



WP 15-04

G. Bampinas

Department of Economics, University of Macedonia, Greece

T. Panagiotidis

Department of Economics, University of Macedonia, Greece
The Rimini Centre for Economic Analysis, Italy

ON THE RELATIONSHIP BETWEEN OIL AND GOLD BEFORE AND AFTER FINANCIAL CRISIS LINEAR, NONLINEAR AND TIME-VARYING CAUSALITY TESTING

Copyright belongs to the author. Small sections of the text, not exceeding three paragraphs, can be used provided proper acknowledgement is given.

The *Rimini Centre for Economic Analysis* (RCEA) was established in March 2007. RCEA is a private, nonprofit organization dedicated to independent research in Applied and Theoretical Economics and related fields. RCEA organizes seminars and workshops, sponsors a general interest journal *The Review of Economic Analysis*, and organizes a biennial conference: *The Rimini Conference in Economics and Finance* (RCEF). The RCEA has a Canadian branch: *The Rimini Centre for Economic Analysis in Canada* (RCEA-Canada). Scientific work contributed by the RCEA Scholars is published in the RCEA Working Papers and Professional Report series.

The views expressed in this paper are those of the authors. No responsibility for them should be attributed to the Rimini Centre for Economic Analysis.

On the relationship between oil and gold before and after financial crisis: Linear, nonlinear and time-varying causality testing

G. Bampinas^a, T. Panagiotidis^{a,*}

^a*Department of Economics, University of Macedonia, Greece*

Abstract

We examine the causal relationship between crude oil and gold spot prices before and after the recent financial crisis. In the pre-crisis period, causality is linear and unidirectional, running from oil to gold. In the post-crisis period, a bidirectional nonlinear causality relationship emerges. Volatility spillover transpires as the source of nonlinearity during this period. The time path of the causal linkages both for the returns and the levels (cointegration) was assessed via dynamic bootstrap causality analysis. We find that the causal linkage from gold to oil is time dependent and that the non-Granger causality null hypothesis rejection rate increased considerably in the post-financial crisis period. The probability of gold Granger causing oil in the short-run increases by more than 30% during the recent financial and euro crisis.

Keywords: gold prices, oil prices, linear and nonlinear Granger causality, financial crisis, volatility spillovers, rolling window causality.

1. Introduction

The increasing global market integration alongside with the financialization process of commodity markets, has made commodity prices more sensitive to extreme market conditions and unforeseen events. While financial crises are not a new phenomenon (see e.g., Reinhart and Rogoff,

*Corresponding author.

Email addresses: bampinasg@uom.edu.gr (G. Bampinas), tpanag@uom.edu.gr (T. Panagiotidis)

2009), the 2007/2008 crisis differs from previous in that it is both severe and global. As a consequence, investors have been questioning previously held beliefs about the risk of equity investing and the benefits of global diversification. Intimately, the shortage of liquid financial assets in the world economy triggered a partial recreation of the interest in precious metals and energy markets (Caballero et al., 2008). Commodity markets have attracted international investor's attention not only as 'safe haven' to avoid financial risk but also as a fundamental investment strategy (Baur and McDermott, 2010).

Oil and gold are the most widely traded commodities and among the most popular economic indicators. During the recent global financial crisis, major commodity prices descend simultaneously at the aftermath of the economic downturn. In the second half of 2008, the West Texas Intermediate (WTI) crude oil price fell from a record high of \$147 to \$30 per barrel while the extent of the reduction exceeded 70% over a period of 6 months. The bankruptcy of Lehman Brothers saw the price of gold soar over 20% within a few weeks, as global risk appetite dramatically deteriorated and precipitated a 'flight to quality' across markets. The gold spot price, which is often used as a measure of storage of value, started its increase in early August 2007 from \$660 per ounce and reached its peak of over \$1000 around March 2008, after which it dropped 10% in a short time. However, ever since 2009, with the emergence of recovery expectations, the demand for commodities began to rise again and both the crude oil and gold prices started a new upward course (Figure 1).

Melvin and Sultan (1990) contend that oil price changes and political unrest are significant determinants of volatility in gold prices. Beahm (2008) supports that the price relationship between gold and oil is one of the five fundamentals that drive the prices of precious metals, particularly gold. Narayan et al. (2010) examine the long-run relationship between gold and oil spot and future prices of different maturities through the inflation channel. They conclude that the oil market can be used to predict gold prices and vice versa. Zhang and Wei (2010) tested the linear and nonlinear (in the sense of Hiemstra and Jones, 1994) relationship between the crude oil and gold markets from January 2000 to March 2008. They provide evidence of a linear interaction between the two commodities and find a significant

unilateral linear Granger causality running from crude oil to gold.

Although the two markets tend to be influenced by common factors, their prices were not completely driven by demand and supply fundamentals but rather by the financial futures and the interactions of international commodity markets. Consistent with this notion, Tang and Xiong (2010) find that as a result of the financialization process, futures prices of non energy commodities became increasingly correlated with oil after 2004. This trend intensified after the financial crisis that was triggered by the US sub-prime crisis. Since then, the importance of the interactions between the financial markets and commodity markets has risen, while the increased presence of index investors led to higher volatility in commodity prices.¹ Ciner (2001) relies on Hiemstra and Jones (1994) nonlinear causality test and provides evidence of a bidirectional feedback relation between stock index returns and oil futures markets, that was more pronounced in the 1990s. Bekiros and Diks (2008a), employ alternative econometric approaches (linear and nonlinear) and find evidence of a bidirectional causality relationship between crude oil spot and futures prices. Their empirical results show that these leads and lags patterns might change over time depending on the different methods or periods employed. Investors' decisions and portfolio rebalancing could also act as a channel to spillover shocks from other markets and across different commodities (Kyle and Xiong, 2001).

One puzzling question is how the relationship between oil and gold has evolved and how the recent crisis has affected them. In contrast to the existing literature, our causality analysis allows for linear, nonlinear and time-varying perspective.² Our contribution to the literature is twofold. First, we assess how the recent financial crisis has affected the relationship between the two most important commodities. Second, we employ linear, nonlinear and time-varying techniques to examine the causal relationship between oil and gold. In a related study, Zhang and Wei (2010) investigated the nonlinear causality relationship between crude oil and gold spot prices. They employed the nonparametric test proposed by Hiemstra and Jones

¹Tang and Xiong (2010) document that the total value of commodity exchanges via index funds for institutional investors soared from \$15 billion in 2003 to \$200 billion in mid-2008.

²Baek and Brock (1992) noted that the standard causality testing procedure that relies on linearity assumptions is inappropriate to detect nonlinear relationships.

(1994), which is a modified version of the test by Baek and Brock (1992). The Hiemstra and Jones test relaxes the strict *i.i.d* hypothesis, allowing for the existence of short-term autocorrelations in time series. However, Diks and Panchenko (2005, 2006) point out that the relationship tested by the Hiemstra and Jones (1994)'s approach is not generally compatible with the definition of Granger causality and it may lead to spurious rejections of the null hypothesis of no Granger causality. Consequently, our attention shifts to the nonlinear procedure proposed by Diks and Panchenko (2006, D&P henceforth).

Full sample and subsample analysis reveals that the bidirectional linear price transmission mechanism detected in pre-crisis period became nonlinear in the post-crisis period. Volatility spillover effects are considered as a potential source of nonlinear causal linkages. Given that linear and nonlinear causal linkages are found to be sample dependent, a dynamic rolling window causality framework is also adopted. We employ the fixed length rolling window bivariate causality test proposed by Hill (2007). Bootstrap rolling causality analysis reveals that the non-causality (gold \rightarrow oil) null hypothesis rejection frequency has increased after the financial crisis. For the opposite direction, the non-causality null (oil \rightarrow gold) has a rejection rate close to 90%, roughly stable throughout the subsamples. Gold emerges as more significant during the period of crisis, as confirmed by the probit analysis that is employed.

The remaining of the paper is organized as follows: Section 2 presents the data and methodology, Section 3 discusses the empirical results and Section 4 concludes.

2. Data and methodology

2.1 Data description

This study employs daily time series data on crude oil and gold spot prices, over the period 2003:01 to 2012:12.³ The crude oil price data are

³Following the notation of Bekiros and Diks (2008b), we have denoted the pre-financial crisis period as PI (2003:01 till 2007:07), the post-financial crisis period as PII (2007:08 till 2012:12) and the entire sample period as PIII.

obtained from the US Energy Information Agency,⁴ quoted in US dollars per barrel, while the gold price data come from the World Gold Council,⁵ quoted in US dollars per ounce (Figure 1).

2.2 Methodology

2.2.1 Linear Granger causality tests

Assume that $\{S_t, R_t, t \geq 1\}$ are two scalar-valued strictly stationary time series (i.e. $I(0)$). We can state that $\{S_t\}$ Granger causes $\{R_t\}$ if past and current values of S contain additional information on future values of R that is not contained in the past and current R_t values. If $F_{S,t}$ and $F_{R,t}$ denote the information sets consisting of past observations of S_t and R_t up to and including time t , and if \sim denote equivalence in distribution, then $\{S_t\}$ Granger causes $\{R_t\}$ if, for $p \geq 1$:

$$(R_{t+1}, \dots, R_{t+p}) | (F_{S,t}, F_{R,t}) \sim (R_{t+1}, \dots, R_{t+p}) | F_{S,t} \quad (1)$$

The value of $p = 1$ is often used, i.e. testing for Granger non-causality comes down to comparing the one-step-ahead conditional distribution of $\{R_t\}$ with and without past and current observed values of $\{S_t\}$. A conventional approach of testing for Granger causality is to assume a linear and parametric time series model for the conditional mean $E(R_{t+1} | (F_{S,t}, F_{R,t}))$. In this case, causality can be tested by comparing the residuals of a fitted autoregressive model of R_t with those obtained by regressing R_t on (infinite) past values of both $\{S_t\}$ and $\{R_t\}$ (Granger, 1969). The approach developed by Toda and Yamamoto (1995, T&Y henceforth) employs a modified Wald test restricting the parameters of the VAR (p) model in levels, where p is the lag length of the VAR. The optimal order of the VAR(p) is augmented by the maximal order of integration (d_{max}). Then one could test only the first $p - d_{max}$ coefficient matrices.⁶

⁴Available at: <http://www.eia.gov>.

⁵Available at: <http://www.gold.org/>.

⁶The novelty of this approach lies on the fact that it does not require pretesting to determine the cointegrating properties of the system hence overcoming the potential unit root and cointegration test bias.

2.2.2 Nonlinear Granger causality test

There is no evidence to suggest that economic relationships are linear. Therefore, it would make sense to consider the case of nonlinearity (for a review on nonlinear models see Ramsey, 1996). For two strictly stationary and weakly dependent variables, R_t and S_t , let \mathbf{Z}_t^κ be the κ -length lead vector of R_t , $\mathbf{S}_t^{l_s}$ the l_s -length lag vector of S_t and finally, $\mathbf{R}_t^{l_r}$ the l_r -length lag vector of R_t ($l_s, l_r \geq 1$). Assuming that the null hypothesis of no causality is a proposition about the invariant distribution of the $(l_s + l_r + \kappa)$ -dimensional vector $\mathbf{X}_t = (\mathbf{S}_t^{l_s}, \mathbf{R}_t^{l_r}, \mathbf{Z}_t^\kappa)$, the time subscript can be dropped.⁷ The joint probability density function $f_{S,R,Z}(s, r, z)$ along with its marginals, under the null should ensure:

$$\frac{f_{S,R,Z}(s, r, z)}{f_{S,R}(s, r)} = \frac{f_{R,Z}(r, z)}{f_R(r)} \quad (2)$$

The latter states that R and Z are independent conditionally on $S = s$ for each fixed value of s . D&P demonstrate that this reformulated null hypothesis implies:

$$q \equiv E [f_{S,R,Z}(S, R, Z) f_R(R) - f_{S,R}(S, R) f_{R,Z}(R, Z)] = 0 \quad (3)$$

where the proposed estimator for q is:

$$T_n(\vartheta_n) = \frac{(2\vartheta_n)^{-d_S - 2d_R - d_Z}}{n(n-1)(n-2)} \sum_i \left[\sum_{k, k \neq i} \sum_{j \neq i} (I_{ik}^{SRZ} I_{ij}^R - I_{ik}^{SR} I_{ij}^{RZ}) \right] \quad (4)$$

where, $I_{ij}^X = I(\|\mathbf{X}_i - \mathbf{X}_j\| \leq \vartheta_n)$, with $I(\cdot)$ to be the indicator function and ϑ_n the bandwidth which depends on the sample size. Hence, if we denote $\hat{f}_X(\mathbf{X}_i)$ as the local density estimator of the vector \mathbf{X} at \mathbf{X}_i , then:

$$\hat{f}_X(\mathbf{X}_i) = (2\vartheta_n)^{-d_X} (n-1)^{-1} \sum_{j, j \neq i} I_{ij}^X \quad (5)$$

⁷As a common practice, it is assumed that κ is equal to 1 and for presentation purposes we set $l_s = l_r = 1$.

The $T_n(\vartheta_n)$ statistic could be written as:

$$T_n(\vartheta_n) = \frac{(n-1)}{n(n-2)} \sum_i \left(\hat{f}_{S,R,Z}(S_i, R_i, Z_i) \hat{f}_R(R_i) - \hat{f}_{S,R}(S_i, R_i) \hat{f}_{R,Z}(R_i, Z_i) \right) \quad (6)$$

D&P demonstrate that if $\vartheta_n = Cn^{-\beta}$ with $(C > 0, 1/4 < \beta < 1/3)$, then $T_n(\vartheta_n)$ converges in distribution to the standard normal distribution:

$$\sqrt{n} \frac{(T_n(\vartheta_n) - q)}{S_n} \xrightarrow{d} N(0, 1) \quad (7)$$

where S_n is the estimated standard error of $T_n(\cdot)$. Overall, the risk of over rejecting the null is reduced substantially with the D&P test relative to the Hiemstra and Jones (1994) nonparametric approach.

2.2.3 Hill (2007) causality test

Hill (2007) establishes a sequential multi-horizon non-causality test strategy, which can be employed to characterize nonlinear causality chains for a trivariate process in terms of linear parametric restrictions. Hill's (2007) efficient causality test is based on Wald type test statistics under joint null hypothesis of zero parameter linear restrictions. It relies on a vector autoregression (VAR) framework of order p at horizon h , named (p, h) -autoregression, as the following:

$$W_{t+h} = \alpha + \sum_{k=1}^p \pi_k^{(h)} W_{t+1-k} + u_{t+h} \quad (8)$$

where W_t is a m -vector stationary process, $m \geq 2$, $\pi_k^{(h)}$ are matrix-valued coefficients, u_t is a zero mean $m \times 1$ vector white noise process with non singular covariance matrix $\Omega = E[u_t u_t']$ and α is the constant term. In this study we consider the bivariate case. Causality occurs at any horizon if and only if it occurs at horizon 1 (see Hill 2007, Theorem 2.1). We assume W_t as a 2-vector stationary process, $W_t = (S_t', R_t')$. R does not linearly causes S at 1-step ahead if and only if the RS -block $\pi_{RS,1}^{(h)} = 0$ for $k = 1$. In the case of nonstationarity for some or all variables in the VAR model, the (p, h) -

autoregression Eq. (8) is extended in the Toda and Yamamoto (1995) and Dolado and Lütkepohl (1996) model by adding d extra lags to the VAR models (discussed in 2.2.1). A straightforward Wald test of linear zero restrictions is performed in order to test for 1-step ahead non-causality. Due to possible inferior performance of the χ^2 distribution in small sample distributions, a parametric bootstrap method for simulating small sample p -values is proposed by Hill (2007, see Appendix B.2).⁸

3. Empirical results

We start our analysis by examining the stationarity properties of the series. The augmented Dickey and Fuller (1979) and the Phillips and Perron (1988) unit root tests together with the Kwiatkowski et al. (1992) stationarity test all indicate that the (log) variables are integrated of order one. Furthermore, the unit root tests allowing for one and two endogenous structural breaks of Lee and Strazicich (2003, 2004) were also conducted but no significant breaks were detected in our series for the three periods under investigation (see Tables 5, 6 and 7 in the Appendix). Therefore, (log) levels of the variables will be used in the T&Y procedure (long-run) and their first differences in the short-run Granger causality test and the D&P test. For Hill's (2007) causality test, VAR models of differences and levels (excess-lag technique following T&Y and Dolado and Lütkepohl, 1996) are both employed in order to control for cointegration of unknown form.⁹

Table 1 presents the linear Granger causality test for the two subperiods and the full sample for both the returns (short-run) and the log prices (long-run, T&Y). The evidence in returns reveals a unidirectional causality running from crude oil to gold spot returns in the pre-financial crisis period. Evidence in favor of bidirectional causality is found in the post-financial crisis period and the entire sample period. These results are further reinforced by the T&Y long-run approach (log prices).

⁸The rolling window regressions adopted in our empirical analysis are based on 5000 bootstrap replications. The GAUSS code is available at: <http://www.unc.edu/~jbhill/software.htm>

⁹Johansen (1995) cointegration test indicates that the crude oil and the gold prices do not cointegrate over the full sample as well as the two subperiods (see Table 8 in the Appendix).

Papana et al. (2013) perform a simulation study of different direct causality measures and stress the importance of investigating linear as well as nonlinear linkages, in order to confirm the existence of causal effects. Accordingly, our next step would be to relax the assumption of linearity. The D&P testing procedure that allows for nonlinearity is carried out for three cases. First, the test is implemented to the stationary (returns) series for nonlinear causality. In the second step, the D&P test is applied on the de-linearized series.¹⁰ Lastly, the procedure is re-applied in the GARCH-BEKK filtered VAR residuals in order to examine volatility spillover effects.¹¹ In a similar vein with Bekiros and Diks (2008a, b) the bandwidth value for the D&P test is set equal to one.

Table 2 presents the D&P test results. In the pre-crisis period, the results fail to reject the null of the nonlinear causality relationship between oil and gold returns (PI in Panel A and B). We cannot reject the non-causality null neither on the raw data nor on the VAR residual series. Given that no evidence of nonlinear dependence emerges in the pre-crisis period (Table 2), we can argue that in PI there is a significant and persistent linear unidirectional causality running from oil to gold (Table 1). In PII (post-crisis period), unfiltered (raw) returns indicate a bidirectional nonlinear causality between oil and gold. The nonlinear causality for the filtered VAR-residuals confirms the results for the raw returns and offers support for the existence of a nonlinear price transmission. The results presented in Table 2 (Panel A and B) reveal a significant two-way nonlinear causal relationship for both the raw data and the VAR residuals, in the post-financial crisis period. In PIII (full sample), the results for raw data indicate a (rather weak) one-way nonlinear causality running from gold to oil returns. This causal linkage seems to be nonlinear in nature since it tends to be stronger in the case of the VAR residuals. Generally, weak evidence is found for the nonlinear causal linkages running from gold to oil returns in PIII (both raw and filtered data).

¹⁰By removing linear predictive power with a VAR model, any causal linkage from one residual series of the VAR model to another can be considered as nonlinear predictive power (Hiemstra and Jones 1994, p. 1648). The lag length of the VAR specification was based on the Akaike and Schwartz information criteria.

¹¹Following Bekiros and Diks (2008a, b), the second-moment filtering has been conducted on the data through a bivariate GARCH-BEKK (1, 1) model.

Next, we employ second moment filtering in the three periods under investigation.¹² Once GARCH-BEKK(1,1) filtering is employed, the nonlinear causality evidence disappears (Table 2). The bidirectional causal relationship has vanished in PII period and the same applies for the unidirectional linkage (gold \rightarrow oil) in PIII. The latter indicates that the nonlinear causality is due to volatility effects. Overall, it seems that the additional explanatory power of volatility spillover is important for the oil and gold markets, especially after the financial crisis period.

The evidence from the previous section highlights the different behavior in the vicinity of recent crisis. However, the dating of the subsamples was imposed and the actual effect could have started before or after these dates. Therefore, one needs to study the evolving patterns of causality over rolling fixed length sample periods. In order to control for any apparent trend, we pass the series through a linear filter. The window length is set equal to 522 days (close to 2 years of daily data), generating a total of 2086 windows. We employ VAR models in differences (returns) and log levels (with excess lags) where the optimal order is selected by minimizing the AIC ($p = 1, \dots, 30$ is considered). We perform bootstrap tests of non-causality for each window and count the non-causality null hypothesis rejection rate for both VAR models in differences and levels. Tests of the null hypothesis that oil does not cause gold and vice versa, are performed at the 5% level. The upper bound of the size tests of $(\Delta)\text{oil} \rightarrow (\Delta)\text{gold}$ and $(\Delta)\text{gold} \rightarrow (\Delta)\text{oil}$ is 0.1 ($0.1 \times h$). Table 3 presents the rolling windows bootstrap p -values.¹³

Allowing the fixed window to move through our sample, we find evidence of (strong) causality running from oil to gold spot returns for all the sample periods under consideration (Panel A in Table 3). The latter is not valid in the opposite causality direction (gold \rightarrow oil). The most prominent characteristic is the significant increase in the number of windows providing evidence of causality from gold to oil in the post-financial crisis period.

¹²Pavlidis et al. (2013) highlight the perils of neglecting multivariate conditionally heteroskedasticity when testing for nonlinear causality in mean. The authors examine the stock return-volume causal relationship for the US, UK and Japan and find that in several cases the test statistics become insignificant when heteroskedasticity robust tests are employed.

¹³The results in Table 3 assume $k = 1$. For $k = 2$ up to 6 the results are also available upon request.

Causality from gold to oil that takes place in pre-crisis period almost doubled in post-crisis period, from 22.5% to 39%. Profound causality running from oil to gold (99%) is found during the NBER recession period. Oil and gold rolling window bootstrap p -values are presented in Figure 2. More close inspection of Figure 2, though, reveals an interesting pattern. The gray line (gold \rightarrow oil) decreases substantially in 2007 compared to 2006. The next two years 2008 and 2009 are characterized by increased volatility in the p -values. The euro crisis that evolved from 2009 to mid-2012 has again increased the significance of gold, with numerous p -values (gray line) laying under the 5% line. Overall, the rejection rates of the gold \rightarrow oil null are clustered around 2007 (financial crisis, Lehman Brothers bankruptcy) and after 2009 when the euro crisis emerged (sovereign debt crisis). The black line (oil \rightarrow gold) remains below the 5% line for the whole period with only few exceptions.

Once we control for cointegration a slightly different picture emerges (Panel B in Table 3). In this case, we examine the non-causality null hypothesis in the levels (log prices) of the variables. The strong causality running from oil to gold found in the short-run (PI in Panel A), decreases to 50.52% in the long-run (PI in Panel B). This causal linkage remained strong in PII (94.34%) and also during the NBER recession period (100%). The causal relationship running from gold to oil increased only marginally in the post-crisis period, compared to the pre-crisis period (26.3% and 25.4% respectively). The rejection rate of the null hypothesis that gold \rightarrow oil for PI, PII and PIII appears stable around 26%. Figure 3 plots the p -values for the VAR in levels. Again, we observe that the gray line (gold \rightarrow oil) falls below the 5% line in 2007 (around financial crisis) and in 2011 (euro debt crisis). The black line remains below the 5% line from 2007 to the end of the sample period with few exceptions in 2010 and 2012. The overall conclusion from the bootstrap rolling causality analysis is that for both cases (returns in Figure 2 and levels in Figure 3) the significance of gold emerges in the specific crisis periods.

The rejection rates differ in periods close to financial crisis of 2007 and after 2009 when the euro crisis emerged. The visual inspection of Figures 2 and 3 is informative but insufficient to draw conclusions. Therefore, we explore how the recent turbulent and crises periods (financial turmoil, financial crisis and the sovereign debt-euro crisis) affect the causal linkage from

gold to oil, by employing a probit regression model:

$$P_i = p(y_i = 1) = \beta_0 + \beta_1 D_1 + \beta_2 D_2 + \beta_3 D_3 + u_i, \quad i = 1, \dots, N \quad (9)$$

The dependent variable, y_i takes the value of 0 if we cannot reject the non-causality null hypothesis (gold \rightarrow oil) and 1 otherwise. D_1 takes the value of 1 during the period from the peak of the credit boom (second quarter of 2007) until the collapse of Lehman Brothers in September 2008, D_2 takes the value of 1 in the peak of financial crisis (fourth quarter of 2008) and D_3 takes the value of 1 on specific dates related to the euro crisis.¹⁴ The goal is to quantify the relationship between the different crises periods and the probability of rejecting the gold \rightarrow oil null hypothesis. Table 4 presents the marginal effects for Eq. (9).¹⁵ When the causal linkage is examined in returns (Panel A), we observe that a unit change in financial crisis and the euro crisis dummy variables, increase the probability of rejecting the gold \rightarrow oil null hypothesis by 0.35 and 0.30 respectively. Similarly, being in the financial turmoil period decreases the probability of rejecting the null hypothesis (gold \rightarrow oil) by 0.26. When cointegration is taken into account (Panel B), the probability of rejecting the null hypothesis increases in the financial turmoil and euro debt crisis periods by 0.12 and 0.06 respectively. In the peak of the recent financial crisis this probability decreases by 0.16.

In general, the marginal effects are higher in Panel A reflecting the fact that the crisis dummy variables affect the short-run more than the long-run. In the short-run, both the financial crisis and the euro crisis increase the probability of no rejection (gold \rightarrow oil). The latter confirms the increased role of gold during these turbulent periods. In the long-run, the effect of financial crisis is not significant at the 1% level. Conversely, the financial turmoil and

¹⁴We construct a timeline of the Euro crisis by combining two external sources: the Financial Times (Interactive Timeline: Greek Debt Crisis), the Wall Street Journal (Europe's Debt Crisis - Timeline), in conjunction with specific events (e.g. increase in sovereign CDS spreads, bailout agreements, political turmoil) occurred in the GIIPS (Greece, Ireland, Italy, Portugal, Spain). The compiled joint timeline covers a period starting October 2009 (Greek elections) and ending July 2012.

¹⁵Logit models provide qualitatively similar results and are available from the authors upon request.

the euro crisis increase the probability of gold→oil rejection count.

4. Conclusions

Oil and gold are the most important commodities. This study examines the causal relationship between the two during the recent financial crisis. The linear and nonlinear Granger causality tests suggest that crude oil and gold markets have become more interrelated after the 2007 turbulence in the financial markets. The linear causal relationship detected on returns disappears after linear filtering. Nonlinear causal linkages in post-financial crisis period are revealed but vanished after GARCH filtering both for the post-financial crisis and the entire sample period. Evidence emerges that the nonlinear linkages between the oil and gold markets can be attributed to volatility spillover effects.

The dynamic analysis enlightens us further on how this relationship evolved. Significant changes in the causality test for the two commodities at the turn of the recent financial crisis demonstrate that their interdependence evolves as economic conditions change. A stable relationship is evident for crude oil which consistently influences gold during all fixed rolling subsamples. The reverse does not hold, as it appears that gold did cause oil mostly during the crisis. The latter justifies the rising importance of gold during volatile periods. This is further confirmed in the case where we allow for cointegration. Probit analysis revealed that during the crisis the probability of rejecting the non-Granger causality null (gold→oil) increased more than 30% in the short-run and more than 6% in the long-run.

Acknowledgments

We would like to thank the Associate Editor and an anonymous referee for their valuable comments and suggestions. The usual disclaimer applies.

References

- [1] Baek, E. and Brock, W. (1992) A general test for non-linear Granger causality: Bivariate model, Working paper, Iowa State University and University of Wisconsin, Madison, WI.
- [2] Baur, D.G. and McDermott, T.K. (2010) Is gold a safe haven? International evidence, *Journal of Banking and Finance*, 34, 1886-1898.
- [3] Beahm, D. (2008) 5 factors Driving Gold in 2008. Available at:http://www.blanchardonline.com/beru/five_factors.php
- [4] Bekiros, S. and Diks, C. (2008a) The relationship between crude oil spot and futures Prices: Cointegration, linear and nonlinear causality, *Energy Economics*, 30, 2673-2685.
- [5] Bekiros, S. and Diks, C. (2008b) The nonlinear dynamic relationship of exchange rates: Parametric and nonparametric causality testing, *Journal of Macroeconomics*, 30, 1641-1650.
- [6] Caballero, J.R., Farhi, E. and Gourinchas, P-O. (2008) Financial crash, commodity prices, and global imbalances, *Brookings Papers on Economic Activity*, 2, 1-55.
- [7] Ciner, C. (2001) Energy Shocks and Financial Markets: Nonlinear Linkages, *Studies in Nonlinear Dynamics and Econometrics*, 5, 203-212.
- [8] Dickey, D.A. and Fuller, W.A. (1979) Distribution of the estimators for autoregressive time series with a unit root, *Journal of the American Statistical Association*, 74, 427-431.
- [9] Diks, C. and Panchenko, V. (2005) A note on the Hiemstra-Jones test for Granger noncausality, *Studies in Nonlinear Dynamics and Econometrics*, 9 (art. 4).
- [10] Diks, C. and Panchenko, V. (2006) A new statistic and practical guidelines for nonparametric Granger causality testing, *Journal of Economic Dynamics & Control*, 30, 1647-1669.
- [11] Dolado, J. and Lütkepohl, H. (1996) Making Wald tests work for cointegrated VAR systems, *Econometric Reviews*, 15, 369-386.
- [12] Granger, C.W.J. (1969) Investigating causal relations by econometric models and cross-spectral methods, *Econometrica*, 37, 424-438.
- [13] Hiemstra, C. and Jones, J.D. (1994) Testing for linear and nonlinear Granger causality in the stock price-volume relation, *Journal of Finance*, 49, 1639-1664.

- [14] Hill, J.B. (2007) Efficient Tests of Long-Run Causation in Trivariate VAR Processes with a Rolling Window Study of the Money-Income Relationship, *Journal of Applied Econometrics*, 22, 747-765.
- [15] Johansen, S. (1995) Likelihood-Based Inference in Cointegrated Vector Autoregressive Models, Oxford University Press, Oxford.
- [16] Kyle, A. and Xiong, W. (2001) Contagion as a wealth effect, *Journal of Finance*, 56, 1401-1440.
- [17] Kwiatkowski, D., Phillips, P.C.B., Schmidt, P. and Shin, Y. (1992) Testing the null hypothesis of stationarity against the alternative of a unit root: how sure are we that economic time series have a unit root?, *Journal of Econometrics*, 54, 159-178.
- [18] Lee, J. and Strazicich, M.C. (2003) Minimum Lagrange Multiplier unit root tests with two structural breaks, *Review of Economics and Statistics*, 81, 1082-1089.
- [19] Lee, J. and Strazicich, M.C. (2004) Minimum LM Unit Root Test with One Structural Break, Department of Economics, Appalachian State University, *Working Paper Series*.
- [20] MacKinnon, J.G (1996) Numerical Distribution Functions for Unit Root and Cointegration Tests, *Journal of Applied Econometrics*, 11, 601-618.
- [21] Melvin, M. and Sultan, J. (1990) South African political unrest, oil prices, and the time varying risk premium in the gold futures market, *Journal of Futures Markets*, 10, 103-111.
- [22] Narayan, P.K., Narayan, S. and Zheng, X. (2010) Gold and oil futures markets: Are markets efficient?, *Applied Energy*, 87, 3299-3303.
- [23] Papan A., Kyrtsov C., Kugiumtzis D. and Diks, C. (2013) Simulation Study of Direct Causality Measures in Multivariate Time Series, *Entropy*, 15, 2635-2661.
- [24] Pavlidis, E., Paya, I. and Peel, D. (2013) Nonlinear causality tests and multivariate conditional heteroskedasticity: a simulation study, *Studies in Nonlinear Dynamics and Econometrics*, 17, 297-312.
- [25] Phillips, P.C.B. and Perron, P. (1988) Testing for a unit root in time series regression, *Biometrika*, 75, 335-346.
- [26] Ramsey, J.B. (1996) If Nonlinear Models Cannot Forecast, What Use Are They?, *Studies in Nonlinear Dynamics and Econometrics*, 1, 65-86.

- [27] Reinhart, C.M. and Rogoff, K.S. (2009) *This Time is Different: Six Centuries of Financial Folly*, Princeton University Press, Princeton, NJ.
- [28] Tang, K. and Xiong, W. (2010) Index investment and the financialization of commodities, *NBER Working Paper w16385*, National Bureau of Economic Research, Cambridge, MA. Princeton, NJ.
- [29] Toda, H.Y. and Yamamoto, T. (1995) Statistical inference in vector autoregressions with possibly integrated processes, *Journal of Econometrics*, 66, 225-250.
- [30] Zhang, Y.J. and Wei, Y.M. (2010) The crude oil market and the gold market: Evidence for cointegration, causality and price discovery, *Resources Policy*, 35, 168-177.

Figures

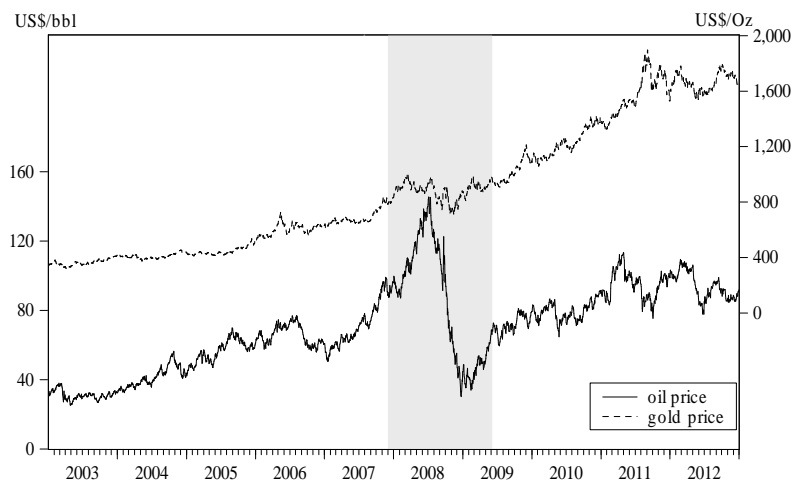


Fig.1. Crude oil and gold spot prices (shaded areas represent NBER recession dates)

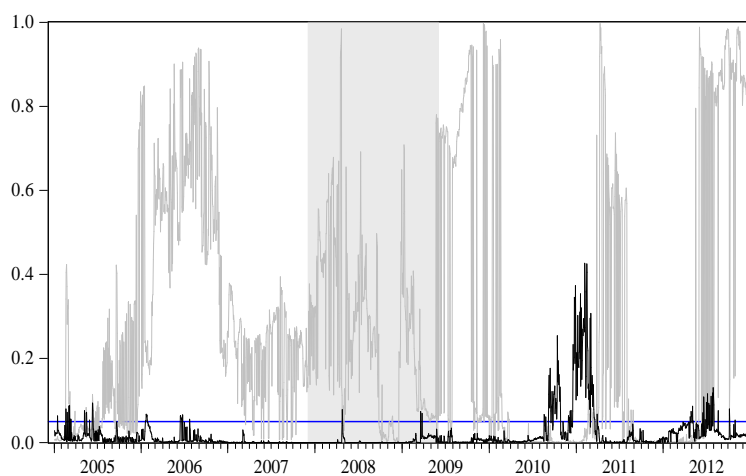


Fig.2. $Oil \rightarrow Gold$ (black line) and $Gold \rightarrow Oil$ (gray line) bootstrap p -values for VAR in first differences (shaded areas represent NBER recession dates). The blue horizontal line denotes the 5% significance level.

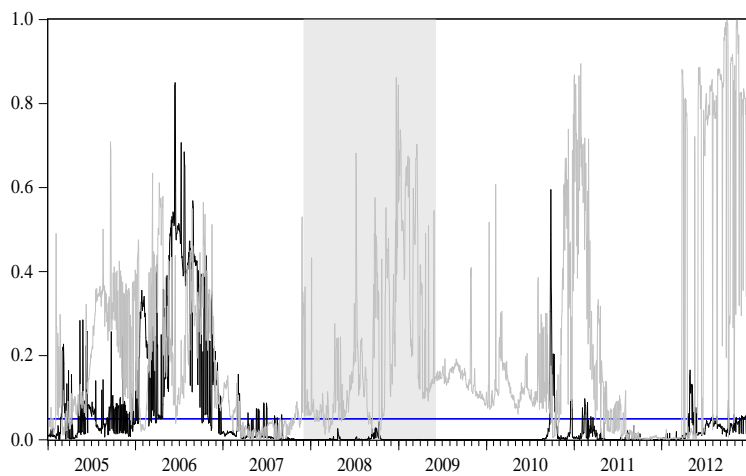


Fig.3. *Oil* \rightarrow *Gold* (black line) and *Gold* \rightarrow *Oil* (gray line) bootstrap p -values for VAR in levels (shaded areas represent NBER recession dates). The blue horizontal line denotes the 5% significance level.

Table 1. Linear Causality tests

	Linear Granger Causality (returns)			T&Y (log prices)		
	PI	PII	PIII	PI	PII	PIII
Oil \rightarrow Gold	20.68 ^a (0.00)	43.25 ^a (0.00)	71.06 ^a (0.00)	20.08 ^a (0.00)	53.94 ^a (0.00)	71.29 ^a (0.00)
Gold \rightarrow Oil	0.918 (0.33)	29.88 ^a (0.00)	26.67 ^a (0.02)	7.03 ^c (0.07)	38.11 ^a (0.00)	25.44 ^b (0.04)

Notes: \rightarrow denotes the non-causality null hypothesis. T&Y denotes the Toda and Yamamoto (1995) approach. The AIC was used to determine the optimal lag lengths for VAR(p) models. Numbers in parenthesis are the corresponding p -values. ^c, ^b, and ^a represent significance at 10%, 5%, and 1% level, respectively. We denote the pre-financial crisis period as PI (2003:01 till 2007:07), the post-financial crisis period as PII (2007:08 till 2012:12) and the entire sample period as PIII.

Table 2. Nonlinear Granger causality test

Lag	Raw data (returns)			VAR filtered series			GARCH-BEKK filtered data		
	PI	PII	PIII	PI	PII	PIII	PI	PII	PIII
<i>Panel A: Oil\rightarrowGold</i>									
1	2.6 ^a (0.00)	4.12 ^a (0.00)	4.2 ^a (0.00)	1.67 ^b (0.04)	3.98 ^a (0.00)	3.36 ^a (0.00)	1.46 ^c (0.07)	1.12 (0.13)	1.52 ^c (0.06)
2	0.29 (0.38)	2.48 ^a (0.00)	1.21 (0.11)	-0.13 (0.55)	3.06 ^a (0.00)	1.39 ^c (0.08)	-0.31 (0.62)	0.51 (0.3)	0.23 (0.4)
3	-0.64 (0.74)	2.27 ^a (0.01)	0.33 (0.36)	-0.73 (0.76)	1.74 ^b (0.04)	0.02 (0.49)	-1.08 (0.86)	-0.74 (0.77)	-1.11 (0.86)
4	-0.57 (0.71)	1.59 ^b (0.05)	0.24 (0.40)	-0.28 (0.61)	1.24 ^c (0.1)	0.05 (0.47)	-0.17 (0.57)	0.93 (0.17)	-0.66 (0.74)
<i>Panel B: Gold\rightarrowOil</i>									
1	-0.72 (0.76)	2.9 ^a (0.00)	1.94 ^b (0.02)	-1.15 (0.87)	2.49 ^a (0.00)	1.44 ^c (0.07)	-0.43 (0.66)	-0.04 (0.51)	0.07 (0.47)
2	-0.43 (0.66)	2.07 ^a (0.01)	1.18 (0.11)	-0.48 (0.68)	2.08 ^a (0.01)	1.33 ^c (0.09)	-0.54 (0.29)	-0.3 (0.61)	0.08 (0.46)
3	0.30 (0.38)	1.77 ^b (0.03)	1.52 ^c (0.06)	0.51 (0.3)	1.47 ^c (0.07)	2.03 ^b (0.02)	1.05 (0.14)	-0.54 (0.7)	0.84 (0.19)
4	0.28 (0.38)	1.43 ^c (0.07)	1.27 ^c (0.1)	0.30 (0.37)	1.96 ^b (0.02)	2.06 ^a (0.01)	1.17 (0.11)	0.76 (0.22)	0.45 (0.32)

Notes: This table reports the results of nonlinear causality test between oil and gold returns. \rightarrow denotes the non-causality null hypothesis. We set the lag length $l^S=l^R=1$ to 4. Numbers in parenthesis are the corresponding p -values. ^c, ^b, and ^a represent significance at 10%, 5%, and 1% level, respectively. PI: 01/2003-07/2001, PII: 08/2007-12/2012, PIII:01/2003-12/2012.

Table 3. Rolling windows rejection rates

	PI	PII	PIII	NBER recession
Date	12:2004-7:2007	8:2007-12:2012	12:2004-12:2012	12:2007-06:2009
<i>Panel A: Returns (log differences)</i>				
Oil→Gold	95%(638)	89%(1258)	91%(1896)	99%(389)
Gold→Oil	22.5%(153)	39%(555)	34%(708)	15%(58)
Total Obs.	672	1414	2086	390
<i>Panel B: Levels (log prices)</i>				
Oil→Gold	50.52%(340)	94.34%(1334)	80.21%(1674)	100%(390)
Gold→Oil	25.4%(171)	26.3%(372)	26.01%(543)	11.79%(46)
Total Obs.	673	1414	2087	390

Notes: → denotes the non-causality null hypothesis. NBER recession denotes the US business cycle contraction period. Rejection counts at the 5% level are in parenthesis. PI: 01/2003-07/2001, PII: 08/2007-12/2012, PIII:01/2003-12/2012.

Table 4. Probit regression coefficients marginal effects

	Marginal Effect	Std. Error	z	P> z
<i>Panel A: Returns (log differences)</i>				
financial turmoil (D_1)	-0.26 ^a	0.03	-6.88	0.00
financial crisis(D_2)	0.35 ^a	0.04	7.91	0.00
euro crisis(D_3)	0.30 ^a	0.01	19.05	0.00
<i>Panel B: Levels (log prices)</i>				
financial turmoil (D_1)	0.12 ^a	0.02	4.58	0.00
financial crisis(D_2)	-0.16 ^b	0.06	-2.56	0.01
euro crisis(D_3)	0.06 ^a	0.02	3.10	0.00

Notes: ^c, ^b, and ^a represent significance at 10%, 5%, and 1% level, respectively. The dependent variable takes the value of 0 if we cannot reject the non-causality null (gold→oil) and 1 otherwise.

Appendix

Table 5. Unit root and stationarity tests

Variables		ADF			PP			KPSS		
		PI	PII	PIII	PI	PII	PIII	PI	PII	PIII
		<i>t</i> -stat.			<i>t</i> -stat.			<i>t</i> -stat.		
Level										
Oil	c	-0.93	-1.76	-1.83	-1.02	-1.86	-1.93	4.04 ^a	0.57 ^b	3.97 ^a
	c, t	-2.77	-1.78	-2.25	-2.86	-1.91	-2.49	0.6 ^a	0.35 ^a	0.53 ^a
Gold	c	-0.63	-1.56	-0.37	-0.66	-1.55	-0.67	4.21 ^a	4.49 ^a	6.23 ^a
	c, t	-2.99	-2.99	-3.89 ^b	-2.75	-2.96	-3.96 ^a	0.5 ^a	0.29 ^a	0.18 ^b
1st differences										
Oil	c	-38.11 ^a	-15.82 ^a	-20.09 ^a	-38.05 ^a	-37.95 ^a	-53.38 ^a	0.03	0.07	0.06
	c, t	-38.09 ^a	-15.81 ^a	-20.1 ^a	-38.04 ^a	-37.93 ^a	-53.38 ^a	0.03	0.07	0.03
Gold	c	-10.98 ^a	-37.64 ^a	-11.79 ^a	-33.61 ^a	-37.69 ^a	-50.7 ^a	0.04	0.08	0.03
	c, t	-10.98 ^a	-37.64 ^a	-11.79 ^a	-33.6 ^a	-37.70 ^a	-50.69 ^a	0.03	0.03	0.02

Notes: The unit root and stationarity tests are applied with (c,t) and without (c) a time trend. Null hypothesis for the KPSS test is stationarity. The critical values for ADF and Phillips-Perron (PP) statistics are taken from MacKinnon (1996). Superscripts ^c, ^b, and ^a represent significance at 10%, 5%, and 1% level, respectively. Lag lengths are determined via AIC. PP was conducted using Bartlett Kernel (Newey-West Automatic). PI: 01/2003-07/2001, PII: 08/2007-12/2012, PIII:01/2003-12/2012.

Table 6. Lee and Strazicich (2004) unit root test with one break

		PI		PII		PIII	
		<i>t</i> -stat	T_B	<i>t</i> -stat	T_B	<i>t</i> -stat	T_B
Oil	Model A	-2.89[5]	5/21/04	-2.01[12]	1/6/09	-2.5[7]	1/6/09
	Model C	-3.75[1]	8/30/05	-2.64[12]	11/12/08	-2.85[7]	3/31/05
Gold	Model A	-2.95[9]	2/16/06	-3.17[11]	9/10/08	-3.05[11]	3/18/09
	Model C	-4.24[9] ^c	12/27/05	-3.29[11]	9/10/08	-4.43[11] ^c	11/14/05

Notes: The tests critical values are obtained from Lee and Strazicich, (2004, table 1). T_B denotes the endogenously determined break points. Numbers in brackets denote the optimal number of lagged first-differenced terms included to correct for serial correlation. PI: 01/2003-07/2001, PII: 08/2007-12/2012, PIII:01/2003-12/2012.

Table 7. Lee and Strazicich (2003) unit root test with two breaks

		PI			PII			PIII		
		<i>t</i> -stat	1st T_B	2nd T_B	<i>t</i> -stat	1st T_B	2nd T_B	<i>t</i> -stat	1st T_B	2nd T_B
Oil	Model A	-3.12[5]	5/21/04	3/25/05	-2.08[12]	1/6/09	5/25/10	-2.6[7]	3/23/05	9/26/08
	Model C	-4.91[1]	12/4/03	9/7/06	-4.00[12]	10/2/08	6/2/09	-3.87[7]	9/26/08	10/19/09
Gold	Model A	-3.09[9]	1/2/06	7/26/06	-3.38[11]	9/10/08	4/3/12	-3.21[9]	5/29/06	8/23/11
	Model C	-5.23[9]	11/4/05	6/8/06	-5.57[9] ^c	7/30/08	8/23/11	-4.78[9]	7/30/08	8/23/09

Notes: The tests critical values are obtained from Lee and Strazicich (2003, table 2). T_B denotes the endogenously determined break points. Numbers in brackets denote the optimal number of lagged first-differenced terms. PI: 01/2003-07/2001, PII: 08/2007-12/2012, PIII:01/2003-12/2012.

Table 8. Johansen trace test

<i>r</i>	Model 2			Model 3			Model 4		
	PI	PII	PIII	PI	PII	PIII	PI	PII	PIII
0	9.04	10.6	13.40	5.2	7.14	6.51	17.19	23.05	23.60
	(0.73)	(0.58)	(0.33)	(0.78)	(0.56)	(0.63)	(0.4)	(0.1)	(0.09)
1	4.02	3.52	4.91	0.27	2.5	0.27	4.81	4.35	5.85
	(0.4)	(0.48)	(0.29)	(0.6)	(0.11)	(0.59)	(0.62)	(0.69)	(0.47)

Notes: Figures in the parenthesis are *p*-values. The AIC was used to determine the optimal lag length. Model 2 includes intercept in the cointegration relation, Model 3 includes deterministic trends in level and Model 4 allows for trend in the cointegrating space. PI: 01/2003-07/2001, PII: 08/2007-12/2012, PIII:01/2003-12/2012.