A Generalized Schwartz Model for Energy Spot Prices -
Estimation using a Particle MCMC Method

Anne Floor Brix\textsuperscript{a}  Asger Lunde\textsuperscript{a}  Wei Wei\textsuperscript{a}

\textsuperscript{a}Aarhus University & CREATES

November 5, 2014

\textbf{Abstract}

We consider a two-factor geometric spot price model with stochastic volatility and jumps. The first factor models the normal variations of the price process and the other factor accounts for the presence of spikes. Instead of using various filtering techniques for splitting the two factors, as often found in the literature, we estimate the model in one step using a MCMC method with a particle filter. In our empirical analysis we fit the model to UK natural gas spot prices and investigate the importance of allowing for jumps and stochastic volatility. We find that the inclusion of stochastic volatility in the process used for modeling the normal price variations is crucial and that it strongly impacts the jump intensity in the spike process. Furthermore, our estimation method enables us to consider both a continuous and purely jump-driven specification of the volatility process, and thereby assess if the volatility specification also influences the spike process and the overall model fit.

\textit{Keywords:} Energy spot price, Multi-factor model, Mean-reversion, MCMC, Ornstein-Uhlenbeck process, Particle filter, Stochastic volatility, Spikes.
1 Introduction

The liberalization of the European energy markets has over the last couple of decades led to highly
deregularized and liquid markets for trading energy commodities, such as gas and electricity. The
introduction of competition, and hence also price risk, has caused the markets to experience a
significant increase in price volatility and a market place for energy-based derivatives, used for
hedging, has emerged. The transition to a competitive market where prices are set according to
supply and demand means that energy spot prices have several distinct characteristics that should
be captured by any proposed model. The most important features are seasonality, mean-reversion,
spikes, multi-scale autocorrelation and stochastic volatility, see for instance Eydeland and Wolyniec
(2003) for empirical evidence on these stylized facts. Seasonality is caused by the seasonal pattern
on the demand side of the market, for instance by an increased need for heating during the winter.
Mean-reversion is a direct consequence of the markets being supply and demand driven, which
means that, unlike the stock market, prices are not allowed to evolve freely but will fluctuate around
a (possibly stochastic) level. This also has the important implication that the deseasonalized spot
prices will be modeled using stationary processes. Due to delivery constraints in the spot market,
sudden imbalances in supply and demand are almost immediately reflected in the spot price, causing
the price to jump because of an inelastic demand curve. These imbalances are typically caused by an
unexpected rise in demand or technical problems on the supply side. After experiencing a jump, the
price quickly mean-reverts to the normal level of production costs, leaving a spike in the price path.
The multi-scale autocorrelation structure that is observed in many markets is often a consequence
of the spike part of the price process having a stronger mean-reversion rate than the so-called base-
signal process that accounts for the normal variations. The inclusion of stochastic volatility in the
modeling framework helps to replicate the time-series properties of the prices, such as a leptokurtic
distribution, and to accurately estimate the jump part of the model. As we will shall see in our
application to the UK natural gas market, failing to include stochastic volatility will drive up the
expected number of jumps, contradicting the fact that jumps are supposed to be rare events.

In the univariate model, proposed in this paper, the logarithmic spot price is given as the sum
of three factors. The first factor is a deterministic mean-level function that models the, possibly
trending, seasonally varying mean-level of the logarithmic spot price. The second factor captures the
base-signal part of the price process and will be modeled using a Gaussian OU process with stochastic
volatility. The third factor is a non-Gaussian OU process that accounts for the spike behavior.
We will consider both a continuous and a purely jump-driven specification of the volatility process.
Instead of using various filtering techniques to split the base-signal and spike process in a first step
before estimating the model parameters, the paper contributes to the existing literature by proposing a method for estimating the model in one step using the particle MCMC (PMCMC) methods developed in Andrieu et al. (2010). In Green and Nossman (2008), a similar model is also estimated in one step using MCMC. In contrast to our approach, the authors in Green and Nossman (2008) condition on future values when computing the posterior distribution, rendering in-sample forecasts a useless tool for model evaluation. The authors also have to include a Brownian component in the specification of the spike factor in order to ensure that the factor has an absolutely continuous distribution when conditioning on the jumps, and thereby simplifying the MCMC estimation. Furthermore, our estimation approach, using the particle marginal Metropolis-Hastings sampler, has the great advantage of being able to accommodate different volatility specifications, including pure jump processes. This will enable us to investigate if the different volatility specifications have the same impact on the filtered spike process and if the volatility specification impacts the overall fit of the model. The method can also handle non-Markovian models, which is essential for effective sampling of the spike process in our proposed two-factor model. Finally, one of the outputs of the particle filter is the likelihood, which makes computation of Bayes factors and model comparison straightforward.

The stepping stone for many of the spot price models found in the literature is the mean-reverting one-factor Schwartz model from Schwartz (1997), where the spot price is defined as the exponential of a Gaussian Ornstein-Uhlenbeck (OU) process. This model was further extended to include a deterministic seasonality factor in Lucia and Schwartz (2002). In Benth et al. (2003), the geometric spot price model from Lucia and Schwartz (2002) are generalized to allow for jumps. A special case of this model, solely based on the NIG distribution is applied to oil and gas in Benth and Šaltytė Benth (2004). Another special case of the model from Benth et al. (2003) is the jump diffusion model, which have been used for modeling electricity spot prices in Cartea and Figueroa (2005) and Benth et al. (2012).

In Benth et al. (2007) an arithmetic multi-factor model based on non-Gaussian OU processes are suggested. The model is able to capture both the spike behavior and multi-scale autocorrelation structure of spot prices, and positivity of prices are ensured by letting the non-Gaussian OU processes be driven by subordinators. The possibility of negative prices in arithmetic models can also be viewed as an advantage since negative prices can actually occur in energy markets, such as the electricity market. Arithmetic models are also advantageous when it comes to pricing of forward contract with delivery being made over a period instead of at a single point in time. Due to the affine structure of the spot price in arithmetic models, forward prices become more analytically tractable than in
geometric models. We will instead consider geometric models, as these are a natural extension of the GBM used in the financial markets. Besides, derivative pricing is not the focus of our paper. It is also easier to model negative spikes, a feature often observed in gas spot prices, and ensure prices above a certain level (for instance zero), when using geometric models. The arithmetic multi-factor model from Benth et al. (2007) are estimated in Meyer-Brandis and Tankov (2008) and Benth et al. (2012) by splitting the spike process and base-signal process using a nonparametric method called hard-thresholding. In Klüppelberg et al. (2010), the filtering of the two mean-reverting processes are instead performed using a method based on extreme value theory. A two-factor extension of the geometric jump diffusion model from Cartea and Figueroa (2005), with a different jump size distribution, can be found in Hambley et al. (2009), but no estimation method is suggested in the paper.

The inclusion of stochastic volatility in the models used for modeling energy markets, was among others suggested by Geman (2005), where a Heston stochastic volatility extension of the Schwarz model is considered, but not estimated. In Green and Nossman (2008) a two-factor extension of the Schwartz model with Heston stochastic volatility is proposed and fitted to electricity spot prices using Markov Chain Monte Carlo (MCMC) techniques. A jump-driven specification of the volatility process is considered in Benth (2011), where the geometric one-factor model from Lucia and Schwartz (2002) is augmented with stochastic volatility given by the sum of non-Gaussian OU processes. However, in Benth (2011) a one-factor volatility process is utilized when the model is fitted to UK natural gas spot prices. The stochastic volatility model from Benth (2011) is extended in Benth and Vos (2013a) to incorporate spikes and a leverage effect, in a multidimensional setting, allowing for the joint modeling of several commodities. The model in Benth and Vos (2013a) only allows for positive jumps in the spot price, as the non-Gaussian OU factors entering the model are driven by subordinators. Estimation of the model from Benth and Vos (2013a) is however still an open question. The estimation method detailed and employed in this paper in a univariate setting, also has potential for usage in the multi-dimensional setup.

With the pure jump driven specification of the volatility process, our proposed model is a univariate version of the geometric model from Benth and Vos (2013a), with the extra flexibility of accommodating both negative and positive spikes. We will use the tempered stable OU process as our jump driven volatility specification. The model also closely resembles the model proposed in Green and Nossman (2008), where a CIR specification is used to describe the volatility dynamics. Instead, we will use a logarithmic OU process as our continuous specification of the volatility process.

The paper is organized as follows: In Section 2 our proposed model and the benchmark models are
presented. Section 3 describes the setup of our empirical application and the PMCMC estimation 
method is outlined in Section 4. In Section 5 the estimation results are presented and various 
methods for model comparisons are performed. Section 6 offers a discussion of possible extensions 
of the model and estimation procedure. Final remarks are given in Section 7.

2 Model Descriptions

In this section we will describe our proposed model. Let \( S(t) \) denote the spot price at time \( t \). The 
dynamics of the spot price will be described using the following geometric OU-based factor model, 
augmented with stochastic volatility:

\[
\begin{align*}
    d \log S(t) &= d \log \Lambda(t) + dX(t) + dY(t), \\
    dX(t) &= -\alpha_x X(t)dt + \sigma(t)dB(t), \\
    dY(t) &= -\alpha_y Y(t)dt + dI(t).
\end{align*}
\]

The first factor, \( \Lambda(t) \), is a deterministic function that accounts for the possible trend and seasonal 
patterns of the data. The specification of \( \Lambda(t) \) and the procedure used for detrending and deseasonalizing 
the spot prices will be described in Section 3. In this section, the focus will be on modeling 
the detrended and deseasonalized process: \( Z(t) = \log S(t) - \log \Lambda(t) \triangleq X(t) + Y(t) \), discretized 
using a time interval of length \( \Delta = 1 \) day to match the data in our empirical application. The 
process, \( X(t) \), is a Gaussian OU process with stochastic volatility, \( \sigma(t) \), that models the continuous 
variations in the logarithmic spot price and will be interpreted as the base-signal process. The last 
factor, \( Y(t) \), is non-Gaussian OU process with a pure jump Lévy process as background driving 
Lévy process (BDLP). \( Y(t) \) will be interpreted as the spike process. The different mean-reversion 
rates, \( \alpha_x > 0 \) and \( \alpha_y > 0 \), make multi-scale autocorrelation possible and help to reproduce the time 
series properties of the data, where a faster mean-reversion rate is observed for the spike process. 
The specification of \( X(t) \) and \( Y(t) \) are given below, and in the following subsections we present our 
benchmark models, that are all nested in the model described in Subsection 2.1.

2.1 The TF-SVJ Model

In our proposed two-factor model with stochastic volatility and jumps, labeled TF-SVJ, and the 
benchmark models we will use \( I(t) = N(t) \) as our BDLP, where \( N(t) \) is a compound Poisson Process 
with intensity parameter \( \lambda_J \) and normally distributed jump sizes. The detrended and deseasonalized
spot price, $Z(t)$, will then solve

$$
\begin{align*}
    dZ(t) &= dX(t) + dY(t) \\
    &= -\alpha_x X(t)dt - \alpha_y Y(t)dt + \sigma(t)dB(t) + dN(t).
\end{align*}
$$

If we assume that at most one jump occurs per day and approximate the variance of the increments in the AR(1) representation of $X(t)$, \( \int_t^{t+1} e^{-2\alpha_x(t+1-s)}\sigma^2(s)ds \), by

$$
\sigma^2(t) \int_t^{t+1} e^{-2\alpha_x(t+1-s)}ds = \sigma^2(t) \frac{1 - e^{-2\alpha_x}}{2\alpha_x},
$$

the discretized model becomes

$$
\begin{align*}
    Z_{t+1} &= X_{t+1} + Y_{t+1} \\
    X_{t+1} &= e^{-\alpha_x} X_t + \epsilon_{t+1} \\
    Y_{t+1} &= e^{-\alpha_y} Y_t + \xi_{t+1} J_{t+1}
\end{align*}
$$

where $\epsilon_{t+1} \sim N \left( 0, \sigma^2(t) \frac{1 - e^{-2\alpha_x}}{2\alpha_x} \right)$, $J_{t+1} \sim Bernoulli(\lambda_J)$ and $\xi_{t+1} \sim Normal(\mu_J, \sigma_J^2)$.

A similar model was suggested in Green and Nossman (2008), using a CIR specification of the stochastic volatility process and including an additional independent Brownian component in the spike process, $Y(t)$. We consider both a purely jump-driven specification of the volatility process and a continuous specification. Note that the base-signal, $X(t)$, will be continuous regardless of the specification of the volatility process and the process $Y(t)$ will therefore account for the spikes. The two specifications of the volatility process $\sigma^2(t)$ are given below.

### 2.1.1 Lévy-driven volatility with tempered stable marginals

The first specification of the volatility process under consideration is the case where $\sigma^2(t)$ is a tempered stable (TS) OU process. That is, $\sigma^2(t)$ solves

$$
    d\sigma^2(t) = -\lambda \sigma^2(t)dt + dL(\lambda t),
$$

and the marginal distribution of $\sigma^2(t)$ follows a tempered stable distribution, $\sigma^2(t) \sim TS(\kappa, \delta, \gamma)$. This process can be simulated recursively from

$$
    \sigma^2(t+1) = e^{-\lambda} \sigma^2(t) + e^{-\lambda} \int_0^1 e^{\lambda u}dL(\lambda u). 
$$
It is shown in Barndorff-Nielsen and Shephard (2001) that the BDLP of the TS-OU is the sum of a TS Lévy process and a compound Poisson process. We use Rosinski’s method to simulate the infinite activity part, and the innovations can be sampled using the following expression:

\[
e^{-\lambda \Delta} \int_0^{\Delta} e^{\lambda \mu} dL(\mu) = \sum_{i=1}^{\infty} e^{(-\lambda \Delta r_i)} \min \left\{ \left( \frac{a_i \kappa}{A \Delta} \right)^{-1/\kappa}, e_i v_i^{1/\kappa} \right\} + \sum_{i=1}^{N(\lambda \Delta)} e^{-\lambda \Delta r_i^*} c_i,
\]

where \(A = \delta^2 \kappa^2 / \Gamma(1 - \kappa)\) and \(B = \frac{1}{2} \gamma^{1/\kappa}\). The sequences of random variables, \(\{r_i\}, \{a_i\}, \{e_i\}, \{v_i\}, \{r_i^*\}\) and \(\{c_i\}\) are all mutually independent. \(\{r_i\}, \{v_i\}\) and \(\{r_i^*\}\) are i.i.d. standard uniforms, and \(\{e_i\}\) are i.i.d. exponential with mean \(1/B\), and \(\{c_i\}\) are i.i.d. Gamma with shape parameter \((1 - \kappa)\) and scale parameter \(1/B\). The \(a_1 < ... < a_i < ...\) are the arrival times of a Poisson process with intensity 1. Finally, \(N(\lambda \Delta)\) is a Poisson random variable with mean \(\lambda \Delta \delta \gamma \kappa\). Further, we have

\[
\sigma^2(0) = \sum_{i=1}^{\infty} \min \left\{ \left( \frac{a_i \kappa}{A_0} \right)^{-1/\kappa}, e_i v_i^{1/\kappa} \right\},
\]

where \(A_0 = \delta^2 \kappa / \Gamma(1 - \kappa)\).

The infinite sums are dominated by the first few terms, as shown in Barndorff-Nielsen and Shephard (2001). We truncate the sum to its first 100 terms as in Andrieu et al. (2010). As a special case, the volatility process becomes an Inverse Gaussian (IG) OU process when \(\kappa = 0.5\). In Benth (2011) a Gaussian OU process with stochastic volatility following a IG OU process is fitted to the logarithm of natural gas spot prices in the UK.

### 2.1.2 Logarithmic volatility model

The second volatility specification is a continuous specification, where we assume that the logarithmic volatility, \(h(t) = \log \sigma^2(t)\), follows a Gaussian OU process

\[
dh(t) = -\alpha_h(h(t) - \mu_h)dt + \sigma_h dB_h(t),
\]

where \(B_h(t)\) and \(B(t)\) are two independent Brownian motions.

### 2.2 The SF-J Model

In this subsection and the following ones, the benchmark models used for comparison are outlined. All the models are nested in the TF-SVJ model and all of them describe the dynamics of the deseasonalized logarithmic spot price \(Z(t)\).

In the first, and most simple, benchmark model we consider a single factor model with jumps and constant volatility. That is, we assume \(\alpha_x = \alpha_y = \alpha\) and \(\sigma(t) = \sigma\) and obtain the following
model

\[ dZ(t) = -\alpha Z(t)dt + \sigma dB(t) + dN(t). \]

If we assume that there is at most one jump in a day, we get the following discretized model

\[ Z_{t+1} = e^{-\alpha} Z_t + \epsilon_{t+1} + \xi_{t+1} J_{t+1} \]

where \( \epsilon_{t+1} \sim N \left( 0, \frac{\sigma^2 e^{-2\alpha}}{2\alpha} \right) \), \( J_{t+1} \sim Bernoulli(\lambda) \) and \( \xi_{t+1} \sim Normal(\mu, \sigma_J^2) \). This model resembles the model proposed in Cartea and Figueroa (2005), with the only difference being the jump size distribution. In Cartea and Figueroa (2005) the authors instead use a log-normal jump distribution.

### 2.3 The SF-SV Model

The next benchmark model we consider, is a single factor Gaussian OU process with stochastic volatility, corresponding to the assumptions \( \alpha_x = \alpha_y = \alpha \) and \( I(t) = 0 \). The detrended and deseasonalized logarithmic spot price now solves

\[ dZ(t) = -\alpha Z(t)dt + \sigma(t) dB(t), \]

and

\[
Z(t+1) = e^{-\alpha} Z(t) + \int_t^{t+1} \sigma(s) e^{-\alpha(t+1-s)} dB(s) \\
\sim N \left( e^{-\alpha} Z(t), \int_t^{t+1} e^{-2\alpha(t+1-s)} \sigma^2(s) ds \right).
\]

If we again approximate \( \int_t^{t+1} e^{-2\alpha(t+1-s)} \sigma^2(s) ds \) by \( \sigma^2(t) \int_t^{t+1} e^{-2\alpha(t+1-s)} ds \), the discretized model becomes the AR(1) model

\[ Z_{t+1} = e^{-\alpha} Z_t + \epsilon_{t+1}, \]

where \( \epsilon_{t+1} \sim N \left( 0, \sigma^2(t) \frac{1-e^{-2\alpha}}{2\alpha} \right) \). We consider the same two specifications of \( \sigma(t) \) as in the full model, the TF-SVJ model. With the TS-OU specification, the SF-SV model is a special case of the model considered in Benth (2011).

### 2.4 The SF-SVJ Model

We now consider adding jumps to the model specification from the previous subsection, resulting in a one-factor version of our proposed model, TF-SVJ. Using the same assumptions and approximations as in the SF-J and SF-SV models, we obtain the following discretized model
\[ Z_{t+1} = e^{-\alpha}Z_t + \epsilon_{t+1} + \xi_{t+1}J_{t+1} \]

where \( \epsilon_{t+1} \sim N \left( 0, \sigma^2(t) \frac{1-e^{-2\alpha}}{2\alpha} \right) \), \( J_{t+1} \sim Bernoulli(\lambda_J) \) and \( \xi_{t+1} \sim Normal(\mu_J, \sigma^2_J) \).

The specifications of the volatility process are again the tempered stable OU process and the logarithmic volatility model.

### 2.5 The TF-J Model

All the benchmark models considered so far have been single factor models, in the sense that the base-signal and spike part have had the same mean-reversion parameter \( \alpha \). The restriction, \( \alpha_x = \alpha_y \), also has the important implication that \( Z(t) \) will be a Markov process. We now consider extending our first benchmark model, the SF-J model, by relaxing the assumption \( \alpha_x = \alpha_y \), and instead allow each factor to have a separate mean-reversion rate. \( Z(t) \) will then solve

\[ dZ(t) = -\alpha_x X(t)dt - \alpha_y Y(t)dt + \sigma dB(t) + dN(t). \]

Once again we assume that there is at most one jump a day and we get the following discretized model

\[
\begin{align*}
Z_{t+1} &= X_{t+1} + Y_{t+1} \\
X_{t+1} &= e^{-\alpha_x}X_t + \epsilon^x_{t+1} \\
Y_{t+1} &= e^{-\alpha_y}Y_t + \xi_{t+1}J_{t+1}
\end{align*}
\]

where \( \epsilon_{t+1} \sim N \left( 0, \sigma^2 \frac{1-e^{-2\alpha_x}}{2\alpha} \right) \), \( J_{t+1} \sim Bernoulli(\lambda_J) \) and \( \xi_{t+1} \sim Normal(\mu_J, \sigma^2_J) \).

### 2.6 Model Overview

In Table 1 we give an overview of our proposed model and the benchmark models described in the previous subsections.

### 3 Data description and Initial Analysis

This section describes the data used in our empirical investigation and the detrending and deseasonalization of the data. We will fit our model and the benchmark models from the previous section to a time series of daily UK gas spot prices ranging from September 11, 2007 to February 10, 2014. The data is collected from Bloomberg\(^1\) and report day-ahead gas spot prices collected at the virtual

\(^1\)Code: NBPGDAHD index
Table 1: Model Overview.

<table>
<thead>
<tr>
<th>Model</th>
<th>Stoc. vol.</th>
<th>Const. vol.</th>
<th>Jumps</th>
<th>Jumps in vol.</th>
<th>Two-factor ($\alpha_x \neq \alpha_y$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TF-SVJ$^{TS}$</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>TF-SVJ$^{Log}$</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SF-J</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SF-SV$^{TS}$</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SF-SV$^{Log}$</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SF-SVJ$^{TS}$</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SF-SVJ$^{Log}$</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TF-J</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

hub NBP (Natural Balancing Point) for trading days (weekdays) in the sample period. This leaves us with a total of 1620 daily price quotes. There are no missing observations in the data set and the log spot price is depicted in Figure 3.

From Figure 3, the presence of both positive and negative spikes become evident. Positive spikes are usually caused by unpredicted weather changes, yielding an increase in demand for gas used in power production. In the UK market, supply uncertainty is also starting to play a role as the UK are becoming more and more dependent on gas import. The dependence on import from mainland Europe, through capacity constrained pipelines, can cause a slower reaction to an increase in demand, and in turn cause a spike in the price process. The negative spikes are often a consequence of poor anticipation of market-wide gas storage levels. Storage is costly and can not fully reconcile the variable seasonal demand for gas with the more constant rate of production. Low inventory levels can also increase price volatility and the risk of spike occurrence.

Like in most of the literature on commodity modeling, we start our investigation of the data characteristics by fitting a deterministic trend and seasonality function to the data. However, before this can be implemented we need to check for outlier in the data as these might influence the parameters in the fitted trend and seasonal components. A visual inspection of Figure 3 already suggested the presence of outliers, i.e. the large price spikes. From the histogram in Figure 3 it becomes clear that the daily changes in the logarithmic spot price are not normally distributed, but
instead follow a leptokurtic distribution. To detect the possible outliers in data that are not normally distributed, the same approach as in Chapter 5 of Benth et al. (2008) is employed. Let the daily change in the logarithmic spot price from day \( t-1 \) to day \( t \) be denoted by \( \Delta s_t = \log(S_t) - \log(S_{t-1}) \) for \( t = 2, \ldots, 1620 \). Now define the interquartile, \( IQR \), as the difference between the upper quartile \( Q_3 \) and lower quartile \( Q_1 \) of the time series \( \{\Delta s_t\} \). Assuming that the first observation in the data is not an outlier, \( \log(S_{t-1}) \) will be labeled as an outlier whenever \( \Delta s_t \) is larger than \( Q_3 + 3 \times IQR \) or smaller than \( Q_1 - 3 \times IQR \). This procedure resulted in 56 detected outliers. The detected outliers are then replaced by the average of the two closest non-outlier observations.

Assuming 250 trading days a year, the trend and seasonal patterns in the logarithmic spot prices are modeled by the following function

\[
\log \Lambda(t) = a_0 + a_1 t + a_2 \cos\left(2\pi(t - a_3)/250\right).
\]

The function represents the average level around which the gas prices fluctuate, and consists of a linear trend describing the inflation in the natural gas prices and a seasonal component modeling the seasonal variation over the year. The function is fitted to the logarithmic spot prices, with the replacement of the outliers, using the \texttt{nlminfit} function in MATLAB. The results are reported in
Figure 2: Histogram of daily changes in the logarithm of gas spot prices.

Table 2: Fitted parameter values and std. errors for $\log \Lambda(t)$.

<table>
<thead>
<tr>
<th></th>
<th>$a_0$</th>
<th>$a_1$</th>
<th>$a_2$</th>
<th>$a_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3.6494</td>
<td>0.0003</td>
<td>0.0761</td>
<td>100.08</td>
</tr>
<tr>
<td>SE</td>
<td>(0.0139)</td>
<td>(1.49e-05)</td>
<td>(0.0098)</td>
<td>(5.1467)</td>
</tr>
</tbody>
</table>

Table 2.

All the estimates in Table 2 are significant at the 5% level. We also fitted a function taking weekly, monthly and quarterly effects into account, but these effects were not significant at the 5% level and will be ignored going forward. The detrended and deseasonalized logarithmic spot price, $Z(t)$, can now be computed by inserting back the detected outliers and subtracting the fitted $\log \Lambda(t)$ function. The resulting time series is depicted in Figure 3 and will serve as the input for our estimation method outlined in the following section.
4 Estimation Method

In this section, the Bayesian techniques underlying our estimation method will be carefully described. Our model is able to account for important features of the spot price dynamics, such as stochastic volatility, jumps, and separate mean reversion rates for the base-signal and the spike process. The flexibility of the model also poses many challenges to the estimation. First, for models with stochastic volatility, evaluating the exact likelihood involves intractable high dimensional integration since volatility is latent. By treating the stochastic volatility as a state variable, these models have a non-linear state space representation, where the measurement equation describes how the logarithmic price changes given state variables, and the transition equation describes the evolution of the states. Jacquier et al. (1994) developed Bayesian MCMC methods for conducting exact inference in stochastic volatility models. Since then, Bayesian methods have been extensively applied to stock return models, including jump-diffusion models, see for example Eraker et al. (2003). Second, contrary to stock prices, energy prices tend to revert to a long run mean determined by the marginal cost of production. When jumps are present, they appear as spikes, meaning that prices revert to the mean level fast after a jump has occurred. Green and Nossman (2008) propose a MCMC algorithm to
handle energy models with these special features.

The third complication comes when we consider stochastic volatility that is driven by a pure jump process. In this case, the volatility process and the parameters governing its dynamics can be highly correlated in their posterior distributions, which results in extremely slowly mixing chains in the above mentioned MCMC algorithms. This problem is referred to as over-conditioning. Roberts et al. (2004) suggest a reparameterization to reduce the correlation. Griffin and Steel (2006) propose an algorithm with dependent thinning and reversible jump MCMC. However, these procedures cannot be easily generalized to the multi-factor models that are popular with commodity prices.

We adopt the particle MCMC methods introduced in Andrieu et al. (2010), in particular the particle marginal Metropolis-Hastings (PMMH) sampler. PMMH algorithms can be easily adapted to accommodate different volatility specifications, including both pure jump OU processes and the logarithmic Gaussian OU process. Furthermore, it can be applied to non-Markovian models, where the measurement density or the transition density may depend on the entire past of the latent process. This allows us to use a non-Markovian representation of the multi-factor model and is essential for effective sampling of the spike process. Last but not least, we use the likelihood obtained from the algorithm to compute Bayes factors and conduct model comparison.

As the name suggests, PMCMC has two components: a particle filter or sequential Monte Carlo (SMC) step and a MCMC step. Specifically, the PMMH sampler employs SMC to approximate the likelihood and latent variables conditional on the model parameters, and then apply MH algorithms to obtain the joint distribution of the parameters and the states, given the observations. We extend the standard PMMH algorithm in two aspects. First, for models with jumps or spikes, advanced SMC techniques need to be employed to alleviate a problem known as sample impoverishment. We propose to deal with this issue by marginalizing out some latent variables, a technique called Rao-Blackwellization; see Doucet et al. (2000). Our approach is closely related to auxiliary particle filters developed by Pitt and Shephard (1999) and illustrated in Johannes et al. (2009). Second, it is costly to evaluate the likelihood using SMC, and we therefore utilize adaptive algorithms to improve the efficiency of the Metropolis-Hasting sampler; see Andrieu and Thoms (2008) for a review on adaptive MCMC. The rest of this section focus on the estimation of the model in Section 2.1, as it is the most complex model, and nests all the benchmark models.

4.1 Sequential Monte Carlo

In the state space representation of our proposed model, the TF-SVJ model, the observed price process, $Z_t$, is the sum of two latent processes $X_t$ and $Y_t$, without any measurement error. We can
not apply particle filters directly in this case since there is no measurement density. One solution is to add a small Gaussian error term to the measurement equation. This is equivalent to assuming that $Y_t$ is a jump-diffusion instead of pure jump process, and it is comparable to the model in Green and Nossman (2008). However, this would still be problematic if particle filters with blind proposals, also called bootstrap filters, are implemented. If the variance of the measurement errors is small compared to the variance of the latent process, then the observations are informative about the latent process and bootstrap filters will perform poorly, see Pitt et al. (2012) for example.

We propose a different approach for solving this problem. Specifically, we use the following representation of Model 1

\[
Z_{t+1} = e^{-\alpha_x} Z_t + Y_{t+1} - e^{-\alpha_y} Y_t + \epsilon_{t+1},
\]

\[
Y_{t+1} = e^{-\alpha_y} Y_t + \xi_{t+1} J_{t+1}.
\]

Notice that this is no longer a Markovian state space model, in the sense that the measurement density depends on both $Y_t$ and $Y_{t+1}$, but we can use SMC methods to evaluate the likelihood and simulate the states given the parameters. Let $\theta$ and $K$ denote the parameters and the latent variables respectively, where $K_{t+1} = \{\sigma^2(t), Y_{t+1}\}$. SMC methods start with approximating the continuous filtering density $p_\theta(K_{1:t}|Z_{1:t})$ by a discrete distribution made of weighted random samples called particles. Given particles and associated weights, $\{K^{(i)}_{1:t}, \omega^{(i)}_t\}_{i=1}^N$, that approximate $p_\theta(K_{1:t}|Z_{1:t})$, SMC obtain samples from $p_\theta(K_{1:t+1}|Z_{1:t+1})$ and compute $p_\theta(Z_{t+1}|Z_{1:t})$ sequentially. Using Bayes Theorem,

\[
p_\theta(K_{1:t+1}|Z_{1:t+1}) = \frac{p_\theta(Z_{t+1}|K_{1:t+1}, Z_{1:t}, K_{1:t}) p_\theta(K_{1:t+1}|K_{1:t})}{p_\theta(Z_{t+1}|Z_{1:t})} p_\theta(K_{1:t}|Z_{1:t}),
\]

the density of interest $p_\theta(K_{1:t+1}|Z_{1:t+1})$ can be sampled using importance sampling techniques. The basic SMC choose the proposal density (importance density) $g_\theta(K_{1:t+1})$ to be $p_\theta(K_{1:t+1}|K_{1:t}) p_\theta(K_{1:t}|Z_{1:t})$, i.e., the new particles $K^{(i)}_{t+1}$ are propagated from $K^{(i)}_t$ using only transition densities and are “blind” to the observations. The importance weights are given by the ratio of the target density and the proposal density. From equation (1), it is therefore easily seen that the weights for the particles $K^{(i)}_{t+1}$ are proportional to $\hat{\omega}^{(i)}_{t+1} \tilde{\omega}^{(i)}_{t}$, where the incremental weights $\hat{\omega}^{(i)}_{t+1}$ are simply given by $p_\theta(Z_{t+1}|K^{(i)}_{t+1}, Z_{1:t}, K^{(i)}_{1:t})$. The likelihood, $p_\theta(Z_{t+1}|Z_{1:t})$, is the normalizing constant for the particles and is equal to $\sum_{i=1}^N \omega^{(i)}_{t+1} \hat{\omega}^{(i)}_{t}$. After normalizing the weights $\hat{\omega}^{(i)}_{t+1} = \hat{\omega}^{(i)}_{t} / p_\theta(Z_{t+1}|Z_{1:t})$, the particles $\{K^{(i)}_{1:t+1}, \omega^{(i)}_{1:t+1}\}_{i=1}^N$ now approximate $p_\theta(K_{1:t+1}|Z_{1:t+1})$.

If the variance of the weights is large, the particles yield a worse approximation to the continuous distribution $p_\theta(K_{1:t+1}|Z_{1:t+1})$, as the number of effective particles has decreased. In bootstrap filters, the incremental weights are simply the measurement density, and the algorithm performs better when
the states are persistent, and when the observations are less informative about the states than the transition density. This is not the case for the spike process $Y_{t+1}$. If there is a jump at time $t+1$, and $Y_{t+1}$ is propagated from a blind proposal, the measurement density, $p_\theta(Z_{t+1}|K_{t+1}, Z_{1:t}, K_{1:t})$, will peak at a few values, resulting in only a few particles having prominent weights. To alleviate this problem, one needs to adapt the proposal density of $Y_{t+1}$, or in other words, to incorporate $Z_{t+1}$ in the proposal density.

We employ the Rao-Blackwellization technique, as the innovations in the spike process can be integrated out conditional on other state variables. The vector $K_{t+1}$ has two components, the stochastic volatility, $\sigma^2(t)$, and spike process, $Y_{t+1}$. Since the innovation in $Y_{t+1}$ is assumed to be Bernoulli distributed jump times with normally distributed jump sizes, $p_\theta(Y_{t+1}|Z_{t+1}, \sigma^2(t), Z_{1:t}, K_{1:t})$ is analytically tractable. Hence, we can rewrite equation (1) as

$$p_\theta(K_{1:t+1}|Z_{1:t+1}) = p_\theta(Y_{t+1}|Z_{t+1}, \sigma^2(t), Z_{1:t}, K_{1:t})p_\theta(Z_{t+1}|\sigma^2(t), K_{1:t}, Z_{1:t})p_\theta(\sigma^2(t)|\sigma^2(t-1))p_\theta(K_{1:t}|Z_{1:t}),$$

and choose the following proposal density,

$$g_\theta(K_{1:t+1}|Z_{1:t+1}) = p_\theta(Y_{t+1}|Z_{t+1}, \sigma^2(t), Z_{1:t}, K_{1:t})p_\theta(\sigma^2(t)|\sigma^2(t-1))p_\theta(K_{1:t}|Z_{1:t}).$$

Here, the stochastic volatility, $\sigma^2(t)$, is still propagated from its transition density, but $Y_{t+1}$ is adapted to $Z_{t+1}$ as we can sample directly from $p_\theta(Y_{t+1}|Z_{t+1}, \sigma^2(t), Z_{1:t}, K_{1:t})$. In particular, we draw $J_{t+1}$ and $\xi_{t+1}$ from

$$p_\theta(J_{t+1}|Z_{t+1}, \sigma^2(t), Z_{1:t}, K_{1:t}) = \frac{p_\theta(Z_{t+1}|J_{t+1}, \sigma^2(t), Z_{1:t}, K_{1:t})p_\theta(J_{t+1})}{p_\theta(Z_{t+1}|\sigma^2(t), K_{1:t}, Z_{1:t})},$$

and

$$p_\theta(\xi_{t+1}|J_{t+1}, Z_{t+1}, \sigma^2(t), Z_{1:t}, K_{1:t}) = \frac{p_\theta(Z_{t+1}|\xi_{t+1}, J_{t+1}, \sigma^2(t), Z_{1:t}, K_{1:t})p_\theta(\xi_{t+1})}{p_\theta(Z_{t+1}|\sigma^2(t), Z_{1:t}, K_{1:t})p_\theta(J_{t+1})},$$

then let $Y_{t+1} = e^{-\alpha_\theta}Y_t + \xi_{t+1}J_{t+1}$. The incremental weights $\tilde{\omega}_{t+1} = p_\theta(Z_{t+1}|\sigma^2(t), K_{1:t}, Z_{1:t})$ do not depend on $Y_{t+1}$ as $\xi_{t+1}$ and $J_{t+1}$ are integrated out. As before, the likelihood $p_\theta(Z_{t+1}|Z_{1:t})$ equals $\sum_{i=1}^N \tilde{\omega}_{t+1}^{(i)}\tilde{\omega}_{t}^{(i)}$.

If importance sampling is carried out sequentially, weights will degenerate and only a few particles will have significant weights after a few iterations. The degeneracy grows exponentially in time and makes particle approximations unreliable. SMC uses a resampling step to deal with this problem. The particles $K_{1:t}^{(i)}$ are resampled with replacement according to their normalized weights $\tilde{\omega}_t^{(i)}$, using the multinomial distribution $\{\tilde{\omega}_t^{(i)}\}_{i=1}^N$ for instance. Particles with higher weights will be duplicated
and particles with lower weights will be eliminated. After resampling the particles, they all have equal weights.

The likelihood computed from SMC is random, and the variance of the likelihood, which is related to the variance of the weights, greatly impacts the acceptance rate in the MCMC step. The Rao-Blackwellization technique described above is the first step we adopt to reduce the variance. Second, the resampling step introduces additional Monte Carlo error, and we implement residual resampling as it has smaller variance than multinomial resampling, see Douc and Cappe (2005). Also, it satisfies the unbiasedness condition, meaning that the expected number of particles is proportional to the weights. Lastly, the variance of the likelihood decreases as the number of particles increases. However, for limited computation time, one faces a trade-off between the number of MCMC iterations to run and the number of particles to use in each iteration. See Pitt et al. (2012) for a guide on how to choose the optimal number of particles.

4.2 MCMC

SMC methods approximate the likelihood and state variables conditional on the parameters, but we are more interested in the joint distribution of the parameters and states. Notice that $p(\theta, K_{1:T}|Z_{1:T})$ can be decomposed into $p(\theta|Z_{1:T})p(K_{1:T}|\theta, Z_{1:T})$. The PMMH sampler suggest the following proposal density,

$$q(\theta, K_{1:T}|Z_{1:T}) = q(\theta|\theta^g)p(K_{1:T}|\theta, Z_{1:T}).$$

The draw, $\theta^{g+1}$, from $q(\theta|\theta^g)$, therefore has the simplified acceptance probability,

$$\alpha^{g+1} = \min \left\{ \frac{p(Z_{1:T}|\theta^{g+1})p(\theta^{g+1})q(\theta|\theta^{g+1})}{p(Z_{1:T}|\theta^g)p(\theta^g)q(\theta^{g+1}|\theta^g)}, 1 \right\},$$

where $p(Z_{1:T}|\theta)$ can be replaced by its particle approximation, as shown in Andrieu et al. (2010). If the marginal distributions of the states are of interest, we also sample $K_{1:T}^{g+1}$ from $p(K_{1:T}^{g+1}|\theta^{g+1}, Z_{1:T})$ in the SMC step and accept it jointly with $\theta^{g+1}$.

The choice of proposal density $q(\theta|\theta^g)$ is another crucial element in determining the efficiency of the MCMC algorithm. We use a random-walk proposal, $\theta^{g+1} \sim \text{TN}(\theta^g, \beta \Sigma^g)$, where $\text{TN}$ denotes truncated normal, as some of the parameters have finite support. In particular:

$$q(\theta^g|\theta^{g+1}) = \frac{f^n(\theta^g; \theta^{g+1}, \beta \Sigma^g)}{F^N(\theta_u; \theta^{g+1}, \beta \Sigma^g) - F^N(\theta_l; \theta^{g+1}, \beta \Sigma^g)},$$

$$q(\theta^{g+1}|\theta^g) = \frac{f^n(\theta^{g+1}; \theta^g, \beta \Sigma^g)}{F^N(\theta_u; \theta^g, \beta \Sigma^g) - F^N(\theta_l; \theta^g, \beta \Sigma^g)}.$$
where $f^N$ and $F^N$ denote pdf and cdf of the multivariate normal distribution respectively, $\theta_u$ is the upper limit of parameters, and $\theta_l$ is the lower limit of parameters. The ratio of proposal densities in equation (2) simplifies to the ratio of normalizing constants as $f^n$ is symmetric:

$$\frac{q(\theta^g|\theta^{g+1})}{q(\theta^{g+1}|\theta^g)} = \frac{F^N(\theta_u; \theta^g, \beta^g \Sigma^g) - F^N(\theta_l; \theta^g, \beta^g \Sigma^g)}{F^N(\theta_u; \theta^{g+1}, \beta^g \Sigma^g) - F^N(\theta_l; \theta^{g+1}, \beta^g \Sigma^g)}.$$ (4)

Notice that if all parameters have support on the whole real line, the proposal density is symmetric, and the ratio of proposal densities becomes 1.

Gelman et al. (1996) show that the efficiency of the random-walk MH algorithm is maximized when $\Sigma^g$ is the covariance matrix of the target posterior distribution, and the scaling factor $\beta^g$ is approximately $2.38^2/d$, where $d$ is the number of parameters.

In practice, we do not know $\Sigma^g$ a priori. Adaptive MCMC allows us to learn $\Sigma^g$ “on the fly”, using previous updates in the chain to construct this covariance. The resulting chain $\{\theta^{g+1}\}_{g=1}^G$ is not Markovian as the proposal density depends on the history of $\theta$, and ergodicity of the chain can be perturbed. Haario et al. (2001) propose an adaptive Metropolis (AM) algorithm, using the whole history of the chain, or any increasing part of the past, which leads to vanishing adaptation and preserves the correct ergodic property. We adopt the AM algorithm with global adaptive scaling as in Andrieu and Thoms (2008). When the chain is starting, $\Sigma^g$ might be a poor initial guess, resulting in too many or too few rejections. Andrieu and Thoms (2008) suggest adapting the scaling factor $\beta^g$ using the the acceptance probability in (2). If the acceptance probability is higher than the optimal acceptance probability, $\beta^g$ increases, and vice versa. The optimal acceptance rate is chosen to be around 24% as suggested in Gelman et al. (1996).

Bayesian inference requires specifying prior distributions for the parameters. For most of the model parameters, we choose diffuse but proper priors. For the jump process, we use a prior that elicit our belief that jumps are large compare to the base-signal. Specifically, we use a gamma distribution for the standard deviation of jump sizes which places lower probability on small jumps.

### 4.3 PMMH Algorithm

We outline the algorithm for the TF-SVJ model in this subsection:

1. For $g = 1, ..., G$, where $G$ is the number of MCMC iterations, sample $\theta^{g+1} \sim TN(\theta^g, \beta^g \Sigma^g)$, then run the following SMC algorithm to obtain $\hat{p}(Z_{1:T} | \theta^{g+1})$ and $K^{g+1}_{1:T}$:

   (a) sample $\sigma^2(0, i)$ and $Y_1^{(i)}$ from their stationary distribution.
i. compute
\[ \omega_1^{(i)} = \frac{p_0(\mathcal{Z}_1|\mathcal{Y}_1^{(i)}, \sigma^2(0,i))}{N} \]
for \( i = 1, \ldots, N \), where \( N \) is the number of particles.

ii. obtain likelihood from \( p(\mathcal{Z}_1) = \sum_{i=1}^{N} \omega_1^{(i)} \), and compute normalized weights: \( \hat{\omega}_1^{(i)} = \frac{\omega_1^{(i)}}{\hat{p}(\mathcal{Z}_1)} \).

(b) at \( t = 1, \ldots, T - 1 \)

i. sample the index \( a_1^{(i)} \) for \( i = 1, \ldots, N \), using \( \tilde{\omega}_t \) and set \( \tilde{\omega}_t = \frac{1}{N} \).

ii. sample \( \sigma^2(t, i) \sim p_\theta(\sigma^2(t)|\sigma^2(t-1, a_1^{(i)})) \), and \( Y_{t+1}^{(i)} \sim p_\theta(Y_{t+1}|\mathcal{Z}_{t+1}, \sigma^2(t, i), Y_t^{a_1^{(i)}}) \).

iii. compute the incremental weights: \( \hat{\omega}_{t+1}^{(i)} = p_\theta(\mathcal{Z}_{t+1}|\sigma^2(t, i), Y_t^{a_1^{(i)}}, \mathcal{Z}_{1:t}) \).

iv. obtain the likelihood: \( \hat{\omega}_t = \sum_{i=1}^{N} \hat{\omega}_{t+1}^{(i)} \).

v. normalize the weights \( \hat{\beta}_{t+1}^{(i)} = \frac{\hat{\omega}_t^{(i)} \hat{\omega}_{t+1}^{(i)}}{\sum_{i=1}^{N} \hat{\omega}_t^{(i)} \hat{\omega}_{t+1}^{(i)}} \).

(c) at \( t = T \)

i. obtain \( \hat{p}(\mathcal{Z}_{1:T}|\theta^{g+1}) = \hat{p}(\mathcal{Z}_1) \prod_{t=1}^{T-1} \hat{p}_\theta(Z_{t+1}|\mathcal{Z}_{1:t}) \).

ii. use \( \hat{\omega}_T \) and \( a_{1:T} \) to draw a realization of states \( K_{1:T}^{g+1} \).

2. accept \( \theta^{g+1} \) and \( K_{1:T}^{g+1} \) with probability:
\[ \alpha^{g+1} = \min \left\{ \frac{\hat{p}(\mathcal{Z}_{1:T}|\theta^{g+1})p(\theta^{g+1})q(\theta^g|\theta^{g+1})}{\hat{p}(\mathcal{Z}_{1:T}|\theta^g)p(\theta^g)q(\theta^{g+1}|\theta^g)}, 1 \right\} \]
where \( \hat{p}(\mathcal{Z}_{1:T}|\theta^{g+1}) \) is computed from the SMC algorithm above, \( p(\theta) \) is the prior density of the model parameters which we specify in Section 5, and \( q \) is the truncated normal proposal density. The ratio of proposal densities is given in Equation (4). If rejected, we set \( \theta^{g+1} \) and \( K_{1:T}^{g+1} \) equal to \( \theta^g \) and \( K_{1:T}^{g} \).

3. update the scaling factor and the covariance matrix for the proposal density:
\[ \nu^{g+1} = 1/(g + 1)^{0.5} \]
\[ \log \beta^{g+1} = \log \beta^g + \nu^{g+1} (\alpha^{g+1} - \alpha^*) \]
\[ \mu^{g+1} = \mu^g + \nu^{g+1} (\theta^{g+1} - \theta^g) \]
\[ \Sigma^{g+1} = \Sigma^g + \nu^{g+1} \left( (\theta^{g+1} - \mu^g)(\theta^{g+1} - \mu^g)^T - \Sigma^g \right) \].
4.4 Model Comparison

The estimation procedure is easily adapted to all the models we considered in Section 2. The question remains, which of the models fits the data better? Specifically, is stochastic volatility important? Which volatility process is more suitable for the UK Gas price data? What is the role of jumps? Is it necessary to have different mean reversion rate for the spike process and the base-signal? To address these important questions, we estimate a large set of models and conduct an extensive model comparison. For nested models, we carry out model specification tests using likelihood ratio statistics. We also compute Bayes factors as models with tempered stable volatility and logarithmic volatility are not nested.

Given two competing models, say the TF-SVJ\textsuperscript{TS} and TF-SVJ\textsuperscript{Log} models, the Bayes factor is then the ratio of the probability of each model given data, i.e.,

\[ BF = \frac{p(\text{TF-SVJ}^{\text{TS}}|Z)}{p(\text{TF-SVJ}^{\text{Log}}|Z)} . \]

If we assume that the competing models are equally probable a priori, the Bayes factor can be expressed as the posterior odds ratio:

\[ BF = \frac{p(Z|\text{TF-SVJ}^{\text{TS}})}{p(Z|\text{TF-SVJ}^{\text{Log}})} . \]

The density \( p(Z|M) \) is termed marginal likelihood, as it is the likelihood of data under model \( M \) obtained by marginalizing over the parameters in model \( M \):

\[ p(Z|M) = \int p(Z|\theta, M)p(\theta|M)d\theta, \tag{5} \]

where \( p(\theta|M) \) is the prior density of parameters in model \( M \).

We use the output from the PMMH algorithm to compute the Bayes factors. The algorithm produces \( \{p(Z|\theta^g, M)\}_{g=1}^G \), where \( \theta^g \) are draws from the posterior density \( p(\theta|Z, M) \). In equation (5), the integration is over the prior density of \( \theta \). Newton and Raftery (1994) propose several estimators to compute marginal likelihood based on importance sampling and Monte Carlo integration. We adopt the version which uses a mixture of the prior and posterior as the importance density, yet does not require further simulation from the prior. Given \( G \) samples of \( \theta \) from the posterior, imagine that additional \( \delta_p G/(1 - \delta_p) \) samples of \( \theta \) are drawn from the prior, resulting in a total of \( G/(1 - \delta_p) \) samples from the mixture density \( \delta_p p(\theta|M) + (1 - \delta_p) p(\theta|Z, M) \). Assume that the draws from the prior all have likelihood \( p(Z|\theta, M) \) equal to their expected value \( p(Z|M) \), we then obtain the following estimator,

\[ \hat{p}(Z|M) = \frac{\delta_p G/(1 - \delta_p) + \sum_{g=1}^G p(Z|\theta^g, M)}{\delta_p G/((1 - \delta_p)\hat{p}(Z|M)) + \sum_{g=1}^G (\delta_p \hat{p}(Z|M) + (1 - \delta_p) p(Z|\theta^g, M))^{-1}}. \tag{6} \]

We compute twice the logarithm of the Bayes factor as it has the same scale as likelihood ratio test statistics. Kass and Raftery (1995) suggest the following interpretation: if \( 2 \log BF \) is between
2 to 6, it is viewed as positive evidence; between 6 to 10, it indicates strong evidence; and a value greater than 10 is interpreted as very strong evidence.

5 Estimation Results

We apply the PMMH algorithm from subsection 4.3 to the deseasonalized and detrended logarithmic gas spot prices. We start with a preliminary run which uses adaptive MCMC, then “freeze” the covariance matrix $\Sigma$ and scaling factor $\beta$ and run a further 20000 iterations to get the posterior distributions of $\theta$ and $K$. The number of particles for the logarithmic volatility specification is set to 4800, while the number of particles for the model with tempered stable volatility is set to 1600 for computational considerations. To conduct likelihood ratio test, we need a point estimate of $\theta$, and we choose both the mean and median of the posterior distribution of $\theta$. To minimize randomness from SMC, we use 300,000 particles to evaluate the likelihood at the point estimates. Marginal likelihood $p(Z|M)$ is computed from equation (6).

The estimation for the benchmark models from Section 2 follows similar procedures. For models with jumps in the logarithmic price, we utilize Rao-Blackwellization and integrate out the jump times and jump sizes analytically for the likelihood. For models with two factors, we use the non-Markovian representation. For the SF-J model, the likelihood can be obtained analytically, and we use the MH step for updating parameters.

The same prior is specified for models with overlapping parameters, and we assume that the parameters are independent a priori and the joint prior is simply the multiplication of prior distributions. In summary, the priors for the mean reversion parameters are $\alpha_x \sim G(1, 1)$ and $\alpha_y \sim G(1, 1)$, where $G$ denotes the Gamma distribution. In the TF-J model, we also impose $\alpha_y > \alpha_x$ to ensure that the mean reversion rate for the spike process is larger than the mean reversion rate for base-signal. For stochastic volatility with tempered stable marginals, we specify $\lambda \sim G(1, 2)$, $\kappa \sim Beta(10, 10)$, $\delta \sim G(1, \sqrt{50})$, and $\gamma \sim G(1, \sqrt{50})$. For logarithmic volatility, $\alpha_h \sim G(1, 1)$, $\mu_h \sim N(-5, 5)$ and $\sigma_{h}^2 \sim G(1, 2)$. For models with constant volatility, we choose $\sigma^2 \sim G(1, 2)$. Finally, the priors for jump parameters are $\mu_J \sim N(0, 2)$, $\sigma_J \sim G(1.5, 0.5)$, and $\lambda_J \sim G(1, 10)$.

\footnote{For the TF-J model we use auxiliary particle filters. This is the case with “perfect adaptation”, and Rao-Blackwellized particle filters and auxiliary particle filters only differs in the order of the sampling and resampling steps.}
5.1 Parameter Estimates

The parameter estimates obtained from fitting the models to the detrended and deseasonalized logarithmic gas spot prices are reported in Table 3. We also report the log-likelihood evaluated at the mean and median of the posterior distribution of the parameters, and the marginal log-likelihood computed using formula (6).

The parameter estimates for the simplest benchmark model, SF-J, indicate that the model does not adequately capture the dynamics of the data. Like in Green and Nossman (2008), failing to include stochastic volatility severely drives up the estimated jump intensity. In our case, we find $\lambda_J = 0.2445$, which does not fare well with our data and the general observation that jumps are rare events. The high jump intensity means that most of the variability in the data is explained by jumps, and as a consequence the estimate of the constant volatility, $\sigma$, is very low. From the estimated spike innovations, plotted in the top panel of Figure 4, we see that there is clustering in the jump times and the assumption of a constant jump intensity does not hold. The estimate of the mean-reversion rate, $\alpha$, are in line with the estimate $\hat{\alpha} = 0.0064$, obtained from fitting the theoretical autocorrelation function, $\exp^{-\alpha |t|}$, to the first 100 lags of the empirical autocorrelation function for $Z(t)$. The mean-reversion rate corresponds to a half-life of approximately 90 days, revealing that most emphasis is put on capturing the mean-reversion of the base-signal and that the logarithmic spot price is very persistent.

In the next two benchmark models, SF-J$^{TS}$ and SF-J$^{Log}$, we consider allowing for stochastic volatility, but neglect to account for the presence of jumps. In Green and Nossman (2008), the authors find that in the constant volatility case, neglecting to account for jumps will drive up the estimate of the mean-reversion rate. This is not the case when there is stochastic volatility in the model, at least not for the data at hand. The estimate of the mean-reversion rate has not changed much, and it appears that most of the variations in the data can be explained by stochastic volatility. This can also be seen from the filtered variance processes in the top panel of Figure 6, where the volatile period in the beginning of the data sample, that was explained by jumps in the SF-J model (see the top panel of Figure 4), is now captured by the stochastic volatility process. The log-likelihood evaluated at the mean and median of the posterior distribution of the parameters has increased a lot, indicating that including stochastic volatility in the model is more important than allowing for jumps. From the plots of the filtered stochastic variance processes, in Figure 6, there does not seem to be much difference across the two specifications. When $\kappa$ equals 0.5 in the TS-OU specification we get a IG-OU process, which was the specification used in the SF-SV model that was fitted to UK natural gas spot prices in Benth (2011). Our estimate of $\kappa$ is 0.4504, and testing
\( \kappa = 0.5 \) yields a p-value of 0.196.

In the benchmark models labeled SF-SVJ, we still consider a one-factor model but this time we include both stochastic volatility and jumps in the model specification. From the results in Table 3, we see that the different volatility specifications now results in different estimates of the mean-reversion rate. The half-life of the observed process is now 125 days in the SF-SVJ\(^{TS} \) model and 88 days for the SF-SVJ\(^{Log} \) model. The jump intensity and jump size distribution also changes with the volatility specification. Perhaps a bit surprisingly, the estimated jump intensity is higher in the SF-SVJ\(^{TS} \) model where the stochastic volatility process is allowed to jump. If we compare the middle and bottom panel in Figure 4, it is clear that the two models especially differ in the modeling of the large innovation in the end of the data set. The volatile period in the beginning the data set is captured (mainly) by stochastic volatility in both models and as a consequence the jump intensity has drastically dropped compared to the estimate from the SF-J model. From the estimated spike innovations in the SF-SVJ\(^{TS} \) model, we see that there is still some clustering in the jump times and allowing for a time-varying jump intensity would be a natural extension of the model. This is however left for future research, but will be further discussed in Section 6. The estimated jump intensity in the SF-SVJ\(^{Log} \) model is extremely low, and from the plots in Figure 4 we see that only one large innovation is categorized as a jump. The low jump intensities in both models also explain the high standard errors on the parameters, \( \mu_J \) and \( \sigma_J \), governing the jump size distribution. The inclusion of jumps does however not seem to impact the parameters governing the volatility process much. However, testing for \( \kappa = 0.5 \) in the TS-OU specification now gives a p-value of 0.0590, making it less plausible that the IG-OU specification could have been used instead.

In our last benchmark model, TF-J, we consider extending the SF-J model to a two-factor model, where the mean-reversion rates are different for the base-signal and spike process. We impose the restriction \( \alpha_y > \alpha_x \), to insure that the estimated mean reversion rate for the spike process is higher than for the base-signal process. We see that the estimate of \( \alpha \) obtained in the SF-J model is a weighted average of \( \alpha_x \) and \( \alpha_y \), with most weight put on the mean-reversion rate for the base-signal. The half-life of the base-signal is 165 days and the half-life of the spike process is 17.5 days. The half-life of the spike process is too high compared to what is normally to be understood by a spike. The estimates for the jump process are unaffected by the inclusion of separate mean-reversion rates for the two factors, and hence the estimated jump intensity is still unrealistically high.

In our proposed model, TF-SVJ, which is the two-factor extension of SF-SVJ, the estimates of the mean-reversion rates are higher than the estimate found for the SF-SVJ model. The half-life of the base-signal equals 80 days for the TF-SVJ\(^{TS} \) model and 90 days for the TF-SVJ\(^{Log} \) model. The
Table 3: Parameter estimates.

<table>
<thead>
<tr>
<th></th>
<th>SF-J</th>
<th>SF-SV^TS</th>
<th>SF-SV^Log</th>
<th>SF-SVJ^TS</th>
<th>SF-SVJ^Log</th>
<th>TF-J</th>
<th>TF-SVJ^TS</th>
<th>TF-SVJ^Log</th>
</tr>
</thead>
<tbody>
<tr>
<td>α_x</td>
<td>0.00773</td>
<td>0.00679</td>
<td>0.00696</td>
<td>0.00553</td>
<td>0.00784</td>
<td>0.00421</td>
<td>0.00867</td>
<td>0.00766</td>
</tr>
<tr>
<td></td>
<td>(0.00293)</td>
<td>(0.00304)</td>
<td>(0.00207)</td>
<td>(0.00268)</td>
<td>(0.00291)</td>
<td>(0.00203)</td>
<td>(0.00226)</td>
<td>(0.00161)</td>
</tr>
<tr>
<td>α_y</td>
<td>0.00773</td>
<td>0.00679</td>
<td>0.00696</td>
<td>0.00553</td>
<td>0.00784</td>
<td>0.03956</td>
<td>2.1008</td>
<td>2.1297</td>
</tr>
<tr>
<td></td>
<td>(0.00293)</td>
<td>(0.00304)</td>
<td>(0.00207)</td>
<td>(0.00268)</td>
<td>(0.00291)</td>
<td>(0.01709)</td>
<td>(0.378)</td>
<td>(0.478)</td>
</tr>
<tr>
<td>σ^2</td>
<td>0.00047</td>
<td></td>
<td></td>
<td>0.00046</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.00004)</td>
<td></td>
<td></td>
<td>(0.00004)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>λ</td>
<td>0.2108</td>
<td></td>
<td>0.1841</td>
<td></td>
<td>0.2011</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.02637)</td>
<td></td>
<td>(0.02701)</td>
<td></td>
<td>(0.02623)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>κ</td>
<td>0.4504</td>
<td></td>
<td>0.4429</td>
<td></td>
<td>0.4108</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.03834)</td>
<td></td>
<td>(0.03024)</td>
<td></td>
<td>(0.04401)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>δ</td>
<td>0.04262</td>
<td></td>
<td>0.04430</td>
<td></td>
<td>0.06809</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.01879)</td>
<td></td>
<td>(0.01541)</td>
<td></td>
<td>(0.03196)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>γ</td>
<td>8.3085</td>
<td></td>
<td>8.4637</td>
<td></td>
<td>8.4585</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.420)</td>
<td></td>
<td>(1.389)</td>
<td></td>
<td>(1.522)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>α_h</td>
<td>0.06049</td>
<td></td>
<td>0.06369</td>
<td></td>
<td>0.06336</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.01346)</td>
<td></td>
<td>(0.01343)</td>
<td></td>
<td>(0.01353)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>µ_h</td>
<td>-7.0472</td>
<td></td>
<td>-7.1098</td>
<td></td>
<td>-7.1014</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.138)</td>
<td></td>
<td>(0.216)</td>
<td></td>
<td>(0.148)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>σ^2_h</td>
<td>0.2737</td>
<td></td>
<td>0.2742</td>
<td></td>
<td>0.2651</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.05783)</td>
<td></td>
<td>(0.05434)</td>
<td></td>
<td>(0.05281)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>µ_J</td>
<td>-0.00189</td>
<td></td>
<td>-0.07008</td>
<td></td>
<td>-0.09599</td>
<td>-0.3270</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.00520)</td>
<td></td>
<td>(0.125)</td>
<td></td>
<td>(0.0526)</td>
<td>(0.328)</td>
<td></td>
<td>(0.358)</td>
</tr>
<tr>
<td>σ_J</td>
<td>0.09342</td>
<td></td>
<td>0.1605</td>
<td></td>
<td>0.7232</td>
<td>0.4330</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.00480)</td>
<td></td>
<td>(0.104)</td>
<td></td>
<td>(0.0274)</td>
<td>(0.230)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>λ_J</td>
<td>0.2445</td>
<td></td>
<td>0.01092</td>
<td></td>
<td>0.00221</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.02340)</td>
<td></td>
<td>(0.00097)</td>
<td></td>
<td>(0.00128)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

LogLikMean 2941.69 3174.17 3194.21 3176.57 3194.26 3194.32 3178.72 3199.63
LogLikMedian 2941.71 3174.82 3194.69 3177.59 3194.49 2944.36 3180.44 3199.95
MarginalLL 2937.97 3172.10 3193.00 3173.35 3191.60 2939.96 3176.43 3196.45

24
estimated half-life for the spike process now equals a third of a day with both volatility specifications. We also fitted the theoretical autocorrelation function, \( w_1 \exp(-\alpha_x |t|) + (1 - w_1) \exp(-\alpha_y |t|) \), to the first 100 lags of the empirical autocorrelation function, and obtained the estimates \( \hat{\alpha}_x = 0.0050 \) and \( \hat{\alpha}_y = 0.4141 \), corresponding to half-lifes of 138 days and 1.7 days. The high estimate of \( \alpha_y \) in our TF-SVJ models, could just be a consequence of the low number of jumps in the spike process. The jump intensity has gone down for the TS-OU specification and up for the log-OU specification, such that it is now roughly equal for the two volatility specifications. Even though the estimates of the mean and variance of the jump size distribution are not the same for the two models, Figure 5 reveals that the filtered jump processes are almost identical. From the filtered variance processes in Figure 6, we see that the volatility processes look very similar to the ones from the SF-SVJ and SF-SV models, except for the high peak in the end of 2009, which is now less pronounced. The parameters governing the TS-OU volatility process has slightly changed compared to the benchmark models, but the log-OU specifications remains unaffected by the inclusion of separate mean-reversion rates. Testing the restriction \( \kappa = 0.5 \) gives a p-value of 0.0386, rejecting that a IG-OU specification could have captured the volatility dynamics equally well.

In order to check if our proposed model is internally consistent we compare the theoretical mean of the variance process, \( \sigma(t) \), to the mean of the filtered variance processes from Figure 6. The theoretical mean of the TS-OU volatility process equals 0.00262, whereas the empirical mean is found to be 0.00224. With the log-OU specification, the theoretical mean equals 0.00230 and the empirical mean is found to be 0.00221. Hence, both models appear internally consistent. From the reported log-likelihood values, we see that the inclusion of separate mean-reversion rates in the model matters more in the TF-SVJ Log model.

The significance of the different model characteristics and the internal ranking of the models will be investigated thoroughly in the next subsection.
Figure 4: The figure depicts the estimated spike innovations, computed from the mean of the posterior distribution $p(\xi_{1:T}, J_{1:T} | Z_{1:T})$. 
Figure 5: The figure depicts the estimated spike processes, computed from the mean of the posterior distribution $p(Y_{1:T} | Z_{1:T})$. 
Figure 6: The figure depicts the estimated stochastic volatility processes, computed from the mean of the posterior distribution $p(\sigma^2(1:T-1)|Z_{1:T})$. 
Figure 7: The figure displays the theoretical and empirical autocorrelation functions for the two stochastic volatility processes in the TF-SVJ model. The theoretical acf’s are computed using the posterior mean of the parameters governing the volatility processes. The empirical acf’s are computed in each MCMC iteration and then averaged.
5.2 Model Evaluation

We conduct a comprehensive comparison between the models using Bayes factors computed from the marginal log likelihood in Table 3. We report in Table 4 twice the logarithm of the Bayes factor as it has the same scale as likelihood ratio test statistics. Let $\text{LBF}(M_1, M_0) = -2(\log p(M_1|Z) - \log p(M_0|Z))$, Kass and Raftery (1995) suggest the following scale for interpretation: if $\text{LBF}(M_1, M_0)$ is between 2 to 6, it is viewed as positive evidence against model $M_0$; between 6 to 10, it indicates strong evidence; and a value greater than 10 is interpreted as very strong evidence. Negative values are interpreted on the same scale while it suggests evidence in favor of $M_0$.

When volatility is assumed to be constant, allowing for different mean-reversion rates in the two factors improves the overall performance of the model, as seen by $\text{LBF}(<\text{TF-J, SF-J}>)$ = 3.989. However, allowing for stochastic volatility has a much larger impact on the model fit, as $\text{LBF}(<\text{SF-SV, SF-J}>)$ are over 400 for both volatility specifications.

Next, we look at the effect of having price jumps in the model when stochastic volatility is being accounted for. With the TS-OU volatility specification, including jumps in the price renders $\text{LBF}(<\text{SF-SVJ_{TS}, SF-SVTS}>)$ = 2.513, indicating positive evidence against SF-SVTS. The two factor version further improves the fit as $\text{LBF}(<\text{TF-SVJ_{TS}, SF-SVJ_{TS}}>)$ is equal to 6.159. For the log-OU volatility specification, the Bayes factor favors SF-SVLog over SF-SVJLog. Although the log-likelihood of SF-SVJLog evaluated at the posterior mean of parameters is higher than that of SF-SVLog, the Bayes factor penalises models with more parameters and in this case selects the simpler model SF-SVLog. However, when we consider the full model, TF-SVJLog, there is strong evidence against both SF-SVJLog and SF-SVLog, with $\text{LBF}(<\text{TF-SVJLog}, SF-SVJLog>) = 6.91$ and $\text{LBF}(<\text{TF-SVJLog}, SF-SVJLog>) = 9.69$. In summary, simply including jumps in the model description is only slightly favorable (TS-OU volatility) or unfavorable (log-OU volatility), while specifying a spike process, by imposing $\alpha_x \neq \alpha_y$, does improve the model fit. This finding is in line with our intuition that jumps are faster decaying and that this feature is important for model building.

The Bayes factors also provide a direct comparison between the two types of volatility specifications. Here $\text{LBF}(M_{\text{Log}}, M_{\text{TS}})$ is ranging from 36.5 to 41.8, with $M$ indicating SF-SV, SF-SVJ or TF-SVJ. For the UK natural gas data, models with the log-OU volatility specification are favored over models with TS-OU volatility across the different jump specifications, with the TF-SVJLog model providing the best fit to the data.
Table 4: Bayes factors.

<table>
<thead>
<tr>
<th></th>
<th>SF-J</th>
<th>SF-SV^{TS}</th>
<th>SF-SV^{Log}</th>
<th>SF-SVJ^{TS}</th>
<th>SF-SVJ^{Log}</th>
<th>TF-J</th>
<th>TF-SVJ^{TS}</th>
</tr>
</thead>
<tbody>
<tr>
<td>SF-SV^{TS}</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SF-SV^{Log}</td>
<td>468.26</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SF-SVJ^{TS}</td>
<td>510.06</td>
<td>41.80</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SF-SVJ^{Log}</td>
<td>470.77</td>
<td>2.513</td>
<td>-39.29</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TF-J</td>
<td>507.28</td>
<td>39.02</td>
<td>-2.788</td>
<td>36.50</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TF-SVJ^{TS}</td>
<td>3.989</td>
<td>-464.27</td>
<td>-506.07</td>
<td>-466.78</td>
<td>-503.29</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TF-SVJ^{Log}</td>
<td>476.93</td>
<td>8.672</td>
<td>-33.13</td>
<td>6.159</td>
<td>-30.35</td>
<td>472.94</td>
<td></td>
</tr>
<tr>
<td></td>
<td>516.97</td>
<td>48.71</td>
<td>6.910</td>
<td>46.20</td>
<td>9.697</td>
<td>512.98</td>
<td>40.04</td>
</tr>
</tbody>
</table>

The table reports twice the natural logarithm of the Bayes factors. The entry \((i, j)\) in the matrix compares the model in the \(i\)th row and the model in the \(j\)th column, with a positive value favoring the first model and a negative value favoring the latter.

We further investigate the performance of the two volatility specifications in different periods of time. Johannes et al. (2009) suggest tracking the likelihood ratio sequentially to help identify when a model fails. We compute the sequential deviance, defined as

\[
D_t = -2(\log p(Z_{1:t}|\hat{\theta},M^{TS}) - \log p(Z_{1:t}|\hat{\theta},M^{Log}))
\]

We choose \(\hat{\theta}\) to be the posterior mean of \(\theta\), since the log-likelihood evaluated at the mean and median are quite close. As before, we use 300,000 particles when computing the likelihood. The results are plotted in Figure 8. By comparing the three panels in Figure 8 we see that while \(D_t\) behaves similarly across the different jump specifications, its dynamics reveal important differences between the TS-OU and log-OU volatility processes. Compared with Figure 6, we see that \(D_t\) is positive during 2007 and the beginning of 2008, indicating that log-OU specification provides a better fit than the TS-OU specification, in this low volatility period. In mid 2008, \(D_t\) drops drastically at the pronounced volatility spikes, and stays negative in the following relatively volatile period. After 2010, as the market enters a more tranquil period, \(D_t\) starts to increase, resulting a in positive value for the full sample.

Our results indicate that the TS-OU specification is well suited for volatile periods where the volatility of volatility is high, while the log-OU specification fits the tranquil periods better. The differences could arise from two aspects: First, while the log-OU volatility process is a continuous process, the TS-OU process is purely jump-driven, allowing it to better explain large price changes.
Second, from Figure 7, we see that the autocorrelation function for the log-OU specification decays slower, inferring a more persistent volatility and providing a better fit to the empirical autocorrelation function. This helps the log-OU specification to outperform the TS-OU specification in the last 4 years of the sample period. The theoretical autocorrelation function for the TS-OU specifications, fits the first few lags of the empirical autocorrelation very well and the faster decay rate of the function stress its ability to better capture large changes in the volatility, as seen in the first period of our data set. Figure 7 also suggests extending the model to a multi-factor specification for the volatility process.

We also conduct LR tests in the nested models. The models SF-J, SF-SVTS, SF-SVJT and TF-J are all nested by TF-SVJT, and we report the LR test statistics and p-values in Table 5. With a 10% significance level, all of the restricted models are rejected. The same conclusion applies to the log-OU specification at a 5% significance level.

![Sequential Deviance](image)

Figure 8: The figure plots the sequential deviance between models with different volatility specifications.
Table 5: LR-tests for the full model.

<table>
<thead>
<tr>
<th></th>
<th>SF-J</th>
<th>SF-SVTS</th>
<th>SF-SVJT</th>
<th>TF-J</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>LR</strong>&lt;sub&gt;mean&lt;/sub&gt;</td>
<td>474.072</td>
<td>9.11</td>
<td>4.31</td>
<td>468.802</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td>(0.058)</td>
<td>(0.038)</td>
<td>(0.000)</td>
</tr>
<tr>
<td><strong>LR</strong>&lt;sub&gt;median&lt;/sub&gt;</td>
<td>477.453</td>
<td>11.239</td>
<td>5.69</td>
<td>472.157</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td>(0.024)</td>
<td>(0.017)</td>
<td>(0.000)</td>
</tr>
</tbody>
</table>

The table reports the LR-test statistic for comparing the models in the columns to the full TF-SVJ model. The p-values for accepting the null model are reported in parentheses.

5.3 Model validation

We conclude our empirical investigation by checking that our proposed model is able to reproduce the statistical properties of the data. For purposes of derivative pricing and risk management, it is very important that the model implied price and return distributions match the empirical distributions. For each estimated model, we simulate 5000 artificial data sets of the same length as the observed spot prices. Then, using these simulated paths, we calculate the empirical model implied distributions of skewness and kurtosis. In Table 6, the 5th, 50th and 90th percentile of the model implied distributions for the skewness and kurtosis of the deseasonalized logarithmic spot price, Z(t), and its returns are reported for each model. The table also reports the values from the observed deseasonalized UK natural gas spot prices.

Table 6 shows that stochastic volatility is crucial for capturing the price skewness and kurtosis. All the models, except for the SF-J and TF-J model, have distributions covering the observed negative price skewness, and small excess kurtosis, of the deseasonalized UK natural gas spot prices. For both the SF-J and TF-J model, the probability that the observed sample value of the skewness and kurtosis are realizations from the model implied distributions, is less than 5%. The best performing model is the TF-SVJ<sub>Log</sub> model, with the SF-SVJ<sub>Log</sub> model being a close runner-up. In fact, as we shall see in Figure 9, there is almost no difference in the distribution of skewness and kurtosis.
Table 6: Distribution of skewness and kurtosis

<table>
<thead>
<tr>
<th>Precile</th>
<th>SF-J</th>
<th>SF-SV \text{TS}</th>
<th>SF-SV \text{Log}</th>
<th>SF-SVJ \text{TS}</th>
<th>SF-SVJ \text{Log}</th>
<th>TF-J</th>
<th>TF-SVJ \text{TS}</th>
<th>TF-SVJ \text{Log}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Price Skewness</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5th</td>
<td>-0.628</td>
<td>-0.827</td>
<td>-1.088</td>
<td>-0.813</td>
<td>-1.113</td>
<td>-0.505</td>
<td>-0.793</td>
<td>-1.129</td>
</tr>
<tr>
<td>50th</td>
<td>-0.011</td>
<td>-0.004</td>
<td>0.005</td>
<td>-0.028</td>
<td>0.006</td>
<td>-0.014</td>
<td>-0.002</td>
<td>-0.026</td>
</tr>
<tr>
<td>95th</td>
<td>0.603</td>
<td>0.808</td>
<td>1.117</td>
<td>0.759</td>
<td>1.141</td>
<td>0.491</td>
<td>0.757</td>
<td>1.059</td>
</tr>
<tr>
<td>Price Kurtosis</td>
<td>3.8611</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5th</td>
<td>2.065</td>
<td>2.106</td>
<td>2.242</td>
<td>2.023</td>
<td>2.271</td>
<td>2.238</td>
<td>2.245</td>
<td>2.296</td>
</tr>
<tr>
<td>50th</td>
<td>2.632</td>
<td>2.832</td>
<td>3.174</td>
<td>2.728</td>
<td>3.245</td>
<td>2.778</td>
<td>3.106</td>
<td>3.300</td>
</tr>
<tr>
<td>95th</td>
<td>3.651</td>
<td>4.445</td>
<td>5.855</td>
<td>4.201</td>
<td>5.972</td>
<td>3.655</td>
<td>4.908</td>
<td>5.867</td>
</tr>
<tr>
<td>Return Skewness</td>
<td>-0.3703</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5th</td>
<td>-0.515</td>
<td>-0.875</td>
<td>-1.425</td>
<td>-1.297</td>
<td>-1.771</td>
<td>-0.622</td>
<td>-2.050</td>
<td>-1.999</td>
</tr>
<tr>
<td>50th</td>
<td>-0.071</td>
<td>0.024</td>
<td>0.012</td>
<td>-0.377</td>
<td>0.039</td>
<td>-0.203</td>
<td>-0.224</td>
<td>-0.540</td>
</tr>
<tr>
<td>95th</td>
<td>0.376</td>
<td>0.850</td>
<td>1.387</td>
<td>0.508</td>
<td>2.411</td>
<td>0.206</td>
<td>1.493</td>
<td>0.798</td>
</tr>
<tr>
<td>Return Kurtosis</td>
<td>21.4775</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>95th</td>
<td>10.593</td>
<td>20.381</td>
<td>41.516</td>
<td>21.153</td>
<td>68.271</td>
<td>10.145</td>
<td>210.076</td>
<td>140.477</td>
</tr>
</tbody>
</table>

The table reports the skewness and kurtosis of the deseasonalized logarithmic spot price ($Z(t)$) and its return series. The table also report the percentiles of the simulated sample skewness and kurtosis for each of the considered models. For each model, the percentiles are computed using 5000 simulated data sets with 1620 observations each. The simulations are performed using the parameter estimates from Table 3.
implied by these two models.

Turning our attention to the return distribution, Table 6 shows that the return skewness can be captured by all the considered models. Once again, we find that inclusion of stochastic volatility enables the model to produce more skewness. The transition from a single factor model to a two-factor model also impacts the distribution of return skewness, in contrast to the results found for the price skewness. The model that captures the return skewness the best, in terms of producing the highest probability of observing an outcome from the model implied distribution that is more extreme than the sample skewness of the data, is the SF-SVJ$^{TS}$ model. As for the kurtosis of the return series, it follows from Table 6 that stochastic volatility is essential in order to produce a high enough level of kurtosis. It is also clear that the volatility specification impacts the distribution of the kurtosis of the returns. For the single factor models, only the models with a log-OU volatility specification are able to fully match the level of kurtosis observed in the data. With the TS-OU specification, the probability of observing an outcome from the distribution of kurtosis, that is more extreme than the sample value from our data, is only around 5%. Allowing the spike process to have its own mean-reversion rate, significantly impact the models ability to generate high levels of kurtosis. In the two-factor models, a jump is now followed by a quick reversion back to the base-signal process, instead of the much slower reversion found in the single factor models, and this behavior increases the kurtosis of the return series. The distribution that captures the observed kurtosis the best, is the distribution implied by the SF-SVJ$^{Log}$ model. The closest runner-up is the TF-SVJ$^{Log}$ model.

![Figure 9: The figure plots the model implied distributions of skewness and kurtosis of the deseasonalized logarithmic spot price $Z(t)$, along with the sample values.](image-url)
In the Figure 9 and Figure 10 the model implied distributions of skewness and kurtosis of the price and return series are plotted for our favoured model, the TF-SVJLog model, and the closest contender, the SF-SVJLog model. The reported p-values should be interpreted as the probability of observing an outcome from the distribution that is more extreme than the sample value. Hence, if the sample value matches the median of the model implied distribution, then we report a p-value of 1.

Following the discussion above, Figure 9 and 10 highlights the fact that allowing for separate mean-reversion rates matters for the return distribution in terms of matching skewness and kurtosis. Also, in the TF-SVJLog model, the estimated jump intensity is about seven times larger than the jump intensity in the SF-SVJLog model, which contributes to the increased spread of the kurtosis distribution. Both models are, however, able to capture the statistical properties of the data at hand.

To sum up, we find that inclusion of stochastic volatility in the models is crucial for matching the price and return distributions. In financial applications, such as risk management, matching the skewness and kurtosis of the return series is often more important than matching the price distribution. In this context, we found that the volatility specification and the inclusion of a separate mean-reversion rate for the spike process greatly impacts the models ability to generate high levels of kurtosis.
6 Discussion: Implications and Extensions

The economical importance of our findings remain to be tested. One approach could be to consider how the different model specifications impacts the forward prices by considering the empirical risk premium, \( RP(t) = F^{\text{observed}}(t, T_1, T_2) - F^{P}(t, T_1, T_2) \), defined as the difference between the observed market price and the predicted spot price over the period of delivery \([T_1, T_2]\). The latter is computed using the theoretical forward prices with \( Q = P \), and then average it over the delivery period. From the forward prices derived in Benth and Vos (2013b), we suspect that the forward prices will depend on the factors \( X(t) \) and \( Y(t) \) as well as the stochastic volatility process \( \sigma^2(t) \). Among other things, this has the implication that even without jumps in the spot price model the forward price can still jump if the volatility process has jumps. The study of forward price dynamics might therefore serve as a way of testing the economical difference between the different volatility specifications. Derivation of forward prices in our full model, TF-SVJ, is therefore a topic for future research.

It would also be of interest to see how the model performs on data from other energy markets. In particular the electricity market, where spikes are larger and more frequent and the overall volatility of the observed spot prices are larger compared to the gas spot prices. In this data it might also be important to consider other jump size specifications as most of the spikes are positive.

The clustered jumps in model SF-J suggest that it could be interesting to incorporate a time-varying jump intensity. For example, we can specify a stochastic process for the jump intensity, or allow the jump intensity to depend on the spot volatility or on exogenous variables such as weather.

From the plot of the autocorrelation functions for the filtered volatility processes, it appears that a better model fit might be obtained from extending the volatility specification to a multi-factor specification with different decay rates for the entering factors.

Another extension of the model specification could be to incorporate leverage effects. In Green and Nossman (2008) the authors introduce leverage in their model by making the driving Brownian motions of the volatility process and base-signal process correlated. This would of course only be possible with the log-OU volatility specification. With the TS-OU specification, another approach could be to investigate the implications of making the volatility and the spike process correlated.

Finally, it would be very interesting to try to adapt our estimation approach to the multi-dimensional model from Benth and Vos (2013a) and investigate how the different model characteristics impact the joint modelling of for instance gas and electricity spot prices.
7 Conclusion

We proposed a two-factor geometric model with stochastic volatility and jumps for the detrended and deseasonalized logarithmic UK gas spot price. We then described how this model could be estimated by using a non-Markovian representation of the model with the spike process and stochastic volatility process being the latent variables. The estimation method employed, adopts the particle marginal Metropolis-Hastings sampler from Andrieu et al. (2010) and has the advantage of being easy to adopt to different volatility specifications, including pure jump-driven specifications. The estimation method is very general and made it easy to estimate and compare our proposed model to other benchmark models in a unified framework. This also made it possible answer question like: What matters the most, including stochastic volatility or jump? Is this conclusion affected by the volatility specification? And is it necessary to allow for jumps in the volatility process? Our empirical application to logarithmic UK natural gas spot prices showed that inclusion of stochastic volatility is much more important than having jumps in the model. Like, in Green and Nossman (2008) we found that neglecting to include stochastic volatility results in a severely overestimated jump intensity. From the results for the two-factor version of the model with jumps and constant volatility we saw that the inclusion of separate mean-reversion rates did not make the jump intensity drop. The results for the single factor model with jumps (SF-J), resembling the model from Cartea and Figueroa (2005), also revealed the need for a time-varying jump intensity when stochastic volatility is not included in the model.

Sticking with a single factor model, but allowing for both stochastic volatility and jumps showed that the volatility specification can impact the estimated jump intensity and jump size distribution. The TS-OU specification, where the volatility process is allowed to jump, actually resulted in a higher jump intensity than with the log-OU volatility specification. From fitting the full model (TF-SVJ), with separate mean-reversion rates, jumps and stochastic volatility, we found that having different mean-reversion rates justifies the inclusion of jumps in the model specification even though the spike process only accounts for a small part of the variations in our data. The model with log-OU volatility specification outperforms the model with TS-OU specification, even when jumps and different mean-reversion rates are included in the first model and not in the latter. By tracking the likelihood ratio sequentially, it became clear that the TS-OU volatility specification is well suited for volatile periods where the volatility process is changing a lot, while the log-OU specification outperforms the TS-OU specification during the more tranquil periods where the volatility of volatility is low.

The estimation method based on particle MCMC is very general and it would be interesting to extend our proposed model by incorporating a time-varying jump intensity to see if this changes the
split between how much of the variations in the data is captured by the base-signal and how much is captured by the spike process. Another path for future research, is to investigate if stochastic volatility also matter the most when other energy spot prices, such as electricity spot, are considered. In electricity prices, spikes are larger and more frequent so the inclusion of a time-varying jump intensity would might be more relevant in this setting. The proposed estimation method also has potential in applications to multi-dimensional models like the one considered in Benth and Vos (2013a) and it would be very interesting to adapt the approach to this setup.
Appendix

7.1 Model Diagnostics

We present the model diagnostic plots for TF-SVJ$^{TS}$ and TF-SVJ$^{Log}$ in Figure 11 and 12. The plots for other models are similar and omitted to save space. The left panels are trace plots of parameter draws against iterations, while the right panels report the prior and empirical posterior distributions (histogram) of the parameters. We first look at the trace plots. For model TF-SVJ$^{TS}$, we ran four chains in parallel from different starting points for speed considerations, and the vertical lines indicates when a new chain is started. For TF-SVJ$^{Log}$, one long chain is used. Visual inspection indicates convergence, although the mixing of some parameters is less satisfactory, for example for $\sigma_j$, the standard deviation of jump sizes. This is likely due to the very low jump intensity in these two models, and hence the algorithms have a hard time estimating the jump sizes.

From the posterior-prior plots, we see that for most of the parameters, prior information is negligible compared with posterior. To assist visual inspections, we use different scales for the densities: ticks on the left axis are the density for posterior, while ticks on the right axis denotes the density for prior. For example in Figure 11, $\alpha_x$ in the TF-SVJ$^{TS}$ model, the prior density at $\alpha_x = 0.0087$ is around 1, while the posterior density is about 180. For $\alpha_x$ in the TF-SVJ$^{Log}$ model, the prior density at $\alpha_x = 0.0077$ is around 1, while the posterior density is about 250. One exception is $\sigma_j$, for which we choose a prior that places lower probability on jumps being small. The posterior-prior plots in both models indicate that the prior for $\sigma_j$ is informative about the posteriors.
Figure 11: Diagnostic Plots for the TF-SVJ model with TS-OU volatility. Left panels are trace plots of parameters. In the right panels, blue lines are posterior densities, while green lines indicate priors.
Figure 12: Diagnostic Plots for the TF-SVJ model with log-OU volatility. Left panels are trace plots of parameters. In the right panels, blue lines are posterior densities, while green lines indicate priors.
References


