Efficiency test of Commodity futures market

Feiyu Liu[1], Shishir Shakya[2]

John Chambers College of Business & Economics

West Virginia University

26506, 1601 University Ave, Morgantown, WV 26501

Abstract

This study examines the return predictability of Commodity futures[3] by testing the quasi-random walk hypothesis using weekly closing prices of each commodity from January 1991 to January 2019. We apply three wild-bootstrapped versions of the variance ratio test (VRT, hereafter) to the raw returns of nine commodities and also employ generalized spectral test (GRT, hereafter) and BDS test[4] to study the nonlinear predictability accounting for conditional heteroscedasticity. The results of three VRT show, on one hand, that there are not only periods of statistically significant return probability, but also periods of non-statistically significant predictability, on the other hand, the GRT and BDS provides arguably statistical significance rejecting the null hypothesis that our time series sample is identically and independently distributed (IID, hereafter). Consequently, our findings reasonably suggest that returns predictability in commodity futures markets change over time in a manner consistent with the adaptive market hypothesis (AMH, hereafter), it seems like investors treat each asset independently given the differences of predictability assets. We also find evidence that returns predictability associated with market conditions.

Keywords: Market Efficiency, Adaptive Market Hypothesis, Commodity futures market, return predictability, Variance ratio test.

[1] Email:filiu@mix.wvu.edu

[2] Email: ss0088@mix.wvu.edu

[3] We select nine commodities based on the Futures Industry Association (FIA) 2017 Annual Report and the availability of data's time span, they are Crude oil, Brent oil, Natural Gas, Corn, Gold, Copper, Coffee, Sugar, Silver, soybean. For consideration of space, we post the results of Crude oil, Gold and Corn. We only post analysis of crude oil , gold and corn in this paper, however, the general results are homogeneous.

[4] Proposed by Brock, Dechert and Schieinkman (1987).

1. Introduction

Since the seminal work of E. Fama (1970) , which summarized the idea of efficient markets hypothesis (EMH, hereafter) by Samuelson (1965) and Fama (1965), the EMH was wildly accepted by academic, financial economists and policymakers. The EMH states that asset prices entirely and promptly reflect all the available and relevant public and private information. As the EMH becomes generally accepted, a question we may need to ask is that why do people care about market efficiency? Three reasons came into our mind. First, the principal function of a capital market is to channel allocation of resources of the economy. Generally, the ideal lies in a market in which prices can provide correct signals for resource allocation: namely, a market in which firms can make production-investment decisions, and investors can choose among the securities that present ownership of firms' activity under the assumption that security prices at any time would "fully" reflect all available information (E. Fama, 1970). Second, the EMH is implicit merit for modern economic research, which is especially the case concerning the widespread use of event studies in modern finance. Whenever researchers conduct event studies, the EMH is treated as an implicit baseline so that any deviation of the actual excess return from the equilibrium return can be ascribed to events over which researchers look. There are no other choices for researchers to consider in event studies. Third, The EMH gives market participants the confidence to trade. If markets are not efficient—that is, they do not reasonably price risk and cash flow—then it is not reasonable to trade in them. Even if mispricing was discovered, in an inefficient market there does have no assurance that prices will correct it and thus compensate market participants for their foresight. Risk will be compensated for efficient markets, meaning investors can invest but not gamble.

While the EMH has become increasingly widely accepted and influential, it is not to say that markets are always right, nor are they perfectly efficient. The mounting empirical evidence or anomalies against it have been emerged in the last decades as well as arising of test and comparison of efficiency across various financial markets. Since the publication of the 1978 special issue of Journal of Financial Economics[5], for example, there is a burgeoning empirical literature suggesting

[5] The purpose of this special issue is to bring together a number of these scattered pieces of anomalous evidence regarding Market Efficiency---Michael C. Jensen. In this special issue, a collection of Ball (197), Watts (1977), Thompson (1978), Galai (1978), Chiras and Manaster (1978), Long (1978), Charest (1978a) and (1978b) is provided.

that market efficiency varies over time and anomalies do exist all the time (for a survey, see Dahabarov and Ziemba (2012)). In the meanwhile, Grossman and Stiglitz (1980) argue that a perfectly efficient market is impossible because the incentive to acquire costly information with a full reflection of all available information by prices disappears. Campbell et al. (1997) also suggest the notion of relative efficient market, which shed lights on measuring the degree of market efficiency, i.e., markets behave relatively efficient but not entirely efficient. The debate between the two troops of market efficiency and inefficiency has been heating up and turning white-hot as Shiller (2003) forcefully argues that the EMH should be replaced by behavioral finance paradigm.

As evidence against the EMH has accumulated, academics have begun to explore alternatives to the standard model of optimizing agents with rational expectations. Lo (2004) has proposed the AMH to develop an underlying theoretical framework for behavioral finance instead of an all-or-nothing argument. The AMH is developed by coupling the evolutionary principle with the notion of bounded rationality (Simon, 1955). Because learning, competition, and evolutionary selection pressures govern the forces that prices drive to their adequate levels, agents are bounded rational rather than hyper-rational. Put another way, a satiated outcome is obtained not analytically, but through an evolutionary process involving trial-and-error and natural selection. The AMH has three testable implications: (1) return predictability of financial assets will generally exist in financial markets, (2) the forces of learning and completion will generally fluctuate or even erode these return predictabilities, (3) the degree of market efficiency is related to market conditions.

The AMH has received increasing attention in the recent academic literature, and there are extensive studies on the stock market and future markets to examine the predictions of the AMH. Todea, Ulici, & Silaghi, (2009) show that profitability of moving average strategies is not constant over time in six Asian markets from 1997 to 2008, and conclude that their findings are in line with those postulated by the AMH. Ito & Sugiyama (2009) study the time-varying autocorrelation of monthly S&P500 returns and show that the degree of market efficiency changes over time, and the market exhibit highly efficient around the year of 2000 and highly inefficient at the end of the 1980s. Kim et al. (2011) examine stock returns predictability of daily Dow Jones Industrial Average index data from 1900 to 2009 using an automatic VRT and automatic portmanteau test. Through using a rolling window, they find strong evidence of time-varying predictability which is driven by market volatility and economic fundamentals. Charles et al. (2012) examine the return predictability of central foreign exchange rates from 1975 to 2009 utilizing daily and weekly nominal exchange rates. Through applying the automatic VRT, GPT and Dominguez-Lobato consistent tests, they find that returns predictability does vary over time relying on changes of market condition, which is consistent with the implications of the AMH. Smith (2012) examines the changing efficiency of fifteen European emerging stock markets and three developed markets. He uses a rolling window VRT and finds that the return predictability changes over time, coinciding with high returns predictability during the Great Recession in countries Croatia, Hungary, Poland, Slovakia, and U.K... Lim, Luo, & Kim, (2013) examine the return predictability for three main U.K. stock indices utilizing the automatic portmanteau Box-Pierce test as well as the wild bootstrapped automatic VRT with a rolling estimation window. They find evidence of time-changing return predictability and that those periods with statistically significant return autocorrelations seem to be primarily associated with main exogenous events, which is consistent with the implications of the AMH as well. Urquhart & Hudson (2013) find strong evidence in support of the AMH, through examining return predictability with a linear and non-linear test in U.S., U.K. and Japanese stock markets. Also, among others, Kim & Shamsuddin (2008), Neely, Weller, & Ulrich (2009), Al-Khazali & Mirzaei (2017) and Noda (2016).

Commodities, such as energy, food or metals, are an essential part of people's life. People who drive a car can be significantly impacted by increasing crude oil prices. The composition of people's next meal may be influenced by the impact of a drought on the soybean supply. Likewise, commodities futures can be an essential way to diversify a portfolio beyond traditional securities – either for the long term investment or as a place to store cash during unusually volatile or bearish stock markets, as commodities traditionally move in opposition to stocks. Furthermore, the estimated 2017 national value of U.S. future markets is about 23 trillion, and commodity futures occupy 32% of the futures trading activity in 2017[6]. Consequentially, the degree of efficiency of the commodity futures market is not negligible.

In this paper, we examine the efficiency of commodity futures market by testing the first and second implication of the AMH. We track the evolution of linear and nonlinear return predictability of nine commodities in the future market using three VRT, the GST, and the BDS. To examine the third implication, we investigate whether the degree of returns predictability of

[6] Commodity Futures Trading Commission 2017 annual report.

commodity futures market depends on market conditions such as bull and bear market, or up and down market. We find strong evidence in support of time-changing returns predictability of the commodity futures market and dependence of returns predictability upon market conditions. Our findings are consistent with the implications of the AMH. In particular, returns predictability is observed to be higher during down and bear time for energy commodity (i.e. crude oil) than it is during up and bull time, but corn and gold seem to show lower returns predictability during downtime than they do during up and bull time. On the other hand, no high degree of uncertainty is associated with measures of return predictability for those commodities. Overall, futures returns appear to have much higher predictability in bad time compared to a good time, which is consistent with the general findings of the previous study in stock and exchange market.

The contribution of this paper to the literature are as follows. First, this is the first study using three wild-bootstrapped VRTs and the non-linear GST on raw returns to examine whether commodity futures' return predictability varies over time. The previous literature on AMH has either studied just a single commodity or has employed alternative methodologies or has utilized a proxy of raw data[7]. Second, this is the first study examining whether commodity futures' return predictability is correlated with market conditions on multiple bases. Third, we employed multiple methods to capture the major dynamics of commodity futures' return predictability to lower the potential bias introduced by a single test may impact our results.

The next section presents the methodology, and section three presents data and the empirical results. We make discussion and conclude section four.

2. Methodology

This section provides a brief description of the statistical tests for return predictability employed in this paper.

2.1 Bootstrapped Chow-Denning variance ratio test

Since the seminal work of Lo & Mackinlay (1988), the VRT has been widely employed to test the weak-form efficiency of financial markets. Suppose that x_t is a time series of asset returns with a sample size T. The variance ratio is defined as

[7] See Kristoufek & Vosvrda (2014), Charles, Darné, & Kim (2015), and Urquhart (2017).

$$VR = \left\{ \frac{1}{Tk} \sum_{t=k+1}^{T} (y_t + y_{t-1} + \ldots + y_{t-k} - k\hat{\mu})^2 \right\} \div \left\{ \frac{1}{T} \sum_{t=1}^{T} (y_t - \hat{\mu})^2 \right\} ,$$

Where $\hat{\mu} = T^{-1}\sum_{t=1}^{T} y_t$, and y_t is IID. More concisely,

$$VR(k) = 1 + 2\sum_{i=1}^{k-1} (1 - \frac{i}{k})\hat{\rho}_i$$

where $\hat{\rho}_i$ is the estimation of the ith order autocorrelation of the returns and k is the holding period. The null hypothesis of the VRT is $VR(k) = 1$ or $\rho_i = 0$ for all k. And the choice of k is often arbitrary[8]. (Choi, 1999) propose a method of estimating the optimal $\hat{k}$. Under the null hypothesis of $VR(k) = 1$, (Choi, 1999) shows that,

$$Z(\hat{k}) = \sqrt{T / \hat{k}} \left[VR(\hat{k}) - 1 \right] / \sqrt{2} \xrightarrow{d} N(0,1) ,$$

Under the assumption that returns are IID. Statistical inference is probably invalid if returns are subject to an unknown form of conditional heteroscedasticity, especially for small sample size. Another problem with the traditional VRT is that under the null hypothesis of no predictability, $VR(k) = 1$ must be right for all k. Therefore, the test for the null hypothesis should be a joint test for all value k or multiple values of k. To overcome this issue, Chow and Denning (1993) propose a multiple VRT where only the maximum absolute value of $VR(k)$ in a set of m test statistics considered. Taking into all issues discussed, a wild-bootstrapped VRT proposed by (Kim, 2006) to improve the quality of VRT in small sample size incorporated with Chow-Denning statistics. The bootstrapped p-values are computed directly from the fraction of replications falling outside the bounds defined by the estimated statistics.

2.2 Bootstrapped Wright variance ratio test

Wright.J.H. (2000) proposes a non-parametric alternative to the conventional VRT using ranks and signs that avoid any asymptotic approximation and may be more powerful than alternative tests if the data are in fact highly non-normal. The signs-based test is exact even under conditional heteroscedasticity and the ranks-based displays low-size distortion under heteroscedasticity. Let

[8] 2, 4, 8 and 16 was used in this paper.

$r(x_t)$ be the rank of log return x_t among $x_1, x_2 \ldots, x_T$ which, under the random walk hypothesis, is just a random permutation of the numbers $1, 2, \ldots T$, each has equal probability. Define the rank-based VRT statistics:

$$R_i(k) = \left(\frac{(Tk)^{-1} \sum_{t=k}^{T} (r_{it} + \ldots + r_{it-k+1})^2}{T^{-1} \sum_{t=1}^{T} r_{it}^{\ 2}} - 1 \right) \left(\frac{2(2k-1)(k-1)}{3kT} \right)^{-1/2},$$

Where $i = 1, 2$, and

$$r_{1t} = \frac{\left[r(x_t) - (T + \frac{1}{2}) \right]}{\sqrt{((T-1)(T+1))/12}},$$

$$r_{2t} = \frac{\Phi^{-1} r(x_t)}{T+1},$$

where Φ is the standard normal cumulative distribution function. The sign-based test statistic is defined:

$$S_i(k) = \left(\frac{(Tk)^{-1} \sum_{t=k}^{T} (s_{it} + \ldots + s_{it-k+1})^2}{T^{-1} \sum_{t=1}^{T} s_{it}^{\ 2}} - 1 \right) \left(\frac{2(2k-1)(k-1)}{3kT} \right)^{-1/2},$$

Where $s_t = 2\mu(x_{t,} 0) = 2\mu(\varepsilon_t, 0)$. And s_t is an IID series with mean 0 and variance 1. Each s_t is equal to 1 with probability $\frac{1}{2}$ and is equal to -1 otherwise.

2.3 Generalized spectral test (test of non-linear predictability)

The preceding three tests are based on autocorrelation, and able to detect linear dependence only. However, evidence of nonlinearity of stock returns is well documented. When the returns follow a general martingale difference sequence, its normalized spectral density function is equal to one at all frequencies. Based on this, (Escanciano & Velasco, 2006) propose a GST, which can capture both linear and non-linear dependence. This test involves wild bootstrapping, in a similar manner to the VRT discussed earlier, where the p-value of the test can be obtained. i.e., if the p-values are less than 0.05, the null hypothesis of no return predictability will be rejected at the 5% significance level.

2.4 BDS test

To detect serial dependence or a nonlinear structure in the return series, a non-parametric method is proposed by Brock, Dechert, and Schieinkman (1987)[9] which examines the null hypothesis of independent and identically distributed (whiteness) against an alternative hypothesis of misspecified model. Given a sample of *i.i.d* observations $\{x_t : t = 1, \cdots, T\}$, the correlation integral $C_n(\varepsilon)$ measures the probability that any two points $\{x_i\}$ meet within distance ε from each other in n dimensional phase space, and is equal to the product of the individual probabilities, provided that pairs of points are independent given as: $C_n(\varepsilon) = \prod_{i,j(i \neq j)} p\left(\left\| x_i - x_j \right\| < \varepsilon\right)$ for $n \to \infty$ and for IID observations, so,

$$C_n(\varepsilon) = C_1(\varepsilon)^n$$

Then statistics

$$B(n,T,\varepsilon) = \sqrt{T}\left(C_{n,T}(\varepsilon) - C_{1,T}(\varepsilon)\right)$$

would converge to an asymptotic normal distribution with zero mean and $V_{n,T}^2(\varepsilon)$ variance. Then BDS statistic is defined as:

$$W_{n,T}(\varepsilon) = \frac{\sqrt{T}\left(C_{n,T}(\varepsilon) - C_{1,T}(\varepsilon)^n\right)}{\sqrt{V_{n,T}^2(\varepsilon)}}$$

Where, $W_{m,n}(\varepsilon)$ is the BDS statistics, T is the sample size, n is the embedding dimension (history) and the metric bound (ε) is the maximum difference between pairs of observations counted in computing the correlation integral. Under moderate conditions the test converges to standardized normal distribution.

To perform BDS test, the return series needs to be whitened due to the easiness of detection of linear dependence, therefore, an $AR(p)$ model is selected based on the Akaike information

---

[9] Also see Brock, Dechert, Scheinkman, & LeBaron (1996)

criterion (AIC) then we tested Ljung-Box Q-Statistic up to 10 lags on standardized residual (to capture if linear correlation exist or not) and residual squared (to capture if ARCH effect exist or not).

As the EMH does not impose any restrictions on the dynamics of the conditional variance, any nonlinear dependence found due to conditional heteroscedasticity is not a violation of the EMH. Because the BDS has been proven to have high power against ARCH and GARCH models where nonlinearity enters through the conditional variance, an AR-GARCH (1,1) was fitted in this paper on the returns and its standardized residuals to test $i.i.d$ using BDS test as follow:

$$r_t = \beta_0 + \sum_{t=1}^{p} \beta_i r_{t-i} + \varepsilon_t \,,$$

$$\varepsilon_t \sim N(0, h_t)$$

$$h_t = \alpha_0 + \alpha_1 h_{t-1} + \alpha_2 \varepsilon_{t-1}^2$$

If the BDS test find the AR-GARCH filtered returns have statistically significant dependence, then there seems to be nonlinear dependence in the return series.

3. Data and Empirical Results

3.1 Data

Our sample data consist of weekly closing prices for nine commodity futures obtained from Quandl Inc. Crude oil, Gold and Corn stand for the highest volume and the most liquid commodities in energy, precious metal, and agriculture future markets respectively. The data spans January 1, 1993, to January 20, 2019. Several reasons determined the choice of weekly observation interval. Firstly, the VRT theories are based wholly on asymptotic approximations, a large number of observations is appropriate. Secondly, while daily sampling yields many observations, the biases associated with non-trading, the bid-ask spread, asynchronous process, etc., are problematics. Even though weekly sampling alone does not ensure against the biases' possibly substantial influences, it is an ideal compromise, yielding a large number of observations while minimizing the biases inherent in daily data data[10].

Insert Table.1

Fig.1 displays the price trends crossing time and log return patterns of the three commodities. Where it's easy to see that returns of gold is with lowest variation, while descriptive statistics in Tables 1 show that the Gold the highest return but with the lowest standard deviation, i.e., the lowest volatility (which is consistent with the observation from Fig.1), and that the Crude oil is the most volatile in our sample. All three return series have negative skewness indicating a longer left tail, which also can be confirmed by the visualizations of the left part of Fig.1. Excess kurtosis is observed for all return series, showing leptokurtic distributions for them. The highly significant Jarque-Bera test statistics indicate a violation of the normality of their returns' distribution, which is consistent with the left-skewness. We apply the LM test to the residuals of a fitted ARMA model for each series for the examination of conditional heteroscedasticity of their returns, and all three return series display trait of conditional heteroscedasticity at 1% significance level.

Insert Fig.1

3.2 Returns predictability over time

Fig.2-4 displays time-varying p-values based on three VRT, i.e., Chow-Denning test, Wright's Ranks test, and Signs test, through a 105-week fixed-length moving subsample window analysis. We evaluate the statistical significance of the three VRT by using p-values where the significance level at 5%, i.e., the null hypothesis of non-predictable returns would be reasonably rejected if the p-value is less than or equal to 5%.

Fig.2 provides evidence from the three VRT of returns of crude oil future. It is clear to see that in some periods the three VRT generate very different p-values, reflecting the difference between these formulations of the three VRT. P-values vary over time throughout our sample, with some periods generating statistically significant p-values and quite high p-values in others. All the three VR tests almost fail to reject the null hypothesis in the first four years of our sample. Starting from

[10] See Lo & MacKinlay (1987).

September 1995 to April 1997, almost all of the p-values for each of the three VRT are statistically significant, suggesting the dependence and predictability of the Crude oil returns. However, from May 1997 to January 2015 the p-values for each of the three VRT are almost insignificant with rare exceptions. A more interesting phenomenon in this period is that the absence of Crude oil's predictability in the Great Recession, which is inconsistent with the findings in the stock market. From January 2015 to January 2017, the Wright's Signs test is persistently significant, the other two VRT, however, fluctuate between being statistically significant and insignificant, the cliff-style drop of the p-values by the end of 2014 is corresponding to the start of a persistent decreasing of the price of crude oil. Therefore, the VRT results for Crude oil are in favor of the AMH regarding the episodic predictability or returns and confirm the first two implications of The AMH.

Insert Fig.2

Fig.3 demonstrates the VRT p-values over time through a 105-week rolling window analysis on the gold future. There are four noteworthy periods which might reveal predictability at some degree; they are the end of May 1995 to May 1996, September 1998 to June 1999, September 2001 to February 2003 and probably most of the time of the Great Recession. For the rest of the time in our sample, the three VRT fail to reject the null hypothesis. Hence, we are reasonably sure that the varying behavior of the commodity returns provides evidence of the AMH.

Paying attention to the four periods of the statistically significant returns predictability of gold future, we find that there are four global economic events with which seems likely being incorporated; they are Asian financial crisis, Dot-com bubbles, U.S. housing, and sub-prime mortgage[11]. Namely, returns predictability of gold future seems to be higher during bad time than it during a good time.

Insert Fig.3

Likewise, the p-values of VRT on corn future also reveal varying predictability over time. In Fig.4, there is evidence that relative returns predictability emerged from the end of 1993 to the first half year of 1994. Interestingly, corn futures show the same pattern of returns predictability as of gold future, i.e., most of the three VRT give back significant p-values during the Asian

[11] The bubbles and crisis time are identified by Bordo (2003)

financial crisis, Dot-com bubbles and sub-prime mortgage time. Despite some mixed results, we are confident in this evidence showing a varying level of returns predictability of corn future which is in favor of AMH.

Insert Fig.4

(Analysis of GST and BDS goes here once we corrected our codes)

Insert Fig.5

Insert Fig.6

Therefore our results show that each market has experienced differing periods of statistically significant returns predictability and periods of non-statistically significant predictability. These fluctuations of predictability ascribe to which test utilized since different VRT capture varied aspects of returns predictability. Moreover, returns predictability of gold and corn futures seem to hold the same pattern, while the return predictability of crude oil future exhibit another pattern.

It is worth to conduct a horizontal comparison for market efficiency if the testing procedure is consistent, because that if the market was efficient, there should be less than 5% time by random chance that the p-values of VRT are significant. To enable the comparison of the returns predictability of the three commodities, we study the relative efficiency based on the proportion of non-statistically significant p-values in our whole sample. In Table.2, we report the percentage of p-values of each VRT which fails to reject the null hypothesis at the 5% significance level. The Gold is deemed to the most efficient of the three commodities since on average 95.44% p-values from the three VRT fail to reject the null hypothesis of efficiency, while the crude oil is the least efficient with only 89.19% of the p-values failing to reject the null hypothesis.

Insert Table.2

3.3 Return predictability and market conditions

Lo (2004) argues that the degree of return predictability of a market varies with changes of market conditions over time. However, he neither suggests any specific indications of market conditions nor any refutable predictions about the direction of the relation between return predictability and the indicators of market conditions (Kim et al., 2011). Despite the lack of a well-defined model, we regress a weekly measure of returns predictability on some commonly used dummy variables proposed by Fabozzi & Francis (1977) and Klein & Rosenfeld (1987) to group our data sample into bull and bear markets, and up, down and normal markets. This procedure yields a mutually exclusive and exhaustive division of our total sample into subsamples.

Explicitly, we define up time in which the average return was greater or equal to zero, and downtime otherwise. Meanwhile, a period is defined as average if the price of futures in this period raised substantially while the price in surrounding periods is on average. If the price of futures either raised or was normal in one period while the price in the surrounding periods is in deemed bearish, then this period is defined as bearish. Likewise, we define the bullish period that if the price of futures in it declined or was normal while the price in the surrounding periods is in deemed bullish.

Table.3 presents our regression results on the market conditions and the measures of return predictability. The results vary with different market condition and different test statistics for different commodities, but they show that the gold future and corn future share a sam pattern, and crude oil goes the opposite way. Specifically, all regression coefficients are statistically significant in the category of up and down period. Crude oil demonstrates higher returns predictability during downtime than it does during up time, but gold and corn show lower returns predictability during downtime than it does during up time. Despite the arguablly non-significant of the regression coefficients in the category of bull, bear and normal market, the pattern holds consistently, while higher uncertainty is revealed here concerning standard errors comparing to the category of up and downtime.

Insert Table.3

5. Conclusions

We examine the degree of returns predictability of nine commodity futures from U.K. and U.S. future market by testing the implications of the AMH using weekly data from January 1991 to January 2019. Regarding measures of the degree of returns predictability, we employ three variance ratio tests. To detect the possible nonlinear dependence in future returns, the generalized spectral test, and BDS test has been implemented. We evaluate the time-changing returns predictability by applying these tests to a fixed moving subsample window over a weekly grid. In the meanwhile, we conduct a regression analysis to determine how our measures of returns predictability are related to varying market conditions.

Despite the arguably statistical significance of GST and BDS, it is clear to see evidence that returns predictability exists and fluctuates over time in these commodity futures market, with each return series going through periods of statistically significant predictability and periods of non-statistically significant predictability. These findings appear to be varying for the three VRT, which suggests that the linear returns predictability does vary over time, and the efficiency of the market is not an all-or-nothing argument. On the other hand, we can see that different commodities experience differing patterns regarding the dissimilarity of significance in different periods, implying that each commodity future evolves independently at some degree and possibly un-correlated.

We also show that several market conditions are associating with the fluctuation of returns predictability and this changes among different commodities. This difference implies that different commodities adapt and interact to market conditions uniquely. Even though we find evidence from commodity futures market which is in line with the AMH, different commodities hold heterogeneous nature which may differ their evolution of price. Hence, market participants are supposed to treat each commodity independently.

Reference:

Al-Khazali, O., & Mirzaei, A. (2017). Stock market anomalies, market efficiency and the adaptive market hypothesis: Evidence from Islamic stock indices. *Journal of International Financial Markets, Institutions and Money*, *51*, 190–208. https://doi.org/10.1016/j.intfin.2017.10.001

Brock, W. ., Dechert, W. ., Scheinkman, J. ., & LeBaron, B. (1996). A test for independence based on the correlation dimension. *Econometrics Reviews*, *15*, 197–235.

Charles, A., Darné, O., & Kim, J. H. (2012). Exchange-rate return predictability and the adaptive markets hypothesis: Evidence from major foreign exchange rates. *Journal of International Money and Finance*, *31*(6), 1607–1626. https://doi.org/10.1016/j.jimonfin.2012.03.003

Charles, A., Darné, O., & Kim, J. H. (2015). Will precious metals shine? A market efficiency perspective. *International Review of Financial Analysis*, *41*, 284–291. https://doi.org/10.1016/j.irfa.2015.01.018

Choi, I. (1999). Testing the Random Walk Hypothesis. *Journal of Applied Econometrics*, *14*(3), 293–308. https://doi.org/10.1016/0165-1765(85)90058-8

Escanciano, J. C., & Velasco, C. (2006). Generalized spectral tests for the martingale difference hypothesis. *Journal of Econometrics*, *134*(1), 151–185. https://doi.org/10.1016/j.jeconom.2005.06.019

Fabozzi, F. J., & Francis, J. C. (1977). Stability Tests for Alphas and Betas Over Bull and Bear Market Conditions. *The Journal of Finance*, *32*(4), 1093–1099. https://doi.org/10.1111/j.1540-6261.1977.tb03312.x

Fama, E. (1970). Efficient Capital Markets : a Review of the Theory. *The Journal of Finance*, *25(2)*, 383–417. https://doi.org/10.1111/j.1540-6261.1970.tb00518.x

Fama, E. F. (n.d.). Fama 1965 The Behaviour Of Stock-Market Prices.pdf.

Grossman, S. J., & Stiglitz, J. E. (1980). On the Impossibility of Informationally Efficient Markets. *The American Economic Review*, *70*(3), 393–408.

Ito, M., & Sugiyama, S. (2009). Measuring the degree of time varying market inefficiency. *Economics Letters*, *103*(1), 62–64. https://doi.org/10.1016/j.econlet.2009.01.028

Kim, J. H. (2006). Wild bootstrapping variance ratio tests. *Economics Letters*, *92*(1), 38–43. https://doi.org/10.1016/j.econlet.2006.01.007

Kim, J. H. (2009). Automatic variance ratio test under conditional heteroskedasticity. *Finance Research Letters*, *6*(3), 179–185. https://doi.org/10.1016/j.frl.2009.04.003

Kim, J. H., & Shamsuddin, A. (2008). Are Asian stock markets efficient? Evidence from new multiple variance ratio tests. *Journal of Empirical Finance*, *15*(3), 518–532. https://doi.org/10.1016/j.jempfin.2007.07.001

Kim, J. H., Shamsuddin, A., & Lim, K. P. (2011). Stock return predictability and the adaptive markets hypothesis: Evidence from century-long U.S. data. *Journal of Empirical Finance*, *18*(5), 868–879. https://doi.org/10.1016/j.jempfin.2011.08.002

Klein, A., & Rosenfeld, J. (1987). The Influence of Market Conditions on Event-Study Residuals. *Journal of Financial & Quantitative Analysis*, *22*(3), 345–351. https://doi.org/10.2307/2330968

Kristoufek, L., & Vosvrda, M. (2014). Commodity futures and market efficiency. *Energy Economics*, *42*, 50–57. https://doi.org/10.1016/j.eneco.2013.12.001

Lim, K. P., & Brooks, R. (2011). The evolution of stock market efficiency over time: A survey of the empirical literature. *Journal of Economic Surveys*, *25*(1), 69–108. https://doi.org/10.1111/j.1467-6419.2009.00611.x

Lim, K. P., Luo, W., & Kim, J. H. (2013). Are US stock index returns predictable? Evidence from automatic autocorrelation-based tests. *Applied Economics*, *45*(8), 953–962. https://doi.org/10.1080/00036846.2011.613782

Lo, A. W. (2004). The Adaptive Markets Hypothesis. *Journal of Portfolio Management*, *30*(617), 15–29. https://doi.org/10.3905/jpm.2004.442611

Lo, A. W., & Mackinlay, A. C. (1988). The Society for Financial Studies Stock Market Prices do not Follow Random Walks : Evidence from a Simple Specification Test Author (s): Andrew W . Lo and A . Craig MacKinlay Source : The Review of Financial Studies , Vol . 1 , No . 1 (Spring , 1988),. *The Review of Financial Studies*, *1*(1), 41–66.

Lo, A. W., & MacKinlay, A. C. (1987). *Stock Market Prices Do Not Follow Random Walks: Evidence From a Simple Specification Test* (Working Paper Series). *NBER Working Paper* (Vol. 20). https://doi.org/10.3386/w2168

Neely, C. J., Weller, P. A., & Ulrich, J. M. (2009). The adaptive markets hypothesis: Evidence from the foreign exchange market. *Journal of Financial and Quantitative Analysis*, *44*(2), 467–488. https://doi.org/10.1017/S0022109009090103

Noda, A. (2016). A test of the adaptive market hypothesis using a time-varying AR model in Japan. *Finance Research Letters*, *17*, 66–71. https://doi.org/10.1016/j.frl.2016.01.004

Samuelson, P. A. (1965). Proof that Properly Anticipated Prices Fluctuate Randomly. *Industrial Management Review*, *6*(2), 41–49. https://doi.org/10.1109/SP.2006.16

Shiller, R. J. (2003). From Efficient Markets Theory to Behavioral Finance Academic finance has evolved a long way from the days when the efficient. *The Journal of Economic Perspectives*, *17*(1), 83–104.

Simon, H. A. (1955). A Behavioral Model of Rational Choice. *Quarterly Journal of Economics*, *69*(1), 99–118.

Todea, A., Ulici, M., & Silaghi, S. (2009). Adaptive markets hypothesis: evidence from Asia-Pacific financial markets. *The Review of Finance and Banking*, *1*(1), 7–14. https://doi.org/10.1103/PhysRevA.38.4537

Urquhart, A. (2017). How predictable are precious metal returns? *European Journal of Finance*, *23*(14), 1390–1413. https://doi.org/10.1080/1351847X.2016.1204334

Urquhart, A., & Hudson, R. (2013). Efficient or adaptive markets? Evidence from major stock markets using very long run historic data. *International Review of Financial Analysis*, *28*, 130–142. https://doi.org/10.1016/j.irfa.2013.03.005

Wright.J.H. (2000). Alternative variance-ratio tests using ranks and signs. *Journal of Business and Economic Statistics*, *18*(18), 1–9.

	Obs	Mean	SD	Skewness	Kurtosis	JB	ARCH(10)
Crude oil	1464	0.000526	0.049226	-0.76259	8.171739	1780.804***	252.7311***
Gold	1464	0.000817	0.022439	-0.02008	6.81014	889.9803***	178.4339***
Corn	1464	0.000333	0.038721	-0.68924	10.33236	3408.375***	60.06319***

Table 1

Descriptive statistics of the weekly returns of the Crude oil, Gold and Corn. ***, ** and * indicates significance level at 1%, 5% and 10% respectively.

	CD	JR	JS	Avg
crude oil	91.18%	92.06%	84.19%	89.14%
gold	98.09%	95.07%	93.16%	95.44%
corn	94.78%	87.65%	90.74%	91.06%

Table.2

The percentage of non-statistically significant p-values based on VRT at the 5% significance level.

	Crude oil			Gold			Corn		
	CD	JR	JS	CD	JR	JS	CD	JR	JS
UP	0.057*** (0.015)	0.025* (0.015)	0.191*** (0.016)	-0.051*** (0.015)	0.055*** (0.015)	-0.067*** (0.016)	-0.013 (0.016)	0.058*** (0.017)	-0.08*** (0.016)
BULL	0.012 (0.034)	-0.025 (0.034)	0.11*** (0.038)	-0.034 (0.034)	-0.004 (0.035)	0.022 (0.037)	-0.016 (0.04)	-0.008 (0.043)	0.091** (0.041)
Normal	0.044 (0.028)	0.006 (0.028)	0.149*** (0.032)	-0.021 (0.03)	-0.008 (0.031)	0.055* (0.032)	0.015 (0.035)	-0.031 (0.037)	0.061* (0.035)

Table.3

Regression of measure of returns predictability and the market conditions. Standard errors are denoted in parentheses. ***, ** and * indicate significance at 1%, 5% and 10% level respectively.

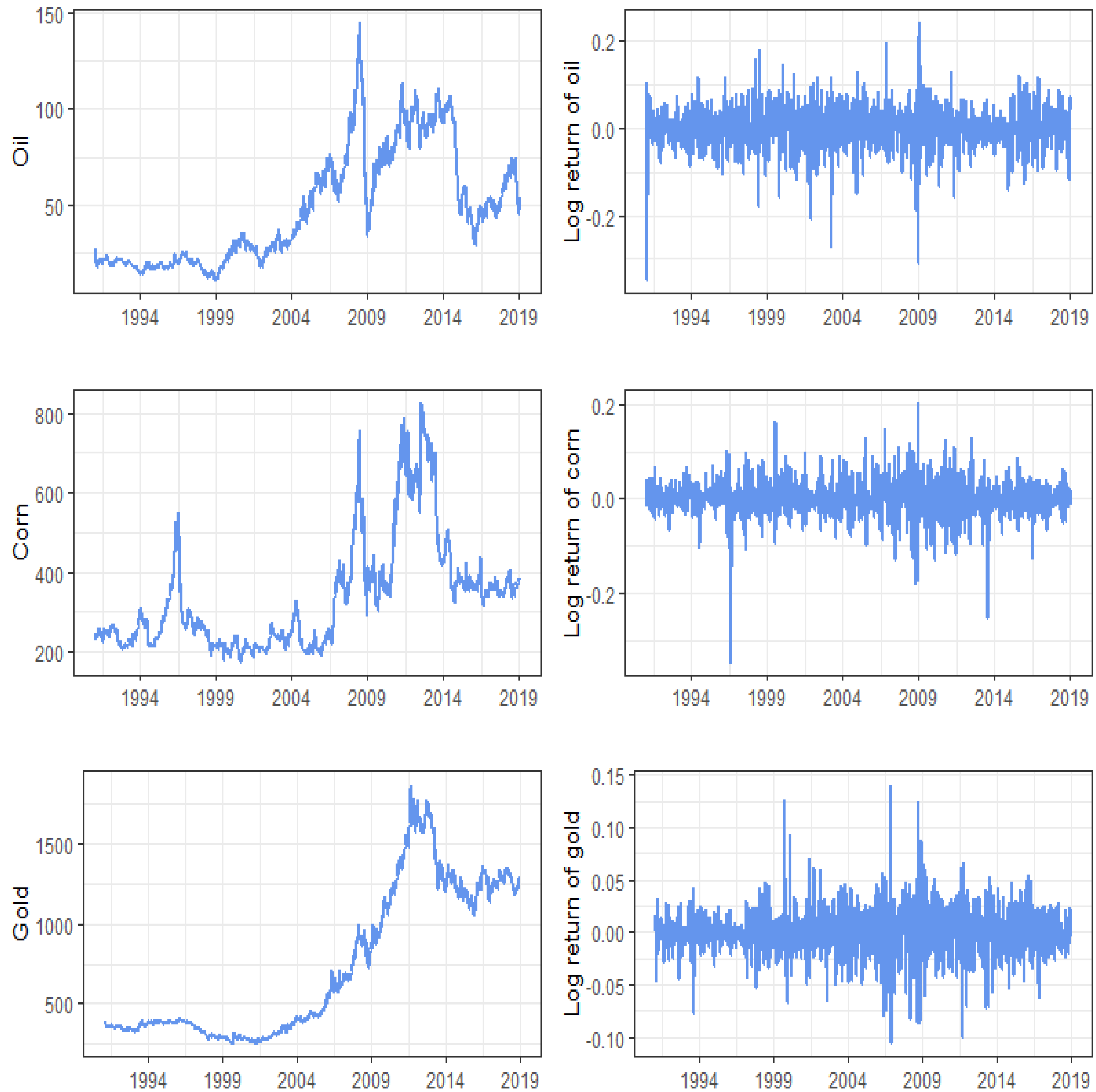

Fig.1. Price trends and log return patterns of Crude oil, Gold and Corn.

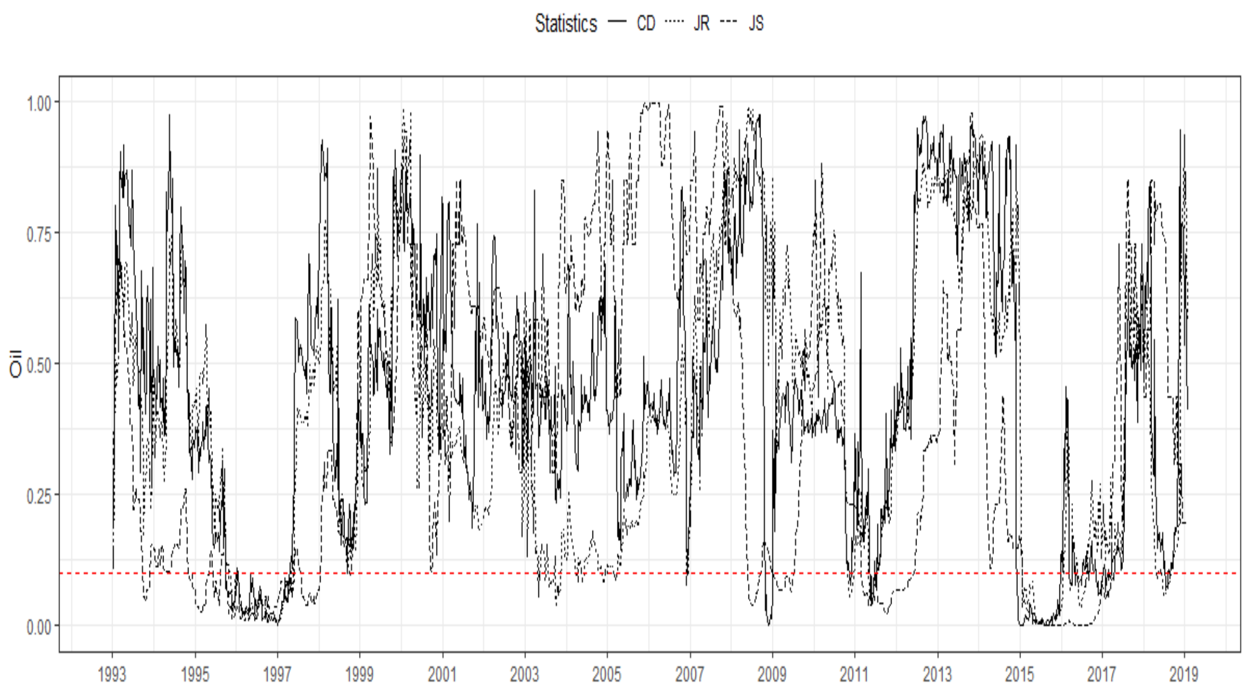

Fig.2 The three different VR joint-test p-values over time for return of Crude oil future based on weekly data and 105-week windows. The red horizontal line is indicating the 5% significance level.

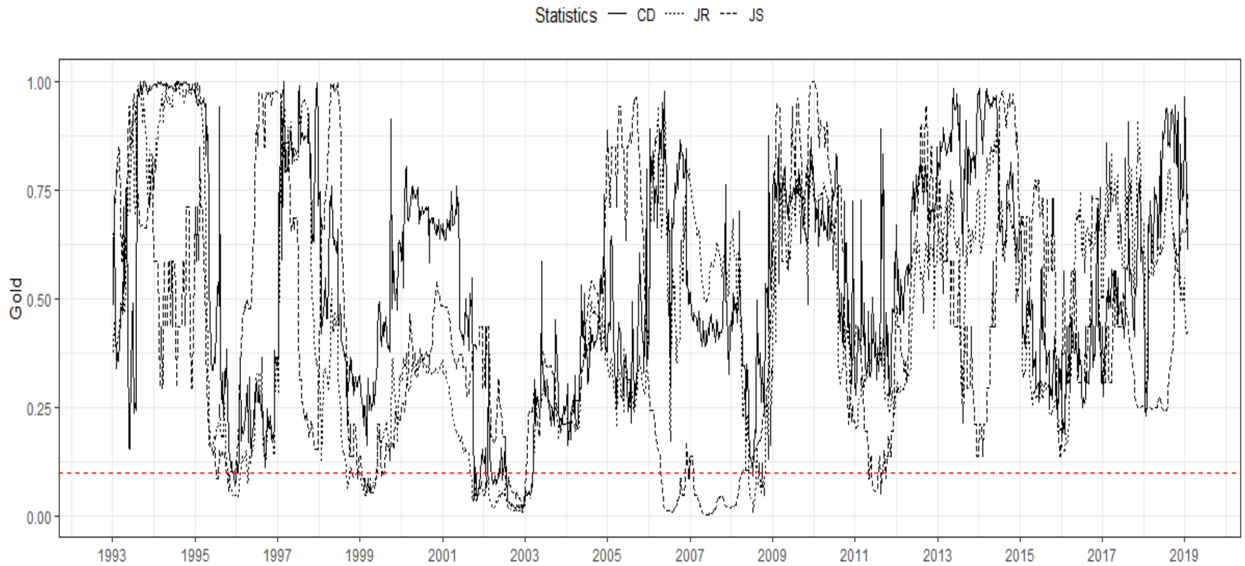

Fig.3. The three different VR joint-test p-values over time for return of Gold future based on weekly data and 105-week windows. The red horizontal line is indicating the 5% significance level.

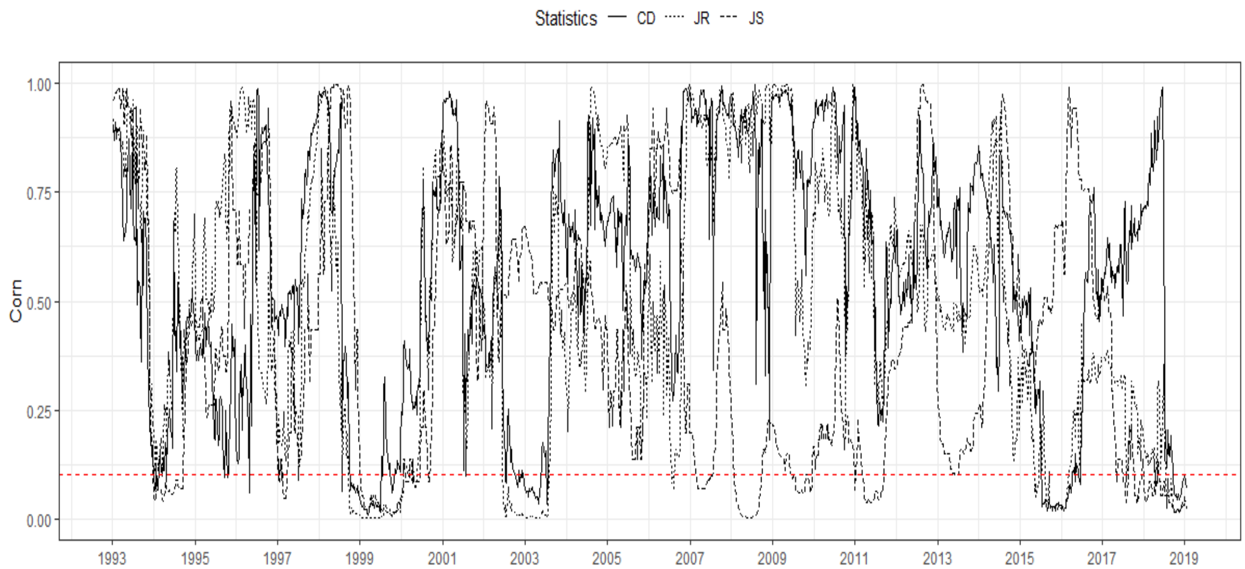

Fig.4. The three different VR joint-test p-values over time for return of Corn future based on weekly data and 105-week windows. The red horizontal line is indicating the 5% significance level.

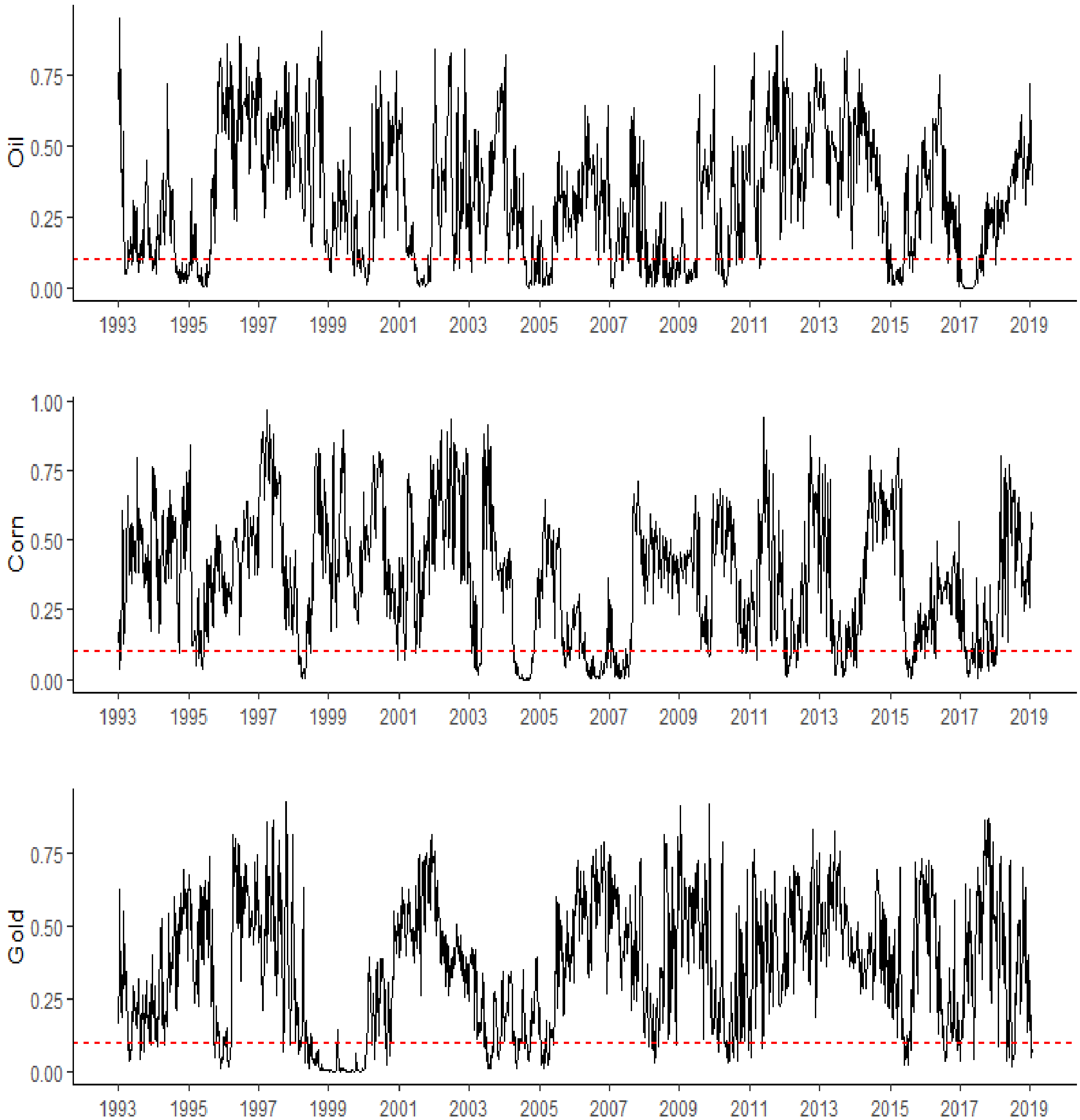

Fig.5. The BDS test p-values over time based on weekly data and 105-week windows. The red horizontal line is indicating the 10% significance level.

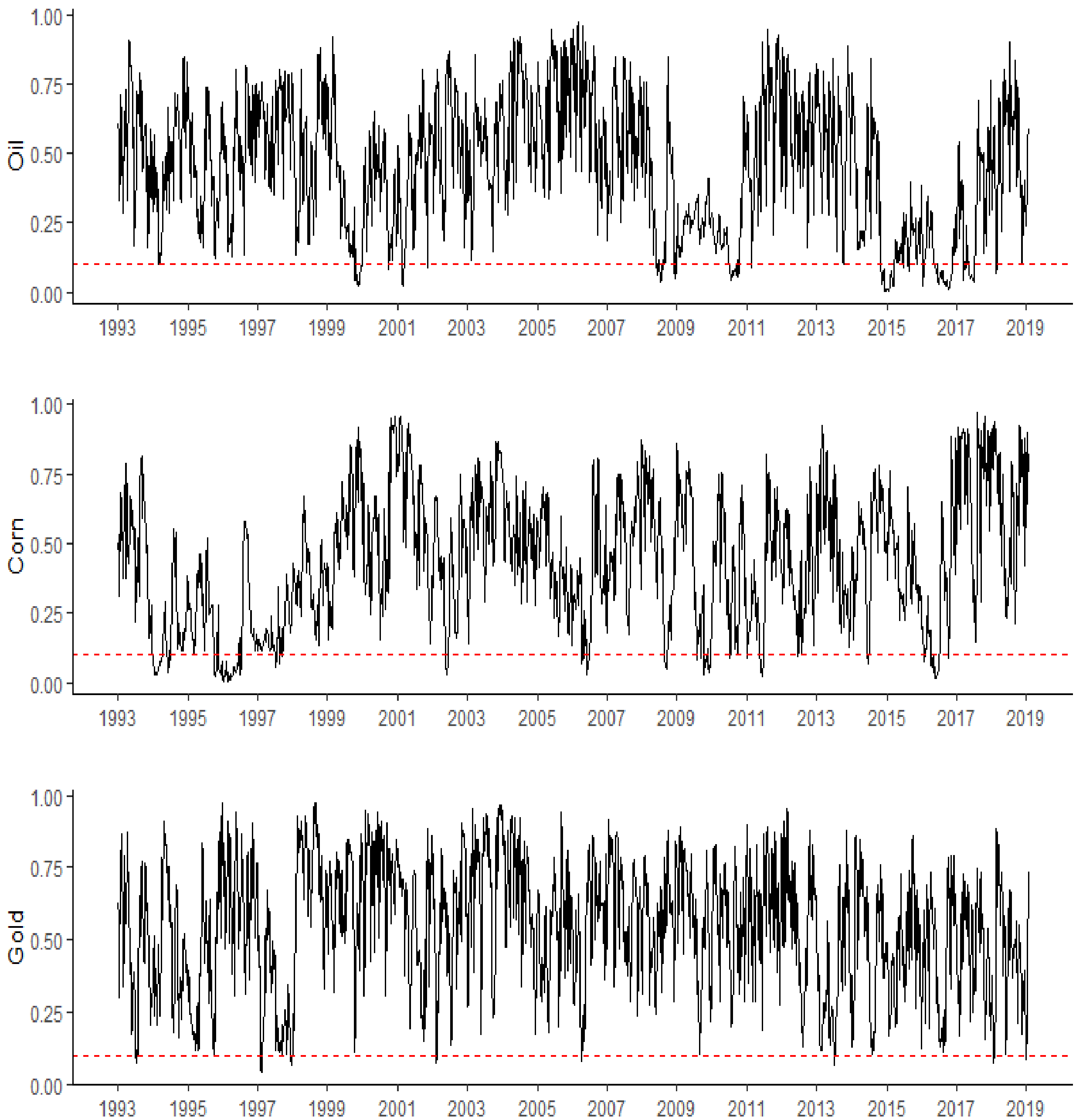

Fig.6. The GST test p-values over time based on weekly data and 105-week windows. The red horizontal line is indicating the 10% significance level.