Time Series Student Project – Summer 2010

Roman Fedoseev

Diesel fuel prices in Luxembourg

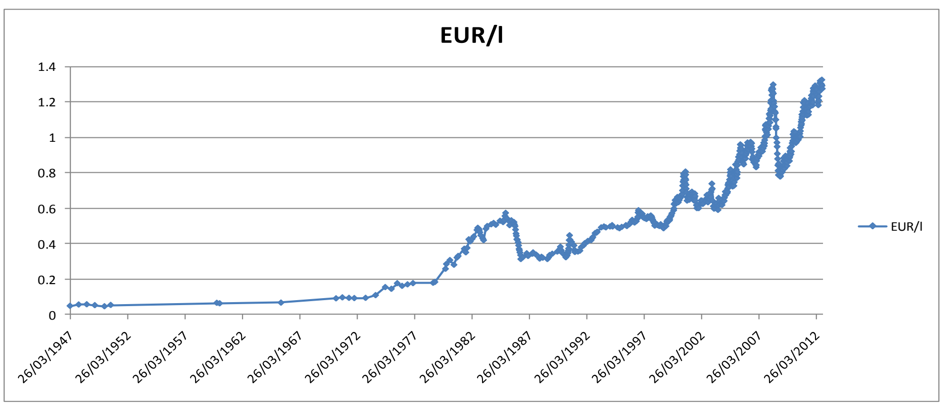
Introduction

The fuel cost and its variation with time receives significant consumer attention in the current distressed times. The fuel prices generally follow oil prices, which have considerably risen in the past years, and their volatility has increased.

This project focuses on diesel fuel prices in Luxembourg. The government sets the maxima prices at irregular intervals following variation of oil prices on the market, and generally the price would stay the same for several days or weeks, unlike the underlying oil price.

The full data in EUR per liter since 1947 is available at http://www.statistiques.public.lu/en/index.html

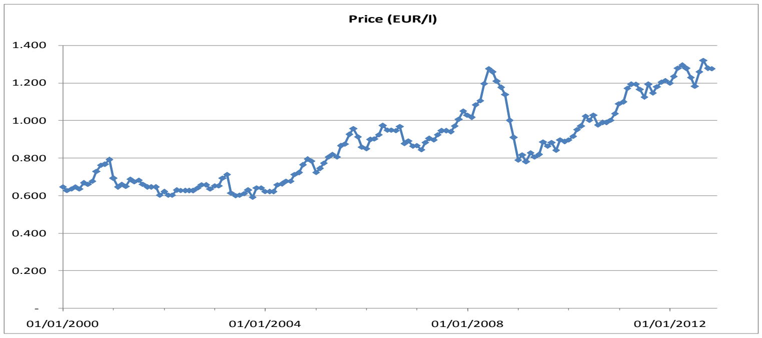
Chart 1 Historic diesel fuel prices in EUR/l



Analysis

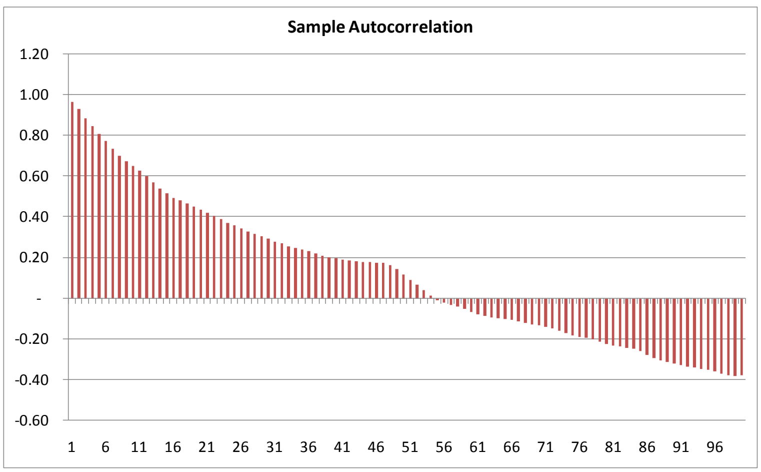
We selected the data at monthly intervals from1/1/2000 to 11/1/2012 (155 data points) to reflect the adoption of Euro in the late 90s and to avoid potential distortion caused by currency conversion from pre-Euro prices. Furthermore, this selection attempts to capture the most relevant recent experience.

Chart 2 Selected data range



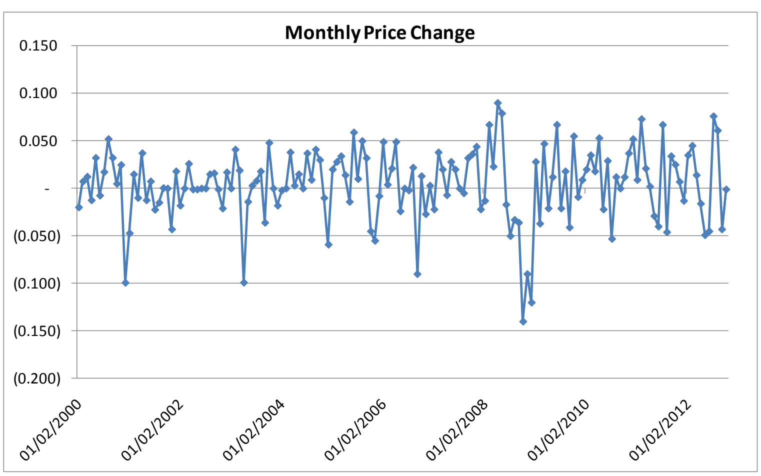
The time series does not appear to be stationary, and this is confirmed by the sample autocorrelation function (ACF) which does not tend to zero after 100 lags:

Chart 3 Sample ACF of original series



We seek to achieve stationarity through various transformations. First, we consider the First Difference of the original series: $\nabla Y_t = Y_t - Y_{t-1}$

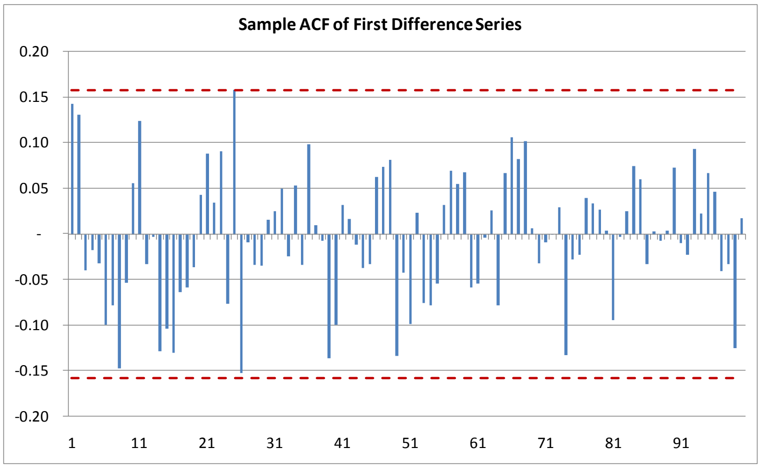
Chart 4 First Difference plot



The First Difference series looks much more stationary than the original series. One can also note some outliers, in particular around end 2008, when a continuous and significant decrease in prices was observed over several months.

The sample ACF appears to confirm stationarity of the First Difference series, as all the correlation coefficients are within the 95% confidence interval (1.96/√*n*, where n = 154) as shown in the following chart:

Chart 5 Sample ACF of First Difference



Although the First Difference of Prices appears to be a good candidate for ARIMA modeling, we consider alternative or additional transformations. One might as well take the natural logarithm of the price and *then* take the first difference. A log-price model would prevent the forecasted series from taking on negative values. Furthermore, a log-price model implies larger variations of prices at higher price levels, and such behavior can be seen in the chart of the full data.

The First Difference of Log Prices and its Sample ACF are as follows:

Chart 6 First Difference of Log Prices

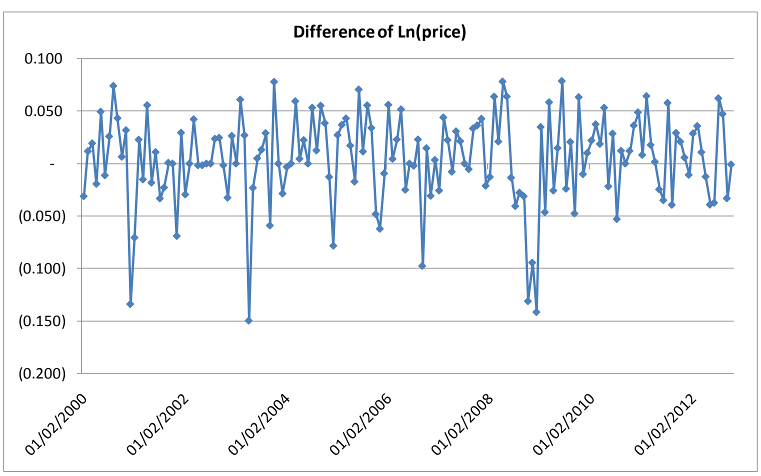
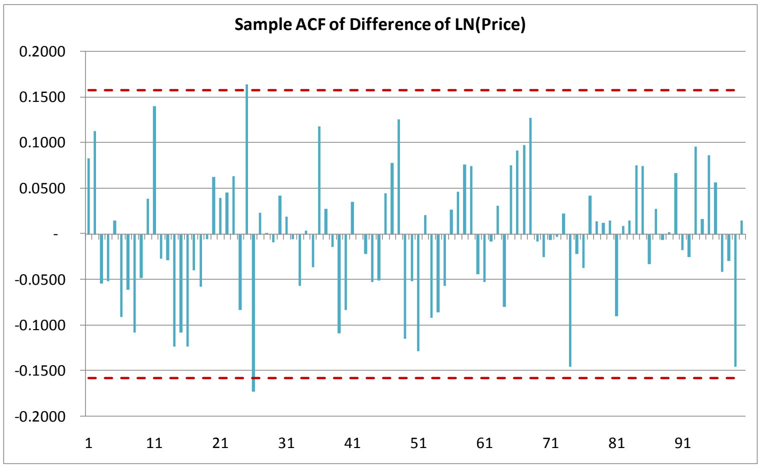


Chart 7 Sample ACF of First Difference of Log Prices



The Log Difference series appears to be slightly worse than the difference of the prices: the outliers from earlier periods with lower prices are more pronounced (although they become comparable with those in the end of 2008), and the correlation coefficients at lags 25 and 26 are slightly outside the 95% confidence interval.

At this stage, one could consider taking the Second difference of either the prices or their logarithms, in attempt to remove the downward skew of the First Difference model. The danger of this approach, however, is a potential overdifferencing of the model.

Indeed, let's have a look at the Second Difference of Log Prices and its Sample ACF:

Chart 8 Second Difference of Log Prices

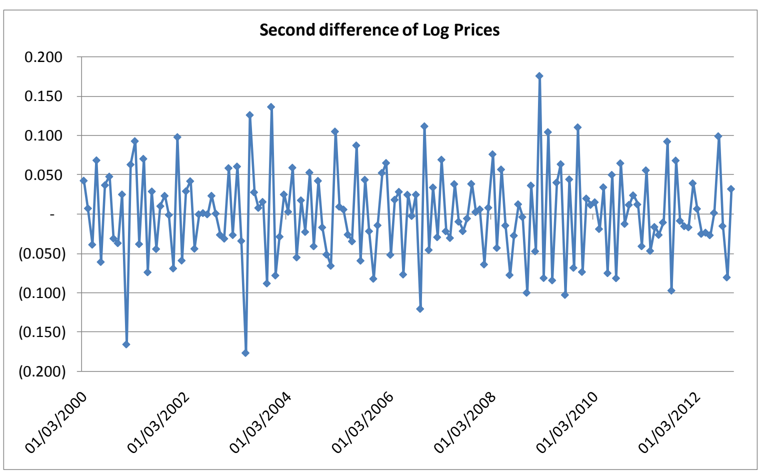
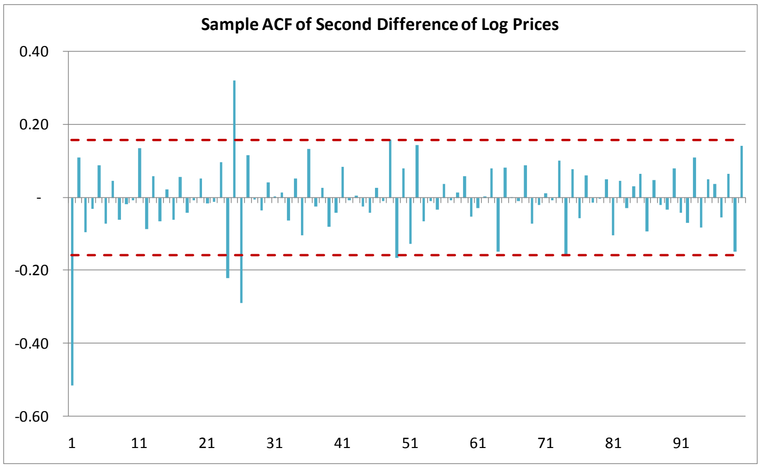


Chart 9 Sample ACF of Second Difference of Log Prices



One can see that the skew appears to have been removed from the series, and the sample ACF suggest that this series might be a good candidate for an MA(1) model. However, strong correlations at lags 24, 25 and 26 would need further treatment and would call for a more complex model.

Nevertheless, let's assume an MA(1) model and estimate the parameter θ. As ρ_1=-0.515, the method of moments fails to produce a real solution for θ. Using the Least Squares method, we calculate the sums of squares of the error terms and use Solver to find θ that minimizes that sum:

$$\left.\begin{aligned} e_1 &= Y_1 \\ e_2 &= Y_2 + \theta e_1 \\ e_3 &= Y_3 + \theta e_2 \\ &\vdots \\ e_n &= Y_n + \theta e_{n-1} \end{aligned}\right\}$$

We obtain θ=0.9405. The variance of the estimate of θ for an MA(1) model is approximately equal to $(1-\theta^2)/n$ = 0.000754, and the standard error of θ is 0.0275. At the 97% confidence level (i.e. at 2.17 x standard error), the upper limit of the confidence level reaches 1, which implies non-invertibility of the model and may indicate that we are dealing with overdifferencing.

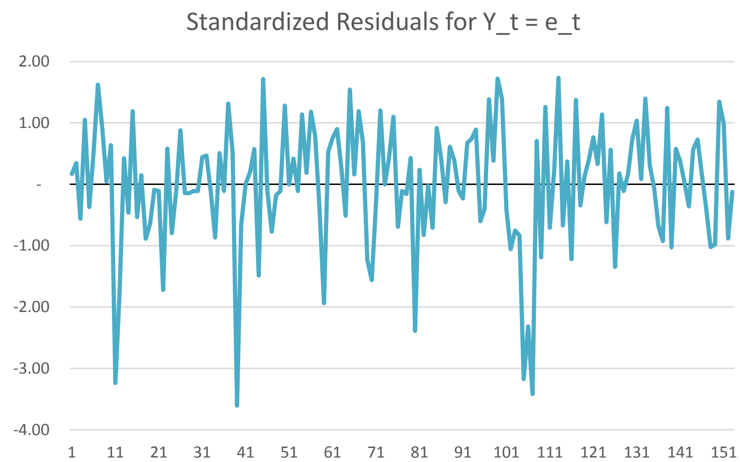
We therefore continue with the First Difference of Log Prices model, and consider an MA(1) model (i.e. an IMA(1,1) model for Log Prices) as suggested by its sample ACF in Chart 7. The lag 1 sample autocorrelation coefficient is within the 95% confidence interval, suggesting that θ may be equal to zero and that our model reduces to:

$$\ln(Y_t) - \ln(Y_{t-1}) = e_t$$

Indeed, as r_1=0.0828, the method of moments produces an invertible solution of θ=-0.0834, and with the Least Squares method we get θ=-0.0746. The standard error of θ is 0.0804, which means that zero is just one standard error away from either estimated value.

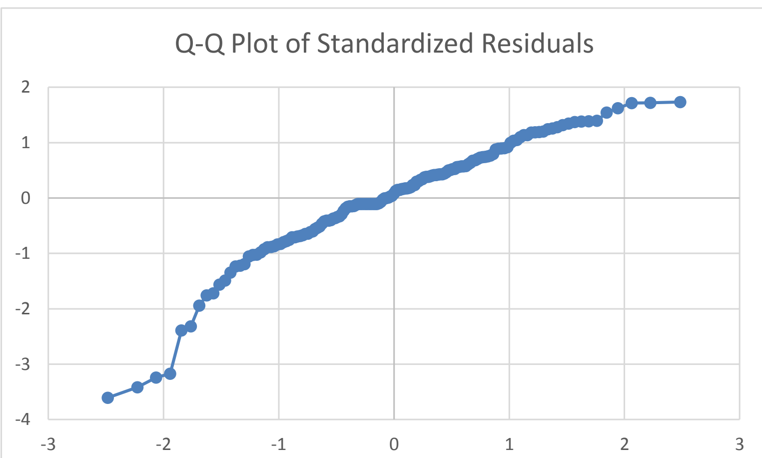
We proceed now to model diagnostics. As a first step, we analyze the residuals for the selected MA(1) model with θ=0. In this case, the residuals are equal to the series (First Difference of Log Prices) itself, since predicted value is simply the mean of the error term, i.e. zero. The standardized residuals are shown in Chart 10.

Chart 10 Standardized Residuals for MA(1)



As observed earlier, there are several outliers with values exceeding 3 standard errors. This departure from normality can also be seen in the quantile-quantile plot below.

Chart 11 Q-Q Plot of Standardized Residuals



We conclude from the diagnostics that the selected MA(1) model could be improved by addition of an outlier term. This project however does not include this analysis.

To conclude the project, we rewrite the selected model equation for the original series:

$$\ln(Y_t) - \ln(Y_{t-1}) = e_t$$
$$Y_t = Y_{t-1} \cdot \exp(e_t)$$

Generating random values for the error term with standard normal distribution, we obtain various projections of the series. The charts below show 4 examples of the projected series (in red) compared against the original one (in blue).

Chart 12 Examples of projected series