Seasonality As An Unobservable Component In South African Agricultural Market Data

Paul Styger, (Email: paul.styger@nwu.ac.za), North-West University, South Africa
André Heymans, North-West University, South Africa

ABSTRACT

The shortcoming of most of the tests for seasonal patterns is that the problem under investigation is formulated in a stringent manner, leading to a test of the null hypothesis of no seasonality against the alternative of deterministic seasonality. These tests would typically involve using models that incorporate deterministic dummy variables that are then used to capture seasonal effects in the data by means of a regression. This implies that the possibility of stochastic seasonality, which is manifested by changing seasonal factors over the sample period (deterministic seasonality implies constant seasonal factors), is ignored. A possible more comprehensive/rigorous approach would thus be to test for the presence of stochastic seasonality versus deterministic seasonality. The aim of this paper is to investigate the possible presence of unobserved seasonal components in South African agricultural market data, with the emphasis to demonstrate the additional insight such an approach can provide to researchers and especially for pairs traders.

1. INTRODUCTION

The study of seasonal patterns in economic and financial time series is not new. The day-of-the-week effect was first observed and documented by M.F.M. Osborne in 1959 (Osborne, 1959). Since Osborne’s discovery, Cross (1973), French (1980), Gibbons and Hess (1981), Lakonishok and Levi (1982), Keim and Stambaugh (1984), and Rogalski (1984) amongst others have confirmed that there are differences in the distributions of stock returns in each of the week days.

Although these studies have been performed on the equity markets in the U.S., the day-of-the-week effect has been investigated for both international equity and non-equity financial markets. Jaffe and Westerfield (1985a, 1985b) found significant negative mean returns on Mondays in the U.S., Canada and the UK stock markets, and significant negative Tuesday returns in the Japanese and Australian stock markets. Aggarwal and Rivoli (1989) studied the emerging markets of East Asia and observed lower mean returns on Mondays and Tuesdays in the stock returns of Hong Kong, Singapore, Malaysia and the Philippines, from September 1976 to June 1988. The day-of-the-week effect is furthermore detected in the commodity and stock futures markets by Cornell (1985), Dyl and Maberly (1986), Gay and Kim (1987); in the Treasury bill market by Flannery and Protopapadakis (1988); and in the foreign exchange market by Corhay, Fatemi, and Rad (1995).

While the focus of these studies has been the seasonal pattern in mean returns, there are several other empirical studies investigating the time-series behaviour of stock prices in terms of volatility by using variations of the GARCH models. French, Schwert, and Stambaugh (1987), Akgiray (1989), Baillie and DeGennaro (1990), Hamao, Masulis and Ng (1990), Nelson (1991), Campbell and Hentschel (1992), and Ogum, Nouyrigat and Beer (2002) tests for these volatility effects by means of GARCH models. Their studies report that the expected returns in stock markets are time-varying and conditionally heteroskedastic.
These studies do not, however, examine the issue of day-of-the-week variations in stock market volatility. French and Roll (1986) pointed out that the variances for the days following an exchange holiday should be larger than other days. Harvey and Huang (1991) observe higher volatility in the interest rates and foreign exchange futures markets during first trading hours on Thursdays and Fridays, while Berument and Kiymaz (2001) tested the presence of the day-of-the-week effect on stock market volatility. Berument and Kiymaz (2001) found a day-of-the-week effect in both volatility and return equations by making use of the S&P 500 index, and Heymans (2005) found high mean returns on Thursdays by making use of South Africa’s All share index.

2. UNOBSERVED SEASONALITY

The shortcoming of the tests for seasonal patterns, such as the ones described above, is that the problem under investigation is formulated in a stringent manner, leading to a test of the null hypothesis of no seasonality against the alternative of deterministic seasonality. These tests would typically involve using models that incorporate deterministic dummy variables that are then used to capture seasonal effects in the data by means of a regression. This implies that the possibility of stochastic seasonality, which is manifested by changing seasonal factors over the sample period (deterministic seasonality implies constant seasonal factors), is ignored.

A possible more comprehensive/rigorous approach would thus be to test for the presence of stochastic seasonality versus deterministic seasonality. Specifically, a general model that allows for all possibilities should be formulated, and by fitting this model to historical data it can be determined whether the seasonality is stochastic or deterministic, or if there is no seasonality present.

Recent studies on developing economies’ market data (Al-Loughani 2003), Al-Saad and Moosa (2005) and Al-Deehani (2006)) show that there are also unobserved seasonality in stock exchange returns data. Significant seasonal factors were found to be stochastic rather than deterministic. Because traditional time series models assume deterministic seasonality only, they cannot test for these unobserved components.

The aim of this paper is to investigate the possible presence of unobserved seasonal components in South African agricultural market data, with the emphasis to demonstrate the additional insight such an approach can provide to researchers and especially for pairs traders.

3. DATA

The data consists of the daily (end-of-the-day) closing prices on the South African Futures Exchange (SAFEX) for both white and yellow maize from 26 March 1998 to 13 July 2006. These spot series consist of 2136 observations, excluding all trading holidays and weekends for both the white and yellow maize series. For the period of observation, there were 413 observations for Monday, 435 observations for Tuesday, 433 observations for Wednesday, 432 observations for Thursday and 423 observations for Friday.

In order to eliminate the effects of an increase in variability around the trend, a log transformation of the data is used. The data are thus transformed to the natural logarithms of the daily closing prices of the maize futures contracts for day t:

\[ P_t = \ln(Price_t) \]

To test for the hidden monthly seasonal effects, the data was converted to monthly prices by taking the arrhythmic mean (average) of the logarithmic daily closing prices.
4. **METHODOLOGY**

4.1 **Estimating The Model – General Model Specification**

In testing for hidden monthly seasonal effects, it is necessary to specify the correct model. This will be done at the hand of the 4 components: the trend, $\mu_t$, the cyclical component, $\phi_t$, the seasonal component $\gamma_t$, and $\epsilon_t$, the random component. The same tests were performed in both STAMP™ and SAS®. Several preliminary models were used in order to decide on the final model. Sections 4.2 to 4.4 will discuss this process.\(^1\)

4.2 **Modelling The Trend Component**

The trend component consists of a level (the actual value of the trend), and a slope (the drift factor). The default setting in STAMP™ allows for both the level and slope to be stochastic. It is however possible to specify some special cases: Random walk with noise (stochastic level and no slope); local level with drift (stochastic level and fixed slope); and a smooth trend (fixed level and a stochastic slope) (Koopman et al, 2006:55). The same options are available when modelling the trend component in SAS®’s PROC UCM.

In order to assess which of the special cases will be used for the purpose of testing the data in the final model, the default setting was specified in the preliminary model to test for the statistical significance of the slope component.

4.3 **Modelling The Cyclical Component**

The parameters of the cycle are estimated from the data and consist of: a variance parameter, $s_k^2$ which is responsible for making the cycle stochastic; a period, $2\pi/\lambda$, a damping factor $\rho$, a frequency (in radians), $c\lambda$, and a variance of the disturbance terms in its stochastic equations (SAS Institute Inc., 2006). Contrary to a deterministic cycle where the sine-cosine wave has a given period, a stochastic cycle is constructed by shocking it with disturbances and introducing a damping factor (Koopman et al, 2006:59).

The statistical specification of the stochastic cycle is:

$$
\begin{bmatrix}
\psi_t \\
\psi_t^v
\end{bmatrix}
= \rho
\begin{bmatrix}
\cos \lambda & \sin \lambda \\
-\sin \lambda & \cos \lambda
\end{bmatrix}
\begin{bmatrix}
\psi_{t-1} \\
\psi_{t-1}^v
\end{bmatrix}
+ \begin{bmatrix}
v_t \\
v_t^v
\end{bmatrix}, \quad t = 1, \ldots, T. \tag{4.1}
$$

In testing for seasonal effects in the agricultural commodities data, a generalised specification of the cycle was used. This cycle is smoother and can be specified as $\psi_t = \psi_t^{(1)}$ where:

$$
\begin{bmatrix}
\psi_t^{(j)} \\
\psi_t^{(j)}^v
\end{bmatrix}
= \rho
\begin{bmatrix}
\cos \lambda & \sin \lambda \\
-\sin \lambda & \cos \lambda
\end{bmatrix}
\begin{bmatrix}
\psi_{t-1}^{(j)} \\
\psi_{t-1}^{(j)}^v
\end{bmatrix}
+ \begin{bmatrix}
\psi_t^{(j-1)} \\
\psi_t^{(j-1)}^v
\end{bmatrix}, \quad j = 1, \ldots, k \tag{4.2}
$$

for $j = 1, \ldots, k$, $t = 1, \ldots, n$, and where $v_t = \psi_t^{(0)}$ and $v_t^v = \psi_t^{(0)}$ are mutually uncorrelated white noise disturbances with zero means and common variance $s_v^2$ (Koopman et al, 2006:63). For this equation, the value of $k$ determines the smoothness of the cycle.

\(^1\) Although several tests were performed, only the final model will be discussed here.
4.4 Modelling The Seasonal Component

Seasonal effects are considered corrections to the general trend of the series due to the seasonal variations. These effects sum to zero over the full season cycle and are therefore modelled as a stochastic periodic pattern of an integer period \( s \) such that the sum \( \sum_{i=0}^{s-1} \gamma_i \) is always zero in the mean.

The seasonal component can be modelled in either a dummy variable or trigonometric form. In SAS\textsuperscript{®}, the dummy variable form is the default. In order to run the test for the trigonometric form the TYPE=TRIG option should be specified in the PROC UCM module. It is furthermore possible to specify the length of the seasonal component. A seasonal component with season length 4 will correspond to quarterly seasonality while season length 12 indicates testing for monthly seasonality (SAS Institute Inc., 2006).

All of these components will be combined to test for any unobserved seasonal patterns in the data. Section 5 below will discuss the testing of unobserved seasonality in white maize, while section 6 discusses yellow maize.

5. WHITE MAIZE

5.1 The Model

The final model used in testing for hidden seasonality in the white maize data takes the form of:

\[
\gamma_t = \mu_t + \phi_t + \gamma_t + \epsilon_t.
\] (5.1)

where \( \mu_t \) is the trend, \( \phi_t \) the cyclical component, \( \gamma_t \) the seasonal component, and \( \epsilon_t \) the random component. During the preliminary testing phase, the \( p \)-value for the slope component was reported at (0.2666) and that of the random component (0.9996). These 2 components were excluded from the final model. For the final model, the trend component thus takes the form of a random walk with noise. The model for \( \mu_t \) is thus given by:

\[
\mu_t = \mu_{t-1} + \eta_t, \quad \eta_t \sim \text{i.i.d. } N(0, \sigma_\eta^2),
\] (5.2)

where the disturbance term \( \eta_t \) is assumed to be independent. The trend component \( \mu_t \) is a random walk (the level) and is thus modelled as a time-varying level component without any persistent upward or downward drift (no slope component is included).

For the cyclical component Harvey and Trimbur (2003) was followed in making use of the smoothed cycle option in SAS\textsuperscript{®}. They prefer the smoothed cycle specification because of its better fit characteristics. In the final model used in testing for unobserved seasonality in the white maize data will thus take the form of \( \psi_t = \psi_t^{(s)} \) where:

\[
\begin{bmatrix}
\psi_t^{(j)} \\
\psi_t^{(j)}
\end{bmatrix} = \rho \begin{bmatrix}
\cos \lambda & \sin \lambda \\
-\sin \lambda & \cos \lambda
\end{bmatrix} \begin{bmatrix}
\psi_t^{(j-1)} \\
\psi_t^{(j-1)}
\end{bmatrix} + \begin{bmatrix}
\psi_t^{(j-1)} \\
\psi_t^{(j-1)}
\end{bmatrix},
\] (5.3)

for \( j = 1, \ldots, k \), \( t = 1, \ldots, n \), and where \( \psi_t = \psi_t^{(0)} \) and \( \psi_t^{*} = \psi_t^{(0)} \) are mutually uncorrelated white noise disturbances with zero means and common variance \( s^2 \). In this case \( \rho < 1 \), and \( \psi_t \) thus has a stationary distribution with zero mean and variance \( \sigma_\psi^2 / (1 - \rho^2) \). The value of \( k \) (the order of the cycle) determines the smoothness of the cycle.

For the seasonal component, Koopman et al (2006:56) suggest that the seasonal component should rather be fixed and not allowed to be stochastic in the case where there is not enough data to allow a changing pattern to be
estimated. Since there are only 103 months (data points) available, the seasonal component will take the trigonometric form. In this form, \( y_i \) is modelled as a sum of cycles of different frequencies. This model is given as (Harvey & Scott, 1994:1328):

\[
\sum_{j=1}^{\lfloor s/2 \rfloor} y_{j,t},
\]

where \( s \) equals \((s - 1)/2\). The cycles \( y_{j,t} \) have frequencies \( \lambda_j = 2\pi j / s \) and are specified by the matrix equation (SAS, 2006:1633):

\[
\begin{bmatrix}
\gamma_{j,t} \\
\gamma'_{j,t}
\end{bmatrix} =
\begin{bmatrix}
\cos \lambda_j & \sin \lambda_j \\
-\sin \lambda_j & \cos \lambda_j
\end{bmatrix}
\begin{bmatrix}
\gamma_{j,t-1} \\
\gamma'_{j,t-1}
\end{bmatrix} + \begin{bmatrix}
\omega_{j,t} \\
\omega'_{j,t}
\end{bmatrix},
\]

where the disturbances \( \omega_{j,t} \) and \( \omega'_{j,t} \) are assumed to be independent. Where \( j \) is fixed, \( \omega_{j,t} \) and \( \omega'_{j,t} \) \( \sim N(0, \sigma^2) \). Since the seasonal component for this model is deterministic, the decomposition of the seasonal effects into these harmonics \( \{ y_{j,t} \}_{j} \) is identical to its Fourier decomposition, and the sum of squares of the seasonal factors is equal to the sum of squares of the amplitudes of the harmonics.

### 5.2 Model Results

The results of the final model used in testing for hidden seasonal patterns in the white maize data indicated that the random component ought to be excluded from the model. This result seems to be common when fitting the different components to time series data (Harvey et al, 2006:2). Even though a higher order cycle was used, the problem of implausible estimates remains.\(^2\) The results of the free parameters can are set out in table 5.1 below.

| Component          | Parameter        | Estimate | Std Error | t - Value | Pr > |t| |
|--------------------|------------------|----------|-----------|-----------|-------|---|
| Level              | Error Variance   | 0.00038242 | 0.0012414 | 0.31      | 0.7582|
| Season             | Error Variance   | 3.27E-12  | 2.40E-09  | 0.00      | 0.9987|
| Cycle              | Damping Factor   | 0.96033   | 0.01877   | 51.16     | <.0001|
| Cycle              | Period           | 43.85977  | 10.58712  | 4.14      | <.0001|
| Cycle              | Error Variance   | 0.00761   | 0.0039904 | 1.91      | 0.0566|

The final results regarding the model indicate that the level and seasonal components are not statistically significant at the 10 percent level of statistical significance. The cycle length was found to be 43.86 periods (months), with a damping factor of 0.9603. Both these parameters were found to be statistically significant at the 1 percent level of statistical significance.

The fact that this model is sensitive to the influence of the noise between seasons can be responsible for this result. These problems might however be overcome by testing for seasonal patterns over longer periods (Venter, 2007).

\(^2\) Making use of a higher order cycle produces a smoother extracted cycle (Harvey and Trimbur, 2003).
5.3 Final Results – White Maize

The results do however correlate well with the actual series, despite the fact that it was not statistically significant. These results might be useful for 2 purposes. The first is that it gives an indication of the underlying seasonal patterns in the data, and the second is that these patterns might be used as a method of identifying pairs. Graph 5.1 illustrates the relationship between the actual monthly average spot price series and the hidden seasonal pattern for the white maize data.

The hidden seasonal pattern extracted for the white maize series does seem to be useful in signalling most of the changes in direction for the actual series. Although this is not always the case, the hidden seasonal pattern seems to be a more accurate representation of the behaviour of the white maize series over the test period when compared to other deterministic methods such as GARCH. Because the UCM series is more consistent over time, it can be used with more confidence than the deterministic type models.

6 YELLOW MAIZE

6.1 The Model

The final model used in testing for unobserved seasonal patterns in the yellow maize data takes the form of:

\[ y_t = \mu_t + \phi_t + \gamma_t + \epsilon_t. \]  

(6.1)

where \( \mu_t \) is the trend, \( \phi_t \) the cyclical component, \( \gamma_t \) the seasonal component, and \( \epsilon_t \) the random component. The slope and the random components were once again statistically insignificant at the 10 percent level of statistical significance with p-values of (0.4300) and (0.9992) respectively, and were therefore excluded from the final model. For the final model, the trend component thus takes the form of a random walk with noise (identical to the model used for white maize). The model for \( \mu_t \) is thus given by:
\[ \mu_t = \mu_{t-1} + \eta_t, \quad \eta_t \sim \text{i.i.d. } N(0, \sigma_\eta^2), \]  
(6.2)

where the disturbance term \( \eta_t \) is assumed to be independent. The trend component \( \mu_t \) is a random walk, modelled as a time-varying level component without any persistent upward or downward drift.

The smooth cyclical component option in SAS\textsuperscript{®} was again used to model for hidden seasonality in the yellow maize data. The smooth cycle specification used in the final model will thus take the form of \( \psi_t = \psi_{t}^{(i)} \) where:

\[
\begin{bmatrix}
\psi_{t}^{(i)} \\
\psi_{t}^{*(i)}
\end{bmatrix} = \rho \begin{bmatrix}
\cos \lambda_t & \sin \lambda_t \\
-\sin \lambda_t & \cos \lambda_t
\end{bmatrix} \begin{bmatrix}
\psi_{t-1}^{(i)} \\
\psi_{t-1}^{*(i)}
\end{bmatrix} + \begin{bmatrix}
\psi_{t}^{(i-1)} \\
\psi_{t}^{*(i-1)}
\end{bmatrix},
\]
(6.3)

for \( j = 1, \ldots, k, t = 1, \ldots, n \), and where \( \psi_t = \psi_{t}^{(0)} \) and \( \psi_{t}^{*} = \psi_{t}^{*(0)} \) are mutually uncorrelated white noise disturbances with zero means and common variance \( s^2_{\psi} \). Since \( \rho < 1 \), \( \psi_t \) has a stationary distribution with zero mean and variance \( \sigma_{\psi}^2 / (1 - \rho^2) \).

Because the amount of observations used for yellow maize is the same as that used for white maize, the seasonal component was fixed and not allowed to be stochastic. The seasonal component will thus take the trigonometric form, as was the case for the tests performed for white maize. As before, \( \gamma_t \) is modelled as a sum of cycles of different frequencies, and is given as (Harvey & Scott, 1994:1328):

\[
\sum_{j=1}^{[t/2]} \gamma_{j,t},
\]
(6.4)

where \( s \) equals \((s - 1)/2\). The cycles \( \gamma_{j,t} \) have frequencies \( \lambda_j = 2\pi j / s \) and are specified by the matrix equation (SAS, 2006:1633):

\[
\begin{bmatrix}
\gamma_{j,t} \\
\gamma_{j,t}^{*}
\end{bmatrix} = \begin{bmatrix}
\cos \lambda_j & \sin \lambda_j \\
-\sin \lambda_j & \cos \lambda_j
\end{bmatrix} \begin{bmatrix}
\gamma_{j,t-1} \\
\gamma_{j,t-1}^{*}
\end{bmatrix} + \begin{bmatrix}
\omega_{j,t} \\
\omega_{j,t}^{*}
\end{bmatrix},
\]
(6.5)

where the disturbances \( \omega_{j,t} \) and \( \omega_{j,t}^{*} \) are assumed to be independent. Where \( j \) is fixed, \( \omega_{j,t} \) and \( \omega_{j,t}^{*} \) \( \sim N(0, \sigma_{\omega}^2) \).

The decomposition of the seasonal effects into these harmonics \( \gamma_{j,t} \) is identical to its Fourier decomposition, and the sum of squares of the seasonal factors is equal to the sum of squares of the amplitudes of the harmonics.

### 6.2 Model Results

The results of the free parameters are set out in table 6.1 below. As was the case with the white maize model, the slope and random components were excluded from the model.

| Table 6.1: Final Estimates Of The Free Parameters For Yellow Maize |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| Component       | Parameter       | Estimate        | Std Error       | t - Value       | Pr > |t|       |
| Level           | Error Variance  | 0.00494         | 0.0009814       | 5.03            | <.0001         |
| Season          | Error Variance  | 3.79E-13        | 5.59E-10        | 0.00            | 0.9995         |
| Cycle           | Damping Factor  | 0.99876         | 0.008734        | 114.35          | <.0001         |
| Cycle           | Period          | 27.78888        | 1.08681         | 25.57           | <.0001         |
| Cycle           | Error Variance  | 0.00005024      | 0.00005537      | 0.91            | 0.3642         |
The final results regarding the model indicate that the level component is statistically significant at the 1 percent level of statistical significance. The seasonal component is not statistically significant. The cycle length was found to be 27.79 periods (months), with a damping factor of 0.9988. Both these parameters were found to be statistically significant at the 1 percent level of statistical significance.

6.3 Final Results – Yellow Maize

Even though the seasonal component was not found to be statistically significant, the results still correlate well with the actual series. Graph 6.1 illustrates the relationship between the actual monthly average spot price series and the hidden seasonal pattern for the yellow maize data.

Graph 6.1: Monthly Average Spot Series Of Yellow Maize Log Returns Prices Vs Hidden Seasonal

The unobserved seasonal pattern extracted for the yellow maize data seems to be useful in signalling the bulk of the changes in direction for the actual yellow maize series. Even though this is not always the case, the hidden seasonal pattern seems to be an accurate representation of the behaviour of the yellow maize series over the test period.

7. CONCLUSION

The UCM results are useful for the trader when implementing a pairs trading strategy. Not only is the information regarding these seasonal patterns in the data useful when compiling a pairs portfolio, it also allows the trader to make informed decisions about the future behaviour of these commodity prices.

In testing for seasonal patterns in the time series data two approaches were used. Tests for the existence of seasonal patterns were first performed by making use of GARCH models – these were not reported above. This deterministic approach did however not render satisfactory results. Even though the results were statistically significant, no meaningful relationship between the results and the actual time series were found.
Tests were also performed in the form of the stochastic UCM method. This approach tests for hidden or unobserved seasonality in time series data. Although no statistically significant results were reported, these patterns seem to describe the actual series more accurately than the deterministic approach. In some cases the unobserved seasonal patterns seems to be a leading indicator to the movements of the actual series. In similar research conducted on stock exchange data, the results were statistically significant.

These seasonal patterns in the agricultural commodities data are important for the trader that considers an agricultural commodity pairs portfolio. If the trader has an idea of the movement of the commodity prices in his pairs portfolio, it makes the decisions of when to change positions easier. Knowledge about the seasonal patterns in the data is not solely crucial in terms of the general trend of the commodity prices, but these patterns might also prove useful in terms of the pairs selection process. Hidden seasonal patterns can be useful to establish whether it is possible to select a commodity pair based on their co-movement over time.

REFERENCES