Unbiasedness and market efficiency of crude oil futures markets: A revisit

Yanan He\textsuperscript{a}, Yongmiao Hong\textsuperscript{b,a}

\textsuperscript{a}The Wang Yanan Institute for Studies in Economics, Xiamen University, Xiamen 361005, China; Email: heyanan@amss.ac.cn;
\textsuperscript{b}Department of Economics and Statistical Science, Cornell University, Ithaca, NY 14850; Email: yhong.cornell@gmail.com

Abstract

Extensive researchers have been interested in testing for the joint hypothesis of unbiasedness and market efficiency of crude oil futures markets. Most of them find evidences supporting the hypothesis. However, issues of econometric tests and data constructing methods make previous conclusions unconvincing. Addressing these issues, this paper presents a rigorous test for the joint hypothesis, and a comprehensive analysis by combining both linearity and nonlinearity. Unlike most previous studies, significant serial dependence in the conditional mean of deviations between crude oil futures price and future spot price is found as evidence against the joint hypothesis. This can be well explained by existence of time-varying risk premium which implies costs for hedging in crude oil futures markets. Both linear and nonlinear serial dependence are discovered in the conditional mean of deviations. Therefore, information of past deviations should be utilized to improve the spot price forecasting inherent in futures price. We also find that overlapping data sets tend to incorrectly reject the joint hypothesis.

Keywords: Crude oil futures market, unbiasedness and market efficiency,
nonlinear, overlapping and nonoverlapping data sets, generalized spectral test

JEL: G14, G17

1. Introduction

Extreme price volatility has fostered the growth of crude oil futures markets especially since 1980s. Futures markets trade oil volumes for future delivery that far overshadow spot markets. The New York Mercantile Exchange (NYMEX) and the International Petroleum Exchange (IPE) together traded an equivalent of 600 million-615 million bbl daily during mid-2005 trading sessions, which is over seven times the volumes of physical oil produced around the world daily.

A joint hypothesis of unbiasedness and market efficiency is an important key to get understanding of the principal functions of crude oil futures markets, namely, risk transferring and price discovery. The hypothesis involves costs and benefits of hedging, diversification benefits that result from including futures in investment portfolios, and also to what extent that economic agents can make decisions by using futures price as an indicator of future spot prices. Unsurprisingly, econometric tests for the joint hypothesis have attracted great attention of economists, such as Samuelson (1965), Serletis and Banack (1990), Green and Mork (1991), Deaves and Krinsky (1991), Deaves and Krinsky (1992), Quan (1992), Crowder (1993), Moosa and Al-Loughani (1994), Schwarz and Szakmary (1994), Peroni and Mcnown (1998), Kellard et al. (1999), Gülen and Gürcan (1998), Abosedra and Baghesani (2004), Chernenko et al. (2004), Chinn et al. (2005), Switzer and EI-Khoury (2007),
Alquist and Kilian (2010), Lean et. al (2010) among others. There are three forms of market efficiency: weak or simple, semi-strong and strong efficiency, of which weak efficiency has been largely studied in the literature. Most empirical evidences consistently support unbiasedness and weak efficiency, except those in Serletis and Banack (1990), Kellard et al. (1999), Green and Mork (1991), and Alquist and Kilian (2010).

With respect to these evidences, two puzzles emerge. First, unbiasedness and market efficiency imply no time-varying risk premium in crude oil futures market, which clearly contradicts another strand of studies. These studies presuppose an existence of time-varying risk premium and endeavor to investigate, both theoretically and empirically, determinants of the risk premium. See, Hirshleifer (1989), de Roon et al. (2000), Dincerler et al. (2003) and Acharya et al. (2010), for example. Second, although some researchers support the hypothesis of unbiasedness and market efficiency, they also find failure of crude oil futures price as an optimal predictor of future spot price at the same time. Morana (2001) shows that one-month futures price fails to predict the sign of the oil price changes in more than 50% of the cases, pointing out that it can not beat a random guess. Chinn et al. (2005) find that, for 3-month horizon, the basis accounts for only 5% of total variation in changes in spot rates. Alquist and Kilian (2010) also find that oil futures prices fail to improve the accuracy of simple no-change forecasts.

A review on previous tests shows that incorrect data constructing and econometric tests make their results unconvincing. Most existing papers construct aggregation data sets by taking sampling intervals smaller than the forecast horizon. However, pooling data carries the danger of biasing
test results towards finding inefficiency markets. This problem is referred to as overlapping data problem, which is omitted by most studies in oil futures market (e.g., Deaves and Krinsky, 1992; Alquist and Kilian, 2010). With regard to those conventional econometric models that have been widely used, models like cointegration and stationary equations require estimation of additional parameters, and tend to suffer a potential problem of endogeneity, which will undermine the usefulness of some estimating method. Meanwhile, the corresponding tests implicitly postulate that the trader’s loss function coincides with econometrician’s quadratic loss function. If that is not the case, forecast efficiency tests tend to be biased in favor of the alternative hypothesis (Elliott et al., 2005; Alquist and Kilian, 2010).

Most importantly, theories in unbiasedness and efficiency hypothesis have implications on conditional mean dynamics of deviations ($\varepsilon_T = S_T - F_{t,T}$) given available information to economic agents. If there is serial dependence, the hypothesis of unbiasedness and market efficiency should be rejected. However, a large number of preceding papers neglect to test whether the sequence of deviations has serial dependence in the conditional mean (e.g., Serletis and Banack, 1990; Gülen and Gürcan, 1998; Chinn et al., 2005; Alquist and Kilian, 2010). Insignificance of parameters cannot be taken as evidences supporting unbiasedness and efficiency. Despite a few papers try to test the serial dependence, they rely solely on the existence of linear dependence in the conditional mean of deviations as an evidence against efficiency. See Deaves and Krinsky (1991), Green and Mork (1991), Peroni and Mohnoven (1998), and Chernenko et al. (2004), for example. From a nonlinear time series perspective, it is important to distinguish serial uncorrelation and
serial independence in the conditional mean. The latter implies the former, but not vice versa. A nonlinear serial dependence in the conditional mean can have zero autocorrelation but a nonzero mean conditional on its past history. An example is a nonlinear moving average process $Y_t = b e_{t-1} e_{t-2} + e_t$, where $\{e_t\}$ is IID. It is a white noise, but with nonlinear serial dependence in its conditional mean. In such situations, misleading conclusions could be reached in favor of the hypothesis when test statistics based on insignificant autocorrelation, variance ratio and power spectrum. Thus, evidence in support of or against unbiasedness and efficiency should rely on both linear and nonlinear serial dependence in the conditional mean of deviations.

By addressing these issues in preceding studies, this paper tests unbiasedness and weak efficiency rigorously with monthly data of 1- and 3-month crude oil futures contracts. It extends existing research in several important ways. It retests the joint hypothesis of unbiasedness and weak efficiency using relatively recent data, and points out the importance of data constructing. Most importantly, this study provides a more comprehensive analysis of efficiency by combining linear and nonlinear tests. We explore serial dependence in the conditional mean of deviations, by using Hong and Lee’s (2005) generalized spectral statistics for conditional mean models in time series with conditional heteroscedasticity of unknown form based on multiple points. In practice, economic theory does not point to a single nonlinear model to describe conditional dynamics of deviations. Tests against a specific alternative may have low or little power against other alternatives. Hence, we use the general spectral statistics to detect any type of pairwise serial dependence in the conditional mean with no prior knowledge of possible alternatives alter-
native. The general spectral statistics can check a large number of lags since the dimension of the conditioning information set may be infinite. Meanwhile, they are robust to conditional heteroscedasticity and higher order time-varying moments of unknown form. As is well known, nonlinearities in mean and in higher order moments have distinct economic implications. Failure to accommodate conditional heteroscedasticity will possibly give a misleading conclusion.

Finally, this paper presents empirical evidences against the joint hypothesis of unbiasedness and market efficiency of crude oil futures markets. Both linear and nonlinear serial dependence are found in the conditional mean of deviations between futures price and future spot price. The rejection contradicts most earlier related tests in the literature, but coincides with findings in another strand of studies on determinants of time-varying risk premium in oil futures market. Significant serial dependence in the conditional mean of deviations can be explained by the existence of time-varying risk premium. Several determinants of the risk premium, such as hedging pressure effects, inventory levels and hedging demand caused by default risks confronted by producers’ managers, are often time varying and highly persistent. Past information of these determinants is incorporated in past values of deviations, resulting in serial dependence in the conditional mean. These results can explain the failure of using current crude oil futures price as a predictor of future spot price. We should consider nonlinear econometric models when futures-based forecasting models are used. Besides, overlapping data observations tend to exaggerate information and incorrectly reject the joint hypothesis of unbiasedness and market efficiency. These findings have signif-
icant implications for the views and practices of market professionals.

The remained paper is organized as follows. Section 2 presents the theory and econometric strategies of unbiasedness and market efficiency of futures markets. In Section 3, we introduce the data used in this paper. Section 4 provides empirical results, and Section 5 concludes.

2. The joint hypothesis of unbiasedness and market efficiency

A financial market can be considered as efficient if prices fully reflect all available information and expected returns are zero. As pointed out by Fama (1991), market efficiency per se is not testable and it must be tested jointly with some pricing assets model. The model that futures prices are unbiased predictors of future spot prices is an appropriate framework. A joint hypothesis of unbiasedness and market efficiency will imply that the market price fully reflects available information and futures price is the best predictor of future spot price. There will be no risk premium.

2.1. Theory

Futures prices will be unbiased forecasts of future spot prices if futures prices equal expected futures spot prices. We have

\[ E_t S_T = F_{t,T}, \]

(1)

where \( F_{t,T} \) is the futures price of a contract that matures at time \( T \) and \( E_t S_T \) is the expected spot price at time \( T \) given all available information at time \( t \). The expected spot price is unobservable in practice. If we assume that expectations are rational, so that:

\[ S_T = E_t(S_T|I_t) + \varepsilon_T, \]

(2)
where \( I_t \) is the information available at time \( t \), and \( \varepsilon_T \) satisfies \( E(\varepsilon_T) = 0 \) and \( E_t(\varepsilon_T|I_t) = 0 \). The information set can be classified into three categories: (1) the information contained in a historical sequence of interested; (2) all publicly available information relevant to futures markets; (3) all information that is known to any investor, including privately held information. In this paper, we only test the weak efficiency which has been largely investigated in the literature. Available information set only includes past values of deviations, which virtually contain past information in crude oil markets.

Consequently, the joint hypothesis of unbiasedness and market efficiency implies a simple model:

\[
S_T - F_{t,T} = \varepsilon_T, \tag{3}
\]

where \( \varepsilon_T \) is a deviation between futures price and future spot price. It means that, if economic agents are risk neutral, costs of transaction are zero, information is used rationally, and the market is competitive, crude oil market will be efficient in the sense that expected returns will be zero. There is no risk premium and any cost for hedgers. Crude oil futures price will be the best predictor of spot price.

Obviously, the joint hypothesis implies restrictions on unconditional and conditional mean dynamics of deviations given available information. That is, \( E(\varepsilon_T) = 0 \) and \( E_t(\varepsilon_T|I_t) = 0 \). Unconditional mean is zero, and there is no serial dependence in conditional mean. When forecast horizon \( h = T - t \) is 1, series of \( \varepsilon_T \) must be an martingale difference sequence (MDS). If \( E_t(\varepsilon_T) \neq 0 \) or \( E_t(\varepsilon_T|I_t) \neq 0 \), the joint hypothesis is rejected. Investors in the oil market, capitalizing on available information, can take appropriate futures market positions in order to earn returns, either constant or predictable time-varying
risk premium, or excess returns from normal (i.e., risk-adjusted). Relevant information could be utilized to improve the spot price forecasts inherent in futures prices.

2.2. Econometric Strategies

Unbiasedness and market efficiency can be tested through formal econometric analysis. Equation (3) has been often extended as

\[ S_T = \alpha + \beta F_{t, T} + \varepsilon_T. \]  

in previous studies. The hypothesis requires that \( \alpha = 0 \) and \( \beta = 1 \) as well as restrictions on deviations \( \varepsilon_T \). Since futures and spot prices are nonstationary, equation (4) is usually estimated with cointegration methods. See, Quan (1992), Crowder (1993), Schwarz and Szakmary (1994), Moosa and Al-Loughani (1994), Peroni and Mcnown (1998), Gülen and Gürcan (1998), and Switzer and El-Khoury (2007), for example. Cointegration between futures and spot prices and significance of \( \alpha \) and \( \beta \) have been widely supported. However, estimation technique of cointegration is a nontrivial issue. Possible presence of a risk premium results in a correlation between the regressor and the error term through lagged futures prices (Keynes, 1930). The regressor will be endogenous, and it would cast doubt on the usefulness of some estimation methods. Using inappropriate methods will yield misleading results. Different estimation techniques may also lead to diverse conclusions (Peroni and Mcnown, 1998; Moore and Copeland, 1995).

In order to handle nonsationary prices in equation (4), one alternative approach is to subtract \( S_t \) from both sides of (4)

\[ S_T - S_t = \beta_0 + \beta_1(F_{t, T} - S_t) + \varepsilon_T. \]  

9
A large number of papers use this equation for the test, such as Serletis and Banack (1990), Chernenko et al. (2004), Abosedra and Baghesani (2004), Chinn et al. (2005), Switzer and EI-Khoury (2007) and Alquist and Kilian (2010). The joint null hypothesis of $\beta_0 = 0$ and $\beta_1 = 1$ cannot be rejected at any reasonable level of significance by most papers with one exception in Alquist and Kilian (2010). Although this equation (5) is easy to estimate, as demonstrated by Liu and Maddala (1992), it also suffers from endogenous bias, rendering tests based on (5) noninformative. Peroni and Mcnown (1998) find that explanatory power of this kind of regressions is abysmally low.

Some empirical studies construct linear autoregressive models to directly test serial autocorrelation in deviations, such as Green and Mork (1991) and Deaves and Krinsky (1991); Deaves and Krinsky (1992). Serial uncorrelation is often taken as evidence supporting unbiasedness and efficiency. This does not require estimation of additional parameters in equations (4) and (5). It can get round potential endogenous issues. However, serial uncorrelation and serial independence are distinguishing definitions. The latter implies the former but not vice versa. A nonlinear serial dependence in the conditional mean can have zero autocorrelation but a nonzero mean conditional on its past values. Hence, misleading conclusions could be reached in favor of the joint hypothesis when test statistics is based on autocorrelation, variance ration and power spectrum are insignificant. Evidences of nonlinearities in daily and intraday oil futures returns have been empirically found by Matilla-Garcia (2007), Moshiri and Foroutan (2006), Shambora and Rossiter (2007), and Wang and Yang (2010), mainly from a perspective of predictability. We should consider to test both linear and nonlinear serial dependence in
the conditional mean of deviations, especially when serial uncorrelation is found. Even if deviations are found to be serial correlated and the joint hypothesis can be rejected, testing for nonlinear serial dependence is still necessary because we can get comprehensive understanding of predictability of crude oil futures prices.

Therefore, we use Hong and Lee’s (2005) generalized spectral statistics to test the null hypothesis $H_0$ of $E(\varepsilon_t | I_{t-1}) = 0$. As introduced in Introduction, the generalized spectral statistics have multiple advantages. The generalized spectrum can capture any type of pairwise serial dependence in the conditional mean over various lags, including those that could be missed by the autocorrelation, variance ration and power spectrum. It does not need any prior knowledge of possible alternatives. Meanwhile, the generalized spectral statistics are robust to conditional heteroscedasticity and higher order time-varying moments of unknown form.

To test the null hypothesis $H_0$ of $E(\varepsilon_t | I_{t-1}) = 0$, the basic idea of the generalized spectrum is to consider the spectrum of a transformed series $\{e^{iux_t}\}$. It is defined as

$$f(\omega, u, v) \equiv \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \sigma_j(u, v)e^{-ij\omega},$$

where $\omega$ is the frequency, and $\sigma_j(u, v)$ is the covariance function of the transformed series:

$$\sigma_j(u, v) \equiv \text{cov}(e^{iu\varepsilon_t}, e^{iu\varepsilon_{t-|j|}}), \quad j = 0, \pm 1, \cdots$$

The function $f(\omega, u, v)$ can capture any type of pairwise serial dependence in $\{\varepsilon_t\}$. The generalized spectrum itself is not appropriate for testing $H_0$, since
it captures the serial dependence in mean and in higher order moments. Just
as the characteristic function can be differentiated to generate various mo-
mments, generalized spectral derivatives can capture various specific aspects
of serial dependence, providing information on possible types of serial depen-
dence. To only capture the serial dependence in the conditional mean, one
can use the derivative

\[ f^{(0,1,0)}(\omega, 0, v) \equiv \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \sigma_j^{(1,0)}(0, v)e^{-ij\omega}, \quad \omega \in [-\pi, \pi], \] (8)

where

\[ \sigma_j^{(1,0)}(0, v) \equiv \sigma_j(u, v)|_{u=0} = \text{cov}(i\varepsilon_t, e^{i\varepsilon_{t-|j|}}). \] (9)

The measure \( \sigma_j^{(1,0)}(0, v) \) checks whether the autoregressive function \( E(\varepsilon_t|\varepsilon_{t-j}) \)
at lag \( j \) is 0.\(^1\)

We can estimate \( f^{(0,1,0)}(\omega, 0, v) \) by a smoothed kernel estimator

\[ f^{(0,1,0)}(\omega, 0, v) \equiv \frac{1}{2\pi} \sum_{j=1}^{T-1} (1 - |j|/T)^{\frac{1}{2}}k(j/p)\hat{\sigma}_j^{(1,0)}(0, v)e^{-ij\omega}, \quad \omega \in [-\pi, \pi], \] (10)

where \( \hat{\sigma}_j^{(1,0)}(0, v) = \frac{\partial}{\partial u} \hat{\sigma}_j(u, v)|_{u=0}, \sigma_j(u, v) = \hat{\varphi}_j(u, v) - \hat{\varphi}_j(u, 0)\hat{\varphi}_j(0, v), \) and

\[ \hat{\varphi}_j(u, v) = \frac{1}{T - |j|} \sum_{t=|j|+1}^{T} e^{iud_t+i\varepsilon_{t-|j|}}. \] (11)

Here, \( p \) is a bandwidth, and \( k : \mathbb{R} \to [-1, 1] \) is a symmetric kernel. The
test statistic proposed in Hong and Lee (2005) that is robust to conditional
heteroscedasticity and other time-varying higher order conditional moments

\(^1\)The hypothesis of \( E(\varepsilon_t|I_{t-1}) = 0 \) is not the same as the hypothesis of \( E(\varepsilon_t|\varepsilon_{t-j}) = 0 \)
for all \( j > 0 \). The former implies the latter but not vice versa.
of unknown form is given as follows:

\[
\hat{M}_1(p) \equiv \left[ \sum_{j=1}^{T-1} k^2(j/p)(T - j) \int |\hat{\sigma}_j^{(1,0)}(0, v)|^2 dW(v) - \hat{C}_1(p) \right] / \sqrt{\hat{D}_1(p)},
\]  

where \( W : \mathbb{R} \to \mathbb{R}^+ \) is a nondecreasing function that weighs sets symmetric about zero equally,

\[
\hat{C}_1(p) = \sum_{j=1}^{T-1} k^2(j/p) \frac{1}{T-j} \sum_{t=j+1}^{T-1} \hat{\varepsilon}_t^2 \int |\hat{\Psi}_{t-j}(v)|^2 dW(v),
\]

\[
\hat{D}_1(p) = 2 \sum_{j=1}^{T-2} k^2(j/p) k^2(l/p) \int \int \frac{1}{T - \max(j,l)}
\times \sum_{t=\max(j,l)+1}^{T} \hat{\varepsilon}_t^2 \hat{\Psi}_{t-j}(v) \hat{\Psi}_{t-l}(v')^2 dW(v) dW(v'),
\]

and \( \hat{\Psi}_t(v) = e^{iv\hat{\varphi}_t} - \hat{\varphi}(v) \), and \( \hat{\varphi}(v) = T^{-1} \sum_{t=1}^{T} e^{iv\hat{\varphi}_t} \). An example of \( W(\cdot) \) is the CDF of \( N(0,1) \), which is commonly used in the characteristic function literature. Under the null hypothesis of martingale and some regular conditions, \( \hat{M}_1(p) \) has a limit distribution of standard normality.

Different combinations of partial derivatives with respect to \((u,v)\) can help explore the nature of serial dependence. If serial dependence exists in mean, different types of serial dependence have valuable implications for predictability of \( \varepsilon_T \). Once generic serial dependence in the conditional mean is discovered use \((m,l) = (1,0)\), we can go further to use various combinations of \((m,l) = (1,l), l = 2, 3, 4\), testing if \( \text{cov}(\varepsilon_T, \varepsilon_{T-j}^l) = 0 \) for all \( j > 0 \). These essentially check whether there exist ARCH-in-mean, skewness-in-mean and kurtosis-in-mean nonlinear effects. Skewness-in-mean can arise due to time-varying asymmetric behavior, while kurtosis-in-mean effects can arise due
to time-varying heavy-tails of the conditional distribution of $\varepsilon_T$ given $I_t$. The corresponding test statistics can be represented as $\hat{M}_3$, $\hat{M}_4$ and $\hat{M}_5$ representatively. The choice of the lag order $p$ can use the data-driven method proposed in Hong and Lee (2005). See Hong and Lee (2005) for more details.

By using $\hat{M}_1$, $\hat{M}_2$, $\hat{M}_3$, $\hat{M}_4$ and $\hat{M}_5$, we can provide a more comprehensive analysis of efficiency and predictability of futures prices. A rejection based on $\hat{M}_1$ means that $E(\varepsilon_t \mid I_{t-1}) \neq 0$. There is serial dependence in the conditional mean of deviations. It is an evidence against unbiasedness and market efficiency. At this point, if $\hat{M}_2$ is insignificant, which implies the series of deviations is serial uncorrelated, there would be only nonlinearities in the conditional mean of deviations. We can go further to find some patterns of serial dependence in the conditional mean by checking $p$-values of $\hat{M}_3$, $\hat{M}_4$ and $\hat{M}_5$.

3. Data

In our empirical study, we use monthly observations of spot prices and futures prices of WTI crude oil. The monthly spot price is the closing spot price for the last trading day of the month. The monthly futures prices for 1- and 3-month ahead contracts are the closing futures prices for the last trading day of the month. The data are from the Energy Information Administration (EIA), spanning from January 1989 to June 2010.

To test market efficiency, we consider the relationship between future spot prices $S_{j,y}$ and current futures prices $F_{j,y-h}$, where contract maturities $j = 1, 2, \ldots, 12$, years $y = 1, 2, \ldots, Y$, and forecast intervals $h = 1$ and 3. It is natural to pool data across different maturity months. This kind of
aggregation data sets are usually constructed by taking sampling intervals smaller than the forecast intervals. However, aggregating observations may exaggerate the importance of exceptional observations caused by some new information in the market. Take 3-month ahead futures for an example. In a given year, deviation $\varepsilon_9$ in September is a difference between the spot price $S_9$ in September and the futures price $F_6$ in June. It contains new information from June to September. Similarly, its adjacent deviation $\varepsilon_8$ in August reflects new information from May to August. Clearly, $\varepsilon_9$ and $\varepsilon_8$ cover a same period from June to August, so they are impacted by some same information in the market. Consequently, we may find significant unconditional mean and/or conditional mean of deviations. Therefore, pooling data carries the danger of biasing test results towards an incorrect rejection. This problem is referred to as overlapping observations problem.

In order to avoid the overlapping problem, a large number of analysts construct nonoverlapping data sets, of which the sampling interval equals the forecast interval (Cornell, 1977; Frenkel, 1979; Geweke and Feige, 1979; Peroni and Mcnown, 1998; McKenzie et al., 2002, etc.). Nonoverlapping sample set can mitigate the effect of overlapped observations. It should be noted that acception of $E(\varepsilon_T|I_t) = 0$ over nonoverlapping sample sets is not a sufficient support of market efficiency. Different sampling intervals constitute different information sets $I_t$. Nonoverlapping samples are no more than subsets of all the current information used to make decisions in practice. Therefore, we should be cautious about choosing sampling intervals. Both overlapping and nonoverlapping data sets have pros and cons. Thus, we construct both to investigate how results differ with different sampling intervals.
Table 1: Data sets

<table>
<thead>
<tr>
<th>Data sets</th>
<th>h</th>
<th>k</th>
<th>Maturity months</th>
<th>Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>$F_{1S1}$</td>
<td>1</td>
<td>1</td>
<td>1,2,⋯,12</td>
<td>258</td>
</tr>
<tr>
<td>$F_{3S1}$</td>
<td>3</td>
<td>1</td>
<td>1,2,⋯,12</td>
<td>258</td>
</tr>
<tr>
<td>$F_{3S3_1}$</td>
<td>3</td>
<td>3</td>
<td>1,4,7,10</td>
<td>86</td>
</tr>
<tr>
<td>$F_{3S3_2}$</td>
<td>3</td>
<td>3</td>
<td>2,5,8,11</td>
<td>86</td>
</tr>
<tr>
<td>$F_{3S3_3}$</td>
<td>3</td>
<td>3</td>
<td>3,6,9,12</td>
<td>86</td>
</tr>
</tbody>
</table>

Deviations are calculated for 1- and 3-month horizons as in equation (3). We consider several sampling intervals, with which different sample sets are listed in Table 1. A data set is represented as $F_{hSk_i}$, where $h$ is the forecast horizon, $k$ is the sampling interval. For example, $F_{3S1}$ denotes a data set of deviations with a forecast horizon of 3 and a sampling interval of 1. $F_{3S1}$ is an overlapping data set, while the others are nonoverlapping sets.

4. Empirical results

Using the data we have described in previous section, we have many series of deviations. We first estimate unconditional mean of the ex-post deviations, i.e., the average of $S_{T} - F_{t,T}$, and test whether it is significantly different from zero. Lastly, we explore serial dependence, especially nonlinear dependence, in conditional mean of deviations using Hong and Lee’s (2005) generalized spectral tests for conditional mean models in time series with conditional heteroscedasticity of unknown form.
4.1. Unconditional mean

Table 2: Unconditional mean of deviations

<table>
<thead>
<tr>
<th>January 1989-June 2010</th>
<th>Data set</th>
<th>Mean</th>
<th>p-value</th>
<th>observation</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1S1</td>
<td>0.01193</td>
<td>0.39</td>
<td></td>
<td>258</td>
</tr>
<tr>
<td>F3S1</td>
<td>0.04558</td>
<td>0.14</td>
<td></td>
<td>258</td>
</tr>
<tr>
<td>F3S3_1</td>
<td>0.04697</td>
<td>0.18</td>
<td></td>
<td>86</td>
</tr>
<tr>
<td>F3S3_2</td>
<td>0.04355</td>
<td>0.23</td>
<td></td>
<td>86</td>
</tr>
<tr>
<td>F3S3_3</td>
<td>0.04252</td>
<td>0.21</td>
<td></td>
<td>86</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>January 1989-February 2007</th>
<th>Data set</th>
<th>Mean</th>
<th>p-value</th>
<th>observation</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1S1</td>
<td>0.01152</td>
<td>0.32</td>
<td></td>
<td>218</td>
</tr>
<tr>
<td>F3S1</td>
<td>0.04874</td>
<td>0.07</td>
<td></td>
<td>218</td>
</tr>
<tr>
<td>F3S3_1</td>
<td>0.04030</td>
<td>0.26</td>
<td></td>
<td>60</td>
</tr>
<tr>
<td>F3S3_2</td>
<td>0.03898</td>
<td>0.32</td>
<td></td>
<td>60</td>
</tr>
<tr>
<td>F3S3_3</td>
<td>0.03575</td>
<td>0.31</td>
<td></td>
<td>60</td>
</tr>
</tbody>
</table>

If unconditional mean of deviations $\varepsilon_T$ is significantly different from zero, we can infer that there is unconditional bias of futures prices relative to the future spot prices, which is often considered as an evidence of constant risk premium and thereby a rejection of unbiasedness and efficiency in crude oil futures prices. We use $t$-statistic based on heteroskedasticity and autocorrelation consistent (HAC) standard errors to test whether the unconditional mean is zero or not. Alquist and Kilian (2010) use overlapping data set and
find a significant biases of 3-month futures over the period from January 1989 to February 2007. In order to compare the performance of overlapping and nonoverlapping data sets, we calculate the average values of several data sets over both the whole sample period and the subperiod used in Alquist and Kilian (2010).

Table 1 provides p-values of the unconditional mean deviations. Over the whole sample period, all mean deviations are insignificant different from zero for both 1-and 3-month ahead contracts. This implies that there is no constant risk premium. However, during the subperiod from January 1989 to February 2007, unconditional mean of the overlapping sample set $F3S1$ is significant as same as in Alquist and Kilian (2010), while other nonoverlapping sets $F3S3_i$, $i = 1, 2, 3$ have insignificant unconditional biases. We can see that overlapping data set tends to have deviations with nonzero unconditional mean. We must be cautious to use overlapping data when testing unconditional mean of deviations, because aggregating observations may exaggerate the exceptional deviations caused by factors external to the market. This point is not novel, but it has been widely neglected by related studies in oil futures market.

4.2. Serial dependence in conditional mean

To explore serial dependence in conditional mean of deviations, we use Hong and Lee’s (2005) generalized spectral tests for conditional mean models in time series with conditional heteroscedasticity of unknown form. A large set of the preliminary bandwidth $\tilde{p}$ is considered from 10 to 50. We use the $N(1,0)$ CDF truncated on $[-3,3]$ for the weighting function $W(\cdot)$, and use Bartlett kernel for $K(\cdot)$ which has a bounded support and is computationally
efficient.  

Table 3: P-values of serial dependence in the conditional mean of deviations

<table>
<thead>
<tr>
<th></th>
<th>ˆM_1</th>
<th>ˆM_2</th>
<th>ˆM_3</th>
<th>ˆM_4</th>
<th>ˆM_5</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1S1</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.4391</td>
<td>0.0831</td>
<td>0.4479</td>
</tr>
<tr>
<td>F3S1</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0935</td>
<td>0.0025</td>
<td>0.0619</td>
</tr>
<tr>
<td>F3S3_1</td>
<td>0.0030</td>
<td>0.0002</td>
<td>0.4693</td>
<td>0.3187</td>
<td>0.4745</td>
</tr>
<tr>
<td>F3S3_2</td>
<td>0.0110</td>
<td>0.0045</td>
<td>0.5901</td>
<td>0.3940</td>
<td>0.5401</td>
</tr>
<tr>
<td>F3S3_3</td>
<td>0.1938</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 3 provides p-values of the five test statistics ˆM_l, l = 1, 2, 3, 4, 5 with ̄p = 10. For F1S1, ˆM_1 is significant, implying this series of ε_T has serial dependence in the conditional mean. Significant ˆM_2 means the existence of serial correlation. Meanwhile, weakly significant ˆM_4 means that there is skewness-in-mean effect supporting the existence of nonlinear serial dependence in the conditional mean of deviations with its own past values.

With respect to F3S1, significant ˆM_1 shows F3S1 has serial dependence in conditional mean, and significant ˆM_2, ˆM_3, ˆM_4, and ˆM_5 imply that there exist both linear and nonlinear serial dependence in conditional mean.  

Hong and Lee (2005) find that W(·) and K(·) have little impact on both the level and the power of the tests. We also find that the tests are robust to the choice of preliminary bandwidth ̄p.

The null hypothesis of Hong and Lee’s (2005) generalized spectral statistics is E(ε_T | I_{T-1}) = 0. Therefore, p-values for F3S1 in Table 1 contain serial dependence in the conditional mean of ε_T with ε_{T-1} and ε_{T-2}, information of which is actually unavailable.
Skewness-in-mean can arise due to time-varying asymmetric behavior, while kurtosis-in-mean effects can arise due to time-varying heavy-tails of the conditional distribution of $\varepsilon_T$ given $I_t$.

For nonoverlapping data sets $F3S3_1$ and $F3S3_2$, significant $\hat{M}_1$ also means that they have serial dependence in conditional mean. Serial correlation exists, which is implied by significant $\hat{M}_2$. Based on these results of nonoverlapping data sets, the joint hypothesis of unbiasedness and efficiency is rejected for both 1- and 3-month futures contracts. However, insignificant $\hat{M}_2$, $\hat{M}_3$, and $\hat{M}_4$, indicate there are no ARCH-in-mean, skewness-in-mean and kurtosis-in-mean effects. Since nonoverlapping data sets do not contain the whole information sets, different results with those of overlapping data are plausible.

We can not find any serial dependence in $F3S3_3$. No strong reasons exist for the degree of unbiasedness and market efficiency to be the same from maturity month to maturity month. Hence, it is plausible that different nonoverlapping data sets have different results.

Empirical results indicate that the hypothesis of unbiasedness and market efficiency is rejected for both 1-and 3-month ahead futures contracts. It contradicts conclusions in Crowder (1993), Gülen and Gürcan (1998), Chernenko et al. (2004), Abosedra and Baghesani (2004), Chinn et al. (2005), among others. These previous papers can not find evidence against unbiasedness and market efficiency, because they neglect to test the serial dependence of deviations. Our result also contradicts Green and Mork (1991), Deaves and

at time $t$. Nonetheless, these results have significant implications for the predictability of crude oil futures price.
Krinsky (1992) and Peroni and Mcnown (1998) which find serial uncorrelated deviations. It can be shown that results in the three papers are not robust to updated sample. The three papers use more earlier data that we can not get, so we are not able to replicate their results using our data and statistics but with their sample period. For the full sample, however, we do reject the hypothesis at longer horizons.

4.3. Interpretations and implications

A rejection of the hypothesis should not equate readily with inefficiency in the crude oil futures market. To deeply investigate underlying reasons, a more visualizing way is to test cross-sectional predictability of deviations, namely to work on forecasting deviations with other variables. In fact, studies focusing on determinants of risk premia have done some of these jobs, although with different aims. Serial dependence in the deviations we found in this paper can be well explained by theoretical and empirical findings in studies of this kind.

According to the risk premium theory, deviations are simply the sum of risk premium and forecasting errors. One simplest possibility to consider is that the risk premium is stable during the sample period. That is, the risk premium is a constant component over time. In this case, the unconditional mean could be used as the estimator for the time-invariant premium. However, we can not find a significant constant term emerging in our results. This confirms the findings in Dominguez (1989) and Bessembinder (1992). There is an extensive literature, both theoretical and empirical, that relates futures risk premium to some determinants: hedging pressure, inventory levels, hedging demand and among others. It’s plausible that time-varying risk
premium may exist, because appropriate compensation for risk depends on factors whose values change over time.

Hedging pressure effects are related to one of the risk transferring function of futures markets, and result from nonmarketable positions, where, for example, the shares in the storage process owned by certain individuals are not tradeable as indicated in Stoll (1979), and also from market frictions such as transaction costs and information asymmetries. If the speculators are risk averse, they will charge a premium as a reward for accepting the price risk which hedgers sought to transfer. Suppose hedgers (or the majorities of hedgers) are net short, as in the case of an oil producer about to pump crude oil out of the ground, in order to insulate themselves from price uncertainties they sell crude oil futures contracts. Correspondingly, speculators will take long positions to supply risk insurance. This means that risk premia are positive. On average, hedgers are net short and speculators are net long. This is supported by the data on positions of large traders reported by the CFTC, and Haushalter (2000, 2001)'s surveys on 100 oil and gas producers over the 1992 to 1994 period which find that close to 50 percent of them hedge, in the amount of approximately a quarter of their production each year. Empirically, hedging pressure is measured using data on positions of large traders published by the CFTC, and is significant to explain risk premia in crude oil futures market (e.g., de Roon et al., 2000; Wang, 2003). Meanwhile, the positions of hedgers and speculators are highly impacted by past futures returns(e.g., Wang, 2003). Hence, futures returns have serial dependence in the conditional mean with its own past values. Therefore, the hedging pressure effects can explain the serial dependence in the conditional
mean of deviations we found in this paper.

Another important determinant of the time-varying risk premium is inventory level of crude oil, which has been studied and proved, both theoretically and empirically, by Hirshleifer (1989), Dincerler et al. (2003) and Gorton et al. (2007). A low storage level means an increased risk of stock-out, resulting in the risk-averse investors demanding a higher risk premium. Moreover, abnormal inventories are slow adjusted as shown in Gorton et al. (2007). Since past deviations are signals of past shocks of inventories, the deviations are expected to be correlated with futures risk premium. This also explains the serial dependence in the conditional mean of deviations we found in the present paper.

Recently, Acharya et al. (2010) study how hedging demand could be another determinant of the time-varying risk premium. General aversion of managers to variance of cash flows is a drivers of hedging demand. By constructing a theoretical model, as well as empirical analysis, Acharya et al. (2010) have shown that a limit on the risk-taking capacity of speculators implies a price impact of the hedging demand of risk-averse producers, who are naturally short commodity futures. The price impact results in a cost of hedging, which affects the optimal inventory holding of commodity producers, and, in turn, the commodity spot price. More specifically, when producers’ fundamental hedging demand increases, the managers’ sensitivity to the risk of holding unhedged inventory increases. Consequently, the manager reduces inventory and increases the proportion of future supply that is hedged. The former raises future spot prices, while the latter leads to a higher variance-adjusted demand for short futures contracts, which is accompanied
by increasing the futures risk premium.

Producers’ fundamental hedging demand can be represented by measures of aggregate default risk of the producers of the commodity. Acharya et al. (2010) find that there is considerable time-variation in measured default risk, which indicates that there is economically significant time-variation in the fundamental hedging demand of producers in crude oil markets. Moreover, measures of default risk are highly persistent. This is another explanation for the serial dependence in the conditional mean of the deviations.

In sum, the serial dependence in the conditional mean of deviations can be explained by the existence of time-varying risk premium. Determinants of risk premium are often time varying and highly persistent. Information of these determinants is incorporated in the risk premium and thereby the deviations. Consequently, as we find, the deviations appear to have serial dependence with its past values.

These findings have significant implications for practical trading in oil markets and futures-based forecasting of spot prices. Hedging in crude oil futures markets have costs, caused by risk premia or forecasting errors. This is also supported by Wang (2003) who find that traders of speculators (hedgers) are positively (negatively) correlated with subsequent abnormal returns. Wang points out that abnormal returns are not caused by superior forecasting power of speculators. The rejection of unbiasedness and weak efficiency also means that crude oil futures price is not the best predictor for future spot price. We can not simply use current futures price as a forecast of the future spot price, because there is still predictive contents in the deviations between futures price and spot price. This explains the failure of
futures-based forecasting models found in Morana (2001), Chinn et al. (2005) and Alquist and Kilian (2010).

5. Conclusions

The joint hypothesis of unbiasedness and market efficiency of crude oil futures can be tested through formal econometric analysis. Most previous papers obtain a consensus conclusion supporting the hypothesis. However, this conclusion contradicts another strand of studies which investigate the risk premium pricing in crude oil futures markets. It also contradicts the failure of forecasts based on crude oil futures price. Motivated by these unsolved contradictions, we have a rigorous review on previous papers which test unbiasedness and market efficiency through econometric methods. We find there are several methodological issues making existing evidences less conclusive.

Be addressing those issues, we retest the joint hypothesis of unbiasedness and market efficiency in this paper. Especially, we make a more comprehensive analysis of serial dependence, both linear and nonlinear, in the conditional mean of deviations between futures prices and future spot prices.

Unlike most previous studies, we find evidences of significant serial dependence in conditional mean of deviations, which are against the joint hypothesis of unbiasedness and market efficiency in crude oil futures markets. With respect to the predictability of oil prices, both linearity and nonlinearity exist. This explains the failure of futures-based forecasting models in the literature. Relevant information should be utilized to improve the spot price forecasts inherent in futures prices. Besides, choice of sampling intervals
has impacts on the test. Overlapping data observations tend to exaggerate information and incorrectly reject unbiasedness and market efficiency.

Since market efficiency and asset-pricing issues are inseparable, we cannot split between market inefficiency and a bad model of pricing herein. Does deviation predictability reflect rational variation through time, irrational bubbles, or some combination of the two? We do not give a precise inference. Nonetheless, the evidence of predictable deviations is related in plausible ways to the existence of time-varying risk premium. Some determinants of risk premium, such as hedging pressure, inventory levels and default risk, are time varying and highly persistent. Past information of these determinants is incorporated in past values of deviations. Consequently, deviations have serial dependence with its past values. This leans us toward the conclusion that the market is weak efficiency. Even if some bubble fans disagree on the market efficiency implication of our results on deviation predictability, the tests in this paper improve our understanding of the behavior of crude oil futures returns which has significant implications for market participants and institutions.

References


Chernenko, S.V., Schwarz, K.B., and Wright, J.H., The information content


Fama, E.F., and French, K., Commodity futures prices: Some evidence on


Lean, H.H., McAleer, M., and Wong, W.-K., Market efficiency of oil spot


