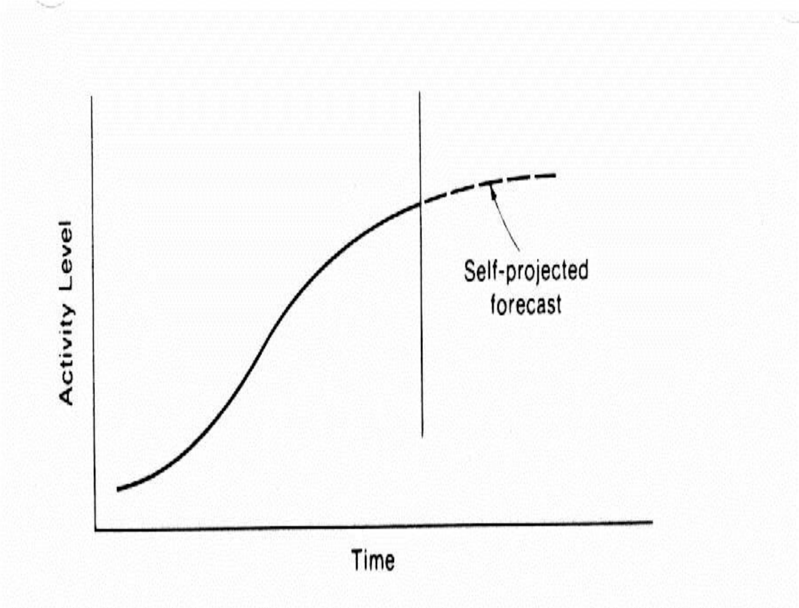
ADVANCED ANALYTICS

Module 9 – Forecasting (Transfer Function)

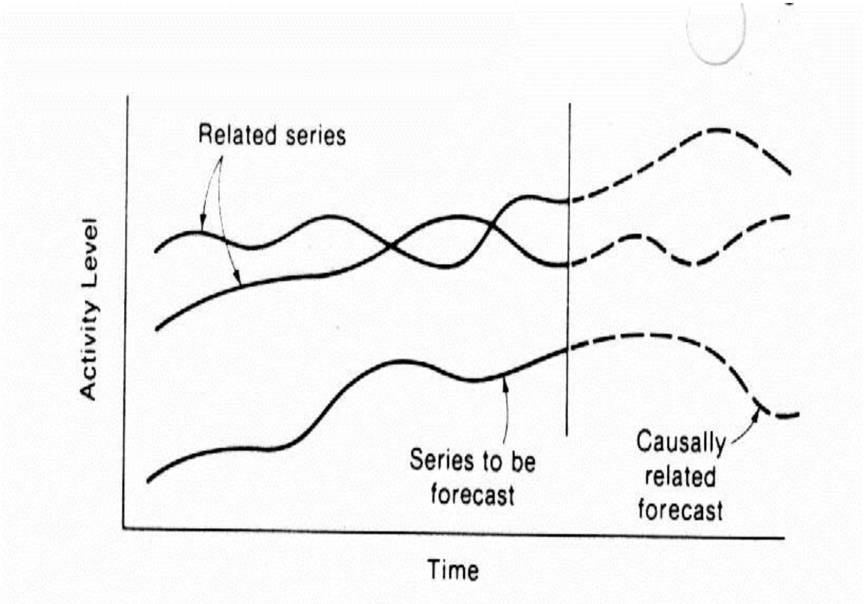
Ms Catherine Khaw
Email: Catherine@dnacapitals.com

Time Series Approach

- Self-projecting approach (univariate)



- Cause-and-effect approach (multivariate)



Introduction

- Extending from univariate time series which is self projecting we shall extend to analyse two or more time series.

$$Y_t \sim X_t$$

- Transfer function models is a ***unidirectional*** relationship between input and output.
- Cointegration provide bi-directional relationship between input and output.

Describe the following relationship between x & y

1. $y_t = a + bx_t + e_t$

2. $y_t = a_0 + b_1x_{t-1} + b_2x_{t-2} + b_3x_{t-3} + e_t$

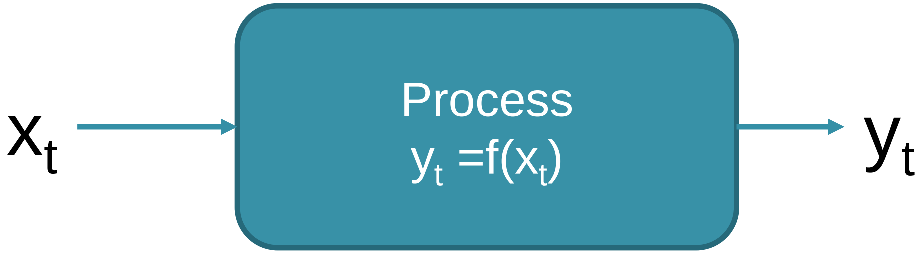
3. $y_t = a_0 + a_1 y_{t-1} + a_2 y_{t-2} + b_1x_{t-1} + b_2x_{t-2} + b_3x_{t-3} + e_t$

Transfer Function Models

- Functional relationship between input (x) and output (y) with *reference to time.*

$$y_t = f(x_t) = a + bx_t + e_t$$

- Respond y may not react immediately to the change of input x, but in a delayed fashion. Such complex relation is called *dynamic transfer functions*.



Application of Transfer Functions

- Transfer functions have wide application in engineering.
  - For example :
    - Gas input and carbon dioxide output

[class:]Give example of transfer functions application in business/financial/economic.

Dynamic Transfer Function

y reacts in a delayed fashion when input x changed.

$$y_t = v_0 x_{t-0} + v_1 x_{t-1} + v_2 x_{t-2} + v_3 x_{t-3} + \ldots + v_n x_{t-n} + n_t$$

$$\hat{v} = r_{\alpha\beta}(j) s_\beta \;/\; s_\alpha$$

$\hat{\alpha}_t$, $\hat{\beta}_t$ are pre-whitened input and output series. In general, the above can be written like

$$y_t = v(B)x_t + n_t$$

Question : If v(j) >0 , what does it mean?

Cross-Correlation

- If the output series, Y responds to X then we would imagine that values of X at time *t*, x_t would be correlated to the response values of Y at a certain time in the future, say *t+1.*
- Observe the below graph which plots the cross-correlations for the Gas Input and Carbon Dioxide output example, in which Gas Input affects the Carbon Dioxide output.

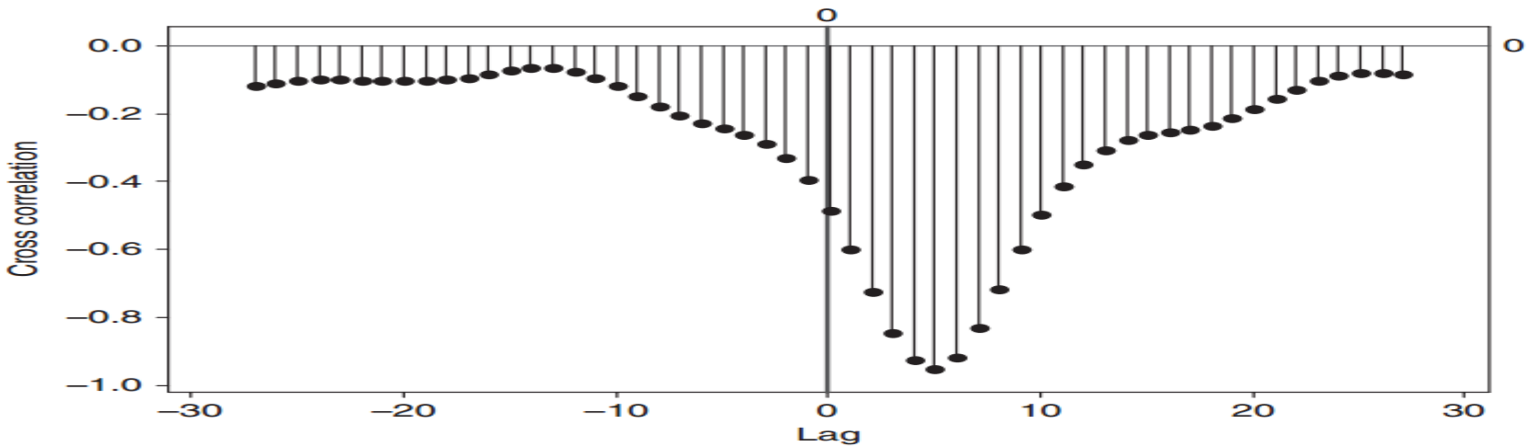


Figure 8.4 The estimated cross correlation function between the input gas feed rate and the output CO_2.

Cross-Correlation

- What is strange about the previous graph?
- Hint: We should expect to see significant cross correlations for positive, but not for negative, lags
- Why?

Prewhitening

- The spurious relationship we observed earlier is due to autocorrelation which exists in the input series.
- If the input is autocorrelated, the effect of any change in the input itself will take some time to play out
- To avoid this kind of spurious relationship, we must “decorrelate” the input data first.
- We do this by fitting a preliminary model to the input data. We then must fit the same model, with the same coefficients for the output data if we want to compare apples to apples.
- Thus, by fitting a preliminary model, we can get rid of the autocorrelation.
- This process is called prewhitening

Example

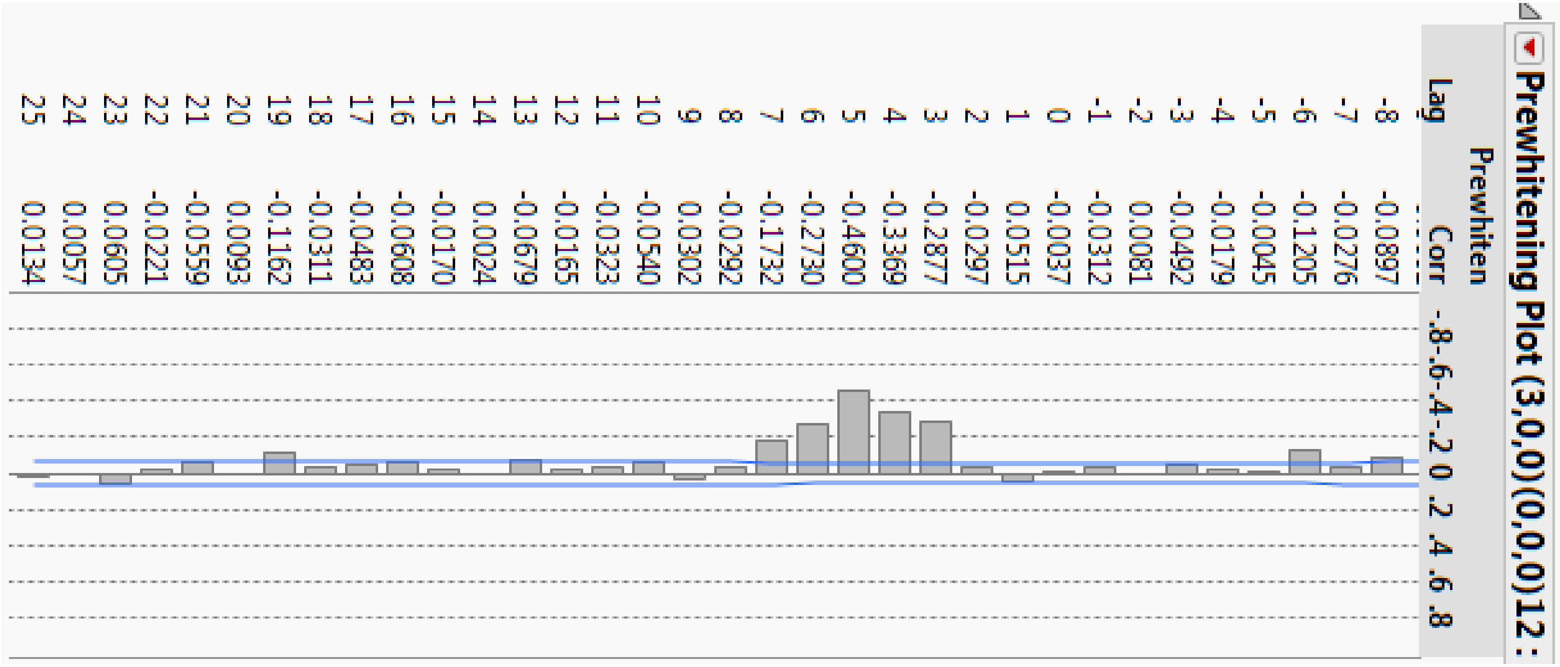
- If we check the Time Series diagnostics for the Input and look at the PACF, this is what we find:

Time Series Basic Diagnostics

Lag	AutoCorr	Ljung-Box Q	p-Value	Lag	Partial
0	1.0000	.	.	0	1.0000
1	0.9525	271.265	<.0001*	1	0.9525
2	0.8341	479.996	<.0001*	2	-0.7880
3	0.6819	619.965	<.0001*	3	0.3390
4	0.5312	705.215	<.0001*	4	0.1212
5	0.4075	755.551	<.0001*	5	0.0590
6	0.3182	786.348	<.0001*	6	-0.1115
7	0.2602	807.011	<.0001*	7	0.0486
8	0.2275	822.865	<.0001*	8	0.0994
9	0.2131	836.818	<.0001*	9	0.0159
10	0.2083	850.204	<.0001*	10	-0.0697
11	0.2028	862.939	<.0001*	11	-0.0943
12	0.1893	874.067	<.0001*	12	0.0414
13	0.1673	882.793	<.0001*	13	0.0880
14	0.1375	888.708	<.0001*	14	-0.1353
15	0.1048	892.155	<.0001*	15	0.0514
16	0.0754	893.947	<.0001*	16	0.0328
17	0.0520	894.803	<.0001*	17	-0.0223
18	0.0371	895.238	<.0001*	18	0.0340
19	0.0340	895.606	<.0001*	19	0.0934
20	0.0424	896.180	<.0001*	20	-0.0408
21	0.0564	897.198	<.0001*	21	-0.0907
22	0.0694	898.748	<.0001*	22	0.0380
23	0.0765	900.641	<.0001*	23	0.0358
24	0.0761	902.519	<.0001*	24	0.0060
25	0.0701	904.119	<.0001*	25	-0.0050

Example

- Since there is a distinct drop in the PACF values after lag 3, we can continue to plot an AR(3) model initially for Pre-whitening.
- Now, look at the below graph after pre-whitening



Example

- Correlations exist only in 5 lags, 3,4,5,6 and 7
- Other correlations look relatively insignificant
- This suggests that our transfer function equation will have terms related to the input series only for lags 3,4,5,6 and 7
- Hence, our transfer function model will look something like the below:

$$y_t = v_3 x_{t-3} + v_4 x_{t-4} + v_5 x_{t-5} + v_6 x_{t-6} + v_7 x_{t-7} + n_t$$

Where $v_j = r_{\alpha\beta}(j) * s_\beta \; / \; s_\alpha$

$r_{\alpha\beta}(j)$ is the estimated cross correlation between the prewhitened input and the prewhitened output

s_β and s_α are the estimated standard deviation of the prewhitened output and the prewhitened input respectively

Another way of writing the Transfer Function Model

- Since, in general there could be a lot of terms of the input coefficient, a better way to represent the model is below:

$$y_t = v(B)x_t + n_t$$

$$v(B) \approx \frac{\omega(B)}{\delta(B)} B^b$$

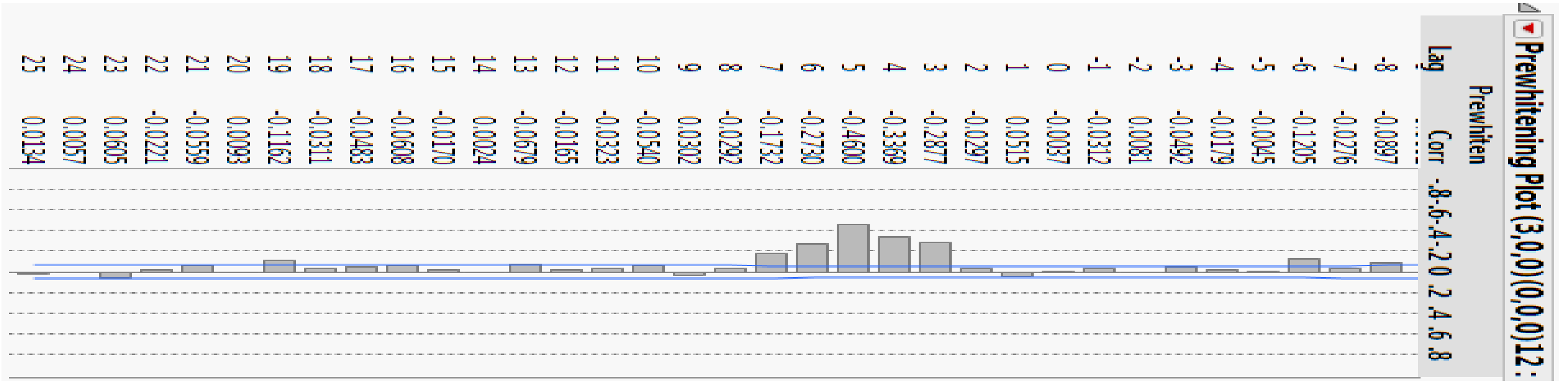
- Where,

$\omega(B) = \omega_0 - \omega_1 B + \cdots - \omega_s B^s$ and $\delta(B) = 1 - \delta_1 B - \cdots - \delta_r B^r$

- Here, s and r are the order of the numerator and the order of the denominator
- b is the lag at which the first non-zero correlation starts

Estimating the parameters

- Let's look at the graph obtained after prewhitening:



- Here, the first significantly non-zero autocorrelation occurs at lag 3. Thus, b is equal to 3
- Also, the values exhibit exponential decay after lag 5, which suggests that s = 5 – 3 = 2
- Usually r is either 1 or 2. Start with r = 2 and check model fit and reduce to 1 if needed

Key Steps to Identify Transfer Functions

Step 1
Fit ARIMA model to input series x_t

Example : AR(3) => $x_t = \Phi_1 x_{t-1} + \Phi_2 x_{t-2} + \Phi_3 x_{t-3} + \alpha_t$

(α_t assumed to be white noise)

Step 2
Prewhiten the input series x_t
(ie compute the *residuals* for x_t)

$\alpha_t = x_t - \Phi_1 x_{t-1} - \Phi_2 x_{t-2} - \Phi_3 x_{t-3}$

Step 3
Prewhiten the output series y_t
using the same model fitted on x_t

$\beta_t = y_t - \Phi_1 y_{t-1} - \Phi_2 y_{t-2} - \Phi_3 y_{t-3}$

Step 4
Compute the cross correlation between the *prewhitened x_t and y_t series*

Correlate β_t & α_t

Step 5
Compute transfer functions weights v_j

Estimate $\hat{v}_j = r_{\alpha\beta}(j) s_\beta \,/\, s_\alpha$

Develop Transfer Functions Model

Step 1
Fit ARIMA model to input series x_t

↓

Step 2
Prewhiten the input series x_t
(ie compute the *residuals* for x_t)

↓

Step 3
Prewhiten the input series y_t
using the same model fitted on x_t

↓

Step 4
Compute the cross correlation between the *prewhitened x_t and y_t series*

↓

Step 5
Compute transfer functions weights v_j

BJK Gas Furnace data

Gas Feed rate$_t$ → Process $y_t = f(x_t)$ → $CO_{2(t)}$

Develop Transfer Functions Model

Step 1
Fit ARIMA model to input series x_t

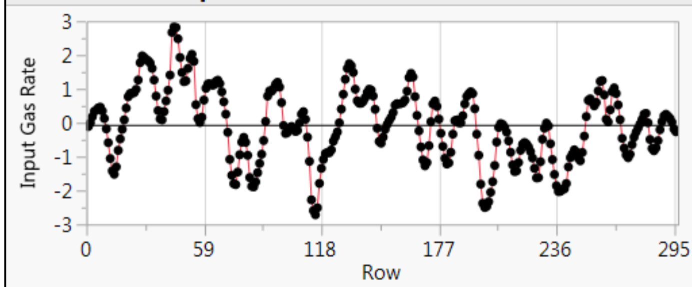
Step 2
Prewhiten the input series x_t
(ie compute the *residuals* for x_t)

Step 3
Prewhiten the input series y_t
using the same model fitted on x_t

Step 4
Compute the cross correlation between
the *prewhitened x_t and y_t series*

Step 5
Compute transfer functions weights v_j

Time Series Input Gas Rate



Mean	-0.056834
Std	1.0709519
N	296
Zero Mean ADF	-2.664823
Single Mean ADF	-2.665476
Trend ADF	-2.824178

Model: AR(3)

Model Summary

DF	292	Stable	Yes
Sum of Squared Errors	10.4475766	Invertible	Yes
Variance Estimate	0.03577937		
Standard Deviation	0.18915436		
Akaike's 'A' Information Criterion	-137.13781		
Schwarz's Bayesian Criterion	-122.37637		
RSquare	0.96919388		
RSquare Adj	0.96887738		
MAPE	.		
MAE	0.13106796		
-2LogLikelihood	-145.13781		

Parameter Estimates

Term	Lag	Estimate	Std Error	t Ratio	Prob>|t|	Constant Estimate
AR1	1	1.969061	0.0543765	36.21	<.0001*	-0.0034434
AR2	2	-1.365134	0.0985395	-13.85	<.0001*	
AR3	3	0.339403	0.0543354	6.25	<.0001*	
Intercept	0	-0.060762	0.1844560	-0.33	0.7421	

Develop Transfer Functions Model

Step 1
Fit ARIMA model to input series x_t

Step 2
Prewhiten the input series x_t
(ie compute the *residuals* for x_t)

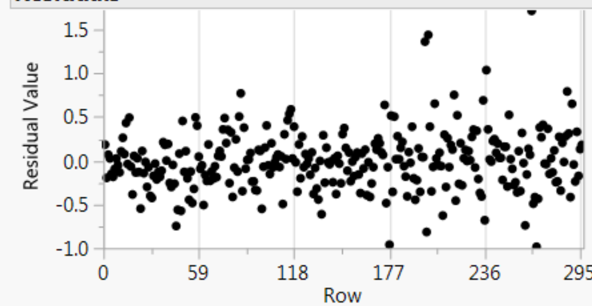
Step 3
Prewhiten the input series y_t
using the same model fitted on x_t

Step 4
Compute the cross correlation between
the *prewhitened x_t and y_t series*

Step 5
Compute transfer functions weights v_j

Model: AR(3)

Residuals



Lag	AutoCorr	-.8-.6-.4-.2 0 .2 .4 .6 .8	Ljung-Box Q	p-Value	Lag	Partial	-.8-.6-.4-.2 0 .2 .4 .6 .8
0	1.0000		.	.	0	1.0000	
1	-0.0921		2.5391	0.1111	1	-0.0921	
2	0.1963		14.1056	0.0009*	2	0.1895	
3	-0.0518		14.9131	0.0019*	3	-0.0206	
4	-0.0405		15.4095	0.0039*	4	-0.0869	
5	-0.1057		18.7946	0.0021*	5	-0.1063	
6	0.0838		20.9286	0.0019*	6	0.0979	
7	-0.0179		21.0264	0.0037*	7	0.0351	
8	0.0443		21.6283	0.0057*	8	-0.0017	
9	-0.0755		23.3805	0.0054*	9	-0.0932	
10	0.0616		24.5527	0.0063*	10	0.0486	
11	-0.0052		24.5610	0.0106*	11	0.0620	
12	0.1010		27.7307	0.0061*	12	0.0848	

Parameter Estimates

Term	Lag	Estimate	Std Error	t Ratio	Prob>|t|	Constant Estimate
AR1	1	1.969061	0.0543765	36.21	<.0001*	-0.0034434
AR2	2	-1.365134	0.0985395	-13.85	<.0001*	
AR3	3	0.339403	0.0543354	6.25	<.0001*	
Intercept	0	-0.060762	0.1844560	-0.33	0.7421	

Develop Transfer Functions Model

Step 1
Fit ARIMA model to input series x_t

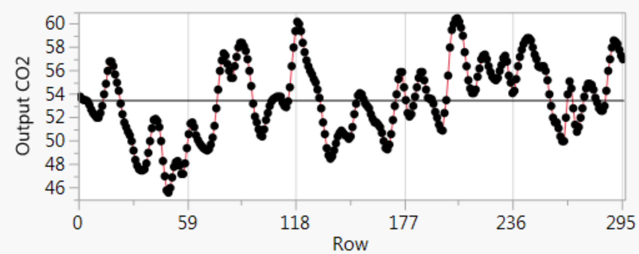
Step 2
Prewhiten the input series x_t
(ie compute the *residuals* for x_t)

Step 3
Prewhiten the input series y_t
using the same model fitted on x_t

Step 4
Compute the cross correlation between
the *prewhitened* x_t *and* y_t *series*

Step 5
Compute transfer functions weights v_j

Time Series Output CO2



Mean	53.509122
Std	3.1967072
N	296
Zero Mean ADF	0.137862
Single Mean ADF	-1.863957
Trend ADF	-2.257128

Time Series Output CO2

Model: AR(3)

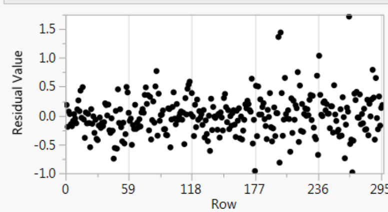
Model Summary

Parameter Estimates

Term	Lag	Estimate	Std Error	t Ratio	Prob>|t|	Constant Estimate
AR1	1	2.19621	0.0512770	42.83	<.0001*	1.46625292
AR2	2	-1.68410	0.0963793	-17.47	<.0001*	
AR3	3	0.46054	0.0513752	8.96	<.0001*	
Intercept	0	53.61491	0.7062130	75.92	<.0001*	

Forecast

Residuals



Lag	AutoCorr	Ljung-Box Q	p-Value	Lag	Partial
0	1.0000	.	.	0	1.0000
1	-0.0921	2.5391	0.1111	1	-0.0921
2	0.1963	14.1056	0.0009*	2	0.1895
3	-0.0518	14.9131	0.0019*	3	-0.0206
4	-0.0405	15.4095	0.0039*	4	-0.0869
5	-0.1057	18.7946	0.0021*	5	-0.1063
6	0.0838	20.9286	0.0019*	6	0.0979
7	-0.0179	21.0264	0.0037*	7	0.0351
8	0.0443	21.6283	0.0057*	8	-0.0017

Develop Transfer Functions Model

Step 1
Fit ARIMA model to input series x_t

Step 2
Prewhiten the input series x_t
(ie compute the *residuals* for x_t)

Step 3
Prewhiten the input series y_t
using the same model fitted on x_t

Step 4
Compute the cross correlation between the *prewhitened* x_t *and* y_t *series*

Step 5
Compute transfer functions weights v_j

Prewhitening Plot (3,0,0)(0,0,0)12 : O

Lag	Prewhiten Corr
-25	0.0177
-24	-0.0288
-23	-0.0245
-22	-0.0467
-21	0.0290
-20	-0.0620
-19	0.0377
-18	
-17	
-16	
-15	
-14	0.0124
-13	0.0125
-12	-0.0048
-11	0.0065
-10	0.0215
-9	0.0001
-8	-0.0897
-7	-0.0276
-6	-0.1205
-5	-0.0045
-4	-0.0179
-3	-0.0492
-2	0.0081
-1	-0.0312
0	-0.0037
1	0.0515
2	-0.0297
3	-0.2877
4	-0.3369
5	-0.4600
6	-0.2730
7	-0.1732
8	-0.0292
9	0.0302
10	-0.0540
11	-0.0323
12	0.0165

$$y_t = v_3x_{t-3} + v_4x_{t-4} + v_5x_{t-5} + v_6x_{t-6} + v_7x_{t-7} + n_t$$

Cross correlation between prewhitened input and output indicate significant correlation between lag 3 to 7

Develop Transfer Functions Model

Step 1
Fit ARIMA model to input series x_t

Step 2
Prewhiten the input series x_t
(ie compute the *residuals* for x_t)

Step 3
Prewhiten the input series y_t
using the same model fitted on x_t

Step 4
Compute the cross correlation between
the *prewhitened x_t and y_t series*

Step 5
Compute transfer functions weights v_j

Parameter to identify :
s, r, b, p, P, q, Q

b= v(b) >0 from step 4 : b=3
s = peak of corr before decaying - b = 5-3 = 2
r = 1 if single decaying or 2 if two decaying terms

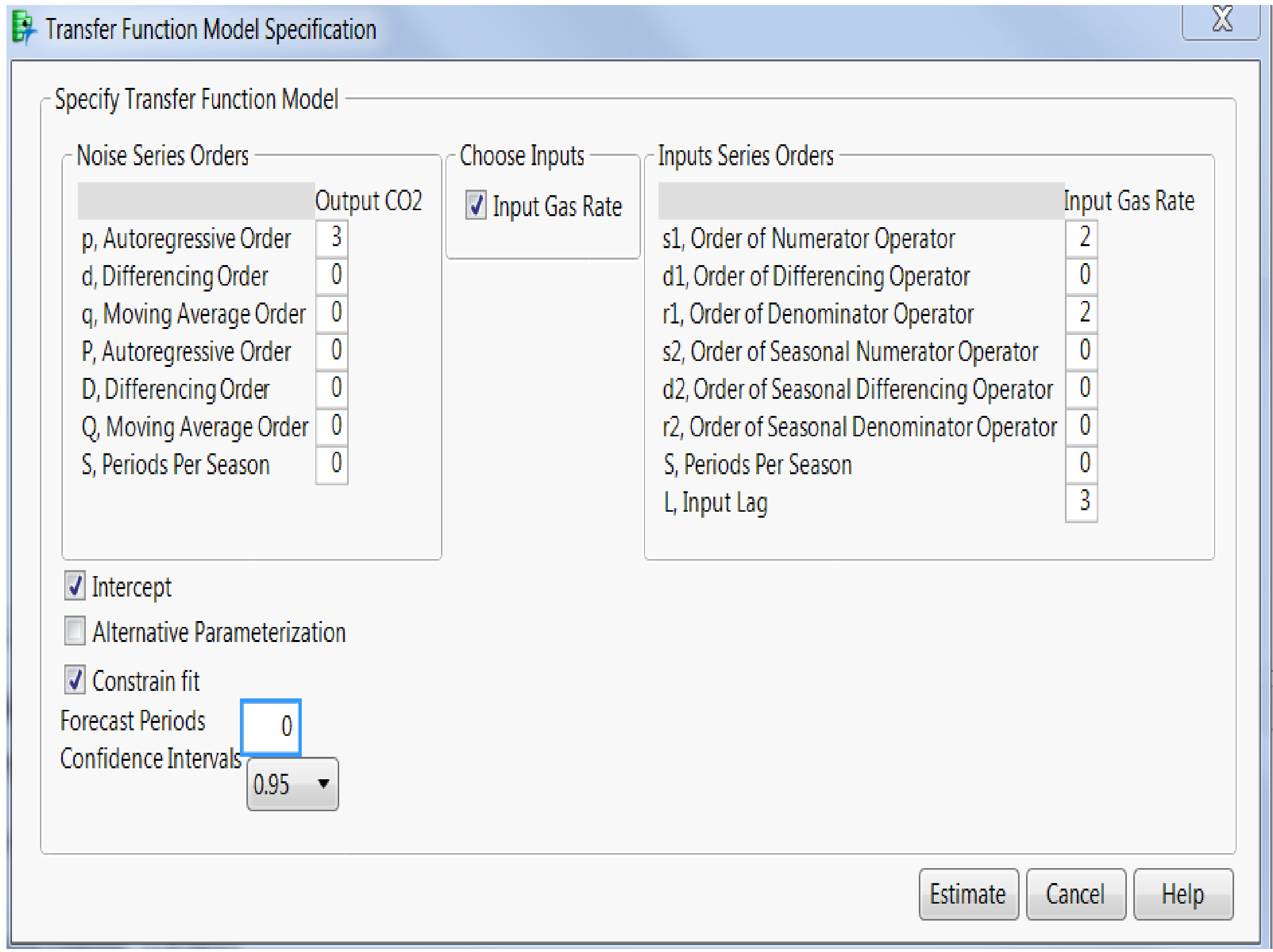
(p,P, q, Q) at the noise (residual check)

Develop Transfer Functions Model

Step 6
Fitting transfer function-noise model (this function can be found in JMP/SAS)*

Step 7
Diagnostic Checks

Model Comparison



Model checking

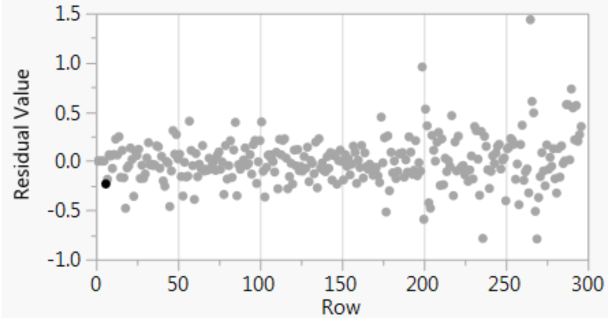
1) Check residuals.

Step 6
Fitting transfer function-noise model (this function can be found in JMP/SAS)*

Step 7
Diagnostic Checks

Model Comparison

Residuals



Lag	AutoCorr	Ljung-Box Q	p-Value	Lag	Partial
0	1.0000	.	.	0	1.0000
1	0.0010	0.0003	0.9868	1	0.0010
2	0.0806	1.9512	0.3770	2	0.0806
3	-0.0605	3.0534	0.3835	3	-0.0611
4	-0.0458	3.6878	0.4499	4	-0.0525
5	-0.0650	4.9669	0.4199	5	-0.0555
6	0.1092	8.5933	0.1978	6	0.1156
7	0.0142	8.6549	0.2784	7	0.0181
8	0.0313	8.9547	0.3461	8	0.0029
9	-0.0817	11.0047	0.2754	9	-0.0812
10	0.0566	11.9936	0.2855	10	0.0667
11	0.0253	12.1918	0.3494	11	0.0581
12	0.1103	15.9714	0.1926	12	0.0870
13	-0.0406	16.4852	0.2239	13	-0.0583
14	0.0573	17.5131	0.2299	14	0.0387
15	-0.0956	20.3829	0.1577	15	-0.0527
16	-0.0043	20.3887	0.2032	16	-0.0058

Model checking

1) Check residuals.
2) Check significant of parameters

Step 6
Fitting transfer function-noise model (this function can be found in JMP/SAS)*

Step 7
Diagnostic Checks

Model Comparison

Model Summary

DF	282
Sum of Squared Errors	16.5906797
Variance Estimate	0.0588322
Standard Deviation	0.24255349
Akaike's 'A' Information Criterion	13.3830428
Schwarz's Bayesian Criterion	46.4429522
RSquare	0.99451513
RSquare Adj	0.99436224
MAPE	0.30834571
MAE	0.16565197
-2LogLikelihood	-4.6169523

Failed: Cannot Decrease Objective Function

Parameter Estimates

Variable	Term	Factor	Lag	Estimate	Std Error	t Ratio	Prob>|t|
Input Gas Rate	Num0,0	0	0	-0.52918	0.0739819	-7.15	<.0001*
Input Gas Rate	Num1,1	1	1	0.37507	0.1071911	3.50	0.0005*
Input Gas Rate	Num1,2	1	2	0.51166	0.1142078	4.48	<.0001*
Input Gas Rate	Den1,1	1	1	0.55873	0.0787780	7.09	<.0001*
Input Gas Rate	Den1,2	1	2	-0.00827	0.0489621	-0.17	0.8660
Output CO2	AR1,1	1	1	1.55212	0.0586173	26.48	<.0001*
Output CO2	AR1,2	1	2	-0.69101	0.1012690	-6.82	<.0001*
Output CO2	AR1,3	1	3	0.04411	0.0626897	0.70	0.4823
	Intercept	0	0	53.37102	0.1477526	361.22	<.0001*

$$\text{Output CO2}_t = 53.371 + \left(\frac{\left(\left(-0.5292 - 0.3751 * B \right) - 0.5117 * B^2 \right)}{\left(\left(1 - 0.5587 * B \right) + 0.0083 * B^2 \right)} \right) * \text{Input Gas Rate}_{t-3} + \left(\frac{1}{\left(\left(\left(1 - 1.5521 * B \right) + 0.691 * B^2 \right) - 0.0441 * B^3 \right)} \right) * e_t$$

Refer handout for the expansion of the above expression

Model comparison

Step 6
Fitting transfer function-noise model (* this function can be found in JMP/SAS)

Step 7
Diagnostic Checks

Model Comparison

Model Comparison

Report	Graph	Model	DF	Variance	AIC	SBC	RSquare	-2LogLH	Weights	.2 .4 .6 .8	MAPE
☑	☐	Transfer Function Model (3)	284	0.0585185	9.877460	35.590723	0.995	-4.12254	0.648656		0.307790
☑	☐	Transfer Function Model (1)	283	0.0587246	11.874851	41.261437	0.995	-4.125149	0.238939		0.307823
☑	☐	Transfer Function Model (2)	282	0.0588322	13.383043	46.442952	0.995	-4.616957	0.112405		0.308346

Interpret the model result

Transfer Function Model (3)

Model Summary

Parameter Estimates

Variable	Term	Factor	Lag	Estimate	Std Error	t Ratio	Prob>|t|
Input Gas Rate	Num0,0	0	0	-0.53110	0.0737809	-7.20	<.0001*
Input Gas Rate	Num1,1	1	1	0.37992	0.1016023	3.74	0.0002*
Input Gas Rate	Num1,2	1	2	0.51798	0.1085541	4.77	<.0001*
Input Gas Rate	Den1,1	1	1	0.54907	0.0391928	14.01	<.0001*
Output CO2	AR1,1	1	1	1.52713	0.0467279	32.68	<.0001*
Output CO2	AR1,2	1	2	-0.62883	0.0494715	-12.71	<.0001*
	Intercept	0	0	53.36276	0.1374725	388.17	<.0001*

$$\text{Output CO2}_t = 53.3628 + \left(\frac{\left(\left(-0.5311 - 0.3799 * B \right) - 0.518 * B^2 \right)}{\left(1 - 0.5491 * B \right)} \right) * \text{Input Gas Rate}_{t-3} + \left(\frac{1}{\left(\left(1 - 1.5271 * B \right) + 0.6288 * B^2 \right)} \right) * e_t$$

Expand the model in full with backshift operator:

$$y_t \sim x_t$$

Forecasting

- Expand the model in full with backshift operator:

$$y_t \sim x_t$$

Refer to the handout for the full expansion of the above transfer function

Summary

- Understand the context between input and output series.
- Plot the two series as well as the cross correlations between the two.
- Prewhitening the series to better understand the relationship between input and output. Cross relation of the prewhitened series should help indicate the relationship between the two.
- Make an estimation for r, s, l , p,P, q ,Q
- Fitting Transfer function and conduct diagnostic check
- Compare models using AIC and SBC

References

“Time Series Analysis and Forecasting by Example” – Soren Bisgaard and Murat Kulahci

Forecasting: principles and practice – Rob J Hyndman, George Athanasopoulos

Analysis of Financial Time Series - Ruey S. Tsay