



# Long memory and structural breaks in modeling the return and volatility dynamics of precious metals

Mohamed El Hedi Arouri, Shawkat Hammoudeh, Amine Lahiani, Duc Khuong Nguyen

## ► To cite this version:

Mohamed El Hedi Arouri, Shawkat Hammoudeh, Amine Lahiani, Duc Khuong Nguyen. Long memory and structural breaks in modeling the return and volatility dynamics of precious metals. 2013. hal-00798033

**HAL Id: hal-00798033**

**<https://hal.science/hal-00798033>**

Preprint submitted on 7 Mar 2013

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Long memory and structural breaks in modeling the return and volatility dynamics of precious metals

Mohamed El Hedi Aroui

*CRCGM, University of Auvergne*

*41 Bld François Mitterrand, 63002 Clermont-Ferrand, France*

&

*EDHEC Business School*

*12 bis, rue de la Victoire, 75009 Paris, France*

Email: [Mohamed.AROURI@edhec.edu](mailto:Mohamed.AROURI@edhec.edu)

Shawkat Hammoudeh

*Lebow College of Business, Drexel University*

*3141 Chestnut Street, Philadelphia, PA 19104, USA*

Email: [hammousm@drexel.edu](mailto:hammousm@drexel.edu)

Amine Lahiani

*LEO, University of Orléans*

*Rue de Blois, BP 267-39, 45067 Orléans cedex 2, France*

&

*ESC Rennes School of Business*

Email: [amine.lahiani@univ-orleans.fr](mailto:amine.lahiani@univ-orleans.fr)

Duc Khuong Nguyen<sup>\*</sup>

*ISC Paris School of Management, France*

*22, Boulevard du Fort de Vaux, 75017 Paris, France*

Email : [dnguyen@iscparis.com](mailto:dnguyen@iscparis.com)

Phone: +33 1 40 53 99 99 | Fax: +33 1 40 53 98 98

<sup>\*</sup> Corresponding author

## Abstract

We investigate the potential of structural changes and long memory (LM) properties in returns and volatility of the four major precious metal commodities traded on the COMEX markets (gold, silver, platinum and palladium). Broadly speaking, a random variable is said to exhibit long memory behavior if its autocorrelation function is not integrable, while structural changes can induce sudden and significant shifts in the time-series behavior of that variable. The results from implementing several parametric and semiparametric methods indicate strong evidence of long range dependence in the daily conditional return and volatility processes for the precious metals. Moreover, for most of the precious metals considered, this dual long memory is found to be adequately captured by an ARFIMA-FIGARCH model, which also provides better out-of-sample forecast accuracy than several popular volatility models. Finally, evidence shows that conditional volatility of precious metals is better explained by long memory than by structural breaks.

**Keywords:** precious metal prices, long memory, structural breaks, ARFIMA-FIGARCH

**JEL classification:** Q47, O13, C22

**Acknowledgement:** we would like to thank two anonymous referees and Editor-in-Chief Hadi Salehi Esfahani for their invaluable and helpful comments. All remaining errors are ours.

## 1. Introduction

Over the last few decades, international financial markets have experienced a succession of serious crisis of different causes and origins. The 1987 stock market crash originated in the United States and affected the world's equity markets. The 1997-1998 Asian crisis started in South Asian economies as a result of short-term capital flows and then spread to other emerging equity and commodity markets. The 2001 U.S. recession was caused by the collapse of the dot com stocks and triggered a push toward greater bank liquidity. Finally, the 2007-2010 global financial crisis which originated in the United States was sparked by the subprime real estate crisis, and then turned into a world financial crisis. Most of these crises are characterized by high volatility and contagion (Forbes and Rigobon, 2002; Lee et al., 2007; Markwat et al., 2009). Moreover, recent studies suggest that these crises stoked greater correlations between the world's equity markets, in particular in periods of high and extreme volatility, and thus lowered the diversification benefit potential from investing in traditional stocks (Chan-Lau et al., 2004; Diamandis, 2009).

The highly volatility and widespread contagion have prompted investors to consider alternative investment instruments as a part of diversified portfolios in order to be used as a hedge to diversify away the increasing risk in the stock markets. Oil and major precious metals including gold, palladium, platinum and silver thus emerged as natural desirable asset classes eligible for portfolio diversification. They offer different volatilities and returns of lower correlations with stocks, at both sector and market levels (Arouri and Nguyen, 2010; Daskalaki and Skiadopoulos, 2011; Arouri et al. 2010,2011,2012). It should be noted that when risk aversion mounts, in particular when the stock markets experience signs of instability or when the price of oil exhibit long swings because of economic uncertainties and geopolitical tensions, the majority of investors is directed towards the precious metals, being viewed as the refugee or safe haven asset in time of crises. Meanwhile, we observe severe speculations on the prices of these precious metals and high elasticity of substitution among them in both consumption and inputs, given the recent increase in their economic uses in the jewelry, electronic and auto industries. Investigating the price dynamics of precious metals is, therefore, of great interest to investors, traders and policy makers.

A large volume of literature deals with oil and other energy price dynamics. These studies have shown significant spillover effects between different commodity prices as well as nonlinearities, asymmetries and other distributional characteristics such as time-varying conditional moments, volatility clustering and long-persistence of commodity price returns (Sa-

dorsky, 2006; Agnolucci, 2009; Akram, 2009; Lescaroux, 2009; Browne and Cronin, 2010). However, only a few attempts have studied the dynamics and distributional characteristics of precious metal prices. So far, modeling volatility properties of precious metals is still of major interest in the financial economics literature as volatility forecast is an important input for asset valuations, hedging, and risk management. One should note that long memory (LM) and structural breaks are at the heart of the debate regarding volatility modeling. While persistence in volatility models deals with exponential decays in the autocorrelation of conditional variance, long memory in volatility processes requires models accommodating volatility persistence over long horizons. But, a presence of structural breaks may reduce the persistence of volatility and hinder the prediction process.

In this article, we extend the existing literature on the dynamics of precious metals prices by examining the relevance of structural breaks and long memory in modeling the conditional returns and volatilities for four major precious metals (gold, silver, palladium, and platinum) traded on the commodity exchange (COMEX) of the New York Mercantile Exchange. Empirically, three long memory tests are implemented to examine the long-range dependence in the conditional mean and variance processes of these precious metals, while a modified version of Inclan and Tiao (1994)'s iterated cumulative sum of squares (ICSS) algorithm is used to detect structural changes in the precious metals time series data. Our results show that long memory is an important empirical feature for the precious metal series, and that the conclusions do not change when potential structural breaks are controlled for. In six out of the eight cases, we find significant evidence that the double long memory and the ARFIMA-FIGARCH class models are more suitable to describe the time-variations in the return and volatility of precious metals. The out-of-sample analysis indicates that the ARFIMA-FIGARCH class models provide more accurate volatility forecasts in most cases than other competing GARCH-based models.

Our research thus constitutes a good venue for understanding the distributional characteristics of precious metals' volatility and has important implications for financial and policy decisions. First, the strong evidence of long memory we found in precious metals implies that the linear return/volatility models are misspecified and cannot be properly used for policy analysis and forecasts. Moreover, accounting for the long memory in a GARCH process reduces volatility persistence. This result is useful for option traders who use volatility in pricing of Call/Put options based on the Black-Scholes formula. Second, testing for the long memory property for the precious metals permits to detect the size of the long memory coef-

ficient. A large coefficient size may indicate that the metal has long positive or negative strays from equilibrium. Thus, the metal with such characteristic is not a good hedge in a group that is known for its safe-haven property. Here comes ultimately the importance of specification of the mean and variance equations in the volatility models. Finally, the LM-based GARCH models have better forecasting quality than the standard GARCH models. Choi and Hammoudeh (2009), for instance, reach similar conclusions for oil and refined products markets.

The remaining part of the article is organized as follows. Section 2 presents a review of the literature. Section 3 describes the empirical framework. Section 4 presents the data used. Section 5 discusses the empirical results. Section 6 provides some concluding remarks.

## **2. Review of Literature**

Most of past studies of the precious metals can essentially be divided into two major categories. The first category has been concerned with the responses of precious metal prices to changes in international institutional and macroeconomic factors (Kaufmann and Winters, 1989; Rookerbie, 1999; Christie–David et al., 2000; Heemskerk, 2001; Ciner, 2001; and Batten et al. 2010). For example, Sjaastad and Scacciavillani (1996) find that fluctuations of floating exchange rates of major currencies, following the breakdown of the Bretton Woods currency arrangements, have led to price instability in the world gold market over the period from January 1982 to December 1990. Batten et al. (2010) find volatility of the precious metals (gold, silver, platinum and palladium) to be sensitive to macroeconomic factors (business cycle, monetary environment and financial market sentiment) but with different degrees. The overall results are consistent with the view that precious metals are too distinct to be considered a single asset class, or represented by a single index. Gold volatility is shown to be explained by monetary variables, but this is not true for silver. Platinum and palladium appear to more likely act as a financial market instrument than gold. Gold also seems to be highly sensitive to exchange rate and inflation, which implies that the yellow metal is the best hedge during inflationary pressures and exchange rate fluctuations. In fact, Hammoudeh, Malik and McAleer (2011) suggest that an optimal portfolio of precious metals that minimizes risk should be dominated by gold.

The second category includes generally more recent studies that have examined the issues of price volatility modeling and information transmission for a broader set of precious

metals, oil and industrial commodities. Some of these studies have considered the implications of the estimated results for portfolio diversification and hedging strategies involving precious metals. To start, Baffes (2007) finds evidence of strong responses of precious metal prices to crude oil price over the period 1960-2005, which is not always confirmed by subsequent studies (e.g., Hammoudeh et al., 2009). Note however that this study uses annual data and oil price is represented by an equally weighted average of Brent, WTI (West Texas Intermediate) and Dubai prices. Hammoudeh and Yuan (2008) employ GARCH-based models to examine the properties of conditional volatility for three important metals (gold, silver, and copper) while controlling for shocks from world oil prices (WTI) and three-month US Treasury bill interest rate. They focus particularly on the following volatility characteristics: persistence, asymmetric reaction to the good and bad news, and transitory and permanent components. Using daily three-month futures prices of the three commodities, these authors find evidence that conditional volatility of gold and silver is more persistent, but less sensitive to leverage effects than that of copper. This result suggests, on the one hand, the importance of accurate volatility modeling especially when gold and silver are used as underlying assets in financial derivatives contracts, and on the other hand the valuable contribution of these two metals in down markets and crisis times. In addition, a rise in short-term interest rates leads to a reduction in the volatility of metals markets, while an increase in the oil prices negatively affects the volatility of some metals. In a related study, Sari et al. (2010) examine linkages among four precious metals, WTI oil price and dollar/euro exchange rate. The empirical results from their short- and long-run analysis based on generalized impulse responses and variance decompositions are consistent with evidence of weak long-run relationships, but strong short-run feedbacks. Spot metal prices are indeed found to be strongly related to exchange rate, but only weakly driven by oil price movements. When considering the case of an emerging market (Turkey), Soytas et al. (2009) find that spot prices of domestic precious metals (gold and silver) are significantly Granger-caused in the short run by domestic interest rate, but not by the changes in the world oil prices (Brent). There is also evidence of unidirectional causality from Turkish Lira/US dollar exchange rate to gold spot prices, thus confirming the reverse and hedging role of gold against exchange rate during crises. As for the long-run analysis, no relationship is found between world oil prices and domestic markets. Finally, based on a multivariate VARMA-GARCH model, Hammoudeh, Yuan, McAleer, and Thompson (2010) document weak volatility spillovers across precious metals, but strong sensitivity of metal volatility to exchange rate variability. They further point out the role of gold as a hedge against exchange rate risk when optimal weights and hedge ratios are computed.

Even though past studies have considerably contributed to improving our understanding of metal price volatility and spillovers based on various extensions of GARCH models (Tully and Lucey, 2007; Hammoudeh and Yuan, 2008; Watkins and McAleer, 2008; Hammoudeh, Yuan and McAleer, 2010), they generally have a major drawback by assuming a stable structure of parameters in the metal volatility process<sup>1</sup>. Differently, the potential of structural breaks is ignored, which might then lead to the detection of “spurious” long memory if long memory is examined (Diebold and Inoue, 2001; Perron and Qu, 2007). Specifically, this assumption implies that the unconditional variance of metal returns is constant, while precious metal markets are very sensitive to fluctuations in supply, demand, and other macroeconomic conditions as reported in previous studies (Radetzki, 1989; Batten et al., 2010; Hammoudeh, Yuan and McAleer, 2010). Moreover, episodes of world geo-political tensions, the Gulf wars, the Asian crisis, worries over Iranian nuclear plans, and the current global economic weaknesses also affect metal prices. These shocks can obviously cause sudden breaks in the unconditional variance of metal returns and, thus, in the parameters of the GARCH dynamics used to model and forecast metal volatility. This possible misspecification should ultimately bias both empirical results and their implications. All in all, neglecting structural breaks in the GARCH parameters induces upward biases in estimates of the persistence of GARCH-based conditional volatility (Mikosch and Stărică, 2004; Hillebrand, 2004).

We should not finish this literature review without indicating that the recent literature on volatility forecasts finds more support for the FIGARCH model over other competing volatility models. Currently, the published work on long memory-based volatility forecasts such as Tansuchat, Chang and McAleer (2009), and Young (2011) is applied primarily to non-precious metal commodity. Our paper deals directly with this issue.

### 3. Empirical Methodology

In this section, we briefly present the tests of long memory and structural changes as well as the GARCH-type specifications we use to account for these stylized facts in the conditional return and volatility of precious metals.

---

<sup>1</sup> The study of Watkins and McAleer (2008) can be viewed as an exception since the authors estimate a rolling GARCH-based volatility model for two non-ferrous metals in order to allow the model’s coefficients to change through time. Such approach, albeit intuitively interesting and meaningful, does not however permit to date the structural changes in the dynamics of metal volatilities.

### 3.1 Long memory tests

Long memory is an important empirical feature of any financial variables. The presence of long memory in the data implies the existence of nonlinear forms of dependence between the first and the second moments, and thus the potential of time-series predictability. Testing for long memory property is an essential task since any evidence of long memory would support the use of LM-based volatility models such as FIGARCH.

We test for long memory components in the return generating process and volatility of precious metals using the Geweke and Porter-Hudak (1983) (GPH), the Robinson and Hendry (1999) Gaussian Semiparametric (GSP), and the Sowell (1992) Exact Maximum Likelihood (EML) test statistics. These tests have been extensively used in the related literature. Note that for long memory in the volatility process, we apply these tests to metals' squared returns, which are commonly regarded as a proxy of conditional volatility (Lobato and Savin, 1998; Choi and Hammoudeh, 2009).

Let  $r_t$  be the precious metal return series. The *GPH estimator* of the long memory parameter  $d$  for  $r_t$  can be then determined using the following periodogram:

$$\log [I(w_j)] = \beta_0 + \beta_1 \log \left[ 4 \sin^2 \left( \frac{w_j}{2} \right) \right] + \varepsilon_j \quad (1)$$

where  $w_j = 2\pi j / T$ ,  $j = 1, 2, \dots, n$ ;  $\varepsilon_j$  is the residual term and  $w_j$  represents the  $n = \sqrt{T}$  Fourier frequencies.  $I(w_j)$  denotes the sample periodogram defined as

$$I(w_j) = \frac{1}{2\pi T} \left| \sum_{t=1}^T r_t e^{-w_j t} \right|^2$$

where  $r_t$  is assumed to be a covariance stationary time series. The estimate of  $d$ , say  $\hat{d}_{GPH}$ , is  $-\hat{\beta}_1$ .

The Robinson and Hendry (1999) *GSP estimator* of the long memory parameter for a covariance stationary series, which is consistent and asymptotically normal under several assumptions, is given by

$$f(w) = Gw^{1-2H} \text{ as } w \rightarrow 0^+ \quad (2)$$



where  $\frac{1}{2} < H < 1$ ,  $0 < G < \infty$  and  $f(w)$  is the spectral density of  $r_t$ . The periodogram with respect to the observations of  $r_t$ ,  $t = 1, \dots, T$  is defined as

$$I(w_j) = \frac{1}{2\pi T} \left| \sum_{t=1}^T r_t e^{i w_j t} \right|^2$$

Accordingly, the estimate of the long memory parameter  $H$  is given by

$$H = \arg \min_{\Delta_1 \leq H \leq \Delta_2} R(H)$$

$$\text{where } \begin{cases} 0 < \Delta_1 < \Delta_2 < 1 \\ R(H) = \log \left\{ \frac{1}{m} \sum_{j=1}^m \frac{I(w_j)}{w_j^{1-2H}} \right\} - (2H-1) \frac{1}{m} \sum_{j=1}^m \log(w_j) \\ m \in (0, [n/2]) \\ w_j = 2\pi j / T \end{cases}$$

The Sowell (1992) *EML estimator* approach to test for long memory is based on the estimation of the ARFIMA( $p, d, q$ ) model using the exact maximum likelihood method. The log-likelihood function takes the following form

$$LL_T(\bar{r}, \theta) = -\frac{T}{2} \log(2\pi) - \frac{1}{2} \log |\Sigma(\theta)| - \frac{1}{2} \bar{r}' \Sigma^{-1} \bar{r} \quad (3)$$

where  $\bar{r}$  is the vector of  $r_t$ ,  $\Sigma$  its covariance-variance matrix, and the EML estimators of the unknown parameter vector  $\theta$  are given by

$$\hat{\theta} = \arg \max_{\theta} LL_T(\bar{r}, \theta)$$

### 3.2 The role of structural breaks

Recent studies establish that structural breaks can severely affect the results of long memory tests and generate spurious long memory in the series (Choi and Zivot, 2006). When structural shifts are effectively present in a stationary short memory process, the estimate of the fractional differencing parameter in LM-based volatility models departs away from zero, and shocks to volatility process only decay at a slowly hyperbolic rate (Diebold and Inoue, 2001; Perron and Qu, 2007). One would then conclude inaccurately in favor of a “spurious” long memory process.

To test for the possibility of structural breaks property in metal returns, we resort to the adjusted version of Inclan and Tiao (1994)'s iterated cumulative sum of squares (ICSS) algorithm.<sup>2</sup> Similar to most structural change tests, the Inclan-Tiao test assumes a normal distribution. For this reason, the unmodified ICSS test may produce spurious changes in the unconditional variance owing to size distortion when the series are leptokurtic and conditionally heteroscedastic. However, in the modified version of their test, Inclan and Tiao (1994) explicitly consider the fourth moment properties of the disturbances and the conditional heteroskedasticity, via a nonparametric adjustment based on the Bartlett kernel. Under general conditions, the modified ICSS statistic exhibits the same asymptotic distribution.

Formally, the null hypothesis of a constant unconditional variance of precious metal returns, which can be modeled by a simple stable GARCH(1,1) specification, is tested against the alternative of presence of structural breaks in the unconditional variance. The ICSS empirical statistic is given by

$$ICSS_a = \sup_k \left| T^{-0.5} F_k \right| \quad (4)$$

where  $F_k = \hat{\gamma}^{-0.5} [C_k - (k/T)C_T]$ ,  $C_k = \sum_{t=1}^k r_t^2$  for  $k = 1, \dots, T$  with  $T$  being the total number of observations,  $r_t$  the precious metal return series,  $\hat{\gamma} = \hat{\sigma}_0^2 + 2 \sum_{i=1}^m [1 - i(m+1)^{-1}] \hat{\beta}_i$ ,  $\hat{\sigma}_i^2 = T^{-1} \sum_{t=1}^k (r_t^2 - \hat{\sigma}^2)(r_{t-1}^2 - \hat{\sigma}^2)$  and  $\hat{\sigma}^2 = T^{-1} C_T$ .  $m$  refers to a lag truncation parameter selected using the procedure in Newey and West (1994). The estimate of the break date is the value of  $k$  that maximizes  $|T^{-0.5} F_k|$ . Under the assumption that return series are zero-mean, normally, independently and identically distributed, the asymptotic distribution of the  $ICSS_a$  statistic is given by  $\sup_c |W^*(c)|$ , where  $W^*(c) = W(c) - cW(1)$  is a Brownian bridge and  $W(c)$  is the standard Brownian motion.

---

<sup>2</sup> Other structural breaks tests have been developed in the literature. Among these tests, CUSUM, and Bai and Perron (2003) tests are frequently used in empirical studies. The CUSUM test is originally designed for testing for variance changes and locating their locations in *iid* samples, but it does not disclose the exact number of breaks and their corresponding dates of occurrence. Similar to the ICSS test, the Bai and Perron (2003)'s testing procedure treats any break points as unknown and permits to test a fixed number of breaks, say  $m$ , versus the alternative  $(m+1)$ . However, Bai and Perron (2003) test has a size-distortion problem when heteroscedasticity is present in the data. In this paper, the modified version of the ICSS is used since it has been corrected for conditional heteroscedasticity.

### 3.3 Long memory versus structural breaks

Long memory and structural changes are often confused. Even though models that accommodate these features separately provide a reasonable description of financial data, the features' presence has different implications for financial modeling exercises. The LM phenomenon suggests constant unconditional volatility, while the structural change implies a significant change in unconditional volatility and thus a structural break model is more plausible. Tests of long memory versus structural breaks are scarce. The existing literature on long memory and structural breaks suggests testing for long memory and structural breaks separately and then estimating a long memory model with breaks, after concluding for the existence of long memory and structural breaks. Several attempts to discriminate between long memory and nonlinearity, we know of, include Van Dijk *et al.* (2002), Lahiani and Scaillet (2009), Baillie and Morana (2009), and Choi *et al.* (2010). Although these methods allow us to decide whether long memory or/and nonlinearity are present in the data, they are based on out-of-sample forecasting and model comparison. In this paper, we use the two tests proposed by Shimotsu (2006) to distinguish between long memory and structural breaks since these tests have the advantage to be the unique in-sample test of long memory against structural breaks, and they also present good power and size.

The first test consists of estimating the long memory parameter over the full sample and over different subsamples, and seeks to examine whether the estimate of the full-sample long memory parameter is equal to the one of each subsample. Let  $b$  be an integer which splits the whole sample in  $b$  subsamples, so that each subsample has  $T/b$  observations.<sup>3</sup> Let also  $\hat{d}^{(i)}$  ( $i = 1, 2, 3, \dots, b$ ) be the local estimator of the true long memory parameter  $d_0$  computed from the  $i^{th}$  subsample, we then define the following expressions:

$$\hat{d}_b = \begin{pmatrix} \hat{d} - d_0 \\ \hat{d}^{(1)} - d_0 \\ \vdots \\ \hat{d}^{(b)} - d_0 \end{pmatrix}, A = \begin{pmatrix} 1 & -1 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 1 & 0 & \dots & -1 \end{pmatrix} \text{ and } \Omega = \begin{pmatrix} 1 & \tau_b' \\ \tau_b & bI_b \end{pmatrix}$$

where  $I_b$  is a  $(b \times b)$  identity matrix and  $\tau_b$  is a  $(b \times 1)$  vector of ones. Following Hurvich and Chen (2000), and Shimotsu (2006), we test the constancy hypothesis of  $d$  ( $H_0: d = d^{(1)} = d^{(2)} = \dots = d^{(b)}$ ) against structural change hypothesis using the following Wald test statistics

---

<sup>3</sup>  $T/b$  is assumed to be an integer.

$$W = 4m \left( \frac{c_{m/b}}{(m/b)} \right) A \hat{d}_b (A \Omega A')^{-1} (A \hat{d}_b)' \quad (6)$$

where  $c_m = \sum_{j=1}^m v_j^2$ ;  $v_j = \log j - \frac{1}{m} \sum_{j=1}^m \log j$ ; and  $m$  is some integer representing the number of periodogram ordinates such that  $m \prec T$ . We consider two values for  $b$  in this study:  $b = 2$  and  $b = 4$ . Note that the above Wald statistic follows a Chi-squared limiting distribution with  $b - 1$  degree of freedom.

The second test requires the estimation of the long memory parameter  $d$ , uses it to take the  $d^{\text{th}}$  difference of the considered return series and tests for the stationarity of the differenced series and its partial sum using the Phillips-Perron test ( $Z_t$ ) and the KPSS test ( $\eta_u$ ).<sup>4</sup> Under the assumptions presented in Shimotsu (2006), the two statistics,  $Z_t$  and  $\eta_u$ , converge towards  $P(W(s, d_0))$  and  $K(W(s, d_0))$  as  $T \rightarrow \infty$  where

$$W(s, d) = W(s) - w(d)(\Gamma(2-d)\Gamma(d+1))^{-1} s^{1-d} W_{d+1}(1)$$

Note that  $P(W(r, d_0))$  is the standard Dickey-Fuller distribution when an intercept is included and  $K(W(s, d_0)) = \int_0^1 (W(s) - sW(1))^2 ds$ .  $W(s, d)$  reduces to the standard Brownian motion  $W(s)$  when  $d = 0$ .  $w(d)$  is a smooth weight function such that  $w(d) = 1$  for  $d \leq 1/2$  and  $w(d) = 0$  for  $d \geq 3/4$ .

At the empirical stage, the above-mentioned tests are carried out as follows. For the first test, the full sample is split into  $b$  subsamples and the long memory parameter  $d^i$  is estimated for each subsample  $i$  ( $i = 1, 2, \dots, b$ ). We consider two values of  $b$  for which the test shows good power and size:  $b = 2$  and  $b = 4$ . Then, the mean of all  $d^i$ , say  $\bar{d}_i$ , is compared to the long memory parameter  $\hat{d}$  estimated over the full sample using Wald tests. As to the second test, the LM parameter  $d$  is estimated over the full sample and then used to take the  $d^{\text{th}}$  difference of the original demeaned series. Finally, the stationarity of the resulting series is tested using the KPSS and Phillips-Perron tests.

### 3.4 The ARFIMA-FIGARCH model

ARCH/GARCH models have been extensively tested for fractional integration in the existing literature (Baillie et al., 1996; Bollerslev and Mikkelsen, 1996, 1999). Past studies

---

<sup>4</sup> See Phillips and Perron (1988) and Kwiatkowski et al. (1992) for more details about unit root tests.

have generally found fractionally integrated models to fit the data better than standard volatility models such as GARCH( $p, q$ ), EGARCH( $p, q$ ), and IGARCH( $p, q$ ). More practically, a fractionally integrated process in both ARMA and GARCH (ARFIMA-FIGARCH) is suitable for modeling any dual LM behavior of financial variables. The main advantage of this model is that it allows a finite persistence of the return and volatility shocks. The econometric specification of the ARFIMA( $p_m, d_m, q_m$ )-FIGARCH( $p_v, d_v, q_v$ ) that will be fitted to each metal return series can be written as follows

$$\begin{aligned}\phi(L)(1-L)^{d_m}(r_t - \mu) &= \beta(L)\varepsilon_t \\ \varepsilon_t &= \eta_t \sqrt[2]{h_t} \\ \alpha(L)(1-L)^{d_v}\varepsilon_t^2 &= w + (1 - \theta(L))v_t\end{aligned}\tag{7}$$

where  $d_m$  and  $d_v$  capture the presence of long memory in the conditional mean and variance of the series, respectively.  $v_t$  represents the skedastic innovation as measured by  $v_t = \varepsilon_t^2 - h_t$ . Note that the ARFIMA( $p_m, d_m, q_m$ ) process is nonstationary when  $d_m \geq 0.5$ , and is said to exhibit long memory for  $0 < d_m < 0.5$ , and short memory for  $d_m = 0$ . The FIGARCH( $p_v, d_v, q_v$ ) process is reduced to a standard GARCH when  $d_v = 0$  and to an IGARCH when  $d_v = 1$ .

The ARFIMA-FIGARCH model is estimated by using the quasi-maximum likelihood (QML) estimation method allowing for asymptotic normality distribution, based on the following log-likelihood function

$$LL_T(\varepsilon_t, \theta) = -\frac{1}{2} \log(2\pi) - \frac{1}{2} \sum_{t=1}^T \left[ \log(h_t) + \frac{\varepsilon_t^2}{h_t} \right]$$

Overall, our empirical approach accounts for long memory in both the mean and variance dynamics of a financial time series. In particular, it permits to test long memory against structural breaks within an in-sample analysis and to show whether the long memory detected in the metal returns is real or is due to the presence of structural breaks and consequently can be considered to be fallacious. Note that the adaptive FIGARCH (A-FIGARCH) model recently developed by Baillie and Morana (2009) is a natural extension of our LM volatility process to incorporate the possibility that the intercept of the conditional variance experiences structural change. However, their method requires an out-of-sample forecasting exercise to confirm the accuracy of the specification of conditional volatility process and may lead to conclude in favor of or against nonlinearity without suggesting the nature of this nonlinearity, i.e., structural break, threshold effects or smooth transition type.

#### 4. Data and Stochastic Properties

We use daily time series data for four major precious metal spot and three-month futures prices (gold, silver, platinum and palladium). These precious metals are traded on the COMEX (Commodity Exchange) in New York, and their prices are measured in US dollars per troy ounce. We use both spot and futures prices because some prices have different distributional characteristics, stylized facts and are followed by different users like investors, traders, physical users and physical producers. The return series are computed as differences in log prices. Data were extracted from Bloomberg database and the whole sample period spans from January 4, 1999 to March 31, 2011. The in-sample estimation period covers the period from January 4, 1999 through December 31, 2009 and is used to estimate the models' parameters. We set the out-of-sample period from January 1, 2010 through March 31, 2011 to evaluate the forecasting performance of the LM-based volatility model, benchmarked against several competing models.

Table 1 summarizes the descriptive statistics for the spot and futures return series as well as their stochastic properties over the in-sample period. Among all the spot and futures returns for the four metals, we find that the highest average returns are for the spot platinum and three-month platinum futures (0.048%), followed closely by the average returns for the spot gold and three-month gold futures (0.046%). The spot and futures palladium returns yield the lowest average, i.e., 0.006% and 0.007%, respectively. It should be noted that for all the metals, the spot returns are not different from their corresponding futures counterparts, with the exception of palladium.

**Table 1. Descriptive Statistics for Returns**

	GOLD	PALL	PLAT	SILV	GOLD3M	PALL3M	PLAT3M	SILV3M
Mean (%)	0.046	0.006	0.048	0.042	0.046	0.007	0.048	0.042
Min. (%)	-7.143	-17.859	-17.277	-16.075	-7.573	-13.201	-14.417	-14.793
Max. (%)	7.382	15.840	16.960	18.278	8.887	15.252	18.678	12.358
Std. dev.	1.143	2.270	1.634	1.956	1.168	2.197	1.603	1.894
Skewness	-0.044	-0.265	-0.434	-0.477	0.244	-0.198	0.266	-0.807
Kurtosis	8.504	6.781	16.059	9.158	9.167	7.681	18.959	11.224
Risk-adjusted return	0.040	0.003	0.029	0.021	0.039	0.003	0.030	0.022
JB	3636.4 <sup>+++</sup>	5531.1 <sup>+++</sup>	20494.7 <sup>+++</sup>	10143.3 <sup>+++</sup>	4589.1 <sup>+++</sup>	2638.8 <sup>+++</sup>	30480.4 <sup>+++</sup>	8404.6 <sup>+++</sup>
ARCH(5)	209.2 <sup>+++</sup>	149.4 <sup>+++</sup>	141.6 <sup>++</sup>	88.1 <sup>+</sup>	146.6 <sup>+++</sup>	188.347 <sup>+++</sup>	201.5 <sup>+++</sup>	171.7 <sup>+++</sup>
Q(5)	3.4	17.6 <sup>+++</sup>	4.6	7.4	7.7	16.4 <sup>+++</sup>	2.4	0.8
Q <sup>2</sup> (5)	313.9 <sup>+++</sup>	226.2 <sup>+++</sup>	194.5 <sup>+++</sup>	114.4 <sup>+++</sup>	216.7 <sup>+++</sup>	259.0 <sup>+++</sup>	286.1 <sup>+++</sup>	252.0 <sup>+++</sup>

Notes: this table reports the descriptive statistics of precious metal returns. GOLD, PALL, PLAT and SILV denote respectively the spot returns of the four precious metals: gold, palladium, platinum and silver. GOLD3M, PALL3M, PLAT3M and SILV3M are returns of three-month metal futures contracts. JB, ARCH, Q(5) and Q<sup>2</sup>(5) refer to the empirical statistics of the Jarque-Bera test for normality, ARCH test for conditional heteroscedasticity, Ljung-Box test for autocorrelation with five lags applied to raw returns, and Ljung-Box test for autocorrelation with five lags applied to squared returns. <sup>+</sup>, <sup>++</sup> and <sup>+++</sup> indicate rejection of the null hypothesis at the 10%, 5% and 1% levels, respectively.

The daily unconditional volatility of all the spot and futures returns, as measured by standard deviations, is substantial, with values ranging from 1.143% (spot gold) to 2.270% (spot palladium). With respect to risk-return profile, spot and futures palladium returns have the highest volatility, but the lowest returns as indicated above, thus historically the higher risk for this precious metal is not compensated for by higher return. This finding also suggests that palladium might not be a good hedge for portfolios of stocks, especially in times of crises and bear markets.

The descriptive statistics also demonstrate that skewness is negative in all cases, except for three-month gold and platinum contracts, and that excess kurtosis is highly significant. Clearly, most of the precious metal returns have fatter tails and longer left tails (extreme negative returns) than the normal distribution. The Jarque-Bera test (JB) confirms these findings since it rejects normality. Results from the ARCH(5) tests for conditional heteroscedasticity provide strong evidence of ARCH effects in all the precious-metal return series, which in turn suggests the usefulness and suitability of GARCH-type models for modeling and forecasting their time-varying conditional volatility. Finally, the Ljung-Box tests,  $Q(5)$  and  $Q^2(5)$ , indicate that autocorrelation is present for the (raw) returns of spot palladium and three-month palladium futures, but autocorrelation in squared returns is highly significant. These results typically show signs of high degree of persistence in the conditional volatility process of precious metals. It is worth noting that the Ljung-Box tests with different lag length indicate the presence of return autocorrelation for series other than palladium.

## 5. Results and interpretations

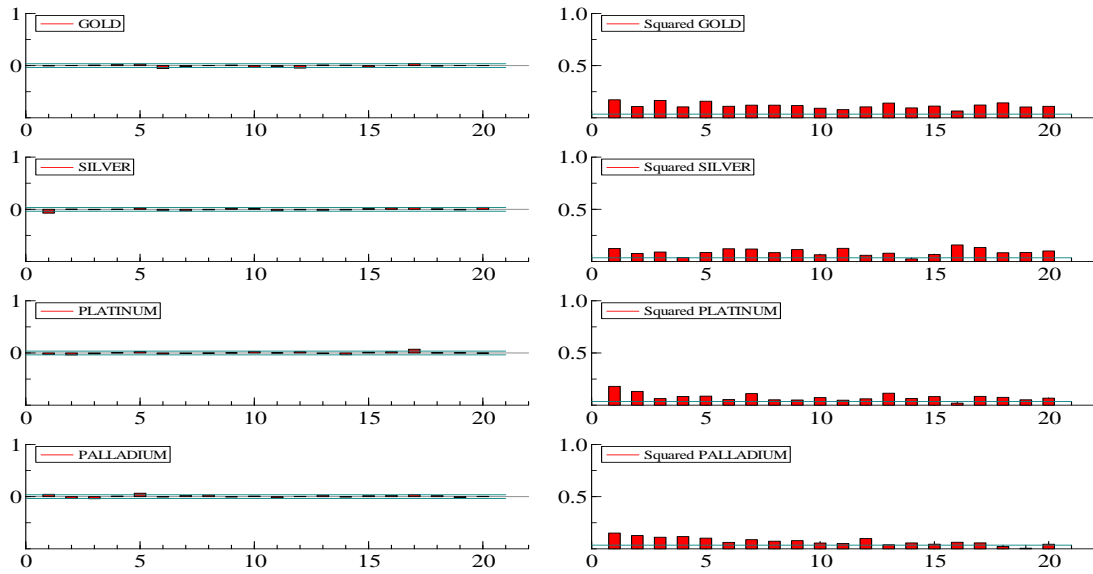
In this section, we discuss the in-sample results obtained from the autocorrelation function analysis, tests of long memory and structural breaks, and LM-based volatility models for precious metals' spot and futures returns. We also report the results of the out-of-sample forecasting analysis where LM-based volatility models are benchmarked against other competing volatility models

### 5.1 Autocorrelation

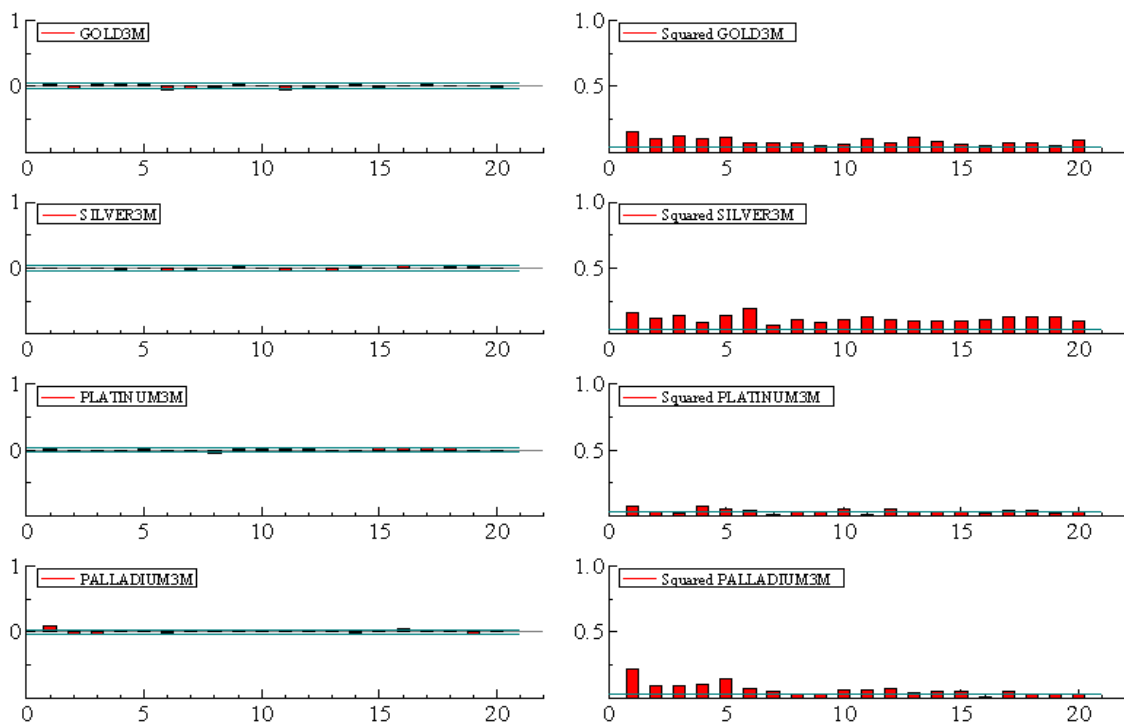
The distributional characteristics of the metal return series can be investigated further by analyzing the behavior of their autocorrelation functions. The results, displayed in Figures 1 and 2, show that the autocorrelation functions of the raw returns are small and have no par-

ticular form. Most of them stay inside the 95% confidence intervals. This is suggestive of their short memory property. The autocorrelation functions of the squared returns are however larger, and they remain positive and significant for many lags. More importantly, they exhibit a very slow decay with a hyperbolic rate, indicating that the time series are strongly autocorrelated up to a long lag. The only exception is observed for the three-month futures palladium contracts which show a faster decay.

**Figure 1. Autocorrelation Functions for Spot and Squared Spot Returns**



**Figure 2. Autocorrelation Function for Futures and Squared Futures Returns**





Overall, our findings shed light on a very persistent behavior in metals squared returns. They are consistent with the common characteristics of squared returns widely documented for financial asset returns (Ballie et al., 1996; Bollerslev and Mikkelsen, 1996; Choi and Hammoudeh, 2009). In addition, it is well argued in the previous literature that these characteristics are suggestive of LM dynamics, and that they can be spuriously generated when structural breaks are ignored in economic modeling of financial series. For example, Diebold and Inoue (2001) emphasize that infrequent stochastic breaks can create strong persistence in the autocorrelation structure of financial series.

### 5.2 Results of long memory tests

We apply the three LM tests (GPH, GSP and EML) to the raw and squared returns of the spot and futures prices of our four precious metals. The obtained results are reported in Table 2. For the (raw) return series, the tests used unanimously show evidence of LM patterns for spot platinum, spot palladium and palladium futures as the null hypothesis of no persistence is always rejected at levels ranging from 1% to 10%. Separately, while the GSP test concludes in favor of the presence of long memory for spot gold and platinum futures, the EML test provides evidence of long memory for spot silver. However, gold and silver futures returns do not have LM properties.

**Table 2. Results of LM Tests for Returns and Squared Returns**

	Returns			Squared returns		
	GPH	GSP	EML	GPH	GSP	EML
GOLD	-0.082 [0.104]	-0.064 [0.082]	-0.002 [0.920]	0.622 [0.000]	0.549 [0.000]	0.187 [0.000]
GOLD3M	-0.082 [0.104]	-0.060 [0.105]	-0.003 [0.851]	0.538 [0.000]	0.475 [0.000]	0.179 [0.000]
SILV	-0.002 [0.960]	-0.022 [0.542]	-0.052 [0.000]	0.509 [0.000]	0.458 [0.000]	0.158 [0.000]
SILV3M	0.005 [0.920]	-0.022 [0.549]	-0.015 [0.318]	0.545 [0.000]	0.529 [0.000]	0.176 [0.000]
PLAT	0.106 [0.034]	0.101 [0.006]	-0.025 [0.067]	0.403 [0.000]	0.399 [0.000]	0.175 [0.000]
PLAT3M	0.065 [0.194]	0.104 [0.005]	-0.001 [0.940]	0.517 [0.000]	0.542 [0.000]	0.119 [0.000]
PALL	0.096 [0.056]	0.087 [0.018]	-0.029 [0.043]	0.325 [0.000]	0.309 [0.000]	0.181 [0.000]
PALL3M	0.092 [0.067]	0.081 [0.029]	-0.054 [0.000]	0.433 [0.000]	0.381 [0.000]	0.199 [0.000]

Notes: this table reports the results from three LM tests: Geweke and Porter-Hudak (1983)'s GPH, Robinson and Hendry (1999)'s Gaussian Semiparametric (GSP), and Sowell (1992)'s Exact Maximum Likelihood (EML). The GPH and GSP tests were carried out with a bandwidth of  $T/16$ , where  $T$  refers to the total number of observations over the in-sample period. The associated  $p$ -values are given in brackets. GOLD, PALL, PLAT and SILV denote respectively spot gold, palladium, platinum and silver. GOLD3M, PALL3M, PLAT3M and SILV3M denote the corresponding three-month metal futures contracts.

The results for squared returns are sensitively different from those for the returns. Indeed, long memory property is found to be highly significant for all the squared returns, whatever the LM tests used. Since squared returns are a good proxy for volatility, these findings thus suggest that the conditional volatility of precious metals would tend to be range-dependent, persist and decay slowly. Intuitively, this volatility persistence can be appropriately modeled by a FIGARCH process because it allows for long memory behavior and slow decay of the impact of a volatility shock.

It is, however, important to note that the GPH and GSP estimates of the LM parameter  $d$  are higher than 0.5 for several spot and futures squared returns (e.g., gold, silver and platinum), and are in contrast to the usual findings. Many explanations for these unusual values of  $d$  are possible. They can firstly arise from the bias inherent in the GPH and GSP estimators. Another explanation, given by Granger and Hyung (2004), is related to the fact that long memory may be the result of various kinds of misspecifications and/or the presence of structural breaks. In this scheme of things, a greater accumulation of misspecifications naturally would lead to greater spurious long memory.<sup>5</sup>

### *5.3 Evidence of structural breaks*

The results from the Inclan and Tiao (1994) test regarding the number and estimated break dates are reported in Table 3. They demonstrate that six out of the eight return series exhibit at least one structural break in their unconditional variance dynamics. Indeed, the ICSS algorithm detects four breaks for the spot gold, three breaks for the spot silver, silver futures and platinum futures, two breaks for the spot platinum, and one break for gold futures. Five of these indentified breaks are a priori associated with the 2008-2009 global financial crisis which was sparked by the US subprime and banking defaults that took place in July 2007. This result can be explained partly by the “flight-to-quality” phenomenon which appears in times of crises when investors rush to buy less risky assets and financial contracts on safe assets such as gold and platinum.

The aforementioned findings suggest that the evidence of long memory in the return volatility of six precious metal price series (spot and futures prices of gold, silver and plati-

---

<sup>5</sup> We also estimated a FIGARCH(1, $d$ ,1) model for metal returns where the conditional mean is modeled by a simple AR(1) process. The results, not reported here for concision purpose, indicate that the estimates of the LM parameters  $d$  are large and highly significant for all the series, and they are very different from unity as well. This finding, in line with the results of long memory tests, thus raises the question about the robustness of LM evidence. The reason is that large values of  $d$  may be due to the ignorance of possibly structural changes in the dynamics of precious metal squared returns (Banerjee and Urga, 2005; Bhardwaj and Swanson, 2006).

num) may be overstated due to the presence of structural breaks which are not accounted for in the LM tests. The discrimination between long memory and structural breaks is however not an easy task. Several studies have examined the nature and causes of volatility persistence for financial series, but the results remain inconclusive. For instance, Bhardwaj and Swanson (2006) find that the LM models give better out-of-sample forecasts than ARMA, standard GARCH and related models. The LM models are also found to outperform models with occasional breaks in out-of-sample analysis (Granger and Hyung, 2004). On the contrary, Choi and Zivot (2007) document that accommodating for structural breaks reduces the volatility persistence.

**Table 3. Results of Structural Break Tests**

	Number of breaks	Break dates
GOLD	4	09/17/1999; 10/05/1999; 10/29/1999; 02/02/2008
GOLD3M	1	12/30/2005
SILV	3	06/21/2000; 01/15/2001; 01/02/2004
SILV3M	3	03/03/2000; 09/12/2001; 01/01/2004
PALL	0	-
PALL3M	0	-
PLAT	2	08/13/2008; 01/06/2009
PLAT3M	3	11/12/2001; 01/22/2008; 07/07/2009

Notes: this table reports the results of the structural break tests based on the application of the modified ICSS algorithm to the metal returns data over the in-sample period. GOLD, PALL, PLAT and SILV denote respectively spot returns for gold, palladium, platinum and silver. GOLD3M, PALL3M, PLAT3M and SILV3M represent the returns on the corresponding three-month metal futures contracts.

Before moving to estimate the LM-based volatility models for precious metals, it is essential to test for the relevance of long memory against structural breaks. For doing so, we rely on the procedure proposed by Shimotsu (2006), which examines the null hypothesis of long memory against the alternative of a structural change. Two and four subsamples are considered since augmenting the number of hypothetical subsamples does not increase the power of the test. The results are reported in Table 4.

**Table 4. Tests of Long Memory versus Structural Breaks**

	$\hat{d}$	$\bar{d}$		$W$		$Z_t$	$\eta_u$
		$b = 2$	$b = 4$	$b = 2$	$b = 4$		
GOLD	0.025	0.036	0.038	2.463	2.724	-2.590	0.145
GOLD3M	-0.001	0.001	0.003	0.013	0.243	-2.378	0.174
SILV	-0.004	-0.008	-0.021	0.358	6.838	-2.087	0.110
SILV3M	0.017	0.016	0.011	0.001	2.462	-2.245	0.088
PALL	0.025	0.025	0.021	0.492	1.160	-1.598	0.109
PALL3M	0.011	0.004	-0.005	0.025	5.461	-2.403	0.054
PLAT	-0.009	-0.011	-0.033	0.676	13.368*	-2.289	0.062
PLAT3M	0.039	0.041	0.035	0.025	7.289	-1.671	0.101

Notes: this table reports the results of statistical tests of the LM hypothesis against structural change.  $b$  denotes the number of subsamples.  $W$ ,  $Z_t$  and  $\eta_u$  are the empirical statistics of the Wald, Phillips-Perron, and KPSS

tests, respectively. \* indicates rejection of the null hypothesis of constancy of the LM parameter  $d$  at the 5% level. The critical values for the Wald test are  $\chi^2_{0.95}(1) = 3.84$  and  $\chi^2_{0.95}(3) = 7.82$ , respectively.

The results of the Wald ( $W$ ) test show that the constancy of the LM parameter  $d$  cannot be rejected for all the series regardless of the number of subsamples, except for platinum's spot returns. This leads us to the conclusion that the evidence against long memory is not invasive. Moreover, we see that for all the series the Philips-Perron test ( $Z_t$ ) does not reject the null hypothesis of  $I(d)$ , while the hypothesis of stationarity cannot be rejected by KPSS test ( $\eta_u$ ). Taken together, our findings suggest that not all the persistence we found in the squared returns and conditional volatility of the precious metals considered is due to the presence of structural change. The evidence of long memory we reported is thus not spurious for almost all series.

#### *5.4 Return and volatility modeling in presence of long memory*

The results of Table 2 and Table 4 show that the ARFIMA-FIGARCH class models can be used to reproduce the LM characteristics in the conditional mean and variance of precious metals return dynamics. In particular, the empirical evidence in Table 2 suggests an ARMA-FIGARCH specification for the three-month gold and silver futures contracts, and an ARFIMA-FIGARCH specification for the remaining return series. With respect to the results of AIC and BIC information criteria, we select one lag for all the specifications and present the estimation results in Table 5.<sup>6</sup>

We first find moderate evidence of persistence in precious metal returns since the LM parameter in the mean equation  $d_m$  is at most significant at the 5% level (spot and three-month platinum). Long memory evidence is however not found for spot silver, as suggested by the EML's LM test. The value of  $d_m$  is negative in all cases and ranges from -0.105 (spot gold) to -0.050 (three-month palladium). The small and negative values of the LM parameter typically imply that the return-generating processes rarely stray far from the mean and have strong tendency to revert to it quickly.

The LM parameters in the conditional volatility processes are all positive and highly significant. Their relatively large values, ranging from 0.328 (three-month gold) to 0.957 (three-month platinum), suggest that these metals' volatility processes display little tendency

---

<sup>6</sup> We also use the Wald tests to examine the hypothesis that  $d_v = 1$ . The obtained results, not reported in the paper, always underscore the rejection of this hypothesis at the 1% level.

to revert towards the volatility mean. Note that  $d_v$  for the three-month platinum is very close to unity, and accordingly an IGARCH process seems to be more suitable for this metal. Finally, it is observed that the ARFIMA-FIGARCH class model appropriately captures the price dynamics of the four precious metals in view of the results of specification tests. Indeed, the ARCH effects and autocorrelations no longer exist in the standardized residuals.

**Table 5. Evidence of Dual Long Memory from the ARFIMA-FIGARCH Class Model**

	GOLD	GOLD3M	SILV	SILV3M	PALL	PALL3M	PLAT	PLAT3M
$\mu_m$	0.034** (0.013)	0.037** (0.016)	0.014 (0.019)	0.017 (0.026)	0.061* (0.036)	0.044 (0.034)	0.070*** (0.014)	0.064*** (0.015)
AR(1)	0.665*** (0.095)	0.956*** (0.016)	-0.353* (0.206)	-0.382* (0.214)	-0.213 (0.179)	-0.110 (0.201)	-0.320* (0.180)	-0.160** (0.076)
MA(1)	-0.546*** (0.095)	-0.966*** (0.016)	0.296 (0.383)	0.360 (0.264)	0.297* (0.162)	0.278* (0.159)	0.369** (0.173)	0.320** (0.157)
$d_m$	-0.105* (0.058)	----	-0.036 (0.037)	----	-0.057* (0.030)	-0.050* (0.030)	-0.065** (0.032)	-0.065** (0.033)
$\mu_v$	0.029* (0.015)	0.070* (0.042)	0.038* (0.021)	0.057* (0.032)	0.287** (0.149)	0.143** (0.072)	0.066** (0.031)	0.006** (0.003)
$\alpha$	0.358*** (0.109)	0.372*** (0.107)	0.280 (0.069)	0.377*** (0.084)	0.183** (0.086)	0.333*** (0.086)	0.310*** (0.097)	0.153*** (0.056)
$\beta$	0.750** (0.377)	0.633*** (0.109)	0.749*** (0.074)	0.728*** (0.095)	0.532*** (0.130)	0.674*** (0.133)	0.630*** (0.122)	0.968*** (0.009)
$d_v$	0.518*** (0.104)	0.328*** (0.089)	0.566*** (0.099)	0.465*** (0.091)	0.483*** (0.168)	0.527*** (0.150)	0.514*** (0.106)	0.957*** (0.046)
LL <sub>T</sub>	-4101.154	-4273.3	-5485.437	-5422.223	-6096.875	-6060.360	-4947.781	-5123.715
AIC	2.864	2.983	3.829	3.784	4.255	4.230	3.454	3.577
SIC	2.881	2.998	3.846	3.799	4.272	4.246	3.471	3.593
Q(5)	2.595	3.084	4.228	4.040	3.656	3.948	3.133	4.233
ARCH(5)	0.300	0.434	1.424	1.180	0.126	0.689	0.165	0.145

Notes: this table reports the results of the quasi-maximum likelihood estimation of the ARFIMA-FIGARCH class model for the daily metals spot and futures returns.  $\mu_m$ ,  $\mu_v$ ,  $d_m$ , and  $d_v$  refer to the constant terms and LM parameters of the mean and variance equations, respectively. GOLD, PALL, PLAT and SILV denote respectively the log spot returns of the four precious metals: gold, palladium, platinum and silver. GOLD3M, PALL3M, PLAT3M and SILV3M are the returns on three-month metal futures contracts. Robust standard errors are given in parenthesis. Q(5) and ARCH(5) are the empirical statistics of the Ljung-Box and Engle (1982) tests for autocorrelation and conditional heteroscedasticity, respectively. \*, \*\* and \*\*\* denote significance at the 10%, 5% and 1% levels, respectively.

### 5.5 Forecasting evaluation

We now turn to examine the ability of the ARFIMA-FIGARCH class model in forecasting the precious metals' returns and volatility. This model's out-of-sample forecasting performance is benchmarked against that of four competing GARCH-based models including GARCH, EGARCH, IGARCH and HYGARCH, which do not accommodate the properties of fractionally integrated time series. The mean equation specifications for all metals' returns are the same as reported in subsection 5.4. That is, the ARMA specification is used for the three-month gold and silver returns, while the ARFIMA is used for the remaining series. Note that standard GARCH (Bollerslev, 1986) and IGARCH (Engle and Bollerslev, 1986) models

are special cases of FIGARCH model when the LM parameter is, respectively, equal to zero and one. The EGARCH model, introduced by Nelson (1991), has the advantage of allowing asymmetry in the reaction of conditional volatility to the sign of shocks to the return series. The hyperbolic GARCH model or HYGARCH proposed by Davidson (2004) is viewed as a more general version of the FIGARCH model with hyperbolic convergence rates, where shock amplitude and long memory are treated separately.

The return and volatility forecasts of the benchmark and competing models are generated over the period from January 1, 2010 through March 31, 2011, yielding a total of 325 daily observations. The prediction error is then compared across models on the basis of three evaluation criteria commonly used in the previous literature (Kang et al., 2009; Weil et al., 2010). These criteria are the mean absolute error (MAE), root mean square error (RMSE), and Theil's inequality coefficient (TIC). Let  $n$  be the number of forecasts, and  $y_t$  and  $\hat{y}_t$  the observed and the predicted values of  $y_t$  at time  $t$ . Here,  $y_t$  refers alternatively to the metal return and volatility series. The evaluation criteria are given in Equations (8)-(10) below.

$$MAE = \frac{1}{n} \sum_{t=1}^n |y_t - \hat{y}_t| \quad (8)$$

$$RMSE = \sqrt{\frac{\sum_{t=1}^n (y_t - \hat{y}_t)^2}{n}} \quad (9)$$

$$TIC = \frac{\sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2}}{\sqrt{\frac{1}{n} \sum_{t=1}^n y_t^2} + \sqrt{\frac{1}{n} \sum_{t=1}^n \hat{y}_t^2}} \quad (10)$$

The best forecasting GARCH-based model is the one that generates the lowest prediction error. The forecasting results for the return series are reported in Table 6, whereas those for the volatility series are presented in Table 7.

Table 6 shows that the EGARCH model provides the best forecasts of the return series in 15 out of the 24 cases (or 62.5%) based on the three evaluation criteria. Indeed, this model is commonly selected by the *MAE*, *RMSE* and *TIC* criteria in four out of the eight precious metal price series. The FIGARCH-based model is identified as the second-best model since it is chosen by the return evaluation criteria in 12 out of the 24 cases (50%). It generates better forecasts than the EGARCH in only three cases (spot gold, spot platinum, and three-month

platinum). The other GARCH-based competing models (GARCH, IGARCH and HYGARCH) have the lowest prediction errors in only three out of the eight metal price series (spot palladium, three-month palladium, and three-month platinum).

**Table 6. Out-of-Sample Predictive Accuracy of Competing GARCH-Based Models for the Return Series**

	GOLD	GOLD3M	SILV	SILV3M	PALL	PALL3M	PLAT	PLAT3M
<b>FIGARCH(1,d,1)</b>								
<i>MAE</i>	<b>0.7318</b>	0.7070	1.6170	1.4570	<b>1.8260</b>	<b>1.7240</b>	<b>0.9267</b>	<b>0.9030</b>
<i>RMSE</i>	<b>0.9772</b>	0.9680	2.1360	1.9550	<b>2.4750</b>	2.3290	<b>1.2484</b>	<b>1.2590</b>
<i>TIC</i>	<b>0.9266</b>	0.9584	0.9940	0.9902	0.9744	<b>0.9770</b>	<b>0.9431</b>	0.9486
<b>GARCH(1,1)</b>								
<i>MAE</i>	0.7321	0.7074	1.6170	1.4590	1.8270	<b>1.7240</b>	0.9274	0.9032
<i>RMSE</i>	0.9774	0.9684	2.1370	1.9570	<b>2.4750</b>	2.3290	1.2490	<b>1.2590</b>
<i>TIC</i>	0.9281	0.9664	0.9970	0.9969	0.9761	0.9790	0.9485	0.9497
<b>IGARCH(1,1)</b>								
<i>MAE</i>	0.7323	0.7076	1.6170	1.4590	1.8360	<b>1.7240</b>	0.9274	<b>0.9030</b>
<i>RMSE</i>	0.9775	0.9685	2.1370	1.9570	2.4850	2.3290	1.2490	<b>1.2590</b>
<i>TIC</i>	0.9298	0.9684	0.9970	0.9970	0.9740	0.9797	0.9490	0.9501
<b>EGARCH(1,1)</b>								
<i>MAE</i>	0.7342	<b>0.7054</b>	<b>1.6150</b>	<b>1.4520</b>	<b>1.8260</b>	<b>1.7240</b>	0.9271	0.9034
<i>RMSE</i>	0.9802	<b>0.9673</b>	<b>2.1340</b>	<b>1.9510</b>	<b>2.4750</b>	<b>2.3280</b>	1.2500	1.2620
<i>TIC</i>	0.9462	<b>0.9390</b>	<b>0.9844</b>	<b>0.9663</b>	<b>0.9732</b>	0.9795	0.9478	<b>0.9422</b>
<b>HYGARCH(1,d,1)</b>								
<i>MAE</i>	0.7369	0.7073	1.6170	1.4570	<b>1.8260</b>	<b>1.7240</b>	0.9272	0.9031
<i>RMSE</i>	0.9815	0.9683	2.1360	1.9550	<b>2.4750</b>	2.3290	1.2490	<b>1.2590</b>
<i>TIC</i>	0.9744	0.9646	0.9946	0.9903	0.9755	0.9798	0.9439	0.9470

Notes: This table reports the results of the one-day out-of-sample prediction errors of metal return series for the benchmark FIGARCH(1,d,1) and the four competing models. For all models, the ARMA(1,1) specification is used for the conditional means of the three-month gold and silver, while the ARFIMA(1,d,1) specification is retained for the remaining return series. A bold entry denotes the model that provides the lowest prediction error for each metal return.

As for the volatility forecasts, Table 7 shows that the FIGARCH volatility model is selected according to the three evaluation criteria for spot silver and future palladium, by at least two criteria for spot palladium, and future silver, and by one criterion for platinum futures. Taken together, the FIGARCH volatility model provides the best volatility forecasts in 11 out of the 24 cases based on all the evaluation criteria for all the metals. The EGARCH model is the second-best volatility model and performs well for 9 out of the 24 cases including spot platinum, spot and future gold. Each of the other GARCH-based competing models (GARCH, IGARCH and HYGARCH) has the lowest prediction errors in at most 2 out of the 24 cases.

**Table 7. Out-of-Sample Predictive Accuracy of Competing GARCH-Based Models for the Volatility Series**

	GOLD	GOLD3M	SILV	SILV3M	PALL	PALL3M	PLAT	PLAT3M
<b>FIGARCH(1,d,1)</b>								
<i>MAE</i>	1.731	1.589	<b>2.044</b>	<b>1.382</b>	<b>1.223</b>	<b>1.327</b>	1.531	1.812
<i>RMSE</i>	1.733	2.345	<b>1.521</b>	<b>1.931</b>	<b>1.159</b>	<b>1.693</b>	1.460	<b>1.031</b>
<i>TIC</i>	0.537	0.599	<b>0.555</b>	0.572	0.672	<b>0.665</b>	0.621	0.586
<b>GARCH(1,1)</b>								
<i>MAE</i>	1.758	1.605	2.249	1.398	1.631	1.492	1.291	<b>1.786</b>
<i>RMSE</i>	1.736	2.454	1.532	1.953	1.217	1.804	1.377	1.245
<i>TIC</i>	0.529	0.592	0.596	0.574	<b>0.661</b>	0.673	0.635	0.599
<b>IGARCH(1,1)</b>								
<i>MAE</i>	1.723	2.196	2.065	1.801	1.741	1.421	1.236	1.968
<i>RMSE</i>	2.405	2.586	1.528	2.402	1.425	1.944	1.278	1.442
<i>TIC</i>	0.593	0.596	0.584	0.554	0.672	0.669	0.694	<b>0.583</b>
<b>EGARCH(1,1)</b>								
<i>MAE</i>	1.734	<b>1.521</b>	2.145	1.389	1.227	1.337	<b>1.227</b>	1.800
<i>RMSE</i>	<b>1.724</b>	<b>2.312</b>	1.525	1.977	1.211	1.752	<b>1.251</b>	1.087
<i>TIC</i>	<b>0.528</b>	<b>0.589</b>	0.593	<b>0.521</b>	0.683	0.667	<b>0.587</b>	0.701
<b>HYGARCH(1,d,1)</b>								
<i>MAE</i>	<b>1.713</b>	1.578	2.945	1.921	1.441	1.901	1.643	1.792
<i>RMSE</i>	1.747	2.392	1.798	2.107	1.241	1.935	1.374	1.207
<i>TIC</i>	0.532	0.591	0.577	0.543	<b>0.661</b>	0.670	0.653	0.587

Notes: This table reports the results of one-day out-of-sample prediction errors of the metal volatility series for the benchmark FIGARCH(1,d,1) and four competing models. For all models, ARMA(1,1) specification is used for the conditional means of the three-month gold and silver, while the ARFIMA(1,d,1) specification is retained for the remaining return series. A bold entry denotes the model with the lowest prediction error.

To sum up, our forecast analysis shows that for predicting the return series, the EGARCH model is the best option seconded by the FIGARCH. The EGARCH model works the best for the prediction of the silver and palladium spot and futures returns, as well as for the gold futures returns. However, the FIGARCH model is empirically identified as the relatively best suitable model in terms of volatility forecasts. Our out-of-sample results thus indicate that the LM evidence is not spurious, and hence accounting for this property in the ARFIMA-FIGARCH class models leads to improvement in the quality of forecasts for some precious metals' spot and futures returns. It is finally worth noting that the superiority of the EGARCH-based models in some returns cases suggests that extending the ARFIMA-FIGARCH models to accommodate the asymmetric volatility effects may increase their predictive power.

## 6. Conclusion

Within the context of the current financial crisis, there is an increasing interest by traders, investors, portfolio managers, physical users and producers, and policy makers to understand better the performance and the distributional characteristics of increasingly important



asset classes. Such enhanced understanding should lead to better returns, greater benefits from portfolio diversification, more adequate pricing of derivatives and improvement in risk management strategies. Among these asset classes are the precious metals which consist of gold, silver, platinum and palladium. These precious metals have been very attractive for portfolio investments over the recent turbulent years owing to their role as reverse or safe haven assets and to the increase in demand for their economic uses.

Several papers in the literature have addressed the issue of volatility modeling for precious metals, but none of them have explicitly investigated the nature and causes of the observed volatility persistence. This paper is an attempt to fill this gap by testing the relevance of long memory against structural breaks in modeling the return and volatility for the spot and futures prices of those four precious metals.

Using a battery of long memory and structural break tests and the Inclan and Tiao (1994) modified ICSS algorithm for dating structural breaks, we find that long memory is particularly strong and plays a dominant role in explaining the spot and futures price dynamics for the four strategic metals. The selection tests also conclude in favor of long memory to the detriment of structural breaks. As such, investors in these precious metals markets can make use of the long-range dependence property to generate better understanding of higher profits through using past information and statistical models such as the linear ARFIMA processes that accommodate LM characteristics.

Comparing the empirical results across the metals, the series of platinum futures returns exhibits the highest long memory in the variance equation, suggesting that the latter may experience long strays away from the mean. Thus, platinum is not a good hedging instrument during bear or crisis markets. Moreover, this series requires an IGARCH modeling for its conditional variance. Among the remaining metals, gold may serve as a good hedge during market downturns because its return has relatively short strays from its mean and variance, confirming the most pronounced safe haven status on this shiny metal. Finally, our out-of-sample analysis indicates that the FIGARCH-based model is the best and the second best model in terms of the predictive power for the volatility and returns, respectively. Our findings also point to the relevance of asymmetry in the dynamics of the precious metal returns and volatility as the EGARCH-based model is the best and second best model for predicting return and volatility, respectively. Thus, extending the ARFIMA-FIGARCH models to accommodate for asymmetry in the return and volatility series may lead to an increase in their predictive power. This empirical feature is left for our future research.

## References

- Agnolucci, P. (2009). Volatility in crude oil futures: A comparison of the predictive ability of GARCH and implied volatility models. *Energy Economics*, 31, 316-321.
- Akram, Q. F. (2009). Commodity prices, interest rates and the dollar. *Energy Economics*, 31, 838-851.
- Arouri, M., & Nguyen, D. K. (2010). Oil prices, stock markets and portfolio investment: evidence from sector analysis in Europe over the last decade. *Energy Policy*, 38, 4528-4539.
- Arouri, M. E.-H. and Nguyen, D. K. . (2010). Time-Varying Characteristics of Cross-Market Linkages with Empirical Application to Gulf Stock Markets. *Managerial Finance*, 36(1), 57-70.
- Arouri M., Dinh TH., and Nguyen DK. (2010), "Time-varying Predictability in Crude Oil Markets: The Case of GCC Countries", *Energy Policy*, Vol. 38, No. 8, pp. 4371-4380, 2010.
- Arouri, M., J. Jouini and D. Nguyen. (2011). "Volatility Spillovers between Oil Prices and Stock Sector Returns: Implications for Portfolio Management." *Journal of International Money and Finance*, 30(7): 1387-1405.
- Arouri, M., Jouini, J. and Nguyen, D.K., 2012. On the impacts of oil price fluctuations on European equity markets: Volatility spillover and hedging effectiveness. *Energy Economics*. 34, 611-617.
- Baffes, J. (2007). Oil spills on other commodities. *Resources Policy*, 32, 126-134.
- Bai, J., Perron, P., 2003. Computation and analysis of multiple structural change models. *Journal of Applied Econometrics*, 18, 1-22.
- Baillie, R., Bollerslev, T., & Mikkelsen, H. (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 74, 3-30.
- Banerjee, A., & Urga, G. (2005). Modelling structural breaks, long memory and stock market volatility: An overview. *Journal of Econometrics*, 129, 1-34.
- Batten, J. A., Ciner, C., & Lucey, B. M. (2010). The macroeconomic determinants of volatility in precious metals markets. *Resources Policy*, 35, 65-71.
- Bhardwaj, G., & Swanson, N. R. (2006). An empirical investigation of the usefulness of ARFIMA models for predicting macroeconomic and financial time series. *Journal of Econometrics*, 131, 539-578.
- Bollerslev, T., & Mikkelsen, H. O. (1996). Modeling and pricing long-memory in stock market volatility. *Journal of Econometrics*, 73, 151-184.
- Bollerslev, T., & Mikkelsen, H. O., 1999. Long-term equity anticipation securities and stock market volatility dynamics. *Journal of Econometrics*, 99, 75-99.
- Browne, F., & Cronin, D. (2010). Commodity prices, money and inflation. *Journal of Economics and Business*, 62, 331-345.
- Chan-Lau, J. A., Mathieson, D. J., & Yao, J. Y. (2004). Extreme contagion in equity markets. *IMF Staff Papers*, 51, 386-408.

- Choi, K., & Hammoudeh, S. (2009). Long memory in oil and refined products markets, *Energy Journal*, 30, 97-116.
- Choi, K., & Zivot, E. (2007). Long memory and structural changes in the forward discount: an empirical investigation, *Journal of International Money and Finance*, 26, 342-363.
- Choi, K., Yu, W. C. & Zivot, E. (2010). Long memory versus structural breaks in modeling and forecasting realized volatility. *Journal of International Money and Finance*, 29, 857-875.
- Christie-David, R., Chaudhry, M., & Koch, T. W. (2000). Do macroeconomics news releases affect gold and silver prices? *Journal of Economics and Business*, 52, 405-421.
- Ciner, C. (2001). On the long run relationship between gold and silver prices: A note. *Global Finance Journal*, 12, 299-303.
- Daskalaki, C., & Skiadopoulos, G. (2011). Should investors include commodities in their portfolio after all? New evidence. *Journal of Banking and Finance*, 35, 2606-2626.
- Davidson, J. (2004). Moment and memory properties of linear conditional heteroscedasticity models, and a new model. *Journal of Business and Economic Statistics*, 22, 16-29.
- Diamandis, P. F. (2009). International stock market linkages: Evidence from Latin America. *Global Finance Journal*, 20, 13-30.
- Diebold, F. X., & Inoue, A., (2001). Long memory and regime switching. *Journal of Econometrics*, 105, 131-159.
- Engle, R.F., & Bollerslev, T. (1986). Modelling the persistence of conditional variances. *Econometric Reviews*, 5, 81-87.
- Forbes, K., & Rigobon, R. (2002). No contagion, only interdependence: measuring stock market comovements. *Journal of Finance*, 57, 2223-2261.
- Geweke, J.P., & Porter-Hudack, S. (1983). The estimation and application of long memory time series models. *Journal of Time Series Analysis*, 4, 221-238.
- Granger, C. W. J., & Hyung, N., 2004. Occasional structural breaks and long memory with an application to the S&P500 absolute stock returns. *Journal of Empirical Finance*, 11, 399-421.
- Hammoudeh, S., & Yuan, Y. (2008). Metal volatility in presence of oil and interest rate shocks. *Energy Economics*, 30, 606-620.
- Hammoudeh, S., Malik, F., & McAleer, M. (2011). Risk management of precious metals. *Quarterly Review of Economics and Finance*. 51, 435-441.
- Hammoudeh, S., Sari, R., & Ewing, B. (2009). Relationships among strategic commodities and with financial variables: a new look. *Contemporary Economic Policy*, 27, 251-269.
- Hammoudeh, S., Yuan, Y., McAleer, M., & Thompson, M. (2010). Precious metals-exchange rate volatility transmissions and hedging strategies. *International Review of Economics and Finance*, 20, 633-647..

- Heemskerk, M. (2001). Do international commodity prices drive natural resource booms? An empirical analysis of small-scale gold mining in Suriname. *Ecological Economics*, 39, 295-308.
- Hillebrand, E. (2005). Neglecting parameter changes in GARCH models. *Journal of Econometrics*, 129, 121-138.
- Hurvich, C. M., & Chen, W. W. (2000). An efficient taper for potentially overdifferentiated long-memory time series. *Journal of Time Series Analysis*, 21, 155-180.
- Inclan, C., Tiao, G. C. (1994). Use of cumulative sums of squares for retrospective detection of changes in variance. *Journal of the American Statistical Association*, 89, 913-923.
- Kang, S. H., Kang, S. M., & Yoon, S. M. (2009). Forecasting volatility of crude oil markets. *Energy Economics*, 31, 119-125.
- Kaufmann, T. D., Winters, R. A. (1989). The price of gold: A simple model. *Resources Policy*, 15, 309-313.
- Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics*, 54, 159-178.
- Lahiani, A., & Scaillet, O. (2009). Testing for threshold effect in ARFIMA models: application to US unemployment rate data. *International Journal of Forecasting*, 25, 418-428.
- Lee, H-Y., Wu, H-C., & Wang, Y-J. (2007). Contagion effect in financial markets after the South-East Asia Tsunami. *Research in International Business and Finance*, 21, 281-296.
- Lescaroux, F. (2009). On the excess co-movement of commodity prices - A note about the role of fundamental factors in short-run dynamics. *Energy Policy*, 37, 3906-3913.
- Lobato, I. N., & Savin, N. E. (1998). Real and spurious long memory properties of stock market data. *Journal of Business and Economic Statistics*, 16, 261-268.
- Markwat, T., Kole, E., & van Dijk, D. (2009). Contagion as a domino effect in global stock markets. *Journal of Banking and Finance*, 33, 1996-2012.
- Mikosch, T., & Stărică, C. (2004). Nonstationarities in financial time series, the long-range dependence, and the IGARCH effects. *Review of Economics and Statistics*, 86, 378-390.
- Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, 59, 347-370.
- Newey, W. K., & West, K. D. (1994). Automatic lag selection in covariance matrix estimation. *Review of Economic Studies*, 61, 631-654.
- Perron, P., & Qu, Z. (2007). An analytical evaluation of the log-periodogram estimate in the presence of level shifts and its implications for stock returns volatility. Working Paper, Boston University.
- Phillips, P. C. B., & Perron, P. (1988). Testing for a unit root in time series regression. *Biometrika*, 75, 335-346.
- Radetzki, M. (1989). Precious metals: The fundamental determinants of their price behaviour. *Resources Policy*, 15, 194-208.

- Robinson, P. M., & Hendry, D. (1999). Long and short memory conditional heteroscedasticity in estimating the memory parameter of levels. *Econometric Theory*, 15, 299-336.
- Rockerbie, D. W. (1999). Gold prices and gold production: Evidence for South Africa. *Resources Policy*, 25, 69-76.
- Sadorsky, P. (2006). Modeling and forecasting petroleum futures volatility. *Energy Economics*, 28, 467-488.
- Sari, R., Hammoudeh, S., & Soytas, U. (2009). Dynamics of oil price, precious metal prices, and exchange rate: Are there relationships. *Energy Economics*, 32, 351-362.
- Shimotsu, K. (2006). Simple (but effective) tests of long memory versus structural breaks. Working Paper, Department of Economics, Queen's University.
- Sjaastad, L. A., & Scacciavillani, F. (1996). The price of gold and the exchange rate. *Journal of International Money and Finance*, 15, 879-897.
- Sowell, F. (1992). Maximum likelihood estimation of stationary univariate fractionally integrated time series models. *Journal of Econometrics*, 53, 165-188.
- Soytas, U., Sari, R., Hammoudeh, S., & Hacihasanoglu, E. (2009). The oil prices, precious metal prices and macroeconomy in Turkey. *Energy Policy*, 37, 5557-5566.
- Tansuchat, R., Chang, C-L. & McAleer, M. (2009). Modelling long memory volatility in agricultural commodity futures. CIRJE-F-680. <http://hdl.handle.net/2261/32452>.
- Tully, E., & Lucey, B. M. (2007). A power GARCH examination of the gold market. *Research in International Business and Finance*, 21, 316-325.
- Van Dijk, D., Franses, P. H., & Paap, R. (2002). A nonlinear long memory model with an application to US unemployment. *Journal of Econometrics*, 110, 135-165.
- Watkins, C., & McAleer, M. (2008). How has volatility in metals markets changed? *Mathematics and Computers in Simulation*, 78, 237-249.
- Wei, Y., Wang, Y., & Huang, D. (2010). Forecasting crude oil market volatility: further evidence using GARCH-class models. *Energy Economics*, 32, 1477-1484.
- Young, J. H. (2011). A long memory conditional variance model for international grain markets. *Journal of Rural Development*, 31, 81-103.