Forecasting crude-oil market volatility: Further evidence with jumps
Amélie Charles, Olivier Darné

To cite this version:

HAL Id: hal-01598141
https://audencia.hal.science/hal-01598141
Submitted on 28 Dec 2020

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.
Forecasting crude-oil market volatility: Further evidence with jumps

Amélie Charles, Audencia Business School
Olivier Darné, LEMNA, Université de Nantes

This paper analyzes volatility models and their forecasting abilities in the presence of jumps in two crude-oil markets - Brent and West Texas Intermediate (WTI) - between January 6th 1992 and December 31st 2014. We compare a number of GARCH-type models that capture short memory as well as asymmetry (GARCH, GJR-GARCH and EGARCH), estimated on raw returns, to three competing approaches that deal with the presence of jumps: GARCH-type models estimated on jump-filtered returns, and two new classes of volatility models, called Generalized Autoregressive Score (GAS) and Markov-switching multifractal (MSM) models, estimated using raw returns. The forecasting performance of these volatility models is evaluated using the model confidence set approach, which allows us to identify a subset of models that outperform all the other competing models. We find that asymmetric models estimated on filtered returns provide better out-of-sample forecasts than do GARCH-, GAS-type and MSM models estimated on raw return series for Brent and WTI returns.

1. Introduction

The price of crude oil is one of the most important global economic indicators, and is monitored by policy-makers, producers, consumers and financial-market participants. From the end of the 1990s oil prices rose steadily, reflecting the rising demand for crude oil, particularly from developing nations. With the Great Recession the demand for energy shrank in late 2008, with oil prices falling from $147 in July 2008 to $32 in December 2008 and then stabilizing in October 2009. From 2010 until mid-2014, world oil prices were fairly stable at around $110. But since June 2014 prices have more than halved, falling below $30 in January 2016. Oil prices have thus been very volatile, changing their trajectories and behavior with respect to the economic situation. Crude-oil prices may significantly affect economic development, social stability and even national security in a country (Wu and Zhang, 2014).

Given the volatility of crude-oil prices, financial-market participants and policy-makers would benefit from a better understanding of how shocks affect volatility over time. The understanding of crude-oil price volatility is important for pricing financial assets, implementing hedging strategies and assessing regulatory proposals to restrict international capital flows. For example, changes in volatility can affect the risk exposure of both the producers and industrial consumers of oil. These changes may feed through to their respective investments in oil inventories and facilities for production and transportation.

Autoregressive conditionally heteroscedastic (ARCH) models, introduced by Engle (1982) and extended to generalized ARCH (GARCH) by Bollerslev (1986), and improved GARCH-type models were developed to capture the most important stylized facts regarding crude-oil returns: heavy-tailed distributions, volatility clustering, asymmetry and long memory volatility (Cheong, 2009; Kang et al., 2009; Mohammadi and Su, 2010; Wei et al., 2010; Arouri et al., 2012; Hou and Suardi, 2012; Salisu and Fasanya, 2013). However, it is well-known that oil markets are subject to a number of drastic shocks (called large shocks, outliers or jumps), such as the invasion of Kuwait by Iraq, Operation Desert Storm, Operation Desert Fox and the Global Financial Crisis (Larsson and Nossman, 2011; Lee et al., 2010; Salisu and Fasanya, 2013; Charles and Darné, 2014), as...
well as OPEC announcements on reductions in production (Lin and Tamvakis, 2010; Demirer and Kutan, 2010). These shocks may pose difficulties for the identification and estimation of GARCH models governing the conditional volatility of returns, and thus out-of-sample volatility forecasts (e.g., Franses and Ghijsels, 1999; Carrero et al., 2007, 2012; Charles, 2008).\(^1\) Harvey and Chakravarty (2008), Harvey (2013) and Creal et al. (2013) have recently and independently proposed a modification to the GARCH model derived from the conditional score of the assumed distribution. The two former called this a Dynamic Conditional Score (DCS) model and the latter a Generalized Autoregressive Score (GAS) model. Harvey and Sucarrat (2014) showed that these models help deal with outliers. Another approach to deal with the outliers is the Markov-switching multifractal (MSM) model of Calvet and Fisher (2001, 2004), which captures the multiscaling behavior or multifractality in financial data. Wang et al. (2016) found that the MSM model can accommodate the presence of outliers.\(^2\) Another way of tackling large shocks is to detect them and estimate models on filtered data. GARCH-type models estimated using filtered returns deliver more accurate out-of-sample forecasts of the conditional variance of stock returns than do GARCH models estimated on raw data (e.g., Charles, 2008; Laurent et al., 2016).

To the best of our knowledge, no work to date has taken the presence of outliers in crude-oil markets into account when forecasting their volatility with GARCH-type models. Cheong (2009), Kang et al. (2009), Mohammadi and Su (2010), Wei et al. (2010), Arouri et al. (2012), Hou and Suardi (2012) and Kang and Yoon (2013) do not account for jumps when they model and forecast crude-oil volatility. In this paper we deal with jumps in two ways. First, we detect jumps in crude-oil returns using a new semi-parametric test for additive jumps in GARCH models proposed by Laurent et al. (2016), and then apply GARCH-type models to the filtered returns.\(^3\) Second, we follow the approach in Harvey and Chakravarty (2008), Harvey (2013), Creal et al. (2013) and Calvet and Fisher (2001, 2004), and apply GAS and MSM models to raw returns. Laurent et al. (2016) and Blazsek and Villatoro (2015) are the only contributions to date that compare GARCH and GAS models when forecasting financial returns. Recently, Wang et al. (2016) and Lux et al. (2016) show that MSM models generate more accurate volatility forecasts than GARCH-type models when forecasting crude oil returns.

As discussed by Lopez (2001), it is not obvious which loss function or criterion is the most appropriate for the evaluation of volatility models, and different loss functions may play different roles in practical applications. Most of the work on oil forecasting uses tests to evaluate forecast performance across various volatility models (see Table 1), such as equality predictive ability (EPA) tests, which are based on the pairwise comparison of forecast performances (Diebold and Mariano, 1995; White, 2000), and the superior predictive ability (SPA) test that allows for multiple comparisons (Hansen, 2005). In this paper, we apply the model confidence set (MCS) procedure proposed in Hansen et al. (2011) to determine the set that consists of a subset of equivalent models, in terms of SPA, which are superior to any other set of competing models. This MCS yields a number of models with equivalent forecasting performance, and therefore produces more robust forecasting than that based on one model only.

Following on from the discussion above, this paper extends previous research in two ways. We first compare the forecasting performance of a number of GARCH-type models (GARCH, GJR-GARCH and EGARCH), which capture short memory as well as asymmetry, estimated on both raw and jump-filtered returns, and GAS-type and MSM models estimated on raw returns, with squared returns used as a proxy for actual volatility. Second, a rolling forecast study is carried out to see how well the volatility models forecast the raw returns, using the model confidence set (MCS) approach proposed by Hansen et al. (2011).

Overall, we find that asymmetric models estimated on filtered returns provide better out-of-sample forecasts than do GARCH-type, GAS-type and MSM models estimated on raw return series for Brent and WTI returns. This finding confirms those in, for example, Carnero et al. (2007) and Charles (2008) that the standard-GARCH models estimated on filtered returns outperform GARCH-type models estimated on raw data, and also Laurent et al. (2016) that standard-GARCH models estimated on filtered returns outperform other models, including GAS-type models. Taking into account volatility jumps may therefore improve forecasting in crude-oil markets.

The remainder of this article is organized as follows. A selected literature review appears in Section 2. Section 3 then briefly describes the GARCH-type, GAS-type and MSM models, as well as the procedure for detecting jumps. The data set is presented in Section 4, and the empirical estimates of volatility models in Section 5. Section 6 presents the out-of-sample forecasting design and the results. Finally, Section 7 concludes.

2. Literature review

A number of contributions have examined the modeling and forecasting of crude-oil spot and futures price volatility, with most taking a GARCH-based approach (see Table 1). Despite extensive work on identifying the most appropriate GARCH model that provides the best out-of-sample forecasting performance, no model has consistently outperformed the others.

Kang et al. (2009) examine the forecasting power of competitive GARCH-volatility models, namely GARCH, IGARCH, FIGARCH and component-GARCH (CGARCH), in three crude-oil markets (WTI, Brent and Dubai). They demonstrate the superiority of FIGARCH and CGARCH over GARCH and IGARCH in modeling and forecasting the persistence of oil-price volatility. Cheong (2009) considers different GARCH specifications, GARCH, APGARCH, FIGARCH, FIEGARCH and FIAPARCH, allowing for both Normal and Student-\(t\) distributions of WTI and Brent oil returns. He finds that the simple GARCH model fits the Brent crude-oil data better than do GARCH models that include asymmetric and/or long memory volatility. Wei et al. (2010) extend the empirical framework of Kang et al. (2009) and estimate both linear and nonlinear GARCH regressions, including RiskMetrics, GARCH, IGARCH, GJR, EGARCH, APARCH, FIGARCH, FIAPARCH and HYGARCH models. None of these models has greater predictive accuracy than the others. Arouri et al. (2012) find that a volatility model with both long memory and structural change (a FIGARCH model with breaks) forecasts best when there are structural breaks. Hou and Suardi (2012) compare a nonparametric GARCH model to an extensive class of parametric GARCH models, as in Kang et al. (2009), in the Brent and WTI markets and show that the out-of-sample volatility forecast of the nonparametric GARCH model is superior to that of parametric GARCH models.

Wang et al. (2016) compare the Binomial MSM model proposed by Calvet and Fisher (2001, 2004) to GARCH-type models, such as GJR, EGARCH and FIGARCH models, in the Brent and WTI markets. They find that MSM models generate more accurate volatility forecasts than either GARCH-type models or the historical volatility model. Lux et al. (2016) compare different types of MSM models.

\(^1\) Harvey and Chakravarty (2008), Harvey (2013) and Harvey and Sucarrat (2014) show that an extreme event in a GARCH model will have a large slowly-decaying effect on future volatility predictions.

\(^2\) We thank an anonymous referee for suggesting the MSM model to deal with the outliers.

\(^3\) Arouri et al. (2012) use GARCH-type models with structural breaks. We also detect the presence of breaks, but do not find variance changes in the filtered return series. Charles and Darné (2014) suggest that crude-oil markets seem to be more affected by outliers and patches of outliers than by structural breaks.
<table>
<thead>
<tr>
<th>Article</th>
<th>Data</th>
<th>Sample</th>
<th>Models</th>
<th>Distrib.</th>
<th>Volatility measures</th>
<th>Horizon</th>
<th>Step</th>
<th>Out-of-sample loss function</th>
<th>Tests</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wang et al. (2016)</td>
<td>Brent, WTI (spot)</td>
<td>January 4th 1993 – September 9th 2013 (D)</td>
<td>GJR, EGARCH, FIGARCH</td>
<td>Normal</td>
<td>Squared returns</td>
<td>Last 4141 days</td>
<td>1, 20 days (rolling)</td>
<td>MAE, MSE, QLIKE, RZLOG, VaR based function</td>
<td>DM, SPA Reality check</td>
</tr>
<tr>
<td>Lux et al. (2016)</td>
<td>WTI (spot)</td>
<td>January 2nd 1985 – 24th March 2014 (D)</td>
<td>GARCH, IGARCH, RiskMetric, EGARCH, GJR-GARCH, MS-GARCH, APARCH, FIGARCH, HYGARCH, FIAPARCH, BMSM, LMSM</td>
<td>Normal</td>
<td>Squared returns</td>
<td>2011–2014</td>
<td>1, 5, 10, ..., 100 days (rolling)</td>
<td>MSE, MAE, HMSE, QLIKE, RZLOG</td>
<td>SPA</td>
</tr>
</tbody>
</table>

Notes: DM = Diebold-Mariano test. SPA = Superior Predictive Ability. MCS = Model Confidence Set. SB-GARCH and SB-FIGARCH denote GARCH and FIGARCH models, respectively, with structural breaks. NPGARCH = Nonparametric GARCH. MS-GARCH = Markov-Switching GARCH. BMSM = Binomial MSM. LMSM = LogNormal MSM. Note that Lux et al. (2016) examine a second subsample, covering the period from January 2, 1875 to December 31, 1895.
3. Methodology

3.1. GARCH-type models

In this paper we present three (linear and nonlinear) GARCH-type models.\(^4\)

3.1.1. The GARCH model

The GARCH model was developed independently by Bollerslev (1986) and Taylor (1986), and allows the conditional variance to depend on previous own lags. Consider the returns series \( r_t \) defined by \( r_t = \log P_t - \log P_{t-1} \), where price at time \( t \), \( P_t \), is described by a Normal ARMA\((p,q)\)-GARCH\((1,1)\) model

\[ r_t = \mu_t + \epsilon_t \quad \mu_t = \mu + \sum_{i=1}^{\infty} \xi_i \epsilon_{t-i} \] \hfill (1)

\[ \epsilon_t = \alpha_0 z_t, \quad z_t \sim i.i.d.N(0,1). \] \hfill (2)

\[ \alpha_t^2 = \omega + \alpha \epsilon_{t-1}^2 + \beta \alpha_{t-1}^2 \] \hfill (3)

where \( \xi_k \) are the coefficients of \( \xi(L) = \phi(L) \Theta(L) = 1 + \sum_{i=1}^{\infty} \xi_i L^i \), \( L \) is the lag operator, and \( \phi(L) \) and \( \Theta(L) \) are the AR and MA polynomials of order \( p \) and \( q \) respectively. The parameters should satisfy \( \omega > 0, \alpha \geq 0 \) and \( \beta > 0 \) to guarantee the positivity of the conditional variance.

The stationarity of the process (the second-order moment condition) is guaranteed by the restriction \( \alpha + \beta < 1 \). Ling and McAleer (2002a, 2002b) derived the regularity conditions of a GARCH\((1,1)\) model as follows: \( E[\epsilon_t^2] = \frac{\omega}{1-\alpha-\beta} < \infty \) if \( \alpha + \beta < 1 \), and \( E[\epsilon_t^4] < \infty \) if \( k\alpha^2 + 2\alpha^2 + \beta^2 < 1 \), where \( k \) is the conditional fourth moment of \( \zeta \).\(^5\) Ng and McAleer (2004) underline the importance of checking these conditions. The sum of \( \alpha \) and \( \beta \) reflects the persistence of shocks to the conditional variance, meaning that the effect of a volatility shock vanishes over time at an exponential rate.

3.1.2. The GJR-GARCH model

The GJR-GARCH model developed by Glosten et al. (1993) was developed to capture asymmetric volatility. The specification of the conditional variance of a GJR-GARCH\((1,1)\) model is

\[ \alpha_t^2 = \omega + \alpha \epsilon_{t-1}^2 + \gamma l_{t-1} \epsilon_{t-1}^2 + \beta \alpha_{t-1}^2 \] \hfill (4)

\[ = \omega + (\alpha + \gamma l_{t-1}) \epsilon_{t-1}^2 + \beta \alpha_{t-1}^2 \]

where \( l_{t-1} = 1 \) if \( \epsilon_{t-1} < 0 \), and 0 otherwise. Volatility is positive if \( \omega > 0, \alpha, \gamma, \beta \geq 0 \) and \( \alpha + \gamma + \beta > 0 \). The process is stationary if \( \alpha + \beta + (\gamma/2) < 1 \). He and Teräsvirta (1999) and Ling and McAleer (2002b) derived the regularity conditions for a GJR-GARCH\((1,1)\) as follows: \( E[\epsilon_t^2] < \infty \) if \( \alpha + \beta + \gamma/2 < 1 \), and \( E[\epsilon_t^4] < \infty \) if \( k\alpha^2 + 2\alpha + \beta^2 + 2\beta^2 + \beta \gamma + k\gamma + k\gamma^2 < 1 \).\(^6\) If the asymmetry coefficient \( \gamma \) is greater than zero then volatility rises more after large negative shocks than after large positive shocks. The GJR-GARCH model reduces to the GARCH model when \( \gamma = 0 \).

3.1.3. The EGARCH model

Nelson (1991) proposes the exponential GARCH model (EGARCH), which can also capture volatility leverage. The EGARCH\((1,1)\) model is defined as follows

\[ \ln(\alpha_t^2) = \omega + g(z_{t-1}) + \beta \ln(\epsilon_{t-1}^2) \] \hfill (5)

\[ g(z_t) = \theta_1 \cdot (|z_t| - E(|z_t|)) + \theta_2 z_t \]

where \( \theta_1 (|z_t| - E(|z_t|)) \) denotes the magnitude effect, and \( \theta_2 z_t \) the sign effect. If \( \theta_2 < 0 \), negative shocks then produce greater volatility than positive shocks of the same size. \( E(|z_t|) \) depends on the assumptions made about the unconditional density of \( z_t \). For the Normal distribution, \( E(|z_t|) = \sqrt{2\pi} \). The specification of volatility in terms of logarithmic transformations implies that the model parameters are not restricted to be positive. As the EGARCH model does not impose positivity restrictions on the volatility coefficients, a sufficient condition for stationarity is \( |\beta| < 1 \). He et al. (2002) consider the fourth-moment structure of the EGARCH\((1,1)\) model.\(^7\)

3.2. Jump detection in GARCH models

We apply the semi-parametric procedure to detect jumps proposed by Laurent et al. (2016) (LLP hereafter). Their test is similar to the non-parametric tests for jumps in Lee and Myklond (2008) and Andersen et al. (2007) for daily data.

LLP assume that the returns \( r_t \) are described by the ARMA\((p,q)\)-GARCH\((1,1)\) model of Eqs. (1) – (3). Consider the return series with an independent jump component \( a_i h \), defined as

\[ r_t^* = r_t + a_i h \] \hfill (6)

where \( r_t^* \) denotes the observed returns, \( h_i \) is a dummy variable for a jump on day \( t \), and \( q_i \) is the jump size. In Eq. (6) a jump \( a_i h \) will not affect \( \alpha_{t-1}^2 \) (the conditional variance of \( r_{t-1}^* \)), so that we can have non-Gaussian fat-tailed conditional distributions of \( r_t^* \).

LLP then use the bounded innovation propagation (BIP)-ARMA proposed by Muler et al. (2009) and the BIP-GARCH\((1,1)\) of Muler and Yohai (2008) to obtain robust estimates of \( \mu_t \) and \( \sigma_t^2 \), respectively, in Eqs. (1) and (3). These are denoted by \( \mu_o \) and \( \sigma_o^2 \) and are robust to

---

\(^4\) See Bauwens et al. (2012) for a survey of volatility models.

\(^5\) Under the assumption of a Normal distribution \( k = 3 \), so the condition becomes \( 3\alpha^2 + 2\alpha z_i^2 + \beta^2 < 1 \). See Ling and McAleer (2002a, 2002b) for other distributions.

\(^6\) Under a Normal distribution and a Student-\( t(\nu) \) distribution, with \( \nu > 5, \delta = \frac{1}{\nu} \).

\(^7\) The asymptotic properties, in particular the asymptotic Normality, of the quasi-maximum likelihood estimator (QMLE) hold under mild conditions for GARCH models (e.g., Lee and Hansen, 1994; Berkes et al., 2003; Francq and Zakoïan, 2004) and for GJR-GARCH models (Hamadeh and Zakoïan, 2011). However, the statistical properties for the QMLE of the EGARCH\((1,1)\) parameters are not available in general, but rather only for special cases under highly-restrictive conditions (Wintenberger, 2013; Kyriakopoulos, 2015; Martinet and McAleer, 2016).
potential jumps $a_{it}$ (i.e. they are estimated on $r_t^*$ and not on $r_t$). The BIP-ARMA and BIP-GARCH(1,1) are defined as

$$
\tilde{\mu}_t = \mu + \sum_{i=1}^{\infty} \xi_i \tilde{\alpha}_{t-1} \omega_{k_{mpv}} (\tilde{J}_{t-1}) \tag{7}
$$

$$
\tilde{\sigma}_t^2 = \omega + \alpha_1 \tilde{\sigma}_{t-1}^2 \omega_{k_{mpv}} (\tilde{J}_{t-1}) + \beta_1 \tilde{\sigma}_{t-1}^2 \tag{8}
$$

where $\xi_i$ are the coefficients from the AR($p$) and MA($q$) polynomials defined in Eq. (1), $\omega_{k_{mpv}}(\cdot)$ is the weight function, and $\omega_k$ a factor ensuring that the conditional expectation of the weighted square of unexpected shocks equals the conditional variance of $r_t$ in the absence of jumps (Boudt et al., 2013).

Consider the standardized return on day $t$ given by

$$
\tilde{J}_t = \frac{r_t^* - \tilde{\mu}_t}{\tilde{\sigma}_t} \tag{9}
$$

LLP detect the presence of jumps by testing the null hypothesis $H_0 : a_{it} = 0$ against the alternative $H_1 : a_{it} \neq 0$. The null is rejected if

$$
\max_j \tilde{J}_t > g_{l,x} \quad t = 1, \ldots, T \tag{10}
$$

where $g_{l,x}$ is the suitable critical value. If $H_0$ is rejected a dummy variable is defined as follows

$$
\tilde{I}_t = I(\tilde{J}_t > k) \tag{11}
$$

where $I(\cdot)$ is the indicator function, with $\tilde{I}_t = 1$ when a jump is detected at time $t$. LLP show that their test does not suffer from size distortions irrespective of the parameter values of the GARCH model in Monte Carlo simulations. The filtered returns $\tilde{r}_t$ are obtained as follows

$$
\tilde{r}_t = r_t^* - (\tilde{r}_t^* - \tilde{\mu}_t) \tilde{I}_t \tag{12}
$$

Laurent et al. (2016) extend this test for additive jumps in ARMA-GARCH models, called BIP-ARMA-BIP-GARCH models, to ARMA-JG-R-GARCH models, called BIP-ARMA-BIP-GJR models, to account for asymmetric effects.

3.3. GAS models

Harvey and Chakravarty (2008), Harvey (2013) and Creal et al. (2013) separately proposed similar modified GARCH models derived from the conditional score of the assumed distribution. The latter called this the Generalized Autoregressive Score (GAS) model and the two former the Dynamic Conditional Score (DCS) model, but it can also be viewed as an unrestricted version of the GAS model.

In the GAS model, the observed returns are described by Eqs. (1) and (2), where $z_t$ is assumed to follow a non-Normal distribution (Student-$t$ or Skewed Student-$t$ distribution), and the conditional variance is

$$
\tilde{\sigma}_t^2 = \omega + \alpha \tilde{r}_{t-1} + \beta \tilde{\sigma}_{t-1}^2 \tag{13}
$$

where $\tilde{\sigma}_t^2$ is a specified function of previous data. The critical part of the GAS and DCS models is the choice of $\tilde{\sigma}_t^2$. In the GAS models of Creal et al. (2013), $\tilde{\sigma}_t^2 = \tilde{S}_t \tilde{V}_t$, with $\tilde{S}_t$ being the score with respect to the parameter $\tilde{\sigma}_t^2$, i.e. $\tilde{S}_t = \partial \log f(\tilde{z}_t \tilde{\sigma}_t^2, Y_{t-1}; \theta)/\partial \tilde{\sigma}_t^2$, where $f(\tilde{z}_t \tilde{\sigma}_t^2, Y_{t-1}; \theta)$ represents the density of $\tilde{z}_t$ and $\theta$ is a vector of unknown parameters describing the joint-distribution function of the data, and $\tilde{S}_t$ is a time-dependent scaling matrix. Creal et al. (2013) recommend using $\tilde{S}_t = -E_{t-1}(\tilde{V}_t \tilde{V}_t')$.

In the DCS models of Harvey and Chakravarty (2008) and Harvey (2013), $\tilde{\sigma}_t^2 = u_t \tilde{\sigma}_t^2$, where $u_t$ is proportional to the (standardized) score of the conditional distribution at time $t$. This variable is a martingale difference by construction. $u_t$ is a function that can reduce the weight on and bound the effect of past shocks. When $z_t \sim i.i.d. \mathcal{N}(0,1)$, $u_t = z_t^2$, and the model reduces to the GARCH(1,1) model in which the impact of the shock is unbounded. However, for non-Normal distributions, large shocks are downweighted and have a smaller effect on future volatilities than in GARCH models. For example, when $z_t$ has a conditional Student-$t$ distribution with $\nu$ degrees of freedom $u_t$ is defined as

$$
u \geq (\nu + 2)^{1/2} - 1, \quad -1 \leq u_t \leq \nu, \quad \nu > 2 \tag{14}
$$

Harvey and Chakravarty (2008) and Harvey (2013) call this model a Beta-$t$-GARCH model because $u_t$ is a linear function of a variable with a Beta distribution ($u_t = 1/\nu + 1$). Harvey and Chakravarty (2008), Harvey (2013), Harvey and Sucarrat (2014) and Laurent et al. (2016) show that, for a DCS model with a Normal distribution (and also for a GARCH model), a jump has a very large effect on $u_t$ and therefore a large and slowly-decaying effect on future volatility predictions. However, with a Student-$t$ distribution the effect of a jump is bounded and therefore has only a small impact. Large shocks are downweighted, since $\tilde{\sigma}_t^2$ appears both in the numerator and the denominator of Eq. (14). A number of extensions have been suggested, including Skewed-Student errors. Here we follow the approach of Harvey and Chakravarty (2008) and Harvey (2013).

Harvey and Chakravarty (2008) and Harvey (2013) also consider an EGARCH-type version of the GAS model (Exponential GAS or EGAS, Beta-$t$-EGARCH). This model is extended to account for the leverage effect as follows,

$$
\log \tilde{\sigma}_t^2 = \omega + \alpha \tilde{r}_{t-1} \tilde{\sigma}_{t-1} + \gamma \tilde{r}_{t-1} + \beta \log \tilde{\sigma}_{t-1} \tag{15}
$$

where $\gamma$ is the asymmetric leverage coefficient describing volatility leverage, and $\tilde{I}_t = \text{sgn}(-z_t) (\alpha_0 / \tilde{\sigma}_{t-1}^2 + 1)$. In the remainder of the paper we call the DCS models of Harvey and Chakravarty (2008) and Harvey (2013) GAS models, and the Beta-$t$-EGARCH an EGAS model.

3.4. MSM models

The Markov-switching multifractal (MSM models) have been introduced by Calvet and Fisher (2001, 2004) and Lux (2008), and are characterized by a multiplicative rather than additive structure of the volatility process. In the MSM framework instantaneous volatility

---

8 The critical values are defined by $g_{l,x} = -\log(-\log(1 - \lambda))b_T + c_T$, with $b_T = 1/\sqrt{2\log T}$, and $c_T = (2\log T)^{1/2} - [\text{log}r + \text{log}(\log T)]/[2(2\log T)^{1/2}]$. Laurent et al. (2016) suggest setting $\lambda = 0.5$.  
9 See Gao and Zhou (2016) for the difference between GAS and DCS models.  
10 Note that in the GARCH(1,1) model, Eqs. (2) and (3), can be written as: $\tilde{\sigma}_t^2 = \omega + \alpha \tilde{r}_{t-1} + (\alpha + \beta) \tilde{\sigma}_{t-1}^2$, where $u_t = (\tilde{z}_t^2 / \tilde{\sigma}_t^2) - 1 = \tilde{z}_t^2 - 1$ is a martingale difference. Harvey and Chakravarty (2008) and Harvey (2013) therefore replace $u_t$ in the conditional variance by another martingale difference that is proportional to the score of the conditional distribution of $\tilde{z}_t$.  
11 The definition can be modified to deal with $\nu < 2$ by measuring volatility in terms of a scale parameter (Harvey and Chakravarty, 2008; Harvey, 2013).
\[ \alpha_i^2 = \sigma^2 \prod_{i=1}^{k} M_i^k \]  

where the multipliers are assumed to be persistent, independent of each other at any time, and satisfy \( M_i^k \geq 0 \) and \( \mathbb{E}[M_i^k] = 1 \). Following the basic hierarchical principle of the multifractal approach, each multiplier \( M_i^k \) is renewed at time \( t \) with probability \( \gamma_i \) depending on its rank within the hierarchy of multipliers, and remains unchanged with probability \( 1 - \gamma_i \). Convergence of the discrete-time MSM to a Poisson process in the continuous-time limit requires to formalize transition probabilities according to

\[ \gamma_i = 1 - (1 - \gamma_k)^{b - 1} \]

with \( \gamma_k \) and \( b \) parameters to be estimated (Calvet and Fisher, 2001). Here we are not interested in the continuous-time process, and therefore, we follow Lux (2008) and use pre-specified parameters \( \gamma_k = 0.5 \) and \( b = 2 \), defining the transition probabilities by \( \gamma_i = 2^{i-k} \). The choice of parameters for \( \gamma_k \) and \( b \) can be motivated by the fact that the in-sample fit and out-of-sample forecasting performance have been found to be almost invariant compared to other (estimated) values (Calvet and Fisher, 2004; Lux, 2008). The MSM model is a Markov-switching process with \( 2^k \) states.

The specification of the MSM model necessitates to specify the distribution of volatility components. As used in Wang et al. (2016) we assume that the random multipliers \( M_i^k \) follow a Binomial distribution taking the values \( m_0 \) and \( 2 - m_0 \) \( (1 \leq m_0 < 2) \), with a probability of 0.5, guaranteeing that \( \mathbb{E}[M_i^k] = 1 \) (Calvet and Fisher, 2004). This model is called the BMSM model. The maximum likelihood approach is employed for estimating only two parameters, the Binomial parameter \( m_0 \) and the scale factor \( \sigma \), although the number of states could be arbitrarily large (for large \( k \)). While, in principle, the ML approach for BMSM is the same as with any Markov-switching model, the high numbers of states \( (2^k \) distinct states) makes it computationally demanding for practical research, it is, therefore, applicable only to one-digit choices for \( k \). As in Lux et al. (2016) and Ben Nasr et al. (2016) we set \( k = 8 \) volatility components for the likelihood approach for estimation of the parameters \( (m_0, \sigma) \).

4. Data and summary statistics

The data we use here consists of the daily closing spot prices on two crude-oil markets: the US West Texas Intermediate (WTI) and UK Brent markets. The data comes from Thomson Financial Datastream and is given in US Dollars per barrel. The data covers January 6th 1992 to December 31st 2014, producing 5998 observations. The WTI and Brent crude-oil returns are calculated as the first differences in the logs of the spot prices \( \tau \). Following Kang et al. (2009), Wei et al. (2010) and Hou and Suardi (2012), daily actual volatility (variance) is assessed by daily squared returns \( (\tau^2) \). The graphical representations of prices, returns and volatility for the two series appear in Figs. 1–3, and reveal the presence of jumps.

Table 2 presents the summary statistics for WTI and Brent crude-oil returns. These have approximately equal mean returns of about 0.04% per day, with Brent returns being marginally smaller than WTI returns. The WTI returns are a little more volatile, as measured by their standard deviation (2.202 versus 2.059). All returns are highly non-Normal, and show evidence of significant positive skewness and excess kurtosis, as might be expected from daily returns. All series are leptokurtic (i.e., fat-tailed) and thus the variance in crude-oil prices principally reflects infrequent but extreme deviations. In addition, both returns have positive skewness, signifying that the series have a longer right tail (extreme gains) than left tail (extreme losses). The Lagrange Multiplier test for the presence of an ARCH effect clearly indicates that crude-oil prices exhibit strong conditional heteroscedasticity, which is common for financial data. As such, there

---

\[ \text{Another distribution assumption such as LogNormal, suggested by Lux (2008), can be used as well. However, to compare our results with those obtained by Wang et al. (2016) we focus only on the Binomial distribution. The comparison with another distribution will be examined in further research.} \]

\[ \text{See Charles and Darné (2014) for economic explanations of jumps in crude-oil markets.} \]
are quiet periods with small price changes and turbulent periods with large oscillations.

The jump-filtered returns \( \tilde{r}_t \) also exhibit excess skewness (except for the Brent returns), excess kurtosis and conditional heteroscedasticity, although the excess kurtosis decreases dramatically. This is consistent with Carnero et al. (2001) and Charles and Darné (2005), who show that jumps may cause significant skewness.

5. Estimation results

The descriptive statistics suggest that an appropriate model of crude-oil return volatility should account for its time-varying nature and the non-Normality of oil returns. We hence estimate the various volatility models that were sketched out in Section 3 using the Normal, Student-\( t \) and Skewed Student-\( t \) distributions (Bollerslev, 1987; Lambert and Laurent, 2001) for GARCH- and GAS-type models\(^{14}\) as well as the BMSM models on the period from January 6th 1992 to December 31st 2007. The different GARCH-type models (GARCH, GJR and EGARCH) are applied to both raw returns and filtered returns, and the GAS-type and BMSM models are applied to raw returns. Note that the filtered returns from the BIP-ARMA-BIP-GARCH models are used to estimate the GARCH models, whereas

\(^{14}\) Note that estimating GARCH models with heavy-tailed distributions may lead to inconsistent (non-Gaussian) maximum-likelihood (ML) estimate of model parameters, while the Gaussian ML estimator may be consistent even if the true distribution is far from being Gaussian, although this gain in robustness comes with a loss of efficiency (Fan et al., 2014). Some GARCH-type models with Student-\( t \) distribution for crude oil volatility appear in Cheong (2009), Marzo and Zagaglia (2010), and Hou and Suardi (2012), among others, whereas Salisu and Fasanya (2013) employ a Skewed Student-\( t \) distribution.
the filtered returns from the BIP-ARMA-BIP-GJR models are used to estimate the asymmetric GARCH models (GJR and EGARCH).\textsuperscript{15} The parameters of the volatility models are estimated by the quasi-maximum likelihood (QML) method of Bollerslev and Wooldridge (1992) – producing robust standard errors – and the quasi-likelihood function is maximized using the quasi-Newton method of Broyden, Fletcher, Goldfarb and Shanno (BFGS).\textsuperscript{16} The estimation results for the Brent and WTI crude-oil price returns appear in Tables 3 and 4, respectively, for GARCH and GAS models. The likelihood estimation of BMSM models is conducted for $k = 8$, fixed from the outset. The estimates of the Binomial parameter, $n_0$, are 1.254 and 1.241, and the scale factor parameter, $\sigma$, are 2.099 and 2.105, for the Brent and WTI returns, respectively.

The comparison between the volatility models is effected via various in-sample criteria: Log-Likelihood (LL), Akaike (AIC) and Hannan-Quinn (HQ). For each table, the best model appears in bold, showing the highest value of the LL and the lowest values of the AIC and HQ. The residual tests are also shown to see whether the chosen volatility model is the most appropriate.

Even though the in-sample criteria for the Brent and WTI markets are very similar to each other under the different GARCH-type models estimated on the raw returns, their volatility is best modeled by a GJR-GARCH process with Student-$t$ and Skewed Student-$t$ distributions, respectively (Tables 3 and 4), suggesting asymmetry in these crude oil markets. Modeling crude-oil returns via (E)GARCH models does not improve the in-sample criteria, except very slightly for the WTI return series with an EGARCH model with a Skewed Student-$t$ distribution. Last, the Brent and WTI data are better fitted when using filtered returns. Tables 3 and 4 show that the best specification for the Brent and WTI data is the GJR-GARCH model with a Student-$t$ and Skewed Student-$t$ distribution, respectively. Note that Brent returns seem to exhibit more asymmetry than WTI returns ($\gamma = 0.024$ and 0.014, respectively) and a similar high degree of persistence (0.997).

### 6. Out-of-sample

As indicated in the estimation and diagnostic evaluations, the GJR-GARCH model appears to fit both the Brent and WTI returns well. However, this does not guarantee that it will perform better in actual forecasting tests. We carry out a rolling out-of-sample forecast for a one-day horizon and assess model forecast performance by out-of-sample one-step ahead prediction errors. The models are estimated using the first 16-year period, from January 6th 1992 to December 31st 2007 ($T = 4171$ observations), and the out-of-sample forecasts of the conditional variance are calculated for January 2nd 2008 to December 31st 2014 ($H = 1826$ observations). This out-of-sample period includes a period of jumps at the beginning of 2009 in the WTI due to US announcements on crude inventories and storage capacity (Charles and Darné, 2014). The models are re-estimated every 50 days, from a rolling window of 4171 observations. As such, the estimation sample size remains fixed and the forecasts do not overlap. One-day out-of-sample volatility forecasts $\hat{\sigma}_T^2$, with $k = 1, \ldots, H$, are obtained for the forecast horizon, and compared to the raw returns $r_{t+1}^2$. The use of squared returns $r_{t+1}^2$ as a proxy for actual volatility is an established practice (see, e.g., Kang et al., 2009; Wei et al., 2010; Hou and Suardi, 2012). This is an unbiased proxy, although noisy (Lopez, 2001).

Patton (2011) shows that there are only two loss functions that yield optimal forecasts when using squared returns as a volatility proxy: the mean squared error (MSE) and the loss implied by a Gaussian likelihood (QLIKE). We thus assess volatility forecast performance by the MSE\textsuperscript{17}.

$$\text{MSE} = \frac{1}{H} \sum_{t=1}^{H} (\hat{\sigma}_t^2 - \sigma_t^2)^2.$$  

The presence of jumps, Hansen and Lunde (2005) and Preminger and Franck (2007) recommend the use of forecast-performance evaluation criteria that are less sensitive to jumps. We hence consider the Mean Absolute Deviation (MAD), defined as

$$\text{MAD} = \frac{1}{H} \sum_{t=1}^{H} |\hat{\sigma}_t^2 - \sigma_t^2|.$$  

Obviously, the simple comparison of MSE- and MAD-values does not take into account the sample uncertainty underlying the observed forecast differences. We therefore also apply the model confidence set (MCS) procedure proposed by Hansen et al. (2011) to determine the set, $\hat{M}_T$, that consists of a subset of equivalent models in terms of superior predictive ability (SPA: Hansen and Lunde, 2005) over the other competing collections of models, $M_0$. The MCS procedure yields a model confidence set, $\hat{M}_T$, which is a set of models that is

\textsuperscript{15} We found the same number of jumps in the jump-detection procedure based on the BIP-ARMA-BIP-GARCH and BIP-ARMA-BIP-GJR models, except for one jump.

\textsuperscript{16} To estimate and forecast these returns series, we use G@RCH 7.0 for Ox, a package dedicated to the estimation and forecasting of GARCH- and GAS-type models. Note that we changed the initial values of the parameters due to convergence problems in the EGARCH model with the Normal distribution for Brent returns. The Berndt-Hall-Hall-Hausman (BHHH) algorithm is another way of optimizing the likelihood function. The BFGS and BHHH algorithms are similar regarding the first derivatives of the likelihood function with respect to the numerically-calculated parameters, but differ in their construction of the Hessian matrix of second derivatives. Brooks et al. (2001, 2003), review some software packages that estimate GARCH-type models and underline that they produce different results (see also Zivot, 2009). We applied the BHHH algorithm with the Marquardt correction (which modifies the BHHH algorithm by adding a correction matrix to the Hessian approximation) available in EViews 9 to estimate the asymmetric GARCH models in order to obtain robust estimated parameters, and found similar results with the BFGS algorithm. The results are available in the Online Appendix. We thank an anonymous referee for bringing this important point to our attention. The ML estimation and forecasting of BMSM models is done on Gauss 17. We thank Thomas Lux and Mawuli Segnon for sharing their Gauss codes.

\textsuperscript{17} We have also used the QLIKE and obtain similar results.
models constructed to contain the best models with a given level of
confidence. This MCS produces a number of models with the same
forecasting performance, and therefore more robust forecasting than
that using only one model. The \( t \)-statistic is defined as

\[
T_{\text{max } \mathcal{M}} = \max_{i \in \mathcal{M}} t_i = \frac{\tilde{d}_i}{\sqrt{V\text{ar}(d_i)}}
\]

(16)

where \( V\text{ar}(d_i) \) denotes the estimate of \( V\text{ar}(\tilde{d}_i) \), \( \tilde{d}_i = m^{-1} \sum_{j=1}^{m} \tilde{d}_{ij} \) and \( \tilde{d}_{ij} = n^{-1} \sum_{i=1}^{n} d_{ij} \), with \( d_{ij} = y_{ij} - \hat{y}_{ij} \) for all \( i, j \in \mathcal{M}_0 \) and \( \hat{y}_{ij} \) is a loss function (here MSE or MAD). \( t_i \) is associated with the null hypothesis of equal predictive ability (EPA: Hansen, 2005) \( H_{0i,\mathcal{M}} : E(d_{ij}) = 0 \) for all \( i \in \mathcal{M}_0 \), \( \mathcal{M}_0 \subseteq \mathcal{M} \). The MCS is a sequential
testing procedure, eliminating at each step the worst model from
\( \mathcal{M} \), until the null hypothesis of EPA is accepted for all the models. If
the null of EPA is rejected for \( \mathcal{M} = \mathcal{M}_0 \), the worst-performing model
is excluded from the set \( \mathcal{M} \).\(^{19}\) The iterative procedure stops when
the null hypothesis of EPA of the models still included in the set
cannot be rejected. If \( H_{0,\mathcal{M}} \) is accepted at level \( \alpha \) then the MCS is the set
\( \mathcal{M}_{\mathcal{M}_0} \).\(^{20}\)

The RMSE and MAD values, as well as results of the MCS test
(\( p \)-values), appear in Tables 5 and 6 for the Brent and WTI returns
series, respectively. The initial model space \( \mathcal{M}_0 \) consists of 19 and
20 models for the Brent and WTI returns series, respectively: these
are the volatility models that satisfy the regularity, significance and
non-negativity conditions. We set the confidence level for the MCS
at \( \alpha = 0.10, 0.50 \) and 0.90.

For the Brent returns, the RMSE and MAD values are close for
each model, but the GJR-GARCH models estimated from the filtered
series display the lowest values. All the volatility models appear in

\(^{19}\) The choice of the worst model to be eliminated uses the following elimination
rule: \( \text{E}_{\text{max }, \mathcal{M}} = \arg \max_{i \in \mathcal{M}} t_i \).

\(^{20}\) The MCS \( p \)-values are calculated using bootstrap implementation with 10,000
resamples (Hansen et al., 2011). The MCS test is carried out using the Ox software
tered return series, is

\[ \text{MSE} \]

is the lag length.

**Notes:** a The parameter of asymmetry in the Skewed-Student distribution is not significant at the 10% level. b The condition for the existence of the fourth moment of the GARCH model is not respected (Ling and McAleer, 2001). The log-likelihood value, AIC and HQ correspond to the Akaike and Hannan-Quinn criteria, respectively. Q(10) and Q(2)** are respectively the Box-Pierce statistics with lag 10 of the standardized and squared standardized residuals. These are asymptotically distributed as \( \chi^2(k) \), where \( k \) is the lag length. LM(10) is the ARCH LM test with lag 10. This is distributed as \( \chi^2(q) \), where \( q \) is the lag length. The robust t-ratios appear in parentheses. The best in-sample criteria appear in bold. 

- **Indicates that the null hypothesis is rejected at the 5% level.**
- **Indicates that the null hypothesis is rejected at the 10% level.**

\[ \hat{\lambda}^{(10)} \]

with the MSE loss function, and only one volatility model, the GARCH model with a Student-t distribution estimated on the filtered return series, is in \( \hat{\lambda}^{(50)} \). Using the MAD loss function, only the GARCH-type models estimated on the filtered returns appear in the \( \hat{\lambda}^{(10)} \), and especially the EGARCH model with a Normal distribution is in the \( \hat{\lambda}^{(50)} \). These results suggest that estimating GARCH-type

\[ \text{Table 5} \]


<table>
<thead>
<tr>
<th>Model</th>
<th>Distrib.</th>
<th>GARCH</th>
<th>GARCH</th>
<th>GJR</th>
<th>GJR</th>
<th>EGARCH</th>
<th>GJR</th>
<th>EGARCH</th>
<th>GAS</th>
<th>GAS</th>
<th>GAS</th>
<th>GJR</th>
<th>GJR</th>
<th>EGARCH</th>
<th>EGARCH</th>
<th>EGAS</th>
<th>EGAS</th>
<th>EGAS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RMSE</td>
<td>0.0980</td>
<td>0.0980</td>
<td>0.0976</td>
<td>0.0975</td>
<td>0.0975</td>
<td>0.0980</td>
<td>0.0981</td>
<td>0.0981</td>
<td>0.0983</td>
<td>0.0980</td>
<td>0.0980</td>
<td>0.0980</td>
<td>0.0980</td>
<td>0.0980</td>
<td>0.0980</td>
<td>0.0980</td>
<td>0.0980</td>
<td>0.0980</td>
</tr>
<tr>
<td>MCS-ME</td>
<td>0.2459</td>
<td>0.2459</td>
<td>0.3773</td>
<td>0.3773</td>
<td>0.3773</td>
<td>0.2459</td>
<td>0.2459</td>
<td>0.2459</td>
<td>0.2439</td>
<td>0.2439</td>
<td>0.2439</td>
<td>0.2439</td>
<td>0.2439</td>
<td>0.2439</td>
<td>0.2439</td>
<td>0.2439</td>
<td>0.2439</td>
<td>0.2439</td>
</tr>
<tr>
<td>MAD</td>
<td>0.0414</td>
<td>0.0414</td>
<td>0.0412</td>
<td>0.0412</td>
<td>0.0411</td>
<td>0.0414</td>
<td>0.0415</td>
<td>0.0415</td>
<td>0.0409</td>
<td>0.0409</td>
<td>0.0409</td>
<td>0.0409</td>
<td>0.0409</td>
<td>0.0409</td>
<td>0.0409</td>
<td>0.0409</td>
<td>0.0409</td>
<td>0.0409</td>
</tr>
<tr>
<td>MCS-MAD</td>
<td>0.0008</td>
<td>0.00080</td>
<td>0.0008</td>
<td>0.0006</td>
<td>0.0000</td>
<td>0.0016</td>
<td>0.0000</td>
<td>0.0006</td>
<td>0.0244</td>
<td>0.0244</td>
<td>0.0244</td>
<td>0.0244</td>
<td>0.0244</td>
<td>0.0244</td>
<td>0.0244</td>
<td>0.0244</td>
<td>0.0244</td>
<td>0.0244</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.0976</td>
<td>0.0976</td>
<td>0.0976</td>
<td>0.0976</td>
<td>0.0976</td>
<td>0.0975</td>
<td>0.0975</td>
<td>0.0975</td>
<td>0.0973</td>
<td>0.0973</td>
<td>0.1003</td>
<td>0.1003</td>
<td>0.1003</td>
<td>0.1003</td>
<td>0.1003</td>
<td>0.1003</td>
<td>0.1003</td>
<td>0.1003</td>
</tr>
<tr>
<td>MCS-ME</td>
<td>0.2459</td>
<td>0.2459</td>
<td>0.3773</td>
<td>0.3773</td>
<td>0.2459</td>
<td>0.2459</td>
<td>0.2459</td>
<td>0.2459</td>
<td>0.2459</td>
<td>0.2459</td>
<td>0.2459</td>
<td>0.2459</td>
<td>0.2459</td>
<td>0.2459</td>
<td>0.2459</td>
<td>0.2459</td>
<td>0.2459</td>
<td>0.2459</td>
</tr>
<tr>
<td>MAD</td>
<td>0.0402</td>
<td>0.0402</td>
<td>0.0403</td>
<td>0.0402</td>
<td>0.0402</td>
<td>0.0402</td>
<td>0.0402</td>
<td>0.0402</td>
<td>0.0402</td>
<td>0.0402</td>
<td>0.0402</td>
<td>0.0402</td>
<td>0.0402</td>
<td>0.0402</td>
<td>0.0402</td>
<td>0.0402</td>
<td>0.0402</td>
<td>0.0402</td>
</tr>
<tr>
<td>MCS-MAD</td>
<td>0.8045**</td>
<td>0.1589**</td>
<td>0.6213**</td>
<td>0.8045**</td>
<td>0.3025**</td>
<td>1.000***</td>
<td>0.1589*</td>
<td>0.3474*</td>
<td>0.0244</td>
<td>0.0244</td>
<td>0.0244</td>
<td>0.0244</td>
<td>0.0244</td>
<td>0.0244</td>
<td>0.0244</td>
<td>0.0244</td>
<td>0.0244</td>
<td>0.0244</td>
</tr>
</tbody>
</table>

Notes: XXXm denotes the GARCH-type model estimated on filtered data. MCS-MSE and MCS-MAD denote the \( p \)-value of the statistic \( T_{\text{max,AH}} \) of Hansen et al. (2011) based on the MASE and MAD loss functions respectively. The lowest RMSE and MAD values appear in bold.

- **Means that the forecasts are in \( \hat{\lambda}^{(10)} \).
- **Means that the forecasts are in \( \hat{\lambda}^{(50)} \).
- **Means that the forecasts are in \( \hat{\lambda}^{(90)} \).
models on filtered data seems to be appropriate for forecasting Brent returns.

For the WTI returns, the EGARCH models with Student-t and Skewed-Student-t distributions estimated on raw data have the best RMSE values. This result is confirmed by the MCS test, where these two EGARCH models and the GJR-GARCH model are in the $\mathcal{X}_{10%}^2$, whereas most of the other models are in the $\mathcal{X}_{50%}^2$, suggesting that these models are equivalent in terms of predictive ability. However, when we use the MAD as the loss function, the models estimated on filtered data are in the $\mathcal{X}_{10%}^2$, and only five models are in the $\mathcal{X}_{50%}^2$, namely a GARCH model with a Normal distribution and the GJR-GARCH and EGARCH models with Student-t and Skewed-Student-t distributions. Among these three, the GJR-GARCH model with a Student-t seems to have the best forecast accuracy for the WTI returns. All the other volatility models, i.e., GARCH-type, GAS-type and BMSM models estimated on raw return data, are excluded from the MCSs, suggesting that they forecast significantly less accurately than GARCH-type models estimated on filtered data. This can be explained by the presence of a number of jumps at the beginning of 2009 due to US announcements regarding crude inventories and storage capacity, which affected the MSE measure as a criterion to evaluate forecast performance (Hansen and Lunde, 2005; Preminger and Franck, 2007). Note that the BMSM model gives the lowest MAD values than the other volatility models estimated on raw return data. This result is in line with those obtained by, for example, Carnero et al. (2007), Charles (2008) and Laurent et al. (2016) for stock and exchange-rate returns. Taking jumps in volatility forecasting into account may therefore improve the forecasting performance of volatility in crude-oil markets.

Our study focuses on GARCH- and GAS-type models, which capture short memory as well as asymmetry. It would be interesting to extend this work to long-memory models. For example, Janus et al. (2014) propose volatility models which introduce long memory into GAS models. Laurent et al. (2016) suggest that their jump-detection procedure can be extended to account for long-memory, as in FIGARCH and FIAPARCH models. Another extension would be to consider the MSM volatility model with a LogNormal distribution as in FIGARCH and FIAPARCH models. Our study has focused on the detection of jumps in volatility models estimated on portfolios of Brent and WTI returns, and also Janus et al. (2014) that the standard-GARCH models estimated on filtered returns outperform GARCH-type models estimated on raw data, and also Laurent et al. (2016) that standard-GARCH models estimated on filtered returns outperform other models, including GAS-type models.

7. Conclusion

This paper has analyzed volatility models and their forecasting performance under the presence of jumps in two crude-oil markets - Brent and West Texas Intermediate (WTI) - between January 6th 1992 and December 31st 2014. We first compare a number of GARCH-type models that capture short memory as well as asymmetry (GARCH, GJR-GARCH and EGARCH), estimated on raw returns, to three competing approaches that deal with the presence of jumps: GARCH-type models estimated from jump-filtered returns, and two new classes of volatility models, called Generalized Autoregressive Score (GAS) and Markov-switching multifractal (MSM) models, estimated from raw returns. The forecasting performance of the volatility models is evaluated via the RMSE and MAD criteria, as well as the model confidence set approach proposed by Hansen et al. (2011), allowing us to identify a subset of models that outperform all the others.

We found that asymmetric models estimated on filtered returns provide better out-of-sample forecasts than GARCH-, GAS-type and BMSM models estimated on raw Brent and WTI returns. This finding confirms those obtained by, for example, Carnero et al. (2007), Charles (2008) and Laurent et al. (2016) for stock and exchange-rate returns. Taking jumps in volatility forecasting into account may therefore improve the forecasting performance of volatility in crude-oil markets.

Acknowledgments

We are grateful for useful comments on a previous version of this paper by anonymous referees. The authors gratefully acknowledge financial support from the Région des Pays de la Loire (France) through the grant PANORisk. Olivier Darné gratefully acknowledges financial support from the Chaire Finance of the University of Nantes Research Foundation.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.eneco.2017.09.002.

21 We have also compared the volatility models estimated only on the raw returns and found that GAS-type and BMSM models are interesting alternative volatility models to GARCH-type models for forecasting Brent and WTI returns based. The results are available upon request.