# Changing patterns of price evolution in commodity markets

**Can it be explained by a rational commodity pricing model?**

by

Holler Zsuzsa

Submitted to

Central European University

Department of Economics

In partial fulfilment of the requirements for the degree of

Master of Arts

Supervisor: Professor Kondor Péter

Budapest, Hungary

2012

I, the undersigned [Holler Zsuzsa], candidate for the degree of Master of Arts at the Central European University Department of Economics, declare herewith that the present thesis is exclusively my own work, based on my research and only such external information as properly credited in notes and bibliography. I declare that no unidentified and illegitimate use was made of works of others, and no part the thesis infringes on any person's or intstitution's copyright. I also declare that no part the thesis has been submitted in this form to any other institution of higher education for an academic degree.

Budapest, 4 June 2012

______________________________

Signature

# Abstract

This empirical work aims to contribute to the debate on the possible causes of increased volatility and rising prices in commodity markets, and whether the role of financial investors is significant or only the changes of fundamental factors induced the changed behavior of prices in commodity markets. My empirical strategy is to test the adequacy of a rational commodity pricing model on the 1991 - 2012 period of wheat and crude oil markets in order to gain evidence on or against different theories explaining the observed events. I estimate linear regressions and vector autoregressive models on commodity prices, futures prices and other relevant variables, and test restrictions implied by the theory. I also use a test to find structural breaks with unknown date in the data. Finally I examine the consistency of the hypotheses on the possible driving forces of commodity prices within the model. My results show that the rational commodity pricing model I apply fits well on wheat but not on oil data. Test results on oil suggest that the assumptions of the rational commodity pricing model are not satisfied. Results for the wheat market suggest that wheat prices might have been affected by the activity of financial investors, but the price couldn't have experienced such a huge spike without suddenly increased demand on spot markets caused by some kind of panic, and without the effect of liquidity problems of market participants. Information on future price movements seems to have become less reliable before the spike, which suggests that fake signals caused by the behavior of financial investors could have

influenced price movements. However, the approach I use is not capable to provide strong evidence on this point because extraordinary events of the 2007-2008 period make hard to filter out the relevant information.

# Contents

# List of Tables

# List of Figures

# Introduction

This empirical work aims to contribute to the debate on the possible causes of recent episodes in commodity markets. My empirical strategy is to test the adequacy of a rational commodity pricing model on wheat and crude oil markets in order to gain evidence on or against different theories explaining the observed events.

Since the beginning of the twentieth century a growing literature of the features of commodity markets evolved. Steeply increasing commodity prices have drawn attention to the role of these markets in the global economy. Several hypotheses were born to explain the rising prices, the high volatility and the increasing volume of trade in commodity markets observed in the past decade. Detailed analysis of the potential contributing factors like demand growth of emerging Asian economies, decreased oil supply of the OPEC countries and climate change can be found in the literature [1].

However, in the past few years the effect of another major factor became the center of attention. The effect of increased volume of financial investment in commodity markets became a serious concern of legislators and market practitioners. The magnitude of the increase in the participation of financial investors in commodity futures markets is debated, but as Irwin and Sanders (2011) states, between 2004 and 2008 at least 100 billion dollar of



[1]Detailed discussion of fundamental factors can be found in a recent study prepared by the secretariat of the United Nations Conference on Trade and Development Unc (2011)

new investment was channeled into the commodity futures markets. The financialization [2] of commodity markets and the influence of this procedure on the pricing mechanisms were posed as a possible cause of the changes. According to this explanation, the changed behavior of the market participants can take the commodity prices away from their natural levels and can lead to commodity price bubbles.

My thesis aims to contribute to the ongoing debate on the causes of changed behavior of commodity prices. The question is important from a policy and regulatory perspective. Since oil and articles of foods are affected seriously, there is a high pressure on policy makers to find explanation and solution to the problems caused by extremely volatile and unpredictable commodity price movements. A recently passed act authorized the Commodity Futures Trading Commission, the agency of the United States government that regulates futures and option markets, to impose limits on futures and swap positions in given commodity markets, even though the role of financial traders is not clarified yet. As long as there is no strong empirical evidence on the reasons of recent changes, it is hard to forecast the effects of these actions. Limits on speculation can ruin the efficient functioning of futures markets, and can have more disadvantages than advantages.

The goal of my study is to investigate whether the hypothesis that the activity of financial investors influence commodity prices can be supported by the data on wheat and crude oil prices. The approach I use is based on the concept of convenience yield and on the relationship between prices, inventory levels and convenience yield. I assume that the price of a commodity in the spot market is determined by an exact, present value relationship, then I test the model against different alternatives, finally I examine the consistency of the hypotheses on the possible driving forces of commodity prices within the model.

Since the test results show that the model I apply fits well on wheat but not on oil, I

[2]The term arises from Domanski and Heath (2007)

can make stronger conclusions on the driving forces of wheat prices. For oil prices, I am able to provide only intuitions suggested by the behavior of the analysed variables.

The final lesson of my investigation is that wheat prices might have been affected by the activity of financial investors, but the price couldn't have experienced such a huge spike without suddenly increased demand caused by some kind of panic, and without the effect of liquidity problems of market participants. My results also show that information on future price movements became less reliable before the spike, which suggest that fake signals caused by the behavior of financial investors might have influenced price movements. However, the approach I use is not capable to provide strong evidence on this point because extraordinary events of the 2007-2008 period make it hard to filter out the relevant information.

The rest of my thesis is organized as follows. First I give a summary of the recent episodes of the history of commodity markets and I review the literature dealing with the topic. I summarize some of those empirical papers that attempt to find evidence supporting one or the other theory. At the end of this chapter, I explain the intuition behind the approach I use to examine the possible sources of changes in commodity markets. In Chapter 2 I give a detailed description of the theory and I discuss the empirical strategy I use. Data description and results are presented in Chapter 3. Finally I draw conclusions.

# Chapter 1

# Facts and literature on financialization

## 1.1 Recent episodes in commodity markets

Figure 1.1 shows the evolution of crude oil [1] and spring wheat [2] prices between the beginning of 1991 and 2012. It can be seen that both wheat and crude oil price reached its all-time high in 2008, and then experienced a sudden fall in the outbreak of the financial crisis. It is not surprising that these extremely high price changes, without precedent in the history of commodity markets have drawn the attention of both policy makers and scientists. The debate on the importance of fundamental factors and on the role of commodity futures markets still proceeds. Several studies were conducted in recent years that tried to explain the evolution of commodity prices and the increased volatility but as I mentioned previously, the role of fundamentals like increased demand or decreased



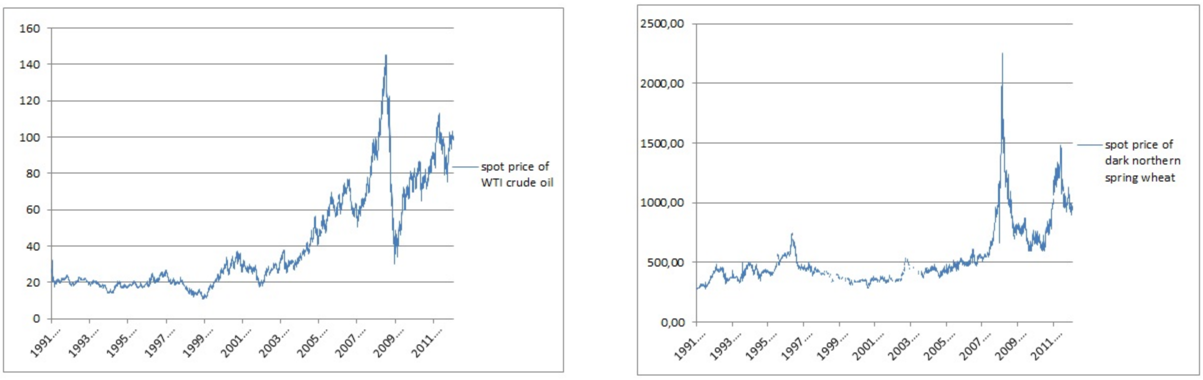
[1] Daily closing price of one barrel WTI crude oil delivered in Cushing, Oklahoma.

[2] Average daily spot price of one bushel 14% Dark Northern Spring Wheat measured in dollar cents.

supply and other factors remained controversial. There are several papers discussing the contradictions in the arguments for the role of financial investors (see for example Krugman (2008) or Hamilton (2009)) and providing mixed or conflicting empirical evidence (see for example Brunetti and Buyuksahin (2009)). Among these studies I discuss three papers with different approaches and empirical strategy in detail. I have chosen these papers because as I see, they demonstrate the most promising procedures that are used to provide empirical evidence on the topic.

Figure 1.1: Evolution of WTI crude oil and spring wheat prices, 1991-2012

(a) (b)




One way to test the role of different factors is to examine the correlation between the returns on assets with unrelated fundamentals. This methodology is followed by Tang and Xiong (2010). Increased cross market correlations suggest that something changed in the commodity markets, but there could be several causes that can be hard to separate. Common, economy-wide shocks can affect the price of assets with unrelated fundamentals, thus the fact in itself that correlations increased does not provide evidence for or against the effect of financialization. The methodology of Tang and Xiong overcomes this problem. As they focused on the effect of index investors, they examined whether the change in cross-market correlation is higher for commodities that are included in major indices. Their results support the influence of index investors on commodity price movements, they found significant difference between commodities included in indices and commodities not included in the indices.

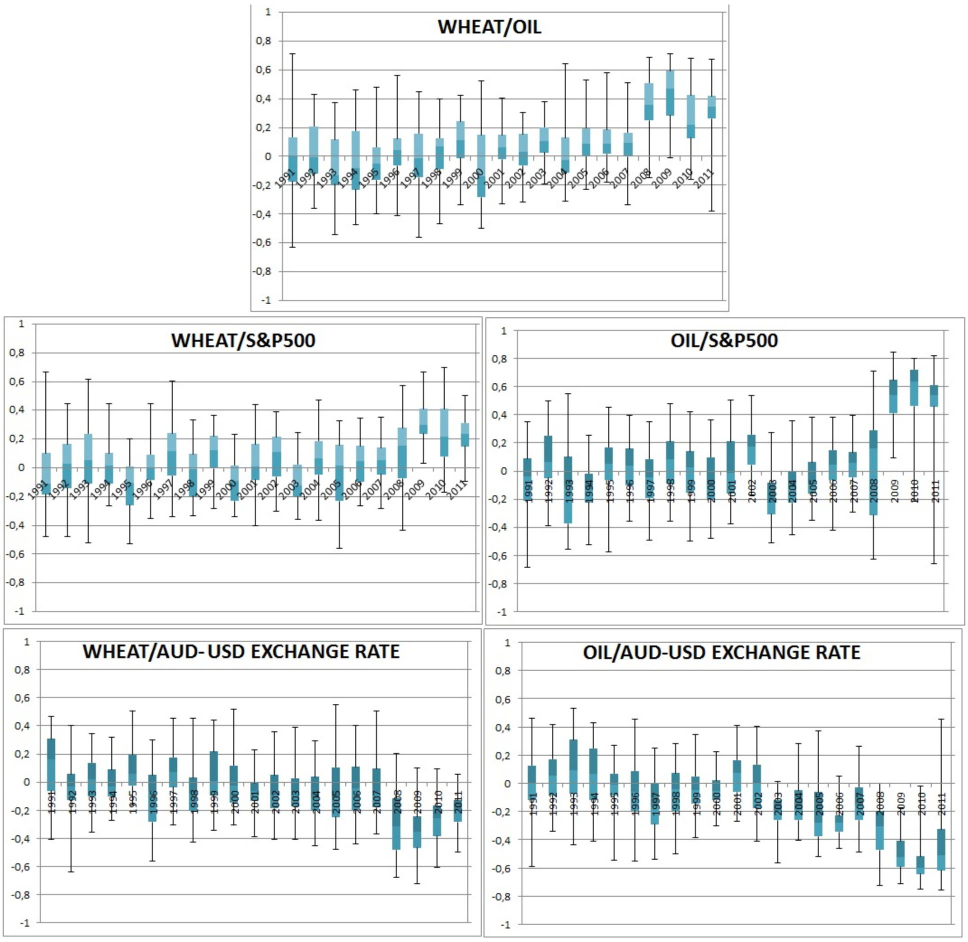
To illustrate the findings of Tang and Xiong (2010), in figure 1.2 I present the box-plot statistics of the thirty day rolling correlations between return on wheat and return on crude oil, the S&P500 index, and Australian-dollar US-dollar exchange rate. The figure shows that correlations between returns increased significantly during the past few years. For example, in the case of wheat and oil, while the median of thirty day rolling correlations is around 0.1 in 2007, it is higher than 0.3 in 2011. As Tang and Xiong showed, this increase is significantly higher among those commodities that are included in major indexes, thus probably caused by the activity of index investors. As I see, while this methodology is very creative, it has a major shortcoming. It concentrates on the influence of index investors, but does not help us to analyse the effect of other types of financial investors on commodity markets. In my analysis I do not separate the effect of different types of financial investors, I examine the influence of the episodes in futures market in general.

The second type of papers tries to support the view that the behavior of financial investors influences the evolution of commodity prices by relating position taking of different types of investors and returns on futures contracts. Gilbert (2009), who followed this methodology, found that position taking of index investors affects the price of commodities and he also gave an estimate on price development that would have prevailed without index investors. However he tends to interpret his results as evidence on the transmitting rather than leading role of futures markets.

The third methodology combines the previous two. Buyuksahin and Robe (2010) analyzed whether positions of different types of financial investors [3] are related to cross-market

[3] They applied the commodity futures market traders classification used in U.S. Commodity Futures Trading Commission's Disaggregated Commitment of Traders Reports.

Figure 1.2: Box plots of thirty day rolling correlations by year, 1991-2011




correlations in commodity futures markets. They found positive relationship between the presence of financial speculators and co-movement of asset returns, but they did not find any relationship between the participation of other kinds of financial investors - like index traders - and cross-market correlations. While their empirical strategy is very promising, it requires detailed data on the position taking of commodity futures market participants that is not available publicly, moreover it has been collected only for a few years. Consequently this methodology can not be used to confront the behavior of commodity prices in the recent years and in previous time periods, thus can not reveal how these patterns changed and how these changes influenced commodity price movements.

The final goal of my study is very similar to that of the discussed papers. I investigate whether the hypothesis that the activity of financial investors influence commodity prices can be supported by the data. On the other hand, my approach is more straightforward. I assume that the price of a commodity in the spot market is determined by an exact relationship, then I test the model against different alternatives, finally I state the implications of the model to gain evidence on the driving forces of commodity prices.

The intuition behind my approach is the following. The level of inventories of a storable commodity carries useful information on the changes in demand and supply and on the expectations of market participants. Increasing inventory level can indicate increasing net supply or an expected price increase. A sudden drop in inventories suggests that something unexpected happened. On the contrary, constant level of stocks indicates that the market is in an equilibrium in some sense. Since it is nearly impossible to properly measure the level of inventories, it is hard to make strong conclusions based on stock data.

Fortunately there is a theory that states that it is possible to deduce how inventory levels change from the value of the so called convenience yield. Marginal convenience yield measures the extra utility that an individual derives by holding an extra unit of stock instead of holding a futures contract. According to the theory, marginal convenience yield decreases as the level of inventories increases, thus the change in marginal convenience yield indicates how stock levels changes. Moreover, based on the no-arbitrage condition on futures and spot prices, one can compute the level of marginal convenience yield at a given period, thus futures and spot commodity price data is sufficient to infer inventory level dynamics.

To sum it up, I use data on futures and spot commodity prices to determine the level of marginal convenience yield and to draw conclusions on the activity of stockholders. Then I use the information on the activity of stockholders to gain evidence on the changes in fundamentals and on the changes in other factors that affect commodity prices.

In the next subsection I discuss the concept of convenience yield and the related theory, then I move on to the rational commodity pricing model I use in my analysis which is based on the concept of convenience yield.

## 1.2 Convenience yield

According to the original definition, convenience yield is the utility obtained from holding the commodity but not obtained from holding a futures contract. The concept of convenience yield was first proposed by Kaldor (1939). He proposed the concept to explain the phenomenon of backwardation that is often observed in commodity markets. Backwardation in this context means that futures prices become lower as maturity approaches.

In the case of goods that are held only for investment purposes, the no-arbitrage condition guarantees that the price of the good in the spot market and the price of the futures contract on the same commodity are tied to each other in the long run. For no arbitrage opportunities to exist, the futures price discounted by the risk free interest rate has to be equal to the spot price plus the cost of storage [4].

Since many types of commodities are not held originally for investment purposes, the standard arguments that support the described relationship of spot and futures prices might fail to work for those goods. The spot price of a commodity can be higher than the futures price discounted by the risk free interest rate and storage costs in the long-run without providing arbitrage opportunities. The following argument is provided by Hull (1993) and it is based on the fact that these goods are used in production and also have consumption value, thus market participants hold the commodity for other reasons than obtaining monetary gain.

[4] The exact relationship depends on the assumptions on the nature of the interest rate process and the storage costs.

Suppose that the spot price is so high that it would be possible to gain riskless profit by selling the commodity, saving the storage costs, investing the money at the risk free interest rate and buying the futures contract. If the owners of the commodity obtain higher utility by holding inventories than by gaining monetary profit, then they might refuse to sell their stocks. As a consequence, the spot price may remain high relative to the futures price in the long run.

Convenience yield simply measures the difference between the discounted futures price and the spot price as a consequence of the consumable nature of commodities. However, depending on whether one treats convenience yield as an additive factor or a multiplicative factor, and on the assumptions on the nature of storage costs, the exact definition of convenience yield might change. In empirical works net convenience yield defined as convenience yield minus physical storage costs is often used instead of convenience yield because of lack and imprecision of information on storage costs.

The relationship of commodity prices, stock levels and convenience yield is developed in the so called theory of storage models. The most influential theory of storage type models that are usually referred are provided by Telser (1958) and Brennan (1958). These models are relatively simple, two period, partial equilibrium models that describe how price expectations, prices and inventory levels should be connected in equilibrium. The most important statements of these models are the following. At first, they state that marginal convenience yield and the level of inventories are inversely related, more precisely, marginal convenience yield is a convex, decreasing function of inventory level. This assumption was supported empirically later by Pindyck (1990). They also stated that the market of storage is in equilibrium if the expected future price of the commodity equals the present price times the expected return minus the marginal convenience yield, otherwise market participants have an incentive to increase or decrease their stocks. The rational commodity pricing model I use is in my analysis is basically based on these two, simple assumptions.

The idea to apply the theory of convenience yield is originally proposed by Einloth (2009). Although I think that Einloth reveals some interesting features of the recent episodes in commodity markets by applying a simple model of storage, he does not provide a really sophisticated theory on the relationship between spot prices, futures prices and convenience yield. The basic idea behind Einloth's model is the same as I use, but because he use a static model to explain price movements he simply neglects the effect of market participants' expectations on commodity price movements. As the rational commodity pricing model I use is a simplified version of a dynamic market equilibrium model, I am able to take into account the role of changing expectations in price formation.

In the next chapter I present the model I use to analyse recent changes in commodity market and I explain the empirical strategy I use.

# Chapter 2

# The Pindyck model of rational commodity pricing

## 2.1 The model

Based on the traditional theory of storage models, Pindyck (1993) establishes a rational commodity pricing model. He applies the present value relation to storable commodities treating convenience yield as the analogue of dividend of stocks. His model gives an explanation why agents hold inventories when the expected capital gain from holding stock is lower than the risk adjusted rate or even negative.

Pindyck (1993) argues that this model is a highly reduced form of a dynamic supply and demand model though the factors of supply and demand, and the market equilibrium conditions are not included explicitly. As long as there are market participants holding stocks of a commodity, there is no need to model supply and demand to determine the equilibrium level of price.

Pindyck's model (Pindyck (1993)) is founded on the same relationship as the theory of storage models. Stockholders incentives guarantee that – as long as there is no stockout - the expected next period price of a commodity equals the present price times the expected rate of return minus the net marginal convenience yield:

$$E_t(P_{t+1}) = P_t(1+\mu) - MCY_t, \tag{2.1}$$

where $\mu$ denotes the one period commodity specific discount rate and $MCY_t$ is the one period net marginal convenience yield. Pindyck (1993) assumes that commodity specific discount rate is constant over time and $\mu = r + \rho$, where $r$ denotes the risk free interest rate and $\rho$ is the risk premium. Expectations denoted by $E_t$ are the mathematical expectations conditional on the full public information set available at time t, which includes every relevant information used to predict commodity prices.

In the context of present value relationship literature this correspondence is derived from a simple identity, the definition of return on an asset. By definition the period $t$ return on a stock is the ratio of the $t+1$ period price plus the dividend and the period t price. However, in theory, the consistency of expectations guarantees that this relationship also holds ex ante. By taking expectations, one can get the analogue of the relationship given in equation 2.1 for stock prices.

In the theory of storage type models, this relationship is obtained in a different way. Kaldor (1939) and Telser (1958) gain this equality as the equilibrium condition of storage, and in their context expectations are simply the subjective anticipations of market participants. However, in general equilibrium, they require the second period price and the expected second period price to be equal. This condition is based on the assumption that market participants can't survive in the long run if they permanently over- or underestimate future prices.

Both of these approaches suggest that an empirical test of this relationship should use mathematical expectations conditional on the t period information set, since the efficiency of a market ensures that market participants use every relevant information to predict prices, at least in the long-run.

It can be shown that the solution of this difference equation is the following:

$$P_t = \delta \sum_{i=0}^{\infty} \delta^i E_t MCY_{t+i}, \tag{2.2}$$

where $\rho = \frac{1}{1+\mu}$. This relationship is the exact counterpart of the present value relationship of stock prices.

This relationship however, and consequently the whole model, holds only if there are no stockouts, since the argument is based on the fact that it is worth it for some agents to hold stocks. If the firms think that there is a positive possibility of a future stockout, this model fails to explain the price evolution.

One can support the no stockout condition of the model with two empirical observations. First, the level of inventories has always been positive for all of the examined commodities since there is documentation on the level of stocks in the United States. In the second place, convenience yield is assumed to be a highly convex function of the level of inventories, thus stockouts have to be accompanied by very high prices. Pindyck in a different paper (Pindyck (1990)) provides empirical evidence on this convexity for different types of commodities. Obviously none of these observations rule out the possibility of a highly unlikely stockout but make the model more credible. Pindyck also makes use of the no-arbitrage condition of forward and spot prices.

In order to avoid arbitrage the following equation has to be satisfied:

$$MCY_{t,T} = (1 + r_T)P_t - f_{T,t}, \tag{2.3}$$

where $f_{T,t}$ is the forward price for the at $t+T$, $MCY_{t,T}$ is the flow of marginal convenience yield net of storage costs over the period $t$ to $t+T$ and $r_T$ is risk-free T-period interest rate. Since futures prices are more available than forward prices, Pindyck (1993) uses futures prices instead of forward prices. He argues that if the risk free interest rate is stochastic then there can be differences between futures and forward prices but the difference for most commodities is small. I follow his method in my analysis and disregard the differences between futures and forward prices.

Pindyck (1993) uses the results of Campbell and Shiller (1987) to deduce the implications of present value relationship on commodity spot and futures prices. He shows that if price and net marginal convenience yield are both integrated of order one, then the following variable has to be stationary, thus futures price and spot price have to be cointegrated.

$$S_t = F_{1,t} - (1-\rho)P_t. \tag{2.4}$$

As a consequence, regressing the spot price on the futures price gives us an estimate of the expected excess return.

It also can be shown that

$$S_t = E_t \Delta P_{t+1} = E_t \sum_{i=1}^{\infty} \delta^i \Delta MCY_{t+i}, \tag{2.5}$$

thus the optimal forecast of $P_{t+1}$ is given by $S_t$ and $P_t$. From this relationship it follows that there should not be any variable in the information set at t that helps to predict price change.

In the next subsections I discuss the empirical methods I used to test this rational commodity pricing model.

## 2.2 Description of the empirical tests

### 2.2.1 Linear regression

The most straightforward way to test the present value relation is to study the basic relation of the model given in equation 2.1. An immediate implication of this relationship is that the difference between the expected value of the future spot price and the realization of the price has to be unpredictable by any of the variables in the information set given at time $t$. This can be easily seen by substituting 2.3 into 2.1:

$$E_t(P_{t+1} - P_t) = F_{1,t} - (1 - \rho)P_t. \tag{2.6}$$

Since $E_t$ denotes mathematical expectations conditional the on period $t$ information set, from 2.6 it follows, that there can't be any variable available in $t$ that predicts price change except the price and the futures price.

To test this relationship I use the regression suggested by Campbell and Shiller (1987):

$$\Delta P_t = \alpha_0 + \alpha_1 P_{t-1} + \alpha_2 F_{t-1,1} + \sum_i \beta_i z_i + \epsilon_t. \tag{2.7}$$

If the model assumptions are satisfied then any included $z_i$ variables have to be insignificant, $\alpha_0$ has to be insignificant, $\alpha_1$ has to be one and $\alpha_2$ has to be equal with $-(1 - \rho)$.

While this is a fairly simple way to provide evidence against the model, the implications of finding variables explaining the unexpected change in price are not clear. The problem with this approach is that there can be several causes explaining the failure of this test. I aim to analyze the effect of two factors that could influence the price and ruin the relationship assumed by Pindyck (1993).

At first if expected return is not constant over time than the relationship between price, convenience yield and expected futures price is not stable over time. As a consequence there can be variables explaining the price change beyond futures and present price. Assuming that, the expected excess return is time varying, 2.6 can be modified in the following way:

$$E_t(P_{t+1} - P_t) = F_{1,t} - (1 - \bar{\rho})P_t + \rho_t P_t, \tag{2.8}$$

where $\bar{\rho}$ denotes the average of expected excess return and $\rho_t$ denotes the time varying component that depends on the changes in risk premium.

I try to proxy the changes in expected excess return (risk premium) with two variables. I use the VIX index as a measure of the market's expectation of stock market volatility over future time periods and the difference between long run and short run treasury yields.

If excess cross-market correlations are caused by the behavior of financial investors, then the overall market risk should affect price changes in commodity markets. According to the traditional view, those risk factors that affect stock markets do not have influence on commodity markets. However, the basic idea behind the theory blaming financial traders for increased volatility and price level is that financial investors do not make their decisions based on fundamentals but on other sources of information. If this is true, then expectations on stock market volatility have to affect expected excess return on commodity markets.

Market inefficiencies can also ruin the assumed relationship. Liquidity constraints can affect the price especially during huge price changes without precedent, like the one in 2007-2008 in the wheat market. I use TED spread to proxy liquidity constraints in general.

TED spread can also be treated as a proxy of counterparty risk in general, which could also affect the relationship of futures and spot commodity prices and ruin the relationship assumed by the rational commodity pricing model. Extremely high value of counterparty risk may prevent the investors from exploiting arbitrage opportunities. As a consequence, the difference between futures and spot prices does not determine the level of convenience yield. Explanatory power of TED spread on price changes can also be interpreted as a consequence of large counterparty risk.

### 2.2.2 Testing for structural change

I also use regression 2.7 to check whether the relationship of price futures price and price change has altered during the sample period. Since my main goal is to gain evidence on the causes and factors of changed price evolution, breakpoint tests present themselves as proper instruments. However, traditional breakpoint tests seem inappropriate because we do not have a presumed, exact date of possible structural break (or breaks). I use Quandt-Andrews statistics that was designed to test the existence of a breakpoint with unknown date. I compute likelihood ratio test statistics, and I use 15 % trimming and 1 % significance level to obtain the critical values.

If I find evidence for a structural break, I follow the methodology of Bai (1997) to detect the date of change or changes. I compute the residual variance for every candidate break date and cut the sample at the point where the residual variance has a minimum. Then I conduct the Quandt-Andrews test separately on the obtained subsamples. If I find evidence on a structural break, I split the subsample again. I proceed until I can accept the null of no structural breaks. Then I reestimate every breakpoint that is obtained from a subsample containing more than one break.

### 2.2.3 Vector autoregression and volatility tests

A different approach suggested by Campbell and Shiller (1987) provides more specific instruments to test the present value model and helps to reveal the source of possible problems. The idea behind this approach is that instead of looking directly at the long-run properties of the data, dynamics of the variables can be estimated directly using simple vector autoregression models. The argumentation is the following.

Suppose that the evolution of variables that carry relevant information on the state of the market can be described by a VAR process. If one can give a consistent estimate on the structure of this VAR model, then optimal forecasts of the variables, including net marginal convenience yield, can be easily derived.

Since only a restricted set of state variables are observable, this kind of general VAR model cannot be estimated. Fortunately, under our model assumptions, equation 2.5 implies that $S_t$ reveals every information that is available for market participants on future evolution of convenience yield. As a consequence, a vector autoregression model on $z_t = (\Delta MCY_t, S_t)'$ gives us the optimal forecast of future changes in $MCY_t$. Since any VAR process can be written in an appropriate linear form, it can be written in the following form without loss of generality:

$$x_t = Ax_{t-1} + u_{t-1}, \tag{2.9}$$

where $x_t = (\Delta MCY_t, \Delta MCY_{t-1}, ...\Delta MCY_{t-p+1}, S_t, S_{t-1}, ..., S_{t-p+1})'$, $A$ is the matrix of vector coefficients, and $u_t$ is a vector of shocks. Given the model assumptions it follows that $i$ period ahead, the optimal forecast of $x$ is $E_t(x_{t+i}) = A^i x_t$. By substituting this expression into 2.5 we obtain the following equality:

$$e_1' x_t = \sum_{i=1}^{\infty} \delta^i e1' A^i x_t = e1' \delta A (I - \delta A)^{-1} x_t, \tag{2.10}$$

where $e_1$ and $e_2$ are column vectors whose first and $p+1$th element are 1 and whose remaining elements are zero respectively.

As this relationship has to hold for every $x_t$, it puts the following restrictions on the

VAR coefficients:

$$\beta 21k = -\beta_{11k}, k = 1, ..., p, \beta_{221} = \frac{1}{\delta} - \beta_{121}, \text{ and } \beta_{22k} = -\beta_{12k} k = 2, ..., p, \tag{2.11}$$

where $\beta_{11k}$ and $\beta_{12k}$ are the coefficients of the k-th lag of $\Delta MCY$ and $S$ on $\Delta MCY_t$ and $\beta 21k$ and $\beta_{22k}$ are on $S_t$ respectively.

Besides testing these linear restrictions Campbell and Shiller (1987) suggest to compare the variance of the discounted sum of the optimal forecasts of futures convenience yield changes given the unrestricted estimated VAR, let's call it $\hat{S}_t$, and $S_t$. If the assumptions of the model are correct, then $\hat{S}_t$, and $S_t$ should be equal apart from measurement error, thus the ratio of their variance should be close to one.

One other interesting variance test compares the volatility of the following two variables. The unexpected change, or true innovation, in price implied by the present value relationship is defined as follows:

$$\psi_t = P_t - E_{t-1}(P_t) = P_t - \left(\frac{1}{\delta}\right) P_{t-1} + MCY_{t-1}, \tag{2.12}$$

where the second equality follows from the present value relationship. Under the assumptions of the present value model $\psi_t$ has to be equal with $\frac{1}{\mu}$ times the innovation from $t-1$ to $t$ in the expected present value of $\Delta MCY_t$, given by the unrestricted VAR model. Denote this variable with $\hat{\psi}_t$. It can be shown easily that:

$$\hat{\psi}_t = \frac{1}{\mu}\hat{S}_t - \frac{1+\mu}{\mu}\hat{S}_{t-1} + \frac{1}{\mu}\Delta MCY_t. \tag{2.13}$$

Comparing these variances helps us to evaluate the performance of the model. While the rejection of a linear restriction does not provide too much information on the sources of failure and the severity of the problems, comparison of the evolution of these variables might be a useful instrument in this sense.

My empirical strategy is, first, to detect the obvious deviations of the data from the model predictions and separate sub-periods that seem to be different based on the parameter estimates by means of the linear regression discussed in the previous subsection. Then for those sample periods in which I did not find evidence against the model I test the restrictions on VAR and check the evolution of the variables discussed recently. Based on these results, I try to detect the source of the possible deviations from the model.

### 2.2.4 Decomposition of price change

In order to shed light on the reasons and components of price movements in those periods where the rational commodity pricing model seems to fit the data, I apply the following methodology. I separate three different components of price movements based on the implications of the Pindyck model (Pindyck (1993)). At first, I separate expected and unexpected price movements, then within unexpected price movements I set apart the effect of an immediate and unexpected change in convenience yield and the effect of the change in expectations regarding future convenience yield change.

The following equality directly follows from the equations presented previously:

$$P_t - P_{t-1} = P_t - E_{t-1}(P_t) + E_{t-1}(P_t) - P_{t-1} =$$
$$= \psi_t + S_{t-1} = \frac{1}{\mu}(S_t - \frac{1}{\delta}S_{t-1}) + \frac{1}{\mu}\Delta MCY_t + S_{t-1}. \quad (2.14)$$

Intuitively the first term, the difference of spreads is the change in expectations since $S_t$ is the sum expected changes in marginal convenience yield (see equation 2.5). The second term is the change in $MCY$ from period $t-1$ to period $t$. The last term is the $t-1$ period expectation of the price change from period $t-1$ to $t$.

# Chapter 3

# Empirics

## 3.1 Data

I use a data set containing daily data on wheat and crude oil spot and futures prices.

The source of data on wheat is the historical data set of the Minneapolis Grain Exchange. In the case of wheat I only used futures price data since appropriate spot price data was not available. The futures prices are the daily closing futures prices of one bushel of wheat in dollar cents specified in Hard Red Spring Wheat Futures contracts sold in the Minneapolis Grain Exchange. The available delivery months of futures contracts are March, May, July, September (New Crop) and December.

Oil price data is obtained from the database of U.S. Energy Information Administration. Spot price refers to the closing price in dollar of one barrel of West Texas Intermediate crude oil traded in the domestic spot market at Cushing, Oklahoma. Futures prices are the daily closing prices specified in futures contracts for one barrel of light sweet crude oil delivered in Cushing, Oklahoma, traded on the New York Mercantile Exchange. Futures contracts are available for every consecutive month for the current year and the next five years.

In the case of oil prices the specification of the futures contracts allows us to compare these spot and futures prices since the underlying asset of the futures contracts is the same type of oil delivered at the same point as the one traded on the cash market.

In my analysis I use data on the period between January 2, 1991 and January 30, 2012.

I also use data on market yield on U.S. Treasury securities at 3-month and at 10-years and on the exchange rate of the dollar against four different currencies, yen, Canadian dollar, franc and pound obtained from the database of the Federal Reserve. 3-Month London Interbank Offered Rate (LIBOR) based on U.S. Dollar, daily VIX index and monthly producer price index data is obtained from the database of the Federal Reserve bank of St. Louis. The source of data on S&P500 index is Google Finance.

While Pindyck used monthly data for his analysis (Pindyck (1993)) I try to make use of the higher frequency data available. Unfortunately there are only a limited number of futures contract specifications thus in order to be able to use more frequent data I have to use estimated futures prices extrapolated from longer horizon contracts.

I use two different data specifications to my estimations. For the first part of my empirical investigation, I use daily data frequency, for the second part I use monthly data frequency. The reasons are detailed below.

In both cases, when actual data is not available, I infer spot and appropriate futures prices from the spot price (if available) and the nearest futures contract, or the nearest and the next to the nearest futures contract using the following formula:

$$\left(\frac{P_t}{F_{1,t}}\right)^{n_1 2} = \left(\frac{F_{1,t}}{F_{2,t}}\right)^{n_{01}}, \tag{3.1}$$

where $P_t$ is the period t spot price, $F_{1,t}$ and $F_{2,t}$ are the prices on futures contracts with given maturity, and $n_{01}$ and $n_{12}$ are, respectively, the number of days between t and the expiration date of the $F_{1,t}$ futures contract, and between the $F_{1,t}$ and the $F_{2,t}$ contract. Assuming that this equality holds for every maturity, it is possible to estimate the spot price or the price specified in any futures contract with known maturity given that at least two futures prices, or the spot price and a futures price is observable. Intuitively I extrapolate the difference between two futures contracts with known maturity date to gain estimated values for the spot price and unknown futures price.

In the case of wheat I assumed that the date of delivery for every futures contract is the first business day after last trading day of the contract. In order to avoid computational difficulties I assumed that the difference between the delivery dates of two consecutive oil futures contracts is exactly one month, and that the oil purchased in the spot market is delivered at the same time as the delivery of the spot contract (the contract expiring in the month) occurs. In order to get precise estimates one should assume a stable relationship between the spot and futures prices with different maturity which might or might not be true. Deviations from the real values might ruin some results of the model.

Measurement error is likely to be high in the case of daily data frequency. Since the number of futures contract specifications is very limited, futures prices and convenience yields for almost all time periods are extrapolated. On the other hand, intuitively, the change in these extrapolated futures prices contains all the information about expectations that we need to test the model and daily data has several advantages over monthly data on the same time period.

Since in the first part of the empirical investigation the most important part of the analyses deals with the significance of variables and I do not put emphasis on the restrictions on the coefficients implied by the model, I decided to use daily data in this part of the analysis.

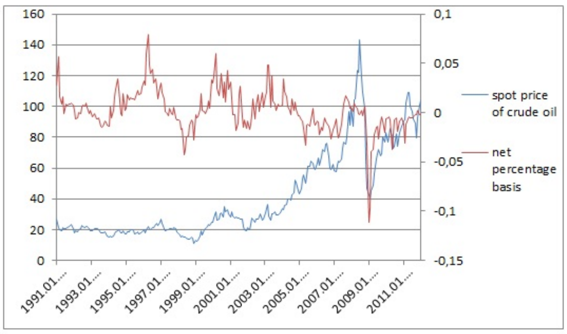
However, the second part of the analysis strongly relies on the accuracy of the data so in that case I use monthly data frequency. In this part of the analysis, real price and futures prices on the first Wednesday of each month are used [1].

## 3.2 Descriptive statistics

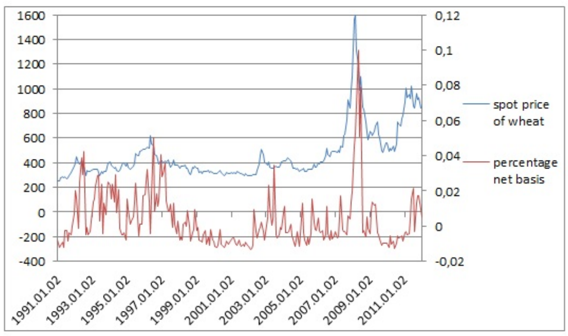
First let us take a look at the evolution of convenience yield during the sample period. Figure 3.1 shows the price and net percentage basis defined as convenience yield over price of crude oil and wheat. It can be seen that for wheat price and net percentage basis tend to move together during the whole sample period. Co-movement of these variables suggests that price changes are expected to be temporary. Loosely speaking common rise of price and percentage net basis implies that convenience yield relatively increases more than price thus market participants expect convenience yield and consequently the price to fall as supply and demand adjust to the change.

Figure 3.1: Price and net percentage basis

(a)




(b)




The case of crude oil seems to be different. While price and net percentage basis seems to move together during the first part of the sample, this is not true after the beginning

[1] Real prices are nominal prices divided by producer price index, the base date is January,1982.

of 2000's. It can be seen that continuous rise of price is not accompanied by the rise of percentage basis. This result suggests that while higher prices were expected to be temporary during the 90's, the changes of the 2000's were not.

Figures also show that convenience yield is quite high relative to commodity price both in the case of wheat and crude oil. It seems that utility derived from stock holding is an important factor for these two commodities. Monthly net convenience yield of wheat is more than 4 % in some periods, for oil it is even higher. This means that firms are willing to give up 4-5 % of the price per month to maintain inventories.

Now let us check whether the price, futures price and net marginal convenience yield follow a unit root process and price and futures price are indeed cointegrated.

Table 3.1 shows the results of unit root tests of prices, futures prices, convenience yield, spread and the estimated values of expected excess return.

Table 3.1: Unit root tests and estimates of $\rho$

| | wheat | | oil | |
|---|---|---|---|---|
| | ADF | KPSS | ADF | KPSS |
| $P_t$ | -2.19 | 3.61 | -0.75 | 7.16 ** |
| $\Delta P_t$ | -16.55 ** | 0.03 | -33.66 ** | 0.1 |
| $F_t$ | -2.19 | 3.62 | -0.79 | 7.16 ** |
| $\Delta F_t$ | -16.54 ** | 0.03 | -33.45 ** | 0.1 |
| $MCY_t$ | -4.1 ** | 0.18 | -6.5 ** | 2.43 |
| $\Delta MCY_t$ | -15.1 ** | 0.31 | -27.05 ** | 0.06 |
| $S(\bar{\rho})$ | -4.85 ** | 0.16 | -4.93 ** | 3.94 |
| $S(\hat{\rho})$ | -4.98 ** | 2.29 | -5.93 ** | 0.55 * |
| $\bar{\rho}$ | 0.000273 | | 0.000337 | |
| $\hat{\rho}$ | 0.002549 | | -0.00049 | |

ADF performed with intercept, without trend
** 1% significance, * 5% significance



It can be seen that in the case of wheat the results are unambiguous. Neither the null of ADF test nor the null of KPSS test can be rejected for price and futures price thus I do not have strong evidence on or against unit root in these time series. For convenience yield stationarity is strongly rejected. At the bottom of the table the estimated values of expected excess return are presented, $\bar{\rho}$ indicates the sample average of excess return, $\hat{\rho}$ is the estimate implied by the estimated cointegration vector of price and futures price. Results also suggest that the spread is stationary no matter which estimate of excess return I use.

Estimates of excess return are far from each other and $\hat{\rho}$ implies unreasonably high annual excess return. This inconsistency might be caused by the fact that price and futures price do not follow a unit root process, rather stationary. The sample average estimate implies approximately 7 % annual expected excess return that seems reasonable.

The results for oil are more distinct. KPSS test provides strong evidence against the stationarity of price and futures price and ADF suggests that first differences are stationary. ADF also strongly rejected for marginal net convenience yield and both spreads. However estimates of expected excess return differ substantially and KPSS test of spread computed with $\hat{\rho}$ is rejected at 5 % level. The sample average estimate of expected excess return seems more reasonable, it implies approximately 9 % annual expected excess return.

To sum up, the test results seems to contradict the model assumptions, since the stationarity of convenience yield implies the stationarity of prices in the rational commodity pricing model. In the case of wheat I cannot reject the assumption that both prices and convenience yield are mean reverting, but I have to be cautious because I also can't reject unit root in prices. For oil the results clearly contradict the model assumptions. On the other hand I do not regard these results as strong evidences against the model. Outlying price levels of the 2007-2008 period might bias the results and hide the real dynamics of prices. One other possible explanation is that some kind of structural change occurred during the sample period that ruin the test results.

Since the sample mean estimate of expected excess return has better properties and it makes the spread clearly stationary both for oil and wheat I use $\hat{\rho}$ and $S(\hat{\rho})$ in the following parts of the analysis.

## 3.3 Linear regression

As I stated in the previous chapter I use a simple linear regression to test the model. I estimated the following equation:

$$\Delta P_t = \alpha_0 + \alpha_1 P_{t-1} + \alpha_2 F_{t-1,1} + \sum_{i=1}^{7} \beta_i \Delta LS_{t-i} + \sum_{i=1}^{7} \gamma_i TED_{t-i} + \sum_{i=1}^{7} \psi_i VIX_{t-i} + \epsilon_t, \tag{3.2}$$

where $\Delta LS_t$ is the first difference of the spread between long run and short run treasury yields, $VIX_t$ is the VIX index and $TED_i$ is the TED spread. Since I work with daily data I include seven lags of all explanatory variables to filter out the intraweek fluctuations common in financial markets.

In table 3.2 I present a summary of estimation results including all observations of the sample period. This table contains only the most important test statistics, including Wald-test results and $R^2$. Detailed outputs with estimated coefficients and individual significance levels can be found in the appendix. It can be seen that in the case of wheat neither price and futures price are individually significant nor the other included variables are jointly significant at a 1% level. For crude oil price and futures price are significant at a 1% level as the model predicts. Other variables are jointly insignificant at a 1% level. $R^2$ values indicate that the included explanatory variables have very low explanatory power on price change.

As one can see I present the Wald statistics both with White and Newey-West standard errors because heteroskedasticity is present and the serial correlation LM-test indicates strong serial correlation in the residuals. From now on, in all of the following estimations

Table 3.2: Full sample regression results

| | wheat | | oil | |
|---|---|---|---|---|
| | White s.e | Newey s.e | White s.e. | Newey West s.e. |
| $P_{t-1}$ | -0.54 | -0.91 | -2.53 * | -2.71 ** |
| $F_{t-1}$ | 0.54 | 0.91 | 2.53 * | 2.71 ** |
| $F(\beta_i, \gamma_i, \psi_i)$ | 1.97 ** | 1.8 * | 1.59 * | 1.81 * |
| $F(\beta_i)$ | 3.01 ** | 1.36 | 2.23 * | 1.85 |
| $F(\psi_i)$ | 1.7 | 1.88 | 1.71 | 0.77 |
| $F(\gamma_i)$ | 2.08 * | 1.62 | 1.07 | 1.77 |
| $R^2$ | 0.02 | | 0.04 | |
| LM test | 35.55 ** | | 7.93 ** | |

LM test is conducted with 7 lags
F() denotes the F-statistic of joint significance test

I use heteroskedasticity consistent standard errors.

Table 3.2 also shows that while the t-value of the difference between the long- and short-run Treasury bill yields and the VIX index are close to each other in the wheat equation, the t-value of VIX index is much lower in the oil equation. In order to decrease the number of coefficients I decided to drop VIX index from the regression and test the reduced model for structural changes.

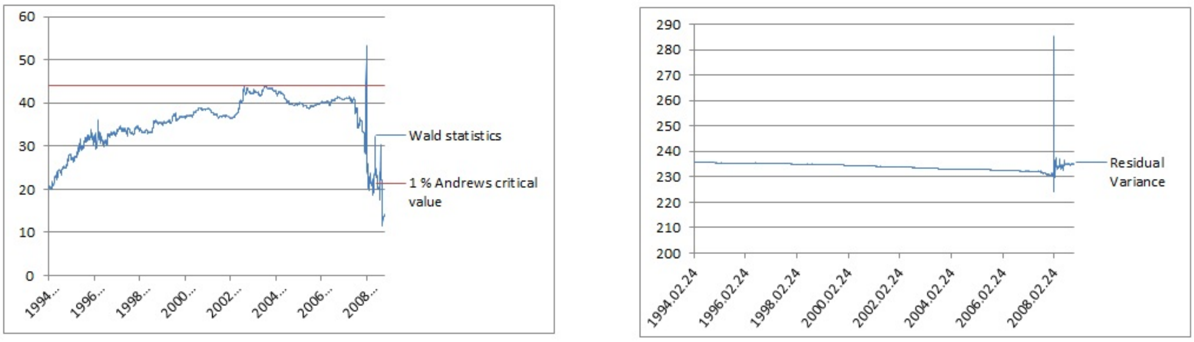
### 3.3.1 Looking for structural break

Figure 3.2 shows that the Quandt-Andrews test indicates structural change for wheat equation at 1% level. The estimated date of the break is 2008. 02. 22 based on the sum of squared residuals.

Testing the first subsample yields ambiguous results. If I do not exclude the observations from 2007 and 2008 the Chow test statistic monotonically increases, the sum of squared residuals monotonically decreases as the breakpoint converges to the right side of the sample.

Figure 3.2: Quandt-Andrews test statistic and residual variance of wheat

(a) (b)




After I exclude the observations from 2007-2008 and test for the first part of the sample (1991-2006) the problem disappears and the Quandt-Andrews test does not indicates structural break at 5% level. I interpret this result as a consequence of a prolonged change during the 2007-2008 period.

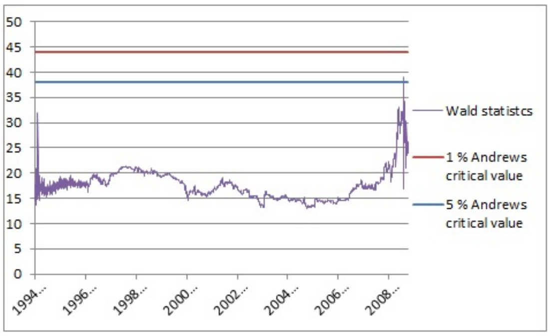
In the case of the second subsample the results are very similar as in the case of the first subsample. The test indicates structural break at 1% level but the residual variance decreases as the breakpoint approaches the left side of the subsample. Just as in the previous case I drop the observations from the 2007-2008 period.

The Quandt-Andrews test for the 2009.01.01 – 2012. 01.30 period indicates structural break at 1 % level. The estimated date of the breakpoint is 2011. 02. 14. For the first part of the second subsample I cannot reject the null of no structural break. However the maximum of the Wald test is a bit higher than the 5 % critical value. For the 2011. 02. 14 – 2012.01.30 period the Quandt-Andrews test still indicates a structural break at 2011. 07. 01. It seems that the last part of the sample is characterized by some prolonged structural change. Since the sample would become too small if I split the subsample into two I do not continue to look for structural breaks.

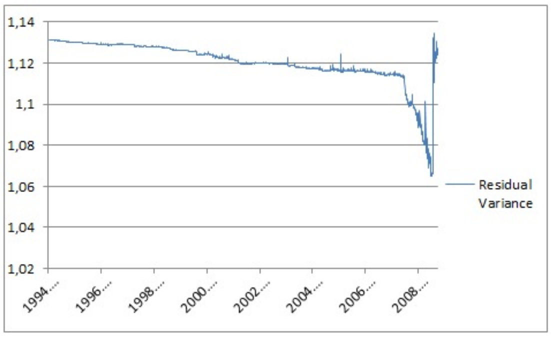
The refinement of my results would require me to reestimate the first breakpoint, but as I concluded that the structural change was prolonged and excluded the data between 2007-2008 it does not make sense. Reestimation of the 2011.02.14. breakpoint using the 2009.01.01 – 2011.07.01 subsample gives exactly the same date.

Figure 3.3: Quandt-Andrews test statistic and residual variance of oil

(a)




(b)




The Quandt-Andrews test does not indicate structural change for oil at 1 % significance level, but the maximum value of Wald statistics is higher than the critical value at 5 % level. The estimated date of structural break is 2008. 08. 26. For the first part of the sample (1991. 01. 02 – 2008. 08. 26.) the test does not indicate a break at 5% significance level.

For the second part there is strong evidence on a structural break. However the sum of squared residuals reaches its lowest values at the left side of the subsample. It is likely that the extreme values of the crisis distort the results thus I exclude the observations before 2009.01.01.

The Quandt-Andrews test still indicates a structural break. The minimum of residual variance is at 2011.02.17. For the 2009.01.01 – 2011.02.17 period the test does not indicate structural change at 5% level. Just as in the case of wheat, for the 2011.02.17 – 2012.01.30 period the test still indicates a structural break but I cannot detect the date of the structural break because the sum of square residuals monotonically decreases as the candidate breakpoint approaches the left side of the subsample.

Just as in the previous case the refinement of my breakpoints would require me to reestimate the breakpoint using subsamples that contains only one structural break. However as I concluded that all of the estimated structural breaks are rather long transitions it would not make sense.

Based on the results of these tests I divide the sample into the following subsamples. For wheat I cut the sample into four subsamples: the pre-crisis period between 1991 and 2006, the years of the crisis between 2007 and 2008, the post-crisis period from 2009.01.01 to 2011. 02. 14 and the last period between 2011.02.14 and 2012.01.30 which seems to be a period of some kind of transition.

For oil the subsamples are similar. I divide the sample into four: the 1990-2006 period, the years of the crisis between 2007 and 2008, the 2009.01.01 – 2011.02.17 period and the transitional period between 2011.02.17 and 2012.01.30.

### 3.3.2 Subsample results

Table 3.3 and table 3.4 show the results of regression 3.2 in subsamples determined in the previous subsection. Just as in the previous case, in the text I present a summary of estimation results that contains only the most important test statistics, including Wald-test results and $R^2$. Detailed outputs with estimated coefficients and individual significance levels can be found in the appendix.

For wheat in the first subsample the results seem to support the model assumptions. Both price and futures price are significant at a 1% level and the Wald restrictions implied by equation 2.6 cannot be rejected. The included additional variables are insignificant.

For the 2007-2008 period the results are different. Price and futures price turn out to be insignificant while TED spread seems to explain price movements. Since this period is characterized by unexpected rise and then fall of the price it is not surprising that futures price, that represents expected price change, cannot explain price movements. The significance of TED spread supports the theory that price moved away from its fundamental level, however the cause is rather the liquidity constraint of market participants.

It can be seen that in the third period neither price and futures price, nor other included variables explain price evolution. This is a bit surprising, especially because in the last period both $LS_t$ and $TED$ turned out to be significant.

Table 3.3: Subsample estimation results of wheat

| WHEAT | | | | |
|---|---|---|---|---|
| | 1991-2006 | 2007-2008 | 2009.01.01-2011.02.14 | 2011.02.15-2012.01.30. |
| | Newey-West | Newey-West | Newey-West | Newey-West |
| $P_{t-1}$ | -3.95 ** | -0.3 | -1.71 | -1.9 |
| $F_{t-1}$ | 3.94 ** | 0.29 | 1.72 | 1.88 |
| $F(\beta_i, \gamma_i)$ | 1.19 | 1.82 * | 0.78 | 4.42 ** |
| $F(\beta_i)$ | 1.51 | 2 | 0.6 | 3.85 ** |
| $F(\gamma_i)$ | 1.28 | 2.06 * | 0.68 | 2.62 * |
| $F(\alpha_i)$ | 1.36 | 0.47 | 1.2 | 2.97 * |
| $R^2$ | 0.02 | 0.06 | 0.03 | 0.19 |
| LM test | 2.08 * | 5.85 ** | 1.94 | 2.54 * |
| ADF($\epsilon_t$) | -53.69 ** | -18.17 ** | -21.96 ** | -10.19 ** |

LM test is conducted with 7 lags
ADF performed with intercept, without trend
F() denotes the F-statistic of joint significance test
** 1% significance, * 5% significance

For crude oil the results of the first subsample are unambiguous. Although price and futures price are significant at 1 % level and other included variables are not significant, the restrictions on the constant term, and the coefficients of price and futures price imposed by the model are rejected at a 5 % level. The estimated coefficients of price and futures price are significantly higher than they should be, but it is likely to be the consequence of measurement error detailed previously.

Just as in the case of wheat price and futures price are insignificant but additional variables are jointly significant. However for oil not the TED spread but the difference between long-run and short-run T-bill yields seems to explain price movements.

It can be seen in figure 3.1 that the evolution of price and net percentage basis is different for wheat and or oil. While in the case of wheat the increase is sudden and followed by an increase in net percentage basis, price increase of oil seems slower and is not accompanied by percentage net increase. Estimation results and these figures together suggest that the sudden jump in wheat price is caused by some kind of panic that generated liquidity problems, but the rise was expected to be temporary.

In the case of oil the estimation outcomes suggest that liquidity constraints do not play a key role in the price evolution. This result is not surprising, since the composition of market participants is pretty different in the oil market. The fact that futures price does not, but $LS_t$ has a significant effect on price movements suggests that the constant expected excess return assumption is not satisfied in this period.

Table 3.4: Subsample estimation results of oil

| OIL | | | | |
|---|---|---|---|---|
| | 1991-2006 White | 2007-2008 Newey-West | 2009.01.01-2011.02.17 White | 2011.02.18-2012.01.30 White |
| $P_{t-1}$ | -3.56 ** | -0.63 | 0.71 | -2.86 ** |
| $F_{t-1}$ | 3.56 ** | 0.63 | -0.74 | 2.85 ** |
| $F(\beta_i, \gamma_i)$ | 1.25 | 2.07 * | 1.28 | 0.85 |
| $F(\beta_i)$ | 1.39 | 2.61 * | 1.24 | 0.78 |
| $F(\gamma_i)$ | 1.08 | 1.58 | 1.93 | 0.73 |
| $F(\alpha_i)$ | 2.74 * | 1.64 | 3.21 * | 2.92 * |
| $R^2$ | 0.02 | 0.14 | 0.06 | 0.13 |
| LM test | 1.78 | 2.2 * | 0.47 | 1.76 |
| ADF($\epsilon_t$) | -57.63 ** | -22.25 ** | -20.27 ** | -13.67 ** |

LM test is conducted with 7 lags
ADF performed with intercept, without trend
F() denotes the F-statistic of joint significance test
** 1% significance, * 5% significance

I need to mention here that the ADF test suggests that the residual is stationary in the 2007-2008 period both for wheat and oil, which implies that price and futures price are cointegrated thus $S_t$ is stationary. Campbell and Shiller (1987) argues that an explosive, rational bubble component in the price causes explosive behavior of the spread. Consequently stationarity of $S_t$ provides evidence against a bubble in commodity prices, provided that the rational commodity pricing model is appropriate.

The results for the last two periods are surprising. While in the third subsample, right after the crisis none of the included variables are significant, in the last period, which was not found to be uniform by the Quandt-Andrews test, price and futures price are strongly significant and the other variables are jointly insignificant. On the other hand, restrictions on price, futures price and the constant are rejected in the last period.

### 3.3.3 VAR restrictions and variance tests

As I mentioned in the beginning of this chapter I use monthly instead of daily data in this part of my analysis in order to avoid the problems arising from measurement errors. For every month in the sample period I used the real price and futures prices prevailing on the first Wednesday of the given month.

In order to make the computation easier I replaced the spread variable, $S_t$, with $S'_t = \frac{1}{\mu} \cdot S_t$ in the vector autoregression. While this technical detail somewhat modifies the precise formula of linear restrictions, the intuition behind the test does not change at all. The new spread variable has the same function as $S_t$ has (for details see Chapter 2), it is only multiplied by a constant for technical reasons.

Before I discuss the outcomes of the test, I need to mention one important detail regarding the results. As I mentioned previously in this chapter, I decided to use the sample average of commodity specific expected returns because it seemed to have better properties (for details see the Descriptive statistics subsection). Though I do not present test results using different estimates of expected excess return, in order to be honest on the reliability of my results, I have to mention that outcomes strongly depend on the value of $\mu$

used. A relatively small change, like using the estimate obtained from the cointegration test instead of the sample average, can changes the results substantially. On the other hand, the estimates of the expected excess return on commodities seem reasonable, and both in the case of oil and wheat they make the spread variable stationary, thus I consider them reliable and rely on the test results when I evaluate the validity of the rational commodity pricing model.

Table 3.5 and 3.6 show the results of the restriction tests on VAR estimates and the estimated variances discussed in the previous section. First, I estimated the vector autoregression, the dynamics of spread and convenience yield, using the whole sample. Next, I excluded the observations after 2007.01.01, finally I considered the last part of the sample. I have chosen these subsamples based on the results of the structural break test, but I did not conduct the estimation for the 2007-2008 period because drawing conclusions about the dynamics of these variables based on a two year period, that seems to be a one time episode, does not make sense. I also treated the last two periods separated by the Quandt-Andrews test, as one homogeneous unit simply because otherwise it would be too short to get meaningful estimates.

In every case I have chosen the number of lags included based on the value of Schwartz-Bayesian Information Criterion. To test the linear restrictions imposed by the rational commodity pricing model I used likelihood ratio test.

It can be seen that changes in marginal convenience yield are fairly predictable in the case of wheat. The explanatory power is higher for the first period, $R^2$ is around 0.24, but it is remarkable in the last period too. On the other hand, estimation results show that in the last period lagged values of $S_t'$ do not help to predict convenience yield as the model predicts. This observation is in line with the results of the linear regression model, it also shows that the assumptions of the Pindyck model (Pindyck (1993)) seem to be violated. For the first sample period the spread variable Granger-causes the $\Delta MCY_t$ as the model

Table 3.5: Test of linear restrictions and variance tests for wheat

| WHEAT | | | |
|---|---|---|---|
| Sample: 1991.01.02 - 2012.01.30 | | | |
| Number of lags included (based on SBIC): | | | 3 |
| $\Delta MCY_t$ equation $R^2$: | | | 0.2913 |
| $S'$ equation $R^2$: | | | 0.6207 |
| LR test statistic for coefficient restriction: | | | $\chi^2(6) = 17.8086$ $p = 0.67\%$ |
| $\sigma(S'_t)$ | 7.883453 | $var(S'_t)/_{var(\hat{S}'_t)}$ | 0.8801 |
| $\sigma(\hat{S}'_t)$ | 8.402894 | | |
| $\sigma(\psi_t)$ | 0.3062986 | $var(\psi_t)/_{var(\hat{\psi}_t)}$ | 0.8677 |
| $\sigma(\hat{\psi}_t)$ | 0.3288199 | | |
| Sample: 1991.01.02 - 2006.12.30 | | | |
| Number of lags included (based on SBIC): | | | 2 |
| $\Delta MCY_t$ equation $R^2$: | | | 0.2352 |
| $S'$ equation $R^2$: | | | 0.4371 |
| LR test statistic for coefficient restriction: | | | $\chi^2(4) = 3.939$ $p = 41.43\%$ |
| $\sigma(S'_t)$ | 4.251739 | $var(S'_t)/_{var(\hat{S}'_t)}$ | 0.89 |
| $\sigma(\hat{S}'_t)$ | 4.512264 | | |
| $\sigma(\psi_t)$ | 0.1856034 | $var(\psi_t)/_{var(\hat{\psi}_t)}$ | 0.5 |
| $\sigma(\hat{\psi}_t)$ | 0.2629943 | | |
| Sample: 2009.01.01-2012.01.30 | | | |
| Number of lags included (based on SBIC): | | | 2 |
| $\Delta MCY_t$ equation $R^2$: | | | 0.1754 |
| $S'$ equation $R^2$: | | | 0.7246 |
| LR test statistic for coefficient restriction: | | | $\chi^2(4) = 1.08$ $p = 89.74\%$ |
| $\sigma(S'_t)$ | 3.978182 | $var(S'_t)/_{var(\hat{S}'_t)}$ | 0.86 |
| $\sigma(\hat{S}'_t)$ | 4.618477 | | |
| $\sigma(\psi_t)$ | 0.3488359 | $var(\psi_t)/_{var(\hat{\psi}_t)}$ | 0.75 |
| $\sigma(\hat{\psi}_t)$ | 0.4633464 | | |

predicts.

Linear restrictions are clearly rejected for wheat when I use the whole sample period, however it is likely to be the consequence of the outlying values in period 2007-2008. For the first and the last subperiod I did not find evidence against the model, however in the second case the lack of evidence is likely to be the consequence of the fact that the variances of parameter estimates are very high.

Table 3.5 also shows that variance of the estimated variable is higher than the variance of the actual variable both in the case of $S_t'$ and $\psi_t$. Intuitively this means that price moves less than it is predicted by the optimal forecast of convenience yield changes. The ratio of the variances clearly differs from one in the first period as well. As I see, this deviation from the model predictions is likely to be the consequence of the poor estimate of the discount factor. I relatively small underestimation of the discount factor can notably blow up the price effect of an expected futures change.

Figures showing the evolution of unexpected change in commodity price implied by the Pindyck model (Pindyck (1993)) and the estimated value of the innovation component predicted by the unrestricted VAR estimate for the whole sample and for the 1991-2006 period separately can be found in the appendix (see figure 3.6 and 3.7).

It can be seen that for the first subperiod the two variables move very closely but at the beginning of the period $\hat{\psi}_t$ is much more volatile. While the model does not fit perfectly the data the strong correlation of these variables from 1998 to 2006 support my assumptions. Deviation in the beginning of the period might be the consequence of the imperfect functioning and low trading volume of futures markets during these years. If the relationship of futures and spot prices does not show the exact value of marginal convenience yield because of the violation of the no-arbitrage condition then the model can give inappropriate results.

Concerning the second part of the sample and the years of the crisis, figure 3.6 supports the results of the previous chapter. It can be seen that the two variables are far from each other during 2007 and 2008. This is not surprising taking into account the results of the previous chapters. During the last period $\hat{\psi}_t$ and $\psi_t$ move close to each other again.

Table 3.6 shows the test results for crude oil. It can be seen that changes in convenience yield are less predictable given the lagged values of the spread and convenience yield than they were in the case of oil. Estimation results also show that $S'_t$ does not Granger-cause $\Delta MCY_t$, neither in the first nor in the last subperiod. This result poses serious doubt on the validity of the model even for the 1991-2006 period.

Linear restrictions on the VAR system cannot be rejected for the whole sample, but contrary to the case of wheat, the p-value is also low for the first subsample. For the last subperiod the restrictions cannot be rejected, but it is likely that this outcome is the consequence of high coefficient estimate variances and does not support the validity of our assumptions.

Unlike in the case of wheat, the variance ratios are similar independently of the sample I use to estimate the VAR system and differ from one by more for all subsamples.

Figure 3.8 and 3.9, included in the appendix, show the evolution of $\psi_t$ and $\hat{\psi}_t$ using the whole sample period and the 1991-2006 period to estimate the dynamics of $MCY_t$ and $S'_t$. Although the different magnitudes of variables can be misleading, it seems that $\psi_t$ is relatively more variable and the correlation is weaker between the two variables.

To sum up the results, considering the whole sample period the Pindyck model (Pindyck (1993)) does not appear to fit the data, the test results contradict the model assumptions. Results for the 2009-2012 subperiod neither for crude oil nor for wheat are conclusive, presumably because of the moderate sample size. In the case of oil the results do not support the model assumptions for the first subperiod.

The test results are only supportive for the 1991-2006 subperiod of wheat. The spread Granger-causes $MCY_t$ as the model suggests, restrictions are not rejected and both $\psi_t$ and

Table 3.6: Test of linear restrictions and variance tests for oil

| OIL | | | |
|---|---|---|---|
| Sample: 1991.01.02 - 2012.01.30 | | | |
| Number of lags included (based on SBIC): | | | 2 |
| $\Delta MCY_t$ equation $R^2$: | | | 0.0958 |
| $S'$ equation $R^2$: | | | 0.7306 |
| LR test statistic for coefficient restriction: | | | $\chi^2(4) = 6.144$ $p = 18.86\%$ |
| $\sigma(S'_t)$ | 0.6298 | $\frac{var(S'_t)}{var(\hat{S}'_t)}$ | 0.7923 |
| $\sigma(\hat{S}'_t)$ | 0.7075604 | | |
| $\sigma(\psi_t)$ | 0.026973 | $\frac{var(\psi_t)}{var(\hat{\psi}_t)}$ | 0.4717 |
| $\sigma(\hat{\psi}_t)$ | 0.0392748 | | |
| Sample: 1991.01.02 - 2006.12.30 | | | |
| Number of lags included (based on SBIC): | | | 2 |
| $\Delta MCY_t$ equation $R^2$: | | | 0.1575 |
| $S'$ equation $R^2$: | | | 0.6845 |
| LR test statistic for coefficient restriction: | | | $\chi^2(4) = 9.38$ $p = 5.23\%$ |
| $\sigma(S'_t)$ | 0.4891547 | $\frac{var(S'_t)}{var(\hat{S}'_t)}$ | 0.82 |
| $\sigma(\hat{S}'_t)$ | 0.5393956 | | |
| $\sigma(\psi_t)$ | 0.0168588 | $\frac{var(\psi_t)}{var(\hat{\psi}_t)}$ | 0.47 |
| $\sigma(\hat{\psi}_t)$ | 0.0245442 | | |
| Sample: 2009.01.01-2012.01.30 | | | |
| Number of lags included (based on SBIC): | | | 2 |
| $\Delta MCY_t$ equation $R^2$: | | | 0.1089 |
| $S'$ equation $R^2$: | | | 0.7842 |
| LR test statistic for coefficient restriction: | | | $\chi^2(4) = 4.22$ $p = 37.71\%$ |
| $\sigma(S'_t)$ | 0.6494168 | $\frac{var(S'_t)}{var(\hat{S}'_t)}$ | 0.86 |
| $\sigma(\hat{S}'_t)$ | 0.7525798 | | |
| $\sigma(\psi_t)$ | 0.031414 | $\frac{var(\psi_t)}{var(\hat{\psi}_t)}$ | 0.44 |
| $\sigma(\hat{\psi}_t)$ | 0.0709591 | | |

$\hat{\psi}_t$, and $S'_t$ and $\hat{S}'_t$ are strongly correlated. The only observation that poses doubts on the model is the inadequate value of variance ratios.

### 3.3.4 Discussion of the implications of the rational commodity pricing model

Since Pindyck's rational commodity pricing model (Pindyck (1993)) is supported by the data on wheat, at least for the period before the extreme rise and then fall in commodity prices, I rely on the implications of the model regarding the underlying causes of recent changes in commodity markets. I support my arguments mainly by the quantitative findings on wheat price changes, but I also present data on crude oil. Despite the fact that I found evidence against the model assumptions in case of crude oil, I still think that my approach helps to better understand recent episodes.

First I describe intuitively how price, convenience yield and spread are expected to behave given the different stories that try to explain the changed behavior of commodity prices. Then I present my empirical findings and confront them with theories.

Let's take first the possibility that commodity prices are led purely by fundamentals, that is increasing commodity price is the consequence of growing demand or falling supply [2]. Decrease of net supply affects marginal convenience yield directly since inventories become more spare thus one additional unit of stock becomes more valuables (see the introduction for more details). If shortage is expected to be temporary then marginal convenience yield grows more than it would grow in the case of a permanent change. The explanation is pretty intuitive, stockholders try to exploit the temporary shortage, sell their stocks at a high price and buy it back when net supply goes back to the original level. As

[2] For detailed discussion of these reasons see: Unc (2011)

a consequence marginal convenience yield jumps to a high level and then falls back after supply and demand adjust. If shortage is expected to last in the long-run stockholders do not have any incentive to sell a high fraction of their inventories immediately. Nevertheless, the given commodity becomes more scarce, thus inventory levels somewhat decrease and marginal convenience yield increases.

A permanent increase in demand or decrease in supply has to be accompanied by increase in net marginal convenience yield and positive (or at least non-negative) values of the spread, since the spread represents the expected changes in marginal convenience yield. However, let me emphasize that this approach does not help me to distinguish between the causes of demand or supply change. The theory that explains recent price movements with speculative inventory hoarding cannot be supported or confuted by this model since increased demand caused by speculation or fundamental change have the same effect both on price and convenience yield.

Demand caused by some kind of panic in the market also cannot be distinguished from other sources of demand. For example, separating the effect of "real" demand change induced by the increasing demand of developing countries and the effect of increased precautionary reserves of states is not possible. As far as I know, the potential effect of these kind of purchases on commodity prices is not taken into account in any other study on the topic. As I see, it is very important to distinguish these "panic" purchases from other sources of demand increase, in order to be able to shed light on the causes of the recent price spike. On the other hand it is pretty hard to define precisely what I mean on "panic" purchase and it is even harder to fit it in a rational commodity pricing model. In some sense it is speculation, but not in the traditional sense, since buyers do not aim to gain monetary profit. It can be also considered simply as a part of demand increase but I think that doing that would make harder to predict future episodes in commodity markets. The motivation of this kind of purchases is rather follows from the consumable nature of the good, thus it can not be defined in stock markets. I define panic purchase as a purchase induced by intensified precautionary motives. While these purchases increase the level of stocks, but do not increase the level of free stocks owned by farmers or others in agriculture trade, thus do not decrease the level of marginal convenience yield.

Consider now the supposed effects of the increased number of financial investors on commodity prices. The hypothesis provided by authors emphasizing the role of financialization in commodity price movements is that the behavior of index investors and other types of financial investors induces commodity prices to move, without any kind of fundamental change. This argument is supported by the increased co-movement of asset returns (see the introduction for details). Let me emphasize, that spot price movements induced by episodes on the futures market do not necessarily contradict the Pindyck model of rational commodity pricing (Pindyck (1993)). If market participants form their expectations on future changes in convenience yield based on these fake futures market signals, then the induced commodity price movements can be in line with the assumptions of the rational commodity pricing model. If futures prices movements are considered as, and actually was, reliable sources of information on future state of fundamentals, then futures price episodes can affect commodity prices without contradicting the assumptions of the rational commodity pricing model.

Let us see how price, convenience yield and spread should move in this case. If futures prices were drawn up by financial investors and stockholders, assume that this movement is a signal of a futures price increase, then they have an incentive to buy on the spot market and increase the level of their inventories. If they do so then spot price increases and convenience yield decreases and in line with an expected increase in price, spread also increases. In principle any kind of change that pushes up the expected value of future price has the same effect both on prices and on convenience yield, since it also induces market participants to build up inventories. Fortunately, there is one important difference that might help us to separate the two cases. If price change is caused by an expected rise in demand, convenience yield has to rise significantly later because of the increased demand. If price increase was the consequence of a fake signal on the futures market, then accumulated inventories are not used up, thus we should not observe any jumps in convenience yield.

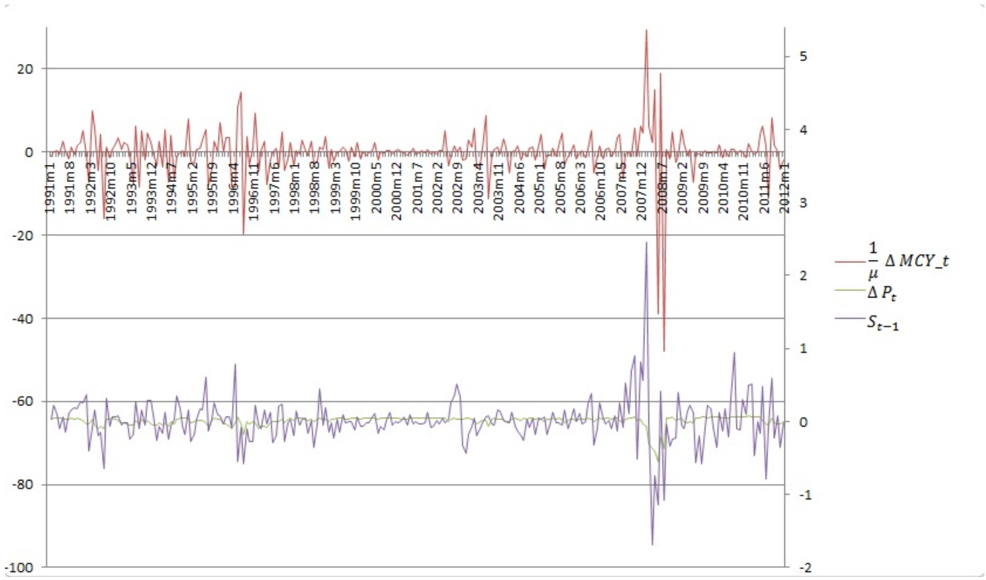
Also one can distinguish the effect of a fake signal and a signal that properly indicates futures demand or supply change based on the relationship of expected and unexpected change in price. If price movements are led by fake signals, caused purely by futures market speculation, then I expect the expected price change and price change to move separately because of the higher uncertainty of the expectations, while anticipated demand change should move price and expected price together.

Since previous research suggest that trading volume of index investors has started to grow rapidly around 2004 and 2005 (see Sanders et al. (2010)), I investigate whether there is any kind of evidence on the effect of these investors on commodity prices during the 2004-2006 period that is not present in the 1991-2003 period.

Figure 3.4 shows the evolution of $\Delta P_t$, $\frac{1}{\mu} \cdot \Delta MCY_t$ and $S_t$, that represents anticipated price movements of wheat, throughout the whole sample period. The most striking feature of the data is that a very small fraction of price movements is expected price change, $S_t$ has much smaller values than $\Delta P_t$ as the lower part of the graph shows.

It can be seen that expected price change and price change are correlated more in the first part of the sample. The correlation of $\Delta P_t$ and $S_t$ for the 1991-2003 period is around 0.27 while for the 2003-2006 period it is only 0.05. This results shows that a higher fraction of price changes is predictable in the first sample period than during the period characterized by the increased presence of financial investors. The result suggests that a higher fraction of price movements is caused by fundamental changes that are usually somewhat predictable.

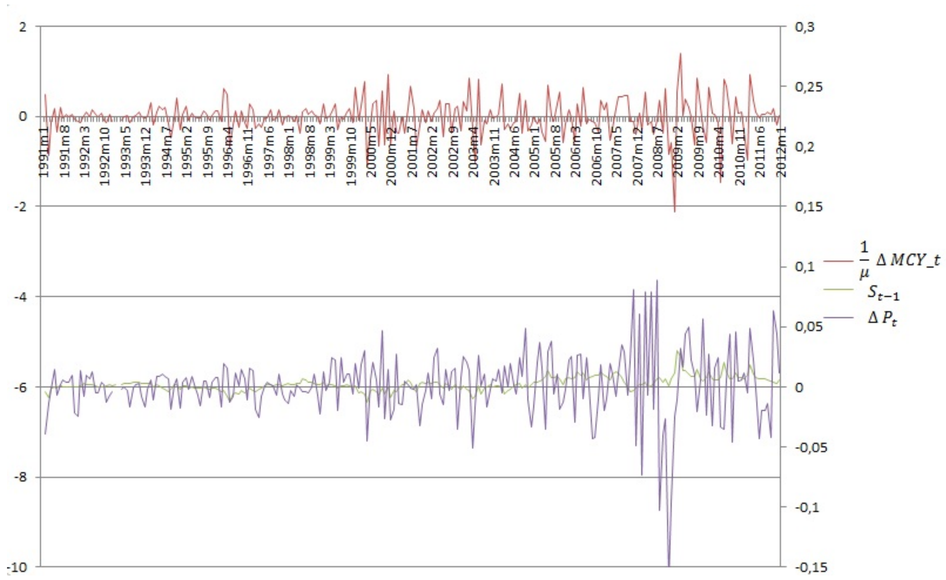
Figure 3.4: Price change decomposition of wheat




It is also striking that while the large increase in 2007 seems to be unpredictable, a substantial part of the subsequent fall is anticipated change. It is also important, that price increase is accompanied by a huge change in convenience yield. This suggests that this sudden, big price increase is caused by a sudden, unexpected change in net supply that seriously decreased the level of inventories. As I see the only possibility that this is the consequence of some kind of panic reaction, that induced big consumers - like governments - to purchase huge amounts of the commodity. This theory can also explain the following, anticipated price decrease. When the fact that increased demand is temporary becomes obvious, price starts to fall but this fall is expected by market participants.

The period just before this sudden increase is interesting. During the first part of 2007 prices continuously and significantly grow. This growth is not followed by convenience yield increase as the boom in 2008, the average value of real marginal convenience yield in 2007 is less than the average of the 1991-2006 period (0.01 and 0.008 respectively). This phenomenon is in line with the theory that during this period new information on future supply or demand changes made spot price to increase, but also suggests that the level of

Figure 3.5: Price change decomposition of crude oil



inventories do not decreased at all which would be a bit surprising in case of substantially decreasing net supply. The only explanation which do not contradict the assumptions of rational commodity pricing model is that stockholders in every period expected higher next period net supply than the actual net supply.

As I discussed previously the new information that caused unexpected price increase can arise both from fake signals, caused by futures market speculations and from new information on market fundamentals. Since this period is followed by a quite confusing period, it would be hard to distinguish between the two possibilities.

While my test results suggest that rational commodity pricing model do not fit the data on the period after the price spike, it worth to take a look at those part of the graph too. Figure shows that while price is much more volatile than it was before the crisis, convenience yield seems to be pretty smooth. This results suggest that changing expectations drive the price.

Figure 3.5 shows the evolution of the same variables for oil. As I stated previously VAR test provided evidence against validity of the rational commodity pricing model for oil, but it can be interesting to mark the difference between this graph and the other for wheat. At first, the relationship of expected price change and actual price change seems to be more stable over time than for wheat. It can be seen also, that the episodes of 2007 and 2008 are not accompanied by as extreme values of marginal convenience yield as in the case of wheat. In addition, a much smaller part of the extreme fall in oil price in 2008 was expected in advance. Both these observations, and the fact that the rational commodity pricing model fits oil data worth, suggest that price formation is different in wheat and crude oil markets.

# Conclusion

This empirical work aims to contribute to the debate on the possible causes of the recent price spike in commodity markets. My empirical strategy is to test the adequacy Pindyck's rational commodity pricing model (Pindyck (1993)) on wheat and crude oil markets in order to gain evidence on or against different theories explaining the observed events. Pindyck's model is based on the concept of convenience yield and can be viewed as a reduced form of a dynamic supply and demand model. I apply this model to gain empirical evidence on the causes of recent events in commodity markets. The intuition behind my approach is that the level of inventories of a storable commodity carries useful information on the changes in demand and supply and on the expectations of market participants.

In the first part of my empirical analysis I summarized the most important properties of the data. I investigated the dynamic properties of price variables and the magnitude of convenience yield. I did not found strong evidence against the rational commodity pricing model, moreover I found that convenience yield is quite high relative to commodity price both in the case of wheat and oil.

After I found that the basic properties of variables do not contradict the model I used linear regressions to test whether price and futures price contains all relevant information to predict price changes as the Pindyck model (Pindyck (1993)) implies. I also used these regressions to check for structural changes in the data. Based on the results of the Quandt-

Andrews I divided the sample into four subperiods both in the case of oil and wheat. The structural change test indicated that the pre-crises period between 1991 and 2006, the years of the crisis between 2007 and 2008, the post-crisis period from 2009 to the beginning of 2011 and the last period until the end of the sample significantly differ. Later I used this decomposition of the sample, except that I treated the last two period together because of the small sample size of them. The linear regression results on these subsamples showed that there were time periods both for oil and wheat when the suggested present value relationship did not hold. For wheat, I found evidence on that in the period of the price spike liquidity constraint of market participants affected the evolution of price. For oil, I did not find similar relationship. In the case of oil evidence on strongly varying risk premium questions the validity of the model.

The test of the restrictions implied by the rational commodity pricing model on vector autoregressions, and volatility tests provided mixed evidence. Considering the whole sample period, I rejected the validity of the Pindyck model (Pindyck (1993)) both for oil and wheat, but the test results were supportive for the 1991-2006 subperiod of wheat.

After the direct tests of the model assumptions, I examined the implications of the rational commodity pricing model regarding the underlying causes of recent changes in commodity markets. Since in the case of wheat, the results were supportive for the period before the extreme rise and then fall in commodity prices I concentrated on this period mainly. Despite the fact that I found evidence against the model assumptions in the case of crude oil,I also present data on crude oil because I think that it helps to better understand recent episodes.

Summing up the observations on crude oil, both VAR test and tests on the linear regression model provided evidence against validity of the rational commodity pricing model. The evolution of convenience yield and the relation of actual and expected price changes suggested that price formation is different in wheat and crude oil markets. In general, I found that price movements can not be explained by a rational commodity pricing model. In one hand it suggests, that there are some other factors than fundamentals moving crude oil prices. On the other hand, since I found that the model thus not fit the data, I am not able to provide reliable evidence on or against the role of financial investors.

For wheat, the relevant implications of the data regarding the role of financial investors and fundamentals in price evolution are the following. First, I found that expected price change and actual price change are correlated more in the first part of the sample, between 1991 and 2003 than in the years characterized by the increased presence of financial investors. This results shows that a higher fraction of price changes is predictable in the first sample period, which suggests that a higher fraction of price movements was caused by fundamental changes that are usually somewhat predictable.

Second, I found that movements of price and convenience yield suggests that the sudden, big price increase in 2007-2008 is caused by a sudden, unexpected change in net supply that seriously decreased the level of inventories. As I see, the most likely possibility is that this increase is the consequence of some kind of panic reaction that induced big consumers - like governments - to purchase huge amounts of the commodity.

Finally, the fact that during the first part of 2007 price continuously and significantly grew while convenience yield did not increase at all suggests that during this period, new information on future supply or demand changes made spot price to increase, but also suggests that the level of inventories did not decrease at all. Unfortunately because this period was followed by a quite turbulent period, it is hard to decide whether the new information that caused unexpected price increase was inferred from fake signals caused by futures market speculations or from new information on market fundamentals.

To conclude my results, the final lesson of my investigation is that wheat prices might have been affected by the activity of financial investors, but the price couldn't have experienced such a huge spike without suddenly increased demand caused by some kind of panic, and without the effect of liquidity problems of market participants. My results also show that information on future price movements became less reliable before the spike, which suggest that fake signals caused by the behavior of financial investors could have influenced price movements. However, the approach I use is not capable to provide strong evidence on this point because the extraordinary events of the 2007-2008 period make hard it to filter out the relevant information.

# Appendix

# Table 3.7: Detailed results of linear regressions for wheat

| Detailed regression results on WHEAT | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Dependent Variable: $\Delta P_t$ | | | | | | | | | | | | | | | |
| Sample (adjusted): | 1/14/1991 – 1/30/2012 | | | 1/14/1991 – 12/26/2006 | | | 1/09/2007 – 12/10/2008 | | | 1/08/2009 – 2/14/2011 | | | 2/14/2011 – 1/30/2012 | | |
| Newey-West HAC Standard Errors & Covariance | | | | | | | | | | | | | | | |
| **Variable** | **Coefficient** | **t-Stat.** | **Prob.** | **Coefficient** | **t-Stat.** | **Prob.** | **Coefficient** | **t-Stat.** | **Prob.** | **Coefficient** | **t-Stat.** | **Prob.** | **Coefficient** | **t-Stat.** | **Prob.** |
| $P_{t-1}$ | -1,45 | -0,91 | 0,36 | -1,81 | -3,95 | 0,00 | -0,98 | -0,30 | 0,77 | -5,62 | -1,72 | 0,09 | -6,78 | -1,90 | 0,06 |
| $F_{t-1}$ | 1,45 | 0,91 | 0,36 | 1,81 | 3,94 | 0,00 | 0,98 | 0,29 | 0,77 | 5,63 | 1,72 | 0,09 | 6,77 | 1,88 | 0,06 |
| $\Delta LS_{t-1}$ | -2,41 | -0,53 | 0,59 | 2,90 | 1,53 | 0,13 | 0,32 | 0,01 | 0,99 | -6,25 | -0,74 | 0,46 | -4,40 | -0,22 | 0,82 |
| $\Delta LS_{t-2}$ | -9,09 | -1,10 | 0,27 | -3,60 | -1,95 | 0,05 | -60,72 | -0,99 | 0,32 | -6,17 | -0,65 | 0,52 | -9,37 | -0,49 | 0,63 |
| $\Delta LS_{t-3}$ | 13,27 | 2,43 | 0,02 | 0,88 | 0,44 | 0,66 | 60,15 | 2,25 | 0,03 | 13,76 | 1,29 | 0,20 | -58,76 | -2,44 | 0,02 |
| $\Delta LS_{t-4}$ | 10,51 | 1,85 | 0,06 | 3,65 | 1,86 | 0,06 | 43,93 | 1,00 | 0,32 | 3,10 | 0,41 | 0,68 | 22,84 | 0,89 | 0,37 |
| $\Delta LS_{t-5}$ | 5,84 | 1,45 | 0,15 | -0,81 | -0,46 | 0,65 | 15,21 | 0,46 | 0,64 | 7,17 | 0,78 | 0,44 | -29,36 | -1,11 | 0,27 |
| $\Delta LS_{t-6}$ | 10,63 | 1,67 | 0,10 | -2,34 | -1,21 | 0,23 | 67,36 | 1,23 | 0,22 | -0,28 | -0,03 | 0,97 | -48,08 | -2,64 | 0,01 |
| $\Delta LS_{t-7}$ | -4,70 | -1,12 | 0,26 | 2,85 | 1,35 | 0,18 | -23,93 | -1,48 | 0,14 | 1,23 | 0,14 | 0,89 | -20,55 | -1,14 | 0,26 |
| $TED_{t-1}$ | -12,69 | -2,29 | 0,02 | 3,99 | 1,50 | 0,13 | -29,26 | -1,62 | 0,11 | 38,93 | 1,01 | 0,31 | 29,68 | 0,19 | 0,85 |
| $TED_{t-2}$ | 12,86 | 1,50 | 0,13 | 1,54 | 0,54 | 0,59 | 72,70 | 1,22 | 0,22 | -91,24 | -1,57 | 0,12 | -270,90 | -1,32 | 0,19 |
| $TED_{t-3}$ | -11,90 | -1,44 | 0,15 | -5,43 | -1,66 | 0,10 | -88,85 | -1,35 | 0,18 | 63,05 | 1,21 | 0,23 | -197,59 | -1,15 | 0,25 |
| $TED_{t-4}$ | -6,18 | -0,73 | 0,47 | 0,47 | 0,15 | 0,88 | -5,47 | -0,15 | 0,88 | -20,00 | -0,44 | 0,66 | 324,96 | 1,92 | 0,06 |
| $TED_{t-5}$ | 13,05 | 0,95 | 0,34 | 2,70 | 1,02 | 0,31 | 34,08 | 0,49 | 0,62 | 17,04 | 0,28 | 0,78 | 92,38 | 0,50 | 0,62 |
| $TED_{t-6}$ | 6,84 | 0,66 | 0,51 | -3,04 | -1,07 | 0,28 | -5,79 | -0,07 | 0,94 | -24,21 | -0,35 | 0,73 | 82,40 | 0,46 | 0,65 |
| $TED_{t-7}$ | -1,34 | -0,22 | 0,82 | -0,63 | -0,28 | 0,78 | 23,63 | 0,50 | 0,62 | 18,10 | 0,49 | 0,63 | -47,94 | -0,38 | 0,71 |
| $VIX_{t-1}$ | -0,09 | -0,56 | 0,58 | | | | | | | | | | | | |
| $VIX_{t-2}$ | 0,18 | 0,85 | 0,40 | | | | | | | | | | | | |
| $VIX_{t-3}$ | 0,27 | 1,36 | 0,17 | | | | | | | | | | | | |
| $VIX_{t-4}$ | -0,37 | -1,62 | 0,11 | | | | | | | | | | | | |
| $VIX_{t-5}$ | 0,21 | 0,97 | 0,33 | | | | | | | | | | | | |
| $VIX_{t-6}$ | 0,15 | 0,80 | 0,43 | | | | | | | | | | | | |
| $VIX_{t-7}$ | -0,35 | -2,03 | 0,04 | | | | | | | | | | | | |
| C | -0,65 | -0,27 | 0,79 | 0,57 | 0,49 | 0,62 | 5,06 | 0,30 | 0,77 | -4,93 | -1,54 | 0,12 | 6,07 | 0,20 | 0,84 |
| **R-squared** | 0,02 | | | 0,02 | | | 0,06 | | | 0,03 | | | 0,19 | | |
| Adjusted R-squared | 0,01 | | | 0,01 | | | 0,02 | | | -0,01 | | | 0,12 | | |
| S.E. of regression | 17,19 | | | 6,22 | | | 49,17 | | | 12,79 | | | 19,89 | | |
| Sum squared resid | 1249981,00 | | | 125006,90 | | | 937885,00 | | | 68734,91 | | | 65679,09 | | |
| Log likelihood | -18127,72 | | | -10550,57 | | | -2143,54 | | | -1725,27 | | | -797,96 | | |
| Durbin-Watson stat | 1,86 | | | 1,92 | | | 1,83 | | | 2,15 | | | 1,80 | | |

## Table 3.8: Detailed results of linear regressions for oil

Detailed regression results on OIL

**Dependent Variable: $\Delta P_t$**

| Sample (adjusted): | 1/14/1991 – 1/30/2012 Newey-West HAC S. E. & Cov. | | | 1/14/1991 –12/26/2006 White S. E. & Cov. | | | 1/09/2007 – 12/10/2008 Newey-West HAC S. E. & Cov. | | |
|---|---|---|---|---|---|---|---|---|---|
| **Variable** | **Coefficient** | **t-Stat.** | **Prob.** | **Coefficient** | **t-Stat.** | **Prob.** | **Coefficient** | **t-Stat.** | **Prob.** |
| $P_{t-1}$ | -1,71 | -2,71 | 0,01 | -2,06 | -3,56 | 0,00 | -2,00 | -0,63 | 0,53 |
| $F_{t-1}$ | 1,71 | 2,71 | 0,01 | 2,05 | 3,56 | 0,00 | 2,00 | 0,63 | 0,53 |
| $\Delta LS_{t-1}$ | -0,25 | -0,75 | 0,45 | -0,04 | -0,20 | 0,84 | 1,88 | 1,13 | 0,26 |
| $\Delta LS_{t-2}$ | 0,18 | 0,38 | 0,71 | 0,10 | 0,48 | 0,63 | 0,55 | 0,25 | 0,80 |
| $\Delta LS_{t-3}$ | 0,75 | 1,48 | 0,14 | 0,02 | 0,10 | 0,92 | 3,03 | 1,40 | 0,16 |
| $\Delta LS_{t-4}$ | -0,96 | -1,85 | 0,06 | 0,06 | 0,28 | 0,78 | -5,69 | -3,09 | 0,00 |
| $\Delta LS_{t-5}$ | 1,21 | 2,48 | 0,01 | 0,61 | 2,63 | 0,01 | 3,99 | 2,17 | 0,03 |
| $\Delta LS_{t-6}$ | 0,09 | 0,22 | 0,82 | -0,27 | -0,70 | 0,49 | 1,80 | 0,95 | 0,34 |
| $\Delta LS_{t-7}$ | 0,03 | 0,06 | 0,95 | 0,32 | 1,38 | 0,17 | -1,14 | -0,72 | 0,47 |
| $TED_{t-1}$ | -0,95 | -1,56 | 0,12 | 0,09 | 0,33 | 0,74 | -3,89 | -2,23 | 0,03 |
| $TED_{t-2}$ | 1,63 | 1,20 | 0,23 | -0,29 | -0,96 | 0,34 | 5,37 | 1,63 | 0,10 |
| $TED_{t-3}$ | 0,10 | 0,14 | 0,89 | -0,36 | -0,72 | 0,47 | -0,16 | -0,06 | 0,95 |
| $TED_{t-4}$ | -1,79 | -1,12 | 0,26 | -0,06 | -0,13 | 0,90 | -0,49 | -0,11 | 0,91 |
| $TED_{t-5}$ | 1,57 | 1,20 | 0,23 | 0,52 | 1,50 | 0,13 | -1,05 | -0,26 | 0,79 |
| $TED_{t-6}$ | -0,33 | -0,48 | 0,63 | -0,05 | -0,14 | 0,89 | -0,24 | -0,08 | 0,94 |
| $TED_{t-7}$ | -0,40 | -0,74 | 0,46 | 0,05 | 0,18 | 0,86 | 0,04 | 0,03 | 0,98 |
| $VIX_{t-1}$ | -0,05 | -2,45 | 0,01 | | | | | | |
| $VIX_{t-2}$ | 0,05 | 1,81 | 0,07 | | | | | | |
| $VIX_{t-3}$ | 0,03 | 0,97 | 0,33 | | | | | | |
| $VIX_{t-4}$ | -0,02 | -0,74 | 0,46 | | | | | | |
| $VIX_{t-5}$ | 0,03 | 1,04 | 0,30 | | | | | | |
| $VIX_{t-6}$ | -0,05 | -1,58 | 0,11 | | | | | | |
| $VIX_{t-7}$ | 0,00 | 0,14 | 0,89 | | | | | | |
| C | 0,13 | 1,93 | 0,05 | 0,16 | 2,58 | 0,01 | 0,45 | 0,98 | 0,33 |
| **R-squared** | 0,04 | | | 0,02 | | | 0,14 | | |
| Adjusted R-squared | 0,03 | | | 0,02 | | | 0,10 | | |
| S.E. of regression | 1,19 | | | 0,71 | | | 2,48 | | |
| Sum squared resid | 5994,73 | | | 1626,06 | | | 2393,79 | | |
| Log likelihood | -6737,38 | | | -3471,30 | | | -934,46 | | |
| Durbin-Watson stat | 2,15 | | | 2,06 | | | 2,26 | | |

| Sample (adjusted): | 1/08/2009 – 2/17/2011 White S. E. & Cov. | | | (non-adjusted): 2/17/2011 – 1/30/2012 White S. E. & Cov. | | |
|---|---|---|---|---|---|---|
| **Variable** | **Coefficient** | **t-Stat.** | **Prob.** | **Coefficient** | **t-Stat.** | **Prob.** |
| $P_{t-1}$ | 0,77 | 0,71 | 0,48 | -9,73 | -2,86 | 0,00 |
| $F_{t-1}$ | -0,80 | -0,74 | 0,46 | 9,68 | 2,85 | 0,00 |
| $\Delta LS_{t-1}$ | 0,25 | 0,24 | 0,81 | 2,58 | 1,06 | 0,29 |
| $\Delta LS_{t-2}$ | -1,87 | -1,83 | 0,07 | -0,68 | -0,29 | 0,78 |
| $\Delta LS_{t-3}$ | -0,86 | -0,74 | 0,46 | 0,52 | 0,24 | 0,81 |
| $\Delta LS_{t-4}$ | 0,11 | 0,10 | 0,92 | 1,70 | 0,70 | 0,48 |
| $\Delta LS_{t-5}$ | 0,29 | 0,29 | 0,77 | -4,35 | -1,76 | 0,08 |
| $\Delta LS_{t-6}$ | -0,46 | -0,42 | 0,67 | 1,06 | 0,47 | 0,64 |
| $\Delta LS_{t-7}$ | 2,08 | 2,13 | 0,03 | -1,64 | -0,65 | 0,52 |
| $TED_{t-1}$ | 8,54 | 2,01 | 0,05 | -8,38 | -0,60 | 0,55 |
| $TED_{t-2}$ | -7,24 | -1,06 | 0,29 | -14,46 | -0,78 | 0,44 |
| $TED_{t-3}$ | -4,80 | -0,67 | 0,51 | 5,73 | 0,29 | 0,77 |
| $TED_{t-4}$ | 3,15 | 0,44 | 0,66 | 5,78 | 0,31 | 0,76 |
| $TED_{t-5}$ | 4,15 | 0,59 | 0,56 | 13,42 | 0,61 | 0,54 |
| $TED_{t-6}$ | -15,98 | -2,00 | 0,05 | 13,96 | 0,77 | 0,44 |
| $TED_{t-7}$ | 11,02 | 2,13 | 0,03 | -15,23 | -1,12 | 0,26 |
| C | 3,00 | 2,62 | 0,01 | 3,74 | 1,44 | 0,15 |
| **R-squared** | 0,06 | | | 0,13 | | |
| Adjusted R-squared | 0,02 | | | 0,05 | | |
| S.E. of regression | 1,56 | | | 1,91 | | |
| Sum squared resid | 1029,41 | | | 594,98 | | |
| Log likelihood | -811,33 | | | -363,01 | | |
| Durbin-Watson stat | 1,97 | | | 2,12 | | |

Figure 3.6: Evolution of variable $\psi_t$ and $\hat{\psi}_t$ for wheat, full sample estimate

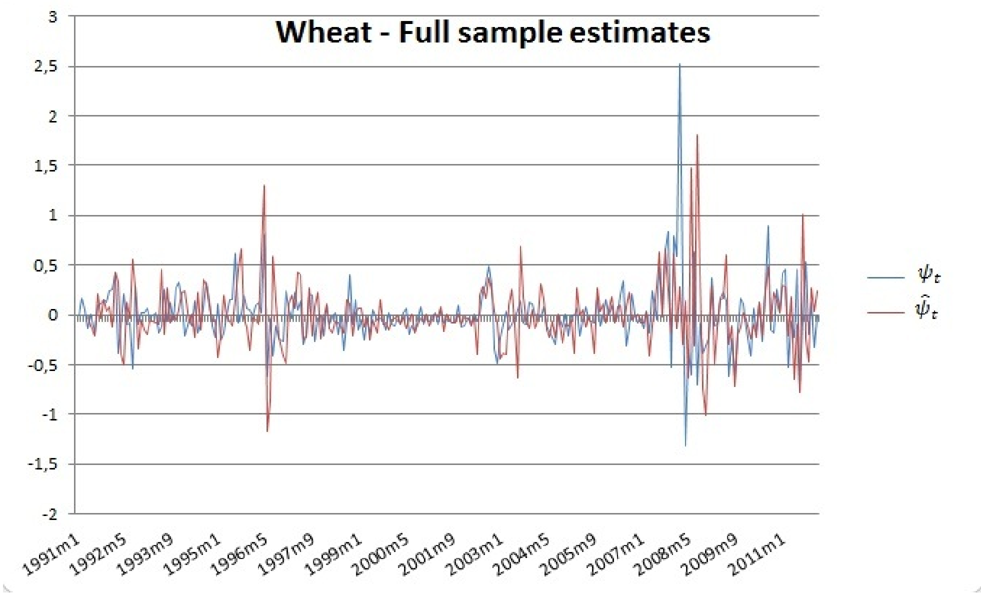


Figure 3.7: Evolution of variable $\psi_t$ and $\hat{\psi}_t$ for wheat, 1990-2006 subperiod estimate

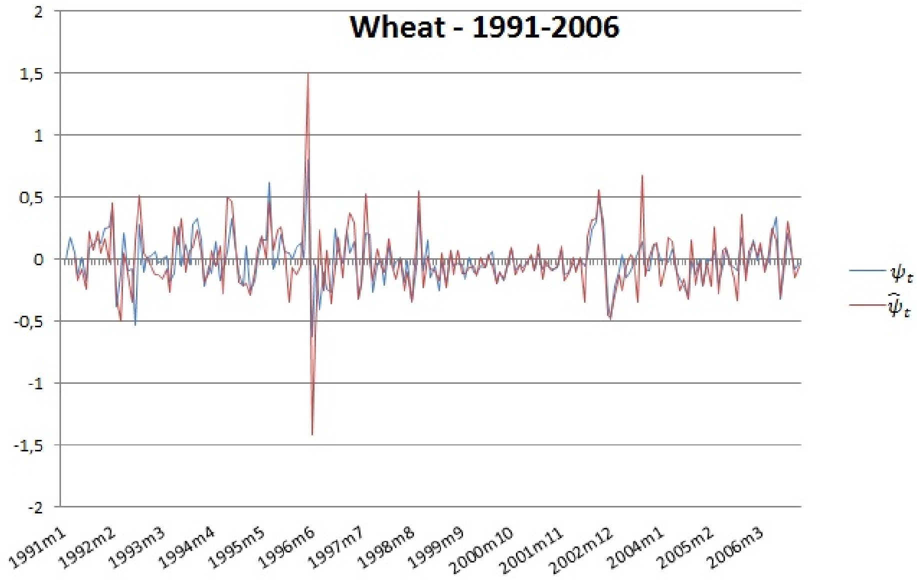



Figure 3.8: Evolution of variable $\psi_t$ and $\hat{\psi}_t$ for oil, full sample estimate

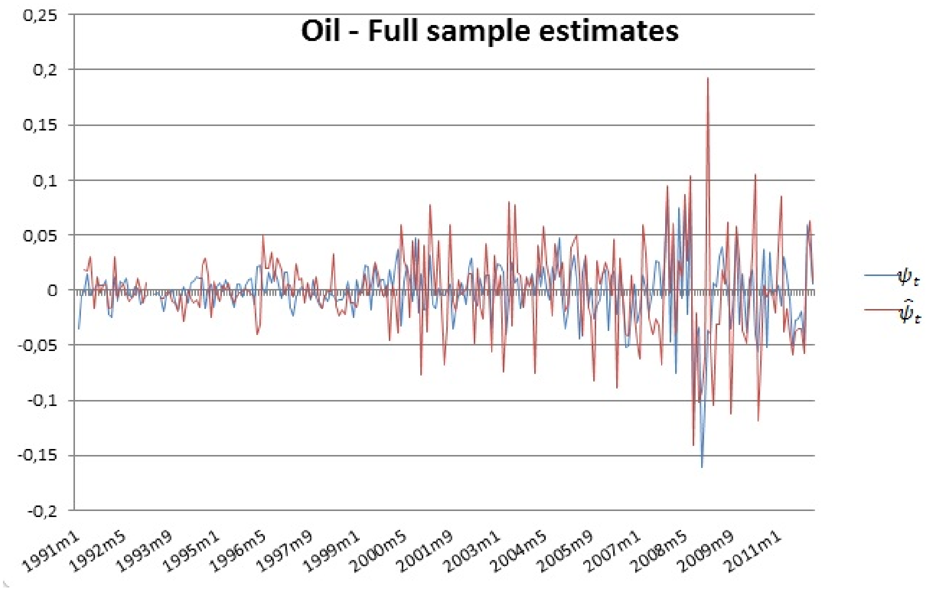
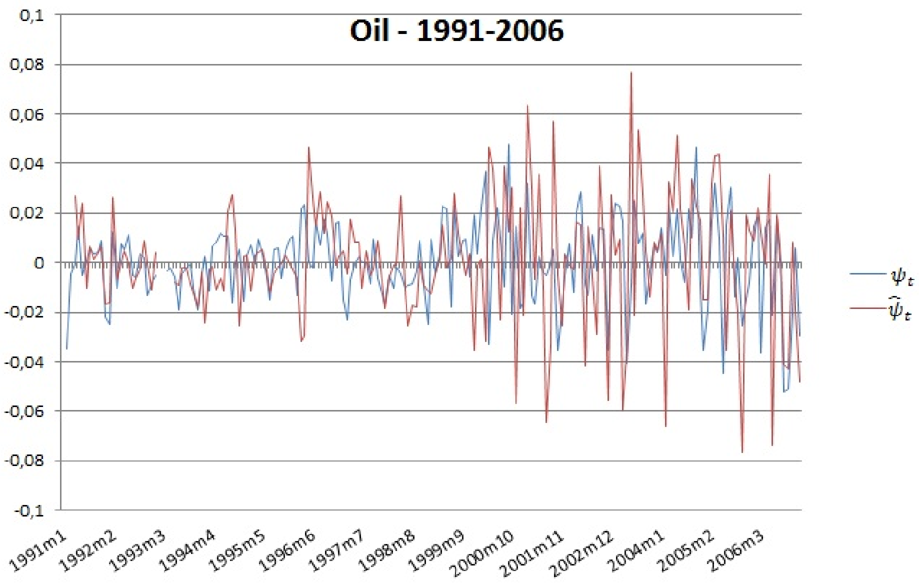


Figure 3.9: Evolution of variable $\psi_t$ and $\hat{\psi}_t$ for oil, 1991-2006 subperiod estimate



# Bibliography

Price formation in financialized commodity markets: The role of information. Technical report, United Nations, 2011.

Jushain Bai. Estimation of a Change Point in Multiple Regression Models. *The Review of Economics and Statistics*, 79(4):551–563, 1997.

Michael J Brennan. The Supply of Storage. *The American Economic Review*, 48(1):50–72, 1958.

Celso Brunetti and Bahattin Buyuksahin. Is Speculation Destabilizing ? 2009.

Bahattin Buyuksahin and Michel A. Robe. Speculators, Commodities and Cross-Market Linkages. 2010.

John Y Campbell and Robert J Shiller. Cointegration and Tests of Present Value Models. *Journal of Political Economy*, 95(5):1062–1088, 1987.

Dietrich Domanski and Alexandra Heath. Financial investors and commodity markets. *BIS Quarterly Review*, March 2007. URL `http://ideas.repec.org/a/bis/bisqtr/0703g.html`.

James Thomas Einloth. Speculation and Recent Volatility in the Price of Oil. *SSRN*

*Electronic Journal*, pages 1–25, 2009. ISSN 1556-5068. doi: 10.2139/ssrn.1488792. URL http://www.ssrn.com/abstract=1488792.

C. L. Gilbert. Speculative Influences on Commodity Futures Prices 2006-08. 2009.

James D. Hamilton. Causes and consequences of the oil shock of 2007-08. Working Paper 15002, National Bureau of Economic Research, May 2009. URL http://www.nber.org/papers/w15002.

John Hull. *Options, Futures and Other Derivative Securities*. 1993.

S. H. Irwin and D. R. Sanders. Index Funds, Financialization, and Commodity Futures Markets. *Applied Economic Perspectives and Policy*, 33(1):1–31, Feb 2011. ISSN 2040-5790. doi: 10.1093/aepp/ppq032. URL http://aepp.oxfordjournals.org/lookup/doi/10.1093/aepp/ppq032.

Nicholas Kaldor. Speculation and Economic Stability. *The Review of Economic Studies*, 7 (1):1–27, 1939.

Paul Krugman. More on Oil and Speculation. *New York Times*, (1): 1–31, May 13. 2008. URL http://krugman.blogs.nytimes.com/2008/05/13/more-on-oil-and-speculation.

Robert S Pindyck. Inventories and the Short-run Dynamics of Commodity Prices. 1990.

Robert S Pindyck. The Present Value Model of Rational Commodity Pricing. *The Economic Journal*, 103(418):511–530, 1993.

Dwight R. Sanders, Scott H. Irwin, and Robert P. Merrin. The adequacy of speculation in agricultural futures markets: Too much of a good thing? *Applied Economic Perspectives and Policy*, 32(1):77–94, 2010. doi: 10.1093/aepp/ppp006. URL http://aepp.oxfordjournals.org/content/32/1/77.abstract.

Ke Tang and Wei Xiong. Index Investment and Financialization of Commodities. 2010.

Lester G Telser. Futures Trading and the Storage of Cotton and Wheat. *Journal of Political Economy*, 66(3):233–255, 1958.