 3: Test for existence of autocorrelation

$$H_0\text{: } \rho = 0$$
$$H_1\text{: } \rho \neq 0$$

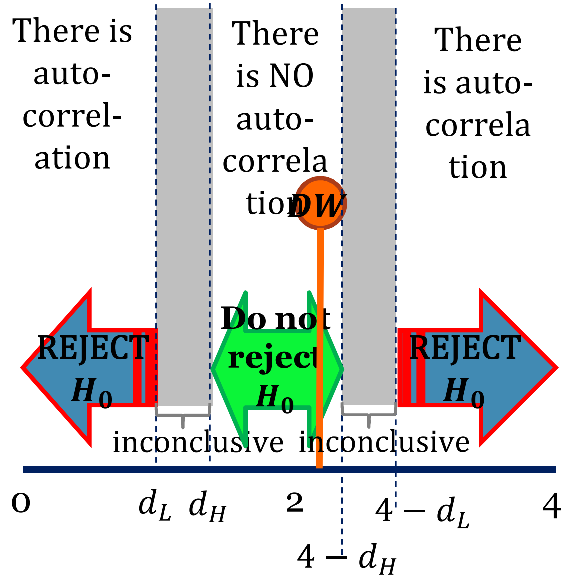
Get crit. vals.:

$d_L = 1.25$
$d_H = 1.38$
for 0.025 sig level

Calculate:

$4 - d_H = 2.62$
$4 - d_L = 2.75$

Conclusion: Since $DW = 2.6$ is between $d_H = 1.38$ and $4 - d_H = 2.62$, we do not reject H_0 and conclude there is no autocorrelation at 5%.



Example: Singapore Old Age People

Year	Year t	>= 60 Years (y, Million)	Predicted (y^)	Error (e_i)	Lag-1 Error (e_i-1)	(e_i)^2	(e_i - e_i-1)^2
1991	1	0.623184	0.51169	0.111493		1.24307E-02	
1992	2	0.645282	0.55403	0.091255	0.111493	8.32748E-03	4.09577E-04
1993	3	0.667948	0.59636	0.071585	0.091255	5.12441E-03	3.86909E-04
1994	4	0.690494	0.63870	0.051796	0.071585	2.68283E-03	3.91605E-04
1995	5	0.715036	0.68103	0.034002	0.051796	1.15614E-03	3.16626E-04
1996	6	0.73501	0.72337	0.01164	0.034002	1.35490E-04	5.00059E-04
1997	7	0.756681	0.76571	-0.00902	0.01164	8.14506E-05	4.27042E-04
1998	8	0.7828	0.80804	-0.02524	-0.009025	6.37159E-04	2.62991E-04
1999	9	0.814367	0.85038	-0.03601	-0.025242	1.29679E-03	1.15971E-04
2000	10	0.848528	0.89271	-0.04419	-0.036011	1.95240E-03	6.68306E-05
2001	11	0.887076	0.93505	-0.04797	-0.044186	2.30150E-03	1.43489E-05
2002	12	0.922683	0.97739	-0.0547	-0.047974	2.99242E-03	4.52794E-05
2003	13	0.912398	1.01972	-0.10732	-0.054703	1.15184E-02	2.76897E-03
2004	14	0.964575	1.06206	-0.09748	-0.107324	9.50294E-03	9.68453E-05
2005	15	1.008085	1.10439	-0.09631	-0.097483	9.27542E-03	1.37828E-06
2006	16	1.054314	1.14673	-0.09242	-0.096309	8.54072E-03	1.51554E-05
2007	17	1.108166	1.18907	-0.0809	-0.092416	6.54481E-03	1.32618E-04
2008	18	1.164941	1.23140	-0.06646	-0.0809	4.41706E-03	2.08485E-04
2009	19	1.235236	1.27374	-0.0385	-0.066461	1.48240E-03	7.81706E-04
2010	20	1.298983	1.31607	-0.01709	-0.038502	2.92102E-04	4.58431E-04
2011	21	1.367142	1.35841	0.008732	-0.017091	7.62478E-05	6.66827E-04
2012	22	1.444036	1.40075	0.04329	0.008732	1.87402E-03	1.19426E-03
2013	23	1.524777	1.44308	0.081695	0.04329	6.67407E-03	1.47494E-03
2014	24	1.615989	1.48542	0.130572	0.081695	1.70490E-02	2.38896E-03
2015	25	1.70532	1.52775	0.177567	0.130572	3.15300E-02	2.20853E-03
			Total:			0.147896	0.015334
						DW =	0.103683

Source: Dept of Statistics, Singapore. Accessed on: 2016 Aug 07,
http://www.tablebuilder.singstat.gov.sg/publicfacing/createDataTable.action?refId=315

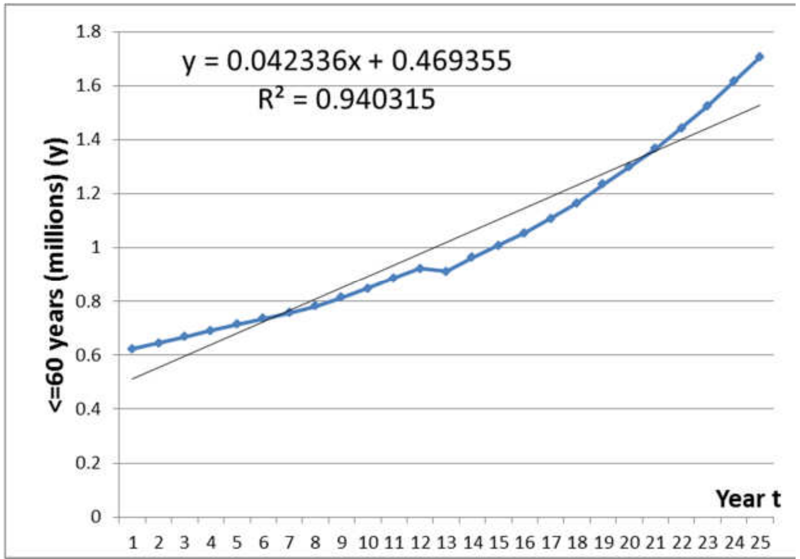
Example

Question: using linear regression, do the errors exhibit significant auto-correlation at 5%?

Linear model is:

$$y_t = 0.469355 + 0.042336\ t$$

$R^2 = 0.9403$ which is not bad. But if there is significant auto-correlation, we could be getting unreliable forecasts.



Example: Singapore Old Age People

Example

- Hypothesis is: $H_0\colon \rho = 0, H_1\colon \rho \neq 0$
- $n = 25, k = 1.$ Significance = 0.05
- Test statistic $DW = \frac{\sum_{i=2}^{n}(\varepsilon_i - \varepsilon_{i-1})^2}{\sum_{i=1}^{n}\varepsilon_i^2} = \frac{0.015334345}{0.147896} = 0.1037$
- From DW table for significance 0.025, we look up d_L and d_H, and calculate upper limits $4 - d_L$ and $4 - d_H$
  - Durbin-Watson initially worked on positive autocorrelation. So tables and significance are listed for positive autocorrelation (ie one-tailed).
  - For equality two-tailed test, we should lookup table with half the required significance.
- $d_L = 1.18, d_H = 1.34.\ 4 - d_H = 2.66, 4 - d_L = 2.82$
- Since $DW < d_L$, REJECT H_0 and conclude errors have autocorrelation at 5% level.

Durbin-Watson Test and Correlation

- After making minor assumptions and exercising tedious derivation, DW is also approximately given by $DW \approx 2(1 - r_{\varepsilon_t,\varepsilon_{t-1}})$
  - Let's check: for Singapore Old Age People linear model errors, $r_{\varepsilon_t,\varepsilon_{t-1}} = 0.94238$
  - $2(1 - r_{\varepsilon_t,\varepsilon_{t-1}}) = 2(1 - 0.94238) = 0.11524$
  - Actual $DW = 0.1037$
- As errors correlate more and more positively, $r_{\varepsilon_t,\varepsilon_{t-1}} \rightarrow +1$ and so $DW \rightarrow 0$
- As errors get more and more uncorrelated, $r_{\varepsilon_t,\varepsilon_{t-1}} \rightarrow 0$ and so $DW \rightarrow 2$
- As errors correlate more and more negatively, $r_{\varepsilon_t,\varepsilon_{t-1}} \rightarrow -1$ and so $DW \rightarrow 4$

- Not bad, but this just gives us better intuition
- Still must use Durbin-Watson test to properly test for autocorrelation at stated significance levels.

Auto-Regressive (AR) Predictive Models

- When DW test shows our regression model has autocorrelation (no good), we can use Auto-Regressive (AR) models to overcome auto-correlation.
  - This just means using the lagged data series as one or more of explanatory variables.
  - Idea is: by "borrowing" from the auto-correlatedness of itself from the nearby past, we can use them to account for future variations much better than other variables, since there is significant correlation with itself.
- If y_t is the future variable to be forecasted, AR(k) models will use $y_{t-1}, y_{t-2}, \ldots y_{t-k}$ past variables to explain away future variations, and thereby reduce auto-correlations.
  - ie $y_t = b_0 + b_1 y_{t-1} + +b_2 y_{t-2} + \cdots + b_k y_{t-k}$
- We will stick to only AR(1) models.

Example: AR(1) Singapore Old Age People

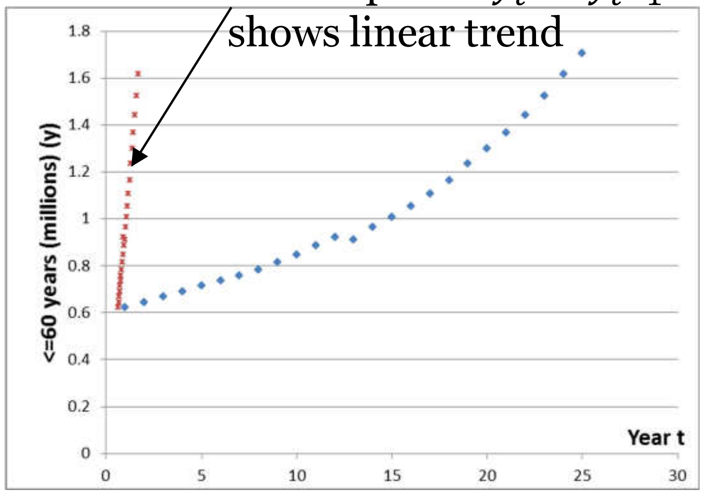
Year	Year t	>=60 Years (y, Million)	Lag-1 y_1	Predicted (y^)	Error (e_i)	Lag-1 Error (e_i-1)	(e_i)^2	(e_i - e_i-1)^2
1991	1	0.623184						
1992	2	0.645282	0.623184	0.63952	0.005758		3.3155E-05	
1993	3	0.667948	0.645282	0.66335	0.004599	0.005758	2.1151E-05	1.3433E-06
1994	4	0.690494	0.667948	0.68779	0.002709	0.004599	7.3387E-06	3.5721E-06
1995	5	0.715036	0.690494	0.71209	0.002943	0.002709	8.6612E-06	5.4756E-08
1996	6	0.73501	0.715036	0.73855	-0.00354	0.002943	1.2546E-05	4.2055E-05
1997	7	0.756681	0.73501	0.76009	-0.00341	-0.003542	1.1601E-05	1.8496E-08
1998	8	0.7828	0.756681	0.78345	-0.00065	-0.003406	4.2380E-07	7.5900E-06
1999	9	0.814367	0.7828	0.81161	0.002756	-0.000651	7.5955E-06	1.1608E-05
2000	10	0.848528	0.814367	0.84564	0.002884	0.002756	8.3175E-06	1.6384E-08
2001	11	0.887076	0.848528	0.88247	0.004602	0.002884	2.1178E-05	2.9515E-06
2002	12	0.922683	0.887076	0.92403	-0.00135	0.004602	1.8225E-06	3.5426E-05
2003	13	0.912398	0.922683	0.96242	-0.05002	-0.00135	2.5024E-03	2.3692E-03
2004	14	0.964575	0.912398	0.95133	0.013241	-0.050024	1.7532E-04	4.0025E-03
2005	15	1.008085	0.964575	1.00759	0.000498	0.013241	2.4800E-07	1.6238E-04
2006	16	1.054314	1.008085	1.05450	-0.00018	0.000498	3.3489E-08	4.6376E-07
2007	17	1.108166	1.054314	1.10434	0.003829	-0.000183	1.4661E-05	1.6096E-05
2008	18	1.164941	1.108166	1.16240	0.002544	0.003829	6.4719E-06	1.6512E-06
2009	19	1.235236	1.164941	1.22361	0.011629	0.002544	1.3523E-04	8.2537E-05
2010	20	1.298983	1.235236	1.29939	-0.00041	0.011629	1.6892E-07	1.4496E-04
2011	21	1.367142	1.298983	1.36812	-0.00098	-0.000411	9.6040E-07	3.2376E-07
2012	22	1.444036	1.367142	1.44161	0.00243	-0.00098	5.9049E-06	1.1628E-05
2013	23	1.524777	1.444036	1.52451	0.000269	0.00243	7.2361E-08	4.6699E-06
2014	24	1.615989	1.524777	1.61156	0.004432	0.000269	1.9643E-05	1.7331E-05
2015	25	1.70532	1.615989	1.70990	-0.00458	0.004432	2.0931E-05	8.1126E-05
				Total:			0.0030158	0.00699943
					r_ei,ei-1 =	-0.170627	DW =	2.32088609

Source: Dept of Statistics, Singapore. Accessed on: 2016 Aug 07,
http://www.tablebuilder.singstat.gov.sg/publicfacing/createDataTable.action?refId=315

Example

Question: using AR(1) regression, do the errors still exhibit significant auto-correlation at 5%?

AR(1) model is: $y_t = -0.03234789 + 1.0781279\ y_{t-1}$



Scatter plot of y_t vs y_{t-1} shows linear trend

Example: AR(1) Singapore Old Age People

Example

- Hypothesis is: $H_0\colon \rho = 0, \quad H_1\colon \rho \neq 0$
- $n = 24, \ k = 1$. Significance = 0.05

Note: We lost 1 sample datum in using lag-1 data as explanatory variable.

- Test statistic $DW = \frac{\sum_{i=2}^{n}(\varepsilon_i - \varepsilon_{i-1})^2}{\sum_{i=1}^{n}\varepsilon_i^2} = \frac{0.00699943}{0.0030158} = 2.3209$
- From DW table for significance 0.025, we look up d_L and d_H, and calculate upper limits $4 - d_L$ and $4 - d_H$
- $d_L = 1.16, d_H = 1.33.\ 4 - d_H = 2.67,\ 4 - d_L = 2.84$
- Since $d_H < DW < 4 - d_H$, we do not reject H_0 and conclude errors have NO autocorrelation at 5% level.
- There, autocorrelation removed by AR(1) model!

Let's check: for Singapore Old Age People AR(1) model errors:

$r_{\varepsilon_t,\varepsilon_{t-1}} = -0.17063$

$2\left(1 - r_{\varepsilon_t,\varepsilon_{t-1}}\right) = 2(1 - -0.17063)$

$= 2.3413$

Actual $DW = 2.3209$