Maxime Maltais
Time Series – Summer 2011
Student Project
September 17, 2011

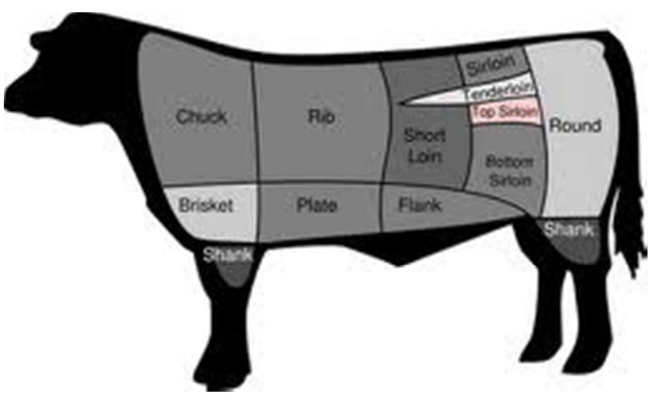
A kilogram of round steak

Introduction

For my time series project, I wanted to analyse data that are unusual and have fun while doing it. With that in mind, I went on Statistics Canada website looking and exploring for a dataset.

I end up being inspired by a table giving the price of a kilogram of round steak for the past months. Maybe that the smell of my neighbour's BBQ triggered something in my brain and that it is why I found this dataset interesting at that particular moment. Right now you probably think that it is a bit ridiculous but you will see that I found this dataset reveal a dark period in bovine history. But before, the following figure shows which part of the beef is the round steak.

Figure 1: Standard beef cuts



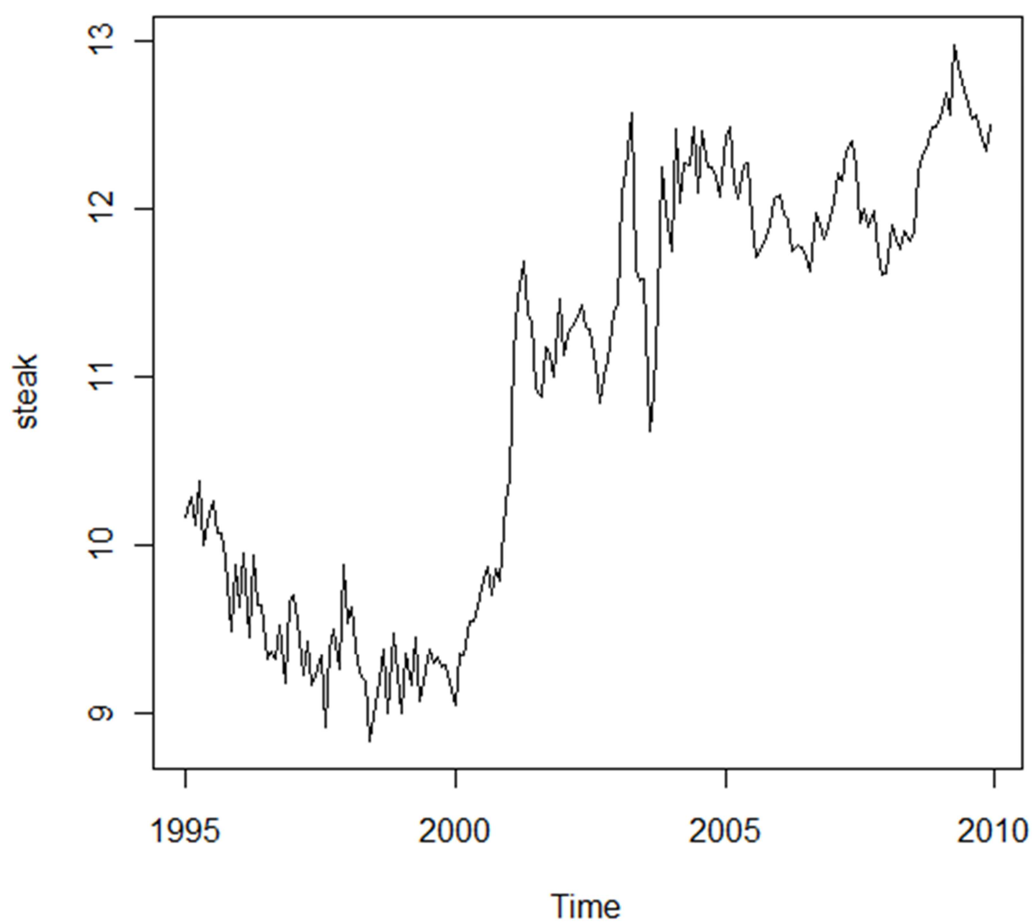
All the calculations and analysis were performed using R. Programming code to replicate figures is presented in Appendix B.

Data

Data used for my analysis has been downloaded from Statistics Canada website (www.statcan.gc.ca, CANSIM table 326-0012 v735165 Canada; Round steak, 1 kilogram (monthly, 1995-01-01 to 2009-12-01)).

The average price of a kilogram of round steak is recorded monthly as a part of the Consumer price indexes (CPI) survey. The following graph shows the evolution of the price from January 1995 to December 2009. Complete data are provided in Appendix A.

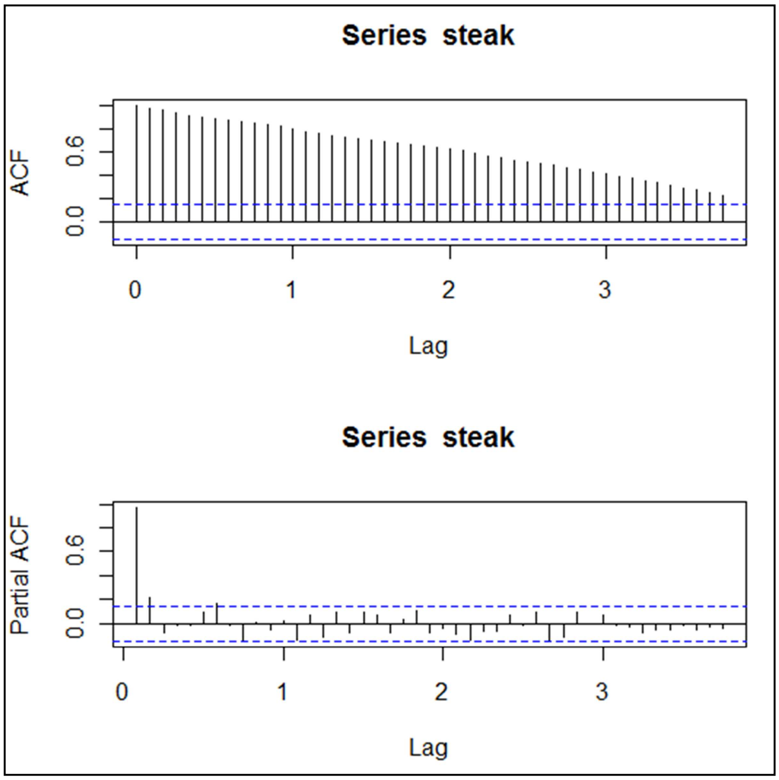
Figure 2: Average retail price (monthly) of a kilogram of round steak (refer as *steak*)



Data analysis and process selection

We can see from figure 2 that our data are not stationary. We will first take a look at the autocorrelations to confirm this intuition.

Figure 3: ACF and PACF of steak



Presence of a trend

The ACF lead us to think that maybe there is a trend in our data. To see if we have to “detrend” our data, we will fit a regression model of the form:

$$y = \alpha + \beta x.$$

Where the independent variable is time and with the assumption that:

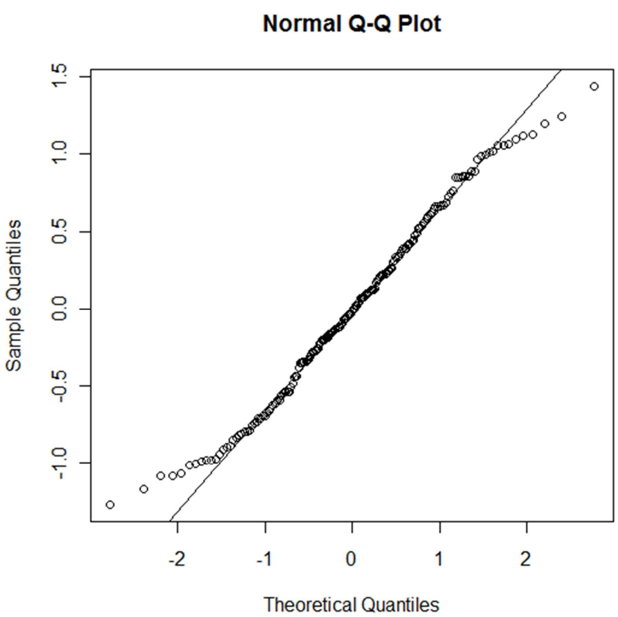
1. Residuals are normally distributed, $\varepsilon \sim N(\mu=0,\sigma^2)$
2. The variance of the error is constant across observations (homoscedasticity)
3. Residuals are independent and not correlated $COV(\varepsilon_i, \varepsilon_j) \neq 0 \ \forall \ i,j$

Figure 4: Regression of steak over time

```
Call:
lm(formula = steak ~ t)

Residuals:
     Min       1Q   Median       3Q      Max 
-1.27399 -0.45526 -0.01156  0.42444  1.44016 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 9.0479100  0.0914641   98.92   <2e-16 ***
t           0.0209193  0.0008765   23.87   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.611 on 178 degrees of freedom
Multiple R-squared: 0.7619,     Adjusted R-squared: 0.7606 
F-statistic: 569.7 on 1 and 178 DF,  p-value: < 2.2e-16
```

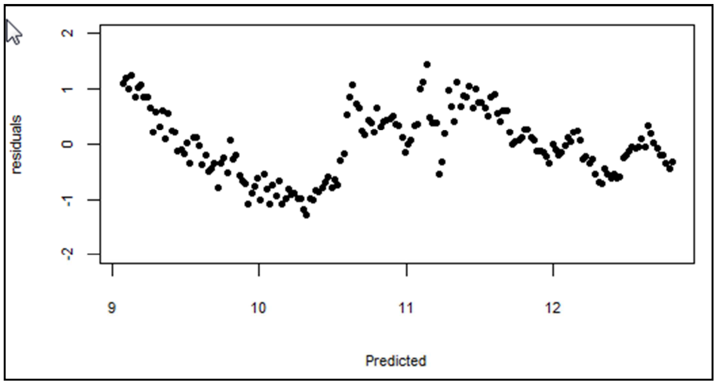
The adjusted R^2 is relatively high so a big part of the variance would be explain and the regression seems significant; small p-value. The following figure will help us assessing if our regression assumptions hold.

Figure 5



The QQ plot of the residuals shows slight normality, although there appears to be a notable deviation from normality in the tails.

Figure 6



A plot of the residuals versus the fitted values does not convince us that the residuals are uncorrelated and have constant variance.

<u>Shapiro-Wilk</u>

Ho: The residuals are normally distributed
Ha: The residuals are not normally distributed

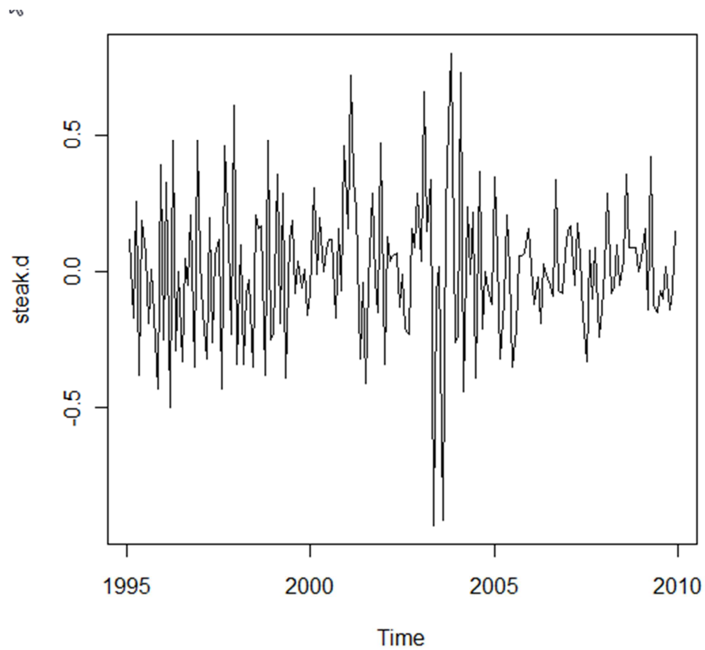
```
> shapiro.test(steak.lm$resid)

        Shapiro-Wilk normality test

data:  steak.lm$resid
W = 0.9849, p-value = 0.04961
```

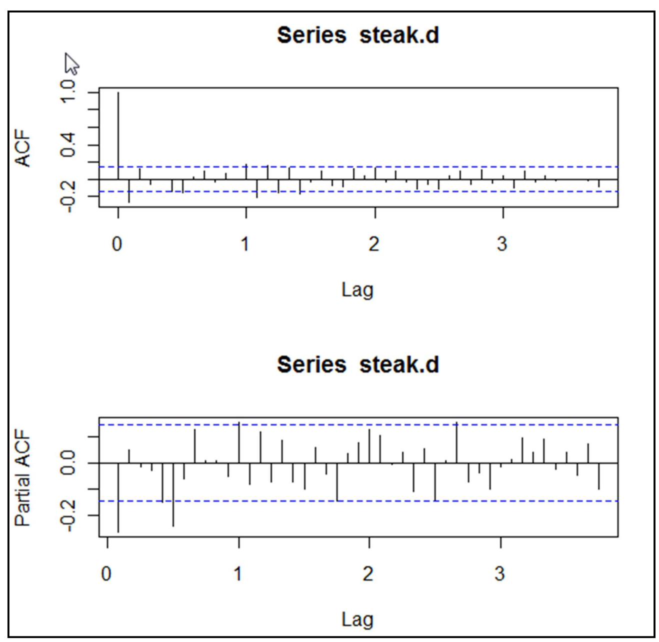
A Shapiro-Wilk normality test produces a p-value of 0.04961 (less than 5 %), which leads us to reject the null hypothesis that the residuals are normally distributed.

First difference

By taking the first difference, we get rid of the appearance of trend and improve toward a stationary series. There is still a lot of variability and it does not seem to be stationary enough.

Figure 7





Differentiation of logarithms

By looking at the percentage of variation from Y_{t-1} to Y_t (see Appendix A-Data), we can observe that it is quite small. Also, since our data represents change in price from time to time, let's try to do the following transformation as suggested in the textbook on page 99:

Figure 8: p99. Jonathan D. Cryer and Kung-Sik Chan, *Time Series Analysis: With Applications in R* (Springer Texts in Statistics, 2009)

$$Y_t = (1 + X_t)Y_{t-1}$$

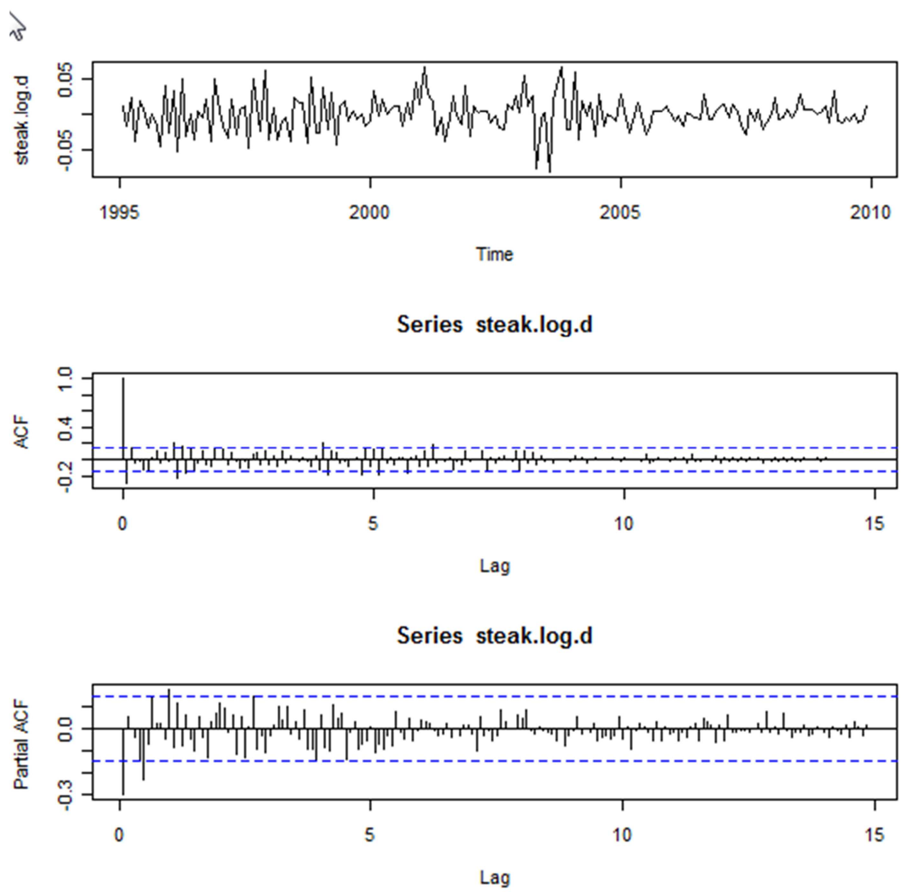
where $100X_t$ is the percentage change (possibly negative) from Y_{t-1} to Y_t. Then

$$\log(Y_t) - \log(Y_{t-1}) = \log\left(\frac{Y_t}{Y_{t-1}}\right)$$
$$= \log(1 + X_t)$$

If X_t is restricted to, say, $|X_t| < 0.2$ (that is, the percentage changes are at most ±20%), then, to a good approximation, $\log(1+X_t) \approx X_t$. Consequently,

$$\nabla[\log(Y_t)] \approx X_t \qquad (5.4.3)$$

Figure 9



By looking at figure 9, the transform series graph and ACF/PACF, we can conclude that the transformation could be adequate and produce a stationary enough dataset to try fitting an ARIMA process on X_t. The ACF/PACF strongly recommend an MA(1) process. ACF peaks at lag 1 and PACF that tails off.

There is still something that’s bugging me with our data. The original series and the transform series have a strange behaviour. From the original series, the price goes down from 1995 to 2000 and after goes up until 2003 and crash. The transform series indicate large volatility from 1995 to 2000. There is a big variation in the middle of 2003 and stabilized after.

What happen in that industry over the last 15 years, why the price of a kilogram of round steak goes up and down like that. We’ve been eating meet for as long the human kind exists; the price is not supposed to be that volatile. We are not in an abnormal inflation era. After a few moment of deep thinking, I found it!. My answer is on the next page.

MAD COW DISEASE

The first appearance of bovine spongiform encephalopathy (BSE) was in 1986 in the UK. From that moment new cases appears quickly mainly Europe. In 1996, it's everywhere in the media, the disease can be transmitted to us, human.

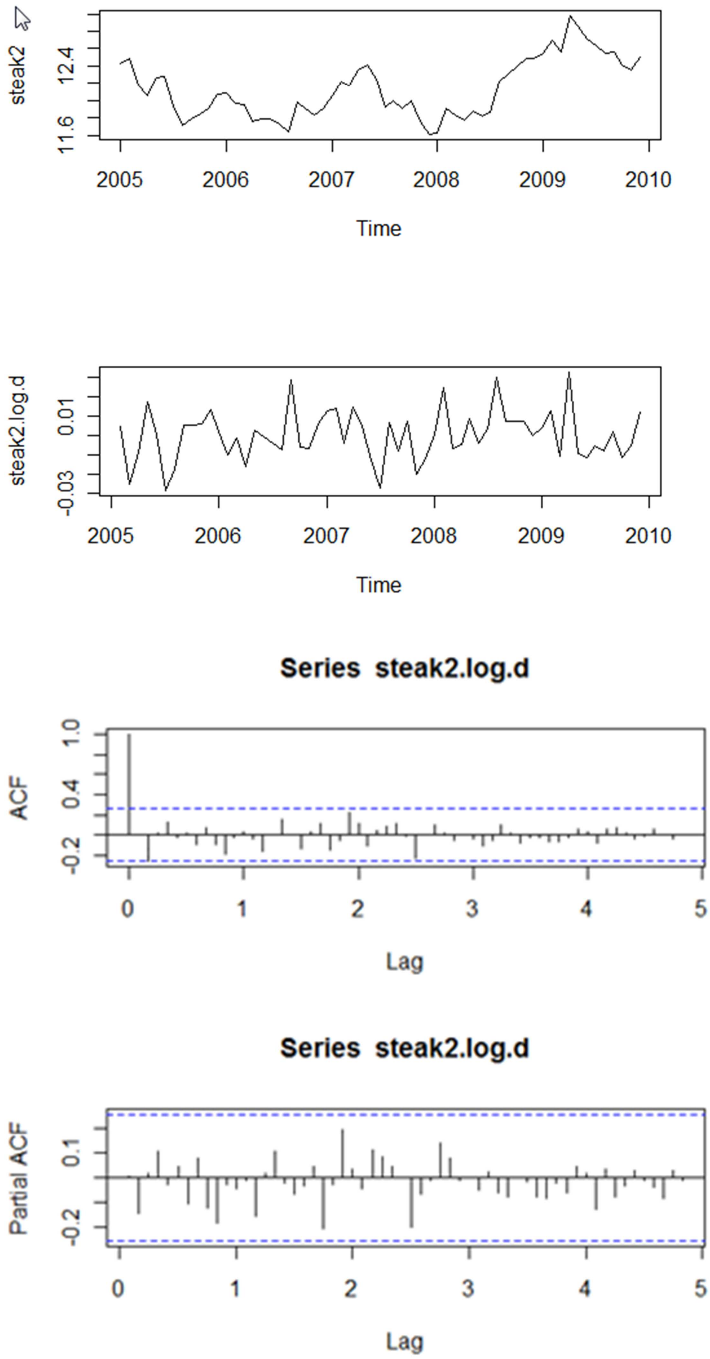
The crisis is huge in Europe and there is a significant reduction of consumption of bovine meet worldwide because people are afraid. Canada exports the major part of its production to the United States and the price goes down. When American and Canadian consumers begin to regain trust, a new case in Alberta is reported (middle of 2003). US banned temporally the importation of Canadian bovine meet, the price crash.

This is a really short summary to better understand our data. For more interesting information, follow this link: http://www.parl.gc.ca/Content/LOP/ResearchPublications/prb0301-e.pdf.

We now better understand our volatility. To remove the effect of the mad cow episode, I kept only data for 2005 and beyond.

Fitting the ARIMA process

Figure 10



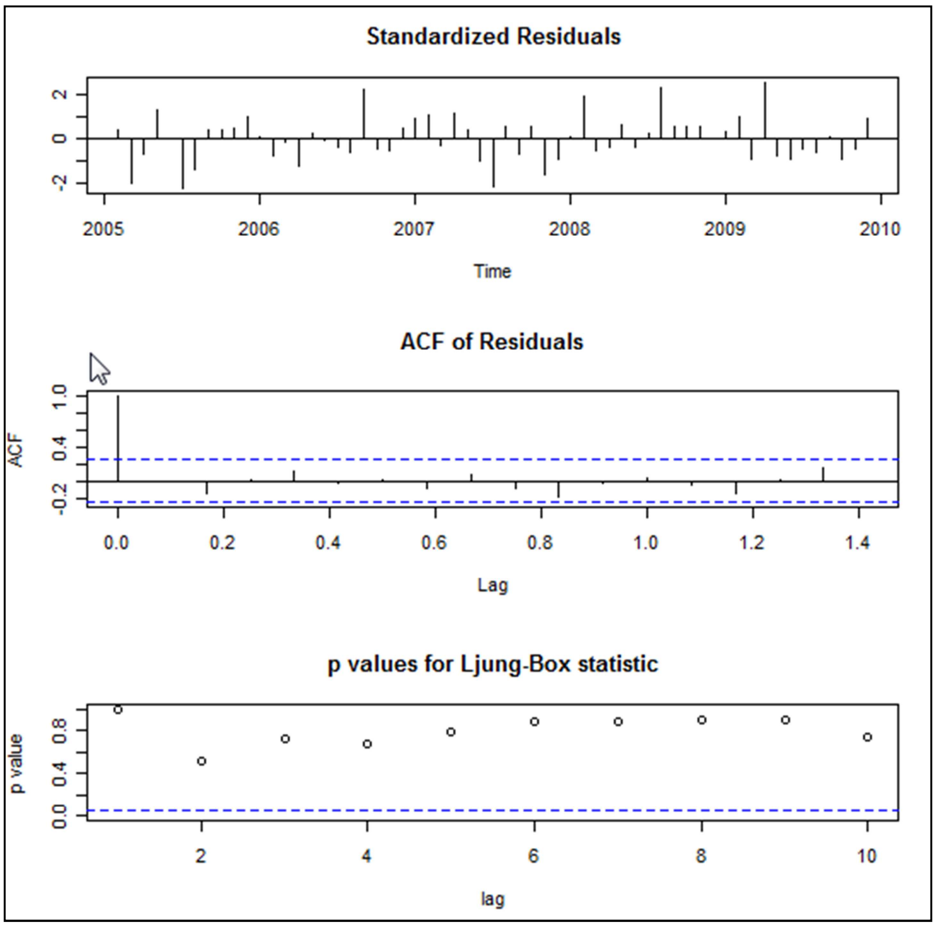
From figure 10, our transformed dataset looks stationary. The ACF still tend to indicate an MA(1) process.

The resulting process is:

```
> arima(steak2.log.d,order=c(0,0,1))

Call:
arima(x = steak2.log.d, order = c(0, 0, 1))

Coefficients:
         ma1  intercept
      0.0102     0.0001
s.e.  0.1542     0.0017

sigma^2 estimated as 0.000169:  log likelihood = 172.51,  aic = -339.03
```

Model diagnostic

Figure 11




```
> shapiro.test(MA1_steak2$resid)

        Shapiro-Wilk normality test

data:  MA1_steak2$resid
W = 0.9798, p-value = 0.4321
```

From Figure 11, we see that the residuals appear random and show no trend. The ACF looks like a random walk, no correlation, the Ljung-Box statistics have large p-values and the Shapiro-Wilk normality test produces a p-value higher than 5 %, which leads us to fail to reject the null hypothesis that the residuals are normally distributed.

We finally conclude that an ARIMA(0,1,1) on the logarithm of our data (starting 2005) is a reasonable fit.

Our final model would be :

$$\nabla(log(Y_t) = X_t$$

Where follows an MA(1) with $\theta = 0.0102$ and $\mu = 0.0001$.

To forecast formulas would be:

$$\hat{X}_t(1) = \mu + \theta e_t$$

$$\hat{X}_t(k) = \mu, \text{ for } k > 1$$

Appendix A – Raw data

n	Period	Price/kg	Change % from previous period	n	Period	Price/kg	Change % from previous period	n	Period	Price/kg	Change % from previous period
1	janv-95	10.17		61	janv-00	9.05	-0,88%	121	janv-05	12.43	2,90%
2	févr-95	10.29	1,18%	62	févr-00	9.36	3,43%	122	févr-05	12.49	0,48%
3	mars-95	10.12	-1,65%	63	mars-00	9.35	-0,11%	123	mars-05	12.17	-2,56%
4	avr-95	10.38	2,57%	64	avr-00	9.55	2,14%	124	avr-05	12.06	-0,90%
5	mai-95	10.00	-3,66%	65	mai-00	9.55	0,00%	125	mai-05	12.27	1,74%
6	juin-95	10.19	1,90%	66	juin-00	9.63	0,84%	126	juin-05	12.28	0,08%
7	juil-95	10.26	0,69%	67	juil-00	9.75	1,25%	127	juil-05	11.93	-2,85%
8	août-95	10.07	-1,85%	68	août-00	9.87	1,23%	128	août-05	11.72	-1,76%
9	sept-95	10.08	0,10%	69	sept-00	9.70	-1,72%	129	sept-05	11.78	0,51%
10	oct-95	9.92	-1,59%	70	oct-00	9.86	1,65%	130	oct-05	11.84	0,51%
11	nov-95	9.49	-4,33%	71	nov-00	9.79	-0,71%	131	nov-05	11.91	0,59%
12	déc-95	9.88	4,11%	72	déc-00	10.25	4,70%	132	déc-05	12.07	1,34%
13	janv-96	9.63	-2,53%	73	janv-01	10.41	1,56%	133	janv-06	12.09	0,17%
14	févr-96	9.96	3,43%	74	févr-01	11.13	6,92%	134	févr-06	11.97	-0,99%
15	mars-96	9.46	-5,02%	75	mars-01	11.47	3,05%	135	mars-06	11.95	-0,17%
16	avr-96	9.94	5,07%	76	avr-01	11.70	2,01%	136	avr-06	11.76	-1,59%
17	mai-96	9.65	-2,92%	77	mai-01	11.38	-2,74%	137	mai-06	11.79	0,26%
18	juin-96	9.65	0,00%	78	juin-01	11.34	-0,35%	138	juin-06	11.78	-0,08%
19	juil-96	9.32	-3,42%	79	juil-01	10.93	-3,62%	139	juil-06	11.73	-0,42%
20	août-96	9.37	0,54%	80	août-01	10.89	-0,37%	140	août-06	11.64	-0,77%
21	sept-96	9.32	-0,53%	81	sept-01	11.18	2,66%	141	sept-06	11.98	2,92%
22	oct-96	9.53	2,25%	82	oct-01	11.15	-0,27%	142	oct-06	11.91	-0,58%
23	nov-96	9.18	-3,67%	83	nov-01	11.00	-1,35%	143	nov-06	11.83	-0,67%
24	déc-96	9.66	5,23%	84	déc-01	11.47	4,27%	144	déc-06	11.90	0,59%
25	janv-97	9.70	0,41%	85	janv-02	11.13	-2,96%	145	janv-07	12.05	1,26%
26	févr-97	9.55	-1,55%	86	févr-02	11.26	1,17%	146	févr-07	12.22	1,41%
27	mars-97	9.23	-3,35%	87	mars-02	11.30	0,36%	147	mars-07	12.17	-0,41%
28	avr-97	9.43	2,17%	88	avr-02	11.36	0,53%	148	avr-07	12.35	1,48%
29	mai-97	9.17	-2,76%	89	mai-02	11.43	0,62%	149	mai-07	12.41	0,49%
30	juin-97	9.23	0,65%	90	juin-02	11.30	-1,14%	150	juin-07	12.25	-1,29%
31	juil-97	9.35	1,30%	91	juil-02	11.29	-0,09%	151	juil-07	11.92	-2,69%
32	août-97	8.92	-4,60%	92	août-02	11.08	-1,86%	152	août-07	12.00	0,67%
33	sept-97	9.38	5,16%	93	sept-02	10.85	-2,08%	153	sept-07	11.90	-0,83%
34	oct-97	9.50	1,28%	94	oct-02	11.01	1,47%	154	oct-07	11.99	0,76%
35	nov-97	9.27	-2,42%	95	nov-02	11.10	0,82%	155	nov-07	11.75	-2,00%
36	déc-97	9.88	6,58%	96	déc-02	11.39	2,61%	156	déc-07	11.61	-1,19%
37	janv-98	9.54	-3,44%	97	janv-03	11.43	0,35%	157	janv-08	11.62	0,09%
38	févr-98	9.64	1,05%	98	févr-03	12.09	5,77%	158	févr-08	11.91	2,50%
39	mars-98	9.30	-3,53%	99	mars-03	12.24	1,24%	159	mars-08	11.83	-0,67%
40	avr-98	9.22	-0,86%	100	avr-03	12.58	2,78%	160	avr-08	11.77	-0,51%
41	mai-98	9.19	-0,33%	101	mai-03	11.65	-7,39%	161	mai-08	11.87	0,85%
42	juin-98	8.84	-3,81%	102	juin-03	11.57	-0,69%	162	juin-08	11.82	-0,42%
43	juil-98	9.05	2,38%	103	juil-03	11.59	0,17%	163	juil-08	11.86	0,34%
44	août-98	9.21	1,77%	104	août-03	10.68	-7,85%	164	août-08	12.22	3,04%
45	sept-98	9.38	1,85%	105	sept-03	10.91	2,15%	165	sept-08	12.31	0,74%
46	oct-98	9.00	-4,05%	106	oct-03	11.45	4,95%	166	oct-08	12.40	0,73%
47	nov-98	9.48	5,33%	107	nov-03	12.25	6,99%	167	nov-08	12.49	0,73%
48	déc-98	9.23	-2,64%	108	déc-03	11.99	-2,12%	168	déc-08	12.49	0,00%
49	janv-99	9.00	-2,49%	109	janv-04	11.75	-2,00%	169	janv-09	12.54	0,40%
50	févr-99	9.36	4,00%	110	févr-04	12.48	6,21%	170	févr-09	12.70	1,28%
51	mars-99	9.17	-2,03%	111	mars-04	12.04	-3,53%	171	mars-09	12.56	-1,10%
52	avr-99	9.46	3,16%	112	avr-04	12.28	1,99%	172	avr-09	12.98	3,34%
53	mai-99	9.07	-4,12%	113	mai-04	12.27	-0,08%	173	mai-09	12.86	-0,92%
54	juin-99	9.19	1,32%	114	juin-04	12.49	1,79%	174	juin-09	12.71	-1,17%
55	juil-99	9.38	2,07%	115	juil-04	12.10	-3,12%	175	juil-09	12.64	-0,55%
56	août-99	9.30	-0,85%	116	août-04	12.47	3,06%	176	août-09	12.54	-0,79%
57	sept-99	9.34	0,43%	117	sept-04	12.26	-1,68%	177	sept-09	12.56	0,16%
58	oct-99	9.28	-0,64%	118	oct-04	12.26	0,00%	178	oct-09	12.42	-1,11%
59	nov-99	9.29	0,11%	119	nov-04	12.20	-0,49%	179	nov-09	12.35	-0,56%
60	déc-99	9.13	-1,72%	120	déc-04	12.08	-0,98%	180	déc-09	12.50	1,21%

Appendix B – R code

Figure 2

```
> steak=ts(read.table("E:/steak.txt",header=T)[,2],start=c(1995),frequency=12)
> plot.ts(steak)
```

Figure 3

```
> acf(steak,180/4)
> pacf(steak,180/4)
```

Figure 4

```
> steak.lm=lm(steak~t)
> summary(steak.lm)
```

Figure 5

```
> qqnorm(steak.lm$resid)
> qqline(steak.lm$resid)
```

Figure6

```
> plot(steak.lm$fitted,steak.lm$resid,xlab="Predicted",ylab="residuals",pch=20,ylim=c(-2,2))
```

Figure 7

```
> steak.d=diff(steak)
> plot.ts(steak.d)

> par(mfrow=c(2,1))
> acf(steak.d,180/4)
> pacf(steak.d,180/4)
```

Figure 9

```
> par(mfrow=c(3,1))
> plot.ts(steak.log.d)
> acf(steak.log.d,180)
> pacf(steak.log.d,180)
```

Figure 10

```
> par(mfrow=c(2,1))
> plot.ts(steak2)
> steak2.log.d=diff(log(steak2))
> plot.ts(steak2.log.d)
> par(mfrow=c(2,1))
> acf(steak2.log.d,60)
> pacf(steak2.log.d,60)
```

Figure 11

```
> tsdiag(MA1_steak2)
```