**United States Energy Dynamics Analysis - 27 Years of Trends and 6 Year Projections**

**SAS Code**

**/\* Importing excel file into SAS \*/**

proc import out=energy datafile="/home/u62277456/sasuser.v94/Forecasting\_Deepa/Energy.xlsx"

dbms=xlsx replace;

run;

data energy;

    set energy;

**/\* Convert year and month to a standard SAS date format \*/**

    Date = mdy(month, 1, year);

**/\* Format the Date column as "MONYY7." (e.g., "MAY23") \*/**

    format Date monyy7.;

run;

**/\* Time-Series Plot - Production \*/**

proc sgplot data=energy;

     series x=date y=TotalProduction;

     title "Total Primary Energy Production (Trillion Btu)";

     xaxis label="Date";

     yaxis label="Total Production";

run;

proc sgplot data=energy;

     series x=consumerpriceindex y=Totalproduction;

     title "Total Primary Energy Production (Trillion Btu)";

     xaxis label="Consumer Price Index (CPI)";

     yaxis label="Total Production";

run;

proc sgplot data=energy;

     scatter x=stockchange y=Totalproduction;

     title "Total Primary Energy Production (Trillion Btu)";

     xaxis label="Primary Energy Stock Change";

     yaxis label="Total Production";

run;

**/\* Time-Series Plot - Consumption \*/**

proc sgplot data=energy;

     series x=date y=Totalconsumption;

     title "Total Primary Energy Consumption (Trillion Btu)";

     xaxis label="Date";

     yaxis label="Total Consumption";

run;

proc sgplot data=energy;

     series x=consumerpriceindex y=Totalconsumption;

     title "Total Primary Energy Consumption (Trillion Btu)";

     xaxis label="Consumer Price Index (CPI)";

     yaxis label="Total Consumption";

run;

proc sgplot data=energy;

     scatter x=stockchange y=Totalconsumption;

     title "Total Primary Energy Consumption (Trillion Btu)";

     xaxis label="Primary Energy Stock Change";

     yaxis label="Total Consumption";

run;

**/\* ACF plot \*/**

proc timeseries data=energy plots=acf out=\_null\_;

var totalproduction;

corr acf/nlag=60;

run;

proc timeseries data=energy plots=acf out=\_null\_;

var totalconsumption;

corr acf/nlag=60;

run;

**/\* Descriptive Statistics (numerical summary) \*/**

proc means data=energy n mean median std min max ;

var totalproduction totalconsumption StockChange ConsumerPriceIndex;

run;

**/\* Winter's Exponential Smoothing \*/**

**/\* Additive Winter's \*/**

proc esm data=energy lead=12 outfor=energyout1 plot=forecasts out=\_null\_ print=all;

id Date interval=month;

forecast TotalProduction/model=addwinters;

run;

proc sgplot data=energyout1;

series x=date y=actual;

series x=date y=predict;

run;

proc esm data=energy lead=12 outfor=energyout1 plot=forecasts out=\_null\_ print=all;

id Date interval=month;

forecast Totalconsumption/model=addwinters;

run;

proc sgplot data=energyout1;

series x=date y=actual;

series x=date y=predict;

run;

**/\* Multiplicative Winter's \*/**

proc esm data=energy lead=12 outfor=energyout2 plot=forecasts out=\_null\_ print=all;

id date interval=month;

forecast TotalProduction/model=winters;

run;

proc sgplot data=energyout2;

series x=date y=actual;

series x=date y=predict;

run;

proc esm data=energy lead=12 outfor=energyout2 plot=forecasts out=\_null\_ print=all;

id date interval=month;

forecast Totalconsumption/model=winters;

run;

proc sgplot data=energyout2;

series x=date y=actual;

series x=date y=predict;

run;

**/\* Forecast accuracy (Production) - using last 72 obs as test set - 6 years  \*/**

proc esm data=energy lead=72 back=72 outfor=energyout1 plot=forecasts out=\_null\_ print=all;

id date interval=month;

forecast TotalProduction/model=addwinter;

run;

proc esm data=energy lead=72 back=72 outfor=energyout2 plot=forecasts out=\_null\_ print=all;

id date interval=month;

forecast TotalProduction/model=winters;

run;

proc sgplot data=energyout1;

series x=date y=actual;

series x=date y=predict;

run;

proc sgplot data=energyout2;

series x=date y=actual;

series x=date y=predict;

run;

**/\* Forecast accuracy (Consumption) - using last 72 obs as test set - 6 years  \*/**

proc esm data=energy lead=72 back=72 outfor=energyout3 plot=forecasts out=\_null\_ print=all;

id date interval=month;

forecast Totalconsumption/model=addwinter;

run;

proc esm data=energy lead=72 back=72 outfor=energyout4 plot=forecasts out=\_null\_ print=all;

id date interval=month;

forecast Totalconsumption/model=winters;

run;

proc sgplot data=energyout3;

series x=date y=actual;

series x=date y=predict;

run;

proc sgplot data=energyout4;

series x=date y=actual;

series x=date y=predict;

run;

**/\* Simple Regression Model (Production) - setting a unique identifier t with a sequence starting from 1**

**since the data doesn't have one. Also setting t>264 since the regression the model needs to be fitted for the first 264 obs. \*/**

data energy1;

set energy;

t=\_n\_;

energy\_p=TotalProduction;

if t>264 then energy\_p=.;

run;

proc sgplot data=energy1;

series x=t y=energy\_p;

run;

**/\*#fitting Regression model\*/**

proc reg data=energy1;

model energy\_p=t/clb dwprob;

run;

**/\*#forecasting production\*/**

proc reg data=energy1;

model energy\_p=t/clb;

output out=energy1out p=energy1\_pre r=energy1\_res;

run;

proc sgplot data=energy1out;

series x=date y=TotalProduction;

series x=date y=energy1\_pre;

run;

**/\* Simple Regression Model (Consumption) - setting a unique identifier t with a sequence starting from 1 since the data don't have one. Also setting t>264 since the regression the model needs to be fitted for the first 264 obs. \*/**

data energy1;

set energy;

t=\_n\_;

energy\_p=Totalconsumption;

if t>264 then energy\_p=.;

run;

proc sgplot data=energy1;

series x=t y=energy\_p;

run;

**/\*#fitting Regression model\*/**

proc reg data=energy1;

model energy\_p=t/clb dwprob;

run;

**/\*#forecasting production\*/**

proc reg data=energy1;

model energy\_p=t/clb;

output out=energy1out p=energy1\_pre r=energy1\_res;

run;

proc sgplot data=energy1out;

series x=date y=Totalconsumption;

series x=date y=energy1\_pre;

run;

**/\* Evaluating model's performance\*/**

data energy1out;

set energy1out;

if t<=264 then

do;

mape\_fit=abs(energy1\_res/energy\_p)\*100;

mae\_fit=abs(energy1\_res);

end;

else if t>264 then

do;

mape\_acc=abs((Totalconsumption - energy1\_pre)/Totalconsumption)\*100;

mae\_acc=abs(Totalconsumption - energy1\_pre);

end;

run;

proc means data=energy1out mean;

var mape\_fit mae\_fit mape\_acc mae\_acc;

run;

**/\* Deseasonalizing \*/**

proc timeseries data=energy1 outdecomp=sa\_energy out=\_null\_;

decomp sa;

id date interval=month;

var Totalconsumption;

run;

data energycombine;

merge sa\_energy energy1;

sa1=sa;

if t>264then sa1=.;

si=Totalconsumption/sa;

run;

proc sgplot data=energycombine;

series x=date y=Totalconsumption;

series x=date y=sa;

run;

**/\*Simple Regression model for deseasonalized dataset energy combine - obtaining seasonally adjusted data \*/**

proc reg data=energycombine;

model sa1=t;

output out=energy2\_out r=energy2\_res p=energy2\_pre;

run;

proc sgplot data=energy2\_out;

series x=t y=sa1;

series x=t y=energy2\_pre;

run;

data energy2\_out;

set energy2\_out;

energy\_reseason=si\*energy2\_pre;

run;

proc sgplot data=energy2\_out;

series x=t y=Totalconsumption;

series x=t y=energy\_reseason;

run;

**/\*calculating MAPE and MAE of model fit and model accuracy\*/**

data energy2\_out;

set energy2\_out;

if t<=264 then

do;

mape\_fit2 =(abs(energy2\_res/sa1))\*100;

mae\_fit2 =abs(energy2\_res);

end;

else if t>264 then

do;

mape\_acc2 =(abs(sa-energy2\_pre)/sa)\*100;

mae\_acc2 =abs(sa-energy2\_pre);

end;

run;

proc means data=energy2\_out mean;

var mape\_fit2 mae\_fit2  mape\_acc2 mae\_acc2;

run;

**/\* Multiple Regression Model \*/**

proc reg data=PrimaryEnergy;

model TotalProduction=TotalConsumption StockChange CPI/clb corrb vif dwprob;

run;

proc reg data=PrimaryEnergy;

model TotalConsumption=TotalProduction StockChange CPI/clb corrb vif dwprob;

run;

**/\* Selecting predictors \*/**

proc reg data=PrimaryEnergy outest=PrimaryEnergyOut1;

model TotalProduction=TotalConsumption StockChange CPI/aic bic;

run;

proc reg data=PrimaryEnergy outest=PrimaryEnergyOut2;

model TotalProduction=TotalConsumption CPI/aic bic;

run;

proc reg data=PrimaryEnergy outest=PrimaryEnergyOut3;

model TotalProduction=CPI/aic bic;

run;

proc reg data=PrimaryEnergy outest=PrimaryEnergyOut4;

model TotalProduction=TotalConsumption/aic bic;

run;

proc reg data=PrimaryEnergy outest=PrimaryEnergyOut7;

model TotalConsumption=TotalProduction StockChange CPI/aic bic;

run;

proc reg data=PrimaryEnergy outest=PrimaryEnergyOut8;

model TotalConsumption=TotalProduction StockChange/aic bic;

run;

proc reg data=PrimaryEnergy outest=PrimaryEnergyOut9;

model TotalConsumption=StockChange CPI/aic bic;

run;

proc reg data=PrimaryEnergy outest=PrimaryEnergyOut10;

model TotalConsumption=StockChange/aic bic;

run;

proc reg data=PrimaryEnergy;

model TotalProduction=TotalConsumption CPI/clb corrb vif dwprob;

output out=PrimaryEnergyOut2 P=TotalProduction\_Prediction;

run;

proc sgplot data=PrimaryEnergyOut2;

series x=Date y=TotalProduction;

series x=Date y=TotalProduction\_Prediction;

run;

proc reg data=PrimaryEnergy;

model TotalConsumption=TotalProduction StockChange/clb corrb vif dwprob;

output out=PrimaryEnergyOut8 P=TotalConsumption\_Prediction;

run;

proc sgplot data=PrimaryEnergyOut8;

series x=Date y=TotalConsumption;

series x=Date y=TotalConsumption\_Prediction;

run;

**/\* Linearity check and Prediction without Seasonality using Multiple regression after splitting the data\*/**

data PrimaryEnergy;

set PrimaryEnergy;

t=\_n\_;

t2=t\*t;

CPI2=CPI\*CPI;

New\_TotalProduction=TotalProduction;

if t>264 then New\_TotalProduction=.;

run;

/\* **this has lessor adjusted r2 though it shows non-linearity**

proc reg data=PrimaryEnergy;

model New\_TotalProduction=TotalConsumption CPI t t2/clb corrb vif dwprob;

output out=PrimaryEnergyOut5 R=TotalProduction\_Residual P=TotalProduction\_Prediction;

run;

proc reg data=PrimaryEnergy;

model New\_TotalProduction=TotalConsumption t t2/clb corrb vif dwprob;

output out=PrimaryEnergyOut5 R=TotalProduction\_Residual P=TotalProduction\_Prediction;

run;

proc reg data=PrimaryEnergy;

model New\_TotalProduction=TotalConsumption CPI CPI2/clb corrb vif dwprob;

output out=PrimaryEnergyOut5 R=TotalProduction\_Residual P=TotalProduction\_Prediction;

run; \*/

proc reg data=PrimaryEnergy;

model New\_TotalProduction=TotalConsumption CPI CPI2 t t2/clb corrb vif dwprob;

output out=PrimaryEnergyOut11 R=TotalProduction\_Residual P=TotalProduction\_Prediction;

run;

proc sgplot data=PrimaryEnergyOut11;

series x=Date y=TotalProduction;

series x=Date y=TotalProduction\_Prediction;

run;

data PrimaryEnergyOut11;

set PrimaryEnergyOut11;

if t<=264 then

do;

mape\_fit=(abs(TotalProduction\_Residual/New\_TotalProduction))\*100;

mae\_fit=abs(TotalProduction\_Residual);

end;

else if t>264 then

do;

mape\_acc=(abs(TotalProduction-TotalProduction\_Prediction)/TotalProduction)\*100;

mae\_acc=abs(TotalProduction-TotalProduction\_Prediction);

end;

run;

proc means data=PrimaryEnergyOut11 mean;

var mape\_fit mape\_acc mae\_fit mae\_acc;

run;

**/\* Prediction of Total Production with Seasonality and non-Linearity using Multiple regression by adding dummy variables\*/**

data PrimaryEnergy;

set PrimaryEnergy;

monthly=month(Date);

if month=1 then month1=1; else month1=0;

if month=2 then month2=1; else month2=0;

if month=3 then month3=1; else month3=0;

if month=4 then month4=1; else month4=0;

if month=5 then month5=1; else month5=0;

if month=6 then month6=1; else month6=0;

if month=7 then month7=1; else month7=0;

if month=8 then month8=1; else month8=0;

if month=9 then month9=1; else month9=0;

if month=10 then month10=1; else month10=0;

if month=11 then month11=1; else month11=0;

run;

proc freq data=PrimaryEnergy;

table month month1 month2 month3 month4 month5 month6 month7 month8 month9 month10 month11;

run;

proc reg data=PrimaryEnergy;

model New\_TotalProduction=TotalConsumption CPI CPI2 month1 month2 month3 month4 month5 month6 month7 month8 month9 month10 month11/aic bic;

output out=PrimaryEnergyOut6 r=TotalProduction\_Resid p=TotalProduction\_Predict;

run;

data PrimaryEnergyOut6;

set PrimaryEnergyOut6;

if t<=264 then

do;

mape\_fit=(abs(TotalProduction\_Resid/New\_TotalProduction))\*100;

mae\_fit=abs(TotalProduction\_Resid);

end;

else if t>264 then

do;

mape\_acc=(abs(TotalProduction-TotalProduction\_Predict)/TotalProduction)\*100;

mae\_acc=abs(TotalProduction-TotalProduction\_Predict);

end;

run;

proc means data=PrimaryEnergyOut6 mean;

var mape\_fit mape\_acc mae\_fit mae\_acc;

run;

proc sgplot data=PrimaryEnergyOut6;

series x=Date y=TotalProduction;

series x=Date y=TotalProduction\_Predict;

run;

**/\* Linearity check and Prediction for Consumption without Seasonality using Multiple regression \*/**

data PrimaryEnergy;

set PrimaryEnergy;

t=\_n\_;

t2=t\*t;

/\* StockChange2=StockChange\*StockChange;

TotalProduction2=TotalProduction\*TotalProduction; \*/

New\_TotalConsumption=TotalConsumption;

if t>264 then New\_TotalConsumption=.;

run;

proc reg data=PrimaryEnergy;

model New\_TotalConsumption=TotalProduction StockChange t t2/clb corrb vif dwprob;

output out=PrimaryEnergyOut8 R=TotalConsumption\_Residual P=TotalConsumption\_Prediction;

run;

/\* **Since the adjusted r2 is less**

proc reg data=PrimaryEnergy;

model New\_TotalConsumption=TotalProduction StockChange StockChange2/clb corrb vif dwprob;

output out=PrimaryEnergyOut12 R=TotalConsumption\_Residual P=TotalConsumption\_Prediction;

run;

proc reg data=PrimaryEnergy;

model New\_TotalConsumption=TotalProduction TotalProduction2 StockChange /clb corrb vif dwprob;

output out=PrimaryEnergyOut13 R=TotalConsumption\_Residual P=TotalConsumption\_Prediction;

run; \*/

proc sgplot data=PrimaryEnergyOut8;

series x=Date y=TotalConsumption;

series x=Date y=TotalConsumption\_Prediction;

run;

data PrimaryEnergyOut8;

set PrimaryEnergyOut8;

if t<=264 then

do;

mape\_fit=(abs(TotalConsumption\_Residual/New\_TotalConsumption))\*100;

mae\_fit=abs(TotalConsumption\_Residual);

end;

else if t>264 then

do;

mape\_acc=(abs(TotalConsumption-TotalConsumption\_Prediction)/TotalConsumption)\*100;

mae\_acc=abs(TotalConsumption-TotalConsumption\_Prediction);

end;

run;

proc means data=PrimaryEnergyOut8 mean;

var mape\_fit mape\_acc mae\_fit mae\_acc;

run;

**/\* Linearity and Prediction Consumption with Seasonality using Multiple regression adding dummy variables\*/**

data PrimaryEnergy;

set PrimaryEnergy;

monthly=month(Date);

if month=1 then month1=1; else month1=0;

if month=2 then month2=1; else month2=0;

if month=3 then month3=1; else month3=0;

if month=4 then month4=1; else month4=0;

if month=5 then month5=1; else month5=0;

if month=6 then month6=1; else month6=0;

if month=7 then month7=1; else month7=0;

if month=8 then month8=1; else month8=0;

if month=9 then month9=1; else month9=0;

if month=10 then month10=1; else month10=0;

if month=11 then month11=1; else month11=0;

run;

proc freq data=PrimaryEnergy;

table month month1 month2 month3 month4 month5 month6 month7 month8 month9 month10 month11;

run;

proc reg data=PrimaryEnergy;

model New\_TotalConsumption=TotalProduction StockChange month1 month2 month3 month4 month5 month6 month7 month8 month9 month10 month11 t t2/aic bic;

output out=PrimaryEnergyOut8 r=TotalConsumption\_Resid p=TotalConsumption\_Predict;

run;

data PrimaryEnergyOut8;

set PrimaryEnergyOut8;

if t<=264 then

do;

mape\_fit=(abs(TotalConsumption\_Resid/New\_TotalConsumption))\*100;

mae\_fit=abs(TotalConsumption\_Resid);

end;

else if t>264 then

do;

mape\_acc=(abs(TotalConsumption-TotalConsumption\_Predict)/TotalConsumption)\*100;

mae\_acc=abs(TotalConsumption-TotalConsumption\_Predict);

end;

run;

proc means data=PrimaryEnergyOut8 mean;

var mape\_fit mape\_acc mae\_fit mae\_acc;

run;

proc sgplot data=PrimaryEnergyOut8;

series x=Date y=TotalConsumption;

series x=Date y=TotalConsumption\_Predict;

run;

**/\* Seasonal ARIMA model for Production\*/**

proc timeseries data=PrimaryEnergy plots=(acf pacf) out=\_null\_;

var TotalProduction;

run;

/\* **TO CHECK ON differentiations**

data PrimaryEnergy;

set PrimaryEnergy;

TotalProductiondiff=dif(TotalProduction);

run;

proc timeseries data=PrimaryEnergy plots=(acf pacf) out=\_null\_;

var TotalProductiondiff;

run; \*/

proc arima data=PrimaryEnergy;

identify var=TotalProduction(1,12) nlag=36 whitenoise=ignoremiss;

estimate p=(2)(12)(24) q=(1)(12); /\* ARIMA(2,1,1)(2,1,1) \*/

\*estimate p=(1)(12)(24) q=(1)(12); /\* ARIMA(1,1,1)(2,1,1) \*/

run;

**/\* we dropped ARIMA model as we didn’t get residual white noise \*/**