PhD in Business Administration and Quantitative Methods 2006/07
Time Series Course ~ Prof. Regina Kaiser
FINAL PROJECT

Sara Degli Esposti

DEMAND/SUPPLY: A LINEAR RELATION OVER TIME?

Insights from Australia

ABSTRACT

Since Demand and Supply are strictly related in economic life, it might be possible that they show a similar behavior over time. Therefore, to test this hypothesis I will analyze different series of goods productions and goods sales (as a proxy of demand). Then, I will test for the presence of cointegration among two different price index related to natural gas and its production. At the end, I will compare the production of grapes, the sales of wine and the price of wine, in order to see if it is possible to find a common pattern. Finally, it is important to notice that all series refer to Australia in order to put constant the institutional and macroeconomic environment.

1. INTRODUCTION

The basic aim of this paper is simply to check if it is possible to identify a common class of forecast models for time series referring to demands and supply of goods. Taking into account that firms try to produce the exact amount of commodities requested by the market, it is easy to see the implications of confirming the existence of a stationary predictable pattern for market demand. Of course, many firms already adopt this approach using their internal data and the market forecast, however which kind of model is the more parsimonious and exhaustive might depends on the type of goods sold. For this reason, in this paper, we will deal with three types of very different physical goods: a raw material (gas), a manufacture (cars), and a food (wine). It could be very interesting to find a common model for forecast of so different goods. Moreover, it might be interesting to observe what happens to the price generated by the interaction of supply and demand, according to the neoclassical model.

2. USING ARIMA FORECAST MODELS

The basic set of model that I am going to use assumes linearity in the path dependence of the series. I mean the ARIMA models, where ARIMA stands for "Auto-Regressive Integrated Moving Average." These type of model, coming from the previous ARMA model (Box and Jenkins), are a method for modelling univariate time series that can be used to extend estimates beyond the end of a series. Autoregressive models are simply a linear regression of the current value of the series against one or more prior values of the series. The value of p is called the order of the model. On the other hand, the primary idea behind the moving average model is that the random shocks are prorogued to future values of the series. So, lags of the differenced series appearing in the forecasting equation are called "auto-regressive" terms, lags of the forecast errors are called "moving average" terms, and a time series which needs to be differenced to be made stationary is said to be an "integrated" version of a stationary series. A non-seasonal ARIMA model is classified as an "ARIMA(p,d,q)" model, where: p is the number of autoregressive terms, d is the number of non-seasonal differences, and q is the number of lagged forecast errors in the prediction equation.

Using these forecast models is allowed only in the case of stationary time series, otherwise, the basic assumptions of the model are not satisfied. Therefore, since very few empirical series are stationary in their original form, we need to *stationarize* the series through the use of mathematical transformations. Let me remind that a stationary time series is one whose statistical properties such as mean, variance, autocorrelation, etc. are all constant over time.

2.1 How construct the right ARIMA model

If the series has a stable long-run trend and tends to revert to the trend line following a disturbance, it may be possible to stationarize it by transforming it into a series of period-to-period and/or season-to-season *differences*. If the mean, variance, and autocorrelations of the original series are not constant in time, perhaps the statistics of the *changes* in the series between periods or between seasons *will* be constant. Such a series is said to be *difference-stationary*. Sometimes it can be hard to tell the difference between a series that is trend-stationary and one that is difference-stationary, and a so-called *unit root test* may be used to get a more definitive answer.

The *first difference* of a time series is the series of changes from one period to the next. If Y_t denotes the value of the time series Y at period t, then the first difference of Y at period t is equal to $(Y_t - Y_{t-1})$. Normally, the correct amount of differencing is the lowest order of differencing that yields a time series which fluctuates around a well-defined mean value and whose autocorrelation function (ACF) plot decays fairly rapidly to zero, either from above or below. One of the most common errors in ARIMA modelling is to "over-difference" the series and end up adding extra AR or MA terms to undo the damage. If the lag-1 autocorrelation is more negative than -0.5 (and theoretically a negative lag-1 autocorrelation should *never* be greater than 0.5 in magnitude), this may mean the series has been over-differenced. The time series plot of an over-differenced series may look quite random at first glance, but if you look closer you will see a pattern of excessive *changes in sign* from one observation to the next (i.e., up-down-up-down, etc.):

Others types of mathematical transformations are allowed in order to get stationary the series and, then, to forecast theirs future values. For instance, the LOG function has the defining property that $\ln(x \times y) = \ln(x) + \ln(y)$, i.e., the logarithm of a product equals the sum of the logarithms. Therefore, logging tends to convert multiplicative relationships to additive relationships, and it tends to convert *exponential* (compound growth) trends to *linear* trends. By taking logarithms of variables which are multiplicatively related and/or growing exponentially over time, we can often explain their behaviour with linear models. Let highlight that the logarithm transformation can be applied only to data that are *strictly positive*, that is you can not take the log of zero or a negative number. Logging a series often has another effect: it dampens exponential growth patterns and reduces heteroscedasticity (i.e., stabilizes variance).

After a time series has been stationarized by differencing, the next step in fitting an ARIMA model is to determine whether AR or MA terms are needed to correct any autocorrelation that remains in the differenced series. By looking at the *autocorrelation function* (ACF) and *partial autocorrelation* (PACF) plots of the differenced series, you can tentatively identify the numbers of AR and/or MA terms that are needed. A partial *auto*correlation is the amount of correlation between a variable and a lag of itself that is not explained by correlations at all lower-order-lags. The autocorrelation of a time series Y at lag 1 is the coefficient of correlation between Y(t) and Y(t-1), which is presumably also the correlation between Y(t-1) and Y(t-2). But if Y(t) is correlated with Y(t-1), and Y(t-1) is equally correlated with Y(t-2), then we should also expect to find correlation between Y(t) and Y(t-2). (In fact, the amount of correlation we should expect at lag 2 is precisely the *square* of the lag-1 correlation.) Thus, the correlation at lag 1 "propagates" to lag 2 and presumably to higher-order lags. The *partial* autocorrelation at lag 2 is therefore the difference between the actual correlation at lag 2 and the expected correlation due to the propagation of correlation at lag 1.

If the PACF displays a sharp cutoff while the ACF decays more slowly (i.e., has significant spikes at higher lags), we say that the stationarized series displays an "AR signature," meaning that the autocorrelation pattern can be explained more easily by adding AR terms than by adding MA terms. In principle, any autocorrelation pattern can be removed from a stationarized series by adding enough autoregressive terms (lags of the stationarized series) to the forecasting equation, and the PACF tells you how many such terms are likely be needed. However, this is not always the simplest way to explain a given pattern of autocorrelation: sometimes it is more efficient to add MA terms (lags of the forecast errors) instead.

The autocorrelation function (ACF) plays the same role for MA terms that the PACF plays for AR terms, that is, the ACF tells you how many MA terms are likely to be needed to remove the remaining autocorrelation from the differenced series. If the autocorrelation is significant at lag k but not at any higher lags (i.e., if the ACF "cuts off" at lag k) this indicates that exactly k MA terms should be used in the forecasting equation. In the latter case, we say that the stationarized series displays an "MA signature," meaning that the autocorrelation pattern can be explained more easily by adding MA terms than by adding AR terms.

If a series is grossly either under or over-differenced (i.e., if a whole order of differencing needs to be added or cancelled), this is often signalled by a "unit root" in the estimated AR or MA coefficients of the model. An AR(1) model is said to have a unit root if the estimated AR(1) coefficient is almost exactly equal to 1, that is not significantly different from 1, in terms of the coefficient's own standard error. When this happens, it means that the AR(1) term is precisely mimicking a first difference, in which case you should remove the AR(1) term and add an order of differencing instead. This is exactly what would happen if you fitted an AR(1) model to the undifferenced UNITS series, as noted earlier. In a higher-order AR model, a unit root exists in the AR part of the model if the sum of the AR coefficients is exactly equal to 1. In this case you should reduce the order of the AR term by 1 and add an order of differencing. A time series with a unit root in the AR coefficients is nonstationary (i.e., it needs a higher order of differencing). If there is a unit root in the AR part of the model (i.e., if the sum of the AR coefficients is almost exactly 1) we should reduce the number of AR terms by one and increase the order of differencing by one. Similarly, an MA(1) model is said to have a unit root if the estimated MA(1) coefficient is exactly equal to 1. When this happens, it means that the MA(1) term is exactly cancelling a first difference, in which case, you should remove the MA(1) term and also reduce the order of differencing by one. In a higher-order MA model, a unit root exists if the sum of the MA coefficients is exactly equal to 1. If there is a unit root in the MA part of the model (i.e., if the sum of the MA coefficients is almost exactly 1) you should reduce the number of MA terms by one and reduce the order of differencing by one.
A forecasting model with a unit root in the estimated MA coefficients is said to be *noninvertible*, meaning that the residuals of the model cannot be considered as estimates of the *true* random noise that generated the time series. Another symptom of a unit root is that the forecasts of the model may *blow up* or otherwise behave bizarrely. If the time series plot of the longer-term forecasts of the model looks strange, you should check the estimated coefficients of your model for the presence of a unit root. If the long-term forecasts appear erratic or unstable, there may be a unit root in the AR or MA coefficients.

Here, I will try to find out the best model in order to fit my data. In most cases, the best model turns out a model that uses either only AR terms or only MA terms, although in some cases a "mixed" model with both AR and MA terms may provide the best fit to the data. However, it is possible for an AR term and an MA term to *cancel each other's effects*, even though both may appear significant in the model (as judged by the t-statistics of their coefficients). For this reason, ARIMA models *cannot* be identified by "backward stepwise" approach that includes both AR and MA terms. In other words, I will not begin by including several terms of each kind and then throwing out the ones whose estimated coefficients are not significant. Instead, I am going to follow a "forward stepwise" approach, adding terms of one kind or the other as indicated by the appearance of the ACF and PACF plots.

According to the frequency of our data, it might be possible to discover pattern also in seasonality. A seasonal ARIMA model is classified as an ARIMA(p,d,q)x(P,D,Q) model, where P=number of seasonal autoregressive (SAR) terms, D=number of seasonal differences, Q=number of seasonal moving average (SMA) terms. In identifying a seasonal model, the *first* step is to determine whether or not a *seasonal difference* is needed, in addition to or perhaps instead of a non-seasonal difference. I shall look at time series plots and ACF and PACF plots for all possible combinations of 0 or 1 non-seasonal difference and 0 or 1 seasonal difference. If the seasonal pattern is both *strong* and *stable* over time (e.g., high in the Summer and low in the Winter, or vice versa), then I often will prefer to use a seasonal difference regardless of whether I have used a non-seasonal difference, since this will prevent the seasonal pattern from "dying out" in the long-term forecasts. For example, a pure SAR(1) process should have spikes in the ACF at lags s, 2s, 3s, etc., while the PACF cuts off after lag s. Conversely, a pure SMA(1) process should show spikes in the PACF at lags s, 2s, 3s, etc., while the ACF cuts off after lag s. Accordingly, an SAR signature usually occurs when the autocorrelation at the seasonal period is *positive*, whereas an SMA signature usually occurs when the seasonal autocorrelation is *negative*.

Basically, the two kind of ARIMA models that I will apply are the simple exponential smoothing in the ordinal and in the seasonal part. The simple exponential smoothing model (i.e. ARIMA(0,1,1)) uses an *exponentially weighted moving average* of past values in order to filter out the noise and more accurately estimate the local mean. In other words, rather than taking the most recent observation as the forecast of the next observation, it is better to use an *average* of the last few observations. The model can be written as $X_t = X_{t-1} - q a_{t-1}$, where a_{t-1}) denotes the error at period t-1. Note that this resembles the prediction equation for the ARIMA(1,1,0) model, except that instead of a multiple of the lagged difference it includes *a multiple of the lagged forecast error*.
On the other hand, the ARIMA(0,1,1)x(0,1,1) model is basically a Seasonal Random Trend (SRT) model fine-tuned by the addition of MA(1) and SMA(1) terms to correct for the mild overdifferencing that resulted from taking two total orders of differencing. This is probably the most commonly used seasonal ARIMA model, as confirmed also in this analysis.

2.2 Using cointegration techniques

For a long time it was common practice to estimate equations involving nonstationary variables in macroeconomic models by straightforward linear regression. It was not well understood that testing hypotheses about the coefficients using standard statistical inference might lead to completely spurious results. In an influential paper, Clive Granger and his associate Paul Newbold (Granger and Newbold (1974)[1]) pointed out that tests of such a regression may often suggest a statistically significant relationship between variables where none in fact exists. However, if economic relationships are specified in first differences instead of levels, the statistical difficulties due to nonstationary variables can be avoided because the differenced variables are usually stationary even if the original variables are not. An alternative approach would involve removing a linear time trend from the variables and specifying the empirical relationship between them using detrended variables. Removing (separate) time trends assumes, however, that the variables follow separate deterministic trends, which does not appear realistic, given the awkward long-run implications. Dynamic econometric models based on linearly detrended variables may, thus, be able to

[1] C.W.J. GRANGER and P. NEWBOLD , “SPURIOUS REGRESSIONS IN ECONOMETRICS.” Journal of Econometrics 2 (1974) 111-120. (6 North-Holland Publishing Company).

characterize short-term dynamics of economic variables but not their long-run relationships. The same is true for models based solely on first differences.
Therefore, if we consider the following bivariate autoregressive system of order p :

$$x_t = \sum_{j=1}^{p} g_{1j} x_{t-j} + \sum_{j=1}^{p} d_{1j} y_{t-j} + e_{1t} \qquad y_t = \sum_{j=1}^{p} g_{2j} x_{t-j} + \sum_{j=1}^{p} d_{2j} y_{t-j} + e_{2t}$$

where x_t and y_t are I(1) and cointegrated, and e_{1t} and e_{2t} are white noise; the Granger representation theorem says that in this case, the system can be written as:

$$\Delta x_t = a_1 (y_{t-1} - b x_{t-1}) + \sum_{j=1}^{p-1} g^*_{1j} \Delta x_{t-j} + \sum_{j=1}^{p-1} d^*_{1j} \Delta y_{t-j} + e_{1t} \quad \Delta y_t = a_2 (y_{t-1} - b x_{t-1}) + \sum_{j=1}^{p-1} g^*_{2j} \Delta x_{t-j} + \sum_{j=1}^{p-1} d^*_{1j} \Delta y_{t-j} + e_{2t}$$

where at least one of parameters a_1 and a_2 deviates from zero. Both equations of the system are "balanced", that is, their left-hand and right-hand sides are of the same order of integration, since $y_{t-1} - b x_{t-1} \sim I(0)$. Suppose that $y_t - b x_t = 0$ defines a dynamic equilibrium relationship between the two economic variables, y and x. Then $y_t - b x_t$ is an indicator of the degree of disequilibrium. The coefficients a_1 and a_2 represent the strength of the disequilibrium correction, and the system is now said to be in error-correction form. A system characterized by these two equations is thus in disequilibrium at any given time, but has a built-in tendency to adjust itself towards the equilibrium.

Engle and Granger (1987) consider the problem of testing the null hypothesis of no cointegration between a set of I(1) variables. They estimate the coefficients of a static relationship between these variables by ordinary least squares and apply well-known unit root tests to the residuals. Rejecting the null hypothesis of a unit root is evidence in favour of cointegration. The performance of a number of such tests is compared in the paper. In other words, the nonstationary time series in Y_t are cointegrated if there is a linear combination of them that is stationary or I(0).

Cointegration naturally arises in economics and finance. In economics, cointegration is most often associated with economic theories that imply equilibrium relationships between time series variables. For instance: (a) the permanent income model implies cointegration between consumption and income, with consumption being the common trend; (b) money demand models imply cointegration between money, income, prices and interest rates; (c) growth theory models imply cointegration between income, consumption and investment, with productivity being the common trend; (d) purchasing power parity implies cointegration between the nominal exchange rate and foreign and domestic prices; (e) covered interest rate parity implies cointegration between forward and spot exchange rates; (f) the Fisher equation implies cointegration between nominal interest rates and inflation; (g) the expectations hypothesis of the term structure implies cointegration between nominal interest rates at different maturities.
The equilibrium relationships implied by these economic theories are referred to as long-run equilibrium relationships, because the economic forces that act in response to deviations from equilibrium may take a long time to restore equilibrium. As a result, cointegration is modeled using long spans of low frequency time series data measured monthly, quarterly or annually.

In this study, I will use cointegration to test the relationships between production and price of goods.

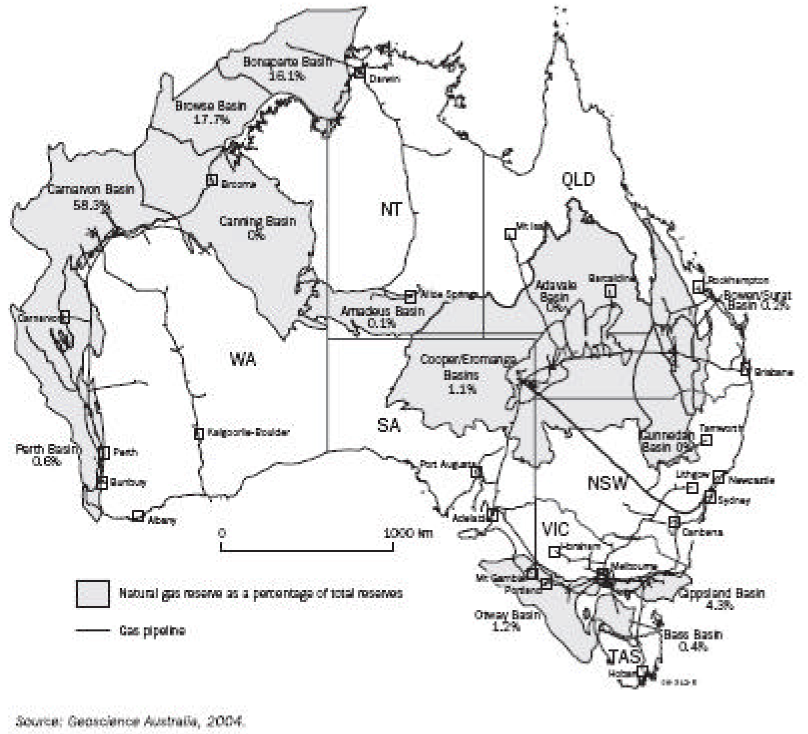
3. DATA AND ANALYSIS

I deal with three types of different goods produced in Australia: natural gas, cars, and wine. The large distance among these categories of goods ought help us to test the validity of an ARIMA (0,1,1) as forecast model using economic variables.

3.1 GAS

Australia has large identified resources of fossil fuels and uranium. It is ranked in the top six countries in the world for economic demonstrated resources (EDR) of black and brown coal, and has the world's largest EDR of uranium. Australia also has significant reserves of natural gas and crude oil.

Even tough, energy sources might be both renewable (energy sources for which the supply is essentially inexhaustible, i.e. solar, wind, hydro-electricity, geothermal and biomass) and non-renewable (energy sources with a finite supply), however, most of Australia's energy comes from non-renewable sources, which include the fossil fuels of oil, natural gas and coal. The Map on the right, shows the extent of access to gas resources and major transmission pipelines in Australia (2004).

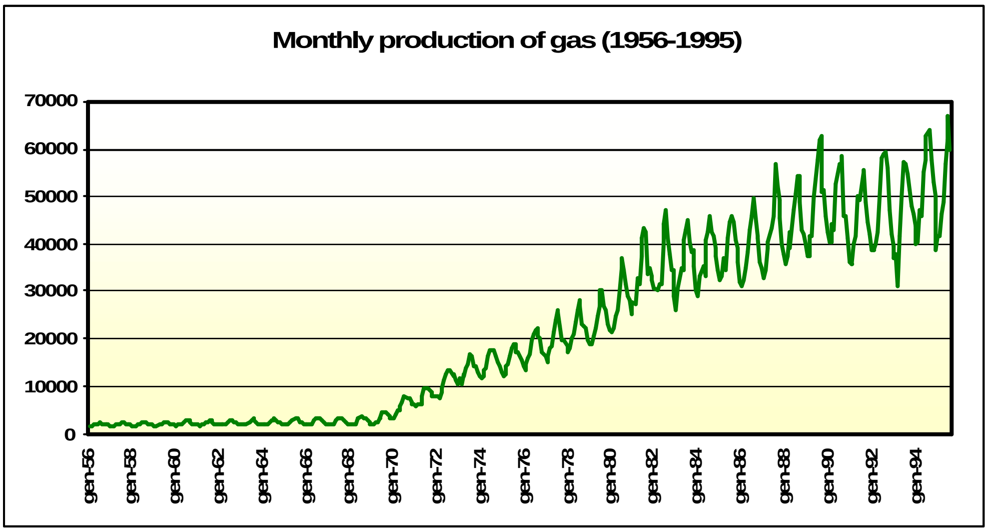


Geoscience Australia, last viewed August 2006 <http://www.ga.gov.au>

In 2004–05 Australia's total primary energy production was estimated at 17,524 PJ of which black coal accounted for nearly half (46%), followed by uranium (30%), natural gas (9%) and crude oil (6%). Renewable energy production (including wood, bagasse, biofuel, hydro-electricity and solar thermal energy) accounted for only 2% (261 PJ) of total production in 2004–05. Let point out that in 2004–05 Australia's total domestic energy use was 5,841 PJ, less than one third of the total energy it produced (17,524 PJ)

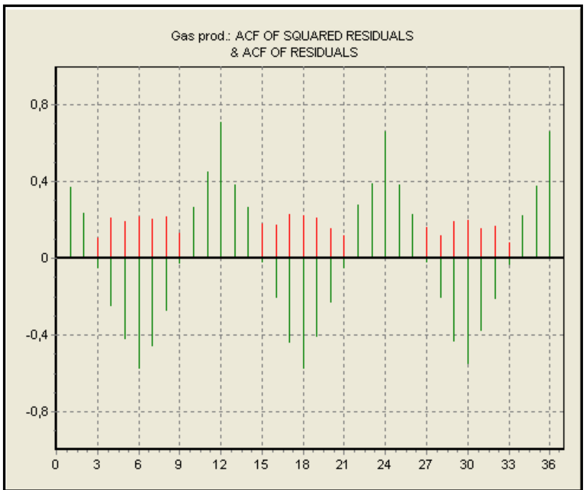
I am going to analyze a series about gas production with three different frequency: monthly, quarterly and annual. As we will see, all these series show the same behavior over time, and, so, they might be forecasted using the same ARIMA model.
Then, I will deal with series about the price of these produced gas and the price of the gas imported in the country, in order to test if these series might be forecast using the same ARIMA model and if there is a common tendency between production, as a proxy of supply, and price.

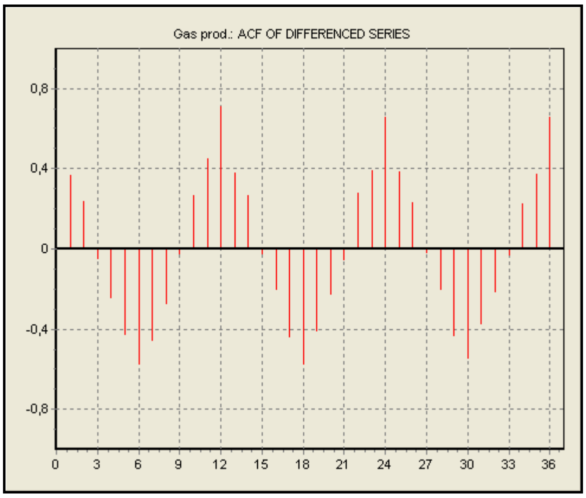
Australia monthly production of gas (million megajoules)

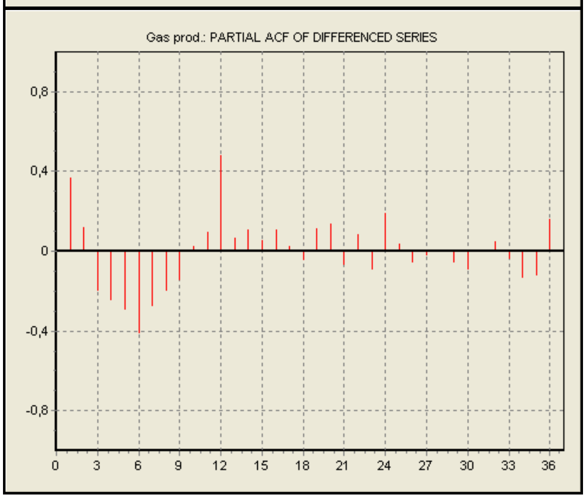


Source: ABS 8301.0
(Jan 1956 - Aug 1995)

After taking the first difference of the logging series, I have observed the following pattern in the ACF and in the PACF. Of course, something is happening in the seasonal part, since the sine-cosine behaviour is clearly showed with decreasing picks in 12, 24 etc. That is obvious the need of differentiating in the seasonal part.
The ACF, in which the red lines represent the squared residuals, on the right, represents residuals before taking the first difference in the ordinal part.







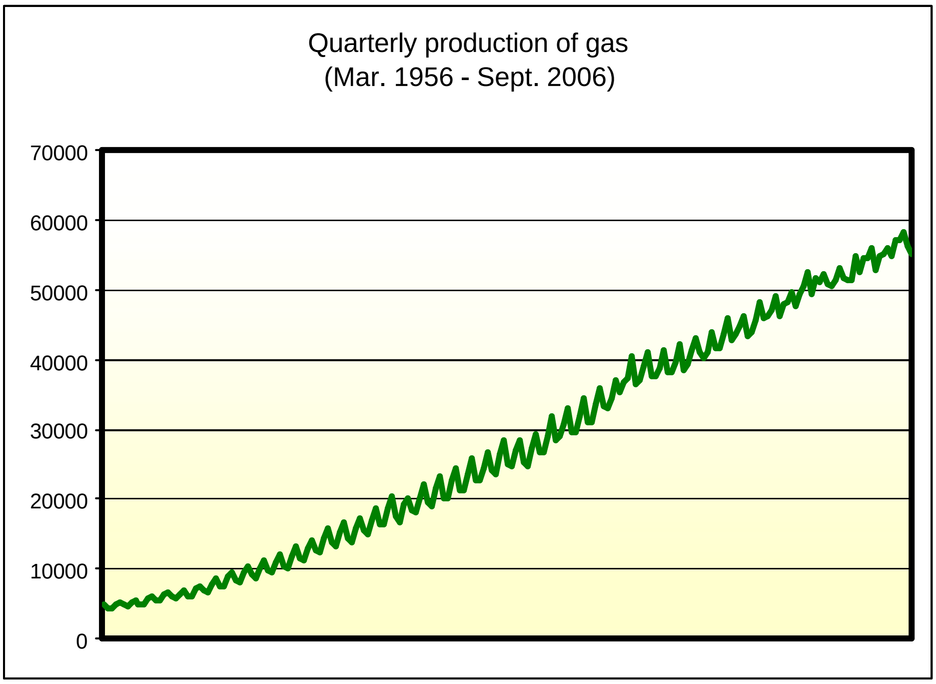
The model finally chosen is an ARIMA (0,1,1) (0,1,1), in which log transformation is needed, also, trading day correction. Moreover, it is required to apply an outliers correction in order to avoid behaviours that seem to be not linear. As showed later, all the outliers are significant, and, also, there are level shifts, which need to be treated as additive outliers.

```
OUTLIERS                      ST. ERROR     T VALUE
 447 AO    ( 3 1993) (      0.00045)    4.48
 170 LS    ( 2 1970) (      0.03224)   -9.31
 149 TC    ( 5 1968) (      0.03528)    7.21
 468 AO    (12 1994) (      0.03530)    4.57
 163 LS    ( 7 1969) (      0.03348)   -4.60
 309 AO    ( 9 1981) (      0.03528)    4.17
 301 LS    ( 1 1981) (      0.03160)   -3.79
 205 AO    ( 1 1973) (      0.03526)    3.76
 192 LS    (12 1971) (      0.03166)    3.83

 METHOD OF ESTIMATION: EXACT MAXIMUM LIKELIHOOD
 PARAMETER      ESTIMATE          STD ERROR      T RATIO    LAG
 MA1   1        -.37608          0.44189E-01      -8.51      1
 MA2   1        -.77810          0.32668E-01     -23.82     12
```

The model appear to fit very well with the data without having problems about normality. Due to outliers and the large difference between the first part of the series and the subsequent tendency, there are still some trouble in the squared residuals, as pointed out in the Ljung-Box test.

```
AIC  -1635.4877   and   BIC  -6.2874
DURBIN-WATSON=  1.8909
LJUNG-BOX Q VALUE OF ORDER 24 IS  35.88 < 36.415
(Squared)  LJUNG-BOX Q VALUE OF ORDER 24 IS  44.99 > 36.415
```

The model finally chosen, an ARIMA (0,1,1)x(0,1,1), is also confirmed when applied at the same series, but with different unit measurement, as it will be shown later.



When used in conjunction with differencing, logging converts absolute differences into relative (i.e., percentage) differences. Thus, the series DIFF(LOG(Y)) represents the *percentage change* in Y from period to period. Strictly speaking, the percentage change in Y at period t is defined as (Y(t)-Y(t-1))/Y(t-1), which is only *approximately* equal to LOG(Y(t)) - LOG(Y(t-1)), but the approximation is almost exact if the percentage change is small, as happened in this case.

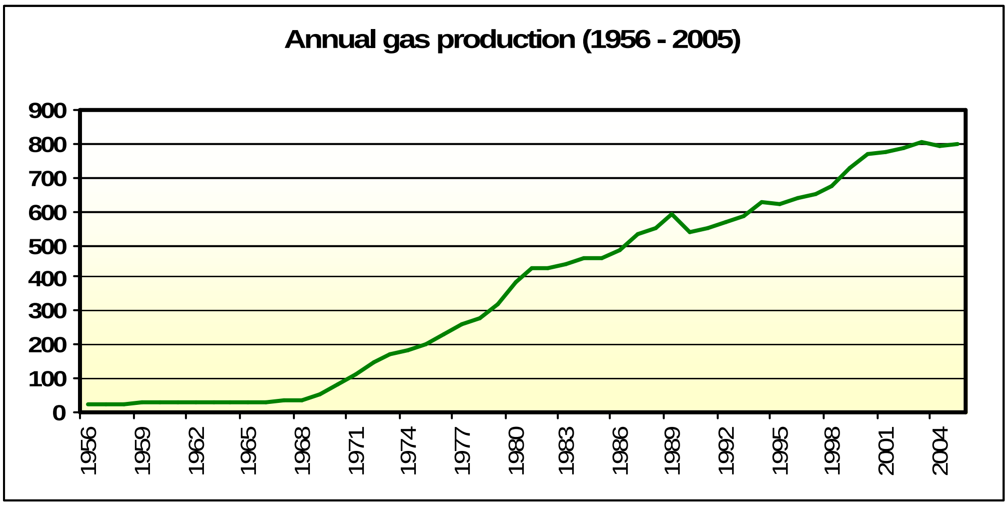
The model, which was finally chosen is the ARIMA (0,1,1)(0,1,1), in which LOGS are Selected and nothing more. Since there is just on outlier (131 AO [3 1988] with a t-value of 3.76) the problems about non-linearity saw in the previous model, are solved.

```
PARAMETER      ESTIMATE          STD ERROR      T RATIO    LAG
MA1   1        -.33646          0.68486E-01      -4.91      1
```

```
MA2  1          -.58153          0.57901E-01      -10.04      4
AIC   -1040.8677 and BIC      -8.0540
DURBIN-WATSON=  1.8554
```

The Ljung-Box test used here is based on the autocorrelation plot, but, instead of testing randomness at each distinct lag, it tests the "overall" randomness based on a number of lags.

```
LJUNG-BOX Q VALUE OF ORDER 16 IS  19.81 < 26.296
Squared LJUNG-BOX Q VALUE OF ORDER 16 IS  19.68 < 26.296
```



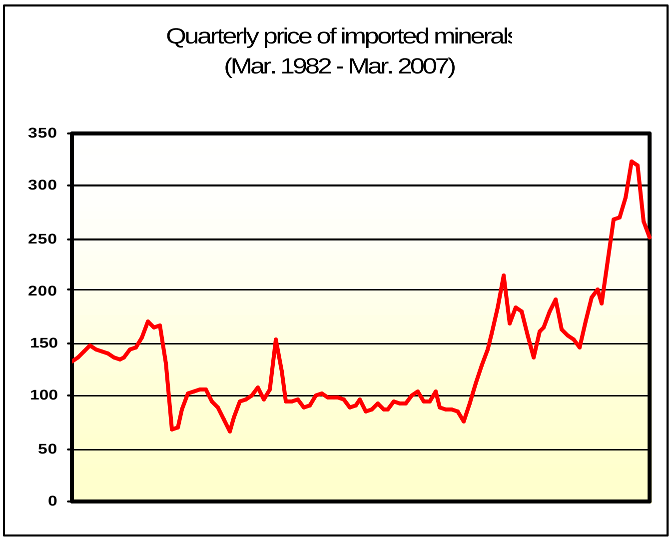
MODEL FINALLY CHOSEN: (0,1,1)(0,1,1)

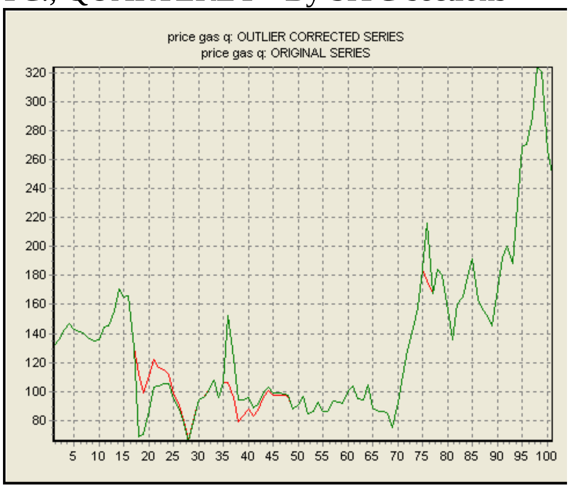
```
In which LOGS
are Selected
WITH MEAN,
WITHOUT
TRADING DAY
CORRECTION,
WITHOUT
EASTER
CORRECTION
NO OUTLIERS
DETECTED
```

```
PARAMETER      ESTIMATE         STD ERROR       T RATIO     LAG
 MA1   1       0.78143           0.10259         7.62        1
 MA2   1      -.95004          0.51320E-01      -18.51      12

AIC  -59.4871   and   BIC  -4.7485
DURBIN-WATSON=  1.4462
LJUNG-BOX Q VALUE OF ORDER 24 IS    36.96
(Squared)    LJUNG-BOX Q VALUE OF ORDER 24 IS    45.01
```

Now, I want to check if the behaviour of the series about the price at which gas has been sold and the price of the imported same raw material, may be forecast using the same ARIMA model.

IMPORT PRICE INDEXES OF MINERALS FUELS, ETC., QUARTERLY - By SITC sections





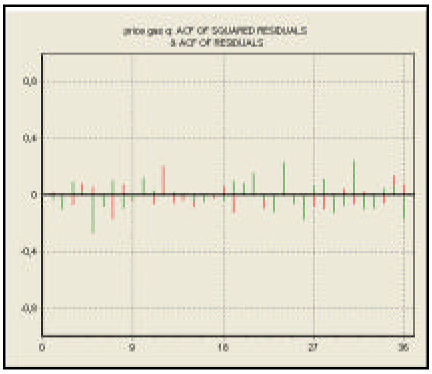
The red line shows the outliers corrected series.

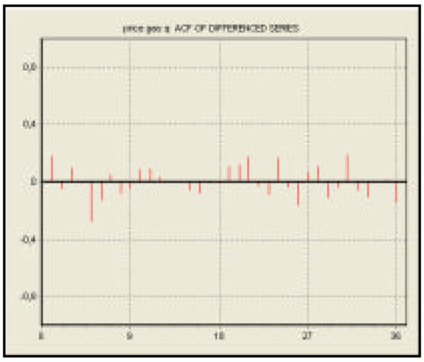
Source: Australian Bureau of Statistics (Jul 2007 Australian Economic Indicators)

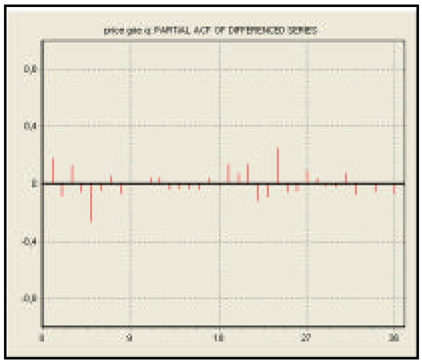
The model finally chosen is an ARIMA(0,1,1)(0,0,0), so without seasonal part. Again, the logs transformation is needed, but without anything else. Just three significant outliers are detected:

```
PARAMETER            VALUE          ST. ERROR     T VALUE
 OUT 1 ( 18)         -.49389        (     0.08023)   -6.16   TC     ( 2 1986)
 OUT 2 ( 36)         0.36072        (     0.08023)    4.50   TC     ( 4 1990)
 OUT 3 ( 76)         0.20328        (     0.05581)    3.64   AO     ( 4 2000)

METHOD OF ESTIMATION: EXACT MAXIMUM LIKELIHOOD
PARAMETER       ESTIMATE           STD ERROR       T RATIO    LAG
MA1   1         0.30754          0.95154E-01         3.23       1

AIC     -183.2731
BIC       -4.5674
DURBIN-WATSON=  2.0598
LJUNG-BOX Q VALUE OF ORDER 16 IS    14.02 < 26.296
(Squared) LJUNG-BOX Q VALUE OF ORDER 16 IS    13.59 < 26.296
```

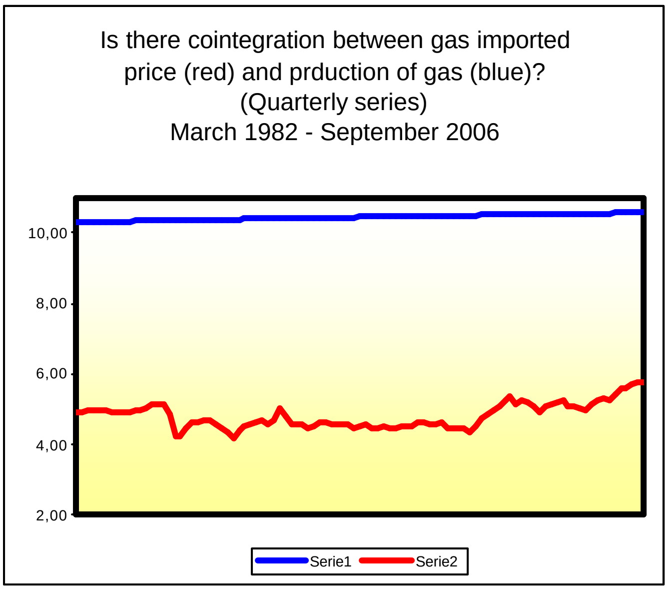
As we can see, this model fits very well our data, without non-linearity problems.







Finally, we want to test if there is a relations between the series about quarterly gas production and the price of imported gas. Actually, we should expected an inverse relation between the two. So, I suggest that we should not find cointegration between these series. As showed in the table below, the hypothesis about the absence of cointegration between the two logged series is confirmed.



Before looking at the cointegration test results, let us make a summary about how we should interpret cointegration and what we mean for it.

The starting point is the definition of degree of integration among series. We called a series integrated of order *p*, if $D^p(X_t) = e_t$, that is the series obtained after p differences is a stationary one. Considering these set of variables such that $b_1x_{1t} + b_2x_{2t} + ... + b_nx_nt = 0$, $b = (b_1, b_2, ..., b_n)$, and $X_t = (x_{1t}, x_{2t}, ..., x_{nt})'$. If the condition above stated is satisfied, we can say that $X_t = (x_{1t}, x_{2t}, ..., x_{nt})'$ is in a long run equilibrium.

Otherwise, changes in the equilibrium condition are called *equilibrium error*, which is defined as $e_t = bX_t$. When this equilibrium error appears to be stationary, we can say that the set of variables $X_t = (x_{1t}, x_{2t}, ..., x_{nt})'$ is in equilibrium. Therefore, for cointegration we mean a linear combination of non-stationary variables. We will focus our attention on the case in which each variable is integrated of order 1, that is I(1), such that its first difference is of type I(0), which is stationary.
Here, I will use Johansen methodology in order to test if there is cointegration among the variables under study. Let us consider an autoregressive vector of order p, such that $Y_t = A_1 Y_{t-1} + ... + A_p Y_{t-p} + BX_t + e_t$, where Y_t is a k-dimensional vector of non-stationary variables of type I(1), X_t is a vector of deterministic variables, and e_t is a innovations vector. We can rewrite the autoregressive vector as $\Delta Y_t = pY_{t-1} + \sum_{i:1}^{p-1} \Gamma_i \Delta Y_{t-i} + bX_t + e_t$, where $p = \sum_{i=1}^{p} A_i - I$, $\Gamma_i = -\sum_{j:i+1}^{p} A_j$, and I is the identity matrix. According to Granger's theorem, if the coefficient matrix π has got rank r < k, then there are two matrices (of k x r dimensions) α and β, each of them of rank r, such that $p = ab'$, with stationary $b'Y_t$. Here, r represents the cointegrating rank, where each column of β is a cointegrating vector, whilst the elements presents in α are known as adjusting parameters. The Johansen's method estimates matrix π in order to verify if we can reject the null hypothesis due to the restricted rank of π. To verify the rank of the matrix π is applied a c^2 test, before whom we should specify some elements of the model, as the presence of intercept and/or trend in the cointegration space. We can chose among the following 5 possibilities applying Johnson's methodology using EViews:

1. neither deterministic trends nor intercept in the series under study, that is in Y_t (VAR).
   $H_2(r) : pY_{t-1} + BX_t = ab'Y_{t-1}$;
2. no deterministic trends in Y_t, but intercepts in the cointegrating equations (EMC)
   $H_1^*(r) : pY_{t-1} + BX_t = a(b'Y_{t-1} + r_o)$;
3. Y_t has linear trends and the ECM has intercept $H_1(r) : pY_{t-1} + BX_t = a(b'Y_{t-1}) + r_o + a_\perp g_0$ where a⊥ is non-singular matrix of dimension *k* x (*k* - *r*) such that a'a = 0 and the rank of a | a⊥ is equal to *k*;
4. both Y_t and ECM have liner trends $H^*(r) : pY_{t-1} + BX_t = a(b'Y_{t-1} + r_o + r_1 t) + a_\perp g_0$;
5. Y_t has a quadratic trend, whilst ECM has a linear trend such that
   $H(r) : pY_{t-1} + BX_t = a(b'Y_{t-1} + r_o + r_1 t) + a_\perp (g_0 + g_1 t)$.

These five cases are link each other according to the following relationship
$H_2(r) \underset{c_r^2}{\subset} H_1^*(r) \underset{c_{k-r}^2}{\subset} H_1(r) \underset{c_r^2}{\subset} H^*(r) \underset{c_{k-r}^2}{\subset} H(r)$, from which we can identify the c^2 distribution needed. Thus, the first step is to choose the series under study, that is Y_t, then we can construct the test statistics in order to test the eventual presence of cointegration. We will use the following statistics, taking into account each of the cases selected.

$$\begin{cases} \text{CASES 3 and 5} & n\sum_{i:r+1}^{k}\log((1-l_i)/(1-l_i^*)) \approx c_{k-r}^2 \\ \text{CASES 1, 2, and 4} & n\sum_{i=1}^{r}\log((1-l_i)/(1-l_i^*)) \approx c_r^2 \end{cases}$$

where l_i and l_i^* represent the biggest self-values for models H(r) and $H^*(r)$ respectively.

Johansen Cointegration Test

Sample: 1982:1 2006:4
Included observations: 96
Test assumption: No deterministic trend in the data
Series: GASPRICEQ GASQLN
Lags interval: 1 to 2

Eigenvalue	Likelihood Ratio	5 Percent Critical Value	1 Percent Critical Value	Hypothesized No. of CE(s)
0.054337	5.377231	12.53	16.31	None
0.000144	0.013829	3.84	6.51	At most 1

*(**) denotes rejection of the hypothesis at 5%(1%) significance level
L.R. rejects any cointegration at 5% significance level

Unnormalized Cointegrating Coefficients:

GASPRICEQ	GASQLN
-0.255307	0.111260
0.060855	-0.015977

Normalized Cointegrating Coefficients: 1 Cointegrating Equation(s)

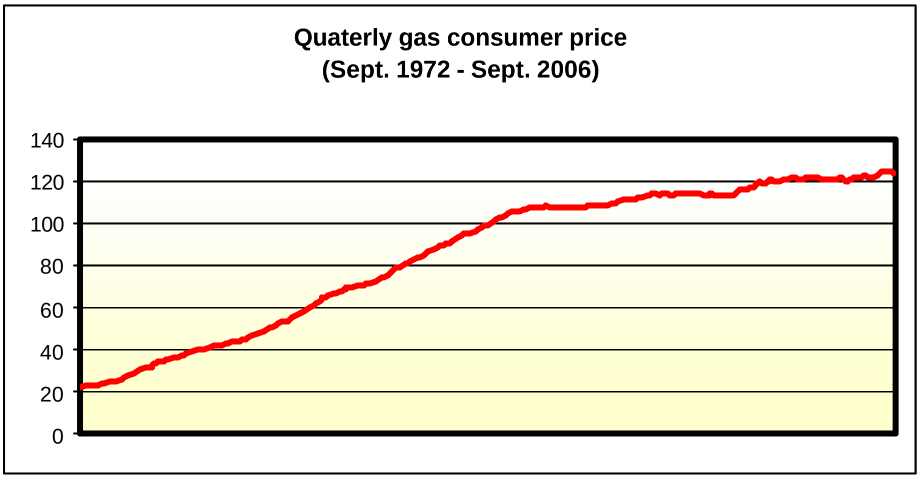
GASPRICEQ	GASQLN
1.000000	-0.435789
	(0.01758)
Log likelihood	36.09459

The cointegration test is applied on the logged series about respectively the price of imported minerals and the Australian production of gas from march 1982 to September 2006.

Here, cointegration is not found, because the unit root test does not reject the hypothesis about the no existence of cointegration.

The next series used in the analysis is about the paid price of household contents and services, among which there is gas provision.

Australian Bureau of Statistics: Australian Economic Indicators (2007)
CONSUMER PRICE INDEX of Household contents and services, QUARTERLY - By group
Original (1989–90 = 100.0)



The model finally chosen is an ARIMA (0, 1, 1)(0, 1, 0), which is different from the ARIMA (2,1,1) model suggested by the TRAMO automatic procedure. As showed in the following table, the unit root test confirms that TRAMO was deceived by the AR coefficient.

```
PARAMETER       ESTIMATE            STD ERROR       T RATIO    LAG
MA1   1         0.17981          0.84352E-01          2.13       1
ESTIMATES OF REGRESSION PARAMETERS CONCENTRATED OUT OF THE LIKELIHOOD
PARAMETER            VALUE              ST. ERROR       T VALUE
MU                       0.76967            (        0.07151)    10.76
OUT 1 (131)              -2.5270            (        0.69610)    -3.63    LS      ( 1 4107)
```

Augmented Dickey-Fuller Unit Root Test on CONPRICEGASQ

ADF Test Statistic	-1.136067	1% Critical Value*	-4.0298
		5% Critical Value	-3.4442
		10% Critical Value	-3.1467

*MacKinnon critical values for rejection of hypothesis of a unit root.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(CONPRICEGASQ)
Method: Least Squares

Sample(adjusted): 1973:4 2006:3
Included observations: 132 after adjusting endpoints

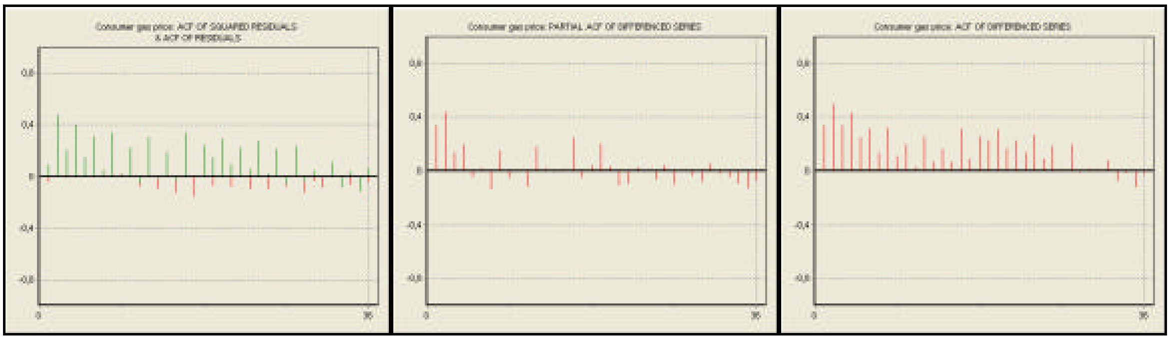
Variable	Coefficient	Std. Error	t-Statistic	Prob.
CONPRICEGASQ(-1)	-0.009279	0.008168	-1.136067	0.2581
D(CONPRICEGASQ(-1	-0.151864	0.088662	-1.712835	0.0892
D(CONPRICEGASQ(-2	0.249130	0.088883	2.802902	0.0059
D(CONPRICEGASQ(-3	0.210279	0.089003	2.362600	0.0197
D(CONPRICEGASQ(-4	0.299020	0.088950	3.361656	0.0010
C	0.795719	0.287690	2.765887	0.0065
@TREND(1972:1)	0.004152	0.007335	0.566074	0.5724

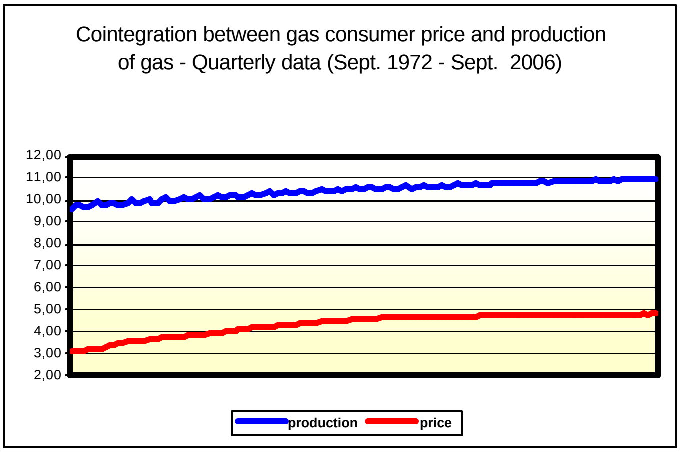
R-squared	0.311444	Mean dependent var	0.757576
Adjusted R-squared	0.278393	S.D. dependent var	0.892045
S.E. of regression	0.757769	Akaike info criterion	2.334696
Sum squared resid	71.77675	Schwarz criterion	2.487572
Log likelihood	-147.0900	F-statistic	9.423222
Durbin-Watson stat	1.922903	Prob(F-statistic)	0.000000

MODEL SUPPOSED: (0, 1, 1)(0, 1, 0)

```
AIC       294.189
 BIC       -0.6107
DURBIN-WATSON=   1.8128
LJUNG-BOX Q VALUE OF ORDER  8 IS
93.63
(SQUARED) LJUNG-BOX Q VALUE OF ORDER
8 IS     36.28
```

For any size level of the test, we can not reject the unit root hypothesis.
As showed in the ACF and in the PCF there is still some presence of path dependence impossible to avoid also using an ARIMA (0,1,1)(0,1,1).





Now, let us check if the quarterly series about gas production and this series about a consumer price index on expenses, among which there is also the gas expenditure, might be cointegrated, as suggested in the following graph.

Therefore, we can test whether they are cointegrated, that is, whether a linear function of these is I(0). As showed in the table in the newt page, the alternative hypothesis seems to be rejected, that is there is integration.

QUARTERLY GAS PRODUCTION

Testing for stationarity in the quarterly data about gas production after took the logarithm, arrives to this result: the series is an I(1) variable.

Correlogram of D(GASQLN)

Sample: 1972:1 2006:4
Included observations: 136

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.144	-0.144	2.8854	0.089
		2	-0.431	-0.461	28.920	0.000
		3	-0.002	-0.207	28.921	0.000
		4	0.284	0.048	40.379	0.000
		5	-0.144	-0.186	43.354	0.000
		6	-0.144	-0.097	46.352	0.000
		7	0.142	0.007	49.281	0.000
		8	-0.001	-0.144	49.281	0.000
		9	-0.001	0.074	49.282	0.000
		10	-0.001	0.001	49.282	0.000
		11	-0.001	-0.032	49.282	0.000
		12	-0.001	0.060	49.282	0.000
		13	-0.001	-0.026	49.282	0.000
		14	0.000	0.004	49.282	0.000
		15	-0.001	0.040	49.283	0.000
		16	-0.001	-0.031	49.283	0.000
		17	0.000	0.019	49.283	0.000
		18	0.000	0.002	49.283	0.000
		19	-0.001	-0.024	49.283	0.000
		20	-0.001	0.015	49.283	0.000
		21	0.000	-0.015	49.283	0.000
		22	0.000	-0.012	49.283	0.001
		23	0.000	0.007	49.283	0.001
		24	0.000	-0.017	49.283	0.002
		25	0.000	-0.001	49.283	0.003
		26	0.000	0.000	49.283	0.004
		27	0.000	-0.012	49.283	0.005
		28	0.000	0.004	49.283	0.008
		29	0.000	-0.004	49.283	0.011
		30	0.000	-0.005	49.283	0.015
		31	0.000	0.003	49.283	0.020
		32	0.000	-0.006	49.283	0.026
		33	0.000	-0.001	49.283	0.034
		34	0.000	0.000	49.283	0.044
		35	0.000	-0.005	49.283	0.055

CONSUMER PRICE INDEX ABOUT GAS

The same happens for the series about consumer price index: the series is an I(1) variable.

Correlogram of D(CONPRICEGASQ)

Sample: 1972:1 2006:4
Included observations: 136

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.061	0.061	0.5176	0.472
		2	0.414	0.411	24.466	0.000
		3	0.200	0.193	30.091	0.000
		4	0.408	0.298	53.710	0.000
		5	0.108	-0.014	55.394	0.000
		6	0.248	-0.031	64.247	0.000
		7	0.074	-0.102	65.045	0.000
		8	0.254	0.067	74.471	0.000
		9	0.073	0.041	75.256	0.000
		10	0.150	0.010	78.629	0.000
		11	-0.063	-0.184	79.232	0.000
		12	0.320	0.220	94.722	0.000
		13	-0.062	-0.019	95.312	0.000
		14	0.195	0.067	101.19	0.000
		15	0.001	0.018	101.19	0.000
		16	0.227	0.048	109.29	0.000
		17	0.078	0.106	110.24	0.000
		18	0.124	-0.051	112.70	0.000
		19	0.148	0.132	116.19	0.000
		20	0.223	0.090	124.22	0.000
		21	0.132	0.024	127.06	0.000
		22	0.195	0.023	133.36	0.000
		23	0.027	-0.136	133.47	0.000
		24	0.265	0.019	145.28	0.000
		25	0.063	0.047	145.96	0.000
		26	0.148	-0.030	149.70	0.000
		27	0.001	-0.043	149.70	0.000
		28	0.154	-0.058	153.82	0.000
		29	0.004	-0.053	153.82	0.000
		30	0.063	0.039	154.52	0.000
		31	-0.028	-0.055	154.66	0.000
		32	0.111	0.095	156.90	0.000
		33	-0.046	-0.076	157.28	0.000
		34	0.049	-0.059	157.71	0.000
		35	-0.139	-0.112	161.32	0.000

Johansen Cointegration Test

Sample: 1972:1 2006:4
Included observations: 132
Test assumption: Linear deterministic trend in the data
Series: CONPRICEGASQ GASQLN
Lags interval: 1 to 4

Eigenvalue	Likelihood Ratio	5 Percent Critical Value	1 Percent Critical Value	Hypothesized No. of CE(s)
0.063053	15.20708	15.41	20.04	None
0.048844	6.610178	3.76	6.65	At most 1 *

*(**) denotes rejection of the hypothesis at 5%(1%) significance level
L.R. rejects any cointegration at 5% significance level

Also here, we can not find out any cointegration between the price of gas and its production.

Unnormalized Cointegrating Coefficients:

CONPRICEGA	GASQLN
-0.004540	0.359760
0.001022	0.207503

Normalized Cointegrating Coefficients: 1 Cointegrating Equation(s)

CONPRICEGA	GASQLN	C
1.000000	-79.24726	700.1187
	(21.3195)	

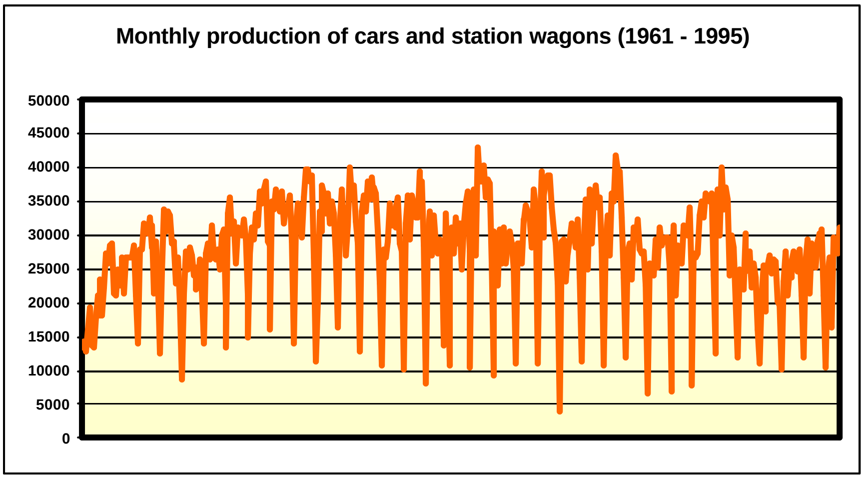
Log likelihood -103.3164

3.2 CARS

The Manufacturing industry contributed a little over 11% to Australia's gross domestic product in 2004–05. Although the value of Manufacturing industry gross value added has grown by 19% over the last ten years, the industry's share of the total production of goods and services in the economy has fallen from 14% to its current level over the period.
In the period 2000–01 to 2004–05, the largest increase in production was for Non-metallic mineral product manufacturing (26%), followed by Other manufacturing (15%) and Machinery and equipment manufacturing (14%). The largest increases between 2001–02 and 2004–05 were experienced in the production of cars and station wagons for fewer than ten persons and Portland cement. Production of these commodities increased by 25% and 23% respectively. Production of unfortified wine continued to increase over this period (22%) whereas the manufacture of beer experienced a slight decrease (3%). On the other hand, Between 2001–02 and 2005–06, the value of exports for transport equipment (excluding road vehicles) fell by 43% ($0.8b). Of course, the major commodity imported into Australia between 2001–02 and 2005–06 was petroleum, but road vehicles (including air cushion vehicles) made up 12% of imports.

The automotive industry is one of Australia's key manufacturing sectors and an important source of employment, and research and development. The increasing exposure of the Australian automotive industry to international competition has seen it develop to where it is now competing successfully in global markets. There is also a strong inter-dependence between the car makers and their suppliers, and strong linkages with the rest of the economy. The Australian automotive industry consists of four motor vehicle producers - Ford, Holden, Mitsubishi and Toyota - which produce large passenger motor vehicles (PMV) and variants, light commercial vehicles and sports utility vehicles. There are also over 200 motor vehicle component manufacturers. The four motor vehicle producers are based in Victoria and South Australia.

Nominal prices of transport equipment (including motor vehicles) are rising at a slower rate than the consumer price index (CPI), indicating a fall in 'real' prices. Over the five years prior to 2000-01, transport equipment prices grew by 7%, whereas the CPI increased by more than 11%. In addition, average weekly earnings increased by over 17% over the same time period, indicating that vehicles are becoming more affordable. Australian production contributes around 55% to total automotive supply. The high proportion of imports in total supply highlights the high level of import penetration in the Australian market. Household consumption and private sector investment provide the primary sources of domestic demand for total automotive output.



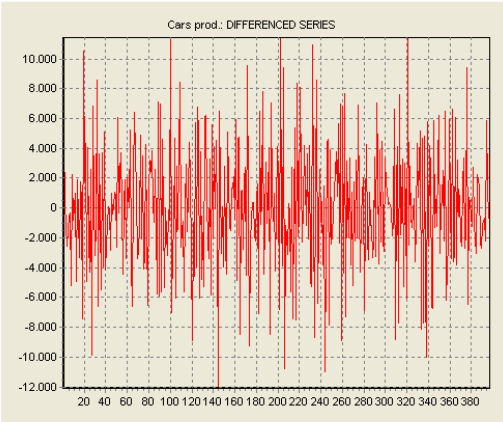
What we are going to test here if that the same ARIMA model can be applied to model both demand and supply of cars, even if the series seem to be very different in theirs original shape.

Source: ABS 8301.0 Jul 1961 - Aug 1995

Model finally chosen is an ARIMA (0,1,1)(0,1,1)

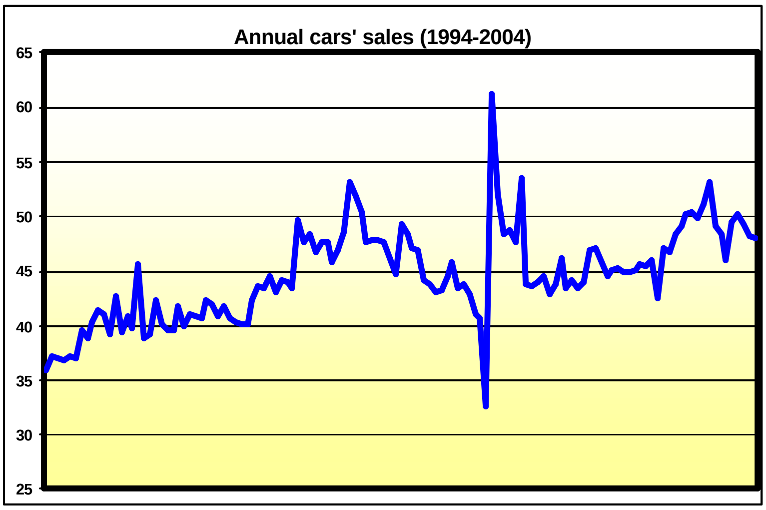
PARAMETER	ESTIMATE	STD ERROR	T RATIO	LAG
MA1 1	-.57221	0.41690E-01	-13.73	1
MA2 1	-.79431	0.35938E-01	-22.10	12

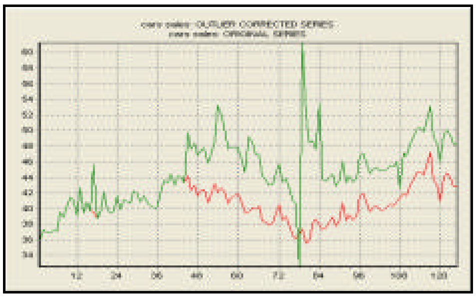
```
AIC      7608.407 and BIC      16.3158
DURBIN-WATSON=  1.9775
LJUNG-BOX Q VALUE OF ORDER 24 IS     71.68 > 36.415
(squared) LJUNG-BOX Q VALUE OF ORDER 24 IS    28.19 < 36.415
```



Actually, there are some problems to apply this model to this data, in the sense that the fit does not perform well as before. In any case, I was not able to find a model able to fit better this data, neither in their entire collection, nor using a part of the series. The suggestion of the TRAMO automatic procedure was an ARIMA (2,1,1) (0,1,1), but the second AR coefficient proposed was too near the unit root to be accepted, so I tested an ARIMA (1,1,1) (0,1,1), which had a better performance in term of AIC (- 491.957) and BIC (- 4.0263), but again problems with randomness: LJUNG-BOX Q value of order 24 was 59.82 > 36.415, and for the squared residuals LJUNG-BOX Q value of order 24 was 57.74 > 36.415. Finally, I preferred to use the first model proposed, I mean the ARIMA (0,1,1) (0,1,1).

Series Title: Motor vehicle sales: Passenger vehicles (thousands, SA): Australia





The red line showed the outlier corrected series

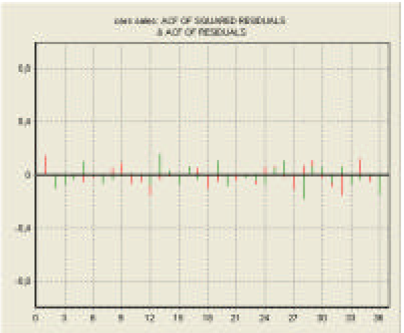
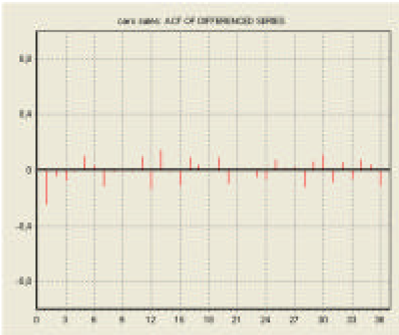
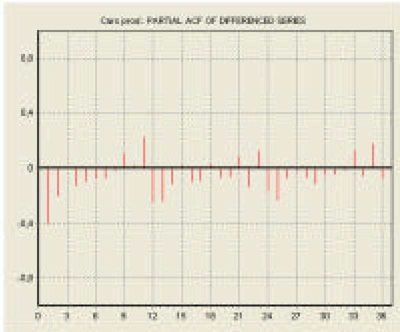
MODEL FINALLY CHOSEN: ARIMA(0,1,1)(0,0,0) in which logs have been selected.

WITHOUT MEAN, WITHOUT TRADING DAY CORRECTION, WITHOUT EASTER CORRECTION

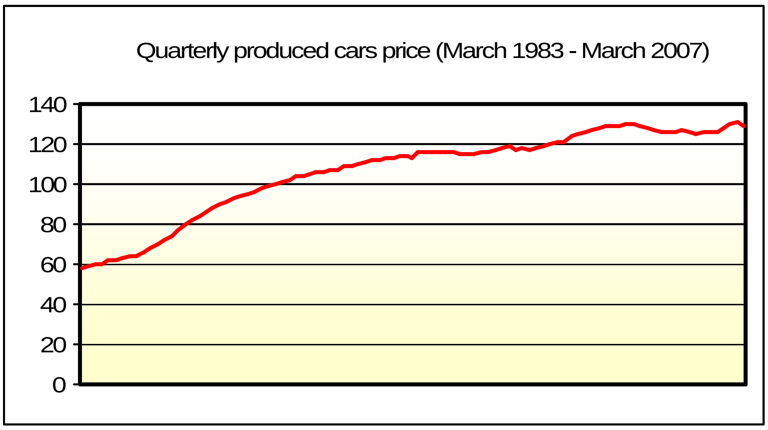
OUTLIERS

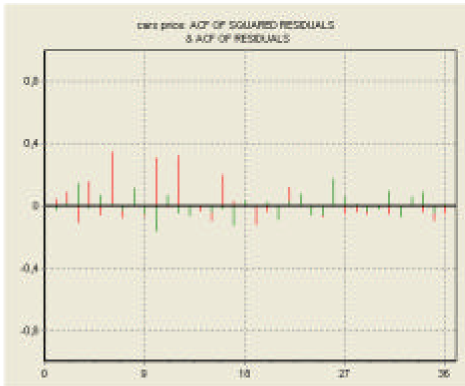
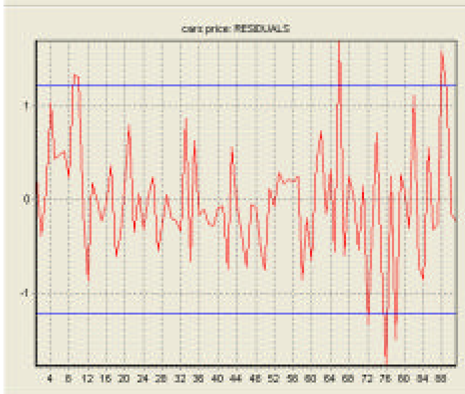
PARAMETER	VALUE	ST. ERROR	T VALUE		
OUT 1 (79)	0.54456	(0.02823)	19.29	TC	(7 2000)
OUT 2 (78)	-.23754	(0.02822)	-8.42	TC	(6 2000)
OUT 3 (84)	0.16599	(0.02479)	6.69	AO	(12 2000)
OUT 4 (17)	0.14218	(0.02477)	5.74	AO	(5 1995)
OUT 5 (45)	0.11658	(0.02788)	4.18	LS	(9 1997)
OUT 6 (54)	0.11728	(0.02785)	4.21	TC	(6 1998)
OUT 7 (108)	-.88081E-01	(0.02477)	-3.56	AO	(12 2002)
OUT 8 (63)	0.95058E-01	(0.02785)	3.41	TC	(3 1999)

```
METHOD OF ESTIMATION: EXACT MAXIMUM LIKELIHOOD
PARAMETER       ESTIMATE           STD ERROR      T RATIO    LAG
MA1  1          -.36635           0.83559E-01      -4.38       1

AIC    -509.2271 and BIC      -6.7410
DURBIN-WATSON=  1.9736
LJUNG-BOX Q VALUE OF ORDER 24 IS    12.56 < 36.415
LJUNG-BOX Q VALUE OF ORDER 24 IS    12.09 < 36.415
```

Thus, in the case of the price of cars we can apply the ARIMA (0,1,1) (0,1,1) obtaining an optimal fit with the data and respect for the residuals randomness assumption.



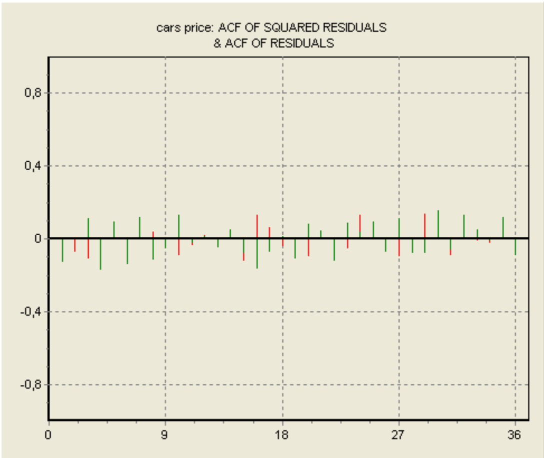



The model finally chosen by TRAMO automatic procedure is an ARIMA (1,1,1)(0,0,0), with the following results

```
LEVELS are Selected, WITH MEAN, WITHOUT TRADING DAY
CORRECTION, WITH EASTER CORRECTION

OUTLIERS
  97 AO     ( 1 2007)
  49 AO     ( 1 1995)
  64 LS     ( 4 1998)

METHOD OF ESTIMATION: EXACT MAXIMUM LIKELIHOOD
PARAMETER       ESTIMATE           STD ERROR      T RATIO    LAG
AR1  1          -.89253           0.63868E-01     -13.97       1
MA1  1          -.43893           0.12931          -3.39       1

AIC     188.9918 and  BIC      -0.6907
DURBIN-WATSON=  2.0425
LJUNG-BOX Q VALUE OF ORDER 16 IS     8.07 < 26.296
(Squared) LJUNG-BOX Q VALUE OF ORDER 16 IS    44.19 > 26.296
```

I propose a model that shows to fit better the data, that is an ARIMA (0,1,1)(0,1,0)

PARAMETER	VALUE	ST. ERROR	T VALUE		
MU	-.35799E-01	(0.10931)	-0.33		
OUT 1 (84)	-2.2118	(0.36693)	-6.03	LS	(4 2003)
OUT 2 (49)	-1.4696	(0.20280)	-7.25	AO	(1 1995)
OUT 3 (64)	-2.1485	(0.36373)	-5.91	LS	(4 1998)
OUT 4 (96)	1.3363	(0.34378)	3.89	AO	(4 2006)
OUT 5 (71)	-1.2540	(0.21164)	-5.93	AO	(3 2000)
OUT 6 (86)	1.0686	(0.20796)	5.14	AO	(2 2004)
OUT 7 (77)	1.1171	(0.32625)	3.42	TC	(1 2002)
OUT 8 (9)	-.69270	(0.20471)	-3.38	AO	(1 1985)
OUT 9 (94)	1.5004	(0.50941)	2.95	LS	(2 2006)



METHOD OF ESTIMATION: EXACT MAXIMUM LIKELIHOOD

PARAMETER	ESTIMATE	STD ERROR	T RATIO	LAG
MA1 1	0.69343	0.75119E-01	9.23	1

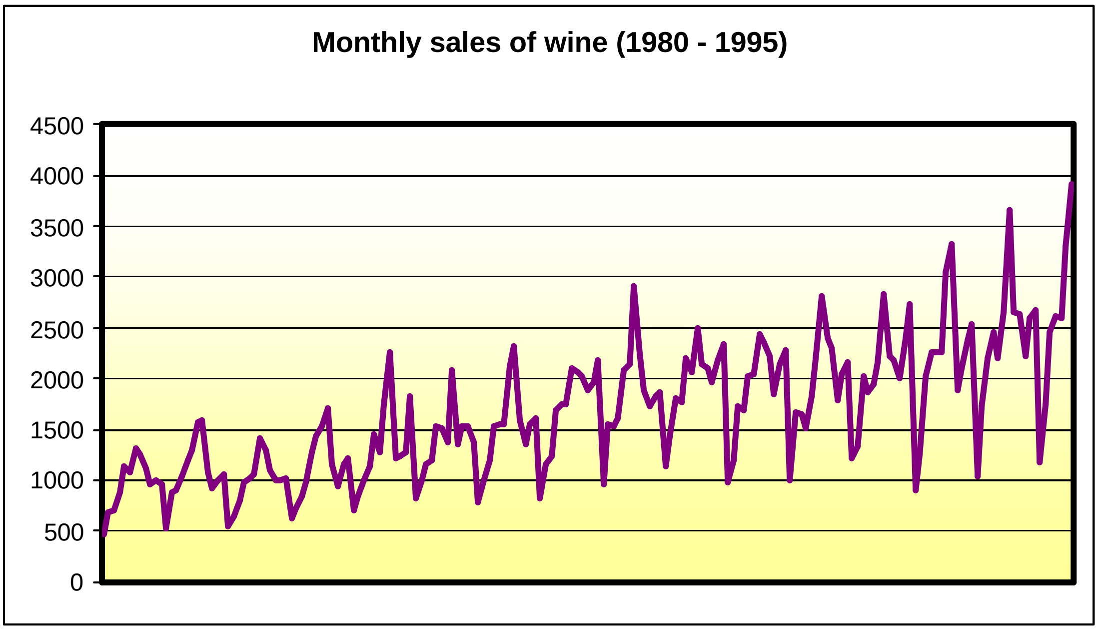
AIC 184.231 and BIC -0.5410
DURBIN-WATSON= 2.2293
LJUNG-BOX Q VALUE OF ORDER 16 IS 15.33 < 26.296
(Squared) LJUNG-BOX Q VALUE OF ORDER 16 IS
7.24 < 26.296

3.3 WINE

Grapes are a temperate crop requiring predominantly winter rainfall and warm to hot summer conditions for ripening. Almost all grape production in Australia depends on irrigation water as a supplement to rainfall. An absence of late-spring frosts is essential if the loss of the developing fruit is to be prevented. Grapes are grown for winemaking, drying, and to a lesser extent, for table use. The better known grape producing areas include the Adelaide Hills, Barossa Valley, Clare Valley, Riverland, McLaren Vale and Coonawarra (all in South Australia); Sunraysia and the Yarra Valley (Victoria); the Hunter and Riverina (New South Wales); the Swan Valley and Margaret River (Western Australia); and the Tamar Valley and Coal River Valley (Tasmania). The gross value of grape production for 2004–05 fell by 11% from the previous year, to $1,508m.
In contrast, Australian wine has won an international reputation for quality and value. Australian wines have taken key international awards, competing favourably against longer-established national wine industries. Australia produces a full range of wine styles from full-bodied reds and deep fruity whites through to sparkling, dessert and fortified styles. In global terms, Australia was ranked 7th in the list of world wine producers in 2003, producing 1,085 million litres of wine.
Wine-grape growing and winemaking are carried out in each of the six States and two mainland Territories of Australia. The principal production areas are located in the south-east quarter of the Australian continent, in the states of South Australia, New South Wales and Victoria. However, wine is produced in over 60 regions, reflecting the wide range of climates and soil types available in the continent. In 2003-04 Shiraz was the most-produced variety, followed by Chardonnay and Cabernet Sauvignon. Premium white varieties other than Chardonnay include Semillon, Riesling and Colombard.

Now, wine is very much a part of Australian life, closely associated with both business and leisure. Wine consumption is often linked to the country's outdoor-oriented lifestyle as well as to the cosmopolitan urban way of life of the bulk of the Australian population. Wine festivals are a feature of cultural life in the major wine producing regions of Australia and draw many Australian holidaymakers and international visitors each year. From an historical point of view, the end of the Second World War saw a rapid influx of migrants from Europe who brought with them a strong culture related to wine and provided an important impetus to the Australian wine industry. However it is the period 1996 to 2004 that has seen spectacular growth in exports following rapidly increasing appreciation of Australian wines overseas. Major wine producers from abroad have invested in Australian wineries and Australian companies have taken controlling interests in wineries in countries such as France and Chile.
In 2003-04 sales of Australian wine totaled approximately 999 million litres, with 414 million litres sold domestically and 584 million litres exported. Australian wine exports were worth $2.5 billion, with represented an increase of 12.7% over the previous year.

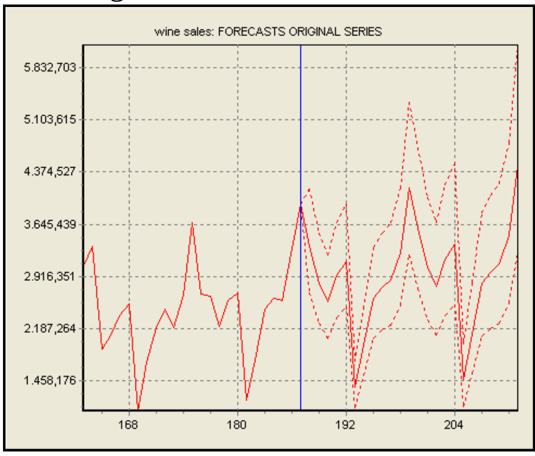
Monthly Australian sales of red wine: thousands of liters.
Jan 1980 - Jul 1995. Source: ABS.

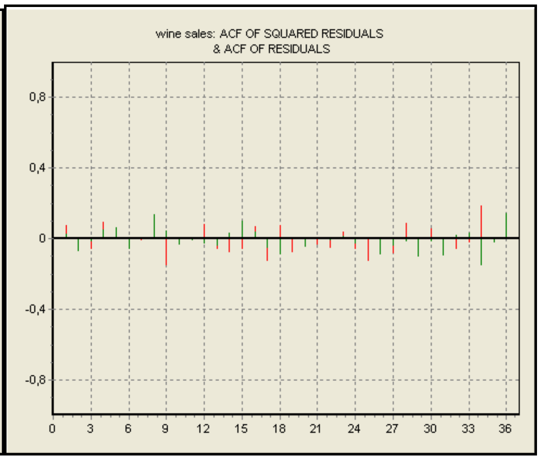


MODEL FINALLY CHOSEN: ARIMA(0,1,1)(0,1,1) in which logs are selected

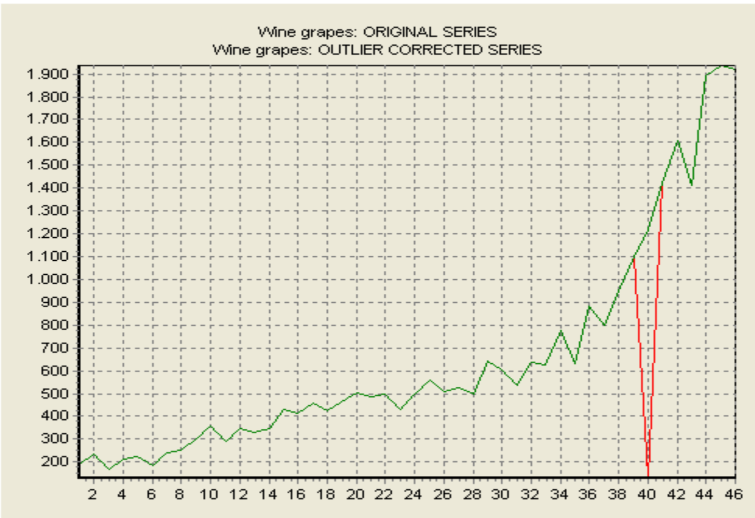
```
WITHOUT MEAN, WITHOUT TRADING DAY CORRECTION, WITHOUT EASTER CORRECTION

METHOD OF ESTIMATION: EXACT MAXIMUM LIKELIHOOD
PARAMETER      ESTIMATE          STD ERROR      T RATIO    LAG
MA1  1         -.81312         0.47548E-01      -17.10      1
MA2  1         -.74146         0.66648E-01      -11.13     12

AIC    -261.4457  and  BIC      -4.3480
DURBIN-WATSON=  1.9507
PARAMETER           VALUE            ST. ERROR    T VALUE
OUT 1 ( 67)        -.34845        (     0.09769)   -3.57     AO    ( 7 1985)

LJUNG-BOX Q VALUE OF ORDER 24 IS  11.66 < 36.415
(Squared)  LJUNG-BOX Q VALUE OF ORDER 24 IS  20.01 < 36.415
```

It is very clear how much this model fit well with our data: the BIC and AIC are very low, there are no problems with normality nor with linearity. Thus, this model also allows to forecast the values of future lags of the series.





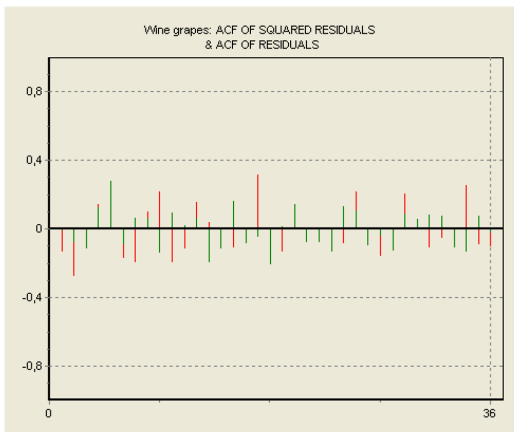
Wine grapes (thousand tonnes): Australia 1961-2006



In this case the green line represents the outliers corrected series, as in the following graph about consumer price for alcohol and tobacco.

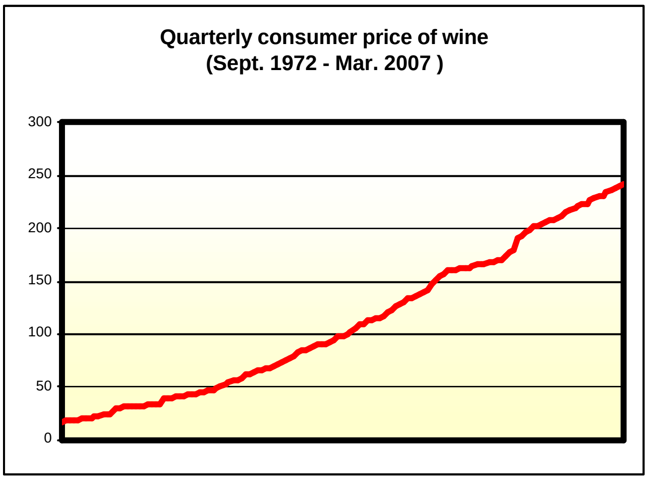
The model finally chosen is an ARIMA (1,1,0) with mean correction.

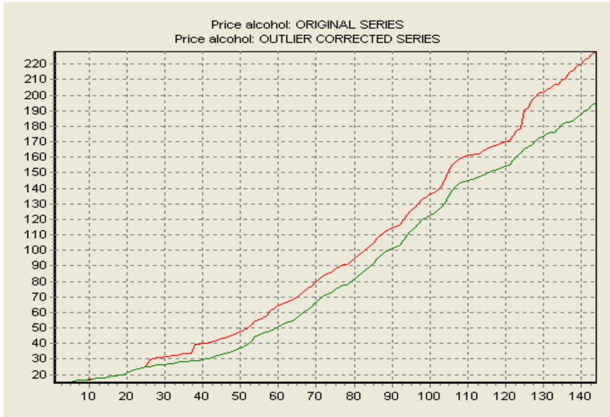
```
OUTLIERS
PARAMETER     VALUE          ST. ERROR    T VALUE
MU            0.50792E-01    (0.01233)     4.12
OUT 1 (40)   -2.2629         (0.10427)   -21.70
AO (1 2000)

METHOD OF ESTIMATION: EXACT MAXIMUM LIKELIHOOD
PARAMETER  ESTIMATE   STD ERROR   T RATIO   LAG
AR1  1     0.56000    0.12350     4.53      1

AIC     -54.0321 and  BIC      -3.9265
DURBIN-WATSON=  1.9968
LJUNG-BOX Q VALUE OF ORDER  8 IS 6.30
(Squared) LJUNG-BOX Q VALUE OF ORDER  8 IS
10.69 < 15.507
```



Consumer Price Index: Alcohol and Tobacco

Australia – Quarterly data from June 1969 to March 2006





The green line represents the outliers corrected series.

The model finally chosen is an ARIMA (1,1,1) (0,1,1), instead the model selected by the TRAMO automatic procedure was an ARIMA (2,1,1) (0,0,0), with a worst fit that our model, and, also, trouble in the unit roots, as showed later.

```
METHOD OF ESTIMATION: EXACT MAXIMUM LIKELIHOOD

 PARAMETER       ESTIMATE           STD ERROR      T RATIO    LAG

 AR1   1         -.62954          0.18066          -3.48       1
 MA1   1         -.26735          0.22439          -1.19       1
 MA2   1         -.79714          0.69028E-01     -11.55       4

PARAMETER            VALUE         ST. ERROR     T VALUE
 MU                  0.44825E-01   (      0.02357)    1.90
 OUT 1 (125)          9.3880       (      0.55997)   16.77    LS      ( 1 2000)
 OUT 2 ( 38)          4.7678       (      0.48083)    9.92    TC      ( 2 1978)
 OUT 3 ( 26)          4.0179       (      0.54974)    7.31    LS      ( 2 1975)
 OUT 4 (104)          3.5928       (      0.57737)    6.22    LS      ( 4 1994)
 OUT 5 (127)          2.4041       (      0.55261)    4.35    LS      ( 3 2000)
 OUT 6 (105)          3.0381       (      0.60277)    5.04    LS      ( 1 1995)
 OUT 7 (121)         -1.6157       (      0.34801)   -4.64    AO      ( 1 1999)
 OUT 8 ( 58)          2.3452       (      0.54732)    4.28    LS      ( 2 1983)
 OUT 9 (113)         -1.4652       (      0.34647)   -4.23    AO      ( 1 1997)
 OUT10 (106)          2.1771       (      0.57924)    3.76    LS      ( 2 1995)
 OUT11 (137)          2.0191       (      0.57097)    3.54    LS      ( 1 2003)
 OUT12 ( 86)          1.8923       (      0.54751)    3.46    LS      ( 2 1990)

AIC       278.845 and   BIC    -0.5227
DURBIN-WATSON=  2.0024
LJUNG-BOX Q VALUE OF ORDER 16 IS     15.24 < 26.296
LJUNG-BOX Q VALUE OF ORDER 16 IS     16.67 < 26.296
```

Augmented Dickey-Fuller Unit Root Test on D(PRICEWINE)

ADF Test Statistic	-4.308355	1% Critical Value*	-3.4804
		5% Critical Value	-2.8832
		10% Critical Value	-2.5782

*MacKinnon critical values for rejection of hypothesis of a unit root.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(PRICEWINE,2)
Method: Least Squares

Sample(adjusted): 1974:1 2007:1
Included observations: 133 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(PRICEWINE(-1))	-0.584333	0.135628	-4.308355	0.0000
D(PRICEWINE(-1),2)	-0.353576	0.132742	-2.663625	0.0087
D(PRICEWINE(-2),2)	-0.029314	0.131954	-0.222155	0.8245
D(PRICEWINE(-3),2)	0.149935	0.120496	1.244316	0.2157
D(PRICEWINE(-4),2)	0.041436	0.088842	0.466397	0.6417
C	0.898569	0.237716	3.780006	0.0002

R-squared	0.503501	Mean dependent var	0.015038
Adjusted R-squared	0.483954	S.D. dependent var	2.033749
S.E. of regression	1.460971	Akaike info criterion	3.640143
Sum squared resid	271.0734	Schwarz criterion	3.770535
Log likelihood	-236.0695	F-statistic	25.75823
Durbin-Watson stat	1.991534	Prob(F-statistic)	0.000000

Correlogram of D(ORG_PRICEWINE)

Included observations: 143

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.160	0.160	3.7470	0.053
		2	0.324	0.306	19.203	0.000
		3	0.230	0.165	27.064	0.000
		4	0.044	-0.105	27.346	0.000
		5	0.079	-0.042	28.293	0.000
		6	0.105	0.096	29.947	0.000
		7	-0.008	-0.022	29.955	0.000
		8	0.098	0.045	31.444	0.000
		9	0.014	-0.018	31.473	0.000
		10	0.147	0.144	34.855	0.000
		11	0.143	0.114	38.089	0.000
		12	0.143	0.050	41.314	0.000
		13	0.041	-0.118	41.586	0.000
		14	-0.001	-0.122	41.586	0.000
		15	0.078	0.112	42.570	0.000
		16	0.019	0.062	42.631	0.000
		17	0.122	0.104	45.088	0.000
		18	0.114	0.035	47.235	0.000
		19	0.182	0.147	52.754	0.000
		20	0.185	0.102	58.546	0.000
		21	0.204	0.052	65.651	0.000
		22	0.114	-0.099	67.865	0.000
		23	0.116	-0.060	70.188	0.000
		24	-0.018	-0.053	70.242	0.000
		25	0.038	0.046	70.498	0.000
		26	0.008	0.043	70.510	0.000
		27	0.095	0.081	72.110	0.000
		28	0.086	0.065	73.451	0.000
		29	0.096	-0.005	75.134	0.000
		30	0.045	-0.109	75.507	0.000
		31	0.139	0.007	79.079	0.000
		32	0.131	0.112	82.263	0.000
		33	0.076	0.025	83.364	0.000
		34	0.077	0.000	84.498	0.000
		35	0.041	0.005	84.826	0.000
		36	-0.016	-0.036	84.875	0.000

The Durbin-Watson statistic is a test for first-order serial correlation, that is the DW statistic measures the linear association between adjacent residuals from a regression model. If there is no serial correlation, the DW statistic will be around 2. The DW statistic will fall below 2 if there is positive serial correlation (in the worst case, it will be near zero). If there is negative correlation, the statistics will lie somewhere between 2 and 4. Positive serial correlation is the most commonly observed form of dependence. As a rule of thumb, with 50 or more observations and only a few independent variables, a DW statistic below about 1.5 is a strong indication of positive first order serial correlation.

4. CONCLUSIONS

We can summarize our analysis saying that series about production or sales are easy to forecast because of their linear behaviour. The model that best fits with the data, independently of the nature of goods sold or produced, is an ARIMA (0,1,1), that is an exponential smoothing model either in the regular or in the seasonal part or in both of them. Focusing our attention on the indexes price series, we have discovered that, with small changes, we can again apply ARIMA (0,1,1) or mixed ARIMA model, like an ARIMA (1,1,1). Finally, we found out that cointegration among series about production and series about the price of this production is a complex one, and maybe it needs much more knowledge about the phenomenon under study and its evolution over time.