





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

Asger Lunde, Anne Floor Brix and Wei Wei

CREATES Research Paper 2015-46

Department of Economics and Business Economics Aarhus University Fuglesangs Allé 4 DK-8210 Aarhus V Denmark Email: oekonomi@au.dk Tel: +45 8716 5515

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

Anne Floor Brix^{a, \dagger} Asger Lunde^{a, \ddagger} Wei Wei^{a, \S}

^aAarhus University & CREATES

October 7, 2015

Abstract

We propose an energy spot price model featuring a two-factor price process and a two-component stochastic volatility process. The first factor in the price process captures the normal variations; the second accounts for spikes. The two-component volatility allows for a flexible autocorrelation structure. 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 an adaptive MCMC method with a Rao-Blackwellized particle filter. We fit the model to UK natural gas spot prices and investigate the importance of spikes and stochastic volatility. We find that the inclusion of stochastic volatility is crucial and that it strongly impacts the jump intensity in the spike process. Furthermore, our estimation method enables us to consider both continuous and purely jump-driven volatility processes, and thereby assess if the volatility specification also influences the spike process and the overall model fit.

Keywords: Energy prices, Multi-factor model, Particle filters, MCMC, Stochastic volatility. *JEL Codes:* C11, C15, C5, C52, Q4

^{*}The authors acknowledge support from the project entitled "Stochastic and Econometric Analysis of Commodity Markets" funded by Aarhus University Research Foundation and CREATES - Center for Research in Econometric Analysis of Time Series (DNRF78), funded by the Danish National Research Foundation. Wei Wei acknowledges financial support from the The Danish Council for Independent Research, Social Sciences (4003-00106B/FSE).

[†]email: afb@jyskebank.dk

[‡]email: alunde@econ.au.dk

[§]email: wwei@econ.au.dk

1 Introduction

The liberalization of European energy markets over the last couple of decades has led to highly deregulated and liquid markets for trading energy commodities, in particular gas and electricity. The introduction of competition has caused the markets to experience a significant increase in price volatility. In addition, a market place for energy-based derivatives 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 Evdeland and Wolyniec (2003) for empirical evidence on these stylized facts. Seasonality is caused by the seasonal pattern on the demand side of the market. 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 a jump, the price quickly mean-reverts to the normal level, leaving a spike in the price path. In other words, the spike process has a faster mean-reversion rate than normal variations, and this contributes to the multi-scale autocorrelation structure observed in many markets. The normal fluctuations, typically referred to as the "base-signal" process, often display volatility clustering. Stochastic volatility helps replicate the time series properties of the prices, such as volatility clustering and leptokurtic distributions.

In the univariate model proposed in this paper, the logarithmic spot price is specified as the sum of a deterministic function and a two-factor stochastic process as in Benth et al. (2008). The deterministic function models the trend and seasonality of the logarithmic spot price. The first stochastic factor captures the base-signal part of the price process, and it is modeled using a Gaussian Ornstein-Uhlenbeck (OU) process with stochastic volatility. The second factor is a purely jump-driven OU process that accounts for the spike behavior.

The role of stochastic volatility is an important issue to address. Different stochastic volatility models have different implications in hedging strategies and derivative pricing, and they have been studied extensively for equity returns, see for example Shephard (2005) and Chernov et al. (2003). The first volatility model we consider is the non-Gaussian OU processes proposed in Barndorff-Nielsen and Shephard (2001a). This class of jump-driven models is widely adopted because of its analytical tractability and its ability to generate large fluctuations. We also consider a continuous specification, which models the logarithmic variance by a Gaussian OU process.

Both volatility specifications can be augmented by allowing multiple components.¹ In the case

 $^{^{1}}$ We use "multi-component" instead of "multi-factor" for volatility throughout this paper to distinguish it from the multi-factor spot price.

of jump-driven volatility, we use the superposition of independent non-Gaussian OU processes as in Griffin and Steel (2006). With continuous volatility, the second component is added to the logarithmic variance in the similar fashion as Alizadeh et al. (2002) and Chernov et al. (2003). Multi-component volatility models have two major benefits. First, different mean-reversion rate in the different components generate flexible dependence structure in the overall volatility. Second, tail of volatility and the persistence can be captured separately. Our proposed model features a two-factor price process and a two-component stochastic volatility process. In addition, we estimate and compare twelve nested models to study the interaction between price dynamics and volatility dynamics and assess the importance of each element.

Our paper contributes to the literature by proposing a method for estimating the model in one step using the particle Markov chain Monte Carlo (PMCMC) methods developed in Andrieu et al. (2010). PMCMC is a Bayesian approach which conducts likelihood-based inference through a particle filter. Flury and Shephard (2011) demonstrate the use of PMCMC in economic models including the lognormal stochastic volatility model we adopt. The complication induced by stochastic volatility has been treated by a variety of methods, such as efficient method of moment (see e.g. Gallant and Tauchen, 1996 and Chernov et al., 2003), simulated maximum likelihood (see e.g. Durham, 2006), and Markov chain Monte Carlo (see e.g. Eraker et al., 2003 and Kalli and Griffin, 2015). In Green and Nossman (2008), a multi-factor model for power prices with Heston-type volatility specification is estimated in one step using MCMC. In contrast to our approach, where an unbiased estimate of the likelihood is obtained from the particle filter, Green and Nossman (2008) condition on future values when computing the posterior distribution, rendering in-sample volatility a useless estimate for likelihood 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.

Another advantage of PMCMC is its ability to handle non-Markovian models, which is an essential element in the effective sampling algorithm we propose for the multi-factor model. Pitt et al. (2012) show that the performance of PMCMC can be severely hindered by the variance of the likelihood estimate, which depends on the efficiency of the particle filter. In the presence of jumps or spikes, the efficiency of the particle filter is particularly important, see for example Johannes et al. (2009). We propose an algorithm based on a non-Markovian representation of the multi-factor model and use Rao-Blackwellization to decrease the variance of the likelihood estimate. Rao-Blackwellized particle filter is closely related to the fully adapted particle filter in Pitt and Shephard (1999) and Pitt et al. (2012), but it could also deal with stochastic volatility models driven by jumps.²

Our proposed model encompasses many of the existing models in the energy literature. The stepping stone for many of the spot price models found in the literature is the mean-reverting one-

 $^{^{2}}$ In Fileccia and Sgarra (2015a,b) the authors use PMCMC with auxiliary particle filter to estimate a non-stationary model with jumps for the crude oil market. They model considers a stochastic mean factor and a stochastic volatility factor, both specified with a continuous process.

factor Schwartz model from Schwartz (1997), where the logarithmic spot price follows a Gaussian OU process. This model is 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) is generalized to allow for jumps. A special case of this model 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 has been used for modeling electricity spot prices in Cartea and Figueroa (2005) and Benth et al. (2012).

The inclusion of stochastic volatility in energy models was suggested by Geman (2005), where a Heston stochastic volatility extension of the Schwarz model is considered. 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 MCMC methods. A jump-driven specification of the volatility process is considered in Benth (2011), where the geometric one-factor model is augmented with stochastic volatility given by the sum of non-Gaussian OU processes. The model is fitted to UK natural gas spot prices using a one-component volatility process. Benth and Vos (2013) extend the stochastic volatility model from Benth (2011) to incorporate spikes and leverage effects in a multidimensional setting, allowing for the joint modeling of several commodities. The model in Benth and Vos (2013) only allows positive jumps in the spot price, since the non-Gaussian OU factors entering the model are driven by subordinators. Estimation of the model from Benth and Vos (2013) is still an open question. The estimation method detailed and employed in this paper also has potential for usage in the multi-dimensional setup.

The paper is organized as follows: In Section 2 our proposed model and the nest alternatives are presented. The PMCMC estimation method is outlined in Section 3, and Section 4 describes the data of our empirical application. 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. Concluding remarks are given in Section 7.

2 Model Descriptions

We adopt a flexible modeling framework proposed in Benth et al. (2008). Let S(t) denote the spot price at time t. The dynamics of the spot price is described using the following geometric OU-based factor model, augmented with stochastic volatility:

$$d \log S(t) = d \log \Lambda(t) + dX(t) + dY(t),$$

$$dX(t) = -\alpha_x X(t) dt + \sigma(t) dB(t),$$

$$dY(t) = -\alpha_y Y(t) dt + dI(t).$$

The deterministic function $\Lambda(t)$ accounts for the possible trend and seasonal patterns of the data. In this section we focus 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 $\Delta = 1$. The first factor, X(t), is a Gaussian OU process with stochastic volatility $\sigma(t)$. It is interpreted as the base-signal process that models the continuous variations in the logarithmic spot price. The second factor, Y(t), is a non-Gaussian OU process with a pure jump Lévy process as background driving Lévy process (BDLP). Y(t) is 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. This model is also a simplified example of superposition of Lévy semistationary processes studied in Barndorff-Nielsen et al. (2013).

	Two-comp SV $$	One-comp SV	Jumps in vol.	Jumps in price	Two-factor
SF-J				\checkmark	
TF-J				\checkmark	\checkmark
$SF-SV^{IG}$		\checkmark	\checkmark		
$\mathrm{SF}\text{-}\mathrm{SV}^{\mathrm{Log}}$		\checkmark			
$SF-SVJ^{IG}$		\checkmark	\checkmark	\checkmark	
$\rm SF\text{-}SVJ^{Log}$		\checkmark		\checkmark	
$\mathrm{TF} ext{-}\mathrm{SVJ}^{\mathrm{IG}}$		\checkmark	\checkmark	\checkmark	\checkmark
$\mathrm{TF}\text{-}\mathrm{SVJ}^{\mathrm{Log}}$		\checkmark		\checkmark	\checkmark
$SF-S2V^{IG}$	\checkmark		\checkmark		
$\mathrm{SF}\text{-}\mathrm{S2V}^{\mathrm{Log}}$	\checkmark				
$SF-S2VJ^{IG}$	\checkmark		\checkmark	\checkmark	
$\rm SF\text{-}S2VJ^{Log}$	\checkmark			\checkmark	
$\mathrm{TF} ext{-}\mathrm{S2VJ}^{\mathrm{IG}}$	\checkmark		\checkmark	\checkmark	
$TF-S2VJ^{Log}$	\checkmark			\checkmark	\checkmark

Table 1: Model Overview.

We consider 14 models in total, which comprises three specifications for price jumps: (1) no-jump; (2) single factor jump-diffusion; (3) two-factor spike process; and four specifications for volatility: (1) jump-driven OU; (2)two-component jump-driven OU; (3) lognormal OU; (4) two-component lognormal OU. We also include two constant volatility models. Table 1 gives an overview of the models we consider.

The next section explains in detail the specifications of X(t) and Y(t) in our proposed model, and the following subsections present alternative models that are nested in the model described in section 2.1.

2.1 Two-Factor Models with Stochastic Volatility

In our proposed two-factor models with stochastic volatility, we will use I(t) = N(t) as the BDLP, where N(t) is a compound Poisson process with intensity parameter λ_J and normally distributed jump sizes. The detrended and deseasonalized spot price, Z(t), will then solve

$$dZ(t) = dX(t) + dY(t)$$

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

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

$$Z_{t+1} = X_{t+1} + Y_{t+1}$$

$$X_{t+1} = e^{-\alpha_x} X_t + \epsilon_{t+1}$$

$$Y_{t+1} = e^{-\alpha_y} Y_t + \xi_{t+1} J_{t+1},$$
(1)

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

A similar model was suggested in Green and Nossman (2008), using a Heston specification of the stochastic volatility process and including an additional independent Brownian component in the spike process, Y(t). Adding a continuous Brownian component in Y(t) simplifies estimation as it gives Y(t) a density, but it contradicts our interpretation of Y(t) as the spike process.

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. The specifications of the variance process $\sigma^2(t)$ are given below.

2.1.1 Inverse Gaussian(IG)-OU Stochastic Volatility

The first specification we consider for $\sigma^2(t)$ is an IG-OU process:

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

where L(t) is an infinite activity jump process and the marginal distribution of $\sigma^2(t)$ follows an inverse Gaussian (IG) distribution, $\sigma^2(t) \sim IG(\delta, \gamma)$. Note that this is a special case of the tempered stable distribution, $TS(\kappa, \delta, \gamma)$, with $\kappa = 0.5$. The IG-OU volatility process is considered in Benth (2011) to fit the logarithm of natural gas spot prices in the UK. He finds that the jump-driven volatility specification offers more flexible dynamics than the continuous Heston model. This process can be simulated recursively from the transition equation

$$\sigma^2(t+1) = e^{-\lambda}\sigma^2(t) + e^{-\lambda}\int_0^1 e^{\lambda u} dL(\lambda u).$$
⁽²⁾

The transitional density is not available analytically. We use an accept-reject algorithm proposed by Zhang and Zhang (2008) to simulate the innovations in (2). This algorithm is based on exact simulation and it is considerably faster than the infinite series representation of Rosiński (see Barndorff-Nielsen and Shephard, 2001b) or the Ferguson and Klass representation adopted in Gander and Stephens (2007).

Equation (1) and (2) form our first model, which features two factors in price with jumps and stochastic volatility with IG density. We term this model TF-SVJ^{IG}.

2.1.2 Two-Component IG-OU Stochastic Volatility

We extend the IG-OU model to a two-component stochastic volatility model where the volatility is the sum of two independent IG-OU processes:

$$\sigma^2(t) = \sigma_1^2(t) + \sigma_2^2(t)$$

$$\sigma_i^2(t+1) = e^{-\lambda_i} \sigma_i^2(t) + e^{-\lambda_i} \int_0^1 e^{\lambda_i u} dL(\lambda_i u),$$
(3)

where i = 1, 2 denotes the two components in the volatility process. The marginal distribution of $\sigma_i^2(t)$ follows $\sigma_i^2(t) \sim \text{IG}(\delta_i, \gamma)$. We restrict γ to be the same for the two components such that their marginal distributions have the same mean/variance ratio, and hence the superposition of $\sigma_1^2(t)$ and $\sigma_2^2(t)$ still follows the IG marginal, $\sigma^2(t) \sim \text{IG}(\delta_1 + \delta_2, \gamma)$. The different mean-reversion rates, λ_1 and λ_2 , give rises to a flexible autocorrelation structure in the volatility process:

$$\operatorname{Corr}(\sigma^2(t), \sigma^2(t-u)) = w_1 \exp(-\lambda_1 u) + w_2 \exp(-\lambda_2 u),$$

where $w_i = \delta_i / (\delta_1 + \delta_2)$. We term this model TF-S2VJ^{IG}.

2.1.3 Log-OU Stochastic Volatility Process

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. The transition density for the logarithmic volatility is given by

$$h(t)|h(t-1) \sim N\left(e^{-\alpha_h}h(t-1), \frac{1-e^{-2\alpha_h}}{2\alpha_h}\sigma_h^2\right)$$

We denote this model TF-SVJ^{Log}.

2.1.4 Two-Component Log-OU Stochastic Volatility

We extend the logarithmic OU model to a two-component stochastic volatility model where the volatility is the sum of two independent log-OU process:

$$\sigma^{2}(t) = \exp\left(\mu_{h} + h_{1}(t) + h_{2}(t)\right)$$

$$h_{1}(t)|h_{1}(t-1) \sim N\left(e^{-\alpha_{h1}}h_{1}(t-1), \frac{1-e^{-2\alpha_{h1}}}{2\alpha_{h1}}\sigma_{h1}^{2}\right)$$

$$h_{2}(t)|h_{2}(t-1) \sim N\left(e^{-\alpha_{h2}}h_{2}(t-1), \frac{1-e^{-2\alpha_{h2}}}{2\alpha_{h2}}\sigma_{h2}^{2}\right).$$
(4)

Let $V_{h1} = \sigma_{h1}^2/(2\alpha_{h1})$ and $V_{h2} = \sigma_{h2}^2/(2\alpha_{h2})$ denote the unconditional variance of h_1 and h_2 , respectively; the autocovariance and autocorrelation function implied by this model is

$$Cov(\sigma^{2}(t), \sigma^{2}(t-u)) = \exp(2\mu_{h} + V_{h1} + V_{h2}) \left(\exp\left(V_{h1}e^{-\alpha_{h1}u} + V_{h2}e^{-\alpha_{h2}u}\right) - 1\right)$$
$$Corr(\sigma^{2}(t), \sigma^{2}(t-u)) = \frac{\left(\exp\left(V_{h1}e^{-\alpha_{h1}u} + V_{h2}e^{-\alpha_{h2}u}\right) - 1\right)}{\left(\exp\left(V_{h1} + V_{h2}\right) - 1\right)}.$$

We denote this model TF-S2VJ^{Log}.

2.2 Two-Factor Model with Constant Volatility

In this subsection and the next one, we consider models that are nested in our proposed models in section 2.1. These nest alternatives serve as a comparison to help us understand what are the important elements in modeling the deseasonalized logarithmic spot price Z(t).

The first alternative model we consider is obtained by imposing constant volatility in (1). The latent $\sigma^2(t)$ in the full model becomes a constant parameter, σ^2 , in this simplified model. The second factor, i.e. the spike process Y(t), is unaffected. We term this model TF-J.

2.3 Single-Factor Models

We consider single factor models in the sense that the base-signal and spikes have the same meanreversion rate, i.e., $\alpha_x = \alpha_y = \alpha$. This restriction has the important implication that Z(t) will be a Markov process:

$$Z_{t+1} = e^{-\alpha} Z_t + \epsilon_{t+1} + \xi_{t+1} J_{t+1}, \tag{5}$$

where $\epsilon_{t+1} \sim N\left(0, \sigma^2(t)(1-e^{-2\alpha_x})/(2\alpha_x)\right)$. The BDLP of Y(t) is still assumed to be compound Poisson with normally distributed jump sizes, hence we obtain the same specification for the innovations in the spike process: $J_{t+1} \sim Bernoulli(\lambda_J)$ and $\xi_{t+1} \sim N(\mu_J, \sigma_J^2)$.

If we further impose $\sigma^2(t) = \sigma^2$, we obtain model SF-J which is a single-factor model with jumps and constant volatility. 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 use a log-normal jump distribution. Alternatively, we can allow stochastic volatility in (5). We adopt the same volatility specification as in section 2.1, and obtain the following four models: SF-SVJ^{IG}, SF-S2VJ^{IG}, SF-SVJ^{Log}, and SF-S2VJ^{Log}.

To investigate the role of jumps, we also consider a single-factor Gaussian OU process with stochastic volatility, corresponding to the restriction Y(t) = 0. The discretized model becomes

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

where $\epsilon_{t+1} \sim N(0, \sigma^2(t)(1 - e^{-2\alpha_x})/(2\alpha_x))$. Depending on the specifications of $\sigma^2(t)$, we have model SF-SV^{IG}, SF-S2V^{IG}, SF-SV^{Log}, and SF-S2V^{Log}. The SF-SV^{IG} model is a special case of the model considered in Benth (2011).

3 Estimation Method

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 nonlinear 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) develop 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 multi-factor and continuous stochastic volatility.

The third complication arises 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 when the volatility is driven by a compound Poisson process. Griffin and Steel (2006) propose an algorithm using dependent thinning and reversible jump MCMC. Gander and Stephens (2007) extend their method to allow more general BDLP. However, these procedures can not be easily generalized to the multi-factor models commonly used for energy prices.

Instead, 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 lognormal 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. In addition, we obtain a likelihood estimate 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, then applys MH algorithms to obtain the joint posterior distribution of the parameters and the states. 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 analytically using a technique called Rao-Blackwellization; see Doucet et al. (2000). Our approach is similar 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 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 focuses on the estimation of the model in Section 2.1, as it is the most complex model and nests all the other models.

3.1 Sequential Monte Carlo

In the state space representation of our proposed two factor 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 as in Green and Nossman (2008). This is equivalent to assuming that Y_t is a jump-diffusion instead of pure jump process. 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, bootstrap filters will perform poorly, see Pitt et al. (2012) for an illustration.

We propose a different approach for this problem. Specifically, we use the following representation of the two factor model

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

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

Note 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 sample the states given the parameters. Let θ 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. We denote the particles by $\{K_{1:t}^{(i)}, \tilde{\omega}_t^{(i)}\}_{i=1}^N$, where $K_{1:t}^{(i)}$ is a sample from random vector $K_{1:t}, \tilde{\omega}_t^{(i)}$ is the associated weight, and N is the number of particles. Given particles that approximate $p_{\theta}(K_{1:t}|Z_{1:t})$, SMC obtains samples from $p_{\theta}(K_{1:t+1}|Z_{1:t+1})$ and computes $p_{\theta}(Z_{t+1}|Z_{1:t})$ sequentially. From Bayes Theorem:

$$p_{\theta}(K_{1:t+1}|Z_{1:t+1}) = \frac{p_{\theta}(Z_{t+1}|K_{t+1}, Z_{1:t}, K_{1:t})p_{\theta}(K_{t+1}|K_{1:t})}{p_{\theta}(Z_{t+1}|Z_{1:t})}p_{\theta}(K_{1:t}|Z_{1:t}),$$
(6)

the density of interest $p_{\theta}(K_{1:t+1}|Z_{1:t+1})$ can be sampled using importance sampling techniques. The bootstrap filters choose the importance density $g_{\theta}(K_{1:t+1})$ to be $p_{\theta}(K_{t+1}|K_{1:t})p_{\theta}(K_{1:t}|Z_{1:t})$, i.e., the new particles $K_{t+1}^{(i)}$ are propagated from $K_t^{(i)}$ using only transition densities and are "blind" to the observations. The importance weights are given by the ratio of the target density and the importance density. From equation (6), it is therefore easily seen that the weights for the particles $K_{1:t+1}^{(i)}$ are proportional to $\hat{\omega}_{t+1}^{(i)} \tilde{\omega}_t^{(i)}$, where the incremental weights $\hat{\omega}_{t+1}^{(i)}$ are simply given by the measurement density $p_{\theta}(Z_{t+1}|K_{t+1}^{(i)}, Z_{1:t}, K_{1:t}^{(i)})$. The likelihood, $p_{\theta}(Z_{t+1}|Z_{1:t})$, is the normalizing constant for the particles and is equal to $\sum_{i=1}^{N} \hat{\omega}_{t+1}^{(i)} \tilde{\omega}_t^{(i)}$. After normalizing the weights $\tilde{\omega}_{t+1}^{(i)} = \tilde{\omega}_t^{(i)} \hat{\omega}_{t+1}^{(i)} / p_{\theta}(Z_{t+1}|Z_{1:t})$, the particles $\{K_{1:t+1}^{(i)}, \tilde{\omega}_{t+1}^{(i)}\}_{i=1}^{N}$ 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 target distribution, as the number of effective particles is small. In bootstrap filters, the incremental weights are simply the measurement density. This algorithm is usually applied when the observations are less informative about the states than the transition density, which is not the case when the states include 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 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 importance density.

We employ the Rao-Blackwellization technique, since the innovations in the spike process can be integrated out analytically conditional on other state variables. The vector K_{t+1} has two components, the stochastic volatility, $\sigma^2(t)$, and the 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 (6) as

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

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

This algorithm is easily adapted to different volatility specifications as long as we can simulate $\sigma^2(t)$ from its transition density, $p_{\theta}(\sigma^2(t)|\sigma^2(t-1))$. In the case of two-component IG-OU volatility, The state vector $K_{t+1}^{(i)}$ becomes $\{\sigma_1^2(t,i), \sigma_2^2(t,i), Y_{t+1}^{(i)}\}$. We simulate $\sigma_1^2(t,i) \sim p_{\theta}(\sigma_1^2(t)|\sigma_1^2(t-1,i))$ and $\sigma_2^2(t,i) \sim p_{\theta}(\sigma_2^2(t)|\sigma_2^2(t-1,i))$ independently from equation (3), then compute the weights as before, with $\sigma^2(t,i) = \sigma_1^2(t,i) + \sigma_2^2(t,i)$. In the two-component log-OU volatility specification, the state vector $K_{t+1}^{(i)} = \{h_1(t,i), h_2(t,i), Y_{t+1}^{(i)}\}$, and $\sigma^2(t,i)$ is simulated using equation (4).

If importance sampling is carried out sequentially, weights will degenerate in time and only a few particles will have significant weights after a few iterations. SMC incorporates 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. 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, resampling introduces additional Monte Carlo error, and we implement residual resampling instead of multinomial resampling, see Douc and Cappe (2005). Residual resampling has smaller variance it satisfies the unbiasedness condition, meaning that the expected number of resampled 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. ³

3.2 MCMC

SMC methods approximate the likelihood and state variables conditional on the parameters, but we are interested in the joint distribution of the parameters and states. Note 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 suggests the following proposal

³See Pitt et al. (2012) for a guide on how to choose the optimal number of particles.

density,

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

In this case, a draw from $q(\theta|\theta^g)$, θ^{g+1} , has the simplified acceptance probability,

$$\alpha^{g+1} = \min\left\{\frac{p(Z_{1:T}|\theta^{g+1})p(\theta^{g+1})q(\theta^{g}|\theta^{g+1})}{p(Z_{1:T}|\theta^{g})p(\theta^{g})q(\theta^{g+1}|\theta^{g})}, 1\right\},\tag{7}$$

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 $p(K_{1:T}|\theta^{g+1}, Z_{1:T})$ in the SMC step and accept it jointly with θ^{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, $\theta^{g+1} \sim TN(\theta^g, \beta^g \Sigma^g)$, where TN denotes truncated normal, as some of the parameters have finite support. In particular:

$$q(\theta^{g}|\theta^{g+1}) = \frac{f^{n}(\theta^{g};\theta^{g+1},\beta^{g}\Sigma^{g})}{F^{N}(\theta_{u};\theta^{g+1},\beta^{g}\Sigma^{g}) - F^{N}(\theta_{l};\theta^{g+1},\beta^{g}\Sigma^{g})}$$
$$q(\theta^{g+1}|\theta^{g}) = \frac{f^{n}(\theta^{g+1};\theta^{g},\beta^{g}\Sigma^{g})}{F^{N}(\theta_{u};\theta^{g},\beta^{g}\Sigma^{g}) - F^{N}(\theta_{l};\theta^{g},\beta^{g}\Sigma^{g})},$$
(8)

where f^N and F^N denote pdf and cdf of the multivariate normal distribution respectively, θ_u is the upper limit of parameters, and θ_l is the lower limit of parameters. The ratio of proposal densities in equation (7) simplifies to the ratio of normalizing constants as f^n is symmetric:

$$\frac{q(\theta^g|\theta^{g+1})}{q(\theta^{g+1}|\theta^g)} = \frac{F^N(\theta_u;\theta^g,\beta^g\Sigma^g) - F^N(\theta_l;\theta^g,\beta^g\Sigma^g)}{F^N(\theta_u;\theta^{g+1},\beta^g\Sigma^g) - F^N(\theta_l;\theta^{g+1},\beta^g\Sigma^g)}.$$
(9)

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 Σ^g is the covariance matrix of the target posterior distribution, and the scaling factor β^g is approximately $2.38^2/d$, where d is the number of parameters.

In practice, we do not know Σ^g a priori. Adaptive MCMC allows us to learn Σ^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 θ , and ergodicity of the chain can be perturbed. Haario et al. (2001) propose an adaptive Metropolis (AM) algorithm which uses an increasing part of the chain 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, Σ^g might be a poor initial guess, resulting in too many or too few rejections. Andrieu and Thoms (2008) suggest adapting the scaling factor β^g using the the acceptance probability in (7). If the acceptance probability is higher than the optimal acceptance probability, β^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 compared to the base-signal. Specifically, we use a gamma distribution for the standard deviation of jump sizes which places lower probability on small jumps.

3.3**PMMH** Algorithm

We outline the algorithm for the two-factor 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 $K_{1:T}^{g+1}$:
 - (a) sample $\sigma^2(0, i)$ and $Y_1^{(i)}$ from their stationary distribution.
 - i. compute

$$\omega_1^{(i)} = \frac{p_{\theta}\left(Z_1 | Y_1^{(i)}, \sigma^2(0, i)\right)}{N}$$

for i = 1, ..., N, where N is the number of particles.

- ii. obtain likelihood from $p(Z_1) = \sum_{i=1}^N \omega_1^{(i)}$, and compute normalized weights: $\tilde{\omega}_1^{(i)} = \frac{\omega_1^{(i)}}{p(Z_1)}$.
- (b) at t = 1, ..., T 1
 - i. sample the index $a_t^{(i)}$ for i = 1, ..., N, using $\tilde{\omega}_t$ and set $\tilde{\omega}_t = \frac{1}{N}$. ii. simulate $\sigma^2(t, i) \sim p_\theta(\sigma^2(t) | \sigma^2(t-1, a_t^{(i)})).$
 - iii. sample $Y_{t+1}^{(i)} \sim p_{\theta}(Y_{t+1}|Z_{t+1}, \sigma^2(t, i), Y_t^{a_t^{(i)}}).$
 - iv. compute the incremental weights: $\hat{\omega}_{t+1}^{(i)} = p_{\theta}(Z_{t+1} | \sigma^2(t, i), Y_t^{a_t^{(i)}}, Z_{1:t}).$

 - v. obtain the likelihood: $\hat{p}_{\theta}(Z_{t+1}|Z_{1:t}) = \sum_{i=1}^{N} \hat{\omega}_{t+1}^{(i)} \tilde{\omega}_{t}^{(i)}$. vi. normalize the weights $\tilde{\omega}_{t+1}^{(i)} = \frac{\tilde{\omega}_{t}^{(i)} \hat{\omega}_{t+1}^{(i)}}{\sum_{i=1}^{N} \tilde{\omega}_{t}^{(i)} \hat{\omega}_{t+1}^{(i)}}$.
- (c) at t = T

i. obtain
$$\hat{p}(Z_{1:T}|\theta^{g+1}) = \hat{p}_{\theta}(Z_1) \prod_{t=1}^{T-1} \hat{p}_{\theta}(Z_{t+1}|Z_{1:t}).$$

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

2. accept θ^{g+1} and $K_{1:T}^{g+1}$ with probability:

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

where $\hat{p}(Z_{1:T}|\theta^{g+1})$ is computed from the SMC algorithm above, $p(\theta)$ is the prior density of the model parameters which is detailed in the Appendix, and q is the truncated normal proposal density. The ratio of proposal densities is given in Equation (9). If rejected, we set θ^{g+1} and $K_{1:T}^{g+1}$ equal to θ^g and $K_{1:T}^g$.

3. update the scaling factor and the covariance matrix for the proposal density:

$$\begin{split} \nu^{g+1} &= 1/(g+1)^{0.8} \\ \log \beta^{g+1} &= \log \beta^g + \nu^{g+1} \left(\alpha^{g+1} - \alpha^* \right) \\ \mu^{g+1} &= \mu^g + \nu^{g+1} (\theta^{g+1} - \mu^g) \\ \Sigma^{g+1} &= \Sigma^g + \nu^{g+1} \left(\left(\theta^{g+1} - \mu^g \right) \left(\theta^{g+1} - \mu^g \right)^T - \Sigma^g \right) \end{split}$$

3.4 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? 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 conduct an extensive model comparison.

Given two competing models, say M_1 and M_2 , the Bayes factor is the ratio of the probability of each model given data, i.e., $BF = p(M_1|Z)/p(M_2|Z)$. If we assume that the competing models are equally probable a priori, the Bayes factor can be expressed as the posterior odds ratio: $BF = p(Z|M_1)/p(Z|M_2)$. 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:

$$p(Z|M) = \int p(Z|\theta, M) p(\theta|M) d\theta, \qquad (10)$$

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

We use the output from the PMMH algorithm to compute Bayes factors. The PMMH algorithm produces $\{p(Z|\theta^g, M)\}_{g=1}^G$, where θ^g are draws from the posterior density $p(\theta|Z, M)$. In equation (10), the integration is over the prior density of θ . 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 θ from the posterior, imagine that additional $\delta_p G/(1 - \delta_p)$ samples of θ 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,

$$\hat{p}(Z|M) = \frac{\delta_p G/(1-\delta_p) + \sum_{g=1}^G p(Z|\theta^g, M) / \left(\delta \hat{p}(Z|M) + (1-\delta_p) p(Z|\theta^g, M)\right)}{\delta_p G/\left((1-\delta_p) \hat{p}(Z|M)\right) + \sum_{g=1}^G \left(\delta \hat{p}(Z|M) + (1-\delta_p) p(Z|\theta^g, M)\right)^{-1}}.$$
(11)

This estimator reduces the instability from a simple harmonic mean estimator which corresponds to $\delta_p = 0$. In practice, we choose δ_p to be a small number such as 0.01.

4 Data description

We fit the models from section 2 to a time series of daily UK gas spot prices ranging from September 11, 2007 to February 10, 2014. The data is collected from Bloomberg⁴ and report day-ahead gas spot prices collected at the virtual hub NBP (Natural Balancing Point) for trading days (weekdays) in the sample period. This leaves us with a total of 1620 daily price quotes. We construct a detrended and deseasonalized logarithmic spot price, Z(t), that is displayed in Figure 1. This series serve as the input for our estimations.⁵



Figure 1: Detrended and Deseasonalized Logarithmic Spot Prices.

5 Estimation Results

We apply the PMMH algorithm in section 3.3 to the deseasonalized and detrended logarithmic gas spot prices. We start with a preliminary run which uses adaptive MCMC, then "freeze" the covariance matrix Σ and scaling factor β and run a further 20000 iterations to get the posterior distributions of θ and K. The number of particles is set to three times the number of observations. To conduct model comparison, we compute marginal likelihood, p(Z|M), using equation (11). We also evaluate the likelihood at the mean and median of the posterior distribution of the parameters using 300,000 particles.

⁴Code: NBPGDAHD index.

⁵A detailed description of the data manipulations are given in a supplementary appendix.

The estimation for the other models from Section 2 follows similar procedures. For models with jumps in the logarithmic price, we utilize Rao-Blackwellization and integrate out the jumps analytically for the likelihood.⁶ For models with two factors, we use the non-Markovian representation.

5.1 Parameter and State Estimates

Table 2, 3, and 4 report the parameter estimates from constant volatility models, IG-OU stochastic volatility models, and log-OU stochastic volatility models, respectively. We also report the log-likelihood and marginal log-likelihood in excess of the worst model in each table.

								U C	
	$lpha_x$	α_y	σ^2	μ_J	σ_J	λ_J	LLMean	LLMedian	MarginalLL
SF-J	0.0077 (0.0029)		0.00047 (0.00004)	-0.0019 (0.0052)	0.0934 (0.0048)	0.2445 (0.0234)	0.00	0.00	0.00
TF-J	0.0042 (0.0020)	0.0396 (0.0171)	0.00046 (0.00004)	-0.0058 (0.0053)	0.0914 (0.0045)	0.2521 (0.0229)	2.69	2.72	2.09

Table 2: Parameter Estimates from Constant Volatility models

The table reports the posterior mean and posterior standard deviation of model parameters. LLmean/LLmedian is the log-likelihood evaluated at the posterior mean/median using 300,000 particles. The marginalLL is computed using equation (11). All three log-likelihood are reported in excess of model SF-J.

The parameter estimates in table 2 indicate that neither SF-J nor TF-J adequately capture the dynamics of the data, although allowing different mean-reverting rates for the spike process yields an increase in likelihood. Most of the variability in the data is explained by spikes which severely drives up the estimated jump intensity. In both models, the estimated jump intensity corresponds to roughly one jump every four days, which does not fare well with the intuition that jumps are rare events. From the estimated spike innovations or spike processes, plotted in the top panel of Figure 2, we see that there is clustering in the jump times and the assumption of a constant jump intensity does not hold.

In the SF-J model, the base-signal and spike process have same mean-reversion rate, $\alpha_x = 0.0077$. This is 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 the logarithmic spot price is very persistent. In the two-factor model TF-J, we impose the restriction $\alpha_y > \alpha_x$ to ensure that the estimated mean-reversion rate of the spike process is higher than that of the base-signal process. We see that the estimate of α obtained in the SF-J model is a weighted average

⁶For 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.

of α_x and α_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.

Table 3 contains the stochastic volatility models with IG marginals. Most of the variation in the data, which was explained by jumps in the SF-J and TF-J models (see the top panel of Figure 2), is now captured by the stochastic volatility, as shown in the top panel of Figure 3 and 4. As a result, the estimates for jump intensity become a lot smaller, ranging from 6-7 jumps per year for the one component models to about 1 jump per year in the two component models. The scarcity of jumps explains the high parameter uncertainty in jump size parameters μ_J and σ_J , as well as the mean-reversion rate α_Y which estimates the half-life of spikes to be about 1/3 of a day. From the left panels of Figure 3, we see that the estimated variance processes in the one-component models are similar for different spike specifications, except for the high peak in the end of 2009 which is less pronounced in the TF-SVJ^{IG} model.

The half-life of the variance process in the one-component models is about 3-4 days. In the three two-component IG models, we observe a slowly mean-reverting process with half-life more than one month, and a faster mean-reverting process with half-life less than two days. The autocorrelation function (ACF) of the superposition is simply a weighted sum of the ACF of its two components, with weights $w_i = \delta_i/(\delta_1 + \delta_2)$. In all three models, the slowly mean-reverting component has a higher weight, ranging from 0.55 to 0.56. The left panel of Figure 4 plots the estimated overall variance from model TF-S2VJ^{IG} as well as its two components, with $\sigma(t)^2 = \sigma_1(t)^2 + \sigma_2(t)^2$.

Next, we look at the models with log-OU stochastic volatility specifications in Table 4 and the right panels of Figure 3 and 4. As in the cases with IG volatility, the inclusion of log-OU volatility greatly reduces the need for spikes, and different spike specifications slightly affect the estimate for volatility. The half-life of the logarithmic variance is about 11 days in the one-component models. In the two-component models, the first log-component h_1 has half-life of two to three months, while h_2 has half-life of less than two days. In Figure 4, the overall variance from the log-OU model is the multiplication of the two components, $\exp(h_1(t))$ and $\exp(h_2(t))$, and a constant. While the first component is more persistent and resembles the overall volatility level in the sample, the fast-moving second component is more volatile and explains the occasional large changes.

	$\mathrm{SF}\text{-}\mathrm{SV}^{\mathrm{IG}}$	$\mathrm{SF} ext{-}\mathrm{SV}\mathrm{J}^{\mathrm{IG}}$	$\mathrm{TF} ext{-}\mathrm{SVJ}^{\mathrm{IG}}$	$\mathrm{SF} ext{-}\mathrm{S2V}^{\mathrm{IG}}$	$\mathrm{SF} ext{-}\mathrm{S2VJ}^{\mathrm{IG}}$	$\mathrm{TF} ext{-}\mathrm{S2VJ}^{\mathrm{IG}}$
α_x	0.0073	0.0065	0.0068	0.0085	0.0080	0.0088
	(0.0029)	(0.0029)	(0.0028)	(0.0031)	(0.0032)	(0.0034)
α_y			3.1706			2.1723
			(1.0716)			(0.9739)
λ_1	0.2134	0.1844	0.1702	0.0208	0.0206	0.0202
	(0.0256)	(0.0286)	(0.0299)	(0.0063)	(0.0062)	(0.0062)
λ_2				0.3603	0.3479	0.3713
				(0.0600)	(0.0529)	(0.0734)
δ_1	0.0235	0.0231	0.0235	0.0125	0.0125	0.0128
	(0.0018)	(0.0018)	(0.0020)	(0.0038)	(0.0033)	(0.0038)
δ_2				0.0102	0.0101	0.0103
				(0.0018)	(0.0019)	(0.0019)
γ	8.9667	9.3327	9.9108	5.5620	5.7398	6.7638
	(1.7092)	(1.8230)	(2.0208)	(1.9211)	(1.8809)	(2.1592)
μ_J		0.0046	-0.0312		0.0999	-0.3615
		(0.0643)	(0.0724)		(0.8237)	(0.7822)
σ_J		0.1306	0.1372		0.6923	0.5917
		(0.0936)	(0.1169)		(0.8333)	(0.4579)
λ_J		0.0182	0.0167		0.0040	0.0022
		(0.0162)	(0.0098)		(0.0055)	(0.0017)
LLMean	0.00	2.24	7.81	24.85	21.29	27.90
LLMedian	0.00	3.33	10.37	24.93	24.59	28.81
MarginalLL	0.00	1.76	6.31	23.77	22.60	22.37

Table 3: Parameter Estimates from IG-OU Volatility Models.

The table reports the posterior mean and posterior standard deviation of model parameters. LLmean/LLmedian is the log-likelihood evaluated at the posterior mean/median using 300,000 particles. The marginalLL is computed using equation (11). All three log-likelihood are reported in excess of model SF- SV^{IG} .

	$\mathrm{SF}\text{-}\mathrm{SV}^{\mathrm{Log}}$	$\mathrm{SF} ext{-}\mathrm{SVJ}^\mathrm{Log}$	$\mathrm{TF}\text{-}\mathrm{SVJ}^{\mathrm{Log}}$	$\mathrm{SF}\text{-}\mathrm{S2V}^{\mathrm{Log}}$	$\rm SF\text{-}S2VJ^{Log}$	$\mathrm{TF} ext{-}\mathrm{S2VJ}^{\mathrm{Log}}$
α_x	0.0070	0.0078	0.0077	0.0073	0.0073	0.0071
	(0.0021)	(0.0029)	(0.0016)	(0.0029)	(0.0029)	(0.0027)
$lpha_y$			2.1297 (0.4785)			2.1937 (0.8928)
_	0.0005	0.0627	0.0624	0.0104	0.0004	0.0076
α_{h1}	(0.0135)	(0.0037)	(0.0034)	(0.0104)	(0.0094)	(0.0076)
	(0.0100)	(0.0110)	(0.0100)	(0.0002)	(0.0000)	(0.0002)
α_{h2}				0.4498	0.4848	0.4062
				(0.1695)	(0.2015)	(0.1783)
σ_{h1}^2	0.2737	0.2742	0.2651	0.0395	0.0369	0.0289
	(0.0578)	(0.0543)	(0.0528)	(0.0265)	(0.0218)	(0.0195)
σ_{h2}^2				0.7950	0.8608	0.7326
102				(0.2508)	(0.3136)	(0.2516)
μ_h	-7.0472	-7.1098	-7.1014	-7.0647	-7.1245	-7.1515
	(0.1379)	(0.2162)	(0.1482)	(0.6388)	(0.6420)	(0.6409)
μ_J		0.0744	-0.3270		-0.8251	-0.3206
		(1.6863)	(0.3583)		(1.7866)	(0.3288)
σ_J		0.4919	0.4330		0.7678	0.4687
		(0.2920)	(0.2301)		(0.6031)	(0.2757)
λ_J		0.00031	0.0022		0.00098	0.0016
		(0.00020)	(0.0013)		(0.0012)	(0.0011)
LLMean	0.31	0.00	5.35	14.00	12.94	18.04
LLMedian	0.25	0.00	5.32	14.73	13.72	19.22
MarginalLL	1.51	0.00	4.60	14.24	11.95	16.52

Table 4: Parameter Estimates from log-OU Volatility Models.

The table reports the posterior mean and posterior standard deviation of model parameters. LLmean/LLmedian is the log-likelihood evaluated at the posterior mean/median using 300,000 particles. The marginalLL is computed using equation (11). All three log-likelihood are reported in excess of model SF-SVJ^{Log}.



Figure 2: Spikes. The left panels plot the estimated spike innovations in the single-factor models, computed from the mean of the posterior distribution $p(\xi_{1:T}J_{1:T}|Z_{1:T})$. The left panels plot the estimated spike processes in the two-factor models, computed from the mean of the posterior distribution $p(Y_{1:T}|Z_{1:T})$.



Figure 3: One-component Volatility. The figure depicts the estimated variance processes from the one-component volatility models, computed from the mean of the posterior distribution $p(\sigma^2(1 : T-1)|Z_{1:T})$.



Figure 4: Two-component Volatility. The figure depicts the estimated variance processes from the TF-S2VJ model and its two components, computed from the mean of the posterior distribution.

5.2 Model Evaluation

We conduct a comprehensive comparison between the models using Bayes factors computed from the marginal log-likelihood. Bayes factor do not require alternative models to be nested, so we can compare different volatility specifications such as log-OU and IG volatility. Another benefit is that it accounts for parameter uncertainty. For example, in Table 3, model TF-S2VJ^{IG} has higher likelihood than the simpler model SF-S2V^{IG} using log-likelihood evaluated at the point estimates of parameters, but it has lower marginal log-likelihood which takes parameter uncertainty into consideration. We compute twice the logarithm of the Bayes factor as it has the same scale as likelihood ratio test statistics. Let $\text{LBF}(M_1, M_2) = -2(\log p(M_2|Z) - \log p(M_1|Z))$, Kass and Raftery (1995) suggest the following scale for interpretation: if $\text{LBF}(M_1, M_2)$ is between 2 to 6, it is viewed as positive evidence against model M_2 ; 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_2 .

Model evaluation allows us to access the importance of each model-building element, in addition to pick out the best model. In the following subsections, we compare models in terms of spike specifications and volatility specifications, respectively.

5.2.1 Spike Specifications

Table 5 compares three possible cases regarding the price jumps: 1) no-jump; 2) one-factor jumpdiffusion; 3) two-factor process, where the base-signal and spike process have different mean reversion rate. The comparison is conducted in four groups; each group has the same volatility specification.

In the group with one-component IG volatility, there is strong evidence in favor of the twofactor model against either the no-jump specification or the one-factor jump-diffusion. As for the two-component IG volatility, no-jump specification is slightly preferred than the other two models. Two-component volatility specification allows for more complex dynamics in volatility, and this reduces the need for jumps in price.

With one-component log-OU volatility, we observe a different pattern: simply including jumps in the model is unfavorable compared to the no-jump case, while the two-factor specification with $\alpha_x \neq \alpha_y$ does improve model fit. The two-component log-OU group offers similar conclusion. This finding is in line with our intuition that jumps are faster decaying and this feature is important for model building.

Table 5. Dayes ractors negaring spike specifications.											
	$SF-SV^{IG}$	$SF-SVJ^{IG}$	$SF-SV^{Log}$ $SF-SVJ^{Log}$								
$SF-SVJ^{IG}$	3.5		$SF-SVJ^{Log}$ -3.0								
$\mathrm{TF} ext{-}\mathrm{SVJ}^{\mathrm{IG}}$	12.6	9.1	$TF-SVJ^{Log}$ 6.2 9.2								
	$SF-S2V^{IG}$	$SF-S2VJ^{IG}$	$SF-S2V^{Log}$ $SF-S2VJ^{Log}$								
$SF-S2VJ^{IG}$	-2.3		$SF-S2VJ^{Log}$ -4.6								
$TF-S2VJ^{IG}$	-2.8	-0.5	$TF-S2VJ^{Log}$ 4.6 9.1								

Table 5: Bayes Factors Regarding Spike Specifications.

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.

5.2.2Volatility Specifications

First, we compare the jump-driven IG-OU volatility to the log-OU continuous volatility, conditional on the same jump specification and number of components. From Table 6, models with the log-OU volatility specification are strongly favored over models with IG-OU volatility across different jump specifications. When we consider two-component models, the difference becomes smaller, indicating that allowing more than one component is particularly important for jump-driven volatility.

Table 6: Bayes Factors Regarding Volatility Specifications.									
$\mathrm{SF} ext{-}\mathrm{SV}^\mathrm{Log}$	$SF-SV^{IG}$ 43.8	$\rm SF\text{-}SVJ^{Log}$	$\mathrm{SF} ext{-}\mathrm{SVJ}^{\mathrm{IG}}$ 37.2	$\mathrm{TF} ext{-}\mathrm{SVJ}^\mathrm{Log}$	TF-SVJ ^{IG} 37.3				
	$\mathrm{SF}\text{-}\mathrm{S2V}^{\mathrm{IG}}$		$SF-S2VJ^{IG}$		$\mathrm{TF} ext{-}\mathrm{S2VJ}^{\mathrm{IG}}$				
$SF-S2V^{Log}$	21.7	$SF-S2VJ^{Log}$	19.5	$TF-S2VJ^{Log}$	29.1				

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.

Next, we focus on the number of volatility components. Table 7 suggests that the two-component models provide better fit to the data regardless of volatility type. Furthermore, we look at the empirical ACF of filtered $\sigma^2(t)$, which is computed using the posterior mean of parameters with 300,000 particles, and the theoretical ACF computed from the same parameter estimates. The top panels in Figure 5 compare the ACF of $\sigma^2(t)$ from the best performing one-component model, TF-SVJ^{Log}, to the best performing two-component model, TF-S2VJ^{Log}. The empirical ACF is decaying fast at small lags and slowly at larger lags, and this feature is better captured by the two-component model. This is further illustrated by the bottom panels of Figure 5, which plots the ACF of the two components separately.

Last, we compare the constant volatility models SF-J and TF-J to the stochastic volatility model with lowest marginal log-likelihood, SF-SV^{IG}. We find LBF(SF-SV^{IG}, SF-J) = 466.9 and $LBF(SF-SV^{IG}, TF-J) = 462.7$. The inclusion of stochastic volatility greatly reduces the need for price jumps and increases the marginal likelihood.

 $SF-SV^{IG}$ SF-SVJ^{IG} TF-SVJ^{IG} SF-S2V^{IG} SF-S2VJ^{IG} TF-S2VJ^{IG} 47.541.732.1 $SF-SV^{Log}$ $SF-SVJ^{Log}$ TF-SVJ^{Log} $SF-S2V^{Log}$ $SF-S2VJ^{Log}$ TF-S2VJ^{Log} 25.423.923.8

Table 7: Bayes Factors Regarding Volatility Components.

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.



Figure 5: The figure displays the theoretical and empirical autocorrelation functions for the variance processes. The theoretical ACF are computed using the posterior mean of the parameters governing the volatility processes. The empirical ACF is computed from filtered variance using the posterior mean of the parameters and 300,000 particles.

5.3 Model Validation

We conclude our empirical investigation by checking whether our models are able to produce price and return distributions that match the empirical distributions of the data. This is particular important for purposes of derivative pricing and risk management. For each estimated model, we simulate 5,000 artificial data sets of the same length as the deseasonalized logarithmic spot price Z(t). Then, we calculate the skewness/kurtosis of prices and returns for each simulated path, and obtain an empirical distribution of skewness/kurtosis implied by the model. Table 8 reports the 1st, 5th, 50th, 95th, and 99th percentiles of the model implied distributions of price and return skewness/kurtosis. The table also reports the values from the observed deseasonalized UK natural gas spot prices.

Table 8 shows that stochastic volatility is crucial for capturing the price skewness and kurtosis. All the models except SF-J and TF-J have distributions covering the observed negative price skewness

prctile	SF-J	$\mathrm{SF}\text{-}\mathrm{SV}^{\mathrm{IG}}$	$\mathrm{SF} ext{-}\mathrm{SV}^\mathrm{Log}$	$\mathrm{SF} ext{-}\mathrm{SVJ}^{\mathrm{IG}}$	$\rm SF\text{-}SVJ^{\rm Log}$	TF-J	$\mathrm{TF} ext{-}\mathrm{SVJ}^{\mathrm{IG}}$	$\mathrm{TF}\text{-}\mathrm{SVJ}^{\mathrm{Log}}$	$\mathrm{SF} ext{-}\mathrm{S2V}^{\mathrm{IG}}$	$\mathrm{SF} ext{-}\mathrm{S2V}^\mathrm{Log}$	SF-S2VJ ^{IG}	$\rm SF\text{-}S2VJ^{Log}$	$\mathrm{TF} ext{-}\mathrm{S2VJ}^{\mathrm{IG}}$	$\mathrm{TF} ext{-}\mathrm{S2VJ}^{\mathrm{Log}}$
	Price Skewness -0.78													
$1 \mathrm{th}$	-0.90	-1.35	-1.75	-1.22	-1.72	-0.73	-1.29	-1.68	-2.19	-2.14	-1.82	-2.32	-2.20	-2.10
5th	-0.61	-0.88	-1.13	-0.83	-1.09	-0.51	-0.87	-1.08	-1.45	-1.42	-1.19	-1.73	-1.40	-1.44
50th	0.01	0.01	-0.01	0.00	0.00	-0.02	0.01	-0.02	0.01	-0.02	0.10	-0.41	-0.02	-0.02
95th	0.59	0.90	1.11	0.84	1.12	0.48	0.92	1.06	1.40	1.47	1.30	1.24	1.39	1.45
99th	0.91	1.31	1.79	1.25	1.67	0.70	1.35	1.67	2.16	2.16	1.99	1.93	2.08	2.21
	Price Kurtosis 3.86													
$1 \mathrm{th}$	1.87	1.94	2.00	1.87	2.03	2.07	1.90	2.04	2.11	2.10	2.08	2.12	2.19	2.11
5th	2.06	2.15	2.23	2.08	2.26	2.23	2.14	2.28	2.40	2.39	2.35	2.45	2.47	2.42
50th	2.63	2.95	3.22	2.83	3.24	2.77	2.96	3.29	3.64	3.69	3.49	3.84	3.71	3.75
95th	3.62	4.83	6.09	4.57	6.05	3.65	4.81	5.92	7.99	7.63	6.62	7.72	7.60	7.60
99th	4.34	6.46	8.36	5.78	8.50	4.26	6.25	8.65	11.59	11.35	9.70	11.21	11.36	11.31
							Retu	rn Skewness -	0.37					
$1 \mathrm{th}$	-0.67	-1.84	-3.02	-1.57	-4.40	-0.79	-1.22	-2.81	-3.03	-3.06	-9.33	-22.85	-2.88	-3.13
5th	-0.49	-1.08	-1.41	-0.97	-1.86	-0.63	-0.77	-1.97	-1.55	-1.56	-5.37	-16.91	-2.05	-2.06
50th	-0.07	-0.00	-0.01	0.02	0.04	-0.21	-0.01	-0.54	0.00	-0.00	0.90	-2.54	-0.53	-0.36
95th	0.36	1.11	1.41	1.04	2.35	0.21	0.77	0.83	1.53	1.64	8.16	1.34	0.89	1.18
99th	0.52	1.84	2.73	1.70	6.14	0.40	1.36	1.79	2.89	3.43	11.83	4.12	1.79	2.48
	Return Kurtosis 21.48													
$1 \mathrm{th}$	7.56	8.35	7.28	8.76	7.31	7.31	8.58	8.86	8.63	7.79	13.39	8.88	10.02	9.02
5th	7.94	9.39	8.54	9.74	8.62	7.65	9.62	11.75	10.35	9.23	18.48	11.19	13.27	11.59
50th	9.04	14.17	14.34	13.64	15.24	8.70	13.32	43.45	17.89	16.29	66.53	62.72	48.81	34.74
95th	10.56	27.01	42.90	23.88	66.23	10.13	21.79	142.35	44.79	48.36	215.15	474.24	179.47	164.69
99th	11.41	42.08	85.26	34.20	169.97	10.91	30.31	210.04	80.41	90.41	336.19	752.56	264.20	254.64

Table 8: Distribution of skewness and kurtosis

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 posterior mean of parameters.

and small excess kurtosis. For both the SF-J and TF-J model, the probability that the observed value of skewness and kurtosis are realizations from the model implied distributions is less than 5%.

Turning our attention to the return distributions, Table 8 shows that return skewness can be captured by all the considered models. Once again, we find that the inclusion of stochastic volatility enables the model to produce more skewness. As for the kurtosis of the return series, Table 8 indicates that stochastic volatility is essential to produce a high enough level of kurtosis. It is also clear that the volatility specification impacts the distribution of return kurtosis. For the one-component models, only log-OU volatility specifications are able to fully match the level of kurtosis observed in the data. Form the kurtosis distribution with IG-OU specification, the probability of observing an outcome that is more extreme than the sample value from our data is only around 5%. Allowing two-component in volatility greatly increases return kurtosis. In the two-component models, both log-OU and IG-OU models are able to generate high levels of kurtosis. Spike specification also impacts return kurtosis. The one-factor jump-diffusion tends to generate too much kurtosis, placing the observed kurtosis near the 5th lower percentile.

To sum up, we find that inclusion of stochastic volatility 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 observed levels of kurtosis.

6 Discussion: Implications and Extensions

The implications of our findings in relation to pricing remain to be tested. One approach is to consider how the different model specifications impacts the forward prices. Specifically, we could compute the empirical risk premium, $RP(t) = F^{\text{observed}}(t, T_1, T_2) - F^{\mathbb{P}}(t, T_1, T_2)$, 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 $\mathbb{Q} = \mathbb{P}$, and then averaged over the delivery period. The forward prices depend on the factors X(t) and Y(t) as well as the stochastic volatility process $\sigma^2(t)$. One interesting implication of this result is 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 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 higher 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.

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

7 Conclusion

We propose a two-factor geometric model with stochastic volatility and jumps for the detrended and deseasonalized logarithmic UK gas spot price. We also provide a general estimation procedure based on the particle marginal Metropolis-Hastings sampler from Andrieu et al. (2010). To efficiently implement PMCMC, we develop a sequential Monte Carlo algorithm using a non-Markovian representation of the model and Rao-Blackwellization. Our estimation method is flexible and can be easily adapted to different volatility specifications, including pure jump-driven processes with superposition. This allows us to estimate and compare fourteen models with different jump or volatility specifications, and address important empirical questions.

Our application to UK natural gas spot prices showed that inclusion of stochastic volatility is more important than having jumps in the model. Failing to include stochastic volatility results in a severely overestimated jump intensity. With regards to jump specification, simple jump-diffusion is either unfavorable or only slightly favorable compared to the models with no jump. Two-factor models with separate mean-reversion rates justify the inclusion of jumps in the model even though the spike process only accounts for a small part of the variations in our data. We also find that the models with log-OU volatility outperforms the models with IG-OU, and the two-component volatility specifications are preferred over the one-component counterparts. Our results suggest that the two-factor model with two-component log-OU volatility, TF-S2V^{Log}, is best suited for modeling UK gas spot prices.

Appendix

Priors

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 the following univariate 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 the base-signal. For stochastic volatility with inverse Gaussian marginals, we specify $\delta_1, \delta_2, \gamma \sim G(1, \sqrt{50}), \lambda_1 \sim G(1, 0.2), \text{ and } \lambda_2 \sim G(1, 1), \text{ so the first component has a slower mean-reversion rate. For logarithmic volatility, <math>\alpha_{h2} \sim G(1, 1), \mu_h \sim N(-7, 2)$ and $\sigma_{h1}^2, \sigma_{h2}^2 \sim G(1, 2)$. α_{h1} is very close to zero so we estimate $\log(\alpha_{h1})$ instead, and the prior is set to $\log(\alpha_{h1}) \sim N(-3, 5)$. For models with constant volatility, we choose $\sigma^2 \sim G(1, 2)$. Finally, the priors for jump parameters are $\mu_J \sim N(0, 2), \sigma_J \sim G(1.5, 0.5), \text{ and } \lambda_J \sim G(1, 10)$.

Model Diagnostics

We present the model diagnostic plots for TF-S2VJ^{IG} and TF-S2VJ^{Log} in Figure 6 and 7. The plots for other models are similar and omitted here. 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. Visual inspection indicates convergence, although the mixing of some parameters is less satisfactory, for example for σ_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 use different scales for the densities: ticks on the left axis are the density for posterior, while ticks on the right axis denote the density for prior. For most of the parameters, prior information is negligible compared with posterior. One exception is σ_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 σ_j is informative about the posteriors.



Figure 6: Diagnostic Plots for the TF-S2VJ model with IG-OU volatility. Left panels are trace plots of parameters. In the right panels, bars are posterior densities, while lines indicate priors.



Figure 7: Diagnostic Plots for the TF-S2VJ model with log-OU volatility. Left panels are trace plots of parameters. In the right panels, bars are posterior densities, while lines indicate priors.

References

- Alizadeh, S., M. W. Brandt, and F. X. Diebold (2002). Range-based estimation of stochastic volatility models. The Journal of Finance 57(3), 1047–1091.
- Andrieu, C., A. Doucet, and R. Holenstein (2010). Particle Markov chain Monte Carlo methods. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 72(3), 269–342.
- Andrieu, C. and J. Thoms (2008). A tutorial on adaptive MCMC. *Statistics and Computing* 18(4), 343–373.
- Barndorff-Nielsen, O. E., F. E. Benth, and A. E. D. Veraart (2013). Modelling energy spot prices by volatility modulated Lévy-driven volterra processes. *Bernoulli* 19(3), 803–845.
- Barndorff-Nielsen, O. E. and N. Shephard (2001a). Non-gaussian ornstein-uhlenbeck-based models and some of their uses in financial economics. *Journal Of The Royal Statistical Society Series* B 63(2), 167–241.
- Barndorff-Nielsen, O. E. and N. Shephard (2001b). Normal modified stable processes. Theory of Probability and Mathematical Statistics 65, 1–19.
- Benth, F. E. (2011). The stochastic volatility models of Barndorff-Nielsen and Shephard in commodity markets. *Mathematical Finance* 4, 595–625.
- Benth, F. E., J. S. Benth, and S. Koekebakker (2008). Statistical Modeling of Electricity and Related Markets. Advanced Series on Statistical Science and Applied Probability. World Scientific.
- Benth, F. E., L. Ekeland, R. Hauge, and B. F. Nielsen (2003). A note on arbitrage-free pricing of forward contracts in energy markets. *Applied Mathematical Finance* 10, 325–336.
- Benth, F. E., R. Kiesel, and A. Nazarova (2012). A critical empirical study of three electricity spot price models. *Energy Economics* 34, 1589–1616.
- Benth, F. E. and J. Šaltytė Benth (2004). The normal inverse Gaussian distribution and spot price modelling in energy markets. International Journal of Theoretical and Applied Finance 07(02), 177–192.
- Benth, F. E. and L. Vos (2013). Cross-commodity spot price modeling with stochastic volatility and leverage for energy markets. Advances in Applied Probability 45, 545–571.
- Cartea, A. and M. Figueroa (2005). Pricing in electricity markets: a mean reverting jump diffusion model with seasonality. *Applied Mathematical Finance* 12(4), 313–335.
- Chernov, M., A. Ronald Gallant, E. Ghysels, and G. Tauchen (2003). Alternative models for stock price dynamics. *Journal of Econometrics* 116(1-2), 225–257.

- Douc, R. and O. Cappe (2005, Sept). Comparison of resampling schemes for particle filtering. In Image and Signal Processing and Analysis, 2005. ISPA 2005. Proceedings of the 4th International Symposium on, pp. 64–69.
- Doucet, A., N. d. Freitas, K. P. Murphy, and S. J. Russell (2000). Rao-Blackwellised particle filtering for dynamic Bayesian networks. In *Proceedings of the 16th Conference on Uncertainty in Artificial Intelligence*, UAI '00, San Francisco, CA, USA, pp. 176–183. Morgan Kaufmann Publishers Inc.
- Durham, G. B. (2006). Monte carlo methods for estimating, smoothing, and filtering one- and two-factor stochastic volatility models. *Journal of Econometrics* 133(1), 273 305.
- Eraker, B., M. Johannes, and N. Polson (2003, 06). The impact of jumps in volatility and returns. *Journal of Finance* 58(3), 1269–1300.
- Eydeland, A. and K. Wolyniec (2003). Energy and Power Risk Management, New Developments in Modeling, Pricing and Hedging. John Wiley.
- Fileccia, G. and C. Sgarra (2015a). A double correlated three factor model for a crude oil market. Available at SSRN: http://dx.doi.org/10.2139/ssrn.2568563.
- Fileccia, G. and C. Sgarra (2015b). Historical and risk-neutral estimation in a two factors stochastic volatility model for oil markets. *Forthcoming in International Journal of Computational Economics and Econometrics*.
- Flury, T. and N. Shephard (2011, October). Bayesian inference based only on simulated likelihood: Particle filter analysis of dynamic economic models. *Econometric Theory* 27(05), 933–956.
- Gallant, A. R. and G. Tauchen (1996). Which moments to match? *Econometric Theory* 12(4), pp. 657–681.
- Gander, M. P. S. and D. A. Stephens (2007). Stochastic volatility modelling with general marginal distributions: Inference, prediction and model selection. *Journal of Statistical Planning and Inference* 137, 3068–3081.
- Gelman, A., G. Roberts, and W. Gilks (1996). Efficient Metropolis jumping hules. Bayesian statistics 5, 599–608.
- Geman, H. (2005). Commodities and Commodity Derivatives. John Wiley.
- Green, R. and M. Nossman (2008). Markov chain Monte Carlo estimation of a multi-factor jump diffusion model for power prices. *The Journal of Energy Markets* 1(4), 65–90.
- Griffin, J. and M. Steel (2006). Inference with non-Gaussian Ornstein-Uhlenbeck processes for stochastic volatility. *Journal of Econometrics* 134, 605–644.

- Haario, H., E. Saksman, and J. Tamminen (2001). An adaptive Metropolis algorithm. Bernoulli 7(2), 223–242.
- Jacquier, E., N. G. Polson, and P. E. Rossi (1994, October). Bayesian analysis of stochastic volatility models. Journal of Business & Economic Statistics 12(4), 371–89.
- Johannes, M. S., N. G. Polson, and J. R. Stroud (2009, July). Optimal filtering of jump diffusions: Extracting latent states from asset prices. *Review of Financial Studies* 22(7), 2559–2599.
- Kalli, M. and J. Griffin (2015). Flexible modeling of dependence in volatility processes. Journal of Business & Economic Statistics 33(1), 102–113.
- Kass, R. E. and A. E. Raftery (1995). Bayes factors. Journal of the American Statistical Association 90(430), pp. 773–795.
- Lucia, J. and E. Schwartz (2002). Electricity prices and power derivatives: evidence from the Nordic power exchange. *Review of Derivatives Research* 5(1), 5–50.
- Newton, M. A. and A. E. Raftery (1994). Approximate Bayesian inference with the weighted likelihood bootstrap. Journal of the Royal Statistical Society. Series B (Methodological) 56(1), pp. 3–48.
- Pitt, M. K., R. dos Santos Silva, P. Giordani, and R. Kohn (2012). On some properties of Markov chain Monte Carlo simulation methods based on the particle filter. *Journal of Econometrics* 171(2), 134 – 151. Bayesian Models, Methods and Applications.
- Pitt, M. K. and N. Shephard (1999, June). Filtering via simulation: Auxiliary particle filters. *Journal* of the American Statistical Association 94 (446), 590–599.
- Roberts, G. O., O. Papaspiliopoulos, and P. Dellaportas (2004). Bayesian inference for non-Gaussian Ornstein-Uhlenbeck stochastic volatility processes. *Journal of the Royal Statistical Society Series* B 66(2), 369–393.
- Schwartz, E. (1997). The stochastic behaviour of commodity prices: Implications for valuation and hedging. *The Journal of Finance* 52(3), 923–973.
- Shephard, N. (Ed.) (2005). Stochastic Volatility: Selected Readings. Oxford University Press.
- Zhang, S. and X. Zhang (2008). Exact simulation of ig-ou processes. Methodology and Computing in Applied Probability 10(3), 337–355.

Research Papers 2013



- 2015-29: Laurent Callot and Johannes Tang Kristensen: Regularized Estimation of Structural Instability in Factor Models: The US Macroeconomy and the Great Moderation
- 2015-30: Davide Delle Monache, Stefano Grassi and Paolo Santucci de Magistris: Testing for Level Shifts in Fractionally Integrated Processes: a State Space Approach
- 2015-31: Matias D. Cattaneo, Michael Jansson and Whitney K. Newey: Treatment Effects with Many Covariates and Heteroskedasticity
- 2015-32: Jean-Guy Simonato and Lars Stentoft: Which pricing approach for options under GARCH with non-normal innovations?
- 2015-33: Nina Munkholt Jakobsen and Michael Sørensen: Efficient Estimation for Diffusions Sampled at High Frequency Over a Fixed Time Interval
- 2015-34: Wei Wei and Denis Pelletier: A Jump-Diffusion Model with Stochastic Volatility and Durations
- 2015-35: Yunus Emre Ergemen and Carlos Velasco: Estimation of Fractionally Integrated Panels with Fixed Effects and Cross-Section Dependence
- 2015-36: Markku Lanne and Henri Nyberg: Nonlinear dynamic interrelationships between real activity and stock returns
- 2015-37: Markku Lanne and Jani Luoto: Estimation of DSGE Models under Diffuse Priors and Data-Driven Identification Constraints
- 2015-38: Lorenzo Boldrini and Eric Hillebrand: Supervision in Factor Models Using a Large Number of Predictors
- 2015-39: Lorenzo Boldrini and Eric Hillebrand: The Forecasting Power of the Yield Curve, a Supervised Factor Model Approach
- 2015-40: Lorenzo Boldrini: Forecasting the Global Mean Sea Level, a Continuous-Time State-Space Approach
- 2015:41: Yunus Emre Ergemen and Abderrahim Taamouti: Parametric Portfolio Policies with Common Volatility Dynamics
- 2015-42: Mikkel Bennedsen: Rough electricity: a new fractal multi-factor model of electricity spot prices
- 2015-43: Mikkel Bennedsen, Asger Lunde and Mikko S. Pakkanen: Hybrid scheme for Brownian semistationary processes
- 2015-44: Jonas Nygaard Eriksen: Expected Business Conditions and Bond Risk Premia
- 2015-45: Kim Christensen, Mark Podolskij, Nopporn Thamrongrat and Bezirgen Veliyev: Inference from high-frequency data: A subsampling approach
- 2015-46: Asger Lunde, Anne Floor Brix and Wei Wei: A Generalized Schwartz Model for Energy Spot Prices - Estimation using a Particle MCMC Method