

Forecasting the Volatility of Coffee Arabica and Crude Oil Prices with a High Frequency Data

Dawit Yeshiwas¹ Endalew Tesfa²

1.MSc. in Statistics, Department Of Statistics, Science College, Debre markos University, Debremarkos, Ethiopia

2.MSc. In Econometrics, Department Of Statistics, Science College, Debre markos University, Debremarkos, Ethiopia

Abstract

Modeling, analyzing, and forecasting volatility has been the subject of extensive research among academics, practitioners and portfolio managers. This paper estimates a variety of GARCH models using weekly closing price (in USD/barrel) of Brent crude oil and weekly closing prices (in USD/pound) of coffee Arabica, and compares the forecasting performance of these models based on a high frequency intra-day data which allows for a more precise realized volatility measurement. The study used weekly price data to explicitly model volatility, and employed high-frequency intra-day data to assess model forecasting performance. The analysis points to the conclusion that GARCH (1,1) for Arabica coffee and GARCH(1,2) crude oil returns were best models, respectively with Student's t distributed innovation terms is the most accurate volatility forecasting models in the context of our empirical setting. We recommend and encourage future researchers studying the forecasting performance of GARCH models to pay particular attention to the measurement of realized volatility, and employ high-frequency data whenever feasible.

Keywords: commodity price, volatility, forecasting, GARCH

DOI: 10.7176/JESD/10-23-04

Publication date: December 31st 2019

1 INTRODUCTION

1.1 Background of the study

The recent global financial crisis has increased the susceptibility of Commodity Dependent Developing Countries (CDDCs) to excessive price volatility in commodity markets. Moreover, structural weaknesses in these countries render their economies more vulnerable to increased commodity market turbulence than developed countries, given their comparatively lower income and high dependence on commodity exports. The World Bank estimates that 119 million additional people have been pushed into hunger as a result of the 2008 food price crisis (World Bank, 2009).

Modeling, analyzing, and forecasting volatility has been the subject of extensive research among academics, practitioners and portfolio managers (Baume et al, .2013). This has been used in risk management, derivative pricing and hedging, portfolio selection and policy making. Similarly the analysis of volatility spillovers between commodity and asset prices has a profound implication for risk management and portfolios maximization by the government and investors alike (Salisu and Oloko, 2015). In view of the current bearish behavior of oil price in the international markets, it is arguably of special interest to study the volatility of oil and coffee prices (Gil-Alana and Yaya, 2014).

In the last two decades there has been an explosion in volatility research. The "Dynamic Volatility Era" began with the introduction of the Autoregressive Conditional Heteroskedastic (ARCH) model (Engle, 2012). The ARCH (p) model expresses the conditional variance as p – th order weighted average of past (squared) disturbances and thus is able to describe volatility clustering in financial series. Following this, an enormous body of research has focused on extending and generalizing the ARCH model, mainly by providing alternative functional forms for the conditional variance. Some of the most important contributors to the dynamic volatility literature have been Engle, Bollerslev, Nelson and Ding. Bollerslev (1986) proposed the Generalized ARCH (GARCH) model, as a more parsimonious way of modeling volatility dynamics.

The most significant contribution of this extensive volatility research is that now we have a much clearer and better understanding of the probabilistic features of speculative returns data. This is of crucial importance in the search for an overarching framework, which will be able to first capture the empirical regularities adequately and ultimately give us reliable volatility forecasts. There are a number of extensive surveys of the literature including for instance Bollerslev (1990), Chou and Kroner (1992), Engle and Nelson (1994), and Ding and Engle (2001). While realized volatility models often demonstrate excellent forecasting performance, there is still much debate concerning optimal approaches. However, recent results reported by Hansen and Lunde (2005) have suggested that, in the context of exchange rate returns, nothing can beat a GARCH (1, 1) model.

Naturally, different papers in the literature tend to focus on different commodity price volatility. The focus of this work is on the volatility of Brent crude oil and Coffee Arabica prices in the commodity futures markets which are markets where one can buy specific quantities of a commodity at a specified price with delivery set at a

specified time in the future. The choice of these commodities is motivated by the fact that these are the most actively traded commodities in the world which also happen to be major importing and exporting commodities of Ethiopia. Existing work appears to focus mostly on in-sample modeling of volatility of commodity prices giving less attention to model forecasting performance. This work will also seek to bridge this gap in the literature.

1.2 Statement of the problem

A number of recent papers have investigated the volatility of energy and commodity price mainly due to increased interest for understanding the drivers and dynamics of such volatilities in these highly volatile times. Alternative estimation approaches are used in the literature, but thus far there is a dearth of work undertaking a thorough “GARCH” analysis –a statistical framework which is particularly suited for modeling asset price volatility. Moreover existing work appears to focus mostly on in-sample modeling of volatility of commodity price giving less attention to model forecasting performance. Accordingly, this work will seek to contribute to the literature by undertaking a systematic analysis of volatility forecasting performance of different GARCH models.

Most of the existing research on volatility spillovers employ statistical models in order to estimate realized volatilities which turned out to be oftentimes poor approximations of true volatilities. An attractive alternative to model-based statistical volatility is to compute realized volatility based on high frequency intra-day or ‘tick’ data. Realized volatility has been found to be more accurate than model based volatilities in predicting latent volatility (Christensen and Prabhala, 1998). Thus, the results provided by previous studies can be potentially misleading in that they may have underestimated or overestimated the extent of the true volatility.

This paper avoids this potential pitfall by using realized volatility based on a 30-minute intra-day data. We argue that doing so represents an important contribution to the literature of volatility measurement within commodity markets.

1.3 Objective of the study

1.3.1 General Objective

The general objective of this study is to model the volatility of the Arabica coffee and crude oil prices using GARCH type models and perform volatility forecasting using realized variance.

1.3.2 Specific Objectives

The specific objectives of this study are:

1. To fit a univariate GARCH models for each commodity prices separately.
2. To forecast the volatility of the two commodities
3. To compare the realized volatility forecasts obtained from weekly return square and a high frequency (a thirty minute interval) data.

2 METHODS

2.1 Stationarity

The concept of stationarity is fundamental in time series analysis. A time series $\{y_t\}$ is said to be *strictly stationary* if the joint distribution of $(y_{t_1}, y_{t_2}, \dots, y_{t_T})$ is identical to that of $(y_{t_1+k}, y_{t_2+k}, \dots, y_{t_T+k})$ for all t , where k is an arbitrary positive integer and (t_1, t_2, \dots, t_T) is a collection of T positive integers. In other words, strict stationarity requires that the joint distribution of $(y_{t_1}, y_{t_2}, \dots, y_{t_T})$ is invariant under time shift. Strict stationarity imposes a very strong condition and is hard to test empirically. Often the concept of weak stationarity is used. A time series $\{y_t\}$ is weakly stationary if both the mean of y_t and the covariance between y_t and y_{t-l} are time invariant, where l is an arbitrary integer. More specifically, $\{y_t\}$ is weakly stationary if (a) $E(y_t) = \mu$, which is a constant and (b) $cov(y_t, y_{t-l}) = \gamma_l$ which only depends on l . In practice, suppose that we have observed n data points $\{y_t \mid t = 1, \dots, n\}$. Weak stationarity implies that the time plot of the data would show that the values fluctuate with constant variation around a fixed level. In application, weak stationarity enables one to make inferences concerning future observations (e.g. Forecasting). The covariance $\gamma_l = cov(y_t, y_{t-l})$ is called the *lag - l* auto-covariance of y_t . It has two important properties: (a) $\gamma_0 = Var(y_t)$ and (b) $\gamma_{-l} = \gamma_l$. In the finance literature, it is necessary to test weakly stationarity of an asset return series. There are three methods of testing stationarity: graphical analysis, unit root test due to Dickey and Fuller (1979) test.

A. Graphical Analysis of the Series

Before pursuing formal tests, it is always advisable to plot the time series under study. Such a plot gives an initial clue about the likely nature of the time series. For instance, if a line-graph of a time series shows an upward trend, then this suggests perhaps that the mean of the data has been changing. This may be a clue that the series is not stationary. Such an intuitive feel is the starting point of more formal tests of stationarity.

B. Unit Root Test

A test for stationarity that has become widely popular over the past several years is the *unit root test*. The Augmented Dickey-Fuller (ADF) test is discussed below. To illustrate the use of Dickey-Fuller test, consider first

an autoregressive process of order one (AR (1)) process:

$$y_t = \omega + \alpha_1 y_{t-1} + \varepsilon_t \quad (1)$$

where ω and α_1 are parameters and ε_t is assumed to be white noise. y is a stationary series if $-1 < \alpha_1 < 1$. If $\alpha_1 = 1$, y is a non-stationary series (a random walk with drift); if the process is started at some point, the variance of y increases steadily with time and goes to infinity. If the absolute value of α_1 is greater than one, the series is explosive. Therefore, the hypothesis of a stationary series can be evaluated by testing whether the absolute value of α_1 is strictly less than one. The augmented Dickey-Fuller (ADF) test takes the unit root as the null hypothesis, $H_0: \alpha_1 = 0$. Since explosive series do not make much economic sense, this null hypothesis is tested against the one-sided alternative $H_a: \alpha_1 < 1$. The test is carried out by estimating an equation with y_{t-1} subtracted from both sides of the equation:

$$\Delta y_t = \omega + \delta y_{t-1} + \varepsilon_t \quad (2)$$

where $\delta = \alpha_1 - 1$, Δ is the first difference operator, and the null and alternative hypotheses are $H_0: \delta = 0$, $H_a: \delta < 0$.

While it may appear that the test can be carried out by performing a t -test on the estimated δ , the t -statistic under the null hypothesis of a unit root does not have the conventional t -distribution. Dickey and Fuller (1979) showed that the distribution under the null hypothesis is nonstandard, and simulated the critical values for selected sample sizes. Recently, MacKinnon (1991) has implemented a much larger set of simulations than those tabulated by Dickey and Fuller. In addition, MacKinnon estimates the response surface using the simulation results, permitting the calculation of Dickey-Fuller critical values for any sample size and for any number of right-hand variables. The simple unit root test described above is valid only if the series is an AR (1) process. If the series is correlated at higher order lags, the assumption of white noise disturbances is violated. The ADF test uses different methods to control for higher-order serial correlation in the series. It makes a parametric correction for higher-order correlation by assuming that the y series follows an AR (p) process and adjusting the test methodology:

$$y_t = \omega + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \dots + \alpha_p y_{t-p} + \varepsilon_t \quad (3)$$

The ADF approach controls for higher-order correlation by adding lagged difference terms of the dependent variable to the right-hand side of the regression:

$$\Delta y_t = \omega + \delta y_{t-1} + \gamma_1 \Delta y_{t-1} + \gamma_2 \Delta y_{t-2} + \dots + \gamma_{p-1} \Delta y_{t-p+1} + \varepsilon_t \quad (4)$$

Where, $\delta = (\sum_{i=1}^p \alpha_i) - 1$, $\gamma_j = -\sum_{k=j+1}^p \alpha_k$, Δ is the first difference operator, and the null and alternative hypotheses are $H_0: \delta = 0$, $H_a: \delta < 0$.

An important result obtained by Fuller is that the asymptotic distribution of the t -statistic on δ is independent of the number of lagged first differences included in the ADF regression. Moreover, while the parametric assumption that y follows an autoregressive (AR) process may seem restrictive, Said and Dickey (1984) demonstrate that the ADF test remains valid even when the series has a moving average (MA) component, provided that enough lagged difference terms are augmented to the regression.

2.2 Tests for autocorrelation

A weakly stationary series is not serially correlated (or is a white noise process) if the autocorrelation function is 0 for all $k > 0$. The ACF at lag k , denoted by ρ_k is defined as:

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{\text{Covariance at lag } k}{\text{variance}} \quad (5)$$

Since both covariance and variance are measured in the same units of measurement, ρ_k is unit less, or pure, number. It lies between -1 and 1. If we plot ρ_k against k , the graph we obtain is known as the population correlogram.

In practice it is only possible to compute the sample *autocorrelation function* (SACF), $\hat{\rho}_k$. To compute this, first computing the sample covariance at lag k , $\hat{\gamma}_k$, and the sample variance, $\hat{\gamma}_0$ is a must.

Therefore, the SACF at lag k is

$$\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0} \quad (6)$$

A plot of $\hat{\rho}_k$ against k is known as the sample *correlogram*.

Financial applications often require to test jointly that several autocorrelations of y_t are zero. the Portmanteau statistics (Q-statistics) is defined as

$$Q = n \sum_{k=1}^m \hat{\rho}_k^2 \quad (7)$$

Where n =sample size and m =lag length. The null hypothesis for this test is, all the ρ_k up to certain lags are simultaneously zero. In large samples, the Q statistic is approximately distributed as the *chi-square* distribution with m degree of freedom. In application, if the computed Q exceeds the critical value from the *chi-square*

square distribution at the chosen level of significance, the null hypothesis is rejected.

Ljung and Box (1978) modify the above Q-statistic as LB statistic (*Ljung-Box Q-statistic*) which is defined as:

$$LB = n(n + 2) \sum_{k=1}^m \left(\frac{\hat{\rho}_k^2}{n - k} \right) \sim \chi_m^2 \quad (8)$$

Although in large samples both Q and LB statistics follow the chi-square distribution with m degrees of freedom, LB statistic has been found to have better performance in small samples than the Q statistic.

2.3 The univariate GARCH model

First of all let the returns r_t be expressed as the change in logarithmic price over a certain period

$$r_t = \log \frac{p_t}{p_{t-1}} \quad (9)$$

Consider the following financial return model;

$$\begin{aligned} r_t &= \mu_t + \varepsilon_t, \quad \varepsilon_t = \sigma_t z_t & (i) \\ z_t &\sim \text{iid } N(0,1) & (ii) \\ \sigma_t^2 &= \alpha_0 + \alpha_1 \varepsilon_{t-1}^2, \quad \alpha_0 > 0, 0 \leq \alpha_1 < 1 & (iii) \end{aligned} \quad (10)$$

Where;

r_t –return series at time t ,

μ_t –the average return at time t ,

ε_t –the innovation term

The autoregressive AR (p), moving average MA (q) and autoregressive moving average ARMA (p, q) models are applicable when the innovation term (ε_t) maintains a constant variance (homoscedasticity). If the error term ε_t is conditionally heteroskedastic, then ARCH/GARCH models are applicable to model the conditional variance of the innovation term. Since the model involves just one lag of ε_t^2 (i.e. ε_{t-1}^2), it is known as ARCH(1) model. The generalized ARCH (here after, GARCH) model is a generalization of the ARCH model in that it includes lagged variances in the conditional variance equation. GARCH models have the advantage of capturing long lags in the shocks by using fewer parameters than ARCH models. The GARCH (1, 1) model is defined as:

$$r_t = \mu_t + \varepsilon_t \quad (11)$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \quad (12)$$

The GARCH model with p number of lagged variance term and q number of lagged squared error terms denoted as GARCH (q, p), and can be written as:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^q a_i \varepsilon_{t-i}^2 + \sum_{j=1}^p b_j \sigma_{t-j}^2 \quad \alpha_0 > 0, a_i \geq 0, b_j \geq 0 \quad (13)$$

Where the volatility term σ_{t-j}^2 denote the variance and j represents the number of lags, and the term ε_{t-i}^2 is the squared error for the period $t - i$. We can write the above equation more compactly:

$$\sigma_t^2 = \omega + \alpha(L) \varepsilon_t^2 + \beta(L) \sigma_t^2$$

where $\alpha(L) = \alpha_1 L + \dots + \alpha_q L^q$

$$\beta(L) = \beta_1 + \dots + \beta_p L^p$$

To ensure that the conditional variance is well defined in a GARCH (p, q) model all the coefficients in the corresponding linear ARCH (∞) should be positive rewriting the GARCH (p, q) modeled as an ARCH (∞):

$$\begin{aligned} \sigma_t^2 &= \left(1 - \sum_{i=1}^p \beta_i L^i \right)^{-1} + \left(\omega + \sum_{j=1}^q \alpha_j \varepsilon_{t-j}^2 \right) \\ &= \omega^* + \sum_{k=1}^{\infty} \phi_k \varepsilon_{t-k-1}^2 \end{aligned}$$

$\sigma_t^2 \geq 0$, if $\omega^* \geq 0$ and all $\phi_k \geq 0$. The non-negativity of ω^* and ϕ_k is also a necessary condition for the non-negativity of σ_t^2 (Rossi, 2004).

The a coefficients of lagged squared returns are interpreted as how fast the model react to, for example, market events, the b –coefficients of lagged conditional variance determines the degree of persistence in the volatility. A large value of a indicates that the conditional variance decays slowly, and that the volatility is persistent. On the other hand, if the a value is relatively higher than the b value, then the volatility is more extreme. The sum of a_i and b_j ($\sum_{i=1}^q a_i + \sum_{j=1}^p b_j$) should be less than one for the above GARCH (q, p) to be stationary. Nevertheless, one regions' volatility has impact on volatility of the other regions. Making estimations that include this contagion effect requires an extended model.

2.4 The model order selection: AIC and BIC

An important step before making the estimations is to determine the order of the models. Theoretically one can do this using the autocorrelation function, but in practice this may be difficult. A more formal way is to use an information criterion and choose the order that minimizes the criterion value. Two common criteria are the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). The formulas for these are

$$\begin{aligned} AIC &= -2 \times LLF + 2M \\ BIC &= -2 \times LLF + M \times LN(N) \end{aligned}$$

where N is the sample size, and M is the number of parameters. LLF is an abbreviation for log likelihood function. The reason for using both criteria is that the BIC is consistent but inefficient, and the AIC is the opposite, not consistent but efficient. No criteria are superior to others, but an overall assessment is needed based on the results showed by the criteria. For a model to be best it should have the smallest information criteria (Brooks, 2008).

2.5 Determining the conditional distribution

When fitting a GARCH-model based on financial data, the conditional distribution of the returns has to be defined. Studies, for example Bollerslev (1987), illustrate that returns are not normally distributed. Instead, the Student- t distribution captures the observed kurtosis in empirical returns in a more sufficient way than the normal distribution. Returns have excess kurtosis and fatter tails than the normal distribution. Therefore, the Student- t distribution is more suitable (Reider, 2009). There are three assumptions about the conditional distribution of the error term commonly employed when working with GARCH models: normal (Gaussian) distribution, student's t -distribution, and the generalized error distribution (GED).

2.6 The ARCH LM test

This test was introduced by Engle in 1982, and it investigates whether the conditional Volatilities of the chosen sample of data contain variation. The test can be used first to determine if there is volatility clustering in the data, and then to determine if there is any $ARCH$ –effects left in the residuals. Firstly, the estimated residuals, $\hat{\varepsilon}_t^2$ are regressed as:

$$\hat{\varepsilon}_t^2 = C + \left(\sum_{i=1}^q a_i \hat{\varepsilon}_{t-i}^2 \right) + v_t \quad (16)$$

and secondly, the hypotheses are

$$\begin{aligned} H_0: a_1 = a_2 = \dots = a_q = 0 \\ H_1: \text{at least one of } a_i \neq 0 \end{aligned}$$

If H_0 is true, then homoscedasticity characterizes the variance indicating no volatility clustering. The test statistic is

$$LM = nR^2 \sim \chi^2 \quad (17)$$

Where n is the number of observations and R^2 is computed from the regression (16) using estimated residuals. If n is large, LM is chi-squared distributed with q degrees of freedom.

2.7 Defining a proxy

In evaluating volatility forecasts the usual proxy for ‘true’ volatility is ‘ex post’ squared returns, or the squared errors. However, as noted by Andersen and Bollerslev (1998) and Andersen *et al.* (1999), although the use of squared returns or the squared errors is justifiable as an unbiased estimate of the volatility process, it provides a very noisy measure due to a large idiosyncratic term. Specifically, the returns innovation can be written as $\varepsilon_t = \sigma_t Z_t$ with Z_t an independent zero mean and unit variance stochastic process and σ_t is the volatility process. If the model for σ_t^2 is correctly specified, then the conditional expectation, $E(\varepsilon_t^2 | I_{t-1}) = E(\sigma_t^2 z_t^2 | I_{t-1}) = \sigma_t^2$, and the squared error is an unbiased estimate of the volatility process; however, it still contains the noisy idiosyncratic term, z_t^2 . This typically resulted in a poor performance, which instigated a discussion of the practical relevance of volatility models.

However, Andersen and Bollerslev (1998) showed that the ‘poor’ performance could be explained by the fact that the squared return is a noisy proxy for the conditional variance. By substituting the realized variance (instead of the squared return), Andersen and Bollerslev (1998) showed that volatility models perform quite well.

Hansen and Lunde (2003) provide another important argument for using the realized variance rather than the squared return. They show that substituting the squared returns for the conditional variance can severely distort the comparison, in the sense that the empirical ranking of models may be inconsistent for the true (population) ranking. So an evaluation that is based on squared returns may select an inferior model as the ‘best’ with a probability that converges to one as the sample size increases. For this reason, our evaluation is based on the realized variance. Building upon this line of research Andersen *et al.* (2003) define the so-called ‘realized volatility’

on day t as:

$$h_t^{rv} = \sum_{i=1}^N r_{t,i}^2 \quad (18)$$

Where;

N —is the number of equally spaced intervals within a day,

$r_{t,i}$ — is a logarithmic return on day t at time interval i ; with $i = 1, 2, \dots, N$ and $t = 1, \dots, T$.

Thus realized volatility is the sum of squared intra-day returns. In principle, letting N tend to large, i.e. continuous time sampling, the measure approaches the true integrated volatility of the underlying continuous time process and is theoretically free from measurement error. Further, this measure allows a market participant to essentially treat volatility as an observed variable and to allow direct estimation.

2.8 Forecast evaluation

2.8.1 The MZ regression

A popular way to evaluate volatility models out-of-sample is in terms of the R^2 from a Mincer –Zarnowitz (MZ) regression,

$$\sigma_t^2 = a + bh_t^2 + v_t, \quad (19)$$

that is, squared returns are regressed on the model forecasts of σ_t^2 and a constant.

Here:

σ_t^2 -is ex-post volatility (e.g. realized volatility) at time t , h_t^2 is estimated (in-sample) or forecasted (out-of sample) volatility at time t ,

v_t -independent and identically distributed; $v_t \sim N(0, 1)$. a and b are parameters to be estimated.

If the model for conditional variance is well specified, we should have: $a = 0$, $b = 1$. According to specific features of financial data series, the value of R^2 is usually low (even less than 5%) (Anderson *et al.*, 2009).

2.8.2 Mean Absolute Error (MAE)

Another way of determining the goodness of the estimations and forecasts is calculating the *MAE*. The approach is to measure how the received conditional covariance are close to their corresponding realized value. The formula is:

$$MAE = \frac{1}{n} \sum_{t=1}^n |\sigma_t^2 - h_t^2| \quad (20)$$

Where the proxy is used as σ_t^2 and the estimated conditional covariance is used as h_t^2 . By comparing the *MAE* between the estimated models, it can give an indication of which model that makes the best estimations.

2.8.3 Root Mean Square Error (RMSE)

The third measure is the Root Mean Square Error (RMSE), which is defined as

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (\sigma_t^2 - h_t^2)^2} \quad (21)$$

Using these methods, the estimated models can be compared, and using the same measurements for estimations and forecasts, one can determine if the relatively best estimations model also makes the best forecast.

3 .RESULTS

3.1 Data

The data considered in this paper were the weekly time series of **Brent crude oil** and **coffee Arabica** futures market closing price, given in US dollars per barrel and US dollars per pound, respectively. Both series contain data spanning between first week of January 2005 and last week of October, 2016 (a total of 616 observations) extracted from Bloomberg database. The full sample is split into two parts: in sample data, in order to estimate the parameters of models and out of sample data, in order to make forecasts. The in-sample period spans from first week of January 2005 up to last week of December, 2015 and the out-of-sample period spans from first week of January, 2016 through last week of October, 2016. For the out of sample period, we have also extracted a 30 minute intra-day data for realized volatility measuring purpose.

3.2 Test of stationarity and features of log return series

Our data consists of two commodity prices; crude oil, and Arabica coffee. **Fig. 3.1** is the trend charts of the two commodity prices in the study period. The two commodity prices present the same trend and direction during the entire study period. Price for both of the commodities have risen from 2005 up to 2008 and fallen from 2009 up to 2010 with a similar pattern. Another impression is that both commodity level series are non-stationary. To visualize the returns series for these two markets, we depict the return time series plots in **Fig.3.2**. The weekly return series display volatility persistence properties, indicating that large changes tend to be followed by large changes of both sign and small changes tend to be followed by small changes, and the differenced series suggests

stationarity.



Figure 3.1: Time series plot for level series, Arabica coffee and crude oil

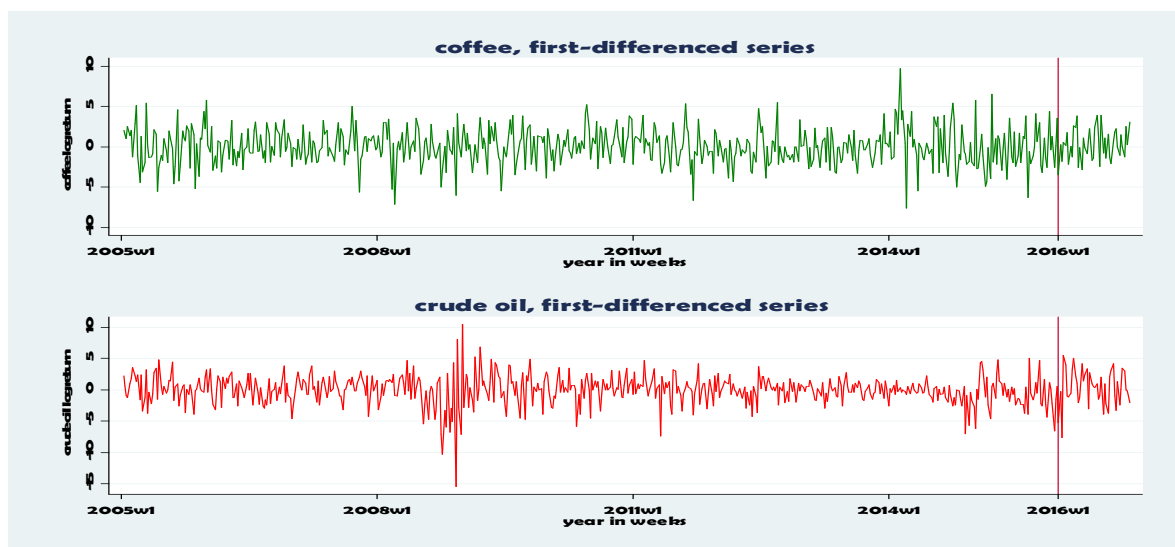


Figure 3.2 Time series plot for differenced series, Arabica coffee and crude oil

The level series plot in Fig 3.1 show that the level prices are non-stationary. Thus, the first logical step is to check the stationarity of these prices using unit root test. In this test the null hypothesis of unit root is rejected if the t-statistic is less than the critical value.

Table 3.1 summarize the DF unit root test results of the two commodity level prices. As can be seen in the table, the t-statistics are greater than the critical values, at 1%, 5% and 10% the significance. Hence, the null hypothesis of unit root would not be rejected, that is, there is a unit root problem in each of the data.

Table 3.1: DF Unit root test of stationarity for level prices, with trend

Series	t-statistic	1%critical value	5% critical value	10%critical value
Crude oil	-1.675	-3.960	-3.410	-3.120
Arabica coffee	-2.053	-3.960	-3.410	3.120

If a time series data is non-stationary, it is necessary to look for possible transformations that might bring stationarity. In practice, econometricians usually transform financial prices in to return forms. This is because often return series are found to be stationary such that analysis is possible. The log return series is obtained by:

$$r_t = 52 * \left(\log \frac{p_t}{p_{t-1}} \right)$$

where r_t is the log return series of the real price multiplied by 52 (which is simply a scaling factor), to

annualize as we are using weekly data and each year contains 52 weeks, and p_t is the real price series and log is the natural logarithm.

Table 3.2 summarize the DF unit root test of the log return series for each of the commodities. The table show that all the t-statistics are less than the critical values at all levels of significance. These indicates that the null hypothesis of unit root would be rejected in both cases. Hence, the log return series are stationary for each of the items as shown by the time series plots.

Table 3.2: DF Unit root test of stationarity for log Returns

Series	t-statistic	1%critical value	5%critical value	10%critical value
Crude oil	-5.791	-3.430	-2.860	-2.570
Arabica coffee	-5.877	-3.430	-2.860	-2.570

3.3 Volatility Modeling

To build a volatility model for the log return series, the first step is specifying the mean equation. Once we specify the mean equation we have to test for ARCH effects using the residuals of the mean equation. If ARCH effects are statistically significant, specifying a volatility model and carrying out a joint estimation of the mean and volatility equations is necessary. The conditional mean specification is, in general, arbitrary for GARCH models of the conditional volatility. Various modifications to the conditional means in GARCH models are possible (see, for example, Asai and McAleer (2003)).

3.3.1 Specification of Mean Equation

Based on two statistics, the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) we have chosen the temporary mean equation to estimate the joint mean and variance parameters. In most applications, lower order ARMA models, say, ARMA (1, 1), ARMA (1, 2), ARMA (2, 1) and ARMA (2, 2) are used. **Table 3.3** gives the alternative ARMA models together with their corresponding AIC and BIC.

Table 3.3: ARMA model selection for the weekly log returns using AIC & BIC

Lag length		Returns			
		Arabica Coffee		Crude Oil	
P	Q	AIC	BIC	AIC	BIC
0	1	2587.192	2600.234	2551.377	2572.674
0	2	2581.203	2598.593	2553.267	2570.656
1	0	2587.093	2600.135	2515.398	2564.44
1	1	2588.146	2605.535	2553.22	2570.609
1	2	2578.419	2600.156	2553.64	2575.377
2	0	2582.588	2599.977	2553.274	2570.664
2	1	2577.982	2606.049	2552.285	2568.674
2	2	2576.96	2599.719	2546.107	2564.419

As can be seen in Table 4.3, the best in-sample results are achieved by ARMA (2, 2) for both commodity log return series. Therefore, ARMA (2, 2) is the best (having minimum AIC & BIC) conditional mean equation for both of the two cases. Hence, the ARMA (2, 2) model given by the following expression is used to model the conditional variance as mean equation of the log return series of prices:

$$r_t = \mu + \rho_1 r_{t-1} + \rho_2 r_{t-2} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} \quad (22)$$

Where;

r_t –return series at time t , either of Arabica coffee or crude oil return, ARMA (2, 2) process

ε_t is an innovation term.

3.3.2 Test for ARCH Effects

Considering the above chosen financial returns model i.e. Equation (22) and on fitting this model, if there is no volatility clustering in each of the returns, the random disturbance term ε_t should be a white noise process. The standardized residual plot from Equation (22) can be an initial insight to judge the heteroskedastic characteristics of the error term. The standardized residual plot from each commodity return series, and is given below (**Figure 3.3**). The figure depicts the residual plots of the two commodity return series generated from the mean equation. We see from the figure that for both of the series there is a prolonged period of low volatility and prolonged period of high volatility. For example for Arabica coffee return there is a long period of low volatility from the first week of 2005 to the end of 2008 and also there exist a long periods of high volatility from first week of 2014 to the end week of 2016. Crude oil return also exhibits a prolonged period of low volatility than the coffee series, which is from the first week of 2005 to the first week of 2009. In other words periods of high volatility are followed by periods of high volatility and periods of low volatility tend to be followed by periods of low volatility, which is known as volatility clustering. This suggests that the residuals or error terms are conditionally heteroskedastic and can be represented by GARCH models.

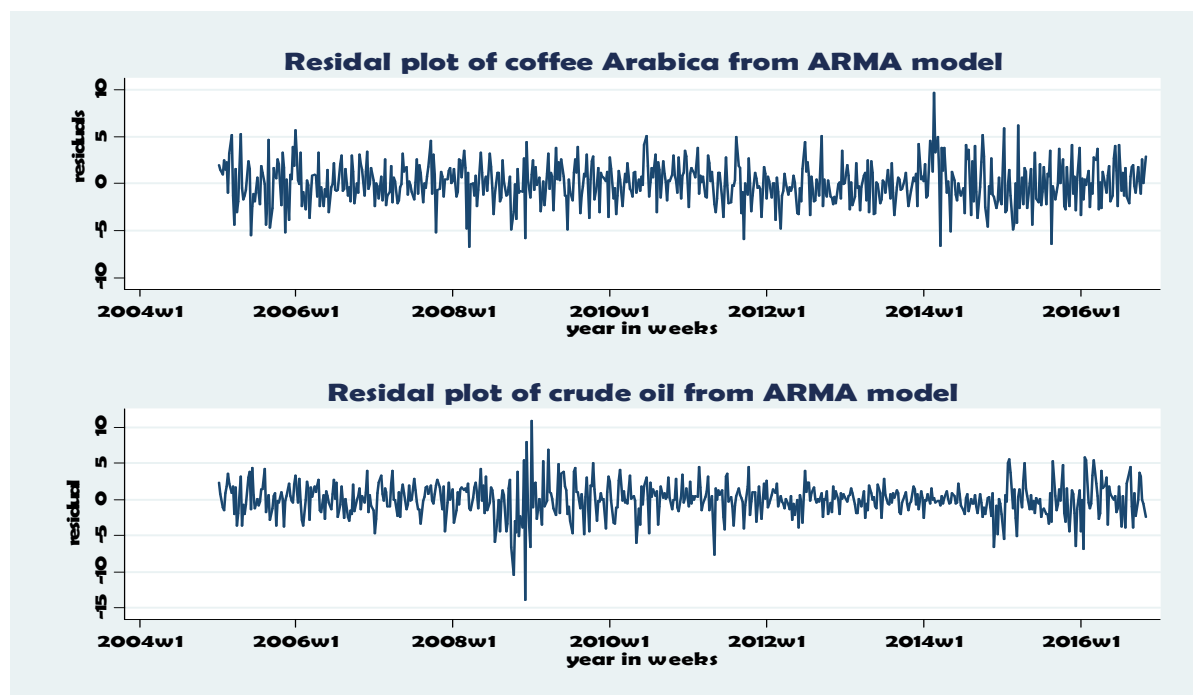


Figure 3.3: Residual plots of the Arabica coffee and crude oil return series from the mean

Based on the residuals from the mean equation chosen above, it is possible to test for the existence of ARCH effect which will allow continuing the analysis using GARCH model.

Table 3.4 shows the results of ARCH LM test for the two commodity returns. The last column of the table includes the p-values that indicate rejection of the null hypothesis that “there is no ARCH effect” up to the fourth lag at 5% level of significance. The results indicate that the two commodities price log return series are volatile and need to be modeled using GARCH models.

Table 3.4: ARCH LM Test Summary Statistics

Series	Obs* R^2	$\chi^2(4)$	Lag(p)	Prob.
Arabica coffee	12.6477	9.488	4	0.0040
Crude oil	105.6419	9.488	4	0.0000

3.3.3 GARCH Model Identification

To estimate and evaluate the forecasts of the competing GARCH models, different p and q values for the standard symmetric GARCH models are tested using different statistics such as Akaike Information Criterion (AIC) and Bayesian information Criterion (BIC) in order to choose the best model based on the in-sample data. The appropriate specification is chosen, and then for this specific p and q , the alternative GARCH models are estimated, tested and finally one GARCH model is chosen based on forecasting performance. For this GARCH (p , q)-model different distributions are evaluated. Namely, normal distribution, student’s t -distribution and generalized error distribution (GED). Finally the parameters for the chosen GARCH model is presented. Table 3.5 summarizes these two statistics computed from different GARCH models. Note that the AIC and BIC of the GARCH models are obtained by estimating the mean return and variance equations simultaneously.

Table 3.5: ARMA (2, 2)-GARCH model selection based on AIC and BIC

Arabica coffee, in sample output

Statistics	GARCH(1,1)	GARCH(1,2)	GARCH(2,1)	GARCH(2,2)
AIC	3290.974 (1)	3291.862 (2)	3292.442 (3)	3293.363 (4)
BIC	3308.363 (1)	3313.599 (2)	3314.179 (3)	3319.447 (4)
Rank Sum	2	4	6	8

Crude oil, in sample output

Statistics	GARCH(1,1)	GARCH(1,2)	GARCH(2,1)	GARCH(2,2)
AIC	3208.934 (4)	3204.773 (3)	3202.883 (2)	3200.631 (1)
BIC	3226.324 (2)	3226.510 (3)	3224.620 (1)	3226.716 (4)
Rank Sum	6	6	3	5

The row ranks indicate the ranking of the various GARCH models based on the two statistics. As a result, the lower the rank sum is the better the model. The best models for the two commodity series are:

- ❖ GARCH(1,1) for Arabica coffee
- ❖ GARCH(1,2) for crude oil

That is, the conditional variance processes are modeled by

$$\sigma_t^2 = \omega + a_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \quad \text{GARCH (1, 1)}$$

$$\sigma_t^2 = \omega + a_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 + \beta_2 \sigma_{t-2}^2 \quad \text{GARCH (1, 2)}$$

3.3.4. Forecast comparison to realized volatility measures

Below in **Table 3.6** and **Table 3.7** we present different forecast accuracy measures. We compare the realized volatility measures by taking the weekly return square and a high frequency (30-minute intra-day) data as realized volatility measure. We use the mincer-zarnowitz regression to our estimates. If the estimated parameters in this regression are as expected, it means *a* close to zero and *b* close to one and *R*² is not lower than 30% from Equation(3.35), indicating that the differences between the compared estimates are not huge and if so both of them are good estimates (Mincer, 1969).

Table 3.6: Forecast error measures, ARMA (2,2)-GARCH (1, 1) models under different distribution and realized volatility, Arabica coffee return.

Statistics	Realized volatility					
	Weekly return square			High frequency		
	N-GARCH	T-GARCH	GED-GARCH	N-GARCH	T-GARCH	GED-GARCH
MZ.reg .R2	0.0667	0.0712	0.010	0.156	0.230	0.155
MSE	21.030	22.6411	21.641	18.44	18.139	18.1755
RMSE	4.586	4.758	4.651	4.2941	4.25894	4.26234
MAE	4.576	4.753	4.581	3.5013	3.553	3.497
AIC	2540.605	2535.28	2536.791	2540.605	2535.28	2536.791
LogL.	-1261.393	-1264.303	-1259.64	-1261.393	-1264.303	-1259.64

Table 3.7: Forecast error measures, ARMA (2,2)-GARCH (1,2) model under different distribution and realized variance, Crude oil return

Statistics	Realized volatility					
	Weekly return square			High frequency		
	N-GARCH	T-GARCH	GED-GARCH	N-GARCH	T-GARCH	GED-GARCH
MZ.reg .R2	0.016	0.022	0.011	0.1763	0.267	0.18
MSE	126.6363	125.6408	125.6408	67.985	64.512	64.56
RMSE	11.2532	11.2089	11.2039	8.2453	8.0319	8.0349
MAE	7.6397	7.5618	7.5817	7.9126	7.739	7.762
AIC	2460.264	2443.797	2443.877	2460.264	2443.797	2443.877
LogL.	-1222.132	-1213.938	-1212.898	-1222.132	-1213.938	-1212.898

Forecast error measures and *R*² values from the Mincer-zarnowitz regression presented in Table 4.6 & Table 4.7 imply that the realized volatility measure from high frequency data is a better measure of unobserved volatility than the commonly used squares of returns (*R*² values are the lowest). This is also confirmed by the other forecast error measures having 44 one step ahead forecasts (2016w1-2016w44). The values of the **RMS** and **MAE** should get lower if the computing model is best performing. With regards to the distribution of the error term, the student's t-distribution combined with high frequency data is best performing for both series.

3.4 Estimations of the univariate GARCH

The values of the coefficients of the two commodities are shown below in **Table 3.8**. The estimation was done assuming that the error term comes from student's t-distribution with eight degree of freedom. From the table, it can be seen that some of the coefficients in both of the returns are not significant at 5% level of significance; for instance, the constant term is not significant in both cases.

Non-negativity rules: As we are modeling variance and the estimated variance must be non-negative, there is

no easy rule that ensures non-negativity for the general GARCH(p, q) process. But Nelson & Cao (1992) give easy checkable condition for GARCH(1,q) model and (somewhat) manageable conditions for GARCH(2,q) model. The non-negativity condition for GARCH (1, 1) and GARCH(2,1) model are:

$$\sigma_t^2 = \omega + a_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \quad \text{GARCH (1, 1)}$$

$$\sigma_t^2 = \omega + a_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 + \beta_2 \sigma_{t-2}^2 \quad \text{GARCH (1,2)}$$

For both of the models to be non-negative the following condition should be satisfied:

$$\omega > 0, \quad a_1 \geq 0, \quad \beta_1 \geq 0, \quad \beta_1 + \beta_2 < 1, \quad \text{and} \quad \beta_2 \geq 0$$

Looking the parameter estimates of ARCH and GARCH terms from the table below they satisfy the non-negativity rule described above for both commodity return series.

Table 3.8: parameter estimates from ARMA (2,2)-GARCH models

Returns	Coffee	Oil
Mean equation		
$\hat{\theta}_1$	-0.5158 (0.000)	-0.7636 (0.000)
$\hat{\theta}_2$	-0.4884 (0.000)	0.6585 (0.000)
$\hat{\rho}_1$	0.5645 (0.000)	0.780 (0.000)
$\hat{\rho}_2$	0.4362 (0.000)	-0.9325 (0.000)
Variance equation		
ω	0.1037 (0.38)	0.0317 (0.160)
$\hat{\alpha}$	0.0249 (0.0097)	0.0391 (0.002)
$\hat{\beta}_1$	0.9556 (0.000)	1.5861 (0.000)
$\hat{\beta}_2$		-0.6194 (0.000)
$\hat{\alpha} + \hat{\beta}$	0.9805	0.9988

**values in parenthesis are p-values at 5% level of significance

3.5 Discussion

The aim has been modeling and forecasting the volatility dynamics of weekly time series of **Brent crude oil** and **coffee Arabica** futures market closing price using GARCH models. From the preliminary analysis over the time period considered, both of the price series show an increasing trend. To determine whether the series are stationary or not, the Augmented Dickey-Fuller (ADF) test was carried out. Often, raw data of commodity prices are non-stationary, which was also the case in this study. For both level time series, the tests indicate to the existence of unit root I(1). The first log difference of each time series was considered as stationary (**Table 3.3**).

As a reference point for the analysis, a standard univariate GARCH models were fitted with ARMA(2,2) mean equation, and it is found that the variance equation for Arabica coffee and crude oil returns were GARCH(1,1) and GARCH(1,2), respectively.

In the two commodity returns it appears that the ARCH term coefficients were significant at 5% level of significance. The result is an implication of the presence of volatility clustering, i.e. large changes followed by large changes and small changes followed by small changes. In the case of Arabica coffee our result goes in line with the Hansen and Lunde (2003) findings. They conclude that GARCH (1,1) model with high frequency data is enough to capture the volatility of exchange rates return volatility, but in the case of crude oil our finding is something different.

This study suggests that using high frequency data results in a more accurate forecast values in both the standard univariate GARCH and multivariate GARCH models. This result agrees with David G. McMillan and Alan E. H. Speight (2010) inference in the context of exchange rate returns. In fact, they consider a daily GARCH (1,1) model and 5-minute intra-day return. In addition to the usual 'ex post' squared returns measure, they also construct realized volatility and two versions of bias-corrected realized volatility. In order to evaluate these forecasts they utilize the Mincer-Zarnowitz regression test of predictive power, comparing forecasting performance on the basis of the R^2 values obtained. In our analysis the procedure is similar, but we use a weekly return series and a 30-minute interval intra-day return series and we stretch their conclusion.

4 CONCLUSION AND RECOMMENDATIONS

This study estimates a variety of GARCH models using weekly closing price (in USD/barrel) of Brent crude oil and weekly closing prices (in USD/per pound) of coffee Arabica, and compares the forecasting performance of these models based on a high frequency intra-day data which allows for a more precise realized volatility measurement. The analysis points to the conclusion that for Arabica coffee and crude oil returns GARCH (1,1) and GARCH(1,2) were best models, respectively with Student's t distributed innovation terms is the most accurate volatility forecasting models in the context of our empirical setting.

We recommend and encourage future researchers studying the forecasting performance of MGARCH models to pay particular attention to the measurement of realized volatility, and employ high-frequency data whenever feasible.

We also recommend that public policy makers interested in foreseeing the price volatility of these two major commodities in the context of the Ethiopian economy consider using the information documented in this study as input in their deliberations given that it is based on some robust econometric work and highly appropriate data.

The scope of the analysis in this study has been limited to the volatility of the two commodities. In order to overcome this limitation and provide a more nuanced analysis, it might be profitable for future researchers to consider incorporating stock, currency and bond markets volatilities into the analysis. The potential complexity of such a research agenda notwithstanding, the results are likely to be rewarding in light of the deeper integration between global financial and commodity markets in recent years, a phenomenon which came to be known as the financialization of commodities.

Author conflict of interest: The authors declare that there is no conflict of interest regarding the publication of this paper.

Funding: this work has not received any specific grant, the authors are Debre markos university employees,

Acknowledgment

First and foremost I thank the Almighty God. I also thank you my Friends and colleagues for helping me in one or another way for the completion of this work

REFERENCES

- Andersen T., Bollerslev T., Diebold F., [2003], *Parametric and nonparametric volatility measurement*.
- Bauwens, L., S. Laurent. and J.V.K. Rombouts (2013). *Multivariate GARCH Models: A Survey*, CORE Discussion Paper 2003/31.
- Bollerslev, Tim, (1990), *modelling the Coherence in Short-run Nominal Exchange Rates: A Multivariate Generalized ARCH Model*, The Review of Economics and Statistics, MIT Press, vol. 72(3), pages 498-505, August.
- Bollerslev, T., R.F. Engle & J.M. Wooldridge, (1986), A Capital Asset Pricing Model with Time-Varying Covariance, *Journal of Political Economy* 96, 116 -131.
- Bollerslev, T. (1987). "A Conditionally Heteroskedastic Time Series Model for Speculative Prices and Rates of Return", *Review of Economics and Statistics*, 69, 542-547.
- Brooks, Chris (2008), *Introductory Econometrics for Finance*, Second Edition, Cambridge University Press.
- Christensen, B. J., & Prabhala, N. R. (1998). The relation between implied and realized volatility. *Journal of Financial Economics*, 50(2), 125-150.
- 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.
- Ding, Z. and R.F. Engle (2001). "Large Scale Conditional Covariance Matrix Modeling, Estimation and Testing", *Academia Economic Papers*, 29, 157-184.
- Engle, R. F., and Gallo, G. M. (2001). A multiple indicators model for volatility using intraday data. *Journal of Econometrics* 131(1), 3-27.
- Gil-Alana, L. A. and Yaya, O. S. (2014). The relationship between oil prices and the Nigerian stock market: An analysis based on fractional co integration. *Energy Economics*, 46: 328-333.
- Hansen PR, Lunde A. (2005). A forecast comparison of volatility models: does anything beat a GARCH (1,1)? *Journal of Applied Econometrics* 20: 873-889.
- Hansen PR, Lunde A, Nason JM. (2003). *choosing the best volatility models: the model confidence set approach*. *Oxford Bulletin of Economics and Statistics* 65: 839-861.
- MacKinnon, J.G. (1991). Critical Values for Co integration Tests, Chapter 13 in Long-run.
- Nelson DB. (1991). Conditional heteroscedasticity in asset returns: a new approach. *Econometrica* 59: 347-370.
- Reider, Rob (2009), Volatility Forecasting I: GARCH Models URL: http://cims.nyu.edu/~almgren/time-series/Vol_Forecast1.pdf (downloaded 25-04-2017).
- Rossi, E. (2004): *Lecture Notes on GARCH Models*, University of Pavia.
- Said, S. E. and. Dickey, D. A. (1984): Testing for Unit Roots in Autoregressive Moving Average Models of Unknown Order, *Biometrika*, 71, 599-607

-
- Salisu, A.A. and Oloko, T.F. (2015). Modeling oil price–US stock nexus: A VARMA–BEKK–AGARCH approach. *Energy Economics*, 50: 1–12.
- World Bank (2009). Global Economic Prospects; commodities at the crossroad. Report to the 12th session of the Human Rights Council.