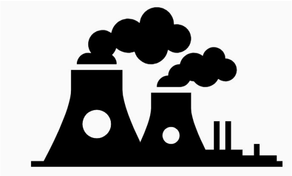
GW | Business

Time Series Analysis of Particulate Matter (PM2.5) Trends in China

Group member: Xu Jiaru, Wang Xinyu,
Wu Jingsi, Liu Yuqi, Li Weiteng

Agenda



I. Introduction and Overview

II. Deterministic Time Series Models

III. Transfer Function Model

IV. Summary and Conclusions

V. Q&A

Introduction and Overview

- **Dataset**: **Daily** data that describes the air quality of Beijing from January 1st, 2013 to December 31st, 2015. (**1095** observations)

- **Source**: Data collected from the **UCI Machine Learning Repository** and originally gathered by the U.S. Department of State.

- **Dependent Variable**: We will model the variable called ***PM_US_POST***, which indicates the PM2.5* concentration level.

- **Independent Variables**: Other Variables in our dataset include Temperature, Humidity, and Wind Speed etc.

*pm2.5:atmospheric particulate matter (PM) that have a diameter of less than 2.5 micrometers

GW | Business

Time Series Plots of Original Dataset

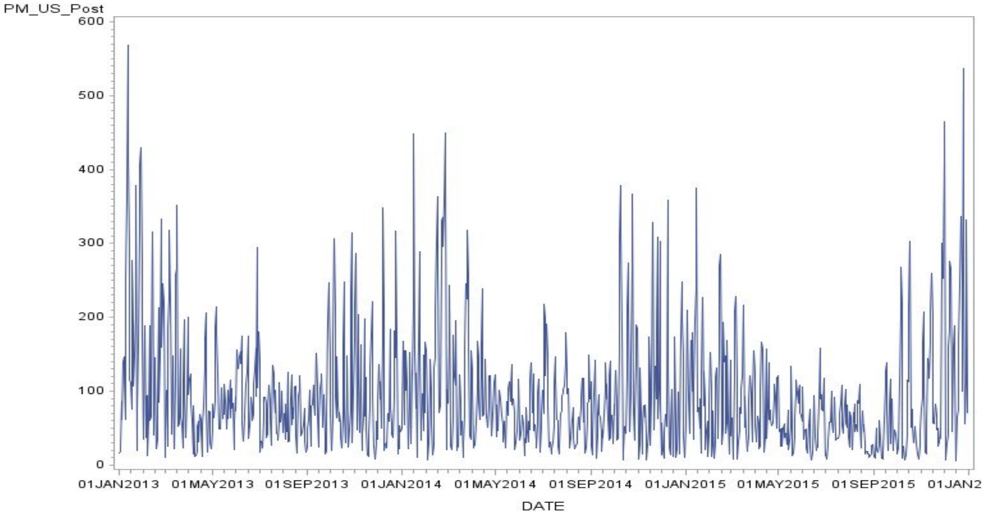


Figure 1: Time Series Plot of Original Series

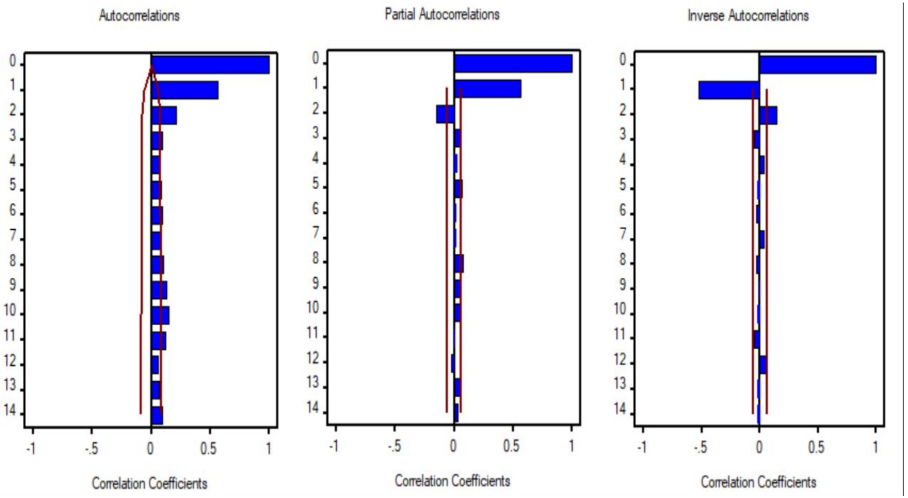


Figure 2: Autocorrelation Plot of Original Series

- ❑ Year Cycle
- ❑ Seasonality
- ❑ No clear trend

- ❑ The sample ACF is decaying fast, we can conclude that the series is stationary.

Seasonal Boxplot

- 1. Lower average in summer
- 2. Extreme high air pollution in winter
- 3. Air quality varies significantly in winter
- 4. Severer air pollution happens from October to next March

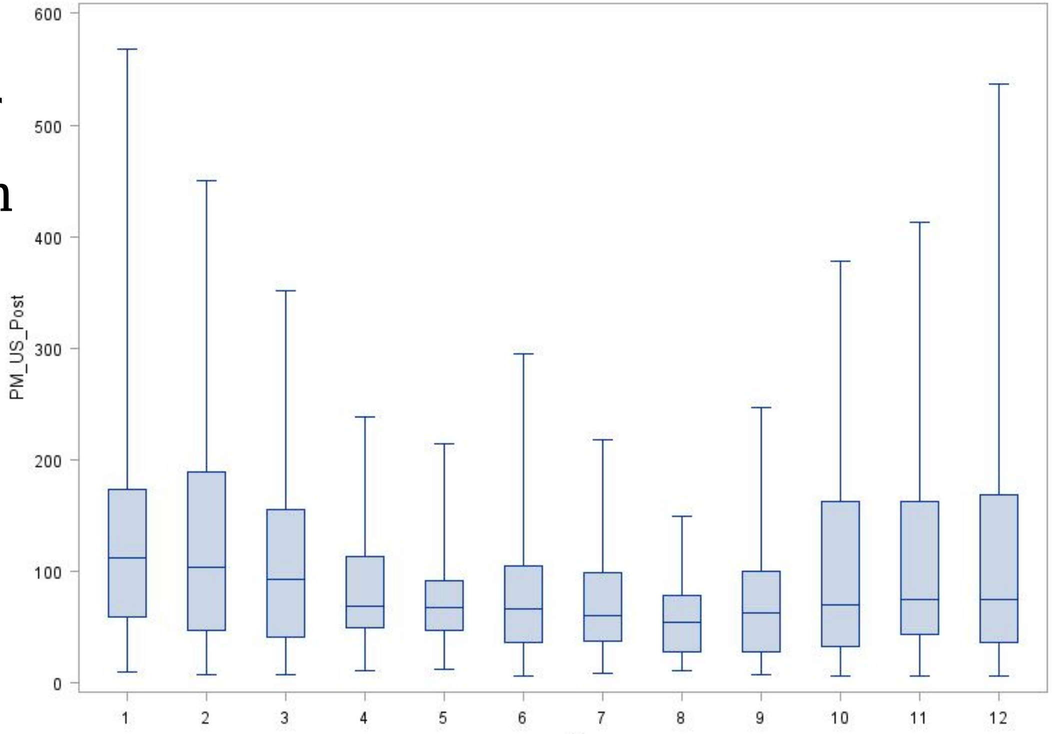


Figure 3: Monthly Boxplot of Air Pollution Series

GW Business

Deterministic Time Series Models – Seasonal Dummy Model

- ❑ Monthly dummy variables
- ❑ Winter season indicator
- ❑ No Trend

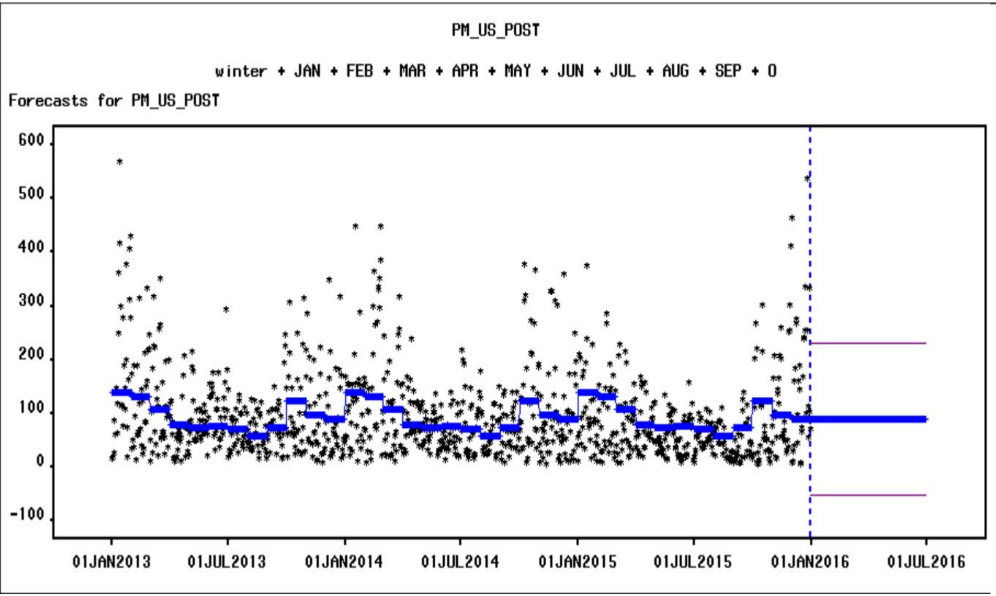


Figure 4: Actual versus Predicted Plot for Seasonal Dummy Model

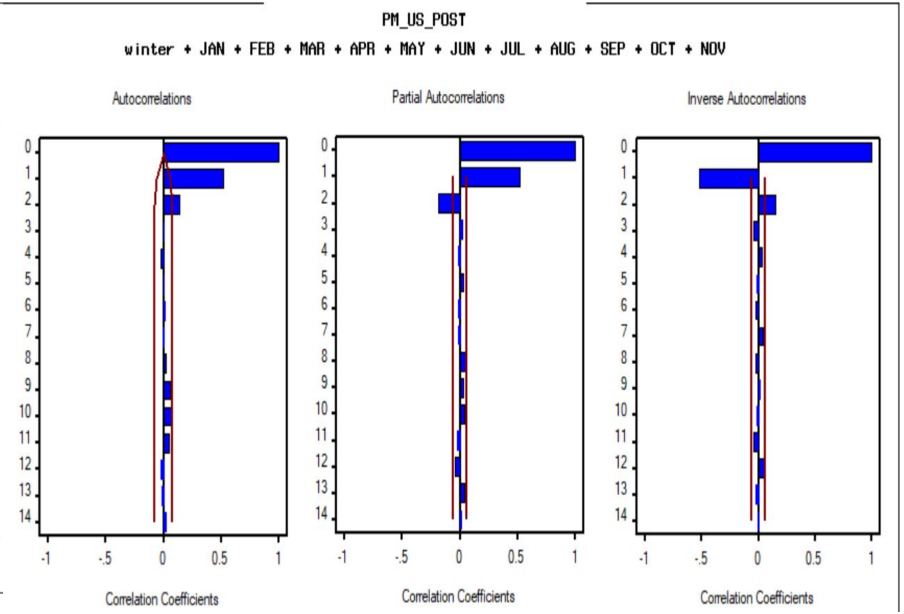


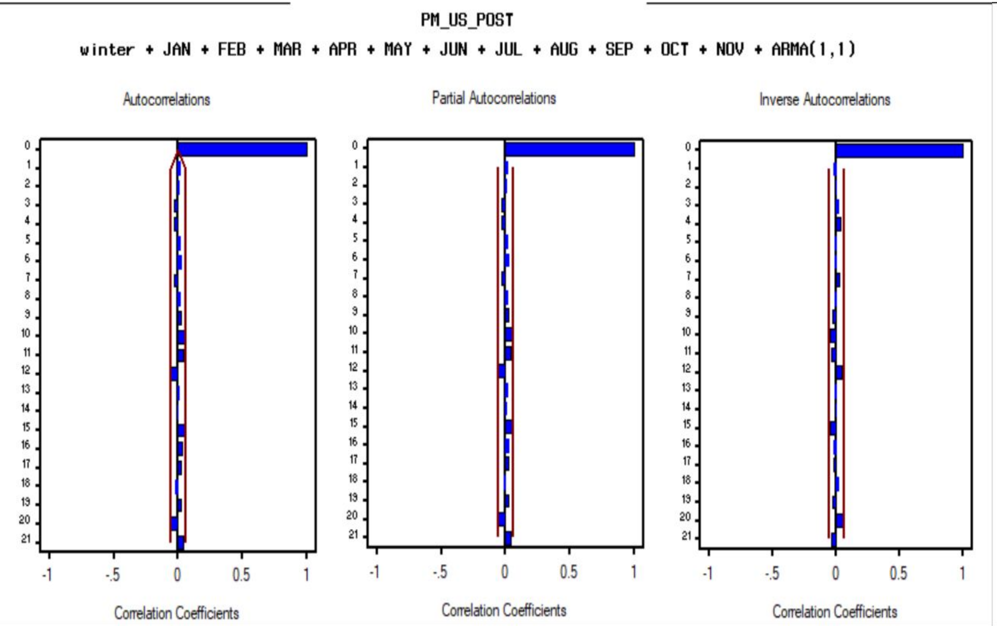
Figure 5: Autocorrelation Plot For Seasonal Dummy Model

🤔 **AR(2) or ARMA(1,1)**

ARMA(1,1)

GW Business

Deterministic Time Series Models – Seasonal Dummy Model



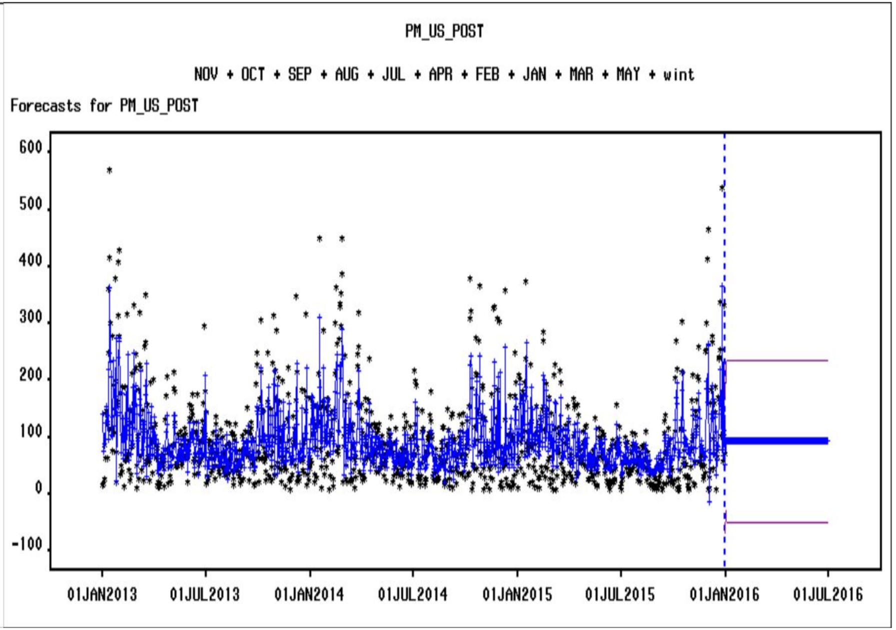


Figure 6: Autocorrelation Plot for Seasonal Dummy Model with ARMA(1,1)

Figure 7: Actual versus Predicted Plot For for Seasonal Dummy Model with ARMA(1,1)

Deterministic Time Series Models – Cyclical Trend Model

1	Obs	FREQ	PERIOD	P_01
2	4	0.01721	365	581002.55
3	6	0.02869	219	133157.46
4	68	0.38445	16.34	121564.12
5	78	0.44183	14.22	95752.81
6	124	0.70578	8.9	93205.04
7	144	0.82054	7.66	90219.3
8	3	0.01148	547.5	84971.01
9	145	0.82628	7.6	82661.63
10	127	0.723	8.69	63930.53
11	114	0.6484	9.69	63774.39
12	29	0.16067	39.11	63042.94
13	32	0.17788	35.32	62152.14
14	54	0.30412	20.66	62103.19
15	147	0.83776	7.5	61966.58
16	188	1.07302	5.86	60499.38

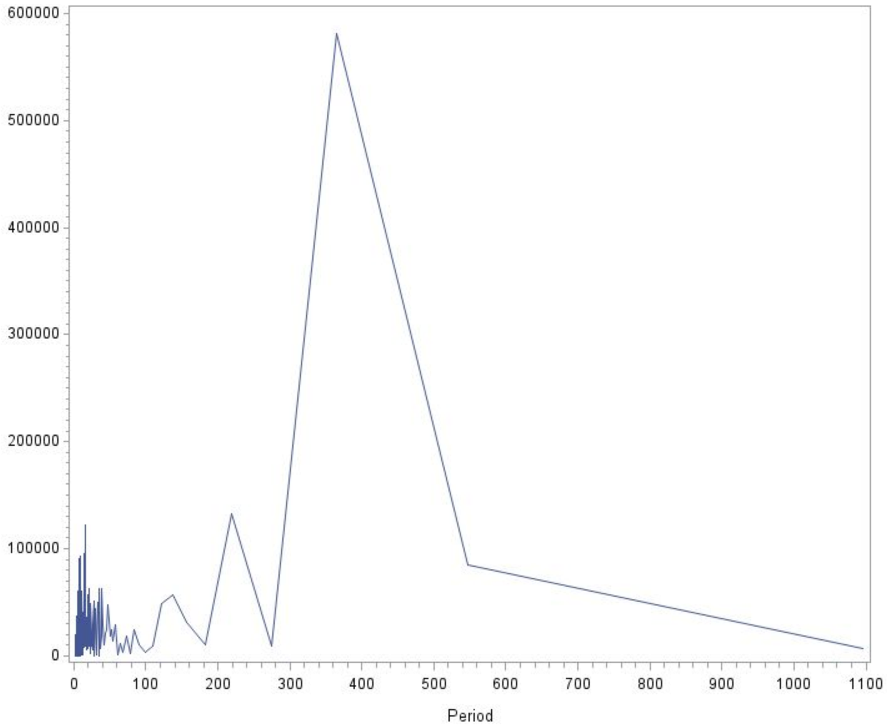


Figure 8: Periodogram Plot for PM2.5 Series

Deterministic Time Series Models – Cyclical Trend Model

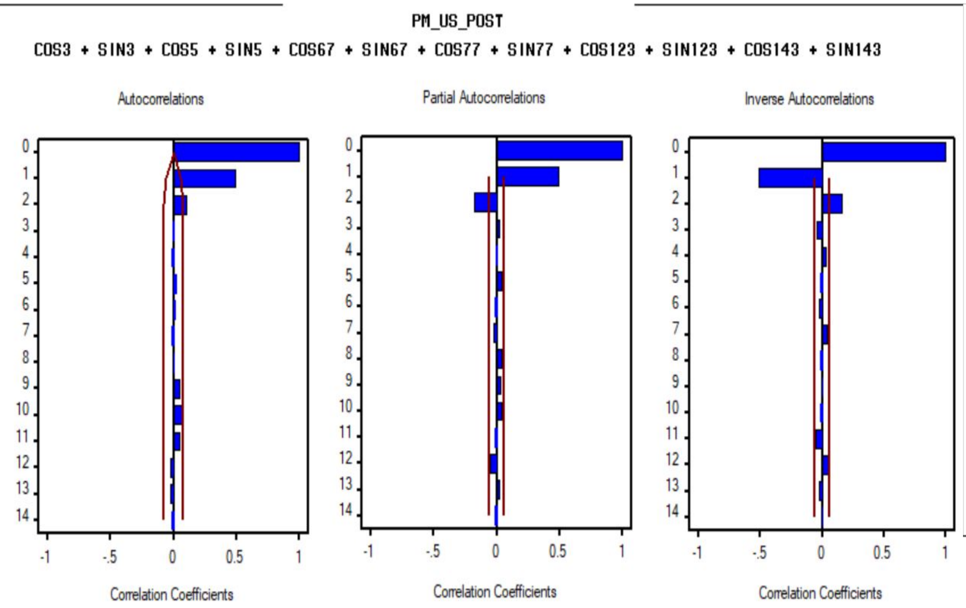


Figure 9: Autocorrelation Plot for Cyclical Trend Model

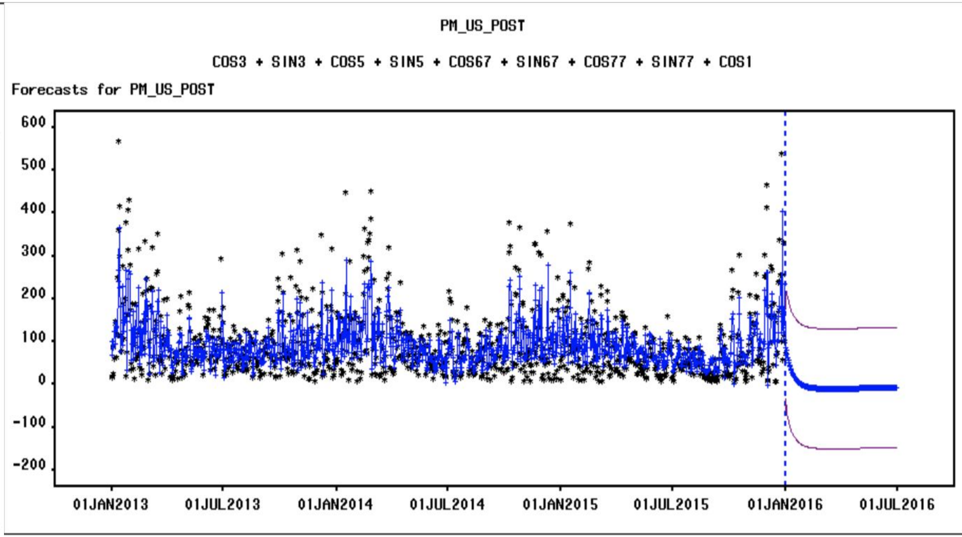


Figure 10: Actual versus Predicted Plot for Cyclical Trend Model with ARMA(1,1)

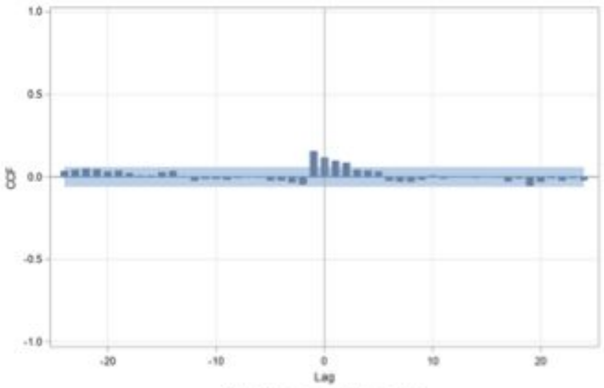
AR(2) or ARMA(1,1)

ARMA(1,1)

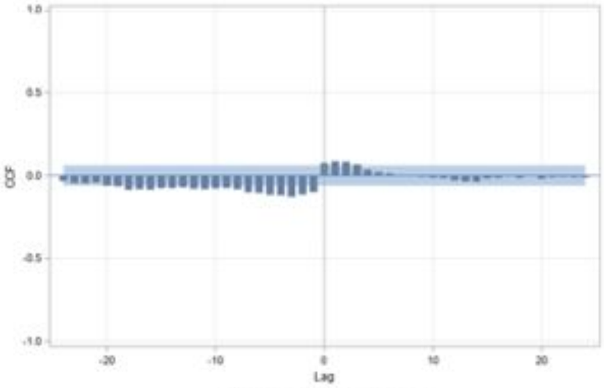
GW Business

Transfer Function Model – Identification

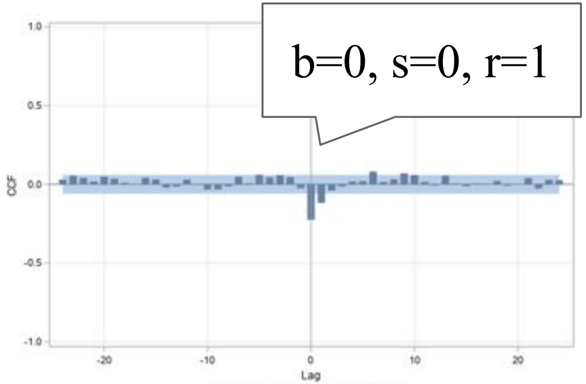
Temperature



Humidity

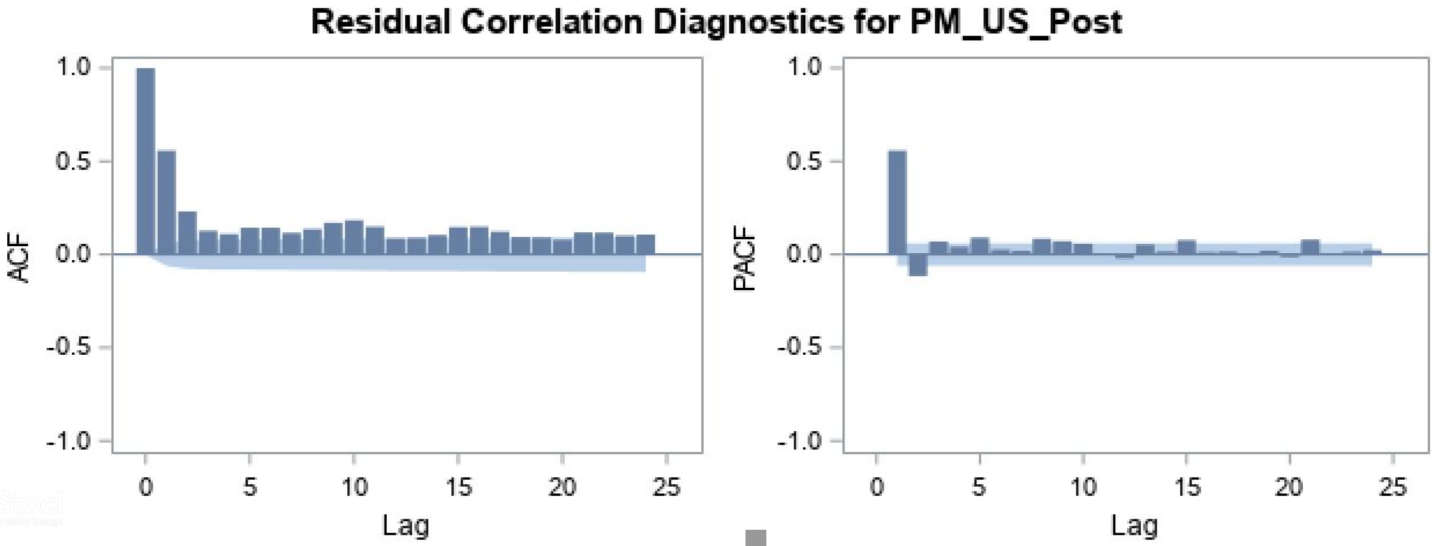


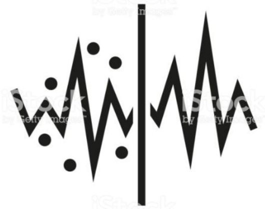
Wind Speed



Transfer Function Model – Identification

AR(2)





TF–noise model

GW Business

Transfer Function Model - Estimation

	Maximum Likelihood Estimation							
	Parameter	Estimate	Standard Error	T Value	Approx. Pr >|t|	Lag	Variable	Shift
	MU	108.9639	4.91213	22.18	<.0001	0	PM_US_POST	0
	AR1,1	0.62287	0.03011	20.69	<.0001	1	PM_US_POST	0
	AR1,2	-0.11647	0.03020	-3.86	0.0001	2	PM_US_POST	0
ω_0	NUM1	-0.44037	0.05459	-8.07	<.0001	0	Iws	0
δ_1	DEN1,1	0.39981	0.10520	3.80	0.0001	1	Iws	0

Recap

1. b=0, s=0, r=1
2. wind speed AR(1)
3. residuals AR(2)

$$(1-0.62287B+0.11647B^2)PM=108.96399+(-0.44037/(1-0.39981))\text{Wind Speed} + \epsilon$$

Transfer Function Model – Validation

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	15.68	4	0.0035	0.009	-0.034	0.033	0.015	0.079	0.074
12	38.91	10	<.0001	0.005	0.047	0.069	0.089	0.076	-0.017
18	55.21	16	<.0001	0.039	0.015	0.079	0.065	0.048	0.012
24	70.24	22	<.0001	0.053	-0.017	0.077	0.051	0.02	0.038
30	91.89	28	<.0001	0.049	0.05	0.003	0.086	0.052	0.065
36	105.42	34	<.0001	0.035	0.084	0.004	-0.027	0.021	0.05
42	116.28	40	<.0001	0.038	0.032	0.014	0.063	0.054	0.008
48	124.82	46	<.0001	-0.001	0.007	0.054	0.06	0.029	0.005

Autocorrelation Check of Residuals

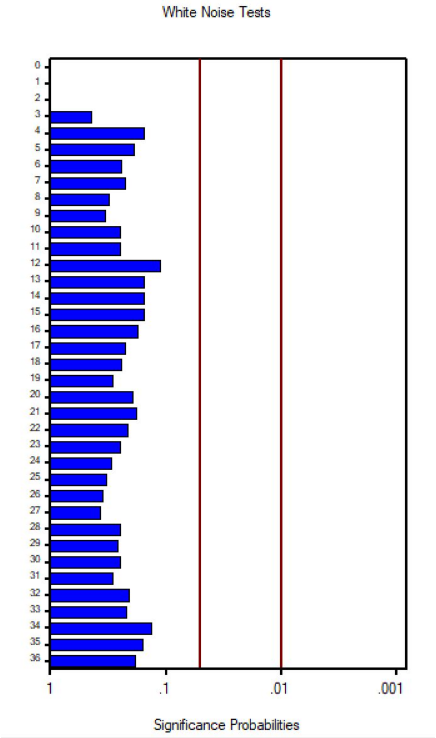
Crosscorrelation Check of Residuals with Input Iws									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
5	1.76	5	0.8806	0.001	-0.02	0.018	0.007	0.017	0.024
11	18.58	11	0.069	0.075	-0.025	0.01	0.071	0.061	-0.01
17	22.32	17	0.1726	0.001	0.058	-0.006	-0.006	0.002	-0.003
23	29.6	23	0.1612	0.036	-0.005	0.02	0.044	-0.051	0.019
29	43.06	29	0.045	0.02	0.068	0.04	0.072	-0.022	0.006
35	46.85	35	0.0869	0.019	-0.032	-0.004	0.034	0.011	0.029
41	49.05	41	0.1816	0.004	0.017	-0.005	-0.03	0.011	-0.025
47	51.75	47	0.2937	-0.008	-0.023	0.005	0.029	-0.03	-0.011

Cross Correlation Check of Residuals with Input

GW Business

Model Comparisons

Model	Model Title	RMSE	Model Variance
1	Seasonal Dummy with ARMA(1,1) and TF Model	86.76707	3525
2	Seasonal Dummy with AR(2) and TF Model	87.94336	3500
3	Seasonal Dummy Model with ARMA(1,1)	92.57457	3796
4	Cyclical Trend with ARMA(1,1) and TF Model	85.07846	3457
5	Cyclical Trend with AR(2) and TF Model	88.39528	3396
6	Cyclical Trend with ARMA(1,1) Model	87.15351	3724



In general, cyclical trend model has a better performance than seasonal dummy model. The choice of best model varies as we select different criteria for evaluation and comparison.

The dynamic regressor of wind speed provides a more significant improvement in RMSE for seasonal dummy model.

GW Business

Summary and Conclusions

- **Monthly** variation with large values occur mainly in winter
- No significant trend within the series
- **Cyclical trend model** captures periods with different length such as year, quarter and month.
- Only **wind speed** is eligible for a transfer function model
- Concerning with hold out samples, **Cyclical Trend with ARMA(1,1) and TF Model** provides the best fit

Thanks!

Q & A