Modeling Energy Price Dynamics: GARCH versus Stochastic Volatility

Joshua C.C. Chan*
Research School of Economics,
Australian National University

Angelia L. Grant
Research School of Economics,
Australian National University

November 2015

Abstract

We compare a number of GARCH and stochastic volatility (SV) models using nine series of oil, petroleum product and natural gas prices in a formal Bayesian model comparison exercise. The competing models include the standard models of GARCH(1,1) and SV with an AR(1) log-volatility process, as well as more flexible models with jumps, volatility in mean, leverage effects, and t distributed and moving average innovations. We find that: (1) SV models generally compare favorably to their GARCH counterparts; (2) the jump component and t distributed innovations substantially improve the performance of the standard GARCH, but are unimportant for the SV model; (3) the volatility feedback channel seems to be superfluous; (4) the moving average component markedly improves the fit of both GARCH and SV models; and (5) the leverage effect is important for modeling crude oil prices—West Texas Intermediate and Brent—but not for other energy prices. Overall, the SV model with moving average innovations is the best model for all nine series.

Keywords: Bayesian model comparison, crude oil, natural gas, moving average, jumps, leverage, t distribution.

JEL classification: C11, C52, Q41

*Financial support from the Australian Research Council via a Discovery Early Career Researcher Award (DE150100795) is gratefully acknowledged.

1 Introduction

The volatility of oil, petroleum product and natural gas prices has broad economic and financial implications, and this has motivated a vast literature on modeling such volatility. Conventionally, this time-varying volatility is modeled—and the fit assessed—using various generalized autoregressive conditional heteroscedastic (GARCH) models, under which the conditional variance is a deterministic function of model parameters and past data.[1]

Alternatively, some recent papers have considered stochastic volatility models, where the volatility is a latent variable that follows a stochastic process (see, e.g., Sadorsky, 2005; Vo, 2009; Trolle and Schwartz, 2009; Larsson and Nossman, 2011; Brooks and Prokopczuk, 2013). These two classes of models are nonnested and the implied time-varying volatilities have very different properties. To the extent that they are compared at all, the literature has mainly focused on their forecasting performance. While volatility forecasting is an important problem, energy prices are widely used in macroeconomic models to analyze the interplay between these prices and the macroeconomy (see, e.g., Kilian, 2009; Peersman and Van Robays, 2012; Blanchard and Riggi, 2013). Consequently, it is of interest to directly compare the model fit of these two classes of time-varying volatilty models in a formal model comparison exercise, but this is rarely done in practice.

We fill this gap by assessing the model fit—while penalizing model complexity—of a number of GARCH and stochastic volatility models for modeling the dynamics of oil, petroleum product and natural gas prices. To that end, we perform a formal Bayesian model comparison exercise to assess the evidence in favor of the GARCH and stochastic volatility models given the data. Specifically, for each model we compute its marginal data density, which evaluates how likely it is for the observed data to have occurred given the model. Using this measure we can further obtain the posterior probabilities of the models (see, e.g., Koop, 2003, for a detailed discussion on Bayesian model comparison).

For the model comparison exercise, we consider seven commonly-used GARCH models in the literature: the standard GARCH(1,1) model, and the more flexible models of GARCH(2,1), GARCH with jumps, GARCH in mean, GARCH with moving average innovations, GARCH with t distributed innovations, and GARCH with an asymmetric leverage effect. We then choose seven stochastic volatility models that are close counter-

[1] See, e.g., Fong and See (2001), Sadorsky (2006), Kang, Kang, and Yoon (2009), Agnolucci (2009), Mohammadi and Su (2010), Nomikos and Andriosopoulos (2012), Mason and Wilmot (2014) and Manera, Nicolini, and Vignati (2014).

parts of these GARCH models. By directly comparing the GARCH and stochastic volatility models pairwise, we can assess whether the deterministic conditional heteroscedasticity under GARCH or the stochastic variance under SV is more favored by the data. In addition, we can compare the more flexible GARCH variants against the standard GARCH—and the flexible stochastic volatility models against the standard SV—to examine what features are most empirically relevant for energy prices, thus providing useful and practical guidelines for practitioners.

The main results can be summarized as follows. First, the stochastic volatility models generally compare favorably to their GARCH counterparts, indicating that the time-varying volatility is better modeled as a latent stochastic process. This result is in line with the finding in the finance literature that stochastic volatility models often fit financial returns better (see, e.g., Kim, Shepherd, and Chib, 1998; Yu, 2002).

Second, both the jump component and t distributed innovations substantially improve the performance of the standard GARCH, but are unimportant for the stochastic volatility model. Since the standard GARCH specifies a deterministic conditional variance process, adding a random jump component or allowing for innovations with heavy-tailed distributions appear to give the model additional flexibility against misspecification. These are apparently unnecessary for the stochastic volatility model. Third, the volatility feedback channel in both the GARCH in mean and stochastic volatility in mean models seems to be superfluous. Fourth, the moving average component substantially improves the fit of both types of models. Fifth, the leverage effect is important for modeling crude oil prices—West Texas Intermediate (WTI) and Brent—but not for other energy prices. Overall, the SV model with moving average innovations is the best model for all nine series. In a recursive out-of-sample forecasting exercise, we confirm these conclusions by comparing the GARCH and stochastic volatility models using their density forecasts.

The rest of this article is organized as follows. Section 2 introduces the two classes of time-varying volatility models—GARCH and stochastic volatility models. In Section 3 we give an overview of Bayesian model comparison and outline an adaptive importance sampling approach to compute the marginal likelihood for comparing models. Section 4 compares the performance of the GARCH models with their stochastic volatility counterparts. Estimation and forecasting results are also reported. Lastly, Section 5 concludes and briefly discusses some future research directions.

2 Time-Varying Volatility Models

In this section we discuss the two classes of time-varying volatility models used in the model comparison exercise. The first class of models is the generalized autoregressive conditional heteroscedastic (GARCH) models, which are developed by Bollerslev (1986) to extend the earlier work on ARCH models by Engle (1982). The second set of models is the stochastic volatility models, which are considered by Taylor (1994). In these models the volatility is specified as a latent stochastic process.

2.1 GARCH Models

In this section we describe various GARCH models that are widely used to model energy prices. The first one is the standard GARCH(1,1) model, which we simply refer to as GARCH:

$$y_t = \mu + \varepsilon_t, \qquad \varepsilon_t \sim \mathcal{N}(0, \sigma_t^2), \tag{1}$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \tag{2}$$

where $\varepsilon_0 = 0$ and σ_0^2 is a constant. To ensure the variance process σ_t^2 is always straightly positive and stationary, we assume that $\alpha_0 > 0$, $\alpha_1 \geqslant 0$, $\beta_1 \geqslant 0$ and $\alpha_1 + \beta_1 < 1$. Note that the conditional variance σ_t^2 is a deterministic function of the model parameters and past data.

The conditional variance σ_t^2 in (2) follows an AR(1) process. Next, we consider the GARCH(2,1) model—in which σ_t^2 follows an AR(2) process—that allows for richer variance dynamics:

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 + \beta_2 \sigma_{t-2}^2,$$

where $\sigma_{-1}^2 = \varepsilon_0 = 0$ and σ_0^2 is a constant. Again, to ensure the variance process σ_t^2 is always straightly positive and stationary, we assume that the parameters $\alpha_0, \alpha_1, \beta_1$, and β_2 are all positive and $\alpha_1 + \beta_1 + \beta_2 < 1$. We refer to this model as GARCH-2.

The third GARCH model allows for the possibility of infrequent "jumps" in the data series, which can accommodate drastic changes in energy prices. More specifically, consider

the following GARCH with jumps (GARCH-J) model:

$$\begin{aligned} y_t &= \mu + k_t q_t + \varepsilon_t, && \varepsilon_t \sim \mathcal{N}(0, \sigma_t^2), \\ \sigma_t^2 &= \alpha_0 + \alpha_1 (y_{t-1} - \mu)^2 + \beta_1 \sigma_{t-1}^2, \end{aligned}$$

where $q_t \in \{0, 1\}$ is a jump variable with success probability $\mathbb{P}(q_t = 1) = \kappa$. Hence, if $q_t = 1$, a jump occurs at time t and its size is determined by k_t, which is modeled as $k_t \sim \mathcal{N}(\mu_k, \sigma_k^2)$.

Next, consider the GARCH in mean (GARCH-M) model, under which the conditional variance enters the conditional mean as a covariate:

$$\begin{aligned} y_t &= \mu + \lambda \sigma_t^2 + \varepsilon_t, && \varepsilon_t \sim \mathcal{N}(0, \sigma_t^2), \\ \sigma_t^2 &= \alpha_0 + \alpha_1 (y_{t-1} - \mu - \lambda \sigma_{t-1}^2)^2 + \beta_1 \sigma_{t-1}^2. \end{aligned}$$

This variant allows for the possibility that the data series depends on its volatility (risk). It is obvious that when $\lambda = 0$, the GARCH-M model reduces to the GARCH model.

The fifth GARCH model combines a first-order moving average model with GARCH innovations:

$$\begin{aligned} y_t &= \mu + \varepsilon_t, \\ \varepsilon_t &= u_t + \psi u_{t-1}, && u_t \sim \mathcal{N}(0, \sigma_t^2), \end{aligned}$$

where the invertibility condition is imposed, i.e., $|\psi| < 1$ and the variance σ_t^2 follows the same GARCH process as in (2). This GARCH model is referred to as GARCH-MA. In contrast to the other GARCH models considered above, this model allows the data series to be correlated over time and might better model the short-run dynamics of the series.

Next, consider the GARCH model with t innovations (GARCH-t):

$$y_t = \mu + \varepsilon_t, \qquad \varepsilon_t \sim t_\nu(0, \sigma_t^2),$$

where σ_t^2 follows the GARCH process as in (2). Since the t distribution has heavier tails than the Gaussian, the GARCH-t model allows for more extreme observations compared to the standard GARCH.

The last model is the GARCH-GJR model of Glosten, Jagannathan, and Runkle (1993) that allows for potentially larger impact of negative excess returns on the conditional

variance. More specifically, the variance process becomes

$$\sigma_t^2 = \alpha_0 + (\alpha_1 + \delta_1 \mathbb{1}(\varepsilon_{t-1} < 0))\varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2,$$

where $\mathbb{1}(\cdot)$ is the indicator function. The asymmetric leverage effect is controlled by the parameter δ_1. It is clear that when $\delta_1 = 0$, this variant becomes the standard GARCH. Various studies, such as Wei, Wang, and Huang (2010), have found that allowing for an asymmetric leverage effect generally improves the forecast performance of the standard GARCH.

2.2 Stochastic Volatility Models

Next we introduce the seven stochastic volatility models that are close counterparts of the GARCH models described in the previous section. The volatility under a stochastic volatility model is a random variable, in stark contrast to GARCH models in which the conditional variance is a deterministic function of the model parameters and past data.

The first model is the standard stochastic volatility (SV) model:

$$y_t = \mu + \varepsilon_t^y, \qquad \varepsilon_t^y \sim \mathcal{N}(0, \mathrm{e}^{h_t}), \tag{3}$$

$$h_t = \mu_h + \phi_h(h_{t-1} - \mu_h) + \varepsilon_t^h, \qquad \varepsilon_t^h \sim \mathcal{N}(0, \omega_h^2). \tag{4}$$

The log-volatility h_t follows a stationary AR(1) process with $|\phi_h| < 1$ and unconditional mean μ_h. The process is initialized with $h_1 \sim \mathcal{N}(\mu_h, \omega_h^2/(1-\phi_h^2))$.

In the second stochastic volatility model, the observation equation is the same as in (3), but the log-volatility h_t now follows a stationary AR(2) process:

$$h_t = \mu_h + \phi_h(h_{t-1} - \mu_h) + \rho_h(h_{t-2} - \mu_h) + \varepsilon_t^h, \qquad \varepsilon_t^h \sim \mathcal{N}(0, \omega_h^2),$$

where we assume the roots of the characteristic polynomial associated with (ϕ_h, ρ_h) lie outside the unit circle. Further, h_1 and h_2 are assumed to follow the unconditional distribution:

$$h_1, h_2 \sim \mathcal{N}\left(\mu_h, \frac{(1-\rho_h)\omega_h^2}{(1+\rho_h)((1-\rho_h)^2 - \phi_h^2)}\right).$$

This stochastic volatility model is referred to as SV-2, which reduces to the standard SV model when $\rho_h = 0$.

Similar to the GARCH-J model, the third stochastic volatility model accommodates the possibility of infrequent jumps. Specifically, under the stochastic volatility model with jumps (SV-J), the observation equation becomes:

$$y_t = \mu + k_t q_t + \varepsilon_t^y, \qquad \varepsilon_t^y \sim \mathcal{N}(0, \mathrm{e}^{h_t}),$$

where the log-volatility h_t follows the same AR(1) process as in (4). The jump indicator q_t and jump size k_t are modeled exactly the same as in the GARCH-J model.

Next we consider the stochastic volatility in mean (SV-M) model of Koopman and Hol Uspensky (2002), under which the stochastic volatility enters the observation equation as a covariate:

$$y_t = \mu + \lambda \mathrm{e}^{h_t} + \varepsilon_t^y, \qquad \varepsilon_t^y \sim \mathcal{N}(0, \mathrm{e}^{h_t}).$$

As before, the log-volatility follows the same AR(1) process as in (4). The parameter λ captures the extent of volatility feedback; when $\lambda = 0$, the SV-M reduces to the standard SV model.

The fifth model is a version of the stochastic volatility models with moving average innovations in Chan (2013). In particular, consider the following first-order moving average model with stochastic volatility:

$$\begin{aligned} y_t &= \mu + \varepsilon_t^y, \\ \varepsilon_t^y &= u_t + \psi u_{t-1}, \qquad u_t \sim \mathcal{N}(0, \mathrm{e}^{h_t}), \end{aligned}$$

where $u_0 = 0$ and $|\psi| < 1$. Again the log-volatility h_t is assumed to follow the AR(1) process as in (4). This stochastic volatility model is referred to as SV-MA.

Next, the counterpart of GARCH-t is the stochastic volatility model with t innovations (SV-t):

$$y_t = \mu + \varepsilon_t^y, \qquad \varepsilon_t^y \sim t_\nu(0, \mathrm{e}^{h_t}),$$

where the log-volatility h_t is again assumed to follow the AR(1) process as in (4).

Similar to GARCH-GJR, the stochastic volatility model with leverage (SV-L) allows a leverage effect. Specifically, the innovations in the observation and state equations can

potentially be correlated:

$$y_t = \mu + \varepsilon_t^y,$$
$$h_{h+1} = \mu_h + \phi_h(h_t - \mu_h) + \varepsilon_t^h,$$

where the innovations ε_t^y and ε_t^h jointly follow a bivariate normal distribution:

$$\begin{pmatrix} \varepsilon_t^y \\ \varepsilon_t^h \end{pmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{pmatrix} \mathrm{e}^{h_t} & \rho \mathrm{e}^{\frac{1}{2}h_t}\omega_h \\ \rho \mathrm{e}^{\frac{1}{2}h_t}\omega_h & \omega_h^2 \end{pmatrix}\right).$$

If $\rho < 0$, given a negative shock to y_t at time t, the volatility at time $t+1$ tends to be larger. It is also clear that when $\rho = 0$, this model reduces to the standard SV.

We summarize the GARCH and stochastic volatility models in Table 1. Both the GARCH and stochastic volatility models are estimated using Bayesian techniques. The estimation is outlined in Appendix A.

Table 1: List of GARCH and stochastic volatility models.

GARCH models	
GARCH	GARCH(1,1) model where σ_t^2 follows a stationary AR(1)
GARCH-2	same as GARCH but σ_t^2 follows a stationary AR(2)
GARCH-J	same as GARCH but the prices equation has a "jump" component
GARCH-M	same as GARCH but σ_t^2 enters the prices equation as a covariate
GARCH-MA	same as GARCH but the observation error follows an MA(1)
GARCH-t	same as GARCH but the observation error follows a t distribution
GARCH-GJR	GARCH with a leverage effect
Stochastic volatility models	
SV	stochastic volatility model where h_t follows a stationary AR(1)
SV-2	same as SV but h_t follows a stationary AR(2)
SV-J	same as SV but the prices equation has a "jump" component
SV-M	same as SV but h_t enters the prices equation as a covariate
SV-MA	same as SV but the observation error follows an MA(1)
SV-t	same as SV but the observation error follows a t distribution
SV-L	SV with a leverage effect

3 Model Comparison Using the Bayes Factor

In this section, we give an overview of Bayesian model comparison via the Bayes factor and outline an efficient approach to compute the Bayes factor using importance sampling.

Suppose we wish to compare a set of models $\{M_1, \ldots, M_K\}$. Each model M_k is formally defined by two separate components: a likelihood function $p(\mathbf{y} \mid \boldsymbol{\theta}_k, M_k)$ that depends on the model-specific parameter vector $\boldsymbol{\theta}_k$ of dimension p_k and a prior density $p(\boldsymbol{\theta}_k \mid M_k)$. One popular Bayesian model comparison criterion is the *Bayes factor* in favor of M_i against M_j, defined as

$$\mathrm{BF}_{ij} = \frac{p(\mathbf{y} \mid M_i)}{p(\mathbf{y} \mid M_j)},$$

where

$$p(\mathbf{y} \mid M_k) = \int p(\mathbf{y} \mid \boldsymbol{\theta}_k, M_k) p(\boldsymbol{\theta}_k \mid M_k) \mathrm{d}\boldsymbol{\theta}_k \tag{5}$$

is the *marginal likelihood* under model M_k, $k = i, j$. This marginal likelihood can be interpreted as a density forecast of the data under model M_k evaluated at the actual observed data $\mathbf{y}$. Hence, if the observed data are likely under the model, the associated marginal likelihood would be "large". Since the marginal likelihood is essentially a density forecast evaluation, it has a built-in penalty for model complexity. In addition, it follows that $\mathrm{BF}_{ij} > 1$ indicates that the observed data are more likely under model M_i compared to model M_j, and is thus viewed as evidence in favor of model M_i—the weight of evidence is proportional to the value of the Bayes factor.

Furthermore, the log Bayes factor is asymptotically equivalent to the *Schwarz information criterion* (SIC) proposed by Schwarz (1978). More specifically, recall that the SIC for model M_k is defined as

$$\mathrm{SIC}_k = \log p(\mathbf{y} \mid \widehat{\boldsymbol{\theta}}_k, M_k) - \frac{p_k}{2} \log T,$$

where $\widehat{\boldsymbol{\theta}}_k$ is the maximum likelihood estimate and T is the sample size. Then, it can be shown that (see, e.g., Kass and Raftery, 1995)

$$\frac{(\mathrm{SIC}_i - \mathrm{SIC}_j) - \log \mathrm{BF}_{ij}}{\log \mathrm{BF}_{ij}} \to 0$$

as $T \to \infty$. In addition, both the Bayes factor and SIC are consistent model selection criteria—i.e., they will asymptotically select the candidate model having the correct structure with probability one.[2]

For a finite sample, the Bayes factor also has a natural interpretation. In particular, it is

[2]On the other hand, the commonly-used Akaike information criterion (AIC) is not consistent. For a more detailed discussion on the differences between the AIC, SIC and Bayes factor, see Kass and Raftery (1995).

related to the *posterior odds ratio* between the two models as follows:

$$\frac{\mathbb{P}(M_i \,|\, \mathbf{y})}{\mathbb{P}(M_j \,|\, \mathbf{y})} = \frac{\mathbb{P}(M_i)}{\mathbb{P}(M_j)} \times \mathrm{BF}_{ij},$$

where $\mathbb{P}(M_i)/\mathbb{P}(M_j)$ is the prior odds ratio. If both models are equally probable *a priori*, i.e., the prior odds ratio is one, the posterior odds ratio between the two models is then equal to the Bayes factor. Then, if, for example, $\mathrm{BF}_{ij} = 10$, then model M_i is 10 times more likely than model M_j given the data. For a more detailed discussion of the Bayes factor, we refer the readers to Kass and Raftery (1995) and Koop (2003).

Since the Bayes factor is simply a ratio of two marginal likelihoods, researchers often only report the marginal likelihoods of the set of competing models. We follow this practice. Next, we outline a method for calculating the marginal likelihoods under the GARCH and stochastic volatility models.

Generally the computation of the marginal likelihood is nontrivial—the integral in (5) is often high-dimensional and cannot be obtained analytically. In this paper we follow Chan and Eisenstat (2015), who implement an improved version of an adaptive importance sampling method called the cross-entropy method (Rubinstein, 1997; Rubinstein and Kroese, 2004) to compute the marginal likelihood.

The main idea is as follows. There is an ideal importance sampling density that would in principle give a zero-variance importance sampling estimator for the marginal likelihood. However, this density is only known up to a constant and therefore cannot be used as an importance sampling density—which would require the normalizing constant to be known. One way to get around this problem is to instead locate a density within a convenient family of distributions such that its Kullback-Leibler divergence—or the cross-entropy distance—to the ideal density is minimized. Once the optimal density is obtained, it is used to construct the importance sampling estimator.

The main advantage of this adaptive importance sampling method is that it is easy to implement and the numerical standard error of the estimator is readily available. The method only requires the evaluation of the prior and the likelihood. For GARCH models the likelihood can be quickly evaluated. For stochastic volatility models, the *complete-data* likelihood—i.e., the joint distribution of the data and the log-volatilities—can be readily evaluated. But the likelihood or more precisely the *observed-data* likelihood—i.e, the marginal distribution of the data unconditional on the log-volatilities—does not have a closed-form expression. Instead, we use the importance sampling algorithms in Chan

and Grant (2014) to evaluate the observed-data likelihood.[3] We leave the technical details to Appendix B.

4 Empirical Results

In this section we compare the performance of the seven GARCH models and their stochastic volatility counterparts discussed in Section 2 in fitting weekly price changes of nine series of energy prices. The main goal of this exercise is to examine which class of time-varying volatility models tends to fit the dynamics of a broad range of energy prices better—while controlling for model complexity. In addition, we also seek to establish the type of features that are useful in modeling these prices. For example, does adding an additional channel of volatility feedback fit the data better? Or is it more important to allow for short-run dynamics via a moving average component?

Table 2: Energy price data.

Crude oil (US Dollars per Barrel)	
S1	Cushing, OK West Texas Intermediate
S2	Europe Brent
Petroleum products (US Dollars per Gallon)	
S3	NY Harbor Conventional Gasoline Regular
S4	US Gulf Coast Conventional Gasoline Regular
S5	NY Harbor No. 2 Heating Oil
S6	Los Angeles, CA Ultra-Low Sulfur CARB Diesel
S7	US Gulf Coast Kerosene-Type Jet Fuel
S8	Mont Belvieu, TX Propane
Natural gas (US Dollars per Million Btu)	
S9	Henry Hub Natural Gas

With these aims in mind, we choose a broad range of energy prices that are commonly used in empirical applications. More specifically, we obtain the nine series of (FOB) spot prices of crude oil, various petroleum products and natural gas from the US Energy

[3]It is also worth noting that the method of Gelfand and Dey (1994) is often used in conjunction with the complete-data likelihood to compute the marginal likelihood. However, using an empirical example, Chan and Grant (2015) show that this approach can have a substantial finite-sample bias in the marginal likelihood estimate.

Information Administration. The specific details of each data series are contained in Table 2. The data frequency is weekly and the sample period is from 3 January 1997 to 6 February 2015. The data are transformed into nominal rates of change by taking the first difference of the logs and multiplying by 100.

4.1 Model Comparison Results

All the models are estimated using the Bayesian techniques outlined in Appendix A. The marginal likelihoods are computed using the improved cross-entropy method of Chan and Eisenstat (2015), which is discussed in Appendix B. The results are reported in Table 3.

A few broad conclusions can be drawn from this exercise. Overall, the best model is the SV-MA for all the nine time series. The second place is less clear-cut; the SV-2 often comes out on top, although the GARCH-MA does better for three series and the SV-L one. Second, with the notable exceptions of the GARCH-J-vs-SV-J and GARCH-t-vs-SV-t pairs, the SV models always outperform their GARCH counterparts. As an example, consider the results for S1 (Cushing, OK West Texas Intermediate). The log marginal likelihoods of the GARCH and SV models are, respectively, -2647.0 and -2632.4. This implies a Bayes factor of 2.2×10^6 in favor of the SV model against its GARCH counterpart, indicating overwhelming evidence for the former model. For the same series, the Bayes factor in favor of the SV-MA against the GARCH-MA is 2.7×10^6, again showing overwhelming evidence in support of the former model.

The exceptions of this general pattern are the GARCH-J-vs-SV-J and GARCH-t-vs-SV-t pairs: both GARCH models perform slightly better than their SV counterparts for two out of the nine series. As mentioned earlier, under GARCH models the conditional variance is a deterministic function of the parameters and past data—in contrast to stochastic volatility models, in which the log-volatility is a random variable. As such, stochastic volatility models are more robust to misspecification and to drastic changes in the time series. This helps explain why they tend to outperform their GARCH counterparts. However, when a heavy-tailed distribution such as the t distribution is used or a jump component—which is a random variable—is added to a GARCH, it gives the model extra flexibility against misspecification and outliers, making the inherent advantage of SV models less apparent. This also explains why both the GARCH-J and GARCH-t do substantially better than the GARCH in the model comparison exercise,[4] whereas the

[4]Mason and Wilmot (2014) also find that allowing for a jump process substantially improves the fit

SV-J, SV-t and SV give very similar results—e.g., in the SV a "jump" can be partially accommodated by a large shock in the stochastic volatility process.

Table 3: Log marginal likelihoods of the GARCH and SV models for the nine series of weekly price changes on energy prices. The numerical standard errors are in parentheses.

	S1	S2	S3	S4	S5	S6	S7	S8	S9
GARCH	-2647.0	-2652.0	-2809.7	-2840.0	-2647.0	-2718.3	-2634.5	-2679.2	-3117.0
	(0.05)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.10)
SV	-2632.4	-2636.7	-2783.8	-2822.6	-2612.2	-2693.5	-2623.1	-2644.0	-3058.6
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.06)	(0.04)
GARCH-2	-2647.2	-2652.2	-2807.6	-2839.4	-2647.3	-2712.1	-2634.4	-2679.7	-3118.1
	(0.10)	(0.09)	(0.04)	(0.05)	(0.11)	(0.04)	(0.06)	(0.04)	(0.04)
SV-2	-2631.9	-2636.4	-2783.5	-2821.9	-2611.8	-2692.2	-2622.6	-2641.2	-3057.9
	(0.08)	(0.07)	(0.06)	(0.11)	(0.06)	(0.10)	(0.10)	(0.08)	(0.06)
GARCH-J	-2634.6	-2638.4	-2789.7	-2822.1	-2613.0	-2700.1	-2627.7	-2650.3	-3061.0
	(0.03)	(0.07)	(0.05)	(0.05)	(0.11)	(0.06)	(0.03)	(0.06)	(0.06)
SV-J	-2632.3	-2637.1	-2784.4	-2823.1	-2613.4	-2694.3	-2624.0	-2645.1	-3059.2
	(0.04)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)	(0.02)	(0.06)	(0.04)
GARCH-M	-2653.6	-2658.8	-2817.0	-2847.3	-2654.0	-2725.4	-2641.3	-2680.6	-3123.6
	(0.02)	(0.03)	(0.04)	(0.02)	(0.03)	(0.02)	(0.02)	(0.03)	(0.04)
SV-M	-2637.8	-2642.9	-2790.3	-2828.8	-2618.8	-2699.9	-2629.1	-2646.2	-3065.3
	(0.02)	(0.02)	(0.13)	(0.04)	(0.02)	(0.04)	(0.03)	(0.04)	(0.03)
GARCH-MA	-2630.7	-2620.4	-2798.6	-2829.5	-2621.8	-2673.5	-2616.3	-2653.6	-3107.7
	(0.03)	(0.04)	(0.02)	(0.02)	(0.16)	(0.07)	(0.04)	(0.02)	(0.04)
SV-MA	-2615.9	-2610.1	-2771.8	-2806.6	-2588.6	-2648.6	-2606.6	-2622.1	-3048.7
	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.02)	(0.01)	(0.05)	(0.02)
GARCH-t	-2636.5	-2641.4	-2788.1	-2824.8	-2616.3	-2698.9	-2627.6	-2641.4	-3058.3
	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)	(0.15)	(0.01)	(0.03)
SV-t	-2632.3	-2636.9	-2784.0	-2823.6	-2612.9	-2693.4	-2623.9	-2644.5	-3059.2
	(0.02)	(0.01)	(0.02)	(0.01)	(0.02)	(0.02)	(0.03)	(0.05)	(0.04)
GARCH-GJR	-2646.6	-2649.8	-2812.5	-2842.5	-2649.9	-2719.3	-2637.1	-2679.7	-3113.0
	(0.04)	(0.04)	(0.07)	(0.03)	(0.03)	(0.03)	(0.06)	(0.11)	(0.09)
SV-L	-2628.5	-2633.5	-2785.3	-2823.8	-2614.0	-2695.3	-2624.7	-2646.2	-3059.7
	(0.02)	(0.02)	(0.01)	(0.02)	(0.01)	(0.02)	(0.01)	(0.03)	(0.03)

Now, we further investigate which features are important in modeling the dynamics of energy prices. By comparing the GARCH with GARCH-2 and the SV with SV-2, we conclude that the richer AR(2) volatility process provides only marginal benefits. For example, the Bayes factor in favor of the SV-2 against the SV is only 2 for S4 (US Gulf Coast Conventional Gasoline Regular). Results for the other series are broadly

of the GARCH model for natural gas spot prices.

similar. Thus, for modeling energy prices at least, one can feel comfortable maintaining the conventional specification of an AR(1) volatility process.

Next, we examine the importance of volatility feedback for modeling energy prices. Although this channel is found to be empirically important for stock returns, it is superfluous for energy data. In fact, adding the volatility feedback component often markedly decreases the marginal likelihood of a model. For instance, the Bayes factor in favor of the GARCH against the GARCH-M is about 1200 for S6 (Los Angeles, CA Ultra-Low Sulfur CARB Diesel). This is in line with the finding in Sadorsky (2006), who finds that GARCH in mean forecasts no better than the standard GARCH for crude oil, heating oil and natural gas volatility. It is also worth noting that even though the GARCH-M nests GARCH as a special case—hence, the GARCH-M would fit the data better—the Bayes factor still prefers the simpler model, highlighting its built-in penalty against model complexity.

To investigate the relevance of the moving average component, we compare the GARCH with GARCH-MA and the SV with SV-MA. For both classes of models, adding the MA component drastically improves the model-fit for all series. For example, the Bayes factor in favor of the SV-MA against the SV is 3.2×10^9 for S8 (Mont Belvieu, TX Propane), indicating that the weekly returns exhibit substantial serial correlation (see also the model diagnostic tests in the next section). Similar results are obtained for the GARCH models and for the other price series.

Finally, by comparing the GARCH with GARCH-GJR and the SV with SV-L, we conclude that the leverage effect is important for modeling crude oil prices (WTI and Brent) but not for other energy prices. Since it is well-known that leverage effects are important for equity returns, these results imply that crude oil prices behave more like equity data than other energy prices.

4.2 Estimation Results

In this section we report the posterior estimates of the model parameters for both the GARCH and stochastic volatility models. Due to space constraint, we only present results for the crude oil price (Cushing, OK West Texas Intermediate), which are broadly representative of the estimates for other energy prices.

Table 4 shows the results for the GARCH models. The parameters governing the evolution

of the conditional variance process have similar estimates across models. In particular, all indicate high persistence with the posterior mean of β_1 estimated to be between 0.74 to 0.9—the lowest estimate is from the GARCH-2, in which the sum of β_1 and β_2 is estimated to be 0.88, implying a persistence level similar to other models. The estimate of β_2 is small and an AR(1) for the conditional variance process appears to be sufficient, which supports the ranking of the marginal likelihood.

The average jump size μ_k is estimated to be negative at about -1.4%. The estimate for the jump probability κ is 0.05, which implies about 2.5 "jumps" per year for weekly data. It is interesting to note that the posterior estimates of λ, ψ and δ_1 all seem to support the ranking of the marginal likelihood. For example, recall that when $\psi = 0$, the GARCH-MA reduces to the standard GARCH. Since the marginal likelihood favors the GARCH-MA relative to the GARCH, one would expect that the posterior distribution of ψ has little mass around zero. In fact, the 95% credible interval of ψ is estimated to be $(0.16, 0.31)$, which excludes 0. Similarly, when $\lambda = 0$, the GARCH-M reduces to the standard GARCH. The 95% credible interval of λ is estimated to be $(-0.01, 0.04)$, which includes 0, supporting the ranking of the marginal likelihood that favors the GARCH over the GARCM-M. Next, the estimate of δ_1 is 0.06 and is statistically different from 0, implying a negative shock at time t would increase the conditional variance at time $t+1$.

Finally, the degree of freedom parameter ν is estimated to be about 11, indicating that the tails of the t distribution are relatively heavy—i.e., outliers occur relatively frequently. This supports the ranking of the marginal likelihood that prefers the GARCH-t model compared to the standard GARCH.

We also present the results of two diagnostic tests. Specifically, we report the Ljung-Box and McLeod-Li statistics of order 20 computed on the standardized residuals and squared standardized residuals, respectively. Except for the GARCH-MA model, all the Ljung-Box tests reject the null hypothesis of no serial correlation in the standardized residuals at the 1% level. This diagnostic result again supports the ranking of the marginal likelihood, which favors the GARCH-MA relative to the GARCH. In addition, all the McLeod-Li tests fail to reject the null hypothesis of no serial correlation in the squared standardized residuals at the 5% level, indicating that the GARCH models adequately capture the time-varying volatility of the data.

Table 4: Parameter posterior means and standard deviations (in parentheses) for the GARCH models (S1: Cushing, OK West Texas Intermediate).

	GARCH	GARCH-2	GARCH-J	GARCH-M	GARCH-MA	GARCH-t	GARCH-GJR
μ	0.15	0.15	0.20	-0.10	0.16	0.18	0.09
	(0.12)	(0.12)	(0.13)	(0.23)	(0.14)	(0.12)	(0.12)
α_0	0.65	0.51	0.41	0.48	0.37	0.39	0.50
	(0.22)	(0.18)	(0.16)	(0.17)	(0.14)	(0.20)	(0.18)
α_1	0.10	0.10	0.08	0.08	0.08	0.08	0.05
	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)	(0.02)
β_1	0.87	0.74	0.88	0.89	0.90	0.88	0.89
	(0.03)	(0.04)	(0.02)	(0.02)	(0.01)	(0.03)	(0.02)
β_2	–	0.14	–	–	–	–	–
	–	(0.03)	–	–	–	–	–
κ	–	–	0.05	–	–	–	–
	–	–	(0.02)	–	–	–	–
μ_k	–	–	-1.39	–	–	–	–
	–	–	(0.93)	–	–	–	–
σ_k^2	–	–	41.26	–	–	–	–
	–	–	(20.82)	–	–	–	–
λ	–	–	–	0.02	–	–	–
	–	–	–	(0.01)	–	–	–
ψ	–	–	–	–	0.24	–	–
	–	–	–	–	(0.04)	–	–
ν	–	–	–	–	–	10.86	–
	–	–	–	–	–	(2.36)	–
δ_1	–	–	–	–	–	–	0.06
	–	–	–	–	–	–	(0.02)
$Q(20)$	66.4	65.5	65.8	67.7	37.5	66.0	64.6
	(1.23)	(0.65)	(1.37)	(2.03)	(1.58)	(1.54)	(1.33)
$Q^2(20)$	15.1	14.7	13.8	15.8	14.9	14.7	13.8
	(1.76)	(1.19)	(1.26)	(3.35)	(1.62)	(1.96)	(1.24)

Notes: $Q(20)$ and $Q^2(20)$ are respectively the Ljung-Box and McLeod-Li statistics of order 20 computed on the standardized residuals and squared standardized residuals. The 5% and 1% critical values are 31.41 and 37.57, respectively.

Next, we present the parameter estimates for the stochastic volatility models in Table 5. Similar to the estimates under GARCH models, the stochastic volatility process is highly persistent for all models. In particular, the posterior mean of ϕ_h is estimated to be between 0.96 to 0.97 across the various models. Other parameters governing the stochastic volatility process are also similar across models.

In contrast to the GARCH-J results, the average jump size μ_k under SV-J is estimated

to be positive (but with a large posterior standard deviation). Seemingly very different "jumps" are identified under the SV-J compared to the GARCH-J. Given the small estimate for μ_k, the marginal likelihood favors the SV model compared to SV-J, whereas among the GARCH and GARCH-J pair it prefers the latter model.

Table 5: Parameter posterior means and standard deviations (in parentheses) for the stochastic volatility models (S1: Cushing, OK West Texas Intermediate).

	SV	SV-2	SV-J	SV-M	SV-MA	SV-t	SV-L
μ	0.17	0.19	0.18	0.47	0.16	0.18	0.10
	(0.12)	(0.12)	(0.14)	(0.24)	(0.14)	(0.12)	(0.12)
μ_h	2.67	2.66	2.63	2.67	2.63	2.63	2.68
	(0.22)	(0.25)	(0.23)	(0.22)	(0.25)	(0.23)	(0.23)
ϕ_h	0.97	0.96	0.97	0.97	0.97	0.97	0.97
	(0.01)	(0.08)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
ω_h^2	0.03	0.05	0.03	0.03	0.03	0.03	0.03
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
ρ_h	–	-0.03	–	–	–	–	–
	–	(0.09)	–	–	–	–	–
κ	–	–	0.04	–	–	–	–
	–	–	(0.03)	–	–	–	–
μ_k	–	–	0.22	–	–	–	–
	–	–	(1.67)	–	–	–	–
σ_k^2	–	–	18.56	–	–	–	–
	–	–	(22.06)	–	–	–	–
λ	–	–	–	-0.02	–	–	–
	–	–	–	(0.02)	–	–	–
ψ	–	–	–	–	0.22	–	–
	–	–	–	–	(0.03)	–	–
ν	–	–	–	–	–	56.13	–
	–	–	–	–	–	(23.32)	–
ρ	–	–	–	–	–	–	-0.38
	–	–	–	–	–	–	(0.11)
$Q(20)$	64.5	64.4	61.6	64.1	34.5	64.2	63.7
	(3.93)	(4.32)	(6.50)	(4.17)	(3.51)	(3.93)	(3.67)
$Q^2(20)$	17.3	19.1	17.8	17.1	16.6	17.1	17.1
	(3.76)	(7.18)	(4.80)	(3.76)	(3.80)	(3.90)	(3.48)

Notes: $Q(20)$ and $Q^2(20)$ are respectively the Ljung-Box and McLeod-Li statistics of order 20 computed on the standardized residuals and squared standardized residuals. The 5% and 1% critical values are 31.41 and 37.57, respectively.

The estimate of the moving average parameter is similar across the GARCH-MA and SV-MA models, with ψ estimated to be 0.22 under the SV-MA. Its 95% credible interval

is (0.15, 0.29), which excludes 0. Hence, these estimation results also support the ranking of the marginal likelihood which favors the SV-MA against the standard SV. Next, the coefficient of the volatility λ is estimated to be -0.02 with a 95% credible interval (-0.05, 0.01), indicating that volatility feedback is unimportant for modeling crude oil returns.

In contrast to the GARCH-t results, the estimate of the degree of freedom parameter ν under the SV-t is substantially larger—the posterior mean is about 56—indicating that the tails of the t distribution are thin and similar to those of the Gaussian distribution. Accordingly, both the SV and SV-t receive similar support from the marginal likelihood. Lastly, the correlation ρ between the observation and state innovations is estimated to be negative at -0.38. This implies a negative shock at time t tends to increase the volatility at time $t + 1$.

Finally, we also report the Ljung-Box and McLeod-Li statistics. Similar to the GARCH results, all the Ljung-Box tests reject the null hypothesis of no serial correlation in the standardized residuals at the 1% level, with the exception of the SV-MA model. This again supports the ranking of the marginal likelihood, which favors the SV-MA compared to the SV.

4.3 Forecasting Results

In this section we perform a recursive out-of-sample forecasting exercise to evaluate the performance of the GARCH and stochastic volatility models. These models are often compared by evaluating their volatility forecasts (e.g., Sadorsky, 2006; Wei et al., 2010) or the Values-at-Risk (e.g., Fan, Zhang, Tsai, and Wei, 2008; Hung, Lee, and Liu, 2008), which are tail quantiles of the return densities. Here we compare the models using the entire forecast densities by computing the log predictive score.

More specifically, given the data up to time t, denoted as $\mathbf{y}_{1:t}$, we compute the one-step-ahead predictive density $p(y_{t+1} \,|\, \mathbf{y}_{1:t})$ under a certain model and use it as the density forecast for y_{t+1}. This density forecast is evaluated by computing the log predictive likelihood $\log p(y_{t+1} = y_{t+1}^{o} \,|\, \mathbf{y}_{1:t})$, i.e., the log predictive density of y_{t+1} evaluated at the observed value y_{t+1}^{o}. Clearly, if the actual outcome y_{t+1}^{o} is likely under the density forecast, the value of the log predictive likelihood will be large, and vice versa. Next, we move one period forward and repeat the whole exercise with data $\mathbf{y}_{1:t+1}$, and so forth. Finally, the log predictive score for the evaluation period $t_0 + 1, \ldots, T$ is defined as the sum of the

log predictive likelihoods:

$$\sum_{t=t_0}^{T-1} \log p(y_{t+1} = y_{t+1}^{o} \,|\, \mathbf{y}_{1:t}).$$

A larger log predictive score indicates better forecast performance. See, e.g., Geweke and Amisano (2011) for a more detailed discussion of the log predictive score and its connection to the marginal likelihood. Below present results for the crude oil price (Europe Brent). The evaluation period for the forecasting exercise is from January 2000 to the end of the sample. The results are reported in Table 6.

Table 6: Log predictive scores of the GARCH and SV models for forecasting changes in crude oil prices (S2: Europe Brent).

GARCH	GARCH-2	GARCH-J	GARCH-M	GARCH-MA	GARCH-t	GARCH-GJR
-2174.9	-2173.8	-2163.1	-2177.6	-2146.6	-2165.2	-2169.0
SV	SV-2	SV-J	SV-M	SV-MA	SV-t	SV-L
-2160.6	-2170.1	-2162.4	-2162.6	-2137.2	-2161.2	-2157.4

These density forecasting results are broadly similar to the model comparison results using the marginal likelihood. In particular, the SV models provide better density forecasts than their GARCH counterparts. Moreover, the jump component and t distributed innovations improve forecasts for GARCH, but they are unimportant for the SV model. Finally, the SV model with moving average innovations is the best forecasting model for Europe Brent.

5 Concluding Remarks and Future Research

We have undertaken a formal Bayesian model comparison exercise to assess a number of GARCH and stochastic volatility models for modeling oil, petroleum product and natural gas prices. Using the marginal likelihood to assess the various models, we find that stochastic volatility models almost always outperform their GARCH counterparts, suggesting that stochastic volatility models might provide a better alternative to the more conventional GARCH models. Overall, the stochastic volatility model with moving average innovations is the best model for all nine series considered.

For future research, it would be worthwhile to compare multivariate GARCH and stochastic volatility models in fitting multiple energy prices. In particular, it would be important

to determine the number of sources of fluctuation in these series. Moreover, it would also be interesting to incorporate macroeconomic variables in the multivariate analysis, as the interplay between energy prices and macroeconomic variables is often of interest.

Appendix A: Priors and Estimation

In this appendix we discuss the priors and outline the estimation methods for fitting the GARCH and stochastic volatility models discussed in Section 2.

Priors

We choose broadly similar priors across the GARCH and stochastic volatility models. In particular, we use the same prior for common parameters. All priors are proper but relatively noninformative.

For the standard GARCH, we assume the following independent priors for μ and $\boldsymbol{\gamma} = (\alpha_0, \alpha_1, \beta_1)'$:

$$\mu \sim \mathcal{N}(\mu_0, V_\mu), \qquad \log \boldsymbol{\gamma} \sim \mathcal{N}(\boldsymbol{\gamma}_0, \mathbf{V}_{\boldsymbol{\gamma}}) \mathbb{1}(\alpha_1 + \beta_1 < 1), \tag{6}$$

that is, $\boldsymbol{\gamma}$ follows a truncated log-normal distribution with the stationarity restriction that $\alpha_1 + \beta_1 < 1$. We set the hyperparameters to be $\mu_0 = 0$, $V_\mu = 10$, $\boldsymbol{\gamma}_0 = (1, \log 0.1, \log 0.8)'$ and $\mathbf{V}_{\boldsymbol{\gamma}} = \text{diag}(10, 1, 1)$. These values imply relatively noninformative priors with prior medians that are similar to typical estimates from financial data. In particular, the prior medians of μ and $\boldsymbol{\gamma}$ are respectively 0 and $(2.72, 0.1, 0.8)'$. For the GARCH-2, we use the same prior for μ, but replace the prior for $\boldsymbol{\gamma}$ with a prior for $\widetilde{\boldsymbol{\gamma}} = (\alpha_0, \alpha_1, \beta_1, \beta_2)'$ where $\log \widetilde{\boldsymbol{\gamma}} \sim \mathcal{N}(\widetilde{\boldsymbol{\gamma}}_0, \mathbf{V}_{\widetilde{\boldsymbol{\gamma}}}) \mathbb{1}(\alpha_1 + \beta_1 + \beta_2 < 1)$ with $\widetilde{\boldsymbol{\gamma}}_0 = (1, \log 0.1, \log 0.8, \log 0.1)'$ and $\mathbf{V}_{\widetilde{\boldsymbol{\gamma}}} = \text{diag}(10, 1, 1, 1)$.

For each of the remaining GARCH models, the priors for μ and $\boldsymbol{\gamma}$ are exactly the same as in (6). Moreover, under the GARCH-J, the jump intensity κ is assumed to have a uniform distribution on the interval $(0, 0.1)$, and the average jump size and the jump variance $\boldsymbol{\delta} = (\mu_k, \log \sigma_k^2)'$ are distributed as a bivariate normal distribution: $\kappa \sim \mathcal{U}(0, 0.1)$ and $\boldsymbol{\delta} \sim \mathcal{N}(\boldsymbol{\delta}_0, \mathbf{V}_{\boldsymbol{\delta}})$. We set $\boldsymbol{\delta}_0 = (0, \log 10)'$ and $\mathbf{V}_{\boldsymbol{\delta}} = \text{diag}(10, 1)$ so that the average jump size is 0. For the GARCH-M, the coefficient of the volatility is assumed to have a normal distribution: $\lambda \sim \mathcal{N}(\lambda_0, V_\lambda)$, where $\lambda_0 = 0$ and $V_\lambda = 100$. Next, the MA(1) coefficient in the GARCH-MA has a normal distribution truncated within the unit interval: $\psi \sim \mathcal{N}(\psi_0, V_\psi) \mathbb{1}(|\psi| < 1)$, where $\psi_0 = 0$ and $V_\psi = 1$. For the GARCH-t, the degree of freedom parameter ν has a uniform distribution on $(2, 100)$: $\nu \sim \mathcal{U}(2, 100)$. Note that we assume $\nu > 2$ to ensure that the first two moments of the t distribution exist. Lastly, for the leverage parameter δ_1 in the GARCH-GJR, we assume a uniform prior conditional on $\boldsymbol{\gamma}$, so long as the variance process is strictly positive and stationary. In particular, we

impose the conditions $\alpha_1 + \delta_1 > 0$ and $\alpha_1 + \beta_1 + \delta_1 < 1$. Hence, we assume the uniform prior $(\delta_1 \,|\, \boldsymbol{\gamma}) \sim \mathcal{U}(-\alpha_1, 1 - \alpha_1 - \beta_1)$.

Next, we discuss the set of priors for each of the stochastic volatility models. In general, we choose the same hyperparameters for parameters that are common across models. For the standard SV, we assume the following independent priors for μ, μ_h, ϕ_h and ω_h^2:

$$\begin{aligned} \mu &\sim \mathcal{N}(\mu_0, V_\mu), & \mu_h &\sim \mathcal{N}(\mu_{h0}, V_{\mu_h}), \\ \phi_h &\sim \mathcal{N}(\phi_{h0}, V_{\phi_h}) \mathbb{1}(|\phi_h| < 1), & \omega_h^2 &\sim \mathcal{IG}(\nu_h, S_h), \end{aligned} \tag{7}$$

where $\mathcal{IG}(\cdot, \cdot)$ denotes the inverse-gamma distribution. We set $\mu_0 = 0$, $\mu_{h0} = 1$, $V_\mu = V_{\mu_h} = 10$, $\phi_{h0} = 0.97$, $V_{\phi_h} = 0.1^2$, $\nu_h = 5$ and $S_h = 0.16$. These hyperparameters are set so that the stochastic volatility process has similar dynamics as the conditional variance under the GARCH models.

For the SV-2, we assume the same priors for μ, μ_h and ω_h^2 as in (7), but replace the prior for ϕ_h with a prior for $\boldsymbol{\theta}_h = (\phi_h, \rho_h)'$: $\boldsymbol{\theta}_h \sim \mathcal{N}(\boldsymbol{\theta}_{h0}, \mathbf{V}_{\boldsymbol{\theta}_h}) \mathbb{1}(\boldsymbol{\theta}_h \in \mathbf{A})$, where $\boldsymbol{\theta}_{h0} = (0.97, 0)'$, $\mathbf{V}_{\boldsymbol{\theta}_h} = \text{diag}(0.1^2, 1)$ and $\mathbf{A} \subset \mathbb{R}^2$ is the set where the roots of the characteristic polynomial defined by $\boldsymbol{\theta}_h$ lie outside the unit circle. For each of the remaining stochastic volatility models, the priors for μ, μ_h, ϕ_h and ω_h^2 are the same as in (7). The additional parameters have exactly the same priors as their GARCH counterparts. Lastly, for the SV-L, the correlation parameter ρ is assumed to have the uniform prior $\rho \sim \mathcal{U}(-1, 1)$.

Bayesian Estimation

All the GARCH and stochastic volatility models are estimated using Markov chain Monte Carlo (MCMC) methods. Specifically, we sample from the posterior distributions of the models by constructing Markov samplers and use the posterior draws obtained to compute various quantities of interest such as the posterior means and the marginal likelihoods.

For the stochastic volatility models, a key step is to jointly sample the log-volatilities. For example, under the standard SV model, we need to sample from the conditional density $p(\mathbf{h} \,|\, \mathbf{y}, \mu, \mu_h, \phi_h, \omega_h^2)$. This is done using the acceptance-rejection Metropolis-Hastings algorithm described in Chan (2015), which is based on the precision sampler of Chan and Jeliazkov (2009). A key feature of this algorithm is its use of fast band matrix routines rather than using the conventional Kalman filter. The former approach is in general more efficient than the latter.

To implement the acceptance-rejection Metropolis-Hastings algorithm, one key ingredient is an appropriate proposal density that well approximates the target $p(\mathbf{h} \,|\, \mathbf{y}, \mu, \mu_h, \phi_h, \omega_h^2)$. The basic idea is to approximate the target using a Gaussian density. To that end, note that $p(\mathbf{h} \,|\, \mathbf{y}, \mu, \mu_h, \phi_h, \omega_h^2) \propto p(\mathbf{y} \,|\, \mu, \mathbf{h}) p(\mathbf{h} \,|\, \mu_h, \phi_h, \omega_h^2)$. It can be shown that the latter density $p(\mathbf{h} \,|\, \mu_h, \phi_h, \omega_h^2)$ is Gaussian. In fact, we have

$$\log p(\mathbf{h} \,|\, \mu_h, \phi_h, \omega_h^2) = -\frac{1}{2}(\mathbf{h}' \mathbf{H}_{\phi_h}' \boldsymbol{\Sigma}_h^{-1} \mathbf{H}_{\phi_h} \mathbf{h} - 2\mathbf{h}' \mathbf{H}_{\phi_h}' \boldsymbol{\Sigma}_h^{-1} \mathbf{H}_{\phi_h} \boldsymbol{\delta}_h) + c_1, \tag{8}$$

where c_1 is a constant independent of $\mathbf{h}$,

$$\mathbf{H}_{\phi_h} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ -\phi_h & 1 & 0 & \cdots & 0 \\ 0 & -\phi_h & 1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & -\phi_h & 1 \end{pmatrix}$$

is a lower triangular matrix, $\boldsymbol{\Sigma}_h = \text{diag}(\omega_h^2/(1-\phi_h^2), \omega_h^2, \ldots, \omega_h^2)$ and $\boldsymbol{\delta}_h = \mathbf{H}_{\phi_h}^{-1} \widetilde{\boldsymbol{\delta}}_h$ with $\widetilde{\boldsymbol{\delta}}_h = (\mu_h, (1-\phi_h)\mu_h, \ldots, (1-\phi_h)\mu_h)'$.

Next, we approximate $p(\mathbf{y} \,|\, \mu, \mathbf{h})$ by a Gaussian density in $\mathbf{h}$. To that end, we expand $\log p(\mathbf{y} \,|\, \mu, \mathbf{h}) = \sum_{t=1}^T \log p(y_t \,|\, \mu, h_t)$ around the point $\widetilde{\mathbf{h}}$—which is chosen to be the mode of $p(\mathbf{h} \,|\, \mathbf{y}, \mu, \mu_h, \phi_h, \omega_h^2)$—by a second-order Taylor expansion:

$$\begin{aligned} \log p(\mathbf{y} \,|\, \mu, \mathbf{h}) \approx & \log p(\mathbf{y} \,|\, \mu, \widetilde{\mathbf{h}}) + (\mathbf{h} - \widetilde{\mathbf{h}})' \mathbf{f} - \frac{1}{2}(\mathbf{h} - \widetilde{\mathbf{h}})' \mathbf{G} (\mathbf{h} - \widetilde{\mathbf{h}}) \\ = & -\frac{1}{2}(\mathbf{h}' \mathbf{G} \mathbf{h} - 2\mathbf{h}'(\mathbf{f} + \mathbf{G}\widetilde{\mathbf{h}})) + c_2, \end{aligned} \tag{9}$$

where c_2 is a constant independent of $\mathbf{h}$, $\mathbf{f} = (f_1, \ldots, f_T)'$ and $\mathbf{G} = \text{diag}(G_1, \ldots, G_T)$ with

$$f_t = \frac{\partial}{\partial h_t} \log p(y_t \,|\, \mu, h_t)|_{h_t = \widetilde{h}_t}, \qquad G_t = -\frac{\partial^2}{\partial h_t^2} \log p(y_t \,|\, \mu, h_t)|_{h_t = \widetilde{h}_t}.$$

That is, $\mathbf{G}$ is the negative Hessian of the log-density evaluated at $\widetilde{\mathbf{h}}$. For the standard stochastic volatility model, $\mathbf{G}$ is diagonal (hence a band matrix).

Finally, combining (8) and (9), we have

$$\begin{aligned}\log p(\mathbf{h} \,|\, \mathbf{y}, \mu, \mu_h, \phi_h, \omega_h^2) &= \log p(\mathbf{y} \,|\, \mu, \mathbf{h}) + \log p(\mathbf{h} \,|\, \mu_h, \phi_h, \omega_h^2) + c_3, \\ &\approx -\frac{1}{2}(\mathbf{h}'\mathbf{K}_\mathbf{h}\mathbf{h} - 2\mathbf{h}'\mathbf{k}_\mathbf{h}) + c_4, \end{aligned} \tag{10}$$

where c_3 and c_4 are constants independent of $\mathbf{h}$, $\mathbf{K}_\mathbf{h} = \mathbf{H}'_{\phi_h}\boldsymbol{\Sigma}_h^{-1}\mathbf{H}_{\phi_h} + \mathbf{G}$ and $\mathbf{k}_\mathbf{h} = \mathbf{f} + \mathbf{G}\widetilde{\mathbf{h}} + \mathbf{H}'_{\phi_h}\boldsymbol{\Sigma}_h^{-1}\mathbf{H}_{\phi_h}\boldsymbol{\delta}_\mathbf{h}$. It can be shown that the expression in (10) is the log-kernel of the $\mathcal{N}(\widehat{\mathbf{h}}, \mathbf{K}_\mathbf{h}^{-1})$ density (see, e.g., Kroese and Chan, 2014, p. 238), where $\widehat{\mathbf{h}} = \mathbf{K}_\mathbf{h}^{-1}\mathbf{k}_\mathbf{h}$. This Gaussian density with mean vector $\widehat{\mathbf{h}}$ and precision matrix $\mathbf{K}_\mathbf{h}$ is then used as the proposal density in the acceptance-rejection Metropolis-Hastings algorithm. It is important to note that $\mathbf{K}_\mathbf{h}$ is a band matrix, and consequently sampling from this proposal density is fast; see, e.g., Chan and Jeliazkov (2009). We refer the readers to Chan and Grant (2014) for the estimation details of the stochastic volatility models.

For the GARCH models, we use Metropolis-Hastings algorithms to sample from the posterior distributions. For example, for the standard GARCH, we group the parameters into two blocks—μ and $\boldsymbol{\gamma} = (\alpha_0, \alpha_1, \beta_1)'$—and we draw from the two full conditional distributions $p(\mu \,|\, \mathbf{y}, \boldsymbol{\gamma})$ and $p(\boldsymbol{\gamma} \,|\, \mathbf{y}, \mu)$ sequentially. Since both conditional distributions are nonstandard,[5] Metropolis-Hastings algorithms are required. To sample μ, we use a Gaussian proposal with mean $\bar{y}$ and variance s^2/T, where $\bar{y}$ and s^2 are the sample mean and sample variance respectively. For $\boldsymbol{\gamma}$, we use a Gaussian proposal centered at the mode of $p(\boldsymbol{\gamma} \,|\, \mathbf{y}, \mu)$ with covariance matrix set to be the outer product of the scores. For other GARCH models with additional parameters, the basic sampler remains the same but with an extra block to sample the additional parameters.

[5]The conditional distribution of μ is not Gaussian as μ also appears in the conditional variance σ_t^2.

Appendix B: Marginal Likelihood Computation

The marginal likelihoods for the GARCH and stochastic volatility models are computed using the adaptive importance sampling approach in Chan and Eisenstat (2015). More specifically, the marginal likelihood $p(\mathbf{y})$ for a given model is estimated using:

$$\widehat{p(\mathbf{y})} = \frac{1}{R}\sum_{i=1}^{R} \frac{p(\mathbf{y} \,|\, \boldsymbol{\theta}^{(i)})p(\boldsymbol{\theta}^{(i)})}{g(\boldsymbol{\theta}^{(i)})}, \tag{11}$$

where $\boldsymbol{\theta}^{(1)}, \ldots, \boldsymbol{\theta}^{(R)}$ are independent draws obtained from the importance density $g(\cdot)$ that dominates the product of the likelihood and the prior $p(\mathbf{y} \,|\, \cdot)p(\cdot)$. The importance sampling estimator (11) is an unbiased, simulation-consistent estimator of the marginal likelihood $p(\mathbf{y})$. The choice of the importance density is critical for the performance of this estimator. As outlined in Chan and Eisenstat (2015), the theoretical zero-variance importance density for estimating $p(\mathbf{y})$ is the posterior density $p(\boldsymbol{\theta} \,|\, \mathbf{y})$—which cannot be used as its normalizing constant is unknown. The improved cross-entropy method generates a procedure to construct an optimal importance density by minimizing the Kullback-Leibler divergence to the zero-variance importance density.

The optimal importance density is typically located within the same parametric class as the prior densities. For example, in the case of the standard GARCH in (1)–(2), the optimal importance density has the form: $g(\boldsymbol{\theta}) = p(\mu)p(\boldsymbol{\gamma})$, where the prior densities are given in Appendix A. Once the optimal importance density $g(\boldsymbol{\theta})$ is constructed, the importance sampling estimator in (11) can be obtained easily for the GARCH models, as the (observed-data) likelihood $p(\mathbf{y} \,|\, \boldsymbol{\theta})$ can be evaluated quickly. For instance, the log-likelihood for the standard GARCH is given by

$$\log p(\mathbf{y} \,|\, \mu, \boldsymbol{\gamma}) = -\frac{T}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{T} \log \sigma_t^2 - \frac{1}{2}\sum_{t=1}^{T} \frac{(y_t - \mu)^2}{\sigma_t^2},$$

where the conditional variance process σ_t^2 is given in (2).

For the stochastic volatility models, the observed-data likelihood $p(\mathbf{y} \,|\, \boldsymbol{\theta})$ is not available analytically and again we evaluate it using importance sampling. Recall that the observed-data likelihood is given by

$$p(\mathbf{y} \,|\, \boldsymbol{\theta}) = \int p(\mathbf{y} \,|\, \boldsymbol{\theta}, \mathbf{h}) p(\mathbf{h} \,|\, \boldsymbol{\theta}) \mathrm{d}\mathbf{h},$$

where $p(\mathbf{y} \mid \boldsymbol{\theta}, \mathbf{h})$ is the conditional likelihood and $p(\mathbf{h} \mid \boldsymbol{\theta})$ is the prior density of the log-volatilities $\mathbf{h}$. If $\mathbf{h}^{(1)}, \ldots, \mathbf{h}^{(R)}$ are independent samples from the importance density $g(\mathbf{h})$, then we can estimate the observed-data likelihood $p(\mathbf{y} \mid \boldsymbol{\theta})$ using the following importance sampling estimator:

$$\widehat{p(\mathbf{y} \mid \boldsymbol{\theta})} = \frac{1}{R} \sum_{i=1}^{R} \frac{p(\mathbf{y} \mid \boldsymbol{\theta}, \mathbf{h}^{(i)}) p(\mathbf{h}^{(i)} \mid \boldsymbol{\theta})}{g(\mathbf{h}^{(i)})}.$$

To choose a suitable $g(\cdot)$, we note that the theoretical zero-variance importance density for estimating $p(\mathbf{y} \mid \boldsymbol{\theta})$ is the conditional density $p(\mathbf{h} \mid \mathbf{y}, \boldsymbol{\theta}) \propto p(\mathbf{y} \mid \boldsymbol{\theta}, \mathbf{h}) p(\mathbf{h} \mid \boldsymbol{\theta})$. Hence, we would like to choose $g(\cdot)$ to be "close" to $p(\mathbf{h} \mid \mathbf{y}, \boldsymbol{\theta})$.

Recall that when we estimate the stochastic volatility models, one key step is to approximate the conditional distribution $p(\mathbf{h} \mid \mathbf{y}, \boldsymbol{\theta})$ using a Gaussian density (see Appendix A for details). For example, in the case of the standard SV model, we can use the Gaussian density in (10) as our importance sampling density. For each of the other stochastic volatility models, we can use a similar Gaussian approximation; see Chan and Grant (2014) for details.

References

P. Agnolucci. Volatility in crude oil futures: A comparison of the predictive ability of GARCH and implied volatility models. *Energy Economics*, 31(2):316–321, 2009.

O. J. Blanchard and M. Riggi. Why are the 2000s so different from the 1970s? A structural interpretation of changes in the macroeconomic effects of oil prices. *Journal of the European Economic Association*, 11(5):1032–1052, 2013.

T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327, 1986.

C. Brooks and M. Prokopczuk. The dynamics of commodity prices. *Quantitative Finance*, 13(4):527–542, 2013.

J. C. C. Chan. Moving average stochastic volatility models with application to inflation forecast. *Journal of Econometrics*, 176(2):162–172, 2013.

J. C. C. Chan. The stochastic volatility in mean model with time-varying parameters: An application to inflation modeling. *Journal of Business and Economic Statistics*, 2015. Forthcoming.

J. C. C. Chan and E. Eisenstat. Marginal likelihood estimation with the Cross-Entropy method. *Econometric Reviews*, 34(3):256–285, 2015.

J. C. C. Chan and A. L. Grant. Issues in comparing stochastic volatility models using the deviance information criterion. *CAMA Working Paper*, 2014.

J. C. C. Chan and A. L. Grant. Pitfalls of estimating the marginal likelihood using the modified harmonic mean. *Economics Letters*, 131:29–33, 2015.

J. C. C. Chan and I. Jeliazkov. Efficient simulation and integrated likelihood estimation in state space models. *International Journal of Mathematical Modelling and Numerical Optimisation*, 1(1):101–120, 2009.

R. F. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50(4):987–1007, 1982.

Y. Fan, Y.-J. Zhang, H.-T. Tsai, and Y.-M. Wei. Estimating value at risk of crude oil price and its spillover effect using the GED-GARCH approach. *Energy Economics*, 30 (6):3156–3171, 2008.

W. M. Fong and K. H. See. Modelling the conditional volatility of commodity index futures as a regime switching process. *Journal of Applied Econometrics*, 16(2):133–163, 2001.

A. E. Gelfand and D. K. Dey. Bayesian model choice: Asymptotics and exact calculations. *Journal of the Royal Statistical Society Series B*, 56(3):501–514, 1994.

J. Geweke and G. Amisano. Hierarchical Markov normal mixture models with applications to financial asset returns. *Journal of Applied Econometrics*, 26(1):1–29, 2011.

L. R. Glosten, R. Jagannathan, and D. E. Runkle. On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance*, 48(5):1779–1801, 1993.

J.-C. Hung, M.-C. Lee, and H.-C. Liu. Estimation of value-at-risk for energy commodities via fat-tailed GARCH models. *Energy Economics*, 30(3):1173–1191, 2008.

S. H. Kang, S. M. Kang, and S. M. Yoon. Forecasting volatility of crude oil markets. *Energy Economics*, 31(1):119–125, 2009.

R. E. Kass and A. E. Raftery. Bayes factors. *Journal of the American Statistical Association*, 90(430):773–795, 1995.

L. Kilian. Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market. *American Economic Review*, 99(3):1053–1069, 2009.

S. Kim, N. Shepherd, and S. Chib. Stochastic volatility: Likelihood inference and comparison with ARCH models. *Review of Economic Studies*, 65(3):361–393, 1998.

G. Koop. *Bayesian Econometrics*. Wiley & Sons, New York, 2003.

S. J. Koopman and E. Hol Uspensky. The stochastic volatility in mean model: Empirical evidence from international stock markets. *Journal of Applied Econometrics*, 17(6): 667–689, 2002.

D. P. Kroese and J. C. C. Chan. *Statistical Modeling and Computation*. Springer, New York, 2014.

K. Larsson and M. Nossman. Jumps and stochastic volatility in oil prices: Time series evidence. *Energy Economics*, 33(3):504–514, 2011.

M. Manera, M. Nicolini, and I. Vignati. Modelling futures price volatility in energy markets: Is there a role for financial speculation? *Energy Economics*, 2014. Forthcoming.

C. F. Mason and N. A. Wilmot. Jump processes in natural gas markets. *Energy Economics*, 46(supplement 1):S69–S79, 2014.

H. Mohammadi and L. Su. International evidence on crude oil price dynamics: Applications of ARIMA-GARCH models. *Energy Economics*, 32(5):1001–1008, 2010.

N. Nomikos and K. Andriosopoulos. Modelling energy spot prices: Empirical evidence from NYMEX. *Energy Economics*, 34(4):1153–1169, 2012.

G. Peersman and I. Van Robays. Cross-country differences in the effects of oil shocks. *Energy Economics*, 34(5):1532–1547, 2012.

R .Y. Rubinstein. Optimization of computer simulation models with rare events. *European Journal of Operational Research*, 99(1):89–112, 1997.

R. Y. Rubinstein and D. P. Kroese. *The Cross-Entropy Method: A Unified Approach to Combinatorial Optimization Monte-Carlo Simulation, and Machine Learning.* Springer-Verlag, New York, 2004.

P. Sadorsky. Stochastic volatility forecasting and risk management. *Applied Financial Economics*, 15(2):121–135, 2005.

P. Sadorsky. Modeling and forecasting petroleum futures volatility. *Energy Economics*, 28(4):467–488, 2006.

G. Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464, 1978.

S. J. Taylor. Modelling stochastic volatility: A review and comparative study. *Mathematical Finance*, 4(2):183–204, 1994.

A. B. Trolle and E. S. Schwartz. Unspanned stochastic volatility and the pricing of commodity derivatives. *Review of Financial Studies*, 22(11):4423–4461, 2009.

M. T. Vo. Regime-switching stochastic volatility: Evidence from the crude oil market. *Energy Economics*, 31(5):779–788, 2009.

Y. Wei, Y. Wang, and D. Huang. Forecasting crude oil market volatility: Further evidence using GARCH-class models. *Energy Economics*, 32(6):1477 – 1484, 2010.

J. Yu. Forecasting volatility in the New Zealand stock market. *Applied Financial Economics*, 12(3):193–202, 2002.