Determinants of Commodity Prices: An Investigation into Structural Break Points and Shifting Regimes

ABSTRACT
Financial returns, although seemingly unpredictable, have been successfully modelled using market and macro-economic factors. The effect of these factors, however, may vary with time and only be statistically significant for specific periods. This paper utilises factors that have been found to capture financial market returns and applies them to the modelling of a broad set of commodities. A change point recognition algorithm is adopted to account for shifting regimes within the commodity return distributions. Specific regimes within each commodity are identified and analysed allowing us to determine the factors that are driving commodity prices within those regimes.

Written and Compiled by Ian Bell* under the supervision of Dr Robert Durand and Dr Alexander Szimayer

August 2004
(Working Paper)

*Corresponding Author: UWA Business School. Crawley, Western Australia. Phone-(08) 94473920, Email-belly82@iinet.net.au
1. INTRODUCTION

The aim of the project is to widen the understanding of the market fundamentals that affect commodity price levels. The factors that drive commodity price returns are likely to differ through time as a result of changes in technology, market structures, consumer preferences, policies and institutions. When assessing the behaviour of commodity prices you are looking at a product that is expressed in a certain currency. For this reason you would expect that commodity prices will be affected in some way by currency exchange rates.

Extensive research has been conducted into the determinants of Commodity prices. The majority of this literature has however neglected the effects of exchange rates and has used US specific macroeconomic data to determine price behaviour. This paper aims to extend the current literature to a more global context by analysing the effects of exchange rates on commodity prices and using broader composite data from the member countries of the Organisation for Economic Co-operation and Development (OECD) and the Group of Seven Nations (G7).

Previous literature has identified non stationarity and structural breaks within commodity price distributions. Structural breaks were initially seen merely as a hindrance to the power and accuracy of measures of determining stationarity and not as a matter of concern in themselves. Cuddington and Urzua (1987, 1989) use the Dickey-Fuller (1981) test to provide evidence of non-stationarity in commodity returns but ignore structural breaks.

This acceptance of the non-stationarity hypothesis was initially accommodated with ad-hoc and crude statistical techniques. The methods used did not determine change points but incorporated structural breaks at points of time that they were known to have occurred due to some historical fact or information. Powell (1991) examines return distributions over a 100 year period and presents a piecewise function with permanent drops in 1921, 1938 and 1975 that reflect breaks in the distribution. Sapsford (1985) introduced dummy variables in an attempt to account for structural breaks.

Following the literature documenting the non-stationarity of commodity prices and indeed other financial assets and instruments we began to see advancement in the econometrics of structural change. Recent literature has delivered strong evidence of single or multiple shifts of conditional means in financial time series. These results followed on from the insurgence of literature that focused on the theory of identifying structural break points at unknown locations. It is this literature that forms the basis of the tests we use for identifying structural break points in our commodity price distributions.

Despite the prevalence of literature emerging on structural break points the majority of techniques used have looked merely at drifts in the conditional mean and neglect the incidence of structural breaks in moments of higher order. The second order moment in particular is of concern given the

---

2 Unit root tests such as the Dickey-Fuller (1981) test and other typical non-parametric tests for stationarity such as those proposed by Cochrane (1988, 1991) and Lo & McKinley (1989) are extremely susceptible to structural breaks.
3 See Bai (1997) and Hansen (2001)
4 See Csorgo and Horvath (1997) for an overview of the techniques used and further references to their earlier development.
recent focus on financial models that relax the assumption of constant variances. Structural breaks in conditional mean and variance have generally in the past been examined in isolation.

The technique we propose isolates multiple unknown change points by examining model residuals and minimising the sum of squared errors. The pattern recognition algorithm adopted will determine whether structural breaks have occurred in the sample, and if so, estimate the time of their occurrence. The breaks in the distribution will be used to identify apparent shifts of regime\(^5\) within the commodity price model. The Bayesian Information Criterion of Schwartz (1978) will then be used to evaluate and derive the optimal model.

By extending the current literature into a global context and relaxing the assumption of a continuous and stationary distribution we can gain a more accurate view of the endogenous and exogenous factors that affect the value of commodities. Further examination of structural breaks in the form of large shocks to prices will also be undertaken with a view to determine if they are driven by news or sentiment.

This understanding is essential in assessing the financial risk that is inherent in holding a commodities position. A more detailed understanding of the origins and dynamics of the risks a commodities position exposes you to will introduce possibilities of more sophisticated hedging techniques that may not purely rely on plain vanilla futures and options positions.

2. PREVIOUS RESEARCH

There has been a great deal of research focused on explaining the apparent erratic behaviour of commodity price movements. This research has resulted in many differing inferences being drawn on the characteristics of commodity price movement. Due to the contradictory nature of many of the theories there has been no single common theory to emerge from the literature on commodity price determinants.

Fama and French (1987, 1988) subscribe to the view that spot commodity prices follow a cyclical mean reverting process and are determined by the interest foregone in storing a commodity, the warehousing costs of storing a commodity and the lagged convenience yield associated with holding a commodity. Schwartz (1997) extended this theory to include the effects of stochastic convenience yield and stochastic interest rates and incorporated them into financial models focusing on option pricing and hedging.

The theory of risk premia\(^6\) suggests that future spot prices can be forecasted by deducing an expected risk premium that an investor should be rewarded with for holding a commodity. This theory is similar to the Arbitrage Pricing Theory (APT) of Ross (1976) that is used to explain equity returns. A risk premium is factored into commodity prices due to the possibility of changes to the dynamics of supply and demand. A number of macroeconomic factors such as GDP growth and inflation as well as factors that are specific to the commodity itself are used to explain these changes in dynamics.

Prebisch (1950) and Singer (1950) through the extrapolation of trends from 1870 -1945 predicted that commodity prices would in future decline by approximately 1% per annum in real terms. The Prebisch-Singer hypothesis (PSH) stipulated that the apparent downward trending nature of

\(^5\) A regime is defined as a sub-series between two structural breaks where the data appears to follow a particular distribution

\(^6\) see Breeden (1980) and Hazuka (1984),
commodity prices is due to a weakening in the terms of trade of developing countries relative to industrialised countries.

The reactions to the theoretical and empirical arguments of the PSH have been widely varied. Leon and Soto (1995) analysed the price of 24 commodities in the period of 1900-1992. Their model tests for non-stationarity in the series by using a technique that allows structural breaks to be determined exogenously. They found that 15 of the commodities exhibited a negative trend, six were trend-less and 3 were positively trended. Sapsford & Singer (1998) and Borensztein et al. (1994) have also found similar results for the period following the analysis of Prebisch and Singer.

Spraos (1980) was the first person to refute the PSH suggesting that the negative trend exhibited by commodities ceased after World War II. Cuddington & Urzua (1989) using stochastic trend models refuted the PSH suggesting that 21 of 24 widely traded commodities exhibited an upward trend or no trend at all in the period following the formulation of the PSH.

Deaton and Laroque (1992) found that commodity price cycles are dominated by long periods where the price deviates below any long-term trend and by sharp upward spikes that break through these long term trends. Cashin, McDermott & Scott. (1999) stated that shocks or sharp erratic movements in commodity prices tend to persist for several years. They added to the research again in 2001 by finding that the probability of the end of a slump (boom) is independent of how long the commodity has been in that slump (boom). This cyclical behaviour and changing of trends is indicative of the regime shifts that are believed to be evident in commodity prices.

Perhaps of most relevance to this study are the findings of Pyndick & Rotemberg (1990) who suggest that price movements exhibit large correlations between individual commodities. They suggest that this herd-like behaviour characteristic of commodity prices gives the appearance that traders have a tendency to be alternatively bullish or bearish on commodities as a whole for no plausible economic reasons. This type of behaviour suggests that there is some external factor(s) driving the prices of commodities as a whole. The behaviour is indicative of large scale macro-economic factors having an effect on commodity prices.

Pyndick and Rotemberg (1990) found that despite controlling for known macro-economic factors commodities still exhibit herd-like behaviour. They suggest three possible explanations to explain this behaviour. i) The macro-economic variables are not completely exogenous; ii) There are liquidity constraints, e.g. a fall in price of one commodity lowers the price of others because it diminishes the available finances of speculators who are long in several commodities at once; iii) There are relevant macro-economic variables missing from the model.

Previous research into structural breaks has been more common in the area of two-state regime shift analysis in stock markets and foreign exchange. This is more than likely due to the lower price volatility exhibited in these markets causing them to be far easier to examine. This was a notion that was confirmed in our analysis. Through the use of macro-economic factors we were able to capture some of that volatility making the analysis more plausible. Our research in commodity markets extends the definition of regimes from the typically used two-state classification of bull or bear markets. Returns have previously been modelled by classifying data and modelling transition probabilities with a dual state Markovian transition matrix. Our technique does not limit the classification of regimes into only one of an upward or downward trending state.

---

7 See for example Chiang (2003) and Bollen et al. (2000).
3. METHODOLOGY

The methodology employed in the project is comprised of five parts

1) Analysis of time series plots
2) Multivariate regressions
3) Implementation of pattern recognition algorithm
4) Implementation of Bayesian Information Criterion (BIC)
5) Assessment of overall model fit and possible explanations for behaviour

3.1 Analysis of time series plots

The analysis begins with the running of time series regressions of returns of a broad range of commodities. Plots are then examined to assess if the commodities follow the generally accepted characteristics of mean reversion and cyclical behaviour. The cumulative sum (cumsum) of returns is also plotted and analysed in an effort to determine if evidence of regime shifts appear in the data.

3.2 Multivariate regressions

A multi-variate regression is then fitted to the data assigning the commodity returns as the variable dependant on multiple explanatory variables. Different macro-economic factors are used in different combinations to determine which set of variables have the greatest explanatory power.³

Because we are also looking at regime shifts it is important to determine if some factors are only relevant at certain periods of time. A factor can be found to be insignificant over the entire sample period and be discarded despite having considerable explanatory power during one (or more) of the as yet unidentified regimes. In order to avoid making this mistake the model residuals are plotted and visually analysed to estimate periods of time where regime shifts seem to have occurred. Cross-sectional regressions are then performed in these sub-periods in order to identify all factors that have had explanatory power over commodity prices whether or not this factor has persisted or has always had explanatory power. A trade-off is then made between a loss of power due to over-fitting of the model and miss-specification.⁹

*The following classical regression equation is employed to determine the macroeconomic factors that have explanatory power*

\[ y_t = x_t^\top \hat{\beta}_t + \sigma \varepsilon_t \quad t = 1,2,...........,T \]  

Where: \( y_t \) is the observed commodity returns  
\( \hat{\beta}_t \) is a vector of the \( n + 1 \) coefficients

³ See section 4 for an explanation of our rationale in the choices of the different macro-economic factors featuring in the initial short-list.
⁹ This problem will be addressed in section 3.4.
\(x_t\) is a vector of the \(n\) identified macroeconomic factor returns
\(\sigma_t\) is the constant variance
\(\varepsilon_t\) is an i.i.d. normally distributed variable with mean 0 and variance 1

### 3.3 Implementation of Pattern Recognition Algorithm

Once the appropriate factors have been determined a pattern recognition algorithm based on the early works of Csorgo and Horvath (1997) is employed to identify possible change-points in the data distribution. The sum of squared errors (SSE) of the model is determined in order to estimate locations of structural breaks in the employed models and their time of occurrence. This technique provides us with estimations of the most likely occurrences of the \(m\) different change-points. Section 3.4 will explain the decision process behind the choice of the number of change-points that will be used in the final model.

**Using equation (1) to model commodity returns for each sub-period and assuming that the series contains multiple breakpoints (\(m\)) the sum of squared errors is determined by equation (2)**

\[
S_m(k) = \sum_{t=1}^{k_1} (y_t - x'_t, \beta_t(k))^2 + \ldots \sum_{t=k_{m+1}}^{T} (y_t - x'_t, \beta_t(m+1)(k))^2
\]

Where: \(k\) is the location of all possible break points

The \(m\) change-points are then identified by minimising (2) using equation (3)

\[
k_{1\ldots m} = \arg \min_{1 < k_1 < \ldots < k_m < T} S_m(k)
\]

### 3.4 Implementation of Bayesian Information Criterion (BIC)

The exact number of change points is unknown so the Bayesian Information Criterion (BIC) of Schwartz (1978) is used to determine the appropriate number of change points in a trade-off between maximum explanatory power and minimum model complexity.

**The following function when minimised will provide the optimal number of change points**

\[
BIC(l) = -2 \ln(L) + .5 N \ln(n) l
\]

Where: \(l\) is the number of change points
\(L\) is the maximised log likelihood
\(N\) is the degrees of freedom of the model
\(n\) is the number of observations

### 3.5 Assessment of Model fit and possible explanations for behaviour

Once the regime shifts have been identified and the models have been fitted to the different commodities. Market conditions, trends and news are then analysed to determine if shifts in
regime pertain to a particular event(s) which may have caused investors to change the way they price commodities. Interest will be shown as to whether different commodities experience shifts in regimes at similar times and whether these shifts as a whole can be explained by news or sentiment.

4. DATA

All commodity price returns have been collected from Thomson Learning’s DataStream database and all Macroeconomic variable data has been collected from both DataStream and the Federal Reserve website (www.federalreserve.gov). The data-set runs for a 24 year period from February of 1980 through to January of 2004. The Commodities analysed were chosen to reflect the large number of commodities that are traded on financial markets. There is a mix of agricultural products, industrial inputs and precious metals. The choice of commodities was somewhat constrained by the requirements for the length of data as well as overall data quality.

The eight commodities are:

- i) Copper LME Grade A
- ii) Gold LME Bullion
- iii) Wheat English Feed (export)
- iv) Coffee Brazilian (NY)
- v) Lead LME
- vi) Platinum London Free Market
- vii) Silver Fix LBM
- viii) Commodity Index The Economist

The macro-economic data collected is designed to be a broad representation of all the possible factors or proxies for factors that could affect the supply or demand for a particular commodity. It is very unlikely that all the factors that data has been collected on will be included in the final model and it is probable that only a handful of the factors will be used. The final factors will be chosen based on tests of significance in the overall sample and also in the identified regimes.

The factors short-listed have all been used in previous studies to model financial returns. It is apparent though that the data used has always been specific to the US. Our inclusion of OECD and G7 for factors such as CPI, GDP, unemployment and Industrial Production results in a model that is more reflective of a global market. See Table 1 for a list of the macroeconomic factors that have been considered for the model and our reasons for their inclusion.
Table 1

<table>
<thead>
<tr>
<th>Factor</th>
<th>Source</th>
<th>Justification</th>
</tr>
</thead>
<tbody>
<tr>
<td>OECD CPI</td>
<td>DataStream</td>
<td>Reflects changes in Supply and Demand</td>
</tr>
<tr>
<td>OECD GDP</td>
<td>DataStream</td>
<td>Reflects changes in Supply and Demand</td>
</tr>
<tr>
<td></td>
<td>(OCDMPPBDP)</td>
<td>See Chen (1991)</td>
</tr>
<tr>
<td>AAA-Baa Spread</td>
<td>FederalReserve.gov</td>
<td>Proxy for health of the economy</td>
</tr>
<tr>
<td></td>
<td></td>
<td>See Chen et al. (1986), Schwert (1990), Chen (1991)</td>
</tr>
<tr>
<td>US 3mth T-Bill</td>
<td>FederalReserve.gov</td>
<td>Affect required ROR on storage and capital investment</td>
</tr>
<tr>
<td>US 10yr-3mth Spread</td>
<td>FederalReserve.gov</td>
<td>Reflects expectations of future aggregate economic activity</td>
</tr>
<tr>
<td>G7 Unemployment Rate</td>
<td>DataStream</td>
<td>Affects demand via changes in income &amp; supply via access to workforce</td>
</tr>
<tr>
<td></td>
<td>(G7OUN014Q)</td>
<td>See Pindyck &amp; Rotenburg (1990)</td>
</tr>
<tr>
<td>OECD Industrial Production</td>
<td>DataStream</td>
<td>Affects demand for industrial commodities and subsequently non industrial commodities via increases in income</td>
</tr>
<tr>
<td></td>
<td>(OCFPR035G)</td>
<td>See Chen et al. (1986)</td>
</tr>
<tr>
<td>S&amp;P 500 Index</td>
<td>DataStream</td>
<td>Proxy for health of the economy</td>
</tr>
<tr>
<td></td>
<td>(S&amp;PCOMP(PI))</td>
<td>See Pyndick &amp; Rotenburg (1990), Borensztein et al. (1994)</td>
</tr>
<tr>
<td>US/YEN Exchange Rate</td>
<td>FederalReserve.gov</td>
<td>Affect supply &amp; demand for commodities by influencing relative prices in domestic currencies</td>
</tr>
<tr>
<td>US/BP Exchange Rate</td>
<td>FederalReserve.gov</td>
<td></td>
</tr>
</tbody>
</table>

5. DATA ANALYSIS AND PRELIMINARY RESULTS

Initial manipulation of the data into appropriate forms was undertaken using Microsoft Excel. The Change Point locations as well as all outputs including descriptive statistics, factor coefficients, t-statistics, factor correlations and the BIC results are computed using the matrix algebra and vector calculus capabilities of MatLab. Due to computer power constraints the algorithm can only be run with a maximum of three change points for any data-set.

The optimal number of change points in the sample period for each commodity is recorded as well as their location. The BIC selection criterion statistics are also included to show how we came about the optimum number of change points. Once this has been done the commodity regimes are analysed independently and the coefficients and R² figures are reported for each individual regime period.
The Optimal model identifies one or more change-points in all but one of the examined commodities. Copper, Gold, Wheat, Coffee and the Economist Commodity index are all found to have 1 change point in the 24 year return distribution corresponding to two different regimes. 2 Change points are identified for both platinum and silver while lead is deemed to have a stationary distribution for the entire sample period.

Five of the original short-listed macro-economic and market factors are found to have considerable explanatory power over commodity returns in particular regimes. The US/EU exchange rate and the S&P 500 appear to do the best job in capturing commodity returns whilst the US 3mth Treasury bill rate, the OECD consumer price index and the US/Yen exchange are also found to have considerable explanatory power over certain commodity returns in particular periods.

Instances occur where a factor can be highly significant in one regime but have no explanatory power in an adjacent regime. This behaviour type is found in both Copper and Gold returns where the US/EU exchange rate is found to be highly significant in one of the two identified regimes but has no explanatory power in the other. Further counterintuitive results such as the reversing of the sign of the US/Yen exchange rate coefficient in the copper regressions whilst still remaining statistically significant have also appeared. Such apparent anomalous behaviour will be examined to determine if it can be explained any economic theory. See table 2 for a summary of the preliminary results and Figures 1-8 for the MatLab visual output highlighting individual commodity price distributions and the locations of identified change-points.

### Table 2

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Copper</td>
<td>1</td>
<td>1</td>
<td>2/80-9/85</td>
<td>0.00526</td>
<td>-1.667</td>
<td>-0.0001</td>
<td>1.08**</td>
<td>0.475**</td>
<td>0.383*</td>
<td>0.329</td>
<td>763.631</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>10/85-1/04</td>
<td>0.00625</td>
<td>-1.333</td>
<td>0.00054</td>
<td>0.0615</td>
<td>-26**</td>
<td>-0.333**</td>
<td>0.037</td>
<td></td>
</tr>
<tr>
<td>Gold</td>
<td>1</td>
<td>1</td>
<td>2/80-01/83</td>
<td>0.0233</td>
<td>-5.34*</td>
<td>0.000616</td>
<td>0.0892</td>
<td>-353</td>
<td>-0.227</td>
<td>0.249</td>
<td>908.458</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>2/83-1/04</td>
<td>0.00968</td>
<td>0.908</td>
<td>-0.0025**</td>
<td>0.186**</td>
<td>-0.206**</td>
<td>-0.116</td>
<td>0.153</td>
<td></td>
</tr>
<tr>
<td>Wheat</td>
<td>1</td>
<td>1</td>
<td>12/87-1/04</td>
<td>0.0249</td>
<td>4.779**</td>
<td>-0.0067**</td>
<td>0.875**</td>
<td>-0.113</td>
<td>0.0338</td>
<td>0.235</td>
<td>642.663</td>
</tr>
<tr>
<td>Coffee</td>
<td>1</td>
<td>1</td>
<td>2/80-9/94</td>
<td>0.0143</td>
<td>1.447</td>
<td>-0.003</td>
<td>0.701**</td>
<td>-0.153</td>
<td>-0.083</td>
<td>0.073</td>
<td>674.584</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>10/94-1/04</td>
<td>0.0368</td>
<td>3.588</td>
<td>-0.014**</td>
<td>-0.022</td>
<td>0.257</td>
<td>-0.031</td>
<td>0.016</td>
<td></td>
</tr>
<tr>
<td>Lead</td>
<td>0</td>
<td>1</td>
<td>2/80-1/04</td>
<td>0.0186**</td>
<td>-1.423</td>
<td>-0.002</td>
<td>0.563**</td>
<td>0.0314</td>
<td>0.325**</td>
<td>0.058</td>
<td>786.106</td>
</tr>
<tr>
<td>Platinum</td>
<td>2</td>
<td>2</td>
<td>8/80-10/87</td>
<td>0.00639</td>
<td>4.098</td>
<td>-0.0048</td>
<td>0.689*</td>
<td>0.552**</td>
<td>-0.077</td>
<td>0.157</td>
<td>771.947</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>11/87-1/04</td>
<td>0.0175*</td>
<td>-1.527</td>
<td>-0.0018</td>
<td>-0.16</td>
<td>-0.194**</td>
<td>-0.308**</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td>Silver</td>
<td>2</td>
<td>2</td>
<td>8/80-1/83</td>
<td>0.219*</td>
<td>6.811</td>
<td>-0.022**</td>
<td>1.434</td>
<td>0.509</td>
<td>0.679</td>
<td>0.335</td>
<td>740.104</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>2/83-1/04</td>
<td>0.0219*</td>
<td>1.794</td>
<td>-0.006</td>
<td>0.0407</td>
<td>-0.084</td>
<td>-0.036</td>
<td>0.023</td>
<td></td>
</tr>
<tr>
<td>Economist</td>
<td>1</td>
<td>1</td>
<td>2/80-3/88</td>
<td>0.0212</td>
<td>2.703*</td>
<td>-0.0045**</td>
<td>0.0837</td>
<td>0.0148</td>
<td>-0.095</td>
<td>0.153</td>
<td>100.794</td>
</tr>
<tr>
<td>Index</td>
<td></td>
<td>2</td>
<td>4/88-1/04</td>
<td>0.0147**</td>
<td>-1.69</td>
<td>-0.0017</td>
<td>-0.12</td>
<td>-0.131**</td>
<td>-0.096</td>
<td>0.075</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Significant at a 10% level
** Significant at a 5% level
6. EXPECTATIONS AND REMAINING WORK

It is apparent from the preliminary regressions and implementation of the change-point technique that structural breaks do exist in commodity price distributions. At this stage it appears that the identified change points are not occurring at the same time for the different commodities. It is perhaps rational to expect that if commodity prices move together due to influences from the same universe of external macroeconomic forces then the regime shifts experienced by different commodities will occur at similar times. This is of course dependant on the characteristics of the supply and demand structure of the individual commodities and how they are influenced by the conditions of the overall economy.

The uniqueness of the change points may represent distinct differences in the market dynamics of individual commodity prices. Thorough analysis of the timing of change points will be undertaken to determine if news or sentiment are responsible for shifts in regimes. Correlations of commodities will be examined in order to determine if those that are highly correlated react similarly to changes in the same factors and exhibit structural breaks at comparable points in time.

The further refinement of the model and the testing of it’s accuracy in picking up distributional breaks is still to be completed. The change point algorithm will be implemented on data that has been randomly generated with known change points in order to gauge the model’s accuracy. Sensitivity analysis will be required to determine how susceptible the model is to large volatility changes in the underlying distributions.

REFERENCES


vol. 96, pp. 893-920.


