

Chapter 3

Preliminary Data Analysis

This chapter describes the elementary character of the data, which comprise the volumes of six kinds of consumption and imports and production of petroleum products in Thailand during the period 1984-1999. These are classified as gasoline, diesel, jet petrol (JP), fuel oil, and liquefied petroleum gas (LPG). The analysis includes:

1. Data summaries of petroleum consumption levels.
2. Data summaries of petroleum imports and production levels.
3. Relation between consumption and imports and production.

3.1 Data Summaries of Petroleum Consumption Levels

Figure 3.1 shows histograms and brief numerical summaries of monthly consumption of petroleum products in Thailand from January 1984 to December 1999 in millions of litres (ML).

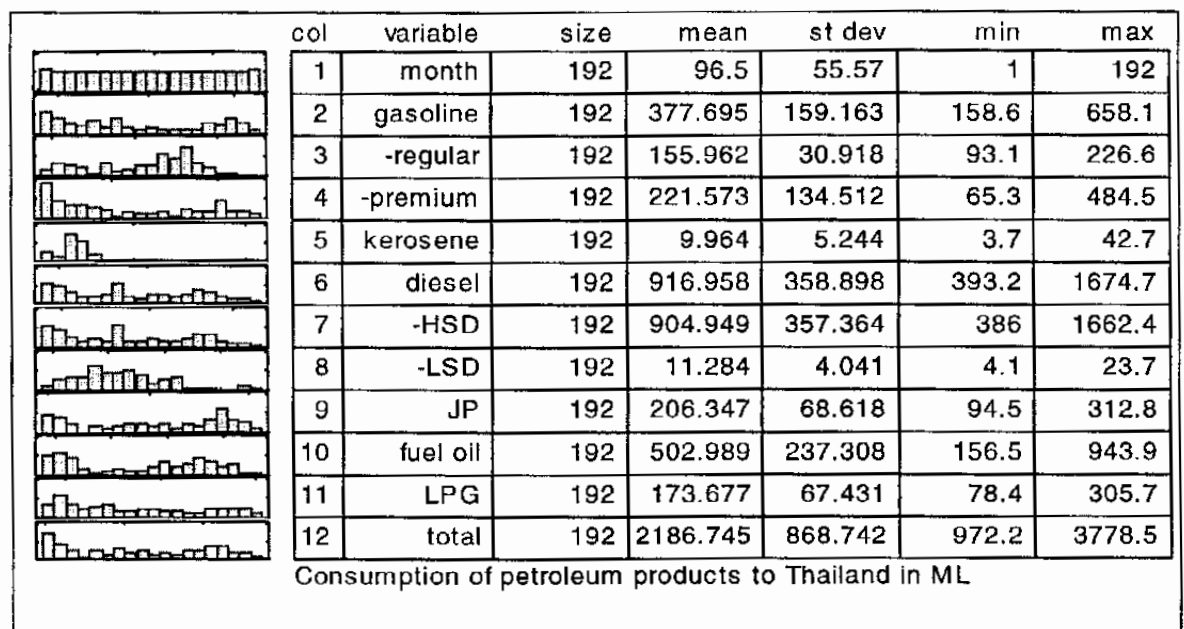


Figure 3.1: Monthly consumption of petroleum products to Thailand, 1984-1999

The figures show the mean, standard deviation, minimum and maximum for each product of the total monthly consumption. The gasoline component is classified into regular and premium, and the diesel component is classified into HSD (high speed) and LSD (low speed).

The average of the total monthly consumption is 2,186.745 ML, and the largest component is HSD (mean = 904.949), which accounts for about 40% of the total on average. At the other extreme, LSD only accounts for 11.284 ML per month on average, and the level of kerosene consumption is only 9.964 ML per month on average. These two components together comprise less than 1% of the total.

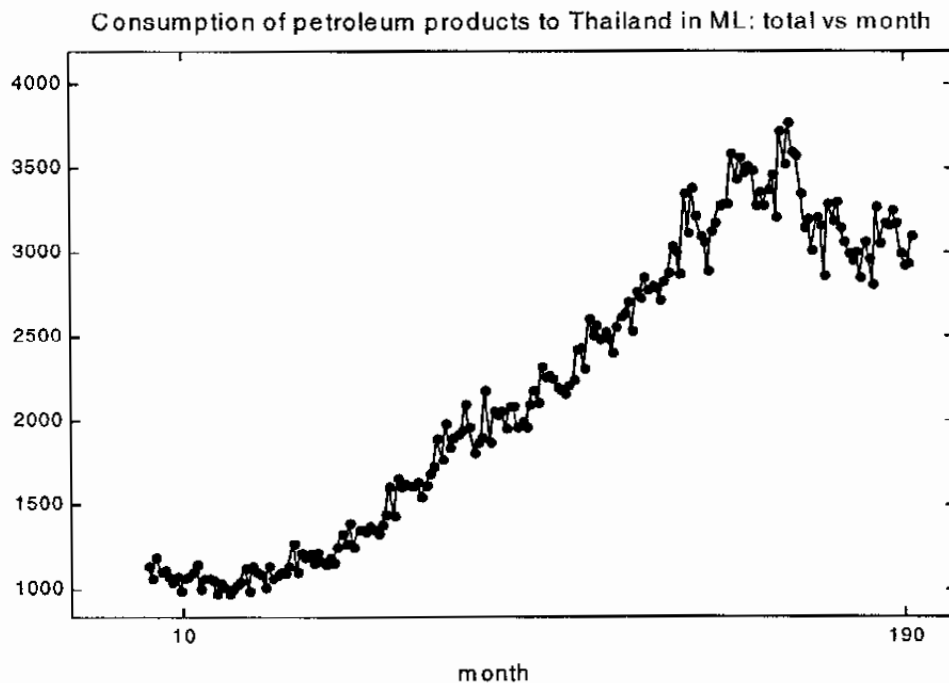


Figure 3.2: Trend of monthly consumption of total petroleum products in Thailand, 1984-1999

Figure 3.2 shows the time series of total monthly consumption. The time series is constant in the first year. Between 1986 and 1996 this series gradually increases, reaching a maximum in 1996, while after 1996 there is a downward trend. In 1999 there is a slight increase.

It appears from Figure 3.2 that there is not constant variability in the monthly consumption levels when the levels themselves are changed. We now consider the possibility of transforming the data, with the aim of making the variability of the

transformed data more constant. This will make it easier to fit a statistical model, because most statistical methods assume constant variance.

Figure 3.3 shows a graph of the magnitude of the change in the consumption levels from one month to the next versus the level itself. The correlation coefficient is 0.363, confirming that the change in level is related to the level.

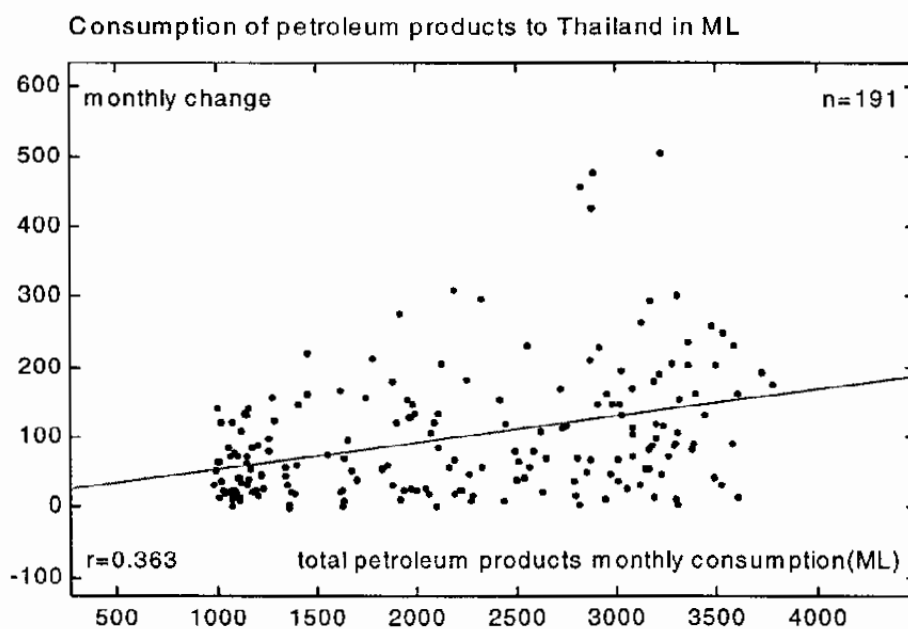


Figure 3.3: Monthly change versus level, for total petroleum consumption

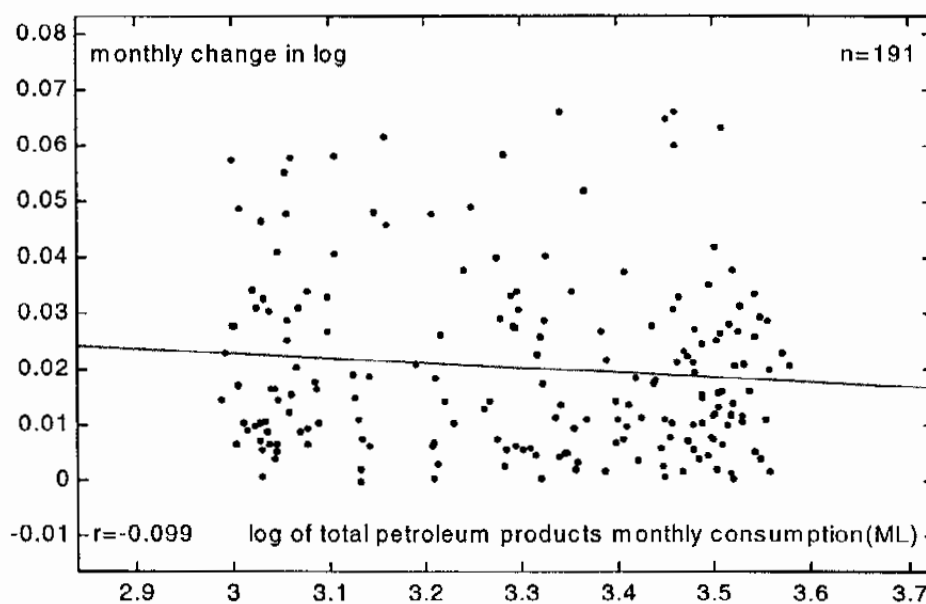


Figure 3.4: Monthly change versus level, for log-transformed total petroleum consumption

Figure 3.4 shows the same graph after taking base-10 logarithms of the total monthly consumption levels. The logarithm transformation is effective in removing the relation between the monthly changes in petroleum consumption levels and the levels themselves.

In this case the correlation coefficient between the change and the level is close to zero (-0.099).

Figure 3.5 shows the trend of the monthly consumption levels after the logarithm transformation. The trend is the same as the trend shown in Figure 3.5, but the variability is now nearly constant.

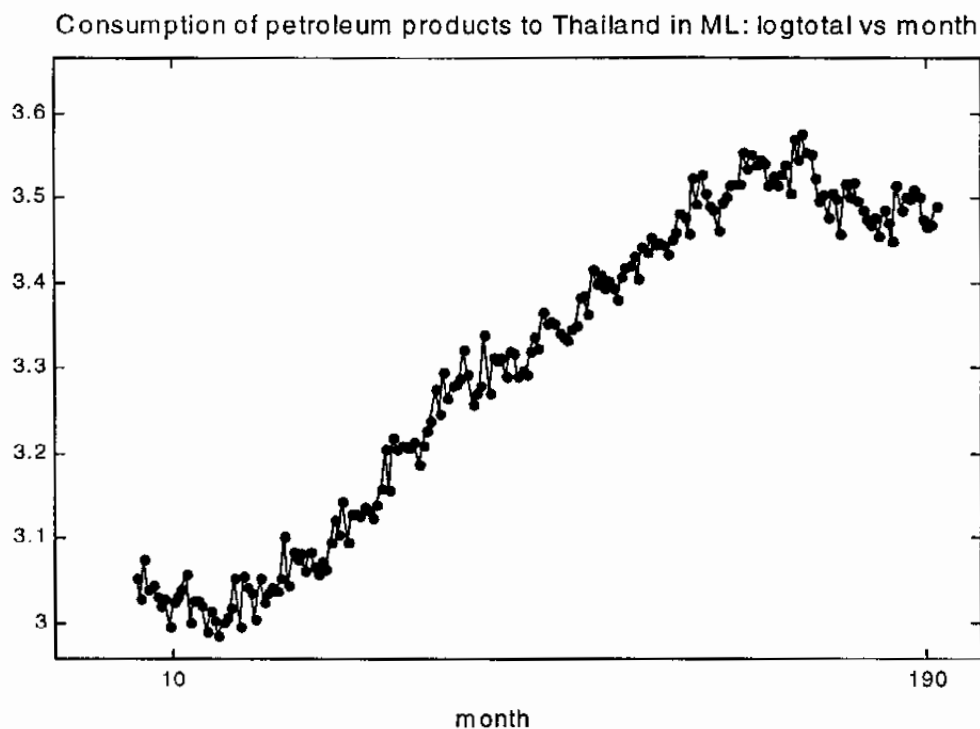


Figure 3.5: Trend of monthly consumption of total petroleum products to Thailand, 1984-1999 after transforming data by taking logarithms

Figure 3.6 summarises the consumption of the main petroleum products after the transformation using base 10 logarithms. The components are ordered according to the mean consumption.

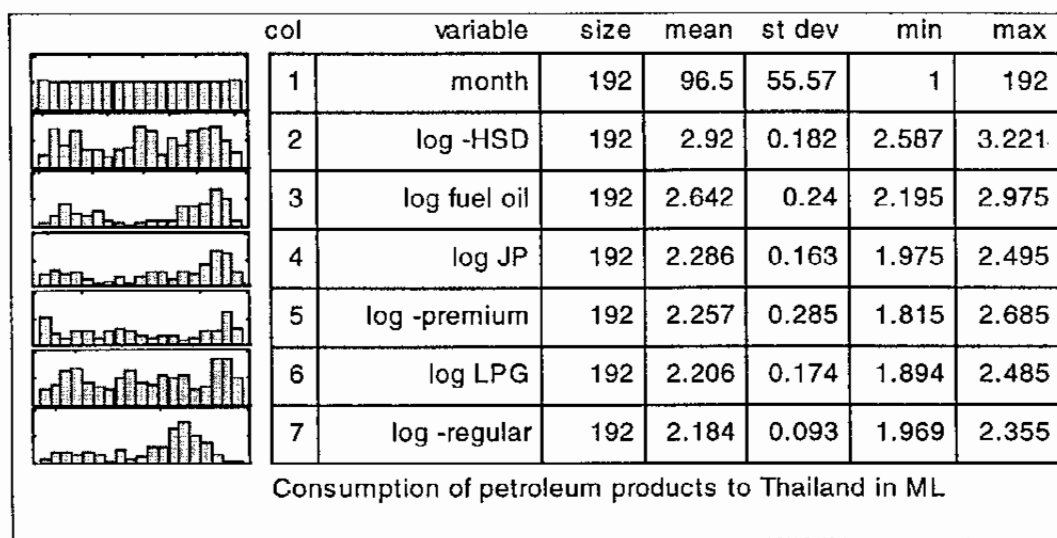


Figure 3.6: Histograms and statistics of base 10 logarithm of consumption of main petroleum products

Figure 3.7 shows the relations between the various components. The relation between all of the various pairs of components is close to linear. The correlation coefficients between all of the pairs of components, shown in Figure 3.7 and Table 3.1, are all close to 1, ranging from 0.855 and 0.859 (between regular and premium gasoline) to 0.983 and 0.988 (between LPG and premium gasoline), respectively.

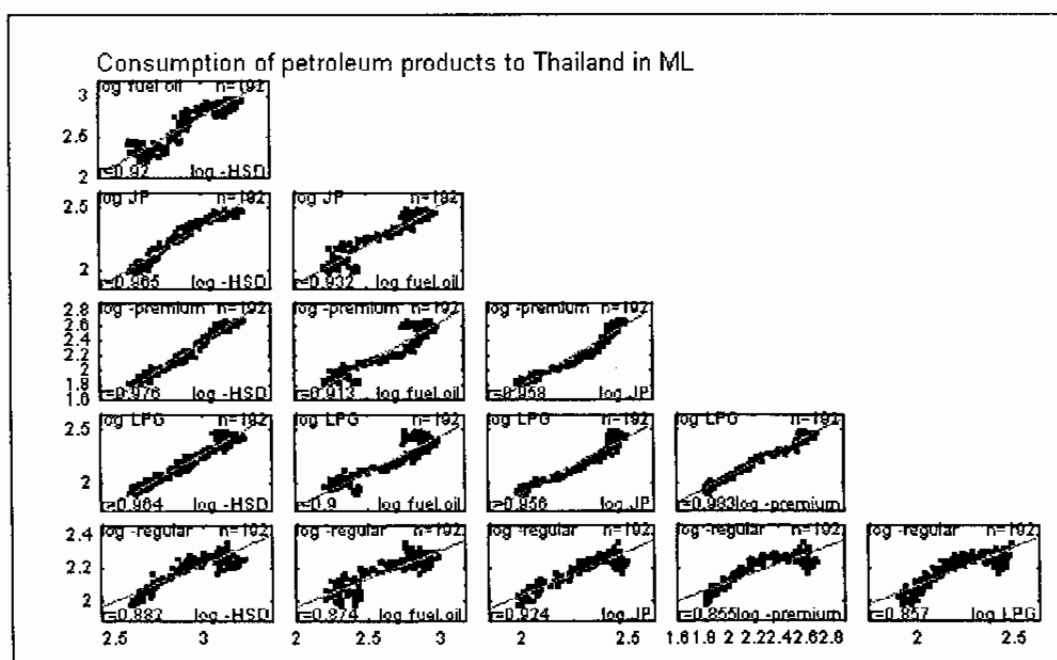


Figure 3.7: Scatterplot matrix showing relations between components of petroleum consumption

	-HSD	fuel oil	JP	-premium	LPG	-regular
-HSD	1					
fuel oil	0.925	1				
JP	0.970	0.937	1			
-premium	0.982	0.918	0.963	1		
LPG	0.969	0.905	0.961	0.988	1	
-regular	0.887	0.879	0.929	0.859	0.861	1

Table 3.1: Correlation matrix for consumption

Figure 3.8 shows how the various components of petroleum consumption change with time over the 16-year period from 1984 to 1999.

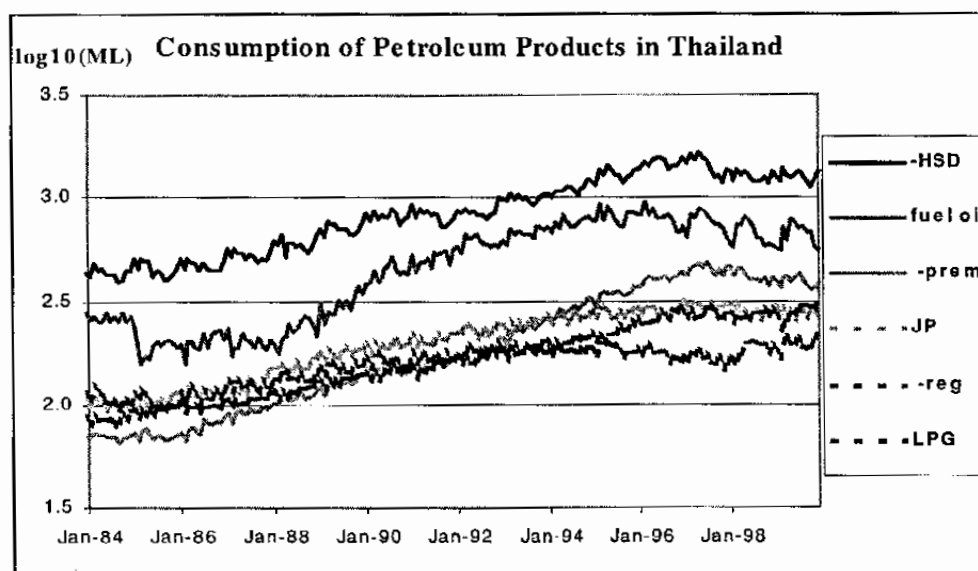


Figure 3.8: Time-series plot of the six main petroleum products in Thailand during 1984-1999

Consumption of petroleum products tends to increase as month increases. There appears to be a positive relation between month and amount. Consumption of JP, LPG and regular increase together.

From the data during 1984-1999, in Figure 3.8 there is an increase in consumption of all petroleum products. There is the most consumption in HSD. The next is fuel oil, and the others are consumed less.

Table 3.2 and Figure 3.9 show the result of fitting a model to the monthly consumption data. In this model we assume that each component increases linearly over the 16-year period, and that the slopes of these fitted lines are the same for each component. Each parameter in this model is highly statistically significant with all the p-values close to 0. The r-squared is 0.934, so this model accounts for 93.4% of the variation in the data.

Relation: petroleum consumption				
outcome: log(ML) 1152 cases				
predictor	coeff	SE	t	p-value
constant	2.6094	0.0077	338.446	0.0000
month	0.0032	0.0000	69.594	0.0000
type	(0 for -HSD)			0.0000
fuel oil	-0.2781	0.0089	-31.291	0.0000
JP	-0.6338	0.0089	-71.315	0.0000
-prem	-0.6628	0.0089	-74.578	0.0000
LPG	-0.7144	0.0089	-80.387	0.0000
-reg	-0.7367	0.0089	-82.892	0.0000
r-sq: .934 Resid SS: 8.682 s: .0871 df: 1145				

Table 3.2: Fitted linear model for petroleum consumption

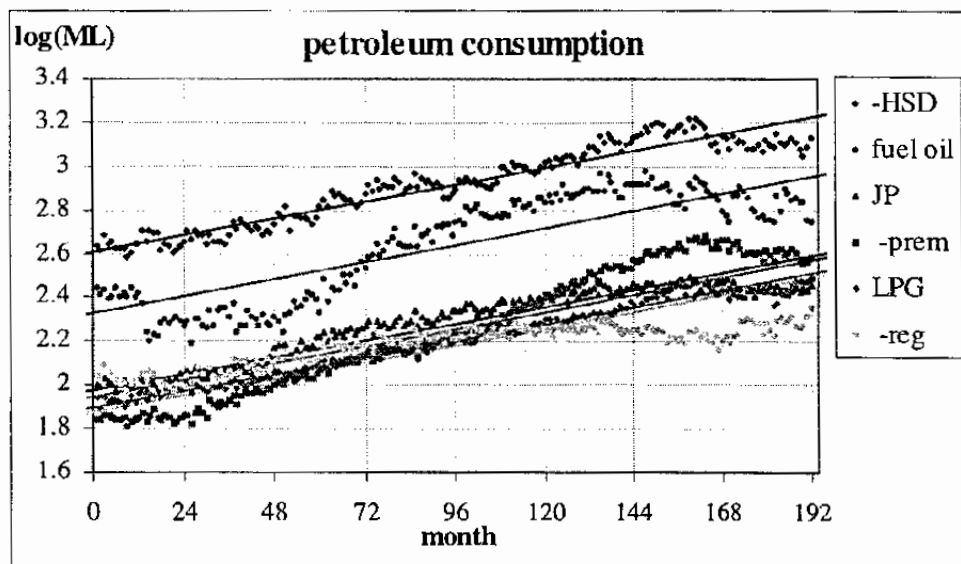


Figure 3.9: Graph of linear model fitted to petroleum consumption: 1984-1999

Figure 3.10 shows a graph of the consumption trends after adjusting for type of component. From this graph it is clear that the linear model does not provide a

satisfactory fit to the data. The trend is not linear, but curved. This curve is flat for the first four years (from 1984 to 1987), then increases linearly over the next eight years, and flattens out again in the last four years (from 1996 to 1999).

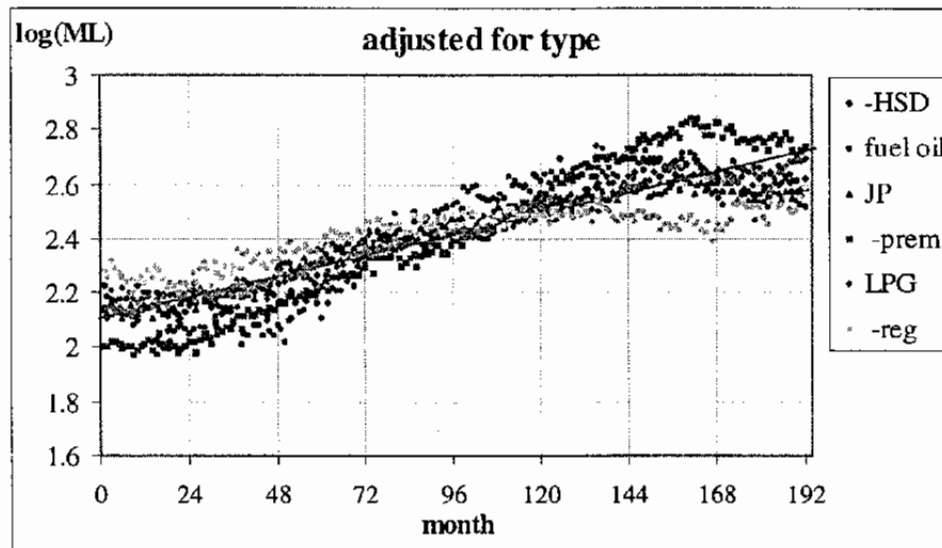


Figure 3.10: Graph of linear model fitted to petroleum consumption after adjusting for the type of fuel: 1984-1999

A less restrictive model, which does not require that the petroleum consumption components follow a linear trend, is provided by two-way analysis of variance. This analysis allows for a separate effect for each of the six main fuel types and adjusts for each of the 192 months. However, since fitting 192 separate period effects seems excessive, we will fit a smaller model containing an adjustment for each quarter rather than an adjustment for each month. This gives rise to a model with six fuel type parameters and 63 quarterly adjustment parameters corresponding to the 64 quarters.

Before fitting the two-way anova, we first show the result from the simple one-way anova, in which no adjustment is made for any time trend. Figure 3.11 shows the graphs of the box plots (on the left) and 95% confidence intervals for the means (on the right), for the base 10 logarithms of the monthly fuel consumption amounts.

The box plots are used to compare amounts of consumption for the different petroleum products. The confidence intervals show the differences in the means of petroleum consumption for the different types. There are two types showing

differences from the others, namely HSD and fuel oil. The mean of HSD is the highest.

Now to compare type after adjusting for quarter, Figure 3.12 shows box plots of the consumptions amounts for the different types after adjusting for quarterly periods. Again, the spreads corresponding of JP, premium, LPG and regular are very similar, they are different from HSD and fuel oil. We see that HSD is the highest.

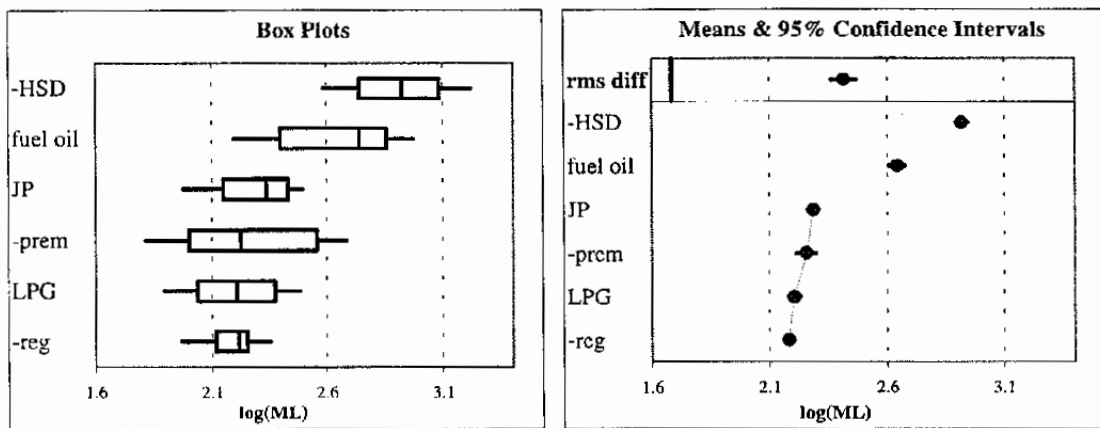


Figure 3.11: Box plots and confidence intervals for means of the six main types of petroleum consumption over months from 1984-1999

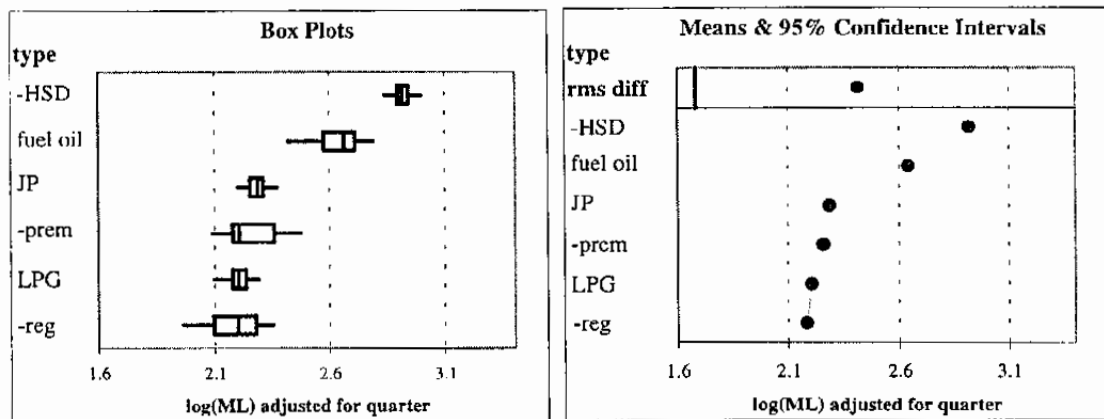


Figure 3.12: Comparison of petroleum consumption by type, adjusted for quarter

Comparing the box plots in Figures 3.11 and 3.12, you can see how adjusting for the quarterly trend substantially reduces the variation within the samples. And the confidence intervals shown in the right-hand panels are even narrower after this adjustment. For example, you can see that the unadjusted means for fuel types JP, premium gasoline, LPG and regular gasoline are not statistically separated. However,

after making the adjustment for trend, Figure 3.12 shows that the all means can be separated except for LPG and regular gasoline.

Table 3.3 shows the means of the log-transformed petroleum consumption amounts for the different fuel types after adjusting for the trend in successive quarters. We see that the differences are statistically highly significantly ($F = 2846.941$, $P\text{-value} = 0.00$). The sum of squares of residuals is 6.508, the residual standard deviation is 0.078, the root-mean-squared difference between the sample means is 0.731, and the 95% confidence interval for this difference ranges from 0.711 to 0.751. However, the p -value for Levene's test (0.000) indicates that the populations do not have the same standard deviation. The table shows that the standard deviations for premium and regular petrol and fuel oil are more than double the standard deviations for the other three fuel types.

Comparison: petroleum consumption								
type	Size	Mean	SE	StDev	Mean	SE	StDev	
-HSD	192	2.920	0.013	0.182	2.920	0.002	0.034	
fuel oil	192	2.642	0.017	0.240	2.642	0.006	0.084	
JP	192	2.286	0.012	0.163	2.286	0.003	0.037	
-prem	192	2.257	0.021	0.285	2.257	0.008	0.108	
LPG	192	2.206	0.013	0.174	2.206	0.003	0.038	
-reg	192	2.184	0.007	0.093	2.184	0.008	0.107	
				log(ML)	adjusted for quarter			
				Twoway Anova: Resid SS: 6.508 r-sq: 0.950				
factor	df	F	p-val	s	rms	CI/2	Lev p	
type	5, 1146	431.798	0.0000	0.199	0.731	0.052	0.000	
quarter+	63, 1083	102.743	0.0000	0.078				
type+	5, 1083	2846.941	0.0000	0.078	0.731	0.020	0.000	

Table 3.3: Results from two-way anova of petroleum consumption

Note that the r -squared goodness-of-fit based on this model is 95%, which is not substantially higher than the r -squared statistic we obtained earlier from fitting the linear regression model. However, a more complex model involving interaction effects would be needed to make a substantial improvement in the goodness-of-fit.

3.2 Data Summaries of Petroleum Imports and Production Levels

Figure 3.13 shows histograms and brief numerical summaries of monthly imports and production of petroleum products in Thailand from January 1984 to December 1999 in million of litres (ML). It shows the mean, standard deviation, minimum and maximum for each product of the total monthly imports and production totals.

The average of the total monthly imports and production is 2,332.271 ML, and the largest component is HSD (mean = 944.647), which accounts for about 40.50% of the total on average. At the other extreme, LSD only accounts for 10.883 ML per month on average, and the level of kerosene imports and production is only 13.235 ML per month on average. These two components comprise together approximately 1% of the total.

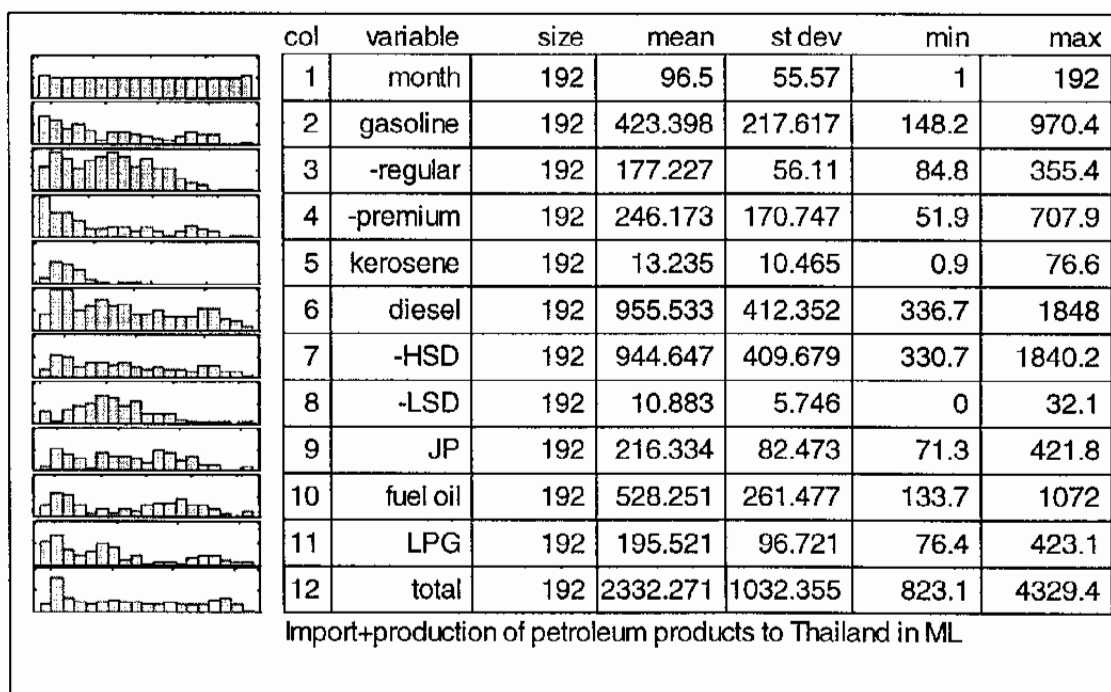


Figure 3.13: Monthly imports and production of petroleum products to Thailand, 1984-1999

Figure 3.14 shows the time series of total monthly imports and production. The time series is constant in the third-first year. Between 1987 and 1997 this series gradually increases, reaching a maximum in 1997, while after 1997 there is a downward trend. In 1999 there is a slight decrease.

It appears from Figure 3.14 that there is not constant variability in the monthly imports and production levels when the levels themselves are changed. Figure 3.15 shows a graph of the magnitude of the change in the imports and production levels from one month to the next versus the level itself. The correlation coefficient is 0.419, confirming that the change in level is related to the level.

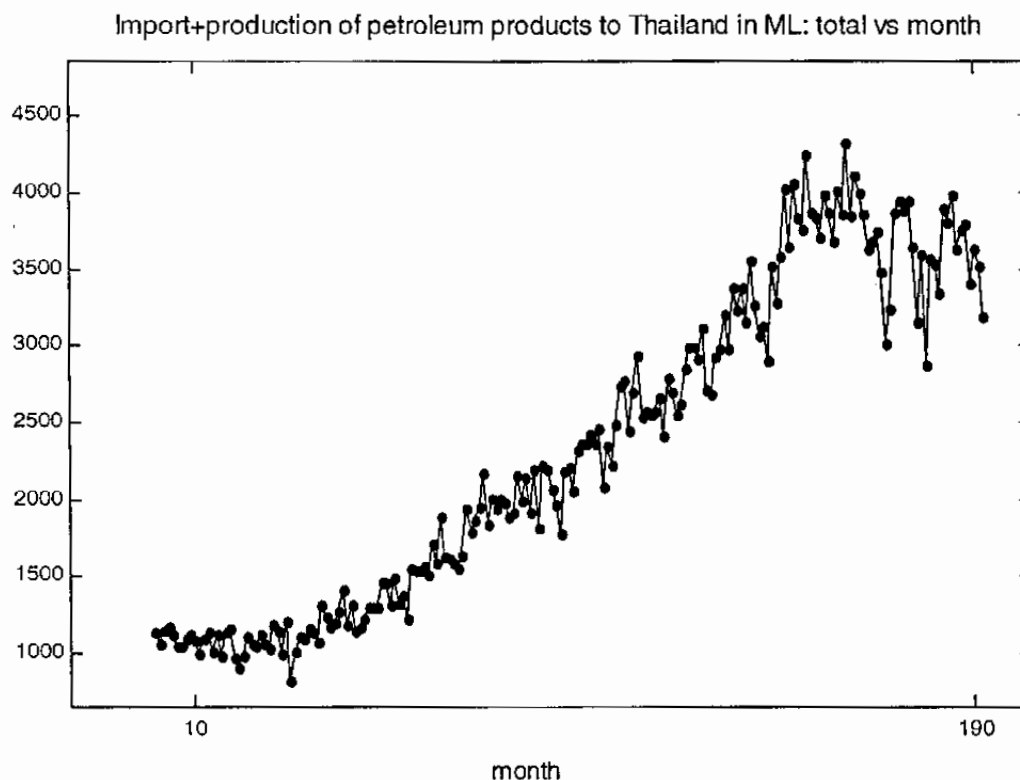


Figure 3.14: Trend of monthly imports and production of total petroleum products in Thailand, 1984-1999

Figure 3.16 shows the same graph after taking base-10 logarithms of the total monthly imports and production levels. That the logarithm transformation is effective in removing the relation between the monthly changes in petroleum imports and production levels and the levels themselves. In this case the correlation coefficient between the change and the level is negative (-0.107).

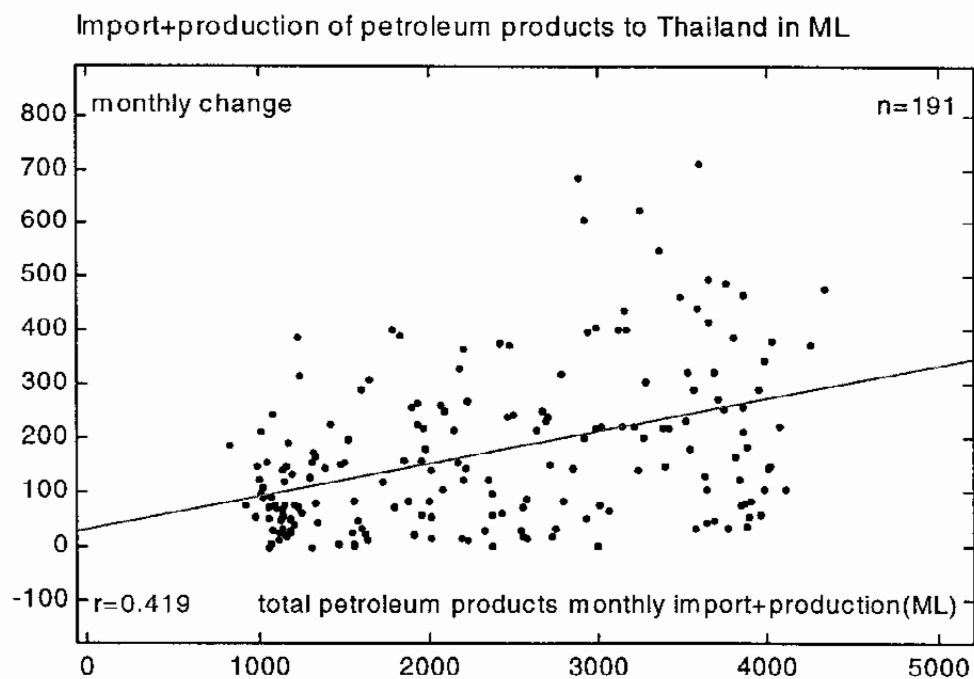


Figure 3.15: Monthly change versus level, for total petroleum imports and production

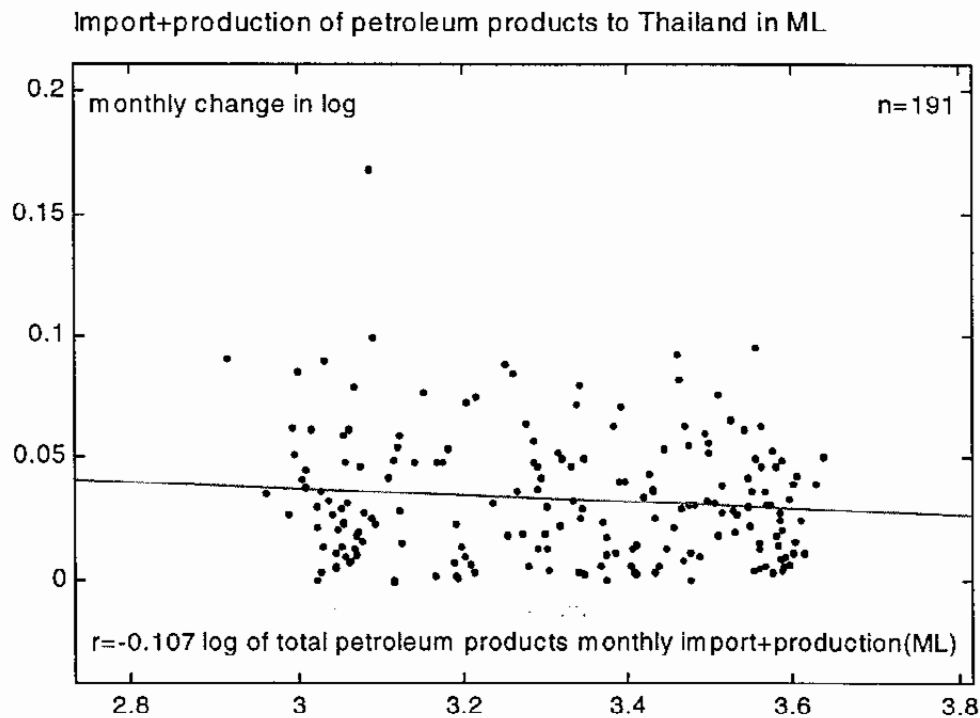


Figure 3.16: Monthly change versus level, for log-transformed total petroleum imports and production

Figure 3.17 shows the trend of the monthly imports and production levels after logarithm transformation. The trend is the same as the trend shown in Figure 3.14, but the variability is now nearly constant.

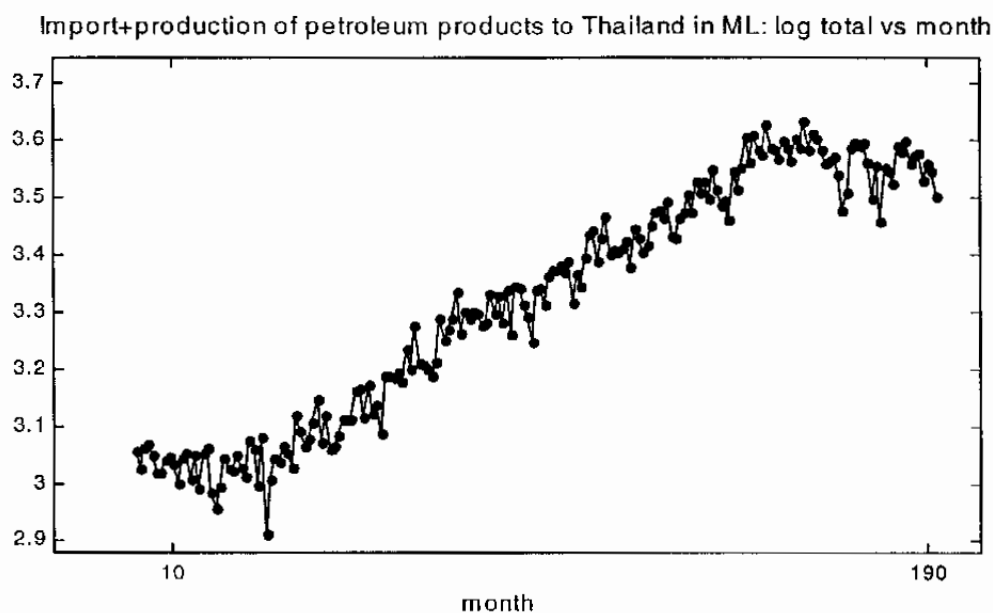


Figure 3.17: Trend of monthly imports and production of total petroleum products to Thailand, 1984-1999 after transforming data by taking logarithm.

Figure 3.18 summarizes the total of imports and production of the main petroleum products after transformation of the total amounts for each fuel type using base 10 logarithms. The components are ordered according to their means.

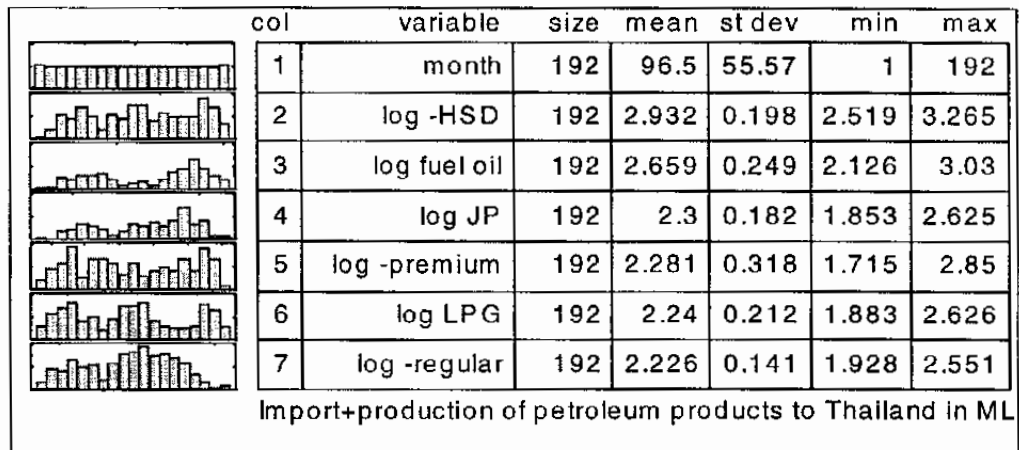


Figure 3.18: Histograms and statistics of base 10 logarithm of imports and production totals of main petroleum products

Figure 3.19 shows the relations between the various components. Most of the relations between pairs of components are close to linear. The correlation coefficients between all of the pairs of components are shown in Figure 3.19 and Table 3.4. These correlation coefficients range from 0.858 and 0.862 (between fuel oil and LPG) to 0.959 and 0.964 (between HSD and premium gasoline), respectively.

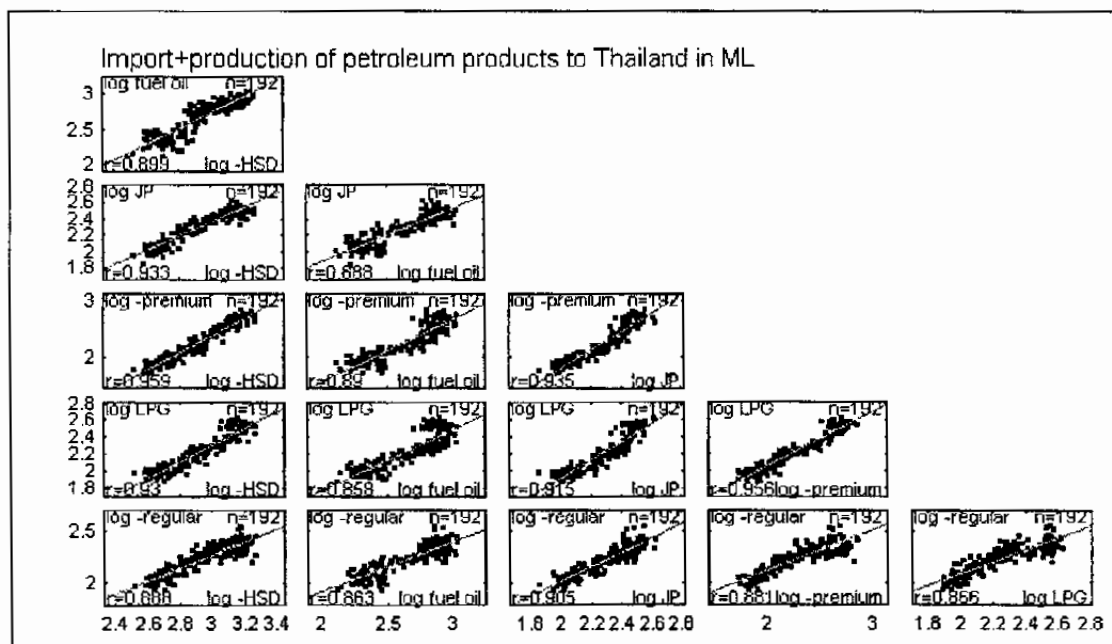


Figure 3.19: Scatterplot matrix showing relations between components of petroleum imports and production totals

Figure 3.20 shows how the various components of petroleum imports and production totals change with time over the 16-year period from 1984 to 1999.

	-HSD	fuel oil	JP	-premium	LPG	-regular
-HSD	1					
fuel oil	0.903	1				
JP	0.938	0.892	1			
-premium	0.964	0.895	0.940	1		
LPG	0.935	0.862	0.920	0.961	1	
-regular	0.893	0.867	0.910	0.886	0.870	1

Table 3.4: Correlation matrix for imports and production totals fuel types

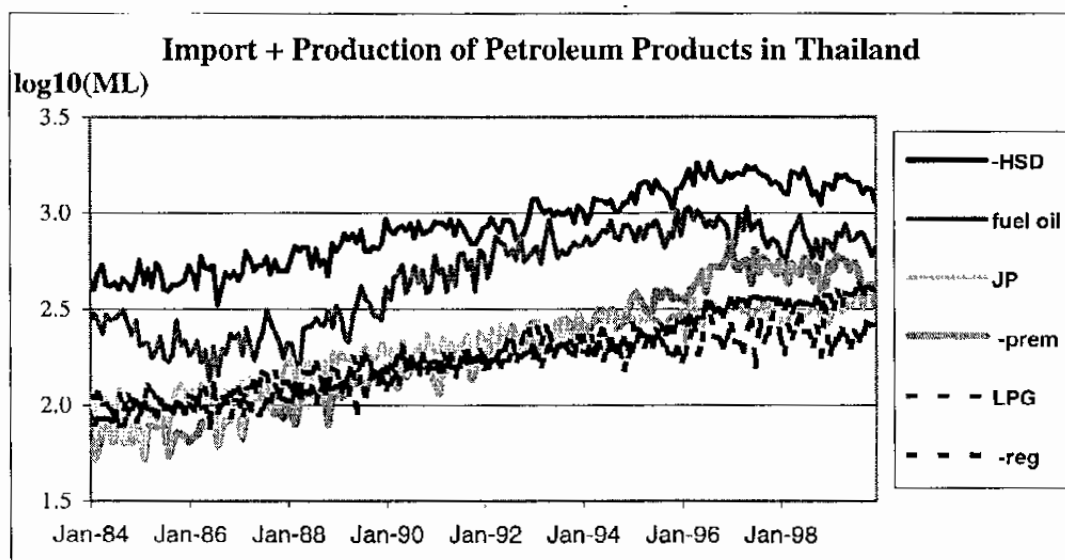


Figure 3.20: Time-series plot of the six main petroleum products in Thailand during 1984-1999

Imports and production totals of petroleum products were increased over the period. JP, premium gasoline, LPG and regular gasoline increase together. HSD is the highest and regular gasoline is the slowest.

Table 3.5 and Figure 3.21 show the results of fitting a model to the monthly imports and production totals. In this model we assume that each component increases linearly over the 16-year period, and that the slopes of these fitted lines are the same for each component.

Each parameter in this model is highly statistically significant with all the p-values close to 0. The r-squared is 0.929, so this model accounts for 92.9% of the variation in the data.

Relation: petroleum import and production				
outcome: log10(ML) 1152 cases				
predictor	coeff	SE	t	p-value
constant	2.5783	0.0082	315.226	0.0000
month	0.0037	0.0000	74.672	0.0000
type	(0 for -HSD)			0.0000
fuel oil	-0.2736	0.0094	-29.016	0.0000
JP	-0.6325	0.0094	-67.083	0.0000
-prem	-0.6513	0.0094	-69.082	0.0000
LPG	-0.6926	0.0094	-73.456	0.0000
-reg	-0.7059	0.0094	-74.874	0.0000
r-sq: .929 Resid SS: 9.771 s: .0924 df: 1145				

Table 3.5: Fitted linear model for petroleum imports and production totals

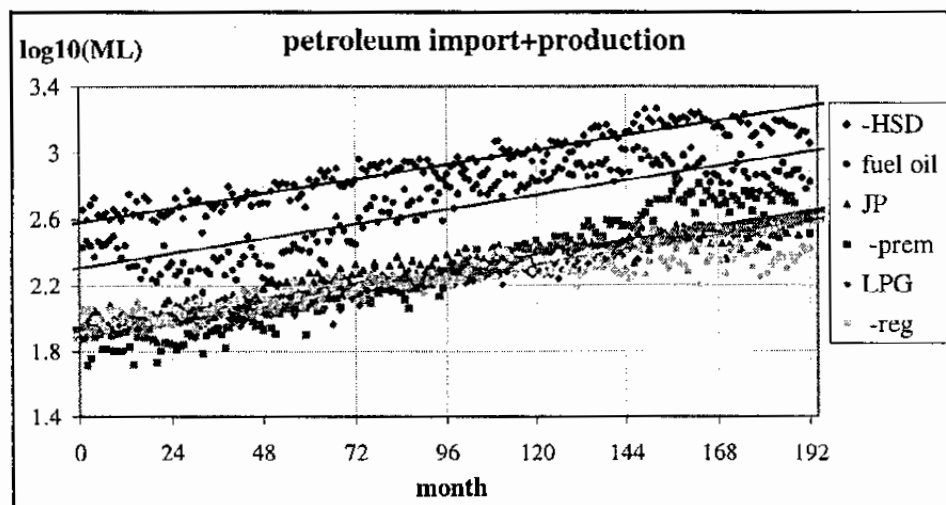


Figure 3.21: Graph of linear model fitted to petroleum supply: 1984-1999

Figure 3.22 shows a graph of the imports and production totals trends after adjusting for type of component. From this graph it is clear that the linear model does not provide a satisfactory fit to the data. The trend is not linear, but curved. This curve is spread not too much for the first four years (from 1984 to 1987) except LPG that is far away from the fit line, then increases linearly over the next eight years, and flattens out again in the last four years (from 1996 to 1999).

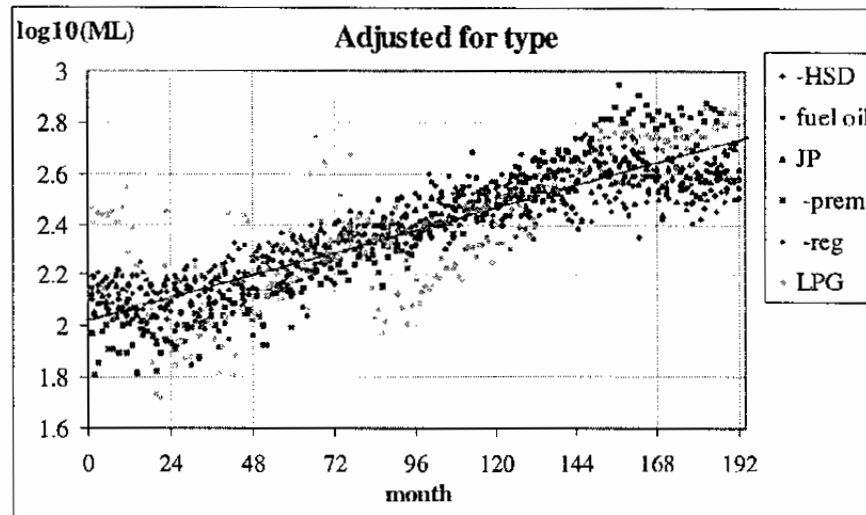


Figure 3.22: Graph of linear model fitted to petroleum imports and production after adjusting for the type of fuel: 1984-1999

Figure 3.23 shows the graphs of the box plots (on the left) and 95% confidence intervals for the means (on the right), for the base 10 logarithms of the monthly fuel imports and production total amounts.

A less restrictive model, which does not require that the petroleum imports and production totals components follow a linear trend, is provided by two-way analysis of variance. This analysis allows a separate effect for each of the six main fuel types and adjusts for each of the 192 months. However, since 192 separate period effects is too many, we will fit a smaller model containing an adjustment for each quarter rather than an adjustment for each month. This gives rise to a model with six fuel type parameters and 63 quarterly adjustment parameters corresponding to the 64 quarters.

Before fitting the two-way anova, we first show the result from the simple one-way anova, with no adjustment is made for any time trend. Figure 3.23 shows the graphs of the box plots (on the left) and 95% confidence intervals for the means (right), for the base 10 logarithms of the monthly fuel imports and production totals amounts.

The box plots are used to compare amounts of imports and production totals for the different petroleum products. The confidence intervals show the differences in the means of petroleum imports and production totals for the different types. There are two types showing differences from the others, namely HSD and fuel oil. The mean of HSD is the highest.

Now to compare type after adjusting for quarter, Figure 3.24 shows box plots of the imports and production totals amounts for the different types after adjusting for quarterly periods. Again, the spreads corresponding of JP, premium gasoline and LPG, regular gasoline are similar, but different from HSD and fuel oil. We see that HSD is the highest.

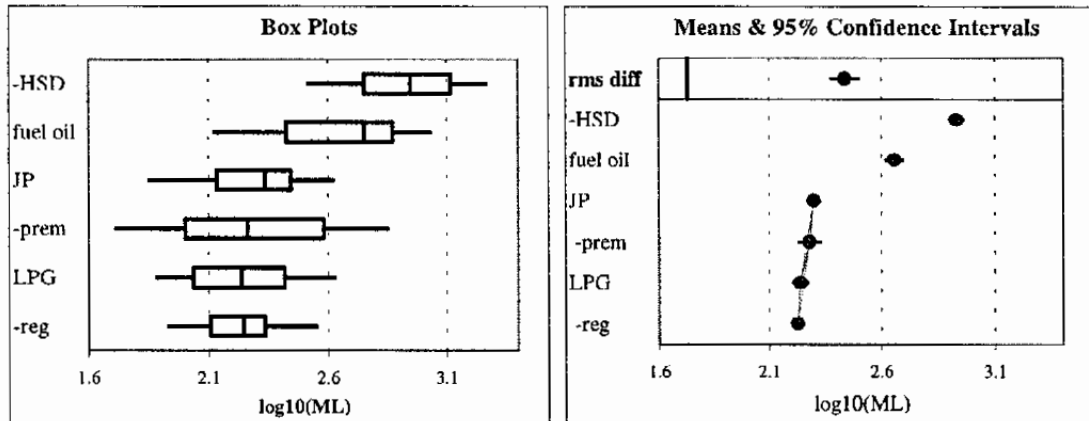


Figure 3.23: Box plots and confidence intervals for mean of six main types of petroleum imports and production over months from 1984-1999

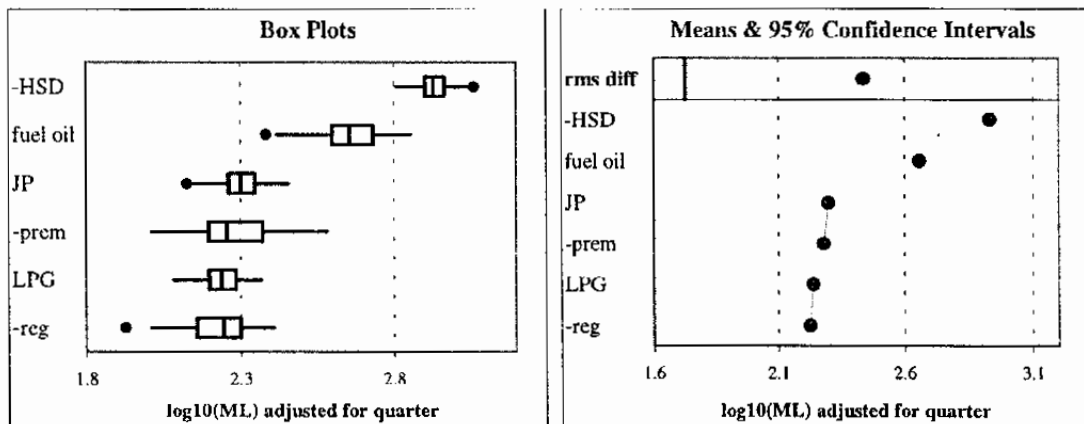


Figure 3.24: Comparison of petroleum imports and production by type, adjusted for quarter

Comparing the box plots in Figures 3.23 and 3.24, you can see how adjusting for the quarterly trend substantially reduces the variation within the samples. And the confidence intervals shown in the right-hand panels are even narrower after this adjustment. For example, you can see that the unadjusted means for fuel types JP, premium gasoline, LPG and regular gasoline are not statistically separated. However, after making the adjustment for trend, Figure 3.24 shows that the all means can be separated except for LPG and premium gasoline.

Table 3.6 shows the means of the log-transformed petroleum imports and production totals amounts for the different fuel types after adjusting for the trend in successive quarters. We see that the differences are statistically highly significantly ($F = 2186.43$, $p\text{-value} = 0.000$). The sum of squares of residuals is 8.005, the residual standard deviation is 0.086, the root-mean-squared difference between the sample means is 0.711, and the 95% confidence interval for this difference ranges from 0.688 to 0.734. However, the $p\text{-value}$ for Levene's test (0.000) indicates that the populations do not have the same standard deviation. The table shows that the standard deviations for premium and regular gasoline are more than twice the value of the standard deviation for fuel oil.

Comparison: petroleum import+production								
type	Size	Mean	SE	StDev	Mean	SE	StDev	
-HSD	192	2.932	0.014	0.198	2.932	0.004	0.050	
fuel oil	192	2.659	0.018	0.249	2.659	0.007	0.092	
JP	192	2.300	0.013	0.182	2.300	0.004	0.061	
-prem	192	2.281	0.023	0.318	2.281	0.009	0.121	
LPG	192	2.240	0.015	0.212	2.240	0.004	0.057	
-reg	192	2.226	0.010	0.141	2.226	0.007	0.097	
		log10(ML)			adjusted for quarter			

Twoway Results:					Resid SS: 8.005 r-sq: .942			
factor	df	F	p-val	s	rms	CI/2	Lev	p
type	51,146	322.937	0.0000	0.224	0.711	0.059	0.000	
quarter+	631,083	105.967	0.0000	0.086				
type+	51,083	2186.430	0.0000	0.086	0.711	0.023	0.000	

Table 3.6: Results from two-way anova of petroleum imports and production totals

Note that the $r\text{-squared}$ goodness-of-fit based on this model is 94.2%, which is not much higher than the $r\text{-squared}$ statistic we obtained earlier from fitting the linear regression model (92.9%). However, a more complex model involving interaction effects would be needed to make a substantial improvement in the goodness-of-fit.

In the next section, we investigate the relation between fuel consumption and fuel imports and production for each of the six main fuel types.

3.3 Relation between Consumption and Imports and Production

In this section we look at the relation between the consumption and the imports and production (that is, the supply) by using time series graphs and scatter plots.

Figure 3.25 shows time series graphs for high speed diesel fuel and fuel oil with both the consumption (heavy line) and imports and production amounts plotted on the same axis. We see that for both HSD and fuel oil, the supply fluctuates much more than the demand, but is mostly sufficient to meet it. However, in the most recent three-year period, from 1997 to 1999 the HSD supply was slightly higher than the demand.

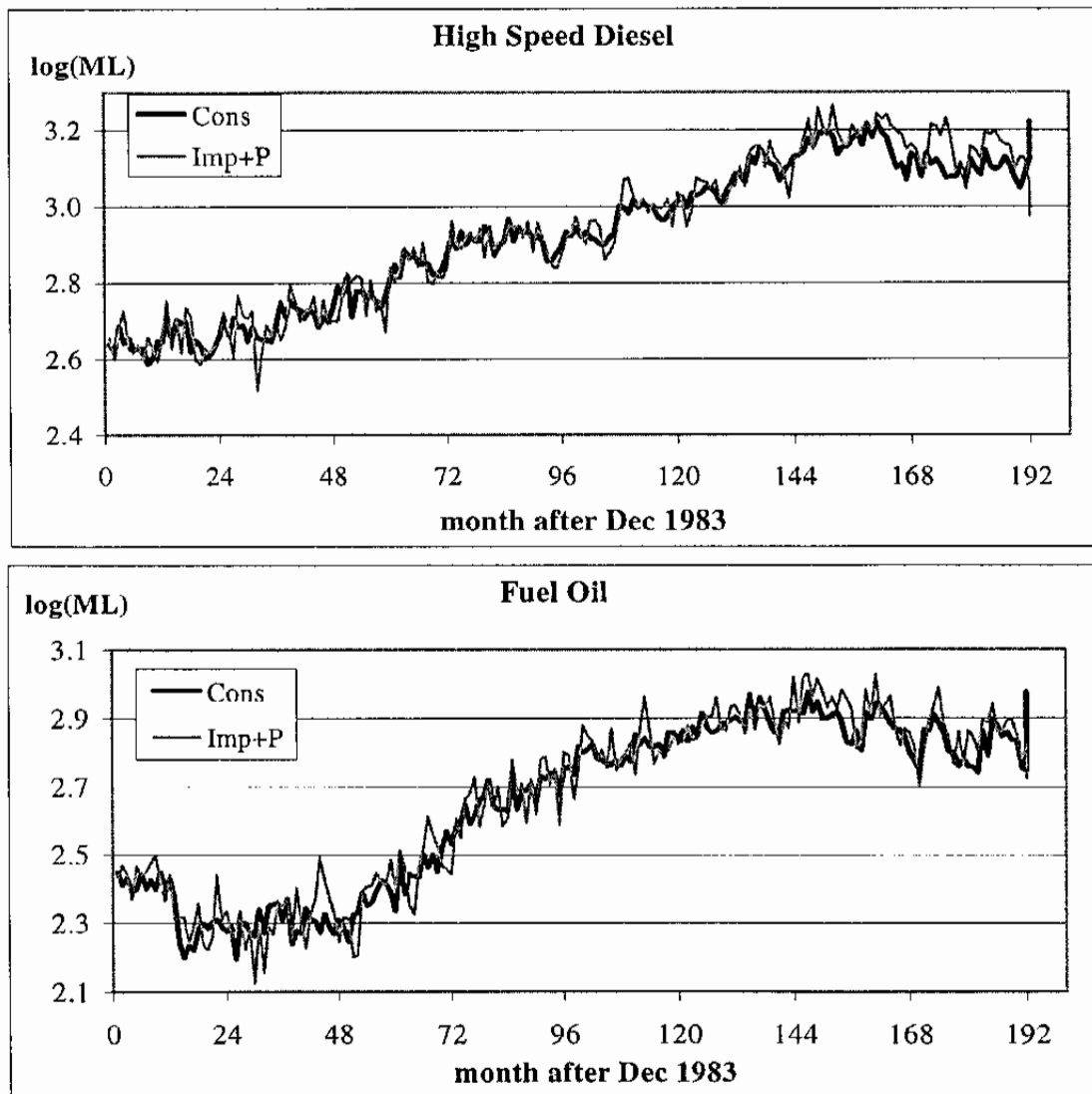


Figure 3.25: Relations between monthly consumption and imports and production amounts for high speed diesel oil and fuel oil

Figure 3.26 shows time series graphs for jet petrol fuel and liquid petroleum gas with both the consumption (heavy line) and imports and production amounts again plotted on the same axis. We see that for jet petrol and LPG the patterns are much the same as for HSD. The supply fluctuates more than the consumption, but are mostly sufficient to meet the demand. However, in the most recent three-year period, from 1997 to 1999 their supplies were slightly higher than the demand. LPG has oversupply more than fuel oil.

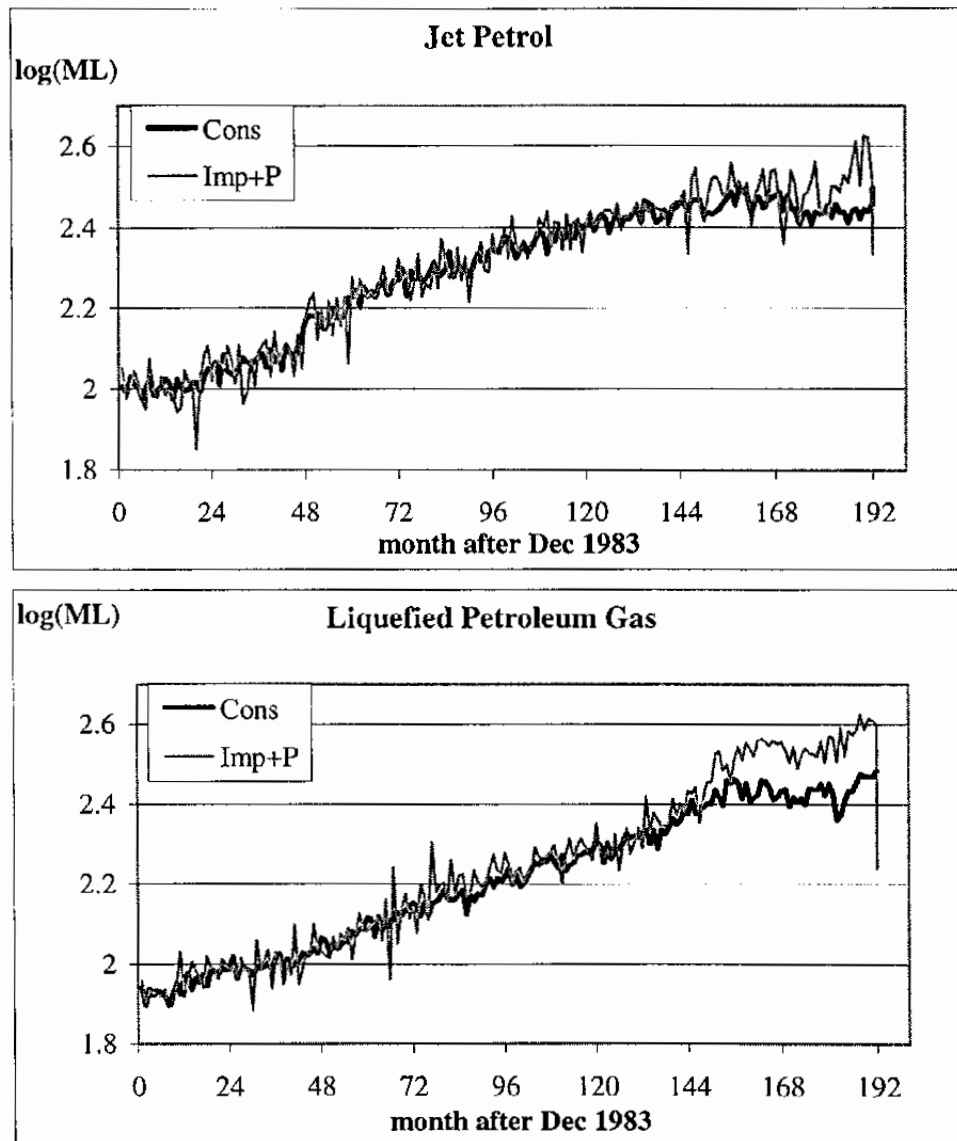


Figure 3.26: Relations between monthly consumption and imports and production amounts for jet petrol and LPG

Figure 3.27 shows time series graphs for premium and regular gasoline with both the consumption (heavy line) and imports and production amounts again plotted on the same axis. It shows that premium is sufficient for the demand of the amount of premium, and the supply outstripped the demand in the more recent years. It also shows that regular is sufficient for the demand of the amount of regular. However, in the eight-year period, since 1992 to 1999, the regular supply was substantially higher than the demand.

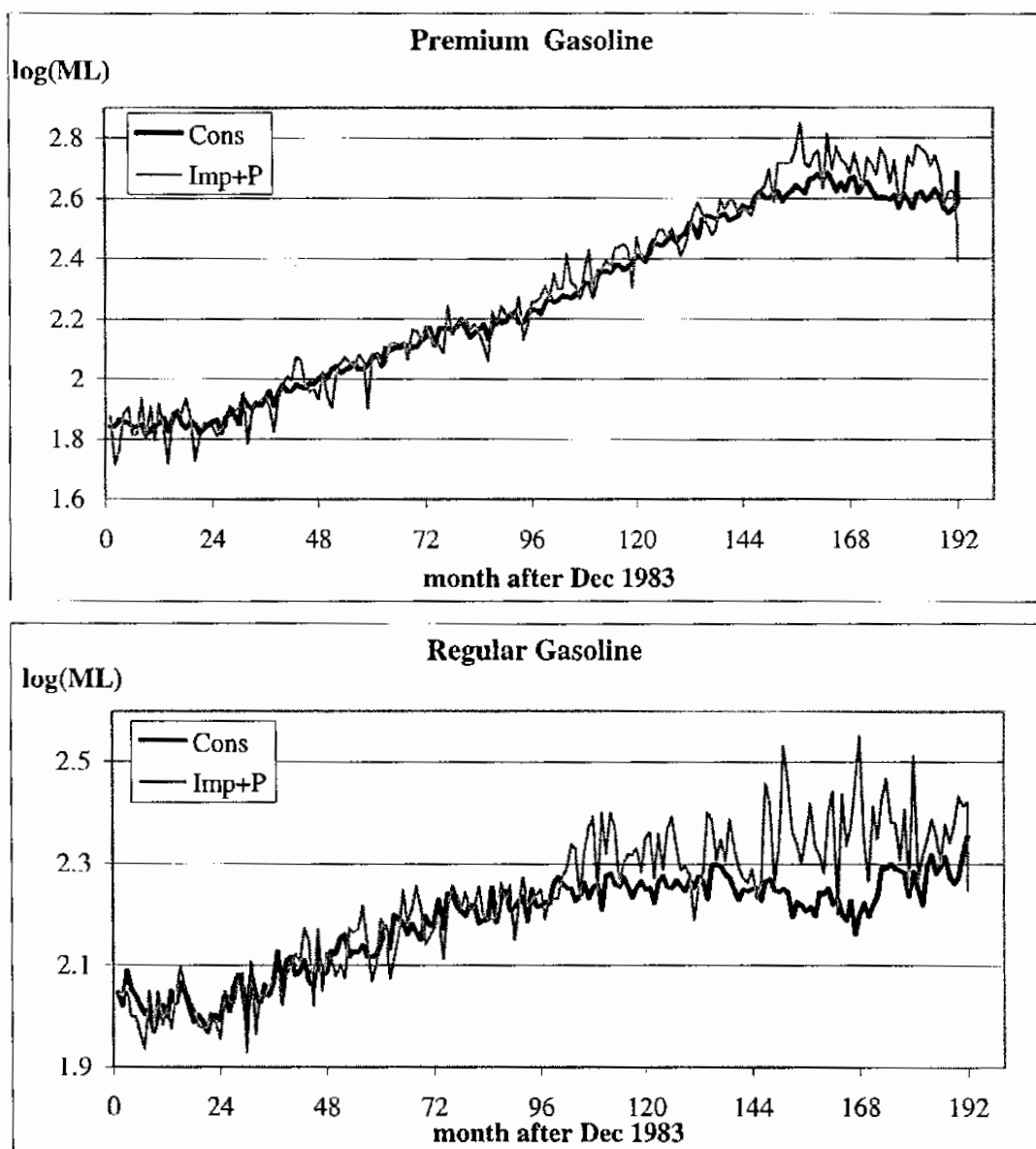


Figure 3.27: Relations between monthly consumption and imports and production amounts for premium and regular gasoline

We expect that the amount of consumption for each fuel type should be at least approximately the same as the total amount imported and produced. If the consumption amounts were less than the corresponding imports and production totals, there would be some stockpiling of resources, with associated storage costs. And if the consumption amounts were greater than the amounts produced and imported, there would have to be a stockpile to cater to the demand.

Next, we investigate the relation between the supply and the demand by using scatter plots. Figure 3.28 to Figure 3.30 show these relations for HSD and fuel oil.

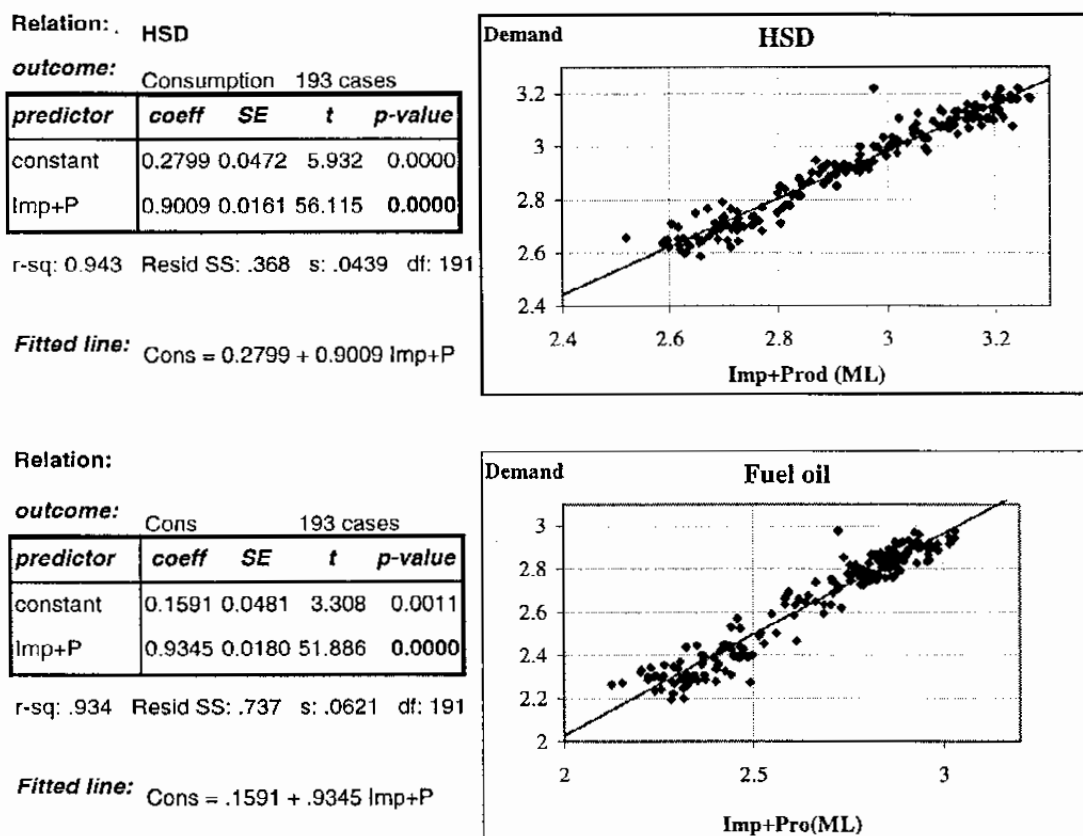


Figure 3.28: Scatter plots and fitted linear relations for HSD and fuel oil

When the supply is increased 1 ML, so the demand of HSD and fuel oil will be increased 0.9009 and 0.9345 ML, and r-square are 94.3% and 93.4%, respectively.

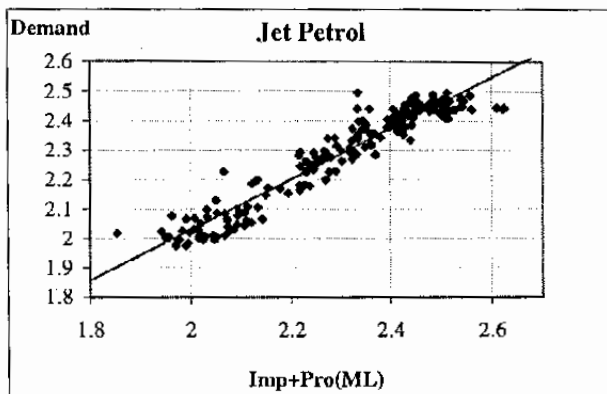
Relation:

outcome: Cons 193 cases

predictor	coeff	SE	t	p-value
constant	0.3044	0.0431	7.069	0.0000
Imp+P	0.8623	0.0187	46.196	0.0000

r-sq: .918 Resid SS: .422 s: .047 df: 191

Fitted line: Cons = .3044 + .8623 Imp+P



Relation:

outcome: Cons 193 cases

predictor	coeff	SE	t	p-value
constant	0.4131	0.0335	12.325	0.0000
Imp+P	0.8011	0.0149	53.761	0.0000

r-sq: .938 Resid SS: .364 s: .0436 df: 191

Fitted line: Cons = .4131 + .8011 Imp+P

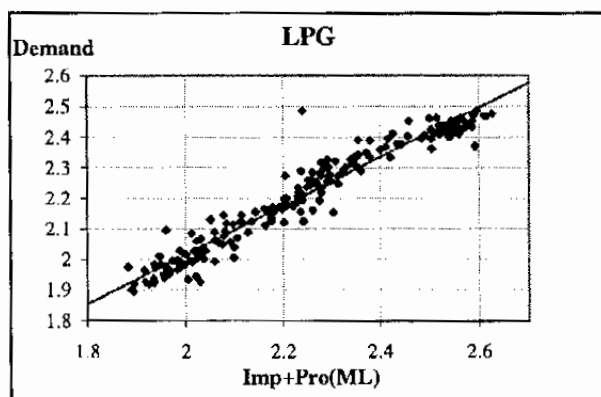


Figure 3.29: Scatter plots and fitted linear relations for jet petrol and LPG

In figure 3.29 and 3.30, we can explain when the is supply increased 1 ML, so the demand of JP and premium gasoline will be increased 0.8623 and 0.885 ML, and r-square are 91.8% and 96.2%, respectively. And, when the supply is increased 1 ML, so the demand of LPG and regular gasoline will be increased 0.8011 and 0.575 ML, and r-square are 93.8% and 74.6%, respectively.

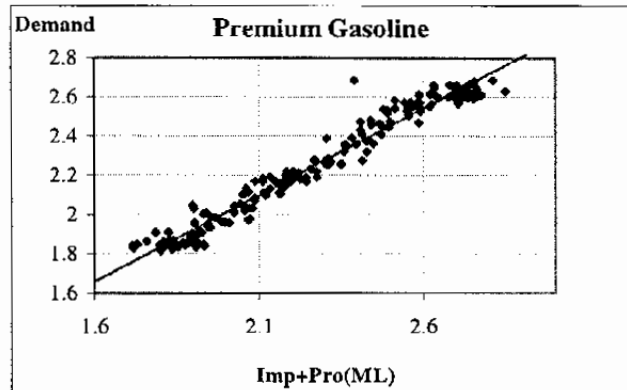
Relation:

outcome: Cons 193 cases

predictor	coeff	SE	t	p-value
constant	0.2406	0.0292	8.252	0.0000
Imp+P	0.8850	0.0127	69.906	0.0000

r-sq: .962 Resid SS: .59 s: .0556 df: 191

Fitted line: Cons = .2406 + .885 Imp+P



Relation:

outcome: Cons 193 cases

predictor	coeff	SE	t	p-value
constant	0.9044	0.0541	16.707	0.0000
Imp+P	0.5750	0.0243	23.693	0.0000

r-sq: .746 Resid SS: .429 s: .0474 df: 191

Fitted line: Cons = .9044 + .575 Imp+P

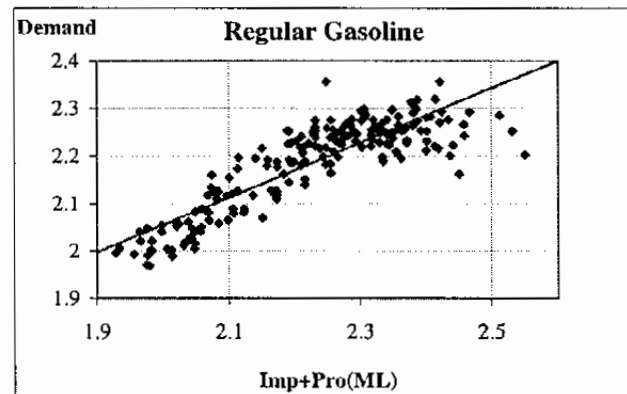


Figure 3.30: Scatter plots and fitted linear relations for premium and regular

In Chapter 4 we use time series analysis to investigate and model the trends in the monthly series of petroleum consumption and imports and production.