The Academy of Economic Studies
The Faculty of Finance, Insurance, Banking and Stock Exchange
Doctoral School of Finance and Banking (DOFIN)

Dissertation Paper

Evidence of the unspanned stochastic volatility in crude-oil market

Student: Răzvan Daniel Tudor
Supervisor Professor: PhD. Moisă Altăr

Bucharest, July 2008

Abstract

The purpose of this dissertation paper is to conduct a comprehensive analysis of unspanned stochastic volatility in commodity markets with focus and empirical evidence on crude-oil market. Using crude-oil futures and options on futures data from *New York Mercantile Exchange (NYMEX)* there are presented model-free results that strongly suggest the presence of unspanned stochastic volatility in the crude-oil market. Sharp oil prices changes exert influence on macroeconomic activity in general and crude-oil industry in particular. The importance of the results is that they show the extent to which volatility risk is spanned by the futures contracts. The extent to which crude-oil futures contracts trading span volatility will indicate if options on futures are redundant securities or there is needed a mixed strategy combining both types of crude-oil market derivatives (futures and options) to fully hedge against volatility risk.

1. Introduction

Over the last few years, the persistent sharp oil prices changes in both the spot and futures markets have represented perhaps the most striking challenge to the forecasting abilities of private and public institutions worldwide. From the demand side increasing crude-oil prices led to new challenges in hedging against volatility risk.

While volatility is clearly stochastic, it is not clear to what extent volatility risk can be hedged by trading in the commodities themselves or, more generally, their associated futures contracts, forward or swap contracts, in other words, the extent to which volatility is spanned.

Existing equilibrium models from commodity markets imply that volatility risk is largely spanned by the futures contracts. Mainly, they suggest that market volatility is embedded in inventories which are the basis for futures price formation. Therefore by construction futures offer a high degree of volatility spanning.

The consequence of these models is that they imply that options on futures are redundant securities. In spite of this, the data provided form Bank of International Settlements – BIS – strongly suggests that the market for commodity derivatives has exhibited phenomenal growth over the past few years. For exchange-traded commodity derivatives, the BIS estimates that the number of outstanding contracts more than doubled from 12.4 million in June 2003 to 32.1 million in June 2006. For over-the-counter (OTC) commodity derivatives, the growth has been even stronger with the BIS estimating that, over the same period, the notional value of outstanding contracts increased five-fold from USD 1.04 trillion to USD 6.39 trillion. Importantly, a large and increasing fraction of the commodity derivatives are options (as opposed to futures, forwards and swaps). According to BIS statistics, options now constitute over one-third of the number of outstanding exchange-traded contracts and almost two-thirds of the notional value of outstanding OTC contracts.

The purpose of this paper is to show that if, for a given commodity, volatility contains important unspanned components it cannot be fully hedged and risk-managed using only the underlying instruments and options are not redundant securities.

The unspanned stochastic volatility evidence research is conducted in crude-oil market because it is by far the most liquid commodity derivatives market. The data for the analysis was provided by New York Mercantile Exchange – NYMEX – and contains a large set of futures and options on futures contracts prices. Since volatility is not directly observable I will use, for different options maturities, straddle returns and implied volatility of the at-the-money options straddles as proxies for the true volatility and I will show the extent to which futures contracts span volatility. If volatility is completely spanned by trading in futures contracts then the equilibrium models for commodity markets are correct in assuming that commodities futures prices formation incorporates market volatility. If shown on contrary, it means that options on futures are not redundant securities and their role is to extend the degree of hedging which futures contracts traditionally offer.

The reason for choosing these two volatility proxies is that straddle returns are not conditioned on a particular pricing model. Returns are obtained from daily options on futures market prices from NYMEX. While using the implied volatility, though it might be more accurate, it involves using a pricing model.

Previously, this approach was used to evidence the unspanned stochastic volatility in fixed-income market, more specifically to show the extent to which trading of bonds span the term structure of interest rates.

The dissertation paper is organized as follows. Section 2 contains a literature review of the models which treated the stochastic volatility in commodity and financial markets. Section 3 briefly presents the crude-oil derivatives data used in this paper and the computational aspects behind data which was used as input for the model. In section 4 the paper contains the model used to evidence of the unspanned stochastic volatility in crude-oil market. Section 5 presents model estimation and analysis. Finally, in Section 6 there are to be found the conclusion which can be drawn from this paper. Section 7 contains the reference list and Section 8 the relevant additional information – Annexes - which are mentioned in the paper content.

2. Literature Review

The first equilibrium models from commodity markets implied that futures contracts provide insurance against price volatility, the level of inventories being negatively related to the required risk premium of commodity futures. The starting point of these models was the traditional Theory of Storage originally proposed by Kaldor (1939). The theory provides a link between the term structure of futures prices and the level of inventories of commodities. This link, also known as "cost of carry arbitrage," predicts that in order to induce storage, futures prices and expected spot prices of commodities have to rise sufficiently over time to compensate inventory holders for the costs associated with storage. Developments in this area were made by Deaton and Laroque (1992), Chambers and Bailey (1996), Routledge, Seppi and Spatt (2000). Their models predict a link between the level of inventories and future spot price volatility. Inventories act as buffer stocks which can be used to absorb shocks to demand and supply, thus dampening the impact on spot prices. Deaton and Laroque show that at low inventory levels, the risk of "stock-out" (exhaustion of inventories) increases and expected future spot price volatility rises. In an extension of the Deaton and Laroque model which includes a futures market, RSS show how the shape of the futures curve reflects the state of inventories and signals expectations about future spot price volatility. DL (1992) and RSS (2000) have explained the existence of a convenience yield as arising from the probability of a stock-out of inventories. Because they study storage in a risk-neutral world, risk premiums are zero by construction, and futures prices simply reflect expectations about future spot prices.

Another reference model, the model in Litzenberger and Rabinowitz (1995) and Ng and Pirrong (1994), incorporates the embedded option in reserves of extractable resource commodities. Finally it has similar implications. The relationship between volatility and the slope of the futures - Litzenberger and Rabinowitz (1995) showed it for crude oil, and Ng and Pirrong (1994), for metals - show that the degree of backwardation is indeed positively related to volatility, implying that volatility does contain a component that is spanned by the futures contracts. However, whether volatility also contains important unspanned components was not shown.

Other papers, which emphasize production/extraction and investment decisions for the formation of futures prices, include those of Casassus, Collin-Dufresne, and Routledge (2003), Kogan, Livdan, and Yaron (2005) and Carlson, Khoker, and Titman (2006).

In their paper, Gordon, Hayashi and Rouwenhorst (2005), analyzed the fundamentals of commodity futures returns and predicted a link between the state of inventories, the shape of the futures curve, and expected futures risk premiums. They showed that show that the convenience yield is a decreasing, non-linear relationship of inventories and also linked the current spot commodity price and the current (nearest to maturity) futures price to the level of inventories, and empirically documented the nonlinear relationship predicted by the existence of the non-negativity constraint on inventories. In particular, they showed that low inventory levels for a commodity are associated with an inverted ("backwardated") term structure of futures prices, while high levels of inventories are associated with an upward sloping futures curve ("contango").

The existence of unspanned volatility factors was first evidenced in fixed income market. Collin-Dufresne and Goldstein (2002) and Heidari and Wu (2003) defined unspanned stochastic volatility as being those factors driving Cap and Swaption implied volatilities that do not drive the term structure of interest rates. In other words they showed that trading in underlying bonds do not span the term structure of interest rates. There are embedded factors in Cap and Swaption that bonds do not contain and make them more valuable in hedging against interest rates volatility risk That is, in contrast to the predictions of standard short-rate models, bonds do not span the fixed income market.

Using Collin-Dufresne and Goldstein (2002) approach, Trolle and Schwartz (2006) extended the problem with existence of unspanned stochastic volatility to commodity markets. They developed a tractable model for pricing commodity derivatives in the presence of unspanned stochastic volatility. The model features correlations between innovations to futures prices and volatility, quasi-analytical prices of options on futures and futures curve dynamics in terms of a low-dimensional affine state vector. Their evidence was on crude-oil market due to its liquidness and showed that in the presence of unspanned stochastic volatility factors options are not redundant securities. The model and the evidence could be extended as well on the other commodity markets.

Richter and Sorensen (2007) have a work in progress for a stochastic volatility model in the presence of unspanned volatility factors for the soybean market.

3. Overview of the Data

As mentioned before, crude-oil market is the most liquid commodity market. The data used in this paper was delivered by New York Mercantile Exchange - NYMEX. It contains large data set of futures and options on futures prices with different maturities and strike prices. The futures data contains daily prices for futures contracts starting with January 1987 and ending with May 2008. Since options on futures prices were available for research purposes only for June 2002 – December 2006 period, I chose to use the futures contracts prices for the same interval.

The NYMEX futures contract trades in units of 1,000 barrels, and the delivery point is Cushing, Oklahoma, which is also accessible to the international spot markets via pipelines. The contract provides for delivery of several grades of domestic and internationally traded foreign crude, and serves the diverse needs of the physical market. The NYMEX symbol for light-sweet crude-oil is CL. Crude oil futures are listed nine years forward using the following listing schedule: consecutive months are listed for the current year and the next five years; in addition, the June and December contract months are listed beyond the sixth year. Additional months will be added on an annual basis after the December contract expires, so that an additional June and December contract would be added nine years forward, and the consecutive months in the sixth calendar year will be filled in. The futures expire on the third business day prior to the 25th calendar day of the month proceeding the delivery month. If the 25th calendar day of the month is a non-business day, expiration is on the third business day prior to the business day proceeding the 25th calendar day.

For my purpose I extracted from various maturities only futures contracts with time-to-maturity 1 Month, 3 Months, 6 Months, 9 Months and 1 Year. The reason is that crude-oil spot prices established on spot markets are not available. Mainly, spot prices are settled on one to one transactions between partners based on current market conditions. Therefore the 1 Month time-to-maturity futures contract serves as a proxy for the crude-oil spot prices. The 1 Year time-to-maturity futures prices are continuous during the June 2002 – December 2006 sample so I chose not to include them in the analysis. The quarterly maturities correspond to the traditional hedging strategy of a crude-oil refining based company. Being given the optimal refining capacity usually the company engages in a rolling futures contract with quarterly maturities providing the company with the necessary crude oil amount at a certain price which can be used for financial forecasts.

While futures offer price protection by allowing the holder of a futures contract to lock in a price level, a major advantage of options is that the holder of an options contract is afforded price protection, but still has the ability to participate in favorable market moves. Because the buyer of an options contract has the options contract but not the market moves against a position, and a trader holds on to this option, the maximum cost is the price he has already paid for the option.

On the other hand, if the market moves in favor of a position, the virtually unlimited profit potential to the buyer of an options contract is parallel to a futures position, net of the premium paid for the options contract. Therefore, protection from unfavorable market moves is achieved at a known cost, without giving up the ability to participate in favorable market moves. Options on futures contracts expire three business days prior to the expiration of the underlying futures.

For the research I chose crude-oil calendar spread options on futures. The reason is that calendar spread options are the most traded crude-oil options derivatives on NYMEX and thus the results of my study will be more representative. Also they imply delivery of the underlying asset as opposed to other derivatives which are only settled in cash, for example European Style options – NYMEX symbol LO. Their NYMEX trading symbol for calendar spread options is WA. The contract is simply an options contract on the price differential between two delivery dates for the same commodity. The price spread between contract months can be extremely volatile because the energy markets are more sensitive to weather and news than any other market. A widening of the month-to-month price relationships can expose market participants to severe price risk which could adversely affect the effectiveness of a hedge or the value of inventory. The calendar spread options can allow market participants who hedge their risk to also take advantage of favorable market moves.

For the corresponding three maturities of the futures prices I extracted corresponding calendar spread straddles. One straddle consists of a call and a put option with the same strike. More, I chose the at-the-money straddles since they are sensitive to market volatility ("Vegas" peak for the at-the-money straddles). At-the-money property of an option means that the option has the strike price equal or near to spot price.

This computation is helpful to extract from the whole options and futures sample the data I need for the evidence of unspanned stochastic volatility.

4. The Model

If equilibrium models are correct and changes in crude oil prices are spanned by the futures contracts then in order to hedge against volatility risk one may construct a portfolio of futures contracts for this purpose.

One simple alternative to evidence if the conclusion of these models is correct is to simply regress changes in volatility from crude oil markets on futures contracts prices and see if they fully explain volatility changes. But volatility in crude oil market as well as in other commodity and financial markets is stochastic and not directly observable. Therefore I will use two reasonable proxies for the true and unobservable volatility – at-the-money calendar spread straddles prices and at-the-money calendar spread straddles implied volatility.

A straddle consists of a call and a put option on the same underlying with the same strike. When purchasing a straddle the investor expects the market to spike in either direction. This is the case of long at-the-money straddle strategy which will be used throughout this paper as opposite to short at-the-money straddle which is used when market is expected to be quiet (expecting minor changes in volatility). Therefore we can say that by purchasing a straddle the investor trades volatility. Straddle profits are unlimited in either direction while losses are limited to the premium paid for both options which form the straddle.

The reason for selecting straddles as volatility proxies is the straddle Greeks indicators. The price of a near-ATM straddle has low sensitivity to variations in the price of the underlying futures contract (since "deltas" are close to zero for ATM straddles) but high sensitivity to variations in volatility (since "Vegas" peak for ATM straddles).

Delta (Δ) and Vega (ν) for at-the-money straddles:

$$\Delta = \frac{\partial V}{\partial S} = \begin{cases} e^{-qT} * \phi(d_2), Call \\ -e^{-qT} * \phi(-d_1), Put \end{cases}$$

$$\nu = \frac{\partial V}{\partial \sigma} = S * e^{-qT} \phi(d_1)\sqrt{\tau}$$, for both Call and Put options.

The indicators were derived from the Black-Scholes option pricing formula where:

V – Value of the option;

S – Stock price;

q – Annual dividend yield;

τ - Time to maturity (T-t);

σ - Volatility;

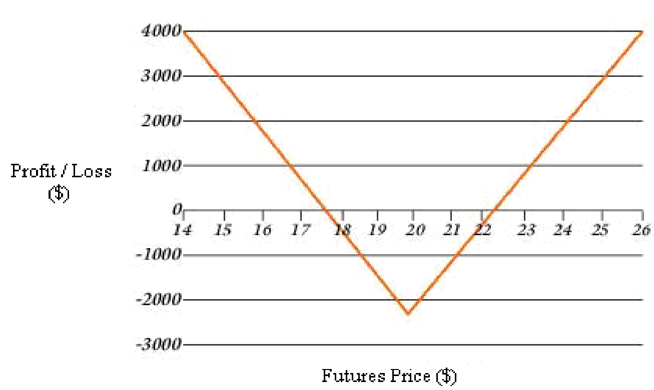
r – Risk free rate;

$\Phi(d_1)$ - The probability of exercise under the equivalent exponential martingale probability measure and the equivalent martingale:

$$\phi(x) = \int_{-\infty}^{x} \frac{e^{-\frac{y^2}{2}}}{\sqrt{2\pi}} dy = \int_{-x}^{\infty} \frac{e^{-\frac{y^2}{2}}}{\sqrt{2\pi}} dy$$

$$y = \frac{\ln(S/K) + (r - q + \sigma^2/2) * \tau}{\sigma\sqrt{\tau}}$$

Bellow graphic exhibits the typical payoff function of an at-the-money straddle:



To avoid the non-stationary problem with straddle and futures prices, which is common to almost all asset prices I will use for further analysis straddle returns and futures returns.

Straddle returns are computed as follows:

$$r_{straddle,i} = \begin{cases} S_i - K - (\pi_{call} + \pi_{put}), S_i > K + (\pi_{call} + \pi_{put}) \\ K - S_i - (\pi_{call} + \pi_{put}), S_i + (\pi_{call} + \pi_{put}) < K \\ 0, else \end{cases}$$

Where:

S_i - The spot price of the underlying commodity. For calendar spread options type the underlying commodity spot price is the price differential from current market price and the futures price of the futures contract which at the maturity of the option.

K – The strike price;

π_{call}, π_{put} - Call and Put option prices;

The futures contracts returns are simply computed this way:

$$futures\_returns_{i,jMonth} = Spot\Pr ice_i - FuturesContract\Pr ice_{i,jMonth}$$

Where:

$Spot\Pr ice_i$ - The crude-oil spot price. In absence of a transparent spot market the spot price is computed as the price of the futures contract with shortest time to maturity, the contract with expiration the following month.

$FuturesContract\Pr ice_{i,jMonth}$ - The today observed market price of the futures contract with expiration in “j” months. As mentioned before j = 3, 6, 9 Months.

There are three alternatives to evidence the presence of unspanned stochastic volatility in crude oil market:

- Investigate how much of the variation in the prices of derivatives highly exposed to stochastic volatility (so-called “straddles”) can be explained by variation in the underlying futures prices;
- Investigate how much of the variation in implied volatilities (which is related to expectations under the risk-neutral measure of future volatility) can be explained by variation in the underlying futures prices;

- Investigate how much of the variation in realized volatility, estimated from high frequency data, can be explained by variation in the underlying futures prices.
- Investigate how much of the volatility of the variance swaps can be explained by variation in the underlying futures prices.

Unfortunately high-frequency data is available only for calendar spread options. Also variance swaps are quite illiquid in the market therefore I will use only the first two approaches for evidence.

For the approach which requires the use of straddle implied volatility, this is computed as the average of straddle component put and call options implied volatilities. The put and call implied volatilities are obtained from put and call formulas of the Black-Scholes model.

Briefly, the formulas for call and put options derived from the Black-Scholes partial differential equation are:

$$\frac{\partial V}{\partial t}+\frac{1}{2}\sigma^2 S^2\frac{\partial^2 V}{\partial S^2}+rS\frac{\partial V}{\partial S}-rV=0 \text{ (Black-Scholes PDE)}$$

$$C(S,T)=S\phi(d_1)-Ke^{-rT}\phi(d_2)$$

$$P(S,T)=Ke^{-r(T-t)}\phi(-d_2)-S\phi(-d_1)$$

Where:

$$d_1=\frac{\ln(S/K)+(r-q+\sigma^2/2)*\tau}{\sigma\sqrt{\tau}}$$

$$d_2=\frac{\ln(S/K)+(r-\sigma^2/2)*T}{\sigma\sqrt{\tau}}=d_1-\sigma\sqrt{T}$$

Φ is the cumulative distribution function. $\Phi(d_1)$ and $\Phi(d_2)$ are the probabilities of exercise under the equivalent exponential risk neutral measure and the equivalent risk neutral probability measure, respectively.

Being given the call and put option prices, the underlying futures price, the current and maturity date, the strike price and the risk free rate, the implied volatility is computed using the Newton-Raphson method. It is a root-finding algorithm that uses the first few terms of the Taylor series of a function $f(x)$ in the vicinity of a suspected root. Newton's method is also known as Newton's iteration.

The Taylor series of $f(x)$ about the point $x = x_0 + \varepsilon$ is given by:

$$f(x_0 + \varepsilon) = f(x_0) + f'(x_0)\varepsilon + \frac{1}{2} f''(x_0)\varepsilon^2 + ...$$

Keeping terms only to first order:

$$f(x_0 + \varepsilon) \approx f(x_0) + f'(x_0)\varepsilon$$

This expression can be used to estimate the amount of offset ε needed to land closer to the root starting from an initial guess x_0. Setting $f(x_0 + \varepsilon) = 0$ and solving the above equation for $\varepsilon \equiv \varepsilon_0$ gives:

$$\varepsilon_0 = -\frac{f(x_0)}{f'(x_0)}$$

This is the first-order adjustment to the root's position. By letting $x_1 = x_0 + \varepsilon$, calculating a new ε_1, and so on, the process can be repeated until it converges to a fixed point (which is precisely a root) using:

$$\varepsilon_0 = -\frac{f(x_0)}{f'(x_0)}$$

Therefore with a good initial choice of the root's position, the algorithm can be applied iteratively to obtain:

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$$

Many commodity markets as well as financial markets are characterized by a high degree of collinearity between returns. In order to extract the most uncorrelated sources of variation in a multivariate system I will use principal components analysis (PCA) for the futures returns.

Mainly, principal components analysis objective is to:

- Reduce dimensionality by taking into account only the most relevant principal components from the whole data set;

- Avoid near multicollinearity issues for the returns and use in further analysis only uncorrelated components. Also, these components characterize the data and are useful for drawing conclusions;

Mathematical background:

The data input to principal component analysis must be stationary. Principal component analysis is based on eigenvalues and eigenvector analysis of V =X'XT, the k x k symmetric matrix of correlations between the variables in X. Each principal component is a linear combination of these columns, where the weights are chosen in such way that:

- the first principal component explains the greatest amount of the total variation in X, the second component explains the greatest amount of the remaining variation, and so on;
- the principal components are uncorrelated to each other;

Denoting by W the k x k matrix of eigenvectors of V. Thus:

$$VW=W\Lambda$$

Where Λ is the k x k diagonal matrix of eigenvalues of V. Then we order the columns of W according to the size of corresponding eigenvalue. Thus if $W=(w_{ij})$ for i,j=1,…,k then the m-th column of W, denoted $w_m=(w_{1m},...,w_{km})$, is the k x 1 eigenvector corresponding to the eigenvalue λ_m and the column labeling has been chosen so that $\lambda_1>\lambda_2>...>\lambda_k$.

Therefore the m-th principal component of the system is defined by:

$$P_m=w_{1m}X_1+w_{2m}X_2+...+w_{km}X_k$$

Where X_i denotes the i-th column of X, the standardized historical input data on the i-th variable in the system. In matrix notation the above definition becomes:

$$P_m=Xw_m$$

Each principal component is a time series of the transformed X variables, and the full T x m matrix of principal components, which has P_m as its m-th column, may be written as:

P=XW

The procedure leads to uncorrelated components because:

P'P=W'X'XW=TW'W Λ

W is an orthogonal matrix, which means $W^{'} = W^{-1}$ and so P'P=T Λ. Since this is a diagonal matrix the columns of P are uncorrelated, and the variance of the m-th principal component is λ_m (sum of eigenvalues). However, the sum of the eigenvalues is k, the number of variables in the system. Therefore, the proportion of variation explained by the first n principal components together is:

$$\sum_{i=1}^{n} \lambda_i / k$$

Because of the choice of labeling in W the principal components have been ordered so that P_1 belongs to the first and largest eigenvalue λ_1, P_2 belongs to the first and largest eigenvalue λ_2, and so on. In a highly correlated system the first eigenvalue will be much larger than the others, so the first principal component alone will explain a large part of variation.

Since $W^{'} = W^{-1}$, is equivalent to X=PW', that is:

$$X_i = w_{i1}P_1 + w_{i2}P_2 + ... + w_{ik}P_k$$

Thus each vector of the data input may be written as a linear combination of the principal components.

To sum up principal component analysis it is a way of identifying patterns in data, and expressing the data in such a way as to highlight their similarities and differences. Since patterns in data can be hard to find in data of high dimension, where the luxury of graphical representation is not available, principal components analysis is a powerful tool to achieve this.

The principal components analysis will be illustrated in the next section on the highly correlated crude-oil futures prices returns.

In order to evidence the presence of unspanned stochastic volatility in crude-oil market I will use the first two approaches mentioned above: investigate how much of straddle returns and straddle implied volatilities variation can be explained by the variation of the futures returns.

The evidence procedure consists of three steps:

a) The first step is principal components analysis of the correlation matrix of daily futures returns. We retain all the principal components identified in the analysis in an attempt to not exclude from evidence any source of variation embedded in a component, though that component might have minor significance.

b) For each futures contract "i" I regress the daily closest to the at-the-money straddle returns on the futures returns principal components.

For each futures contract "i" I regress the daily straddle implied volatilities on futures returns principal components. The daily straddle implied volatility is related to the average expected (under the risk-neutral measure) volatility of the underlying futures contract over the life of the option.

Since in commodity and financial markets the returns dependency is rarely linear I will introduce in the regression equation the squared principal components also, in an attempt to take into account non-linearities between straddle returns and implied volatilities and futures returns. Therefore if I take into account just one principal component the regression equation to catch non-linearity will be:

$$y = \alpha + \beta_1 x + \beta_2 x^2 + \varepsilon$$

The coefficient for the squared principal component will be important (as long as it is significant) just for the sign indicating the convexity or concavity of the dependency.

Another aspect is taking into account the cross-products dependencies of the straddle returns and implied volatility. These dependencies reflect changes in the marginal effect of one explanatory variable given others. Considering straddle returns and first two principal components the transformation can be written as:

$$y = \alpha + \beta_1 x + \beta_2 w + \beta_3 x^2 + \beta_4 w^2 + \beta_5 xw + \varepsilon$$

By rewriting the equation above as:

$$y = (\alpha + \beta_2 w + \beta_4 w^2) + x(\beta_1 + \beta_3 x + \beta_5 w) + \varepsilon$$

We can interpret the intercept as a function of *w* and the slope of *x* as changing with *w* and *x*.

To sum up, taking into consideration both squared components and cross-product between components the regression equations for both approaches may be written as:

Straddle returns regression:

$$y^i = \alpha^i + \beta^i{}_1 x_1 + \beta^i{}_2 x_2 + \beta^i{}_3 x_3 + \beta^i{}_4 x_1{}^2 + \beta^i{}_5 x_2^2 + \beta^i{}_6 x_3^2 + \beta^i{}_7 x_1 x_2 + \beta^i{}_8 x_1 x_3 + \beta^i{}_9 x_2 x_3 + \varepsilon^i{}_t$$

Implied volatility regression:

$$z^i = \alpha^i + \beta^i{}_1 x_1 + \beta^i{}_2 x_2 + \beta^i{}_3 x_3 + \beta^i{}_4 x_1{}^2 + \beta^i{}_5 x_2^2 + \beta^i{}_6 x_3^2 + \beta^i{}_7 x_1 x_2 + \beta^i{}_8 x_1 x_3 + \beta^i{}_9 x_2 x_3 + \varepsilon^i{}_t$$

Where:

$x_i, i = 1,2,3$ - The principal components of the futures return data. They numbered according to they variation explanatory power from principal component with highest eigenvalue to the principal component with smallest eigenvalue.

y^i - The straddle returns at "i" maturity. In my case i = 3, 6, 9 Months.

z^i - The implied volatility of the straddles at "i" maturity.

Both regressions will indicate the extent to which volatility is spanned by the futures contracts.

c) Finally, I will analyze the principal components of the time series of residuals from the straddle return regressions and the implied volatility regressions. The principal components of the residuals are, by construction, independent of those of the futures returns. If there is unspanned stochastic volatility in the data, there should be at least one significant explanatory principal component for the variation due to unspanned factors. If the residuals are simply due to noisy data, there should not be one principal component with high explanatory power among residuals.

5. The Model Estimation and Analysis

Commodity futures prices are characterized by some important properties:

- Commodity futures prices are often "backwardated" in that they decline with time to delivery,

- Spot and futures prices are mean reverting;

- Commodity prices are strongly heteroscedastic and price volatility is correlated with the degree of backwardation;

- Unlike financial assets, many commodities have pronounced seasonality in both price levels and volatilities.

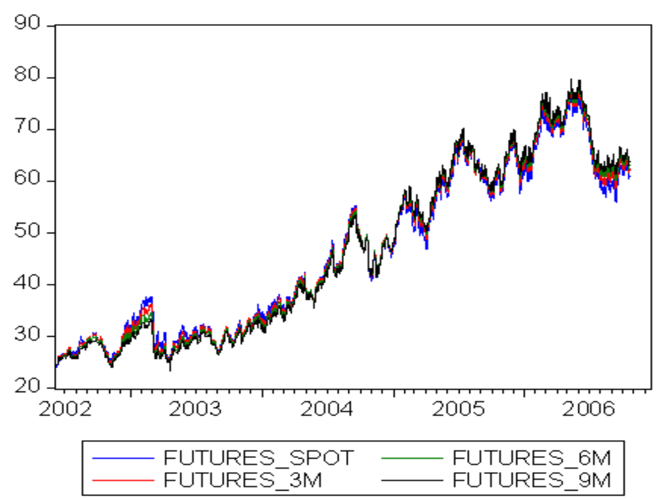
Being given S(t) the time-t crude-oil spot price and F(t, T) [P(t, T)] the time-t price of a crude-oil futures contract [zero-Coupon bond] with maturity T - t. The futures contract is backwardated if S(t) - P(t, T)F(t, T) > 0 and strongly backwardated if S(t) - F(t, T) > 0.

For our futures date the results confirm the above "backwardation" property:

Backwardation type vs. Maturity of the Futures Contract	**3 Months**	**6 Months**	**9 Months**
Backwardation Degree(%)	45.6	52.7	55.3
Strongly Backwardation Degree(%)	94.3	95.4	96.2

Table 1 – The simple and strong "backwardation" degree

As time to maturity increases so does the backwardation degree. If we take a look on the futures prices graphical representation for various maturities we see that clearly the market was in contango, although market expectations derived from the prices of futures contracts for the same maturities were bearish.



Graph 1 – Futures prices

The strongly heteroscedasticity property of commodities, which can be translated as the property of futures prices to have time dependant functions for mean ($\mu(t)$) and variance ($\sigma^2(t)$) poses a serious problem for out further econometric estimations.

In order to test the validity of property I will use Augmented Dickey-Fuller (ADF) test for the 3 Months futures prices. The ADF test is a unit root test which is carried out by estimating the following equation:

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \delta_1 \Delta y_{t-1} + ... + \delta_1 \Delta y_{t-p} + \varepsilon_t$$

This is the most restrictive form of the test, which includes the intercept (α) and the trend (β). The null hypothesis is that the coefficient of the level variable (γ) is 0, which means the series is non-stationary, or less than 0 otherwise. I carried out the test for the futures prices with 3 Month maturity in levels using only the intercept. The results were:

Null Hypothesis: FUTURES_3M has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic based on SIC, MAXLAG=22)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-1.200488	0.6763
Test critical values:	1% level	-3.435876	
	5% level	-2.863868	
	10% level	-2.568060	

*MacKinnon (1996) one-sided p-values.

Table 2 – The ADF test results for 3 Months Futures Prices

The value of the ADF test is larger than the critical values for all levels of confidence meaning that we cannot reject the null hypothesis of the futures prices series being non-stationary.

Therefore I will further use futures returns instead of futures prices. Futures returns are computed as shown in 4th section as:

$$futures\_returns_{i,jMonth} = SpotPrice_i - FuturesContractPrice_{i,jMonth}$$

Building the futures returns time series offer the advantage of stationary. Indeed if we carry out once again the ADF test for the 3 Months futures returns, the value of the ADF test will reject the null hypothesis.

Null Hypothesis: FCR_3M has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 0 (Automatic based on SIC, MAXLAG=22)

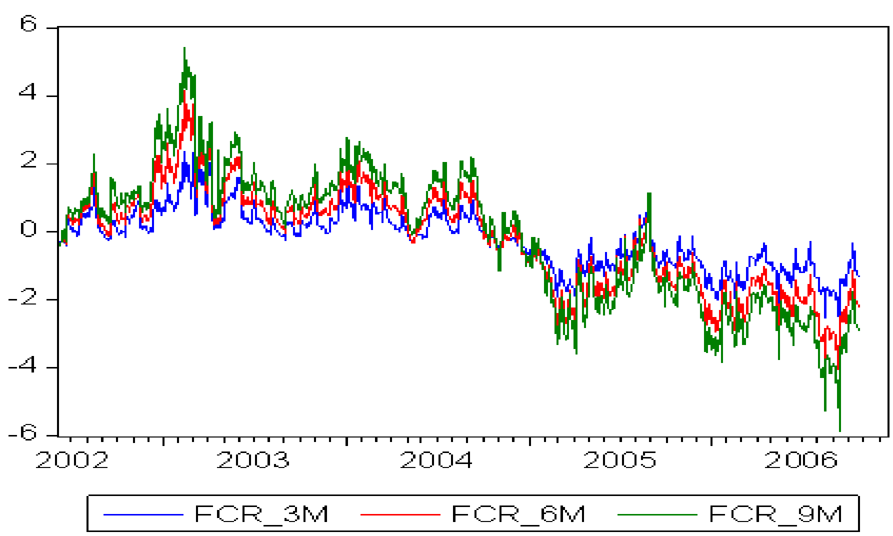
		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-6.498157	0.0000
Test critical values:	1% level	-3.966124	
	5% level	-3.413762	
	10% level	-3.128951	

*MacKinnon (1996) one-sided p-values.

Table 3 – The ADF test results for 3 Months Futures Returns

The ADF unit root tests for the 6 Months and 9 Months futures returns may be found in the Annex section of this paper (Table 13, 14, 15).

The graphical representation of the crude-oil futures returns for the chosen maturities show that in crude-oil market futures returns are highly correlated.



Graph 2 – Futures returns (3 Months, 6 Months, 9 Months)

This highly correlation suggests that for example the 9 Months futures returns are not only influenced by the crude-oil spot price but also by intermediate maturities futures returns. Next I will present the futures returns correlation matrix. The correlation coefficients are closed to 1 indicating as well a high correlation degree.

	FCR_3M	FCR_6M	FCR_9M
FCR_3M	1.000000	0.984578	0.965624
FCR_6M	0.984578	1.000000	0.995120
FCR_9M	0.965624	0.995120	1.000000

Table 4 – Futures returns correlation matrix

If we examine the first column of the model we see that the correlation degree tends to decrease with maturity though very slightly.

Since crude-oil futures returns are explanatory variables in my classical linear regression model, the highly correlated returns pose the problem of near multicollinearity. In this case, it is not possible to estimate all of the "betas" from the model. In the presence of multicollinearity, it will be hard to obtain small standard errors.

Therefore I will use futures returns principal components analysis as a solution for the near multicollinearity problem.

The starting point in identifying the futures returns principal components is the futures returns correlation matrix. Bellow there is presented the eigenvalues and eigenvectors of the futures returns correlation matrix. The eigenvectors are ordered after the corresponding eigenvalue, starting with the highest.

Date: 07/06/08 Time: 17:16
Sample (adjusted): 6/10/2002 10/20/2006
Included observations: 1140 after adjustments
Correlation of FCR_3M FCR_6M FCR_9M

	Comp 1	Comp 2	Comp 3
Eigenvalue	2.963599	0.035471	0.000930
Variance Prop.	0.987866	0.011824	0.000310
Cumulative Prop.	0.987866	0.999690	1.000000

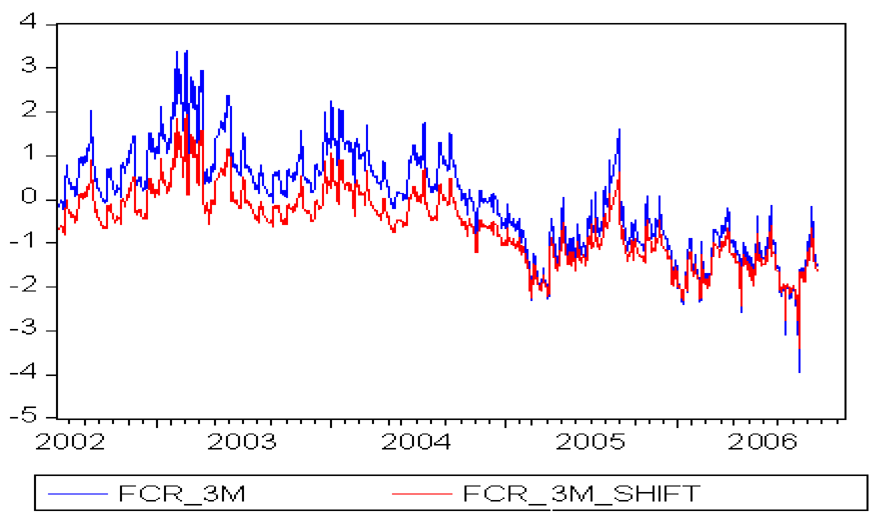
Eigenvectors:

Variable	Vector 1	Vector 2	Vector 3
FCR_3M	-0.574725	-0.769887	0.277426
FCR_6M	-0.580496	0.144590	-0.801323
FCR_9M	-0.576815	0.621585	0.530016

Table 5 – The eigenvalues and eigenvectors of futures returns correlation matrix

The first principal component, which will be further denoted as PC_1, has the highest eigenvalue, which is responsible for explaining 98.76% (λ_1 / k, where k is the matrix dimension, in my case 3) of the variation of the future returns. If we look at corresponding eigenvector weights they are quite similar due to strong correlation between futures returns.

The significance of the first principal component corresponding eigenvector weights is that an upward shift in the first principal component induces a downward parallel shift of the futures returns curve. For this reason first principal component is called the trend component.



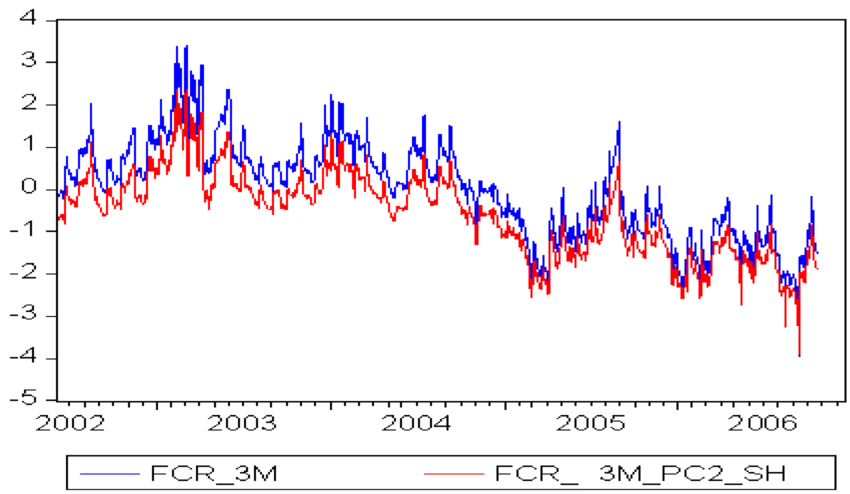
Graph 3 – 3M futures return curve reaction to PC_1 upward shift.

As shown in the theoretical section of my paper starting from the eigenvectors we can get to the original data applying the following formula:

$$X_i = w_{i1}P_1 + w_{i2}P_2 + \ldots + w_{ik}P_k$$

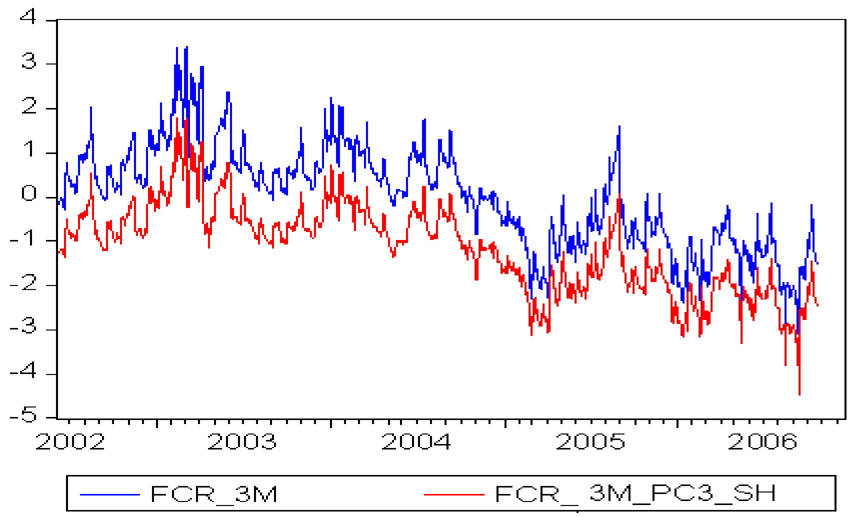
The graph above shows by comparison the 3M futures return curve after inducing an upward shift in the first component. The downward parallel shift is explainable due to negative and similar weights of the eigenvector.

The second principal component, which will be further denoted as PC_2, explains only 1.18% of futures returns variation. The weights are increasing from "-" to "+". Thus an upward movement of the second principal component induces a change in slope of the futures returns, where short maturities move down and long maturities move up. The second principal component significance is that 1.18% of the total futures return variation is attributed to changes in slope.



Graph 4– 3M futures return curve reaction to PC_2 upward shift.

The third principal component, which will be further denoted as PC_3, explains only 0.03% of the futures returns variation. The weights are positive for the short term returns, negative for the medium term returns and positive for the long term returns. Therefore we can say that the third component influences the convexity of the returns curve. The significance of the third principal component is that 0.03% of the total variation is due to changes in convexity.



Graph 4– 3M futures return curve reaction to PC_3 upward shift.

Given the variance explained by each principal component I may choose to drop the third component and use only the first two components in the regression, since they cumulated explain 99.97% of the futures returns variation. I chose not to drop it since I want to see if changes in convexity have significance in explaining volatility in crude-oil markets as well.

As I previously mentioned the main purpose when using principal components analysis was to eliminate the strong correlation among futures returns. Indeed if we check the correlation matrix of principal components we see that correlation indices are close to 0 indicating that we managed to extract patterns from original data which move independently.

	PC1	PC2	PC3
PC1	1.00000000000000	0.00000000000000	-0.00000000000002
PC2	0.00000000000000	1.00000000000000	-0.00000000000011
PC3	-0.00000000000002	-0.00000000000011	1.00000000000000

Table 6 – Principal components correlation matrix

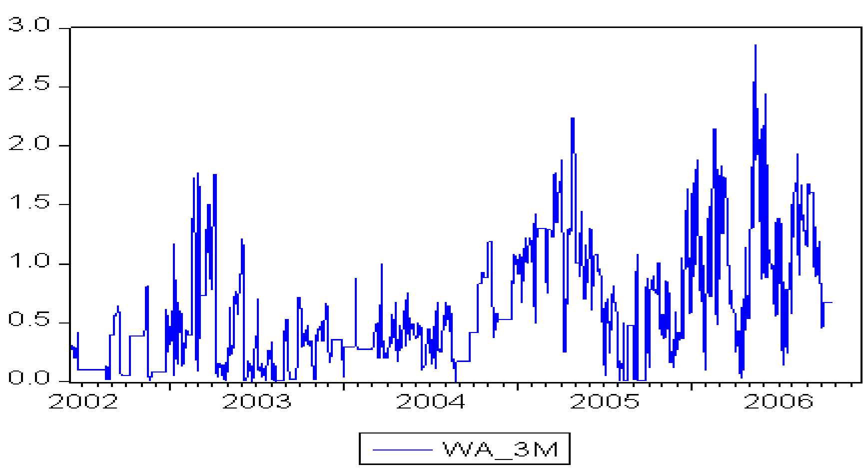
I decided to include squared principal components and principal components cross-product in an attempt to take into account possible non-linearity between volatility proxies and futures returns. Though, this may lead as well to near multicollinearity issue. In Annex part of this paper you may find the correlation matrix of principal components (Table 16). The correlation indices are not high, the highest values are for correlations between squared principal components of the first two components and cross-product between them (0.336674).

Next, I will compute the time series of volatility proxies I chose. The first volatility proxy is straddle returns for the same maturity as futures returns (3 Months, 6 Months and 9 Months). Straddle returns were computed using:

$$r_{straddle,i} = \begin{cases} S_i - K - (\pi_{call} + \pi_{put}), S_i > K + (\pi_{call} + \pi_{put}) \\ K - S_i - (\pi_{call} + \pi_{put}), S_i + (\pi_{call} + \pi_{put}) < K \\ 0, else \end{cases}$$

The period of the sample from which I extracted the straddles was 10/6/2002 – 12/14/2006. When I built the straddles I looked mainly for at-the-money straddles (straddles with strike price near or equal to spot price). The daily frequency of the data was not so high, therefore there were days when just only straddle may be computed from the put and call options available. I decided to take it into account for further evidence imposing though the condition that strike price divided by the underlying asset spot price to be in the interval (0.75; 1.25). Where straddle could not be computed due to lack of data or values outside the (0.75; 1.25) interval I used for the missing daily straddle information the previous available straddle returns value.

The 3Months straddle returns series is represented in the bellow graphic.



Graph 5– 3M straddle returns curve

Further, I will carry out the ADF unit root test to see if we can work with the level series or there is needed at least on difference in order to obtain a stationary series. The output of the ADF test (carried out with both intercept and trend included) is:

Null Hypothesis: WA_3M has a unit root

Exogenous: Constant, Linear Trend

Lag Length: 2 (Automatic based on SIC, MAXLAG=22)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-7.405778	0.0000
Test critical values:	1% level	-3.966139	
	5% level	-3.413769	
	10% level	-3.128955	

*MacKinnon (1996) one-sided p-values.

Table 7 – Calendar spread straddle returns ADF test result

The value of the ADF test is lower than test critical values. Therefore the null hypothesis can be rejected leading to conclusion that calendar spread options straddle returns for the mention period are stationary.

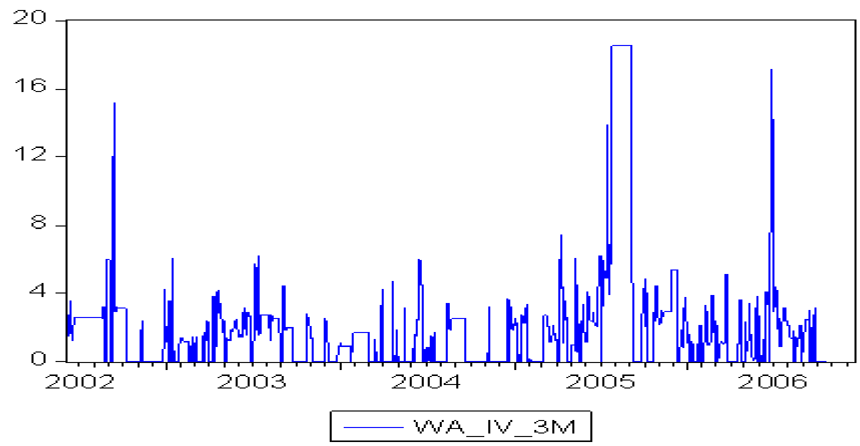
In Annex part of this paper there are shown unit root tests results carried out for straddle returns for 6 Months and 9 Months maturities (Table 17, 18).

The second volatility proxy is the straddle implied volatility. As mentioned, straddle implied volatility is computed as the average of Call and Put options which form the straddle implied volatility.

$$Straddle_{IV} = \frac{Call_{IV} + Put_{IV}}{2}$$

The implied volatility is derived from Black-Scholes formulas for Call and Put options using the options market prices and the risk free rate of the US T-Bills with 3 Months and 6 Months maturities. For the 9 Months maturity the risk free rate was not available, therefore I computed it as the average of 6 Months and 1 Year risk free rate.

The graphical representation of 3M straddle implied volatility.



Graph 6– 3M straddle implied volatility curve

The layout of 3M of straddle implied volatility ADF unit root test shows the series is stationary allowing use it as volatility proxy for our unspanned stochastic volatility research.

Null Hypothesis: WA_IV_3M has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 4 (Automatic based on SIC, MAXLAG=22)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-5.262452	0.0001
Test critical values:	1% level	-3.966153	
	5% level	-3.413776	
	10% level	-3.128960	

*MacKinnon (1996) one-sided p-values.

Table 8 – Calendar spread 3M straddle returns ADF test result

Next, I will regress calendar spread (NYMEX symbol WA) straddle returns on principal components of the futures returns, squared principal components of the returns and cross-products between components. Since R^2 is the square of the correlation coefficient between the values of the dependant variables and corresponding fitted values from the regression model. Using straddle returns as a volatility proxy the R^2 will indicate the extent to which volatility is spanned by trading in the futures contracts, which information is emphasized by the principal components used as regressors. However, there are some issues around the R^2 as goodness of fit measure.

- If we change the order of the regressors the value will change;
- R^2 will never fall if we add extra regressors ;

Therefore, I will rely on Adjusted R^2 as goodness of fit measure since it takes into account the loss of degrees of freedom associated with adding extra variable (squared principal components and cross-products between them).

$$\overline{R^2} = 1 - \left[\frac{T-1}{T-k}(1-R^2) \right]$$

Where, k is the number of degrees of freedom.

The layout of the 3M straddle returns regression on futures returns principal components is presented bellow.

Dependent Variable: WA_3M
Method: Least Squares
Date: 06/30/08 Time: 21:04
Sample (adjusted): 6/10/2002 10/20/2006
Included observations: 1140 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.425578	0.013792	30.85750	0.0000
PC1	0.173028	0.005772	29.97657	0.0000
PC2	0.275921	0.057700	4.782031	0.0000
PC3	2.162054	0.385089	5.614420	0.0000
PC1*PC1	0.057807	0.002972	19.45140	0.0000
PC2*PC2	0.583068	0.196789	2.962914	0.0031
PC3*PC3	-1.675284	2.113742	-0.792568	0.4282
PC1*PC2	0.311658	0.028326	11.00240	0.0000
PC1*PC3	0.643695	0.181391	3.548663	0.0004
PC2*PC3	-1.974297	1.457376	-1.354693	0.1758

R-squared	0.612262	Mean dependent var	0.616018
Adjusted R-squared	0.609173	S.D. dependent var	0.501939

S.E. of regression	0.313793	Akaike info criterion	0.528565
Sum squared resid	111.2664	Schwarz criterion	0.572764
Log likelihood	-291.2818	F-statistic	198.2596
Durbin-Watson stat	0.685380	Prob(F-statistic)	0.000000

Table 9 – Calendar spread 3M straddle returns regression on principal components of futures returns output

We notice that the coefficients for squared third principal component and product between second and third component are not significant. Though, the third principal component (the convexity influence on variance) explained only 0.03% of the total futures returns variation. Lack of significance for the coefficient does not influence the results.

We see that straddle returns have a non-linear dependency on the futures returns trend component. The values of the coefficient for PC_1 and $PC^2{}_1$ are both positive which means the straddle returns dependency on trend component takes the shape of an increasing convex function. Since PC_1 is responsible for explaining 98.76% of the futures return variation we might say that an upward movement in PC_1 will lead to a parallel downward shift of the straddle returns. The slope (PC_2) coefficient is significant as well. An upward movement in (PC_2) will make straddle returns to decrease for short maturities and increase for long maturities. This change has also a degree of convexity ($PC^2{}_2$ coefficient is significant), but since PC_2 is responsible only with 1.18% explanation for the whole futures returns variation the convexity is slight. Also the marginal influence of the components is significant – trend component change on slope component change and trend component change on convexity component change).

The regression both R^2 and $\overline{R^2}$ are low 0.61 and 0.60, which indicates that trading in futures contracts do not span much of crude-oil prices volatility embedded in our volatility proxy – straddle returns. For commodity and financial markets high R^2 and $\overline{R^2}$ should exceed 0.85, whereas R^2 and $\overline{R^2}$ bellow 0.7 indicate that volatility risk cannot be hedged using only futures contracts.

One problem which may appear is the residuals autocorrelation. A key assumption for Ordinary Least Squares method is that residuals have the following property:

$$\text{cov}(u_i, u_j) = 0, \forall i \neq j$$

Since by u_i we denote the residuals of the regression estimation the property assumes that covariance between errors over time are 0. Having autocorrelation among residuals it is not a problem by itself. But since it is a key assumption of OLS if it is violated it means that estimated coefficients may not be significant. In the Annex part of this paper there are presented the test performed to evidence and eliminate residuals autocorrelation (Table 19,20).

If we examine the squared residuals correlogram we see that we have partial autocorrelation among squared residual at lag 1 and 2. We try to model the residuals in order to get rid of the partial autocorrelation by introducing two MA (Moving Average) terms – MA (1) and MA (2). Running the regression with second order MA terms leads to a different regression output. We eliminate residuals autocorrelation (Durbin-Watson test is close to 2). The R^2 and $\overline{R^2}$ increase (0.76 and 0.75) but the most important fact is the significance of the main coefficients remains unchanged - PC_1, PC_1-squared, $PC_1 * PC_2$.

Later on I will use regression residuals to evidence the presence of unspanned stochastic volatility in crude-oil market. Therefore I will retain the residuals form the original regression.

Running the regressions for the 6 Months and 9 Months maturities exhibits same low values for R^2 and $\overline{R^2}$ for 6 Months regression – 0.64 and 0.63 – whereas for 9 Months the results are even lower – 0.24 and 0.23. The results indicate that most volatility risk cannot be hedged by trading in the futures contracts.

Regression Output		Futures Returns (Principal Components)			
		R^2	Adjusted R^2	S.E. From Regression	Sum of the Squared Residuals
Straddle Returns (Volatility Proxy)	3M Straddle Returns	0.612262	0.609173	0.313793	111.2664
	6M Straddle Returns	0.639721	0.636852	0.478931	259.1941
	9M Straddle Returns	0.239385	0.233327	1.041255	1225.16

Table 10 – R^2 and $\overline{R^2}$ from straddle return regressions

We notice that the explanatory power of the futures contracts decreases with maturity. The standard errors from regression as well as the sum of the squared residuals increases with maturity meaning that the gap between $y - \hat{y}$ (actual versus fitted of straddle returns) increases while time-to-maturity increases.

Now we want to investigate how much of the variation in straddle implied volatilities (which is related to expectations under the risk-neutral measure of future volatility) can be explained by variation in the underlying futures prices.

I used the same approach as in straddle returns regressions. The coefficient significance it is quite similar. There is the same partial autocorrelation problem for the squared residuals. Introducing MA terms in the regression equation does not change the significance of the estimated coefficients therefore we may assume that the regression estimation was successful.

In straddle implied volatility case the capacity of futures contracts variation to span crude-oil market volatility is even lower which confirms what the conclusion from straddle returns regression that one cannot hedge much against volatility risk using only trading of futures contracts.

Implied volatility regressions output as well as the procedure for eliminating partial autocorrelation among residuals are shown in the Annex part of this paper (Table 21-25). I will retain the implied volatility residuals as well for further evidence.

Regression Output		Futures Returns (Principal Components)			
		R^2	Adjusted R^2	S.E. From Regression	Sum of the Squared Residuals
Straddle Implied Volatility (Volatility Proxy)	3M Straddle Implied Volatility	0.179701	0.173167	0.447045	491.46
	6M Straddle Implied Volatility	0.064359	0.056907	0.850868	818.0939
	9M Straddle Implied Volatility	0.055037	0.04751	1.424951	2294.449

Table 11 – R^2 and $\overline{R^2}$ from straddle implied volatility regressions

The first remark is that the highest explanatory power is for the shortest maturity (3M) but still very low. Straddle implied volatilities R^2 and $\overline{R^2}$ evidence that there is a very low extent in which futures returns can be used to hedge against volatility. Also the explanatory power of the futures contracts decreases with maturity. The standard errors from regression as well as the sum of the squared residuals increases with maturity meaning that the gap between $y-\hat{y}$ (actual versus fitted of straddle returns) increases while time-to-maturity increases.

There are both advantages and disadvantage for using these approaches to evidence the presence of unspanned stochastic volatility.

Straddle returns:

"+" Straddle returns are not conditioned on a particular pricing model. They are computed based on NYMEX observed call and put premiums, corresponding strike prices. The only assumption in straddle computation is the choice of the shortest time-to-maturity futures contract as a proxy for the crude-oil spot price.

"-" Straddles have high gammas ($\frac{\partial^2 V}{\partial S^2} = \frac{\varphi(d_1)}{S\sigma\sqrt{T}}$, where V is the option premium – Call or Put). Gamma shows how much will vary the value of the option at high changes in crude-oil spot price. It indicates the convexity of the option value. Since straddles are built to hedge against significant changes in crude-oil prices they are subject to high gammas. The assumed significant variant spot price is used in computation of both straddle returns and futures returns. As shown in futures returns principal components analysis and in the significance of estimated coefficients from the straddle returns regression straddle returns are convex in futures returns. Though, even if volatility is completely unspanned by the futures contracts the presence of squared principal components (measuring convexity of the dependencies) may not lead to results close to 0. This may be one of the explanations for higher R^2 and $\overline{R^2}$ in straddle returns regressions than in straddle implied volatility regressions.

Straddle implied volatilities:

"+" If volatility is completely unspanned by futures contracts result will be 0 or closed to 0. If we look at the results this is the case for 9M straddle implied volatility regression.

"-" The results for straddle implied volatilities are conditioned on the accuracy of the pricing model we use. In our case we conditioned on Black-Scholes model.

The third and final step in my evidence is analyzing of the residuals from the regression. I retained the three sets of residuals from each regression type. Next, I will extract the principal components out of each set of regression residuals. If there is unspanned stochastic volatility in the data, there should be large common variation in the residuals. Using the principal components analysis properties this should lead us to a first principal component which embeds most of the variation from the residuals. If the residuals are simply due to noisy data, there should not be common variation in the residuals.

The output of principal components analysis for the two data sets containing regression residuals are presented in the Annex part of this paper (Table 26, 27). Bellow can be found the synthesis of the analysis.

Common Variation Among Residuals	Principal Components Analysis		
	PC1 explanatory power for the variance among residuals (%)	PC2 explanatory power for the variance among residuals (%)	PC3 explanatory power for the variance among residuals (%)
Straddle Returns Residuals	76.69	16.86	6.43
Straddle Implied Volatility Residuals	77.64	16.62	5.72

Table 12 – explanatory power of the first three principal components of the regression residuals

For the straddle return regressions, the first principal component explains 76.69% of the variation in the residuals across maturities, while for the implied volatility regressions, it explains 77.64%. The main property of principal component analysis is that it identifies patterns in data. The strong explanatory power of the first component evidence the presence of large common variation in the residuals, which strongly indicates that low R^2 and $\overline{R^2}$ from the regressions are primarily due to an unspanned stochastic volatility factor rather than noisy data.

One potential weakness of above procedure is that it assumes the estimated coefficients are constant over the 1140 observation length sample. In reality this is not the case, they are time varying. To compensate this I will split the entire sample in four “windows” of 285 observations each. I will repeat the procedure for these rolling windows and see if the new results are consistent with previously illustrated unspanned stochastic volatility evidence. The aggregated results are displayed in the Annex part of this paper (Table 28, 29).

Briefly, the rolling window results display the same low R^2 and , $\overline{R^2}$ meaning that futures variance has low explanatory power on straddle returns– the volatility proxy - even if we split the sample. R^2 and $\overline{R^2}$ are higher for 6 Months maturity than for 3 Months but sensible lower for the 9 Month maturity. The sum of the squared residuals increases with maturity.

Analyzing the principal components of the residuals of the rolling window straddle return regression we notice that first component explanatory power ranges from

49.5% to 92%. This suggest as well the presence of large common variation in the residuals, the signal that low R^2 and $\overline{R^2}$ are due to an unspanned stochastic volatility factor rather than noisy data.

6. Conclusions

In this dissertation paper I presented evidence of the unspanned stochastic volatility in crude-oil market. The results are important since they contradict the general commodity equilibrium models derived mainly from Kaldor's (1939) Theory of Storage, models applied to crude-oil market as well, which suggest that crude-oil market spot prices volatility is determined by the levels of inventories. On the other hand these models suggest the inventories levels are the basis for futures prices formation.

If we rely on these approaches it will mean that trading in futures contracts will be enough to protect against volatility risk. Though, the data obtained from BIS – Bank of International Settlements – states that the number of options on futures derivatives is highly increasing for crude-oil market.

Secondly, there are oil refining companies who still use hedging strategies based on entering a rolling futures contract with different maturities to protect against volatility risk. In my example I simulated one of these strategies with a rolling futures contract with quarterly maturities. The results obtained from the evidence procedure suggest that there is at least one unspanned stochastic volatility factor which cannot be hedged. The low results for R^2 and $\overline{R^2}$ clearly show the low extent to which futures contracts hedge against volatility risk. Therefore, the rolling futures contract strategy is not of much help.

The results are important because they do not rely on a particular pricing model – the case of straddle returns. The implied volatility regression results, though they are model dependant, emphasize the straddle returns regression result.

Further direction in this area will mean to extend the evidence procedure taking into account high frequency data as Andersen and Benzoni (2005) did for fixed income market.

Also a good direction will be to develop an option pricing model to take into account the unspanned stochastic volatility.

There could be interesting to investigate the unspanned stochastic volatility in other commodity markets less liquid where the futures contracts trading covers the most part of the transactions. For example metals commodities markets.

7. Bibliography:

Alexander, C. (2001): "Market Models. A Guide to Financial Analysis", John Wiley & Sons Ltd.

Andersen, T. G. and L. Benzoni (2005): "Can bonds hedge volatility risk in the U.S. treasury market? A specification test for affine term structure models," Working paper, Kellogg School of Management, Northwestern University

Brooks, C. (2002): "Introductory Econometric for Finance", Cambridge University Press

Carlson, M. , Khoker, Z. , Titman, S. (2006): "Equilibrium Exhaustible Resource Price Dynamics", NBER Working Paper 12000.

Cassassus, J. and P. Collin-Dufresne (2005): "Stochastic convenience yield implied from commodity futures and interest rates," Journal of Finance, 60:2283–2331.

Deaton, A. and G. Laroque (1992): "On the behaviour of commodity prices," Review of Economic Studies, 59:1–23.

Deaton, A. and G. Laroque (1996): "Competitive storage and commodity price dynamics," Journal of Political Economy, 104:896–923.

Elekdag, S., Lalonde, R., Laxton, D., Muir, D. and Pesenti, P. (2008): "Oil prices movements and the global economy: A model-based assessment", NBER Working Paper 13792.

Gibson, R. and E. S. Schwartz (1990): “Stochastic convenience yield and the pricing of oil contingent claims,” Journal of Finance, 45:959–976.

Heidari, M. and L. Wu (2003): “Are interest rate derivatives spanned by the term structure of interest rates?,” Journal of Fixed Income, 13:75–86.

Kogan, L., D. Livdan, and A. Yaron (2005): “Futures prices in a production economy with investment constraints,” Working paper, NBER # 11509.

Litzenberger, R. H. and N. Rabinowitz (1995): “Backwardation in oil futures markets: Theory and empirical evidence,” Journal of Finance, 50:1517–1545.

Miltersen, K. (2003): “Commodity price modeling that matches current observables: A new approach,” Quantitative Finance, 3:51–58.

Newey, W. and K. West (1987): “A simple, positive semi-definit, heteroscedasticity and autocorrelation consistent covariance matrix,” Econometrica, 55:703–708.

Nielsen, M. J. and E. S. Schwartz (2004): “Theory of Storage and the Pricing of Commodity Claims,” Review of Derivatives Research, 7:5–24.

Richter, M. and C. Sørensen (2002): “Stochastic volatility and seasonality in commodity futures and options: The case of soybeans,” Working paper, Copenhagen Business School.

Routledge, B. R., D. J. Seppi, and C. S. Spatt (2000): “Equilibrium forward curves for commodities,” Journal of Finance, 55:1297–1338.

Schwartz, E. S. (1997): “The stochastic behavior of commodity prices: Implications for valuation and hedging,” Journal of Finance, 52:923–973.

Trolle, A. and E. Schwartz (2006): "A general stochastic volatility model for the pricing and forecasting of interest rate derivatives," Working paper, UCLA and NBER # 12337.

8. Annex

Table 13 - ADF test for crude-oil spot futures prices

Null Hypothesis: FUTURES_SPOT has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic based on SIC, MAXLAG=22)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-1.337612	0.6138
Test critical values:	1% level	-3.435876	
	5% level	-2.863868	
	10% level	-2.568060	

*MacKinnon (1996) one-sided p-values.

Table 14 – ADF test for crude-oil 6 Months futures returns

Null Hypothesis: FCR_6M has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 0 (Automatic based on SIC, MAXLAG=22)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-5.408393	0.0000
Test critical values:	1% level	-3.966124	
	5% level	-3.413762	
	10% level	-3.128951	

*MacKinnon (1996) one-sided p-values.

Table 15 – ADF test for crude-oil 9 Months futures returns

Null Hypothesis: FCR_9M has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 0 (Automatic based on SIC, MAXLAG=22)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-5.131246	0.0001
Test critical values:	1% level	-3.966124	
	5% level	-3.413762	
	10% level	-3.128951	

*MacKinnon (1996) one-sided p-values.

Table 16– Principal components, Squared principal components and Cross-products between principal components correlation matrix.

	PC1	PC2	PC3	PC1^2	PC2^2	PC3^2	PC12	PC13	PC23
PC1	1.000000								
PC2	1.96E-15	1.000000							
PC3	0.00	0.00	1.000000						
PC1^2	0.076823	-0.15	-0.14	1.000000					
PC2^2	-0.22	-0.37	-0.01	0.207121	1.000000				
PC3^2	-0.05	0.100960	0.480248	0.015384	0.096237	1.000000			
PC12	-0.12	-0.36	0.125049	0.336674	0.314062	-0.02	1.000000		
PC13	-0.17	0.083183	-0.37	0.080146	0.016955	-0.14	-0.12	1.000000	
PC23	0.081019	0.002317	0.429687	-0.08	-0.25	0.627921	0.016908	-0.14	1.000000

Table 17 – ADF test for calendar spread 6 Months straddle returns

Null Hypothesis: WA_6M has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 0 (Automatic based on SIC, MAXLAG=22)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-5.758844	0.0000

Test critical values:	1% level	-3.966124	
	5% level	-3.413762	
	10% level	-3.128951	

*MacKinnon (1996) one-sided p-values.

Table 18 – ADF test for calendar spread 9 Months straddle returns

Null Hypothesis: WA_RET_9M has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic based on SIC, MAXLAG=15)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-3.715733	0.0044
Test critical values:	1% level	-3.454626	
	5% level	-2.872121	
	10% level	-2.572482	

*MacKinnon (1996) one-sided p-values.

Table 19 – Partial autocorrelation among residuals of 3M straddle returns regression on principal components of the futures returns

Date: 07/07/08 Time: 13:41
Sample: 6/10/2002 10/20/2006
Included observations: 1140

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.452	0.452	233.08	0.000
		2	0.366	0.204	386.40	0.000
		3	0.295	0.094	486.39	0.000
		4	0.226	0.027	544.93	0.000
		5	0.215	0.062	598.19	0.000
		6	0.228	0.089	657.78	0.000
		7	0.134	-0.055	678.33	0.000
		8	0.145	0.032	702.66	0.000
		9	0.127	0.017	721.10	0.000
		10	0.083	-0.024	729.13	0.000

Table 20 – Calendar spread 3M straddle returns regression on principal components of futures returns output including which models the residuals in order to eliminate partial autocorrelation

Dependent Variable: WA_3M
Method: Least Squares
Date: 07/07/08 Time: 13:48
Sample (adjusted): 6/10/2002 10/20/2006
Included observations: 1140 after adjustments
Convergence achieved after 23 iterations
White Heteroskedasticity-Consistent Standard Errors & Covariance
Backcast: 6/06/2002 6/07/2002

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.454981	0.022262	20.43761	0.0000
PC1	0.161534	0.011063	14.60119	0.0000
PC2	0.119649	0.108860	1.099107	0.2720
PC3	1.606924	0.519199	3.095005	0.0020
PC1*PC1	0.051177	0.006797	7.529449	0.0000
PC2*PC2	0.298509	0.257420	1.159618	0.2464
PC3*PC3	-1.498973	2.788953	-0.537468	0.5911
PC1*PC2	0.243046	0.061034	3.982122	0.0001
PC1*PC3	0.488360	0.373483	1.307581	0.1913
PC2*PC3	-0.440279	2.290440	-0.192225	0.8476
MA(1)	0.576291	0.048558	11.86812	0.0000
MA(2)	0.299662	0.044152	6.787111	0.0000

R-squared	0.759935	Mean dependent var	0.616018
Adjusted R-squared	0.757594	S.D. dependent var	0.501939
S.E. of regression	0.247128	Akaike info criterion	0.052653
Sum squared resid	68.88970	Schwarz criterion	0.105693
Log likelihood	-18.01213	F-statistic	324.6113
Durbin-Watson stat	1.831519	Prob(F-statistic)	0.000000

Inverted MA Roots	-.29+.47i	-.29-.47i

Table 21 – Calendar spread 6M straddle returns regression on principal components of futures returns output.

Dependent Variable: WA_6M
Method: Least Squares
Date: 06/30/08 Time: 21:04
Sample (adjusted): 6/10/2002 10/20/2006
Included observations: 1140 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.891877	0.021050	42.36981	0.0000
PC1	0.182752	0.008810	20.74417	0.0000
PC2	-0.204198	0.088065	-2.318715	0.0206
PC3	-0.470916	0.587749	-0.801218	0.4232
PC1*PC1	0.149071	0.004536	32.86512	0.0000
PC2*PC2	1.707215	0.300352	5.684044	0.0000
PC3*PC3	-6.490724	3.226136	-2.011919	0.0445
PC1*PC2	-0.201998	0.043234	-4.672244	0.0000
PC1*PC3	0.750564	0.276851	2.711075	0.0068
PC2*PC3	2.792259	2.224345	1.255317	0.2096

R-squared	0.639721	Mean dependent var	1.388184
Adjusted R-squared	0.636852	S.D. dependent var	0.794751
S.E. of regression	0.478931	Akaike info criterion	1.374214
Sum squared resid	259.1941	Schwarz criterion	1.418414
Log likelihood	-773.3022	F-statistic	222.9399
Durbin-Watson stat	0.233306	Prob(F-statistic)	0.000000

Table 22 – Calendar spread 6M straddle returns regression on principal components of futures returns output.

Dependent Variable: WA_9M
Method: Least Squares
Date: 06/29/08 Time: 12:58

Sample (adjusted): 6/10/2002 10/20/2006

Included observations: 1140 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.180095	0.045765	47.63684	0.0000
PC1	-0.157929	0.019154	-8.245442	0.0000
PC2	0.067175	0.191464	0.350847	0.7258
PC3	-3.488171	1.277838	-2.729743	0.0064
PC1*PC1	0.148414	0.009861	15.04992	0.0000
PC2*PC2	0.444897	0.653002	0.681310	0.4958
PC3*PC3	11.68019	7.014011	1.665266	0.0961
PC1*PC3	0.051355	0.601908	0.085320	0.9320
PC1*PC2	-0.089425	0.093995	-0.951386	0.3416
PC2*PC3	2.515891	4.835997	0.520243	0.6030

R-squared	0.239385	Mean dependent var	2.646579
Adjusted R-squared	0.233327	S.D. dependent var	1.189192
S.E. of regression	1.041255	Akaike info criterion	2.927464
Sum squared resid	1225.160	Schwarz criterion	2.971664
Log likelihood	-1658.654	F-statistic	39.51549
Durbin-Watson stat	0.060293	Prob(F-statistic)	0.000000

Table 23 – Calendar spread 3M straddle implied volatilities regression on principal components of futures returns output.

Dependent Variable: WA_IV_3M

Method: Least Squares

Date: 06/30/08 Time: 20:42

Sample (adjusted): 6/10/2002 10/20/2006

Included observations: 1140 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.134059	0.133923	15.93501	0.0000

PC1	0.179456	0.056049	3.201744	0.0014
PC2	-4.806310	0.560285	-8.578327	0.0000
PC3	-14.46602	3.739363	-3.868579	0.0001
PC1*PC1	-0.192425	0.028858	-6.668055	0.0000
PC2*PC2	12.69585	1.910893	6.643937	0.0000
PC3*PC3	-27.16623	20.52523	-1.323553	0.1859
PC1*PC3	-2.207645	1.761375	-1.253365	0.2103
PC1*PC2	-0.942972	0.275059	-3.428250	0.0006
PC2*PC3	44.47415	14.15167	3.142678	0.0017

R-squared	0.179701	Mean dependent var	1.988859
Adjusted R-squared	0.173167	S.D. dependent var	3.350966
S.E. of regression	3.047045	Akaike info criterion	5.074955
Sum squared resid	10491.46	Schwarz criterion	5.119154
Log likelihood	-2882.724	F-statistic	27.50511
Durbin-Watson stat	0.311573	Prob(F-statistic)	0.000000

Table 24 – Partial autocorrelation among residuals of 3M straddle implied volatility regression on principal components of the futures returns

Date: 07/07/08 Time: 14:43
Sample: 6/10/2002 10/20/2006
Included observations: 1140

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.852	0.852	829.58	0.000
		2	0.741	0.054	1456.9	0.000
		3	0.669	0.098	1970.1	0.000
		4	0.622	0.076	2412.8	0.000
		5	0.600	0.113	2826.0	0.000
		6	0.567	0.001	3195.1	0.000
		7	0.551	0.085	3543.6	0.000
		8	0.510	-0.060	3842.5	0.000
		9	0.486	0.060	4114.5	0.000
		10	0.466	0.007	4364.2	0.000

Table 25 – Calendar spread 3M straddle implied volatility regression on principal components of futures returns output including which models the residuals in order to eliminate partial autocorrelation

Dependent Variable: WA_IV_3M
Method: Least Squares
Date: 07/07/08 Time: 14:51
Sample (adjusted): 6/10/2002 10/20/2006
Included observations: 1140 after adjustments
Convergence achieved after 19 iterations
Backcast: 6/06/2002 6/07/2002

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.259644	0.173734	13.00635	0.0000
PC1	0.218204	0.074695	2.921279	0.0036
PC2	-3.371287	0.595331	-5.662876	0.0000
PC3	-13.22784	3.732385	-3.544072	0.0004
PC1*PC1	-0.147461	0.035052	-4.206874	0.0000
PC2*PC2	4.726614	1.762007	2.682517	0.0074
PC3*PC3	-2.313252	17.28295	-0.133846	0.8935
PC1*PC3	-2.026224	1.718391	-1.179140	0.2386
PC1*PC2	0.025651	0.309410	0.082902	0.9339
PC2*PC3	20.02444	11.98333	1.671025	0.0950
MA(1)	0.841608	0.026104	32.24109	0.0000
MA(2)	0.490377	0.026290	18.65293	0.0000

R-squared	0.692256	Mean dependent var	1.988859
Adjusted R-squared	0.689255	S.D. dependent var	3.350966
S.E. of regression	1.867979	Akaike info criterion	4.098062
Sum squared resid	3935.981	Schwarz criterion	4.151101
Log likelihood	-2323.895	F-statistic	230.6716
Durbin-Watson stat	1.616334	Prob(F-statistic)	0.000000

Inverted MA Roots	-.42-.56i	-.42+.56i

Table 26 – Principal components analysis for the straddle returns regression residuals

Date: 07/07/08 Time: 15:50
Sample (adjusted): 6/10/2002 10/20/2006
Included observations: 1140 after adjustments

Covariance of R_WA_3M R_WA_6M R_WA_9M

	Comp 1	Comp 2	Comp 3
Eigenvalue	1.131444	0.248877	0.094989
Variance Prop.	0.766920	0.168695	0.064386
Cumulative Prop.	0.766920	0.935614	1.000000

Eigenvectors:

Variable	Vector 1	Vector 2	Vector 3
R_WA_3M	0.004976	0.129672	0.991544
R_WA_6M	0.253241	0.959071	-0.126696
R_WA_9M	0.967390	-0.251730	0.028066

Table 27 – Principal components analysis for the straddle implied volatility regression residuals

Date: 07/07/08 Time: 15:52
Sample (adjusted): 6/10/2002 10/20/2006
Included observations: 1140 after adjustments
Covariance of R_WA_IV_3M R_WA_IV_6M R_WA_IV_9M

	Comp 1	Comp 2	Comp 3
Eigenvalue	9.266107	1.983520	0.683712
Variance Prop.	0.776489	0.166217	0.057294
Cumulative Prop.	0.776489	0.942706	1.000000

Eigenvectors:

Variable	Vector 1	Vector 2	Vector 3
R_WA_IV_3M	-0.995729	0.088116	-0.027571
R_WA_IV_6M	0.038706	0.127277	-0.991112
R_WA_IV_9M	0.083823	0.987945	0.130144

Table 28 – R^2 and $\overline{R^2}$ for straddle returns regression – Rolling Windows

Regression Output		Futures Returns Principal Components			
		R^2	Adjusted R^2	S.E. From Regression	Sum of the Squared Residuals
First Window	3M Straddle Returns	0.595727	0.582496	0.250818	17.3002
	6M Straddle Returns	0.84287	0.837728	0.364208	36.47813
	9M Straddle Returns	0.499059	0.482665	1.213151	404.7273
Second Window	3M Straddle Returns	0.174372	0.147351	0.17336	8.264747
	6M Straddle Returns	0.264857	0.234253	0.482294	63.96694
	9M Straddle Returns	0.105283	0.076002	1.250616	430.111
Third Window	3M Straddle Returns	0.535728	0.520534	0.333389	30.56583
	6M Straddle Returns	0.680356	0.669895	0.309532	26.34772
	9M Straddle Returns	0.0899	0.053443	0.101605	0.102334
Fourth Window	3M Straddle Returns	0.474501	0.456853	0.395743	41.9721
	6M Straddle Returns	0.646467	0.634594	0.419518	47.1668
	9M Straddle Returns	0.448315	0.431207	0.434085	50.49913

Table 29 – Explanatory power of straddle returns residuals principal components

Common Variation Among Residuals	Principal Components Analysis		
	PC1 explanatory power for the variance among	PC2 explanatory power for the variance among residuals (%)	PC3 explanatory power for the variance among residuals (%)

	residuals (%)		
Straddle Returns Residuals First Window	92.07	4.38	3.54
Straddle Returns Residuals Second Window	85.84	9.92	4.23
Straddle Returns Residuals Third Window	5.35	45.21	1.18
Straddle Returns Residuals Fourth Window	49.51	31.62	18.86