Time Series Final Report

Group Members:

Xinyi Wang

Xin Yuan

Youdan Shi

Yunning Zhu

Yuanjing Han

1.Data overviews:

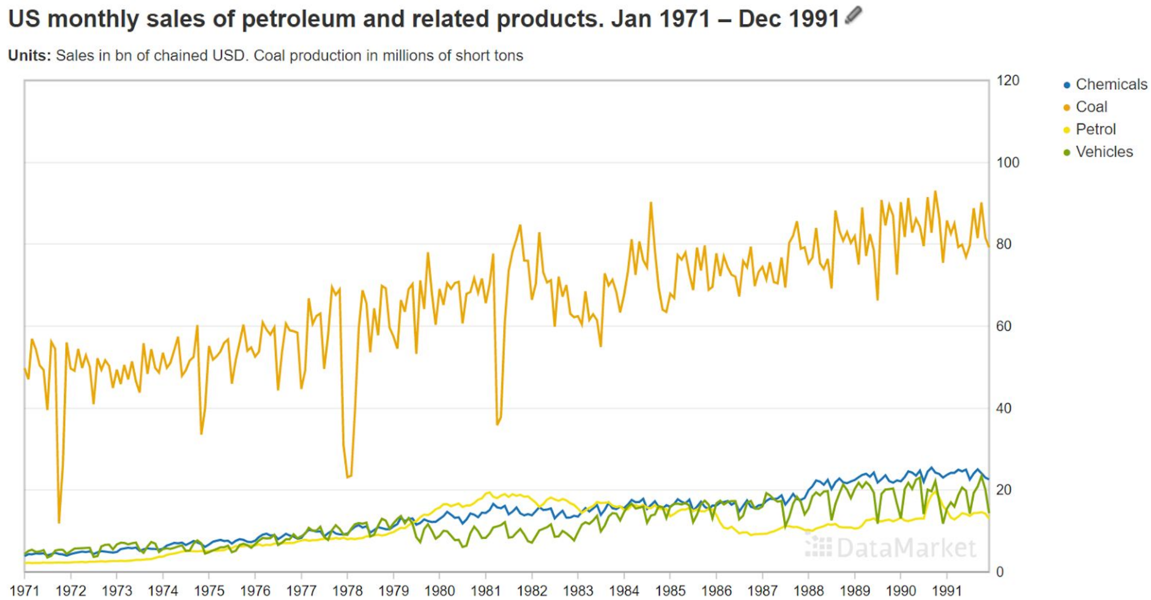


Fig. 1 US monthly sales of petroleum and related products. Jan 1971 - Dec 1991

This dataset contains US monthly sales of petroleum and related products from Jan 1971 to Dec 1991 and the sales is in billions of chained USD and coal production is in millions of short tons (See Fig. 1). Generally, coal production has more clearly increasing trend with some extreme volatility than the sales of other products. As for the sales of the three products, they increased stably with some small ups and downs until the summer of 1979.

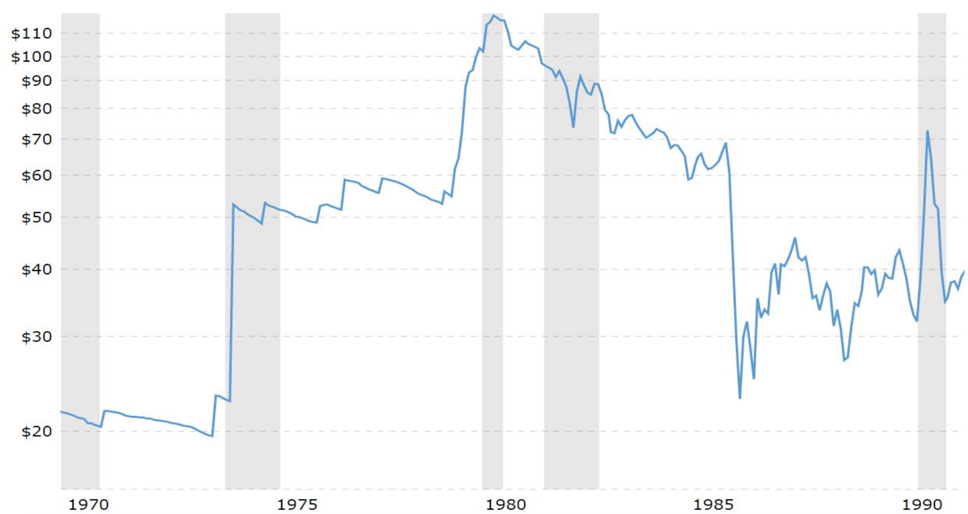


Fig. 2 Real price of crude oil. Jan 1971 - Dec 1991.

This series is about the real price of crude oil, called “real” (see Fig. 2). Based on this original time series plot we can see that there is a sharp increase at around 1980s in Real. After checking these series, we are interested in variables: sales of petrol, sales of vehicles and the real price of crude oil. We identify petrol sales as our main series and Real(real price of crude oil) as our input variable.

2. Univariate models of Petrol

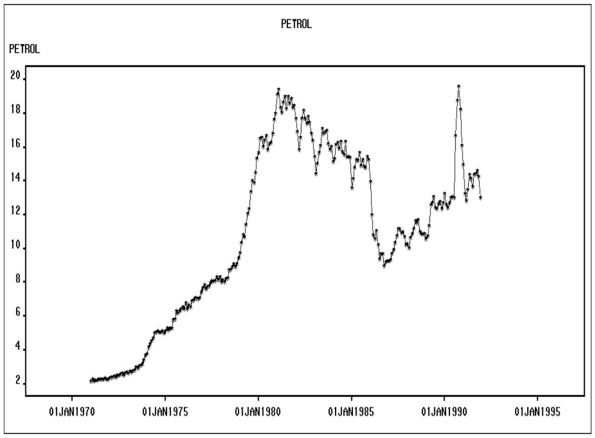


Fig. 3 Original time series of Petrol

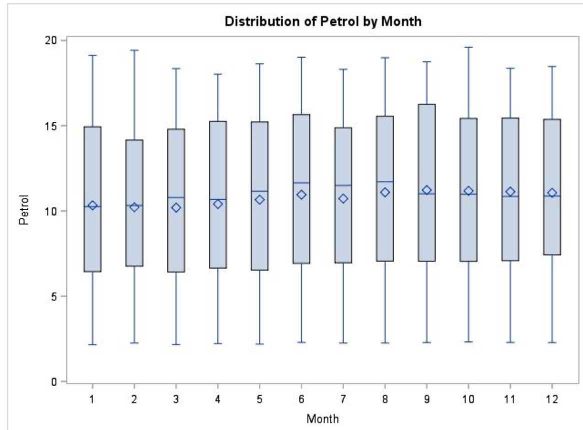


Fig. 4 Seasonal boxplot of petrol

First, we plot the original series and box plot of petrol sales (see Fig. 3 and Fig. 4). From these plots, there is no obvious seasonality and no obvious outliers. Also, the original series shows a general upward trend, which means this is a non-stationary time series. Next, we'll try several models to deal with this series.

a. Deterministic time series model

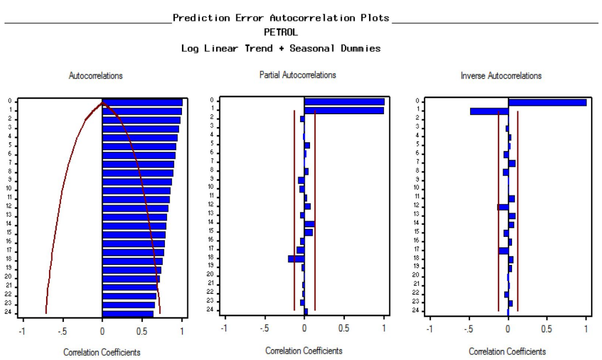


Fig. 5 ACF plot of Residuals of deterministic model

Parameter Estimates
PETROL
Log Linear Trend + Seasonal Dummies

Model Parameter	Estimate	Std. Error	T	Prob>|T|
Intercept	1.33256	0.1065	12.5134	.
Linear Trend	0.00720	0.000393	18.3072	.
Seasonal Dummy 1	-0.03186	0.1334	-0.2389	.
Seasonal Dummy 2	-0.02973	0.1334	-0.2229	.
Seasonal Dummy 3	-0.03841	0.1334	-0.2880	.
Seasonal Dummy 4	-0.02499	0.1333	-0.1874	.
Seasonal Dummy 5	-0.01524	0.1333	-0.1143	.
Seasonal Dummy 6	0.01297	0.1333	0.0972	.
Seasonal Dummy 7	-0.00866	0.1333	-0.0649	.
Seasonal Dummy 8	0.01444	0.1333	0.1083	.
Seasonal Dummy 9	0.01698	0.1333	0.1274	.
Seasonal Dummy 10	0.00404	0.1333	0.0303	.
Seasonal Dummy 11	0.00133	0.1333	0.009960	.
Model Variance (sigma squared)	0.17770	.	.	.

Fit Range: JAN1971 to DEC1990

Fig. 6 Parameter estimates of deterministic model

In this part, we transfer petrol into log petrol. Then, we add linear trend and seasonal dummies in this model. Based on SAS results, we found that residuals are non-stationary (see Fig. 5). Moreover, all the seasonal dummy variables are not statistically significant indicating that there is no seasonality in petrol (see Fig. 6). This conclusion also confirms what we were conjecturing at the beginning of this report.

b. Exponential smoothing model

Parameter Estimates

PETROL

Log Linear (Holt) Exponential Smoothing

Model Parameter	Estimate	Std. Error	T	Prob>¦T¦
LEVEL Smoothing Weight	0.99900	0.0473	21.1293	<.0001
TREND Smoothing Weight	0.04845	0.0151	3.2134	0.0093
Residual Variance (sigma squared)	0.00177	.	.	.
Smoothed Level	2.77772	.	.	.
Smoothed Trend	0.01027	.	.	.

Fig. 7 Parameter estimates of Holt’s linear exponential smoothing model

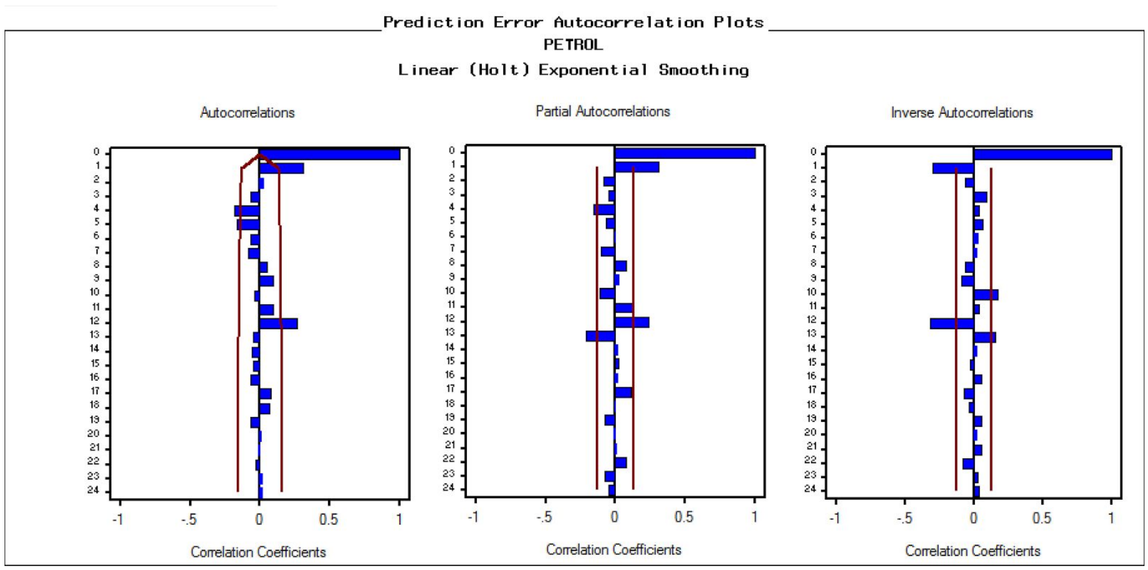
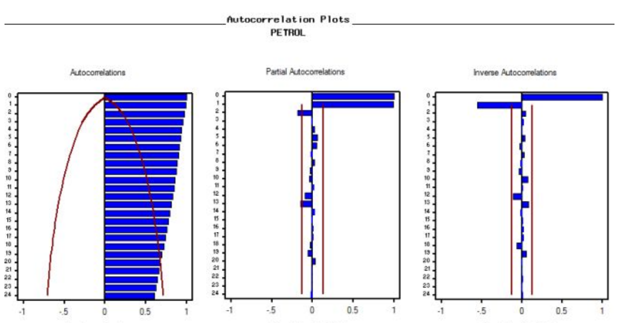


Fig. 8 ACF plot of Residuals of Holt’s linear exponential smoothing model

In this part, we use exponential smoothing to test this serie. Based on SAS result, we find that the residuals also are not white noise (see Fig. 8) although Linear Exponential Smoothing got significant parameters(see Fig. 7), which means this exponential smoothing model is not appropriate model for this variable.

c. ARIMA model



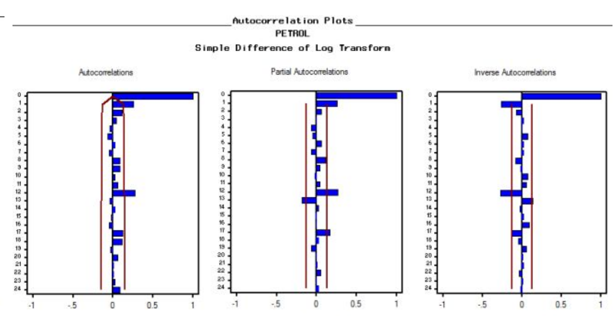
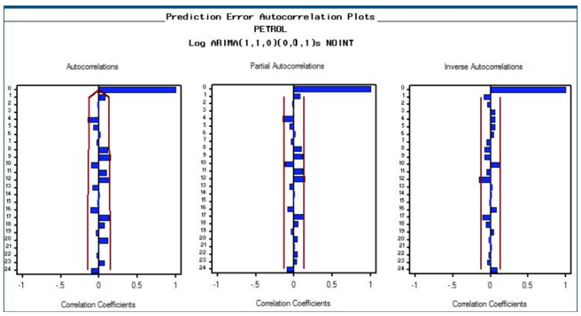


Fig. 9 ACF plot of original series

Fig. 10 ACF plot of simple difference of log transform

Firstly, from acf of origin petrol series(see Fig. 9), we could see it is not stationary. After taking the first simple difference, the series became stationary seasonally and non-seasonally since the acf chopped off after lag1 at seasonal level and decayed quickly at non-seasonal level (see Fig. 10). Besides, the pacf of the first-differenced series chopped off after lag1 at non-seasonal level and decayed quickly at seasonal level(see Fig. 10). Given the behavior of acf and pacf at both level, we derived log ARIMA(1,1,0)(0,0,1)s model without an intercept term.



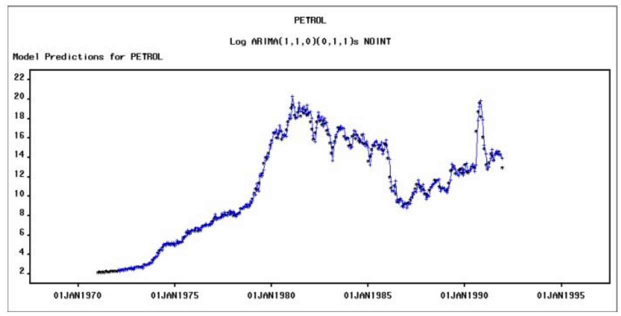


Fig. 11 ACF plot of residuals

Fig. 12 Model prediction for Petrol

Parameter Estimates

PETROL

Log ARIMA(1,1,0)(0,0,1)s NOINT

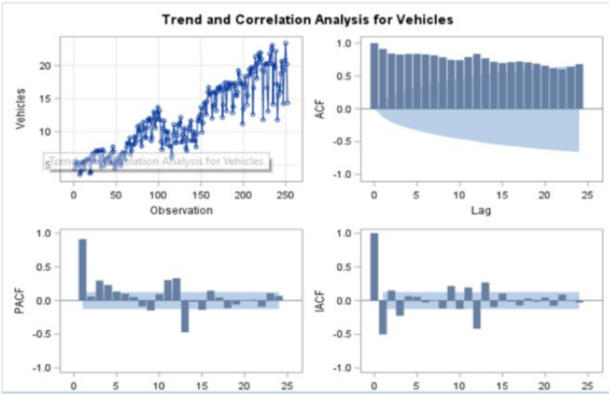
Model Parameter	Estimate	Std. Error	T	Prob>¦T¦
Seasonal Moving Average, Lag 12	-0.27113	0.0721	-3.7611	0.0037
Autoregressive, Lag 1	0.26821	0.0639	4.1988	0.0018
Model Variance (sigma squared)	0.00163	.	.	.

Fig. 13 Parameter estimates for Petrol

After checking the model prediction errors' ACF plot(see Fig. 11), Model prediction for Petrol(see Fig. 12) and parameter estimation (see Fig. 13), we found that the residuals are white noise series, the model provides a very good fit for the data and the parameters are all significant which all indicate that ARIMA model does a very good job in predicting the Petrol sales.

3. Bivariate model of petrol

    a. Bivariate model of petrol and vehicles



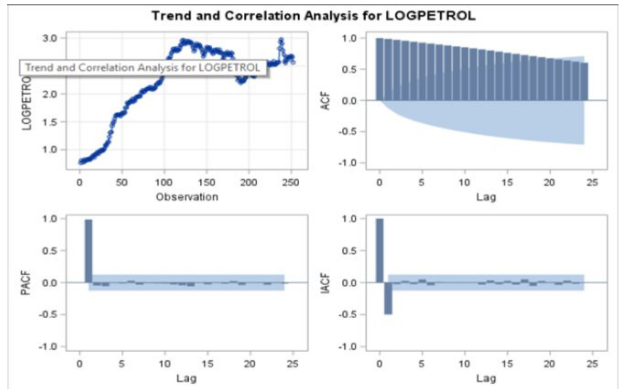
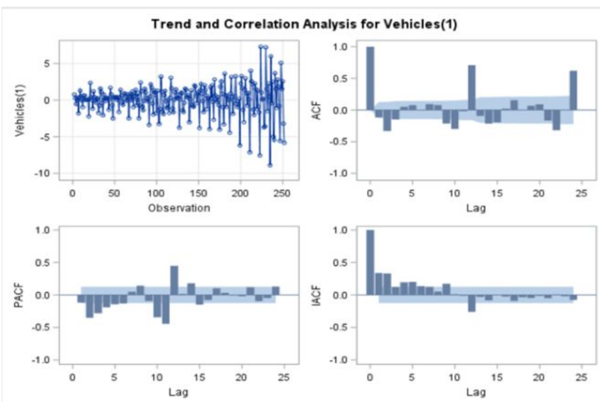


Fig. 14 Trend and Corr Analysis for Vehicles Fig. 15 Trend and Corr Analysis for LogPetrol

The two acf plots above shows that both vehicles and LogPetrol were nonstationary since each acf decayed slowly and was out of the 2 standard-error boundaries until lag 20(see Fig. 14 and Fig. 15).



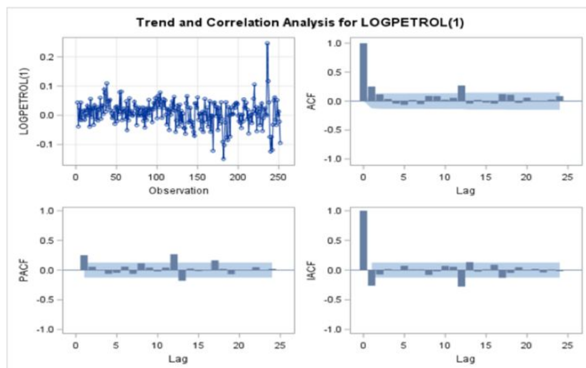


Fig. 16 Trend and Corr Analysis for Vehicles(1) Fig. 17 Trend and Corr Analysis for LogPetrol(1)

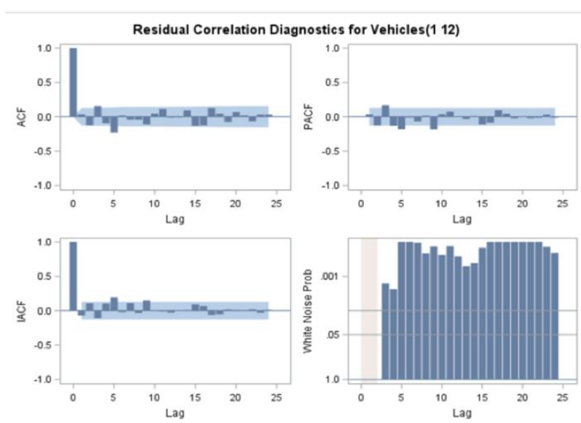
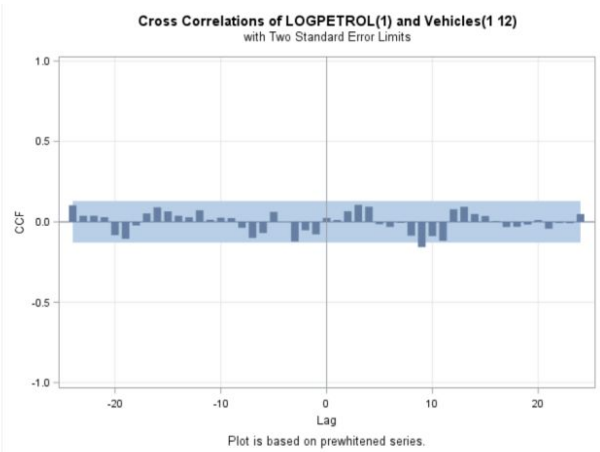


Fig. 18 Residual Corr Diagnostics for Vehicles(1,12)

After taking the first differencing, the updated vehicles(1) series was still not seasonally stationary since it decayed slowly at seasonal lags(see Fig. 16) whereas the new logpetrol(1) series was stationary at both seasonal and nonseasonal lags since it decayed quickly(see Fig. 17). Then we applied first-seasonal differencing to the vehicles(1) series and the vehicle(1,12) series became a stationary series(see Fig. 18).

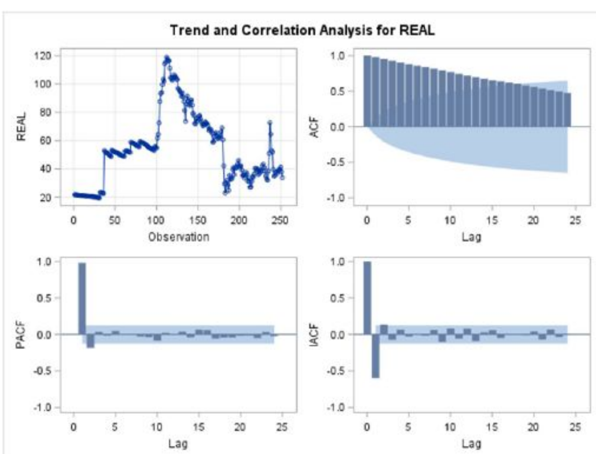


Crosscorrelation Check Between Series									
To Lag	Chi-Square	DF	Pr > ChiSq	Crosscorrelations					
5	6.04	6	0.4189	0.024	0.011	0.066	0.105	0.094	-0.015
11	19.16	12	0.0847	-0.031	-0.006	-0.086	-0.157	-0.088	-0.118
17	23.84	18	0.1602	0.079	0.093	0.049	0.036	0.005	-0.032
23	24.62	24	0.4267	-0.031	-0.016	0.011	-0.042	-0.007	-0.008

Fig. 19 CCF of LogPetrol(1) and Vehicles(1,12) Fig. 20 Crosscorrelation check between series

Unfortunately, there's no cross-correlation between input variable-vehicles(1,12) and output variable-logpetrol(1) as it's obvious to tell from the cross-correlation checking plot that there's no ccf outside the boundaries and we cannot reject the null hypothesis (see Fig. 19 and Fig. 20). In the other way, it means that input variable-vehicles(1,12) could not explain and predict the value of output variable-logpetrol(1).

b. Bivariate model of petrol and real



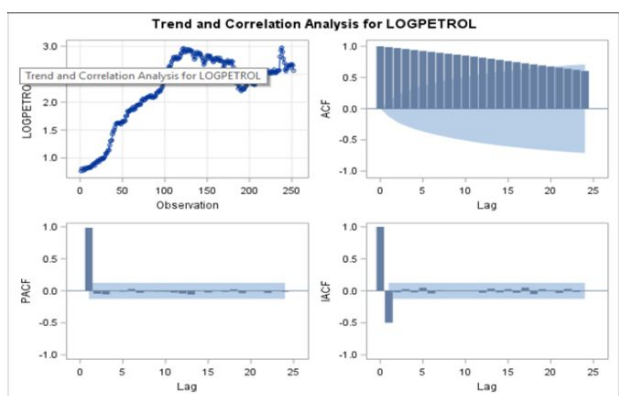


Fig. 21 Trend and Corr Analysis for Real Fig. 22 Trend and Corr Analysis for LogPetrol

The slowly-decayed acf of real and logpetrol indicated that real and logpetrol were not stationary(see Fig. 21 and Fig. 22). After getting natural logarithm and first differencing, variable PETROL could be considered a stationary time series at both seasonal and nonseasonal lags(see Fig. 23) while after getting first differencing and MA(1) prewhitening process, the variable REAL could be considered a white noise series as below since the white noise p-values were all more than .05 and we couldn't reject the null hypothesis(see Fig. 24).

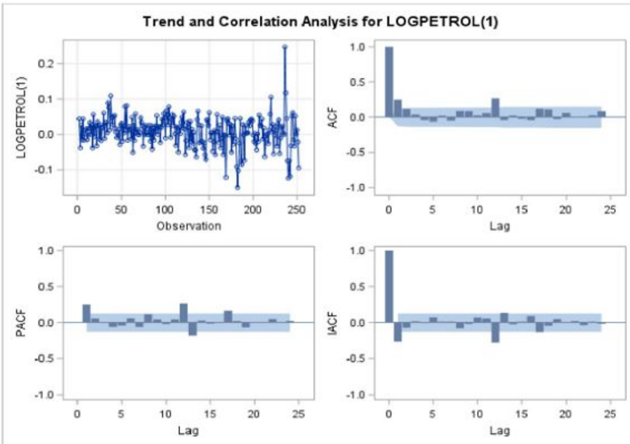


Fig. 23 Trend and Corr Analysis for LogPetrol(1)

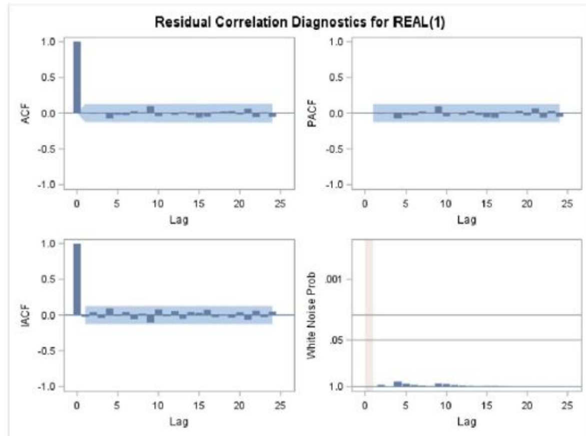


Fig. 24 Residual corr diagnostics for Real(1)

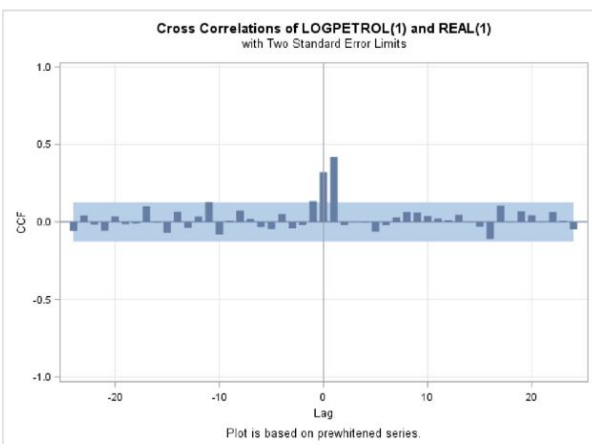


Fig. 25 CCF of LogPetrol(1) and Real(1)

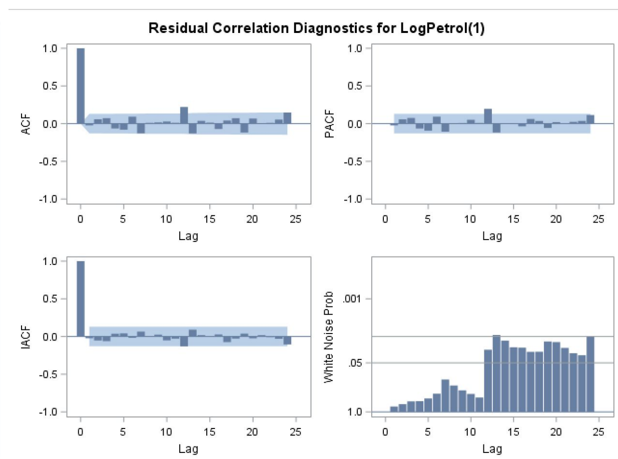


Fig. 26 Residuals corr Diagnostics for LogPetrol(1)

We could tell from the cross-correlation function plot (see Fig. 25) that first non-zero cross-correlation (response to input variable) at lag 0, the cross-correlation chopped off to 0 after lag 1. Therefore, here b equals to 0, r equals to 0 and s equals to 1. Moreover, Through the Residuals corr Diagnostics for LogPetrol(1) (see Fig. 26) we could identify an error model MA(1) on seasonal lag since ACF drops to 0 after lag 12 on seaonal lags

After identifying the order of Transfer Function Model in SAS Time Series Forecasting System, we estimated the TF in SAS Time Series Forecasting System (Fit Custom Model) and used a hold-out sample of 12 observations in order to prepare for further model comparison.

Parameter Estimates

PETROL

Log REAL[N(1)] + ARIMA(0,1,0)(1,0,0)s NOINT

Model Parameter	Estimate	Std. Error	T	Prob>|T|
Seasonal Autoregressive, Lag 12	0.30229	0.0685	4.4105	0.0017
REAL[N(1)]	0.00333	0.000506	6.5820	0.0001
REAL[N(1)] Num1	-0.00358	0.000509	-7.0348	<.0001
Model Variance (sigma squared)	0.00114	.	.	.

Fig. 27 Parameter Estimates

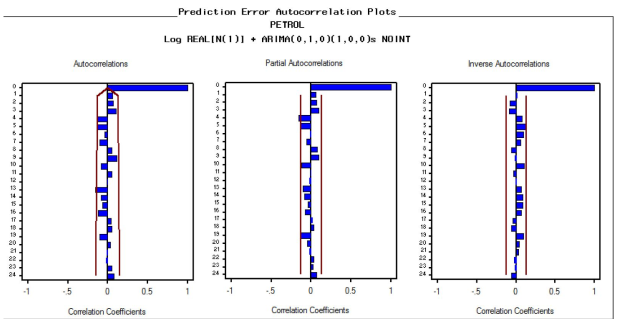


Fig. 28 ACF plot of residuals

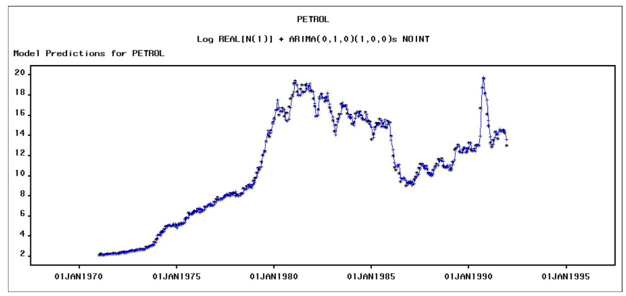


Fig. 29 Model prediction for Petrol

As we can see from the outputs above, the final Transfer Function model would be:

$$LOGPETROL(1)_t = (0.00333 - 0.00358 * B)REAL(1)_t + \varepsilon_t$$
$$(1 - 0.30229 * B^{12})\varepsilon_t = a_t$$
$$REAL(1)_t = (1 + 0.26493 * B)\, a_{t-1}$$

All estimates are statistically significant (see Fig. 27). Besides, the residuals of this transfer function model are white noise series as acf of error terms are not significantly different than 0 at all lags(see Fig. 28). We also conducted the whole process using PROC ARIMA commands and got the same estimates with following outputs, in which no residuals are correlated with the input variables(see Fig. 31) and residuals are white noise series(see Fig. 30 and Fig. 32).

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	6.90	5	0.2281	0.059	0.092	0.090	-0.025	-0.047	0.066
12	9.77	11	0.5509	-0.063	0.034	0.051	0.031	0.024	0.042
18	14.79	17	0.6105	-0.087	0.002	0.022	-0.040	0.070	0.064
24	29.92	23	0.1516	-0.079	0.095	0.020	0.011	0.067	0.185
30	43.73	29	0.0389	-0.006	0.191	-0.004	-0.028	0.101	0.037
36	47.08	35	0.0834	0.073	-0.027	-0.058	0.043	0.012	0.008
42	52.15	41	0.1137	0.073	0.053	-0.018	0.091	-0.001	0.017
48	61.55	47	0.0755	-0.031	0.017	-0.058	0.067	-0.028	0.143

Fig. 30 ACF check of residuals

Crosscorrelation Check of Residuals with Input REAL									
To Lag	Chi-Square	DF	Pr > ChiSq	Crosscorrelations					
5	0.55	4	0.9682	0.012	0.018	0.011	0.004	0.018	-0.036
11	2.98	10	0.9819	-0.009	0.056	0.055	0.043	0.040	0.010
17	10.90	16	0.8156	0.041	0.074	0.020	0.004	-0.089	0.127
23	13.96	22	0.9027	0.014	0.043	0.061	-0.008	0.064	0.048
29	15.96	28	0.9664	-0.044	0.002	-0.069	-0.001	-0.015	0.033
35	20.08	34	0.9721	-0.032	0.011	-0.021	-0.114	0.043	0.012
41	26.46	40	0.9507	0.023	0.087	-0.020	-0.107	0.057	0.048
47	28.49	46	0.9801	-0.061	0.040	0.042	0.031	0.005	-0.011

Fig. 31 Crosscorrelation check of residuals with Real

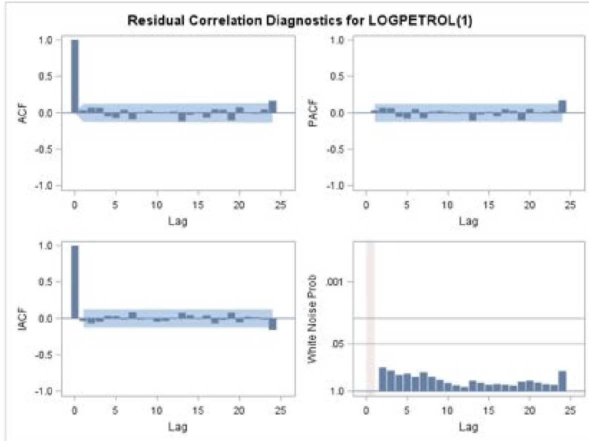


Fig. 32 Residual correlation diagnostics for LogPetrol(1)

4. Model Comparison

(1) Deterministic Time Series Model

Statistics of Fit

PETROL

Log Seasonal Dummies + Linear Trend

Statistic of Fit	Value
Mean Square Error	68.99054
Root Mean Square Error	8.30605
Mean Absolute Percent Error	59.41590
Mean Absolute Error	8.24742
R-Square	-164.511

Fig. 33 Statistics of fit of deterministic model

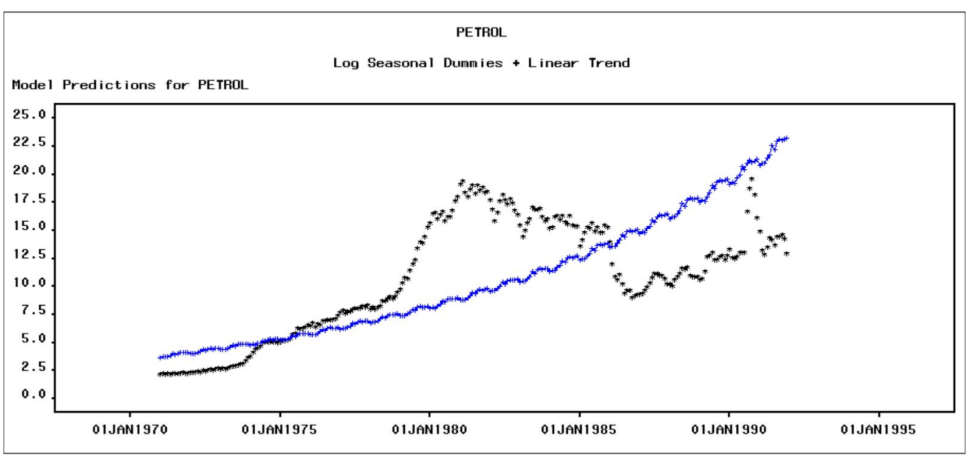


Fig. 34 Model prediction for Petrol

As we can see from Fig. 33 and Fig. 34, RMSE of deterministic model is relatively high and it has a bad performance in terms of prediction.

(2) Exponential Smoothing Model

Statistics of Fit

PETROL

Log Linear (Holt) Exponential Smoothing

Statistic of Fit	Value
Mean Square Error	0.74572
Root Mean Square Error	0.86355
Mean Absolute Percent Error	5.07417
Mean Absolute Error	0.69865
R-Square	-0.789

Fig. 35 Statistics of fit of Holt's linear exponential smoothing model

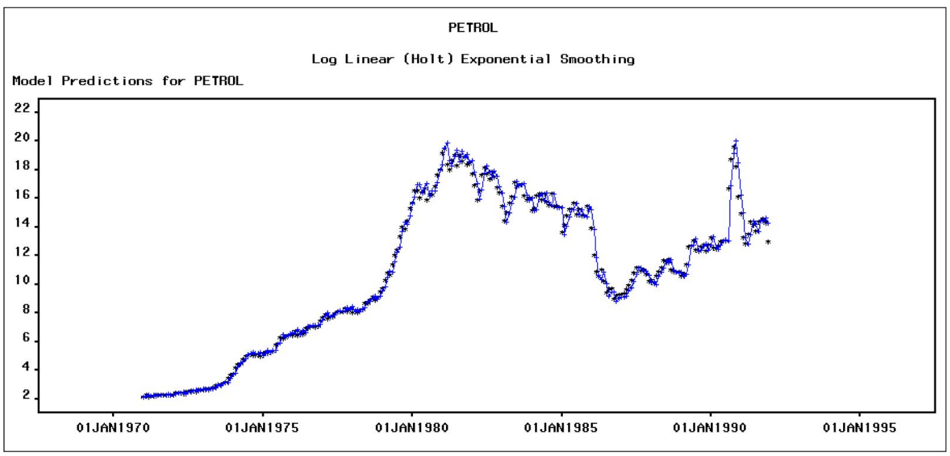


Fig. 36 Model prediction for Petrol

As we can see from Fig. 35 and Fig. 36, RMSE of Holt's linear exponential smoothing model is much lower than deterministic model and it also has a better performance in terms of prediction.

(3) ARIMA model

Statistics of Fit

PETROL

Log ARIMA(1,1,0)(0,0,1)s NOINT

Statistic of Fit	Value
Mean Square Error	0.32485
Root Mean Square Error	0.56995
Mean Absolute Percent Error	3.34525
Mean Absolute Error	0.46194
R-Square	0.221

Fig. 37 Statistics of fit of ARIMA model

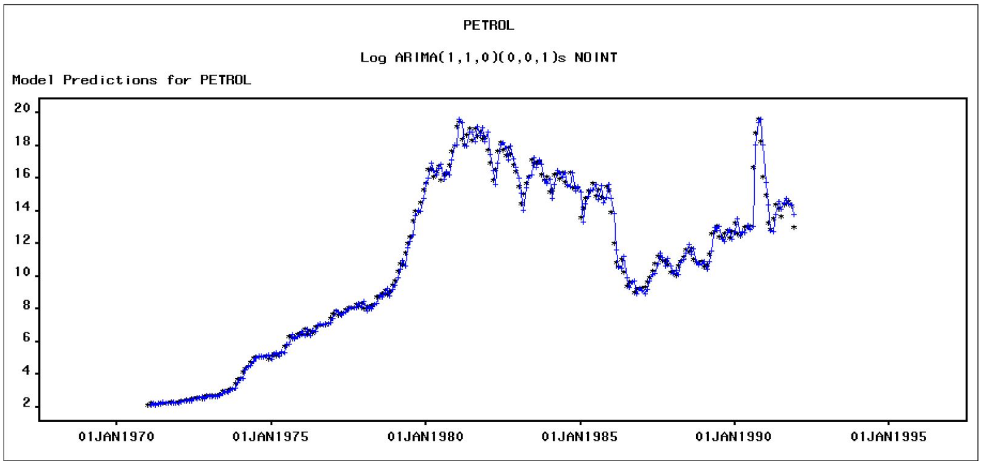


Fig. 38 Model prediction for Petrol

As we can see from Fig. 37 and Fig. 38, RMSE of ARIMA model is much lower than Holt's linear exponential smoothing and it also has a better performance in terms of prediction.

(4) Transfer Function Model

Statistics of Fit
PETROL
Log REAL[N(1)] + ARIMA(0,1,0)(1,0,0)s NOINT

Statistic of Fit	Value
Mean Square Error	0.15418
Root Mean Square Error	0.39265
Mean Absolute Percent Error	2.50239
Mean Absolute Error	0.34525
R-Square	0.630

Fig. 39 Statistics of fit of TF model

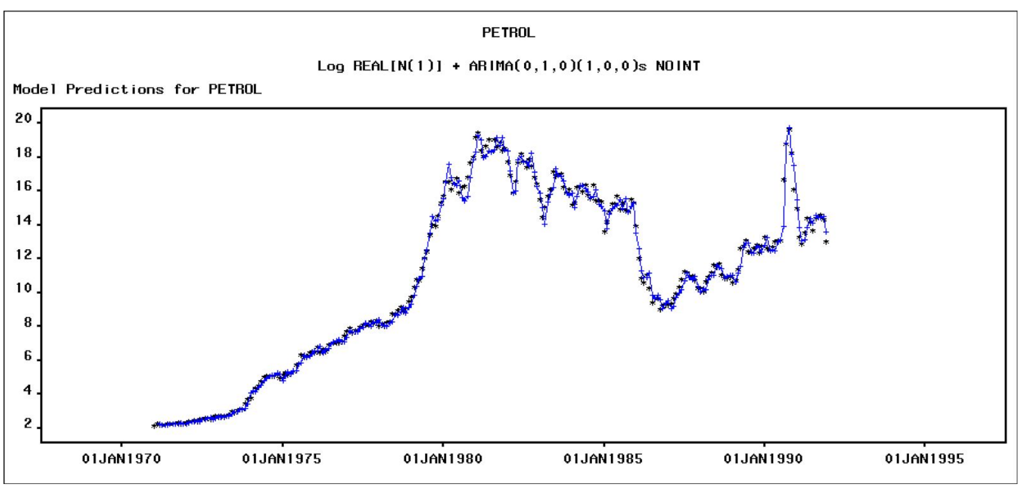


Fig. 40 Model prediction for Petrol

As we can see from Fig. 39 and Fig. 40, RMSE of TF model is much lower than ARIMA model and it also has a better performance in terms of prediction.

	Deterministic Model	Exponential Smoothing Model	ARIMA Model	Transfer Function Model
RMSE	8.30605	0.86355	0.56995	0.39265
Model Variance	0.17770	0.08683	0.00163	0.00114

Table. 1

Based on the output plots above(see Table. 1), we can conclude that univariate ARIMA model does the best job among all the univariate models(lowest RMSE and Model Variance) and Transfer Function Model does a better job as compared to univariate ARIMA model since it has lower RMSE and the lower model variance(in terms of fit). Therefore, using Petrol price to predict Petrol sales is more accurate than using the past observations of Petrol sales to predict.