Norwegian School of Economics

Bergen, Fall 2014

Fundamentals and Spot Return Volatility

An empirical study using SUR model with GLS estimation

Debashis Senapati

Tore Sannes

Supervisor: Terje Lensberg

Master Thesis within the profile of Energy, Natural Resources and the Environmet

NORWEGIAN SCHOOL OF ECONOMICS

Foreword

Working on the thesis has given us a unique possibility to get an insight in global commodity markets. The combination of reading relevant literature and collecting and analyzing data is something that we believe will come in handy in future work related issues.

In addition to this, working together as a team has also given us a lot.

We would like to thank our supervisor Terje Lensberg for great guidance and constructive conversations. We would also like to thank Øyvind Anti Nilsen and Michail Chronopoulos for their timely assistance on technical matter.

Abstract

Our thesis investigates whether fundamental factors- inventory and demand condition- are the main determinant of spot return volatility for 31 commodities in the period 2009-2013. We have followed the *theory of storage* approach and used the adjusted-spread between futures and spot prices for commodities to represent these fundamental factors. We develop a structural model to test the empirical relevance of adjusted spread along with volatility of nominal interest rate and movements in market liquidity on spot return variance. We have used Seemingly Unrelated Regression (SUR) model for panel data with Generalised Least Square (GLS) estimation technique. The adjusted-spread is found to be statistically significant and has positive effect on the spot return variance across the panel data for all the commodities. Our results suggest fundamental factors have an overwhelmingly large impact on the spot return variance as compared to other explanatory variables in our regression. Our results are consistent with both theory of storage and the existing literature related to this topic.

Contents

Contents

Figures

Tables

1. Introduction

Commodity prices are volatile and they change over time. “Most economists have traditionally argued that fundamental (supply & demand) factors determine volatility in commodity market. Others assert, however, that prices are driven by “animal spirits” and other random forces which induce volatility”. (Pirrong & NG 1994) In our thesis we investigate empirically the relevance of fundamental factors (supply & demand conditions), monetary policy and changes in market liquidity on spot return variance of 31 commodities from 2009 to 2013.

To understand how fundamental factors interact with the spot return volatility we have utilized certain implications from the “theory of storage”. The theory of storage implies that the inventory & demand conditions affect the variances of commodity spot prices and the spread between spot and futures prices (Working 1949). The spread between spot and futures price is observable on a daily basis unlike actual inventory positions and demand conditions. The close relationship between spread and fundamental factors has been established through past empirical research[1]. So the spread (adjusted for interest rate and storage cost) is our main explanatory variable representing the fundamental factors.

Recent studies into commodities price dynamics have suggested factors other than fundamentals factors to influence commodity prices. Frankel (2006) has suggested that there is a negative effect of interest rates on the desire to carry commodity inventories and thus lower real interest rates leads to higher commodity prices. Also Irwin and Sanders (2012) have attributed the rapid expansion of derivatives market and commodity index have expanded the market participation into commodities and may have decreased risk premiums, and hence, the cost of hedging thus reducing price volatility in commodity markets.

In our thesis we have collected daily data for spot & futures prices for 31 different commodities in the period from 2009 to 2013. The list of commodities included in our research consists of the “traditional” commodities classes from agriculture, metals and energy and the more recently available commodities from electricity and shipping.

[1] See Working (1949), Williams and Wright (1982), Pindyck (1990) and Brennan (1991)

We have diverged[2] from past research in this field in two significant ways. Firstly we have used the 1-month constant maturity futures prices[3] (generic futures contract) instead of active futures contract price data and market determined spot prices instead of interpolating from near futures contract. The price of active futures contract becomes more volatile as the maturity date for the futures contract approaches. Therefore the generic futures contract prices are more accurate and convenient for our research. Secondly, we have used Seemingly Unrelated Regression (SUR) model with Feasible Generalized Least Square (GLS) estimation techniques to understand commodities. We chose to use the SUR model for two reasons: 1) it was convenient and accurate to model correlation among commodities under this framework 2) The GLS estimation technique provides us with consistent estimates for our panel data structure. To the best of our knowledge, testing the theory of storage using SUR model with GLS estimation technique on a broad class (31) of commodities has not been attempted before.

Our model predicts the spot return variance using the adjusted spread variance as the main explanatory variable and the variance of 1- month USD Libor rate and change in open interest positions in futures market as control variables.

Our results for the adjusted spread variance are statistically significant and positive for the whole cross section of commodities. The movement in market liquidity represented by the changes in open interest is statistically significant and negative. However the results for interest rate variance are not significant. While the effects of market liquidity in commodities is significant, they however do not appear to be overwhelmingly large. Therefore we conclude that the fundamental factors (inventory and demand conditions) are the prime determinant of commodity spot return volatility from 2009 to 2013. Lastly we perform regression on two separate cross section of storable and non-storable (electricity & shipping) commodities to investigate whether the theory of storage is equally applicable to non-storable commodities. The results for the both these cross sections for the adjusted spread variance are statistically significant, positive and comparable in magnitude.

---

[2] (Fama & French 1987), (Pirrong & NG 1994) and (Pindyck 2004)

[3] A constant-maturity price series indicates, for each time *t*, an interpolated price reflecting a specific time-to-expiration that is constant over time.

Our thesis is organized as follows: Chapter 2 gives an overview of the different commodity classes studied. Chapter 3 presents theory relevant for our research. Chapter 4 presents the methodology and chapter 4 describes the data. The results are analysed in chapter 6. Finally, we present our conclusions and compare them with past research and empirical evidence in chapter 7.

2. Overview of Commodities

A commodity can be defined as a "*consumption asset whose scarcity, whether in the form of exhausting underground reserves or depleted stocks, has a major impact on the world and country-specific economic development*". (Geman 2005b)

2.1 Commodity Classes

The commodities investigated in our thesis have been classified into 4 groups viz. agricultural, metals, energy and non-storables. In the following sections we will briefly go through the general characteristics of each class of commodity and describe their expected price and volatility behavior.

2.1.1 Agriculture

The agricultural commodities that we have included in our data set are corn, soy bean, wheat, soybean oil, soybean meal, lean hogs, sugar, coffee and cotton.

An important characteristic of agricultural prices is that they are usually seasonal. This is because storage is generally expensive, and there is often a relatively short limit to how long you can store the product. Furthermore, the prices of agricultural commodities are highly weather dependent. Pre-harvest volatility is usually higher than the volatility during the harvest (when the size of the crop is known). Agricultural commodity price time series show considerable positive autocorrelation and cross correlation on each other. For example, the price of livestock products is influenced by the price of agricultural feed products, like soybean meal. (Hull 2012b)

2.1.2 Metals

In the category of metals we have included 9 commodities: gold, silver, platinum, palladium, copper, aluminum, zinc, nickel and lead.

Unlike agricultural commodities, the supply in this group is not affected by weather and seasons. They are extracted from the ground, are relatively cheap to store and there is no practical limit to how long you can store them. We have classified metals in two groups; consumption and investment assets- depending on their industrial usage.

Inventory ratios are frequently in use for metals, and they are considered important in forecasting short term volatility. Important determinants of metal prices are demand trends (e.g. increased economic activity in developing countries, especially China), discovery of new sources, changes in exploration and extraction methods, etc. In addition to autocorrelation and cross sectional dependence metal prices are associated with price spikes. (Hull 2012b)

2.1.3 Energy

Energy products are among the most liquid and actively traded commodities. The commodities included in our dataset for this class are crude oil (Brent & WTI), gasoline, heating oil, propanol and natural gas.

We have included the two most important bench marks for the crude oil price namely Brent (North Sea) and West Texas Intermediate (WTI). Crude oil is mostly refined into gasoline, heating oil and propanol. Natural gas is often found in association with crude oil. However the price volatility of natural gas is significantly higher than crude oil price volatility.

The crude oil market is integrated and supply tends to follow demand closely and inventories adjust for the differences between supply and demand. The *forward cover*[4] data gives us an indication of global inventory levels in term of days and have stabilized at a higher level in the aftermath of the financial crisis. In contrast to crude oil, we cannot talk about an integrated global gas market. This is mainly due to the transportation challenges, and we therefore see big regional price differences. Since a big share of the natural gas is used for residential and commercial heating, demand for gas is very weather dependent with “peaks” in the winter months (Broxson et al. 2006).

2.1.4 Non-Storable (Electricity & Shipping)

The most important aspect of this class of commodities is the inability to carry inventories forward from one time period to another. Hydropower can indeed be “carried” in the form of water in the reservoirs, but if we look at global electricity markets in general the possibility

[4] The forward cover is calculated dividing stocks at the end of a given period by the expected consumption in the following period. (Amic 2005)

for storing is limited. The spot prices for these commodities must react instantaneously to balance supply and demand making them highly volatile. In this class we have included the Norwegian and German (one month, three month and 12 month) base electricity contracts and the time charter price for the TC2 Handymax tanker route.

Recent development in electricity derivatives market, deregulations and elimination of governmental monopolies has created liquid markets that were non-existent a decade ago. The electricity markets are still highly fragmented and this constrains the supply side. Local supply is vulnerable to disruptions caused by unforeseen power plant shut downs. Regional supply and demand are matched, and possible excess power is sold to other areas. The export of electricity is limited by the transmission line capacity, network charges and energy losses. Because demand is highly dependent on the weather there are occasionally large movements in the spot price (spikes). The increased share of wind and solar power in the generation mix contributes to making supply more weather dependent.

The shipping market has many features in common with electricity markets: Price and volatility in spot and forward freight markets move together, and it is the forward market that “leads” the spot market. Jumps in demand for capacity cause volatile freight rates, as shipping capacity becomes scares and *spot* shipping prices rise quickly. This is very similar to what we observe in the electricity market. (Geman 2005b) Global trade is the prime demand driver for shipping services. Freight rates become volatile when full capacity has been reached and the supply of shipping capacity is quite inelastic.

2.1.5 Summary of Commodity Classes

Supply and demand clearly follows a different pattern for the different commodity classes. The behavior of agricultural products, natural gas and electricity prices can be highly attributed to seasons. Metals, crude oil and shipping, on the other hand, are more influenced by business cycles and there is a cross sectional dependence of volatility within each class.

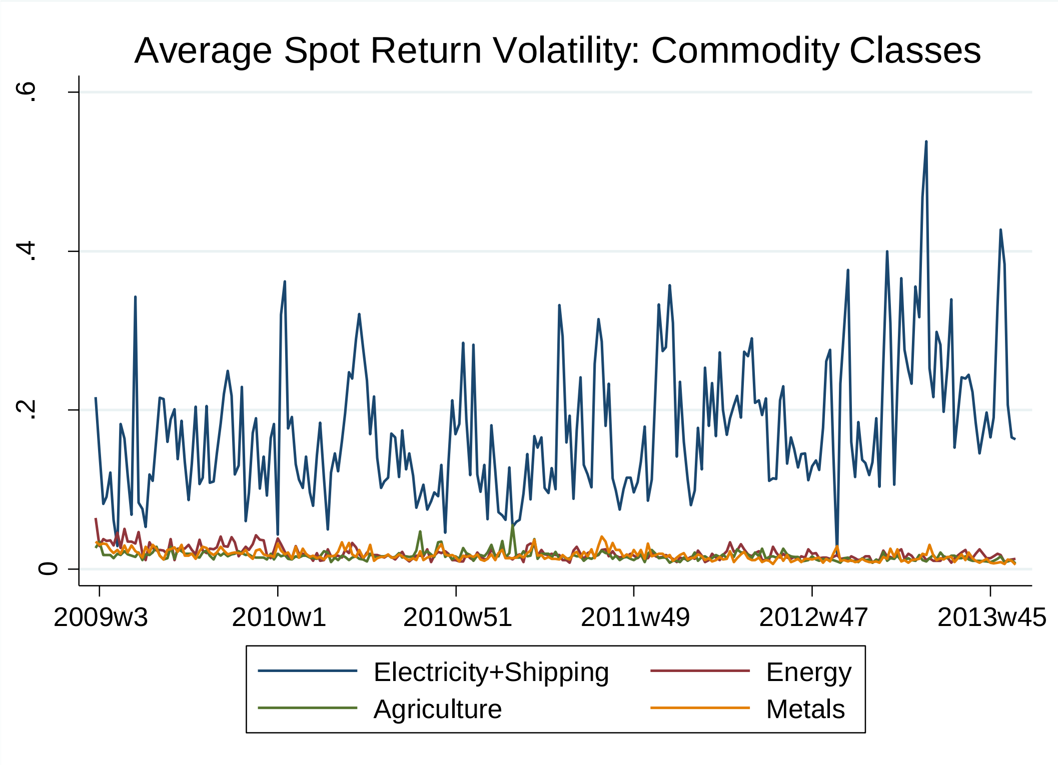

Figure 1: Average Spot Return Volatility in our data set.

The cost of storage is also highly distinct depending on the specific type of commodity: It is typically lowest for precious metals, low for metals, large for agricultural products and very large for animal products. The cost of storing energy products are also generally high, but together with minerals, these commodities can be strategically extracted from the ground whenever needed. (Carpantier & Dufays 2013) From Figure 1 we observe that the spot return volatility of non-storable commodities is substantially higher than the others, and volatility itself changes over time.

3. Theoretical Background

3.1 Relationship between Spot and Futures Prices

In our investigation of the spot return volatility, we employ the spread between spot and futures prices. Therefore, it is important to present the theoretical relationship between spot and futures and introduce the concept of *convenience yield*.

Forwards and futures contracts are heavily traded in most commodity markets in the world. The most liquid Futures contracts have short maturities (a few months), something that is applicable for most commodities. In commodity markets, unlike financial markets, the trading in the spot market is hindered by the transportation and quality requirements. Therefore trading in commodity spot market is generally dominated by supplier and real consumer of the physical asset, whereas the futures market has a substantial participation of market participants that don't have the intention to get involved in the physical delivery of the commodity.

"A commodity futures contract is an agreement to buy (or sell) a specified quantity of a commodity a specific date in the future, to a price agreed upon when entering into the contract" (Gorton & Rouwenhorst 2005). The price of entering into a futures contract is zero, because no transaction is done when the contract is agreed upon. The transaction takes place at maturity of the contract. Why would anyone trade Futures? To reduce risk, a farmer might agree to sell wheat to a broker in the future at a specific price agreed upon today. If the price goes up, the farmer loses money, but if the price goes down he is protected from losses. Holders of a futures contract will benefit when the future spot price at time T (S_T) turns out to be higher than the Futures contract with maturity at time T ($F_{0,T}$), and lose when $S_T < F_{0,T}$.

Commodities can be divided into investment assets and consumption assets. Typical examples of investment assets are gold and silver. Being an investment asset doesn't necessarily mean that it can't be consumed. For example, silver is being used as industrial input. What is required is that some individuals hold the asset for investment purposes, and that these individuals are prepared to sell their holdings and go long on futures contracts, if the latter looks more attractive (Hull 2012a).

3.1.1 No-arbitrage relationship and Convenience Yield

Investment assets can provide income to the holder, but as other commodities, they also have storage costs (insurance, warehousing expenses, maintenance, etc.). Storage costs and the risk-free interest rate represent what is called the *cost of carry*. The futures price, $F_{0,T}$, for a commodity is given by:

$$F_{0,T} = (S_0 + U)e^{rT} \qquad (1)$$

Where S_0 is the current spot price, U is the present value of all storage costs, r is the risk-free interest rate and T is the length of the contract. If we are in a situation where,

$$F_{0,T} > (S_0 + U)e^{rT} \qquad (2)$$

an arbitrager can make money by borrowing an amount equal to $S_0 + U$, and use this to buy one unit of the commodity and pay storage costs. At the same time the arbitrager can short a futures contract on one unit of the commodity. This riskless operation will provide a profit of $F_{0,T} - (S_0 + U)e^{rT}$. As more people take advantage of the arbitrage opportunity, the tendency will be that F_0 decreases and S_0 increases, until equation (2) is not true anymore.

When

$$F_{0,T} < (S_0 + U)e^{rT} \qquad (3)$$

selling one unit of the commodity, invest this money at the risk-free interest rate r, and take a long position in a futures contract will lead to a profit of $(S_0 + U)e^{rT} - F_{0,T}$. For the same reason as the previous example, equation (3) cannot hold for a long time, and we end up in the long term steady state given by equation (1).

The argument above doesn't hold for a *consumption asset*, because holders of a consumption asset normally plan to use it in some way. They are reluctant to sell the commodity in the spot market and buy futures contracts, because actual ownership of a physical asset makes it possible for manufacturers to use the commodity as an input in the production process at any time. It is also beneficial for the owner to be in possession of a consumption asset during periods of temporary local shortages caused by unexpected rise in demand. Therefore equation (3) might also hold, giving us this relationship for consumption assets:

$$F_{0,T} \leq (S_0 + U)e^{rT} \tag{4}$$

The benefit of holding the commodity itself is called the *convenience yield* (c), and was introduced by Kaldor (1939). This can be included in our equation:

$$F_{0,T}e^{cT} = (S_0 + U)e^{rT} \tag{5}$$

The convenience yield is a reflection of the expectations in the market about future availability of the commodity, and it increases together with the probability of a shortage. In periods of high inventories in the market, the probability of a shortage in the near future is low. The stock holder is more able to respond flexibly to unexpected excess demand or supply disruptions. The opposite happens when inventories are low.

3.2 The Theory of Storage

The importance of storage on commodity price movements was first introduced theoretically by Working (1933). He initialized the development of the *theory of storage*, looking at the relationship between inventory levels and the price behavior of wheat. Over the years research has also been done for other commodity classes. Here are some important implications of the theory:

- There is a tendency that the volatility is inversely related to the level of inventories. During periods of low inventories, the spot price of the commodity increases dramatically because there is no buffer to smooth supply (see Figure 2). (Geman 2005a)

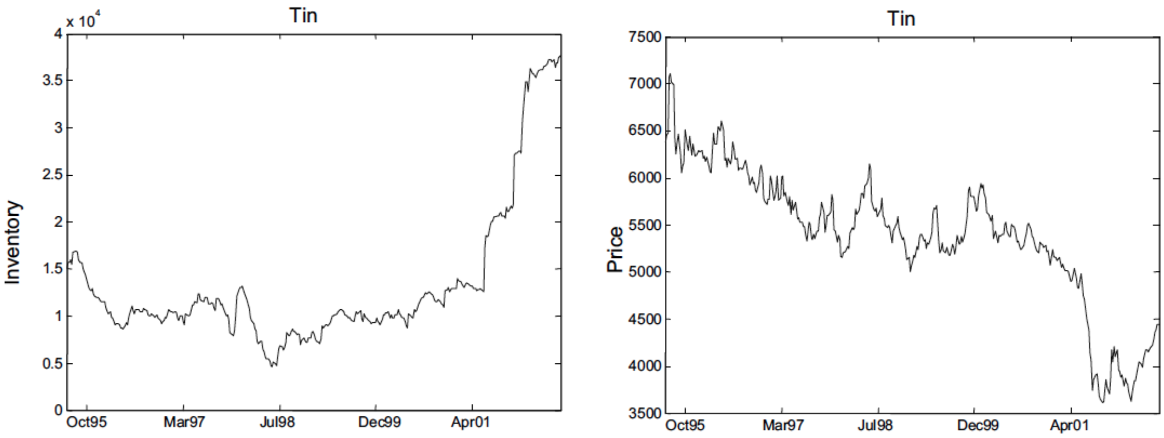

Figure 2: Inventories (in tons) and spot prices (in $ per ton). Source: (Geman 2005a)

- We have a positive correlation between the price and its volatility, because they are both negatively correlated to the level of inventories. This means that higher stocks lead to decreased volatility. (Geman 2005a)
- Consistent with the Samuelson effect (1965), the spot price is more volatile than the Futures price when inventories are low. This is because supply is less elastic in the short run. In the long run, adjustments in production are likely to take place, and Futures contracts are priced accordingly. Market participants know that in the long run, a high spot price will lead to a rebuilding of inventories. When inventories are high, spot and forward prices become equally volatile. (Geman 2005a)

It's a major challenge, if not impossible, to obtain information about inventory levels on a day to day basis for the 31 commodities.

The theory of storage implies that foregone interest rate, storage costs and convenience yield can be captured by the difference between futures and spot commodity prices (referred to as the *basis* or the *spread*). The spread is frequently being used as a proxy[5] for the level of inventories:

$$F_{t,T} - S_t = W_{t,T} + r_{t,T}S_t - C_t(I_t) \qquad (6)$$

Where S_t is the spot price and $F_{t,T}$ is the Futures price at time t with maturity at time T. Further, W_{T-t} represents the cost of storing the commodity from t to T, $r_{t,T}$ the risk free interest rate during the same period, C_t the convenience yield and I_t the state of inventories. When *the spread* is negative ($F_{t,T} < S_t$), the convenience yield exceeds the sum of the interest rate and storage costs, and the market is said to be in *backwardation*. When *the spread* is positive ($F_{t,T} > S_t$), the market is in *contango*. (Carpantier & Dufays 2013)

Equation (7) shows the continuous spot-forward relationship:

$$F_{t,T} = S_t e^{(r+w-c)(T-t)} \qquad (7)$$

[5] This relationship is extremely well documented empirically. See Pirrong and NG (1994) p. 213.

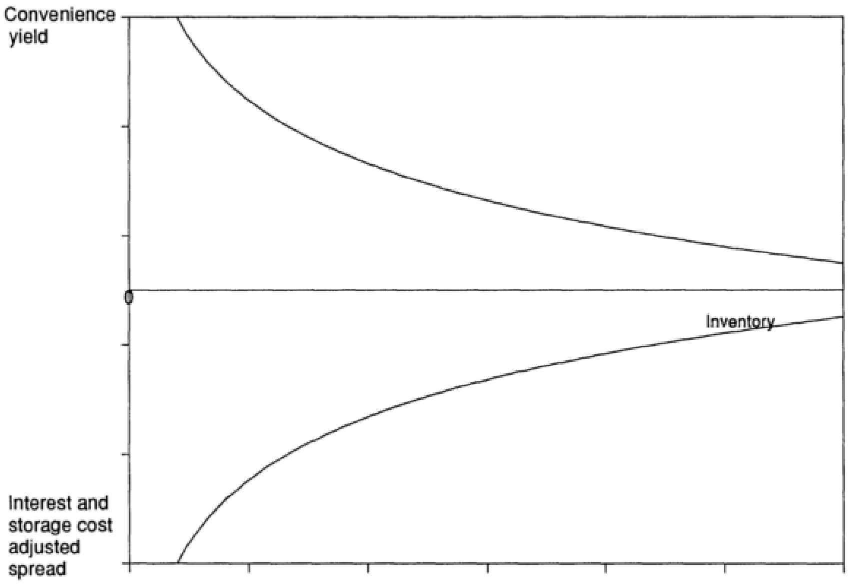

Figure 3: The relationship between inventories, convenience yield, and the interest and storage cost adjusted spread[6] *between futures and spot prices. Source: (Pirrong & NG 1994)*

The *spread* contains information about short term supply and demand. Regardless of whether there is a shortage of inventories or a sudden increase in demand, or both, the spread expands. According to the theory of storage, *supply and demand fundamentals are the main determinants of the volatility of commodity prices*. This hypothesis can therefore be tested empirically by using the spread, and testing whether or not the movements of this variable can explain much of the movements in the commodity price.

The development and emergence of electricity and shipping markets is quite new. The existing literature on this topic is divided regarding the applicability of convenience yield to non-storable commodities. Routledge et al. (2001) have argued that the theory of storage models can be extended to include goods which are not directly storable. They show that most intriguing empirical features of electricity prices follow naturally from the underlying economics of supply and demand. However Geman (2005c) claims that the convenience

[6] The adjusted spread equals the annualized percentage difference between the forward and spot prices at $t = 0$, net of storage costs (c) and interest costs (r) to hold inventory from t to T (Pirrong & NG 1994).

yield cannot be extended to non-storable commodities in its original definition and broader interpretation for convenience yield is required if it is to be extended to this class of commodities. Lautier (2009) has suggested that for electricity the so called committed generation units[7], kept as a reserve, have a role in power markets which is comparable to that of inventories.

3.2.1 Literature review on storage

The literature on the theory of storage is extensive. We have presented the most relevant ones below.

In the ,already mentioned, paper by Working (1933) it was found that in years of low inventories, the wheat price was much higher for July Futures than for September Futures. This was reflected by a negative spread. In years of high inventories, the difference between July and September Futures was only slightly negative, only separated by an amount approximately equal to the cost of storing wheat for two months. Another important observation was that spread tended to become more negative as harvest time approached, because a situation of impending scarcity or abundance only became clear towards the end of the crop year.

An empirical study on the theory of storage was done by Fama and French (1987). They found that seasonal[8] commodities with high storage costs and limited storage period, like agricultural products, had the highest *spread* standard deviation. Metals, which are not subject to seasonality in supply and demand and have relative low storage costs, was found to have the lowest *spread* standard deviation. The authors have used approximations for futures and spot prices and Ordinary Least Square estimation for 21 commodities.
Deaton and Laroque (1992) have taken a more rigorous approach to study the applicability of standard rational expectations of competitive storage model on thirteen commodities from

[7] “A generation unit is said to be committed if it can be turned on, brought up to the desirable speed and connected to the system in order to deliver power to the network, all these steps taking place in a very short amount of time” (Geman 2005c)

[8] Spot prices for agricultural commodities usually increase between harvests and fall across harvests (Fama & French 1987).

1900 to 1987. For most of the thirteen commodity prices, the behavior of prices from one year to the next conforms to the predictions of the theory about conditional expectations and conditional variances.

Pirrong and NG (1994) found that the variance of the adjusted *spread* has a statistically significant effect on the variance of both spot and forward returns and on the correlation between these returns. The authors have investigated using a Generalized Method of Moments (GMM) estimation on individual commodities. Their results were only applicable for industrial metals (consumption assets), not for precious metals (investment assets).

Pindyck (2004) investigate the petroleum complex[9], and finds that changes in volatility help to explain changes in the spot-futures spread. In her book, Geman (2005a) points out that whenever there is a downward adjustment of the estimated oil reserves in the US or another region, the volatility of oil prices increases sharply.

Benavides (2010) extended the work of Pirrong and NG (1994), and obtained results that support the theory of storage for the two seasonal commodities, corn and wheat.

Carpantier and Dufays (2013) found support for the implication that volatility increases in times of low inventories by investigating 16 different commodities. However, the inventory effect was not observed for all commodities, and not a specific type of commodity. The effect was found for precious metals, challenging the results obtained by Pirrong and NG (1994). In this study, past positive returns was used as a proxy for the states of inventories instead of the *spread*. The reasoning behind this choice was that positive price shocks could signal declines in inventories.

[9] Crude oil, heating oil and gasoline.

3.3 Effect of Monetary Policy & Market Liquidity on Commodity Prices

The recent extreme price movements in commodity prices have given credence to a growing amount of literature that the prices in commodity markets are not entirely determined by the fundamentals of supply and demand. The role of global monetary conditions has often been cited as a driving factor of commodity prices. High interest rate reduces the marginal benefit derived from inventories, and makes capital expenditure more costly. With high interest rates, the incentive to extract today rather than tomorrow is increasing while the incentive to carry inventories is decreasing. Similarly the development of derivatives markets and commodity indexes has made it convenient for financial speculators and traders to invest and take position in the commodity markets. But it also decreased risk premiums and better integrated markets.

Anzuini et al. (2010) investigated the empirical relationship between US monetary policy and commodity prices by means of a standard VAR system, commonly used in analyzing the effects of monetary policy shocks. The results suggest that expansionary US monetary policy shocks drove up the broad commodity price index and all of its components. While these effects are significant, they however do not appear to be overwhelmingly large. Frankel (2006) has also empirically found a relationship between real interest rates and real commodity prices. He has suggested negative relationship between commodity prices and interest rate exists due to the fact higher rates interest rates create disincentives to carry commodity inventories.

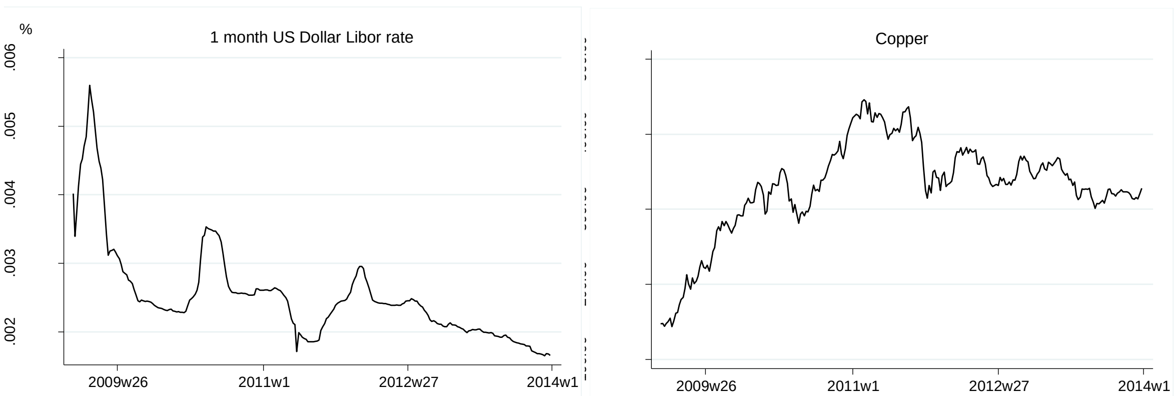

Figure 4: Copper Spot Price and 1 month USD Libor rate (2009 – 2013) for our data set.

Nominal interest rates also play a significant role in global carry trade that leads that can have significant impact on the commodities price volatility when sudden unwinding of contracts take place.

As mentioned, financial markets allow market participants to hedge their exposures to price movements, and thereby serve as a helpful complement to the physical commodity market. Even though these markets provide the opportunity to manage the risk of volatile prices, it has been suggested that opening up for speculators/risk seekers actually contribute to an increase in the level of volatility. (Dwyer et al. 2011)

Due to deregulations and the development of new financial products and electronic trading, financial markets have grown significantly the last decade. In addition to this, there is also a diversification benefit from including commodities in a portfolio, attracting even more market participants. Gorton and Rouwenhorst (2005) found strong evidence of a negative relationship between commodities and stocks/bonds.

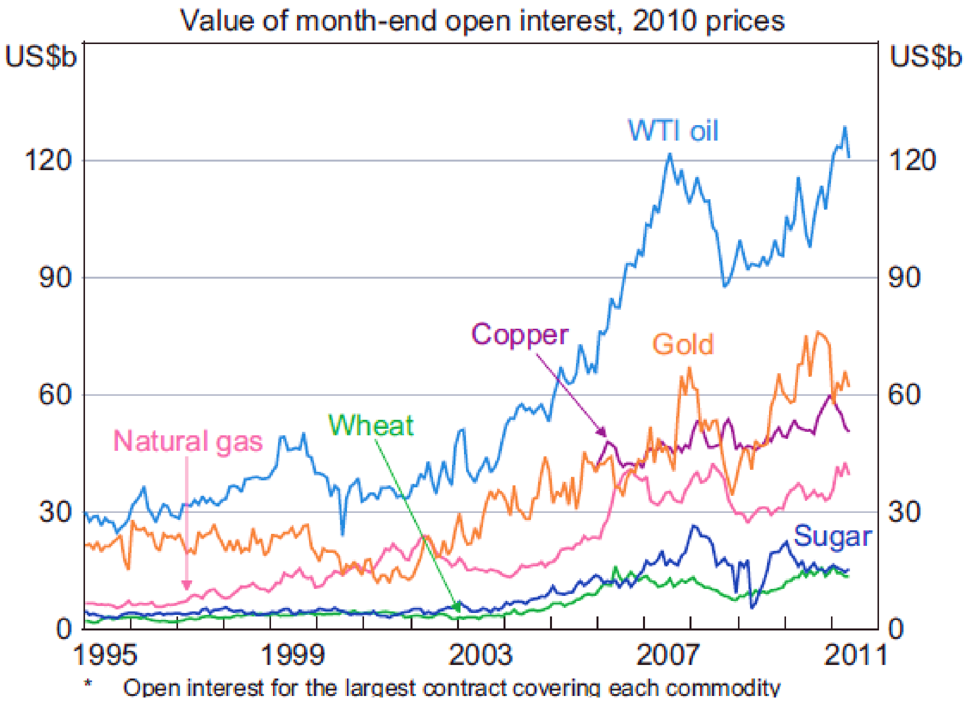

Figure 5: Commodity Futures Market Size. Source: (Dwyer et al. 2011)*

Open interest and volume has been the most often used indicator for market liquidity in commodity derivatives market. “Open interest represents the total number of contracts, either long or short, that have been entered into and not yet offset by delivery. Each open transaction has a buyer and seller, but for calculation of open interest, only one side of the contract is counted." (CMEgroup 2014) In fact, most Futures contracts are not held until maturity, as the position is being closed out by doing an opposite trade. This means that many market participants don't have the intention to get involved in the physical delivery of the commodity, using it only as a financial risk management instrument. As we can see from Figure 5 open interest has clearly been ascending the last decades, and it is applicable across all commodity classes.

However, Irwin and Sanders (2012) found the expanding market participation may have decreased risk premiums, and hence, the cost of hedging, reduced price volatility, and better integrated commodity markets with financial markets. But the empirical evidence for the causality of financial trading on volatility is rather low[10]. The price increases for iron ore and coal, which have relatively small derivatives markets, have not been different from the price increases for commodities with highly developed derivatives markets. The prices also fell together during the financial crisis. (Dwyer et al. 2011)

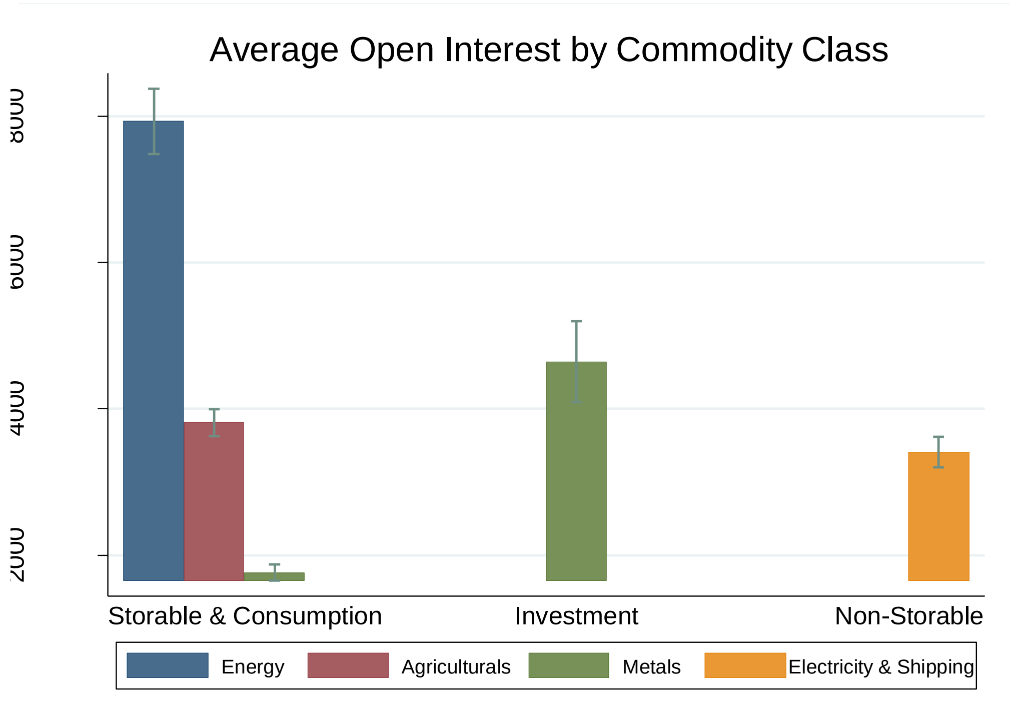

Figure 6: Average open interest by commodity class 2009-2013(95% conf. interval)

[10] Studies conducted by the U.S. Commodity Futures Trading Commission (2009), OECD (2010) and IMF (2011) did not find evidence on speculation activity driving the commodity prices.

4. Data

In our panel data analysis we obtained daily observations of *spot prices* and generic *futures* prices with constant maturity of 1 month. In the futures commodity markets, 1-month generic contract/constant maturity contracts were preferred because they were found to be the most liquid contract available for almost all commodities. In addition, the number of missing data on daily *open interest* for the 1-month generic futures contracts was minimal. All data were collected from the Bloomberg Terminal. The complete summary of the commodities and the list of exchanges these commodities were traded on are presented in Appendix B, together with the description and contract units. Our database covers the period January 2nd 2009 to December 17th 2013, giving us a total of 1 158[11] daily observations for 31 different commodities and a total of 258 weekly observations.

As explained later in details, most of our specification test required us to have a balanced panel i.e. to have the same time periods for each cross section observation. In most cases our panel data set is almost complete, that is, missing observations are infrequent, or only few items were missing from each observation. So we could justify in most cases that the data are randomly missing, thus converting this type of unbalanced panel into a balanced panel leads to a little loss of efficiency. However, if missing data occurs systematically, then the exogeneity assumption doesn't hold and can lead to biased estimators. Therefore the cause of missing data is important. We observed our data came from market exchanges from different countries and the holiday schedule of these exchanges appeared to systematically affect our data on an annual basis. Thus in these few cases we noticed some missing data are nonrandom, therefore converting into a balanced panel may result in biased sample. So in these few cases where we have near complete panel we have approximated the missing data using linear interpolation[12]. In spite of our best effort there were a lot of unexplained missing dates so to obtain a balanced panel we dropped these dates from our entire panel data. These missing observations are listed in Appendix A.

---

[11] We have estimated the total number of trading days in the selected period that matches all exchanges to be 1 222, meaning that 64 random trading days have been taken out of our time series.

[12] One observation missing: $x_t = \frac{X_{t+1} + x_{t-1}}{2}$. If two consecutive observations were missing:

a) $x_t = \frac{X_{t+2} + 2 * x_{t-1}}{3}$ b) $x_{t+1} = \frac{2 * X_{t+1} + x_{t-2}}{3}$

As a proxy for global liquidity we have used the weekly average of 1 month US Dollar *Libor rate*. The Libor interest rate is the standard financial index used in U.S. capital markets, and it is the interest rate at which large international banks are willing to lend each other money on a short-term basis.

4.1 Active vs Generic/Constant Maturity Contracts:

A futures contract will bear different risk characteristics along its maturity, even if all the market and portfolio conditions remained the same. Fama and French (1987) had mentioned this problem in their data that futures prices that do not account for this behavior would produce misleading results as, in general, futures become more volatile as expiration date comes closer. To overcome the constant changes in volatility of active contracts we have used the generic /constant-maturity contracts in our analysis that maintains the invariance characteristics required for the analysis. A constant-maturity futures price series indicates, for each time *t* (1 Month in our thesis) an interpolated price reflecting a specific time-to-expiration that is constant over time. For example, the one month constant maturity forward is at all times based on a combination of contracts with the middle of their delivery periods approximately one months from the date of calculation. We obtained our data from Bloomberg data terminal which follows the Bloomberg Constant Maturity Commodity Index which determine the composition and component weights for these generic contracts[13].

[13] http://www.bloombergindexes.com/content/uploads/sites/3/2013/05/CMCI-Methodology.pdf

Table 1: Descriptive Statistics

Variable	Mean	Std. Dev.	Min	Max	Obs
Spot Return Volatility	0.0529	0.1050	0	0.8927	7998
Spot Return Variance	0.0139	0.0578	0	0.7970	7998
Futures Return Volatility	0.0179	0.0142	0	0.2498	7998
Futures Return Variance	0.0005	0.0016	0	0.0624	7998
Adjusted Spread	0.2955	1.3817	-7.9841	19.8297	7998
Variance of Adjusted Spread	1.0456	5.0305	0	80.5895	7998
1 Month USD Libor Rate	0.0025	0.0007	0.0017	0.0056	7998
Variance of 1 M USD Libor Rate	4.20e-09	4.01e-08	0	6.33e-07	7998
Change in Open Interest	8 500 385	1.19e+09	-1.26e+10	4.24e+10	7998

4.2 Data Transformation and Variable

We used the daily data to calculate weekly averages, giving us a total of 258 observations[14] per variable.

4.2.1 Explained Variable

Spot Return Variance ($\sigma^2_{S_t}$)

To estimate volatility, we have used the sample standard deviations of adjusted daily log changes in spot and futures prices. As Campbell et al. (2001) pointed out, in addition to its simplicity, this approach has the advantage that it does not require a parametric model of the evolution of volatility.

Daily returns for both spot and Futures with maturity in 1 month were calculated in the following way:

$$R_t^S = ln\frac{S_t}{S_{t-1}} \tag{8}$$

Further, weekly volatilities were computed using the formula:

$$\boldsymbol{\sigma}^2_{\mathbf{S_t}} = \frac{1}{L-1}\sum_{i=1}^{L}\left[\left(R_i - \overline{R}\right)^2\right] \tag{9}$$

Where *i* is the first day of the week and *L* is the last day of the week.

and, $\bar{R} = \frac{\sum R_t + \cdots + R_L}{L}$

[14] We had no data for week 41 in 2013.

Table 2: Correlation Matrix

Variable	σ_{S_t}	$\sigma^2_{S_t}$	σ_{F_t}	$\sigma^2_{F_t}$	Z_t	$var(Z_t)$	r_t	$var(r_t)$	ΔOI_t
σ_{S_t}	1								
$\sigma^2_{S_t}$	0.9187	1							
σ_{F_t}	-0.0033	-0.0488	1						
$\sigma^2_{F_t}$	0.0305	-0.0021	0.8186	1					
Z_t	0.3562	0.3232	-0.0253	0.0317	1				
$var(Z_t)$	0.7421	0.7775	-0.0238	0.0127	0.3721	1			
r_t	-0.0199	-0.0482	0.1900	0.0789	-0.0189	-0.0374	1		
$var(r_t)$	-0.0020	-0.0106	0.0524	0.0221	0.0030	-0.0144	0.0075	1	
ΔOI_t	-0.0022	-0.0027	0.0196	0.0096	0.0007	-0.0013	-0.0028	-0.0039	1

4.2.2 Explanatory Variables

Variance of Adjusted Spread $(\boldsymbol{var}(\boldsymbol{Z_t}))$

In Equation (7) we introduced the relationship between spot and Futures prices, and referred to this spread as *the adjusted spread*. This relationship included risk free interest rate, the convenience yield and storage costs. Because of the difficulties in obtaining storage costs for the storable commodities, we decided to leave them out of our analysis[15]. Pindyck (2004) Fama and French (1987) have suggested that leaving out storage cost does not have any impact on the overall results as long as they remain constant over the time period of observation. Including a constant storage cost may have impacts on the intercept estimate of the regression. By leaving out the storage cost we expect to observe more observations of a positive spread. We obtain the daily spread adjusted for interest rate on an annual basis (z_t) by the following formula, which is derived from equation (7):

$$z_t = 12 * ln\frac{F_{t,T}}{S_t} - r_{t,T} = w_{t,T} - c_{t,T} \tag{10}$$

Where $r_{t,T}$, S_t and $F_{t,T}$ represent the 1 month Libor, the spot price and the futures price respectively. The length of the time interval t to T is one month. As pointed out by Pirrong and Ng (1994) and Brennan (1991) the variance of adjusted spread follows the spot return variance closely. We also observed this particular feature from our correlation matrix presented in Table 2, where the correlation between Adjusted Spread Variance and the Spot return variance is 0.7775.Therefore we used the variance of adjusted spread instead of the volatility of the adjusted spread.

So we calculated the sample variance of the adjusted spread as follows:

$$var(Z_t) = \frac{1}{L-1}\sum_{i=1}^{L}\left[\left(z_{t_i} - \overline{z_{t\imath}}\right)^2\right] \tag{11}$$

and, $$\overline{Z_{t\imath}} = \frac{\sum z_{t_i} + \cdots + z_{t_L}}{L}$$

[15] Fama and French (1987) also use the interest-adjusted spread as a proxy for inventories.

Variance of Nominal Interest Rate $(var(r_t))$

For r^w we use weekly observations of 1 month US Dollar Libor interest rate and calculated the sample variance as illustrated above for adjusted spread

Change in value (USD) of Open Interest (ΔOI_t)

Open interest refers to the number of futures contracts outstanding or not delivered on a particular date. As standalone open interest position in every commodity are not comparable we used the USD value of these open interest position to calculate in this analysis has a *contract unit*, and must be transformed into US Dollars. (See Appendix B). The value (USD) of daily open interest is computed:

$$\mathrm{oi_t} = F_{t,T} * \text{ contract unit }_{\mathrm{t}} * \text{ open interest }_{\mathrm{t}} \tag{12}$$

This was averaged on a weekly basis:

$$oi^w = \frac{oi_t + \cdots + oi_L}{L} \tag{13}$$

The change in value of open interest position was calculated as follows:

$$\Delta OI_t = \mathrm{oi_t} - \mathrm{oi_{t-1}} \tag{14}$$

5. Methodology

The salient and distinct feature of our research has been to study commodity as a broad class unlike previous research where structural models on theory of storage were empirically tested on individual commodities. Alvarez et al. (1991) have pointed out, in their study of cross country economic performance over time, that when correlation across units becomes a natural part of the specification panel data models provide more consistent results as compared to individual estimation of regression parameters. The primary challenge to this method was the absence of quality spot and futures data for a broad cross section of commodities. The necessity for a balanced panel made the problem of missing data even more difficult for estimation.

The Fixed Effect (FE) Model and Random Effect (RE) Model are among the most common panel data models. However, their estimators (RE model with OLS estimator) are consistent when the cross sectional dimension approaches infinity. In our panel data the time series dimension (T =258) is relatively larger than cross sectional dimension (N= 31). Therefore we choose an alternative model: the Seemingly Unrelated Regression (SUR) Model using Feasible Generalized Least Square (GLS) estimation techniques. The SUR model was preferred because the consistency of the SUR estimator is based on the large-sample properties of “large T, small N" datasets as T approaches infinity. However, the SUR Model assumes no endogeneity (correlation between explanatory variable and error term) to give unbiased estimators. Therefore if endogeneity is assumed to be absent then the GLS estimators from the SUR model provides us with more efficient and consistent estimator than the Maximum Likelihood Estimator used by Pirrong and Ng (1994) and Pindyck (2004).
In our analysis we have used all three models and compared their results. In the next sections we will introduce briefly about the three models viz: Seemingly Unrelated Regression, Fixed Effect and Random Effect Model briefly and the Generalized Least Square Estimation technique with a brief overview of the different covariance structures.

5.1 Panel Data Models

A panel dataset has two dimensions; a cross-section (N) and a time series (T). In our data set the 31 different commodities represent the cross section and the weekly observations (258) from 2009-2013 form the time series. We have employed three different panel data models:

1. Seemingly Unrelated Regression Model (SUR)
2. Fixed Effect Model. (FE)
3. Random Effect Model(RE)

The heterogeneity (individual characteristics) of cross sections in SUR is modelled by assuming difference in covariance between panels and within panel whereas in FE and RE model it is modelled using shifts in the mean (different intercepts). The SUR and Random effect models assume that there is no endogeneity (no correlation between the error term and one or more regressors) whereas the Fixed Effect Model makes no such assumptions. The SUR and RE model can use Feasible Generalized Least Square (GLS) estimation techniques whereas the FE model uses FE Estimators(Within Estimators). The GLS estimators are more consistent and efficient than the Fixed Effect estimator. The FGLS provides consistent estimators when $T \geq N$ while the FE estimators are consistent when $N \gg T$ (Greene (2003). Our panel data has temporal dimension (T) larger than the cross sectional (N).

5.1.1 Seemingly Unrelated Regression Model

For situation in which we want to estimate a similar specification for a number of different units: for instance, the estimation of a production function or cost function for each industry. If the equation to be estimated for a given unit meets the zero conditional mean assumption, we may estimate each equation independently. However, in instances we may want to estimate the equations jointly for two reasons: 1) Firstly to allow cross-sectional correlation to be imposed or tested, and 2) Secondly, to gain efficiency, since we might expect the error terms across equations to be contemporaneously correlated. Such equations are often called Seemingly Unrelated Regressions (SUR) Model.

In our panel data the cross-sectional units are relatively small compared to the number of time periods which is relatively large ($N \leq T$). Another important characteristic of our large panel data sets is that there is presence of heteroscedasticity and correlation across panels and time (i.e. commodities return are correlated across time and with other commodities). Taking into account these features we have used the framework SUR Model and estimated the regression parameters using Feasible Generalized Least Square estimation techniques as suggested by Greene (2003). In such model it is reasonable to specify a common conditional mean function across the groups, with heterogeneity taking the form of different covariance structures rather than shifts in the means. An essential feature is that we have also assumed the coefficients of regression equal across all commodities.

5.1.2 Feasible Generalized Least Square Estimation:

The estimation technique used for both SUR and RE Model is the Generalized Least Square estimation. For this analysis is the generalized regression is represented as :

$$y_{it} = \boldsymbol{x_{it}}\beta + \varepsilon_{it} \qquad (15)$$

Where

$$\boldsymbol{y_i} = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{\varepsilon}_i \qquad (16)$$

Under the assumptions $$E[\boldsymbol{\varepsilon i} \mid \boldsymbol{X}] = \boldsymbol{0} \qquad (17)$$

and

$$\mathrm{E}[\boldsymbol{\varepsilon}\mathbf{i}\boldsymbol{\varepsilon}\mathbf{j} \mid \mathbf{X}] = \boldsymbol{\sigma}\mathrm{ij}\ \boldsymbol{\Omega}\mathrm{i\ j} \qquad (18)$$

And the heterogeneity in its most general form the covariance matrix can be represented as:

$$E\left[\boldsymbol{\varepsilon_i}\boldsymbol{\varepsilon_j} \mid \boldsymbol{X}\right] = \Omega = \begin{bmatrix} \sigma_{1,1}\Omega_{1,1} & \cdots & \sigma_{1,m}\Omega_{1,m} \\ \vdots & \ddots & \vdots \\ \sigma_{m,1}\Omega_{m,1} & \cdots & \sigma_{m,m}\Omega_{m,m} \end{bmatrix} \qquad (19)$$

Where $\boldsymbol{\Omega}$ is the cross sectional covariance across the groups.

Then the generalized least squares estimator of β is based on the assumptions that determine Ω and is given by the equation:

$$\widehat{\boldsymbol{\beta}} = [\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X}]^{-1}\,[\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{y}] \qquad (20)$$

As specified by Greene (2003) in the generalized linear regression model, the regression coefficients, can be consistently, if not efficiently estimated by ordinary least squares. A consistent estimator of $\sigma_{i\,j}$ can be based on the sample analog to the result

$$\boldsymbol{E}[\boldsymbol{\varepsilon}_{it}\boldsymbol{\varepsilon}_{jt}] = \boldsymbol{E}[\boldsymbol{\varepsilon}_{it}\boldsymbol{\varepsilon}_{jt}/\,\boldsymbol{T}] = \boldsymbol{\sigma}_{it} \qquad (21)$$

This is estimated by using the residuals obtained from ordinary least squares residuals on our regression model:

$$\widehat{\boldsymbol{\sigma}}_{ij} = \frac{\boldsymbol{e}_i^T\boldsymbol{e}_j}{\boldsymbol{T}} \qquad (22)$$

5.1.3 Covariance Structures

Different models under SUR model using GLS estimation differ by the different assumptions that are used to model the heterogeneity of the covariance matrix Ω. We have briefly presented the 4 covariance matrix that we have used in our regressions.

Pooled Ordinary Least Square (OLS) Estimation

When the data set for individual panels are pooled together and the slope coefficient is obtained by simple regression it is called the pooled OLS estimation. It is the simplest model but requires the assumption of conditional mean independence, homoscedasticity and no autocorrelation for the regression to be efficient and consistent. The cross sectional covariance matrix is for pooled OLS is:

$$\Omega = \begin{bmatrix} \boldsymbol{\sigma^2 I} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \boldsymbol{\sigma^2 I} \end{bmatrix} \tag{23}$$

GLS (I) assuming Heteroscedasticity

In many cross-sectional datasets if the variance for each of the panels differs then there is heteroscedasticity. The $\boldsymbol{\Omega}$ is the cross sectional covariance across the groups' heteroscedastic model is:

$$\Omega = \begin{bmatrix} \boldsymbol{\sigma^2{}_1 I} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \boldsymbol{\sigma^2{}_n I} \end{bmatrix} \tag{24}$$

GLS (II) assuming Heteroscedasticity and Cross Sectional Correlation

In GLS (II) we assume that the error terms of panels are correlated, in addition to having different variances. In our data sets related commodities like energy, metals and agriculture and electricity prices and returns show high correlation within the cross sections. Therefore this appears to be a valid assumption while modelling for commodities. The $\boldsymbol{\Omega}$ is the cross sectional covariance across the groups' heteroscedastic model is:

$$\Omega = \begin{bmatrix} \boldsymbol{\sigma_1^2 I} & \cdots & \boldsymbol{\sigma_{1,m} I} \\ \vdots & \ddots & \vdots \\ \boldsymbol{\sigma_{m,1} I} & \cdots & \boldsymbol{\sigma_1^2 I} \end{bmatrix} \tag{25}$$

GLS (III) assuming Autocorrelation (Prais-Winstein Method)

It is simplest to begin with the assumption that no serial correlation within panels exists.

$$Corr[\varepsilon_{it}\ \varepsilon_{is}] \neq 0,\ if\ i = j \qquad (26)$$

However if this condition is violated (almost all commodities show autocorrelation) within the time series the covariance matrix that allows for autocorrelation to be modelled is:

$$\sigma_{ij}\Omega_{ij} = \frac{\sigma_{uij}}{1-\rho_i\rho_j}\begin{bmatrix} 1 & \cdots & \rho_j^{T-1} \\ \vdots & \ddots & \vdots \\ \rho_i^{T-1} & \cdots & 1 \end{bmatrix} \qquad (27)$$

5.2.1 Fixed Effect Panel Data Models

If we have reasons to assume the presence of time-constant factors like specific format of a futures market or a certain storage characteristics of a commodity that remains constant over time but contribute to its volatility then it is proper to estimate it by fixed effects model

$$y_{it} = x_{it}\beta + (\alpha + u_i) + \varepsilon_{it} \qquad (28)$$

In applications, u_i is referred to as unobserved heterogeneity and η_{it} ($= u_i + \varepsilon_{it}$) is the unit specific error term. This unobserved heterogeneity term differs between units, but for any particular unit, its value is constant.

5.2.1 Random Effect Panel Data Models

As an alternative to the individual fixed effects model, we may consider a random effects model.

$$y_{it} = \beta_0 + x_{it}\beta + (\alpha + u_i) + \varepsilon_{it} \qquad (29)$$

In random effect model we explicitly include an intercept so that we can make the assumption that the unobserved effect has a zero mean. The bracketed term or unit specific error term is now assumed to have an individual-specific component and an idiosyncratic component. In Fixed effects unobserved heterogeneity is treated “nuisance parameter" which, if ignored, causes bias and inconsistency in our estimators because it is correlated with one or more of the regressors but in case of Random Effect Model we assume that this unobserved heterogeneity is a random variable distributed independently of x all of the regressors. Then we can use GLS estimator to estimate the model using a covariance matrix

$$\Omega = \begin{bmatrix} \boldsymbol{\sigma_1^2 + \sigma_\varepsilon^2} & \cdots & \boldsymbol{\sigma_\varepsilon^2} \\ \vdots & \ddots & \vdots \\ \boldsymbol{\sigma_\varepsilon^2} & \cdots & \boldsymbol{\sigma_N^2 + \sigma_\varepsilon^2} \end{bmatrix} \tag{30}$$

Where the error term is normally distributed as $\varepsilon_{it} \sim N(0,\ \sigma_\varepsilon^2)$.

6. Analysis

In the previous sections we have discussed the effect of storage, interest rate and market liquidity on spot return variance of commodities. To test the explanatory power of these three factors on spot return variance we employed three corresponding explanatory variables viz. the variance of adjusted spread (var (Z_t)), the 1 month USD Libor interest rate variance (var (r_t)) and the changes in value of open interest position in futures markets (ΔOI_t). In the first part of our analysis we used the whole cross section (N = 31) of commodities to test the significance of the model and the explanatory power of the individual explanatory variable. In the second part we have divided the cross section into two categories viz. storable (N = 24) and non-storable (N = 7) commodities. We evaluated the regression models separately over both these cross sections and compared the regression results of the non-storable commodities against the regression results obtained from storable commodities and the full cross section. In the final part, we have selected a cross sections consisting of 4 precious metals (N = 4 viz. Gold, Silver, Platinum and Palladium). The same regressions were performed on this group to check if the adjusted spread did not have any impact on the spot return variance of, as predicted by the theory of storage.

We have used the Seemingly Unrelated Regression and Random Effect Models using GLS and Fixed Effect Model using within estimators respectively for panel data. These three models provide us with the flexibility to model under different assumptions. In order to determine the suitable regression model that best represents the behavior of our panel data we performed three specification tests viz. Likelihood Ratio-test (hetereoscedasticity across panel), Wooldridge Autocorrelation Test for Panel Data (autocorrelation across panels) and Lagrange Multiplier- Breusch Pagan test (cross sectional correlation across panels).

The presence of heteroscedasticity and serial correlation in panel-data models biases the standard errors and causes the results to be less efficient. Commodity prices are volatile, and volatility itself varies over time (Pindyck 2004). This means that we expect the presence of heteroscedasticity and autocorrelation within panels in our dataset. To test for heteroscedasticity we performed a likelihood ratio (LR) test. Similarly, to test for serial correlation in error term in the panel-data model we have used the Wooldridge Autocorrelation specification test (Drukker 2003).

As noted earlier, scientific literature on commodities has suggested that there is a considerable positive correlation among spot and future prices within commodities of the same class. There are strong reasons to assume the effects of random shocks in some commodity markets may affect the prices in other commodity markets. Our data set has several commodities from the same classes, so in addition to heteroscedasticity and within panel correlation we also expect correlation across panels. To test the independence between cross-section estimating Panel data models, Hoyos and Sarafidis (2013) suggested the use of the Lagrange Multiplier (LM) test, developed by Breusch and Pagan (1980) when the temporal dimension (T) of the panel is larger than the cross-sectional dimension (N) as is our case.

6.1 Structural Model

6.1.1 SUR Model

The SUR model assumes a common conditional mean function across the cross section but individual heterogeneity (individual commodity characteristics) take the form of different covariance structures (Wooldridge 2010). The SUR model with GLS estimation allows us to control for heteroscedasticity, autocorrelation of error term and most importantly the cross correlation of cross sections (i.e. commodities) are present in the data set. In the GLS estimation technique most asymptotic results are obtained with respect to T approaches infinity. Hence this model is preferred for data sets where the number of cross sectional observations are smaller than the number of time periods. The SUR model assumes that the error of the regression model is uncorrelated with the explanatory variables and the temporal dimension is larger than the cross sectional dimension.

$$y_{it} = \boldsymbol{\beta_0} + \boldsymbol{\beta_z var(Z_t)} + \boldsymbol{\beta_r var(r_t)} + \boldsymbol{\beta_{oi}} \Delta \boldsymbol{OI_t} + \boldsymbol{\varepsilon_{it}} \tag{29}$$

The GLS estimator, when correlation across cross sections is present, are more efficient than the Fixed Effect estimator. Moreover in our data set the temporal dimension (T=258) is greater than the cross sectional dimension (N= 31) so the FGLS provides more consistent (asymptotic results) estimators. Therefore the GLS estimation model is preferred in our case.

6.1.2 Fixed Effect & Random Effect Model

In both Fixed Effect and Random Effect models, each individual cross section has an unobserved heterogeneity that captures all time constant factor that affects the dependant variable. The presence of the unobserved heterogeneity basically means that each cross section (commodity) has a different intercept, but has the same slope coefficient and the same idiosyncratic error distribution. This fixed effect estimator gives consistent result when the cross sections (N) are relatively larger than the observation periods (T). Heteroscedasticity and autocorrelation within panels can be modelled but the cross panel dependence is always assumed away in this model (Greene 2003). Driscoll and Kraay (1997) have specified that the cross panel dependence in panel data can be dealt within the scope of Fixed effect model but they have serious limitations if the temporal dimension exceeds the cross sectional dimension of the panel data. The FE model we used is:

$$y_{it} = \boldsymbol{\beta_z var(Z_t)} + \boldsymbol{\beta_r var(r_t)} + \boldsymbol{\beta_{oi}} \Delta \boldsymbol{OI_t} + (\boldsymbol{\alpha} + \boldsymbol{u_i}) + \varepsilon_{it} \tag{30}$$

The distinguishing feature between the Random Effect model from Fixed Effect model is that the that the unobserved heterogeneity in the Random Effect estimation is assumed to be independent of each explanatory variable in all time periods. The RE model we used is:

$$y_{it} = \boldsymbol{\beta_0} + \boldsymbol{\beta_z var(Z_t)} + \boldsymbol{\beta_r var(r_t)} + \boldsymbol{\beta_{oi}} \Delta \boldsymbol{OI_t} + (\boldsymbol{\alpha} + \boldsymbol{u_i}) + \varepsilon_{it} \tag{31}$$

We have used the Hausman specification test to determine the more efficient estimator between Random Effect and Fixed effect estimators.

6.2 Analysis of the whole cross-section (N =31)

All the regressions in this section are based on weekly observation from 2009 to 2013 for 31 different commodities. The results of these regressions are presented in Table 3.

The LR chi-squared test statistics for testing heteroscedasticity (GLS estimates) over the whole cross section is 49989.14 with a corresponding p-value of 0.000. Hence the test statistics are significant so heteroscedasticity across panels is likely to be present in our dataset. The Wooldridge F- test statistics for first order serial correlation within panel over the whole cross section is 924.280 with a corresponding p-value of 0.000. Hence we reject the null hypothesis that states no first order serial autocorrelation exist within the cross section. This indicates the presence of first-order serial correlation within cross sections in our dataset. Finally the LM- Breusch and Pagan test fail to reject the null hypothesis of cross sectional independence at the 1% level. In sum, the specification tests indicate that there is a presence of hetereoscedasticty across panels, first order serial correlation across panels and cross sectional correlation between panels. Hence the GLS (III) estimators which incorporate all these properties are the preferred estimation technique.

The first column of Table 3 is the Pooled OLS regression. Pooled OLS estimators are inefficient in our case due to violation presence of heteroscedasticity and autocorrelations. However the goodness of fit statistics R-square (60.5) gives a rough view of how the model fits the data. In column 2 and column 3 of Table 3 the estimates from Fixed and Random Effect models are tabulated. As indicated by the specification test, we have controlled these models for heteroscedasticity and first order serial autocorrelation. The F-test for the Fixed Effect and the Wald-Chi square test statistics for the Random Effect are very high so both models are significant. The coefficients for adjusted spread variance and market liquidity coefficient are statistically significant but the coefficient for interest rate variance was not statistically significant at the 10% level. The Hausman test statistic is negative. We interpret this result as strong evidence that we cannot reject the null hypothesis that the random effect estimator is efficient and consistent under the assumption being tested. Our data set consists of a diverse set of commodities whose supply and demand patterns, storage properties and structure of market and where they are traded appear random and independent of the explanatory variables we used. The Hausman test result confirms this fact.

Table 3: Spot return variance model estimates (31 commodities)

Dependent Variable	Spot return variance ($\sigma^2_{S_t}$)
Cross-section units	31
Time period	2009 - 2013
Number of observations	258

	Pooled OLS	Fixed Effect	Random Effect	GLS(I)	GLS(II)	GLS(III)
$var(Z_t)$	0.00892***	0.00710***	0.00892***	0.00740***	0.00656***	0.00485***
	(8.09E-05)	(-8.78E-05)	(-8.09E-05)	(-1.61E-05)	(-1.43E-05)	(-1.09E-05)
$var(r_t)$	507.6	-2487.7	507.6	554.9	464.1**	273.0
	(10138.3)	(9307.0)	(10138.3)	(1407.9)	(231.9)	(266.5)
ΔOI_t	-2.56E-04***	-7.50E-06*	-2.56E-04**	-5.00E-06	1.40E-06	-6.00E-06***
	(-1.05E-04)	(-9.82E-05)	(-1.05E-04)	(-1.32E-05)	(-4.60E-06)	(-1.36E-06)
Commodity FE	No	Yes	Yes	No	No	No
GLS	No	No	Yes	Yes	Yes	Yes
Heteroskedastic	No	No	No	Yes	Yes	Yes
Panel Specific Autocorrelation	No	No	No	No	Yes	Yes
Correlated Error Structure	No	No	No	No	No	Yes
R^2	0.605	0.452				
F-test	251.40	274.20				
Prob. > F	0.0000	0.0000				
Wald Chi2			1208.37	2109.33	2113.76	1997.03
Prob. > Chi2			0.0000	0.0000	0.0000	0.0000

Standard errors in parentheses: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

1. GLS(I) refers to an estimation with a covariance structure allowing heteroskedasticity but uncorrelated error structure.
2. GLS(II) refers to an estimation with a covariance structure allowing heteroskedasticity but uncorrelated error structure and panel-specific AR1 autocorrelation structure.
3. GLS(III) refers to an estimation with a covariance structure allowing heteroscedasticity and correlated error structure, and panel-specific AR1 autocorrelation structure.
4. Hausman Test: chi2(1) = -0.55; asymptotic assumptions for the test not satisfied.
5. Breusch-Pagan LM test: chibar2(01) = 0.00; Prob. > chi2bar = 1.0000
6. Test for heteroscedasticity: chi2(30) = 49989.14; Prob. > Chi2 = 0.0000
7. Wooldridge Test for autocorrelation: F(1, 30) = 924.280; Prob. > F = 0.0000

In column 4, 5 and 6 of Table 3 the estimates for three GLS estimations (excluding the Pooled OLS) are tabulated. The GLS (I) model allows estimation in presence of heteroscedasticity, the GLS (II) model allows estimation in presence of heteroscedasticity and panel-specific autocorrelation AR(1) structure and the GLS (III) model allows estimation in presence of heteroscedasticity with correlated error structure, and panel-specific AR(1) autocorrelation structure. The Wald-Chi square test statistics for all three GLS estimations are very high meaning that all three GLS models are significant.

The coefficient for the variance of adjusted spread across all the regression models and estimation techniques is statistically significant at the 1% level and positive. The GLS (III) estimates predict that one unit increase in variance in adjusted spread will increase the variance of spot return variance by 0.00485 units. The result from the regression supports the premise of theory of storage and is consistent with the existing literature on commodities price behavior. In the context of storable commodity the adjusted spread variance increases as inventories become scarce (Lautier 2009). This implies that as inventories become scarce the volatility in the spot market increases.

The coefficient for the market liquidity is negative across all the regression models and is statistically significant at the 10% level except for GLS (I) and GLS (II) estimations. The GLS (III) estimates predict that one unit increase in variance of market liquidity will decrease the variance of spot return variance by 6.00E-06 units. This is consistent with the empirical results obtained by Irwin and Sanders (2012) (see chapter 3.3). The coefficient variance of market liquidity is much smaller than the coefficient of the variance of adjusted spread. This suggests that the fundamentals of supply and demand and inventory positions of a commodity have a dominant role in determining the spot price volatility than movement in market liquidity and volumes.

Moreover, the coefficient for the variance in global liquidity/monetary policy is not statistically significant at the 10% level across all the regression models except for GLS (II) estimations. Its sign is positive for all except for the fixed effect model.

The GLS estimation in accordance to the above discussion produces a larger number of significant coefficients. However, these results, are subject to the criticism advanced by Beck and Katz (1995).

6.3 Analysis of Non Storable Commodities

In this section we have classified the entire cross section into two category viz. non-storable (N=7) commodities (electricity and shipping) and storable (N=24) commodities. We evaluated the regression models separately over both these cross sections and compared the regression results. The results of these regressions are presented in Table 4 and Table 5.

The LR test and the LM- Breusch Pagan test results for non-storable cross section and storable cross section are similar to the test results from the dataset consisting of full cross section indicating the presence of heteroscedasticity across the panels and cross correlation between panels. However The Woolbridge test result for first order serial correlation within panel differs for non-storable compared to the storable cross sections.

The Woolbridge F- test statistics for first order serial correlation within panel over the whole cross section is 2.926 with a corresponding p-value of 0.1006. Hence we cannot reject the null hypothesis that states no first order serial autocorrelation exist within the data for cross section of storable commodities at 10% significance level. In sum the test results indicate the presence of heteroscedasticity and cross sectional correlation for both storable and non-storable commodities in the panel data. Therefore the GLS (III) estimators which incorporates all properties is the preferred estimation technique for non-storable commodities panel data whereas for estimating storable commodities we have dropped the autocorrelation assumption while evaluating the GLS (III) model.

The coefficient for the variance of adjusted spread across all the regression models and estimation techniques is statistically significant the 1% level and positive for both storable and non-storable commodities. The regression estimates predict that increase in the variance of adjusted spread will increase the variance of spot return variance for both storable and non-storable commodities.

Table 4 : Spot return variance model estimates (24 storable commodities)

Dependent Variable	Spot return variance ($\sigma^2_{S_t}$)
Cross-section units	24
Time period	2009 - 2013
Number of observations	258

	Pooled OLS	Fixed Effect	Random Effect	GLS(I)	GLS(II)	GLS(III)
$var(Z_t)$	0.00701**	0.00716*	0.00713*	0.00422***	0.00440***	0.00543***
	(0.00334)	(0.00408)	(0.00404)	(-1.52E-04)	(-1.55E-04)	(-1.12E-04)
$var(r_t)$	414.9*	411.0	411.8	292.7**	220.2*	84.49
	(246.6)	(302.2)	(301.2)	(143.1)	(130.2)	(242.8)
ΔOI_t	-6.36E-06	-6.14E-06	-6.20E-06	-1.08E-06	1.27E-07	-2.25E-06
	(-5.67E-06)	(-5.61E-06)	(-5.65E-06)	(-1.23E-06)	(-1.11E-06)	(-9.47E-07)
Commodity FE	No	Yes	No	No	No	No
GLS	No	No	Yes	Yes	Yes	Yes
Heteroskedastic	No	No	No	Yes	Yes	Yes
Panel Specific Autocorrelation	No	No	No	No	Yes	Yes
Correlated Error Structure	No	No	No	No	No	Yes
R^2	0.439	0.441				
F-test	4.95	3.75				
Prob. > F	0.0020	0.0250				
Wald Chi2			11.46	786.17	818.94	1702.73
Prob. > Chi2			0.0095	0.0000	0.0000	0.0000

Standard errors in parentheses: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

1. GLS(I) refers to an estimation with a covariance structure allowing heteroskedasticity but uncorrelated error structure.
2. GLS(II) refers to an estimation with a covariance structure allowing heteroskedasticity but uncorrelated error structure and panel-specific AR1 autocorrelation structure.
3. GLS(III) refers to an estimation with a covariance structure allowing heteroscedasticity and correlated error structure, and panel-specific AR1 autocorrelation structure.
4. Hausman Test: chi2(1) = 0.00; Prob. > Chi2 = 0.9564
5. Breusch-Pagan LM test: chibar2(01) = 262.64; Prob. > chi2bar = 0.0000
6. Test for heteroscedasticity: chi2(3) = 10867.41; Prob. > Chi2 = 0.0000
7. Wooldridge Test for autocorrelation: F(1,3) = 2.926; Prob. > F = 0.1006

Table 5 : Spot return variance model estimates (7 non-storable commodities)

Dependent Variable	Spot return variance ($\sigma^2_{S_t}$)
Cross-section units	7
Time period	2009 - 2013
Number of observations	258

	Pooled OLS	Fixed Effect	Random Effect	GLS(I)	GLS(II)	GLS(III)
$var(Z_t)$	0.00823***	0.00710***	0.00823***	0.00745***	0.00633***	0.00302***
	(-4.85E-04)	(-2.83E-04)	(-2.21E-04)	(-1.89E-04)	(-1.86E-04)	(-1.83E-04)
$var(r_t)$	-3998.6	-12560.9*	-3998.6	2697.8	1975.6	1820.0
	(16532.5)	(6212.1)	(5503.6)	(23227.2)	(4965.2)	(4919.4)
ΔOI_t	-0.0162***	-0.00268	-0.0162*	-0.00188	-0.00169	-6.27E-04
	(-3.21E-03)	(-3.76E-03)	(-8.72E-03)	(-4.52E-03)	(-1.32E-03)	(-8.92E-04)
Commodity FE	No	Yes	No	No	No	No
GLS	No	No	Yes	Yes	Yes	Yes
Heteroskedastic	No	No	No	Yes	Yes	Yes
Panel Specific Autocorrelation	No	No	No	No	Yes	Yes
Correlated Error Structure	No	No	No	No	No	Yes
R^2	0.543	0.452				
F-test	110.42	265.73				
Prob. > F	0.0000	0.0000				
Wald Chi2			7081.59	1557.34	1159.20	271.57
Prob. > Chi2			0.0000	0.0000	0.0000	0.0000

Standard errors in parentheses: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

1. GLS(I) refers to an estimation with a covariance structure allowing heteroskedasticity but uncorrelated error structure.
2. GLS(II) refers to an estimation with a covariance structure allowing heteroskedasticity but uncorrelated error structure and panel-specific AR1 autocorrelation structure.
3. GLS(III) refers to an estimation with a covariance structure allowing heteroscedasticity and correlated error structure, and panel-specific AR1 autocorrelation structure.
4. Hausman Test: chi2(1) = -0.36; asymptotic assumptions for the test not satisfied.
5. Breusch-Pagan LM test: chibar2(01) = 0.00; Prob. > chi2bar = 1.0000
6. Test for heteroskedasticity: chi2(30) = 2437.71; Prob. > Chi2 = 0.0000
7. Wooldridge Test for autocorrelation: F(1,6) = 789.300; Prob. > F = 0.0000

We are particularly interested in the estimates for the variance of adjusted spread because the convenience yield and hence the adjusted spread have different economic interpretation for both these categories of commodities (Williams & Wright 1991). There is no agreement over the applicability of convenience yield into non-storable commodities However, there exist extensions and interpretations to the theory of convenience yield that can be applied to non-storable where the unused capacity and information on future supply & demand conditions perform a similar role of inventory as in storable commodities (Benth & Meyer-Brandis 2009). Routledge et al. (2001) have argued that the theory of storage models can be extended to include goods which are not directly storable. The estimates for the variance of adjusted spread variable for both the storable and non-storable panel data are statistically significant, positive and comparable in magnitude. Therefore our results conforms with the approach that suggest a broader interpretation of convenience yield can be applied to non-storable commodities and adjusted spread for non-storable has a similar explanatory power over the spot return variance as that of the adjusted spread variance of storable commodities.

6.4 Analysis of Precious Metals

In this section we chose a cross section of metals that have no or negligible industrial application and whose adjusted spread is expected by the theory of storage to be close to zero. For this analysis we have chosen 4 metals viz. Gold, Platinum, Silver and Palladium. We estimated the regression models over this selected cross section of metals and the regressions are tabulated in Table 6.

The LR test, the LM- Breusch Pagan and the Woolbridge Autocorrealations test results for this cross section are similar to the full cross section of all 31 commodities. These specification tests indicate the presence of hetereoscedasticty across panels, first order serial correlation across panels and cross sectional correlation between panels. Therefore the GLS (III) estimators are the preferred estimation technique for this cross section also.

The coefficient for the variance of adjusted spread across all the regression models and estimation techniques is statistically significant the 1% level but the rest of the explanatory variables are statistically insignificant (10% significance level). The GLS (III) estimates predict one unit increase in variance of adjusted spread increases the spot return variance by 0.0142. The magnitude of this coefficient is the highest among the coefficient obtained from the regression from the other cross-sections. The adjusted spread variance for this cross section of metals is close to zero as predicted by theory but its effect on spot variance run counter to it.

Table 6 : Spot return variance model estimates (4 precious metals)

Dependent Variable	Spot return variance ($\sigma^2_{S_t}$)
Cross-section units	4
Time period	2009 - 2013
Number of observations	258

	Pooled OLS	Fixed Effect	Random Effect	GLS(I)	GLS(II)	GLS(III)
$var(Z_t)$	0.0170***	0.0174***	0.0172***	0.0139***	0.0144***	0.0142***
	(0.00324)	(0.00196)	(0.00178)	(-5.75E-04)	(-5.63E-04)	(-5.54E-04)
$var(r_t)$	-1111.2**	-1138.3	-1126.2	-415.0*	-448.7**	-402.8
	(472.1)	(719.5)	(704.5)	(233.0)	(227.0)	(270.1)
ΔOI_t	8.22E-06*	8.24E-06	8.23E-06*	2.72E-06*	2.61E-06	1.91E-06
	(-4.88E-06)	(-4.97E-06)	(-4.91E-06)	(-1.63E-06)	(-1.61E-06)	(-1.43E-06)
Commodity FE	No	Yes	No	No	No	No
GLS	No	No	Yes	Yes	Yes	Yes
Heteroskedastic	No	No	No	Yes	Yes	Yes
Panel Specific Autocorrelation	No	No	No	No	Yes	Yes
Correlated Error Structure	No	No	No	No	No	Yes
R^2	0.537	0.502				
F-test	9.17	7128.22				
Prob. > F	0.0000	0.0000				
Wald Chi2			20352.10	589.76	653.83	662.11
Prob. > Chi2			0.0000	0.0000	0.0000	0.0000

Standard errors in parentheses: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

1. GLS(I) refers to an estimation with a covariance structure allowing heteroskedasticity but uncorrelated error structure.
2. GLS(II) refers to an estimation with a covariance structure allowing heteroskedasticity but uncorrelated error structure and panel-specific AR1 autocorrelation structure.
3. GLS(III) refers to an estimation with a covariance structure allowing heteroscedasticity and correlated error structure, and panel-specific AR1 autocorrelation structure.
4. Hausman Test: chi2(1) = 0.66; Prob. > Chi2 = 0.4157
5. Breusch-Pagan LM test: chibar2(01) = 0.14; Prob. > chi2bar = 0.3537
6. Test for heteroscedasticity: chi2(3) = 814.20; Prob. > Chi2 = 0.0000
7. Wooldridge Test for autocorrelation: F(1,3) = 0.051; Prob. > F = 0.8357

Finally, we have plotted the spot return variance together with the negative, scaled value of the adjusted spread variance for some selected commodities. Figures 7-10 illustrates clearly the positive relationship that we obtained from our regression results. Especially when the spread widens this relationship becomes clear.

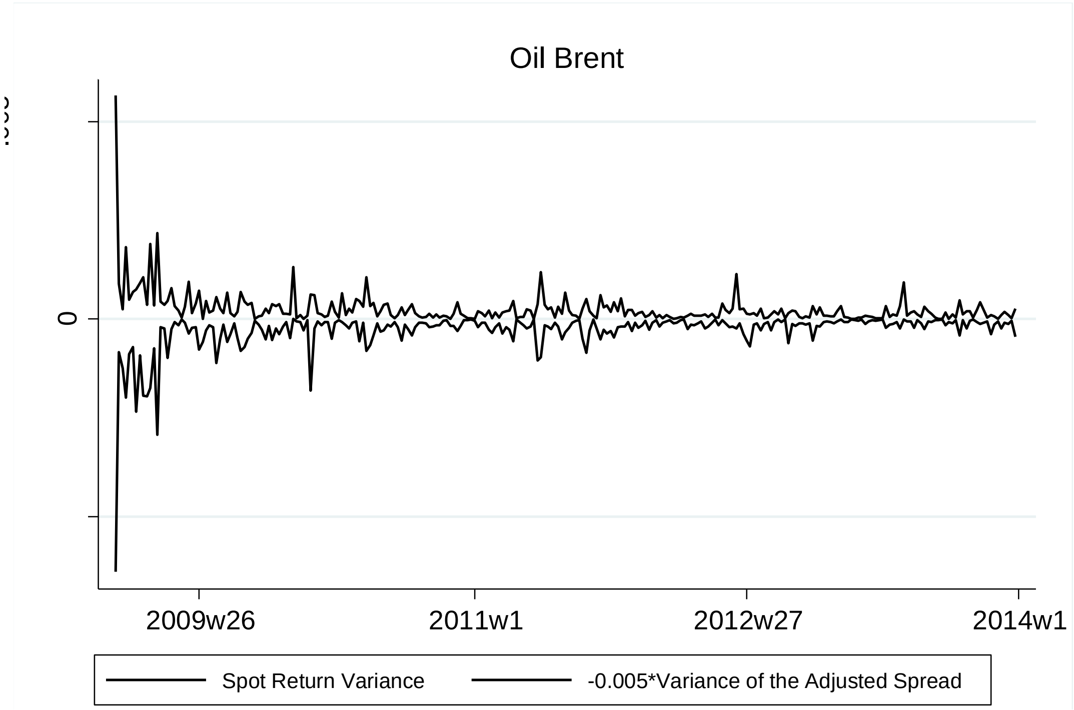

Figure 7: Oil Brent spot return variance, and the variance of the adjusted spread.

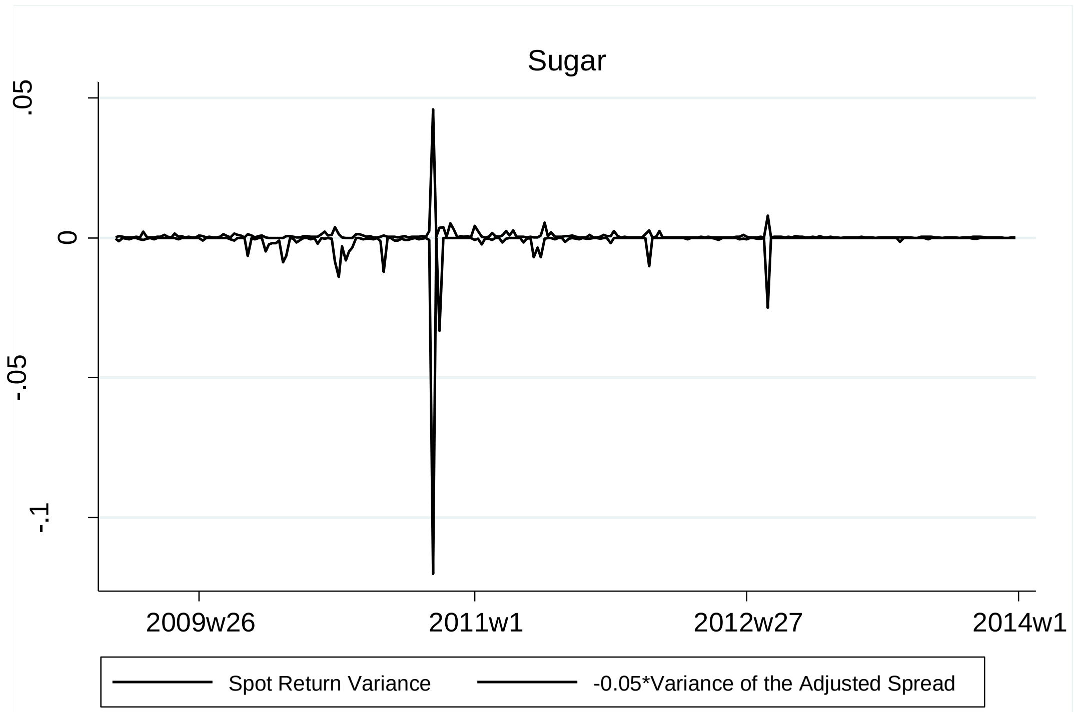

Figure 8: Sugar spot return variance, and the variance of the adjusted spread.

.

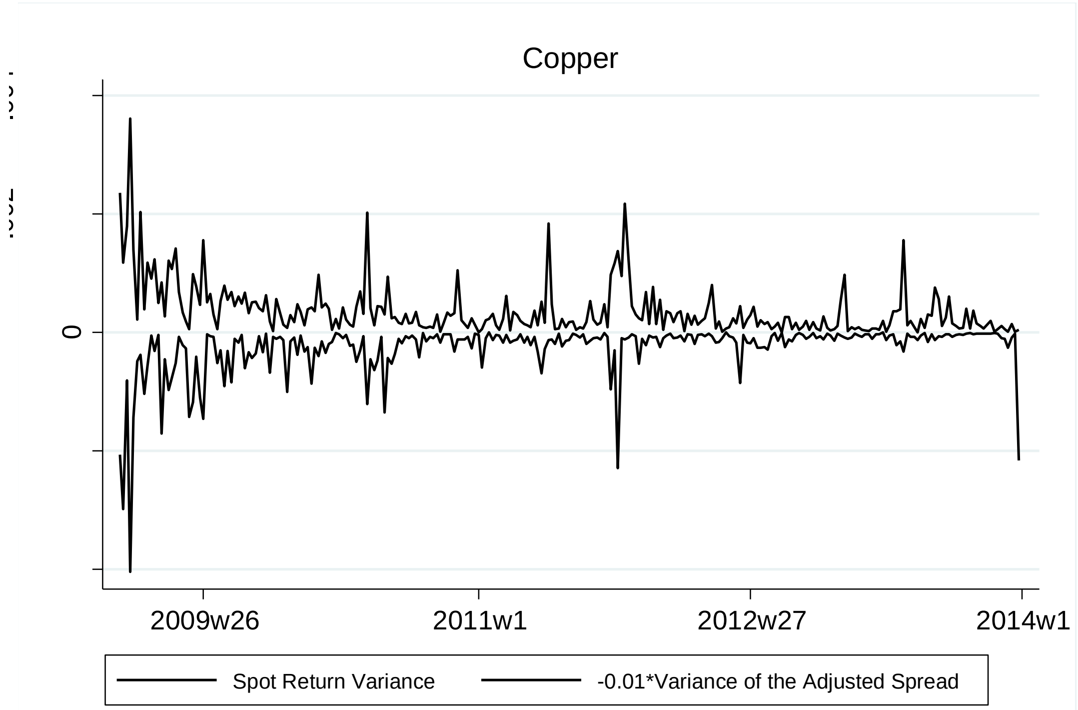

Figure 9: Copper spot return variance, and the variance of the adjusted spread.

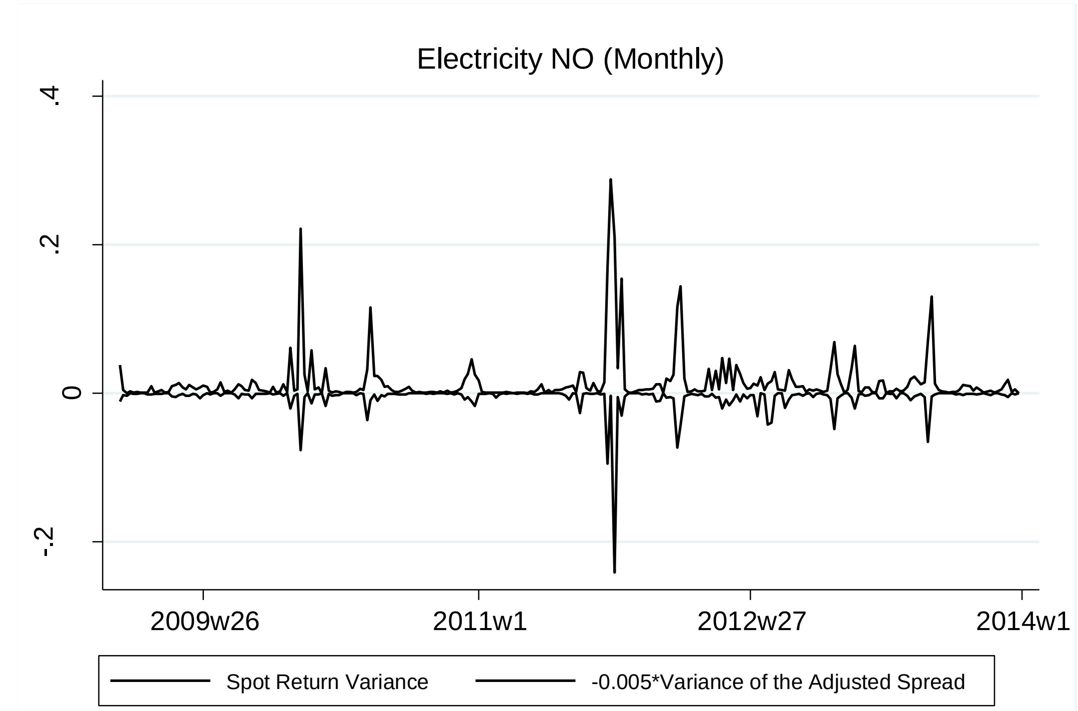

Figure 10: Electricity NO (Monthly) spot return variance, and the variance of the adjusted spread.

7. Conclusion

In our thesis we investigate empirically the relevance of fundamental factors (supply & demand conditions and inventory levels), monetary policy and changes in market value of liquidity from futures market on spot return variance of 31 commodities from 2009 to 2013. Following the *theory of storage*, we have used the *adjusted spread* derived from the 1-month generic futures contract and the corresponding spot price, as a proxy for inventory levels and fundamental factors[16]. The three explanatory variables above were empirically investigated within the framework of SUR Model using Feasible Generalized Least Square estimation (GLS) that allows for heteroscedasticity, autocorrelation, and cross panel correlation on our panel data. The results were also compared using the Random Effect model and the Fixed Effect models.

We find that the variance of the adjusted spread is statistically significant and positively related to the variance of the spot return, i.e. the volatility in spot return is expected to increase when inventories become scarce or demand condition shifts upward. Changes in market value of liquidity is also found to be significant, and has a negative impact on the spot return variance. We did not find the interest rate to have any predictive power. Thus our results suggest fundamental factors have an overwhelmingly large impact on the spot return variance as compared to other explanatory variables in our regression.

The same regression models were also applied to two subsets of the original panel: A group of non-storable commodities (electricity and shipping) consisting of 7 panels and a group of precious metals (gold, silver, etc.) consisting of 4 panels. For both groups we find that the variance of the spot return is statistically significant and positively related to the variance of the adjusted spread. This means that the adjusted spread (i.e. inventory levels) not only has explanatory power on the spot return volatility behaviour of storable consumption commodities, but also on the spot price behaviour of the other classes. None of the other two explanatory variables were significant for any of the panels. This result suggests that a broader definition of theory of storage models can be extended to include commodities that are not directly storable as suggested by Routledge et al. (2001) among other. The unused

[16] It implies that the adjusted spread widens as inventories fall relative to demand. (Pirrong & NG 1994)

capacity in these commodities like the committed generation units in electricity and reserve tonnage in shipping have an equivalent role as inventories in conventional commodities in the theory of storage model (Lautier 2009)

Our results are consistent with past research done by Working (1949), Brennan (1991), Fama and French (1987), Pirrong and NG (1994) and Pindyck (1991, 2004) which suggest that adjusted spread(or adjusted spread variance) has a significant explanatory power over spot return volatility (variance). We agree that fundamental factors are the main determinants of spot price volatility. Hence upward shifts in demand curves and lowering of inventory levels should cause higher spot return volatility. Similarly our results on changes in market value of liquidity in futures markets for commodities suggest that market participation may have decreased risk premiums, and hence, the cost of hedging thus reducing price volatility in commodity markets. However the role of monetary policy on spot return volatility is not significant for our model.

References

Alvarez, R. M., Garrett, G. & Lang, P. (1991). Government Partisanship, Labor Organization, and Macroeconomic Performance. *The American Political Science Review*, 85 (2): 539-556.

Amic, E. (2005). The Oil Market as a World Market. In *Commodities and Commodity Derivatives*, pp. 201-224: John Wiley & Sons.

Anzuini, A., Lombardi, M. J. & Pagano, P. (2010). *The Impact of Monetary Policy Shocks on Commodity Prices*. In Bank, E. C. (ed.). Available at: http://www.ecb.europa.eu/pub/pdf/scpwps/ecbwp1232.pdf (accessed: 06.12.2014).

Beck, N. & Katz, J. N. (1995). What to do (and not to do) with Time-Series Cross-Section Data. *American Political Science Review*, 89 (3): 634-647.

Benavides, G. P. (2010). The Theory of Storage and Price Dynamics of Agricultural Commodity Futures: the Case of Corn and Wheat. 1-22 pp.

Benth, F. E. & Meyer-Brandis, T. (2009). The information premium for non-storable commodities. *The Journal of Energy Markets*, 2: 111-140.

Brennan, D. (1991). The Price of Convenience and the Pricing of Commodity Contingent Claims. *Stochastic Models and Options Values*.

Broxson, B., Peterson, S. & Augustine, C. (2006). *Understanding Natural Gas Markets*. Lexecon. Available at: http://www.spectraenergy.com/content/documents/media_resources_pdfs/understanding_natural_gas_markets.pdf (accessed: 17 November 2014).

Buyuksahin, B., Brunetti, C. & Harris, J. (2009). *Is Speculation Destabilizing?* . U.S. Commodity Futures Trading Commission. Available at: http://www.cftc.gov/ucm/groups/public/@swaps/documents/file/plstudy_03_cftc.pdf (accessed: 14.12.2014).

Campbell, J. Y., Lettau, M., Malkiel, B. G. & Xu, Y. (2001). Have Individual Stocks Become More Volatile? An Empirical Exploration of Idiosyncratic Risk. *The Journal of Finance*, 56 (1): 1-43.

Carpantier, J.-F. & Dufays, A. (2013). Commodity Inventory Effect.

CMEgroup. (2014). *Daily Exchange Volume and Open Interest*. Available at: http://www.cmegroup.com/market-data/volume-open-interest/exchange-volume.html (accessed: 17 November 2014).

Deaton, A. & Laroque, G. (1992). On the Behaviour of Commodity Prices. *The Review of Economic Studies*, 59: 1-23.

Driscoll, J. C. & Kraay, A. C. (1997). Consistent Covariance Matrix Estimation with Spatially-Dependent Panel Data. *Review of Economics and Statistics*, 80 (4): 549-560.

Drukker, D. M. (2003). Testing for serial correlation in linear panel–data models. *The Stata Journal*, 3: 168-177.

Dwyer, A., Gardner, G. & Williams, T. (2011). Global Commodity Markets – Price Volatility and Financialisation: Reserve Bank of Australia, International Department. 49-57 pp.

Fama, E. F. & French, K. R. (1987). Commodity Futures Prices: Some Evidence on Forecast Power, Premiums, and the Theory of Storage. *Journal of Business*, 60: 55-73.

Frankel, J. (2006). The Effect of Monetary Policy on Real Commodity Prices. *Asset Prices and Monetary Policy*: 291-333.

Geman, H. (2005a). Equilibrium Relationships between Spot and Forward Prices. In *Commodities and Commodity Derivatives*, pp. 23-45: John Wiley & Sons.

Geman, H. (2005b). Fundamentals of Commodity Spot and Futures Markets: Instruments, Exchanges and Strategies. In *Commodities and Commodity Derivatives*, pp. 1-19: John Wiley & Sons.

Geman, H. (2005c). Spot and Forward for Electricity Markets. In *Commodities and Commodity Derivatives*, pp. 251-269: John Wiley & Sons.

Gorton, G. & Rouwenhorst, K. G. (2005). Facts and Fantasies about Commodity Futures: Yale University. 1-40 pp.

Greene, W. H. (2003). *Econometric Analysis*, vol. 5. Upper Saddle River, New Jersey: Pearson Education.

Hoyos, R. E. D. & Sarafidis, V. (2013). Testing for Cross-sectional Dependence in Panel Data Models. *The Stata Journal*, 6 (4): 482-496.

Hull, J. C. (2012a). Determination of Forward and Futures Prices. In Battista, D. (ed.) vol. 8 *Option, Futures, and other Derivatives*, pp. 101-127. University of Toronto: Pearson Education.

Hull, J. C. (2012b). Energy and Commodity Derivatives. In Battista, D. (ed.) vol. 8 *Option, Futures, and other Derivatives*, pp. 748-764. University of Toronto: Pearson Education.

IMF. (2011). A Barrel of Oil or a Bottle of Wine: How Do Global Growth Dynamics Affect Commodity Prices? . *IMF Working Paper*, 11 (1): 1-19.

Irwin, S. H. & Sanders, D. R. (2012). Financialization and Structural Change in Commodity Futures Markets. *Journal of Agricultural and Applied Economics*, 44 (3): 371–396.

Kaldor, N. (1939). Welfare propositions of economics and interpersonal comparisons of utility. *Economic Journal*, 49: 549–552.

Lautier, D. (2009). Convenience Yield and Commodity Markets. *Les Cahiers de la Chaire*, 22.

OECD. (2010). The Impact of Index and Swap Funds on Commodity Futures Markets. *OECD Food, Agriculture and Fisheries Working Papers* 27: 1-31.

Pindyck, R. S. (1990). Inventories and the Short Term Dynamics of Commodity Prices. *Massachusetts Institute of Technology*.

Pindyck, R. S. (2004). Volatility and Commodity Price Dynamics. *The Journal of Futures Markets*, 24: 1029–1047.

Pirrong, C. & NG, V. K. (1994). Fundamentals and Volatility: Storage, Spreads, and the Dynamics of Metals Prices. *Journal of Business*, 67: 203-230.

Routledge, B. R., Spat, C. S. & Seppi, D. J. (2001). The "Spark Spread:" An Equilibrium Model of Cross-Commodity Price Relationships in Electricity: Tepper School of Business.

Williams , J. C. & Wright, B. D. (1982). The Economic Role of Commodity Storage. *The Economic Journal*, 92: 596-614.

Williams , J. C. & Wright, B. D. (1991). *Storage and commodity markets*, vol. 1: Cambridge University Press.

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*. 2 ed., vol. 1. Michigan State University: The MIT Press.

Working, H. (1933). Price Relations Between July and September Wheat Futures at Chicago Since 1885. Stanford University: Food Research Institute

Working, H. (1949). The Theory of Price of Storag. *American Economic Review*, 39: 1254-1262.

8. Appendices

8.1 Appendix A

Missing dates

Contract	Spot	Futures	Open Int
Tanker Route TC 2			
Oil Brent			
Oil WTI			
Natural Gas			29.07.2009
			27.08.2010
			24.11.2010
Gasoline			16.01.2009
			30.01.2009
			27.02.2009
			31.03.2009
			30.04.2009
			29.05.2009
			30.06.2009
			31.07.2009
			30.09.2009
			30.10.2009
			26.02.2010
			30.04.2010
			28.05.2010
			30.07.2010
			30.08.2013
Heating Oil			30.01.2009
			27.02.2009
			31.03.2009
			30.04.2009
			29.05.2009
			30.06.2009
			31.07.2009
			30.10.2009
			26.02.2010
			30.04.2010
			28.05.2010
			30.07.2010
			30.08.2013
Ethanol	29.05.2012		30.09.2009
Ethanol	23.10.2012		19.02.2010
	05.11.2012		26.02.2010
	07.03.2013		30.04.2010
			28.05.2010
			04.06.2010
			25.06.2010
			09.07.2010
			16.07.2010
			23.07.2010
			30.07.2010
			20.08.2010
			27.08.2010
			10.09.2010
			17.09.2010
			24.09.2010
			30.09.2010
			30.08.2013
Corn			13.03.2009
			12.03.2010
Soybean			13.03.2009
			14.08.2009
			14.09.2009
			13.11.2009
			12.03.2010
			14.05.2010
			13.08.2010
			14.01.2011
			13.01.2012
			13.07.2012
Wheat			13.03.2009
			12.03.2010
			14.05.2010
Gold			25.02.2009
			28.04.2009
			26.06.2009

07.06.2012 30.10.2009 Silver 28.01.2009

Contract	Spot	Futures	Open Int
Silver			27.03.2009
			27.05.2009
			29.07.2009
			28.09.2009
			27.01.2010
Platinum			28.01.2009
			28.04.2009
			29.07.2009
			27.04.2011
			26.07.2011
Palladium			27.03.2009
			26.06.2009
			28.06.2010
			29.12.2010
			28.12.2011
Copper			27.03.2009
			27.05.2009
			29.07.2009
			28.09.2009
Al, Zi, Ni, Pb			16.03.2009
			18.05.2009
			22.05.2009
			15.06.2009
			13.07.2009
			17.08.2009
			14.09.2009
			19.10.2009
			16.11.2009
			14.12.2009
			15.03.2010
			19.04.2010
			17.05.2010
			14.06.2010
			19.07.2010
			13.08.2010
			16.08.2010
			13.09.2010
			18.10.2010
			15.11.2010
			13.12.2010

Contract	Spot	Futures	Open Int
Al, Zi, Ni, Pb			14.02.2011
			14.03.2011
			18.04.2011
			16.05.2011
			18.07.2011
			15.08.2011
			19.09.2011
			17.10.2011
			14.11.2011
			19.12.2011
			13.02.2012
			19.03.2012
			16.04.2012
			14.05.2012
			18.06.2012
			16.07.2012
			13.08.2012
			17.09.2012
			15.10.2012
			19.11.2012
Soybean Oil			13.03.2009
			14.08.2009
			12.03.2010
			14.05.2010
			13.08.2010
			14.01.2011
			13.01.2012
Soybean Meal	06.04.2011		13.03.2009
	23.08.2011		14.05.2009
	25.11.2011		14.07.2009
	17.07.2012		14.08.2009
	09.11.2012		12.03.2010
	23.11.2012		14.05.2010
	08.01.2013		13.08.2010
	29.11.2013		14.01.2011
			13.01.2012
			13.07.2012
			14.08.2012
			14.05.2013
			13.09.2013

Contract	Spot	Futures	Open Int
Lean Hogs	11.03.2009		13.02.2009
	20.07.2009		15.04.2009
	12.02.2010		14.05.2009
	24.02.2010		12.06.2009
	02.12.2010		15.07.2009
	18.08.2011		14.08.2009
	23.08.2011		12.02.2010
	16.11.2011		14.05.2010
	12.12.2011		13.08.2010
	28.02.2012		
	03.04.2012		
	13.08.2012		
	10.09.2012		
	08.01.2013		
	01.03.2013		
	27.06.2013		
	23.07.2013		
	23.08.2013		
	30.08.2013		
	12.12.2013		
Sugar	02.04.2009		27.02.2009
	22.05.2009		30.04.2009
			30.06.2009
			30.09.2009
Coffee (Arabica)	06.02.2009	13.02.2009	13.02.2009
	30.04.2010	16.12.2010	18.03.2009
	21.12.2012	30.12.2010	15.05.2009
	17.02.2009	03.02.2011	16.07.2009
		11.05.2011	14.09.2009
		15.06.2011	11.12.2009
		17.08.2011	30.12.2010
		07.11.2011	03.02.2011
		30.01.2012	11.05.2011
		16.03.2012	15.06.2011
		13.06.2012	17.08.2011
		13.08.2012	07.11.2011
		01.11.2012	30.01.2012
		02.05.2013	16.03.2012

Contract	Spot	Futures	Open Int
Coffee (Arabica)		11.06.2013	27.04.2012
		08.08.2013	13.06.2012
		05.11.2013	28.06.2012
			13.08.2012
			01.11.2012
			02.05.2013
			11.06.2013
			08.08.2013
			05.11.2013
			25.11.2013
Cotton			09.03.2009
			06.05.2009
			09.07.2009
			08.10.2009
			09.10.2012
Electricity DE (M,Q,Y)			30.01.2009
			27.02.2009
			30.03.2009
			29.04.2009
			29.05.2009
			29.06.2009
			30.07.2009
			29.10.2009
			30.10.2009
			27.11.2009
			30.11.2009
			29.12.2009
			30.12.2009
			28.01.2010
			29.01.2010
			25.02.2010
			26.02.2010
			29.03.2010
			30.03.2010
			28.04.2010
			29.04.2010
			28.05.2010
			28.06.2010

Contract	Spot	Futures	Open Int
Electricity DE (M,Q,Y)			29.06.2010
			29.07.2010
			30.07.2010
			27.08.2010
			28.09.2010
			29.09.2010
			28.10.2010
			29.10.2010
			30.12.2010
			29.04.2011
			29.06.2011
			29.07.2011
			30.08.2011
			29.09.2011
			11.10.2011
			31.10.2011
			29.11.2011
			30.12.2011
			30.01.2012
			28.02.2012
			02.03.2012
			30.03.2012
			20.04.2012
			30.05.2012
			29.06.2012
			30.07.2012
			30.08.2012
			28.09.2012
			30.10.2012
			29.11.2012
			28.12.2012
			30.01.2013
			27.02.2013
			29.04.2013
			30.05.2013
			28.06.2013
			30.07.2013
			30.08.2013

Contract	Spot	Futures	Open Int
Electricity DE (M,Q,Y)			30.09.2013
			29.11.2013
Electricity NO (M,Q,Y)		17.05.2011	22.05.2009
		05.04.2012	28.05.2009
		17.05.2013	02.06.2009
			23.07.2009
			04.04.2011
			26.03.2012
			29.07.2013

8.2 Appendix B

Bloomberg: Commodities with belonging Spot and 1-month Generic Futures.

			Spot
Commodity	**Bloomberg Ticker Code**	**Exchange**	**Price**
Shipping			
Tanker Route TC 2	TANKRATB	NYMEX	USD/Mt Tonnes
Energy			
Oil Brent	COY	ICE	USD/Barrel
Oil WTI	USCRWTIC	Blm	USD/Barrel
Natural Gas	NGUSHHUB	ICE	USD/MMBtu
Gasoline	DOE	NYMEX	USD/Gallon
Heating Oil	DOE	NYMEX	USD/Gallon
Propanol	DOE	CBOT	USD/Gallon
Metals (Investment)			
Gold	GOLDLNPM	CMX	USD/troy oz
Silver	SLVRLND	CMX	USD/troy oz
Platinum	PLTMLNPM	NYM	USD/troy oz
Palladium	PLDMLNPM	NYM	USD/troy oz
Metals (Consumption)			
Copper	LMCADS03	CMX	USD/lb
Aluminium	LMAHDS03	LME	USD/Mt Tonnes
Zinc	LMZSDS03	LME	USD/Mt Tonnes
Nickel	LMNIDS03	LME	USD/Mt Tonnes
Lead	LMPBDS03	LME	USD/Mt Tonnes
Agricultural Products			
Corn	CORNILNC	CBT	USD/Bushel
Soybean	SOYBCH1Y	CBT	USD/Bushel
Wheat	WEATTKHR	CBT	USD/Bushel
Soybean Oil	SOYPIOIL	CBT	USD/lbs
Soybean Meal	SOYPIT48	CBT	USD/lbs
Lean Hogs	ISOSDALY	NYB-ICE	USD/lbs
Sugar	COFECMNY	NYB-Ice	USD/lbs
Coffee	HOGSNATL Index	CME	USD/lbs
Cotton	COTNMAVG	NYB-Ice	USD/lbs
Electricity			
Electricity DE	LPXBHRBS	EEE	EUR/MwH
Electricity NO	ENWSSPAV	NPE	EUR/MwH

Generic 1 month Futures contracts/Constant Maturity

Commodity	Bloomberg Ticker Code	Exchange	Price	Contract Size
Shipping				
Tanker Route TC 2	OX1	NYMEX	USD/Mt Tonnes	1000 MT Tons
Energy				
Oil Brent	CO1	ICE	USD/Barrel	1000 Barrel
Oil WTI	CL1	NYMEX	USD/Barrel	1000 Barrel
Natural Gas	NG1	NYMEX	USD/MMBtu	10,000 MMBtu
Gasoline	XB1	NYMEX	USD/Gallon	42,000 US Gallon
Heating Oil	HO1	NYMEX	USD/Gallon	42,000 US Gallon
Propanol	BAP1	NYMEX	USD/Gallon	42,000 US Gallon
Metals (Investment)				
Gold	GC1	CMX	USD/troy oz	100 troy oz
Silver	SI1	CMX	USD/troy oz	5000 troy oz
Platinum	PL1	NYM	USD/troy oz	50 troy oz
Palladium	PA1	NYM	USD/troy oz	100 troy oz
Metals (Consumption)				
Copper	HG1	CMX	USD/lb	25,000 lb
Aluminium	LA1	LME	USD/Mt Tonnes	25 MT
Zinc	LX1	LME	USD/Mt Tonnes	26 MT
Nickel	LN1	LME	USD/Mt Tonnes	6 MT
Lead	LL1	LME	USD/Mt Tonnes	25 MT
Agricultural Products				
Corn	C1	CBT	USD/Bushel	5000 Bushel
Soybean	S1	CBT	USD/Bushel	5000 Bushel
Wheat	W1	CBT	USD/Bushel	5000 Bushel
Soybean Oil	BO1	CBT	USD/lbs	60,000 lbs
Soybean Meal	SM1	CBT	USD/tones	100 short tons
Lean Hogs	LH1	CME	USD/lbs	40,000 lbs
Sugar	SB1	NYB-ICE	USD/lbs	112,000 lbs
Coffee	KC1	NYB-Ice	USD/lbs	37,500lbs
Cotton	CT1	NYB-Ice	USD/lbs	50,000 lbs
Electricity[17]				
Electricity DE (Monthly)	GI1	EEE	EUR/MwH	720 MwH
Electricity DE (Quarterly)	GT1	EEE	EUR/MwH	2,184 MwH
Electricity DE(Yearly)	HP1	EEE	EUR/MwH	8,784 MwH
Electricity NO (Monthly)	NEL1M	NPE	EUR/MwH	720 MwH
Electricity NO (Quarterly)	NEL1Q	NPE	EUR/MwH	2,184 MwH
Electricity NO(Yearly)	NEL1Y	NPE	EUR/MwH	8,784 MwH

[17] For electricity we use three different 1-month generic Futures, only distinguished by the size of the contract. The contracts are claims on a flow of power delivered at a constant rate for 1 month (720 MwH), 3 months (2 184 MwH) or 1 year (8 784 MwH).