Energy Economics 46 (2014) 422–434

Contents lists available at ScienceDirect

Energy Economics

journal homepage: www.elsevier.com/locate/eneco

Tail risk in energy portfolios[☆]

Carlos González-Pedraz [a,*,1], Manuel Moreno [b], Juan Ignacio Peña [a]

[a] Department of Business, Universidad Carlos III de Madrid, Getafe 28903, Spain
[b] Department of Economic Analysis and Finance, Universidad de Castilla La Mancha, Toledo 45071, Spain

ARTICLE INFO

Article history:
Received 17 January 2013
Received in revised form 9 March 2014
Accepted 10 May 2014
Available online 6 June 2014

JEL classifications:
C46
G11
Q41

Keywords:
Asymmetric DCC
Multivariate generalized hyperbolic distributions
Tail risk
Skewness
Risk measure backtests

ABSTRACT

This article analyzes the tail behavior of energy price risk using a multivariate approach, in which the exposure to energy markets is given by a portfolio of oil, gas, coal, and electricity. To accommodate various dependence and tail decay patterns, this study models energy returns using different generalized hyperbolic conditional distributions and time-varying conditional mean and covariance. Employing daily energy futures data from August 2005 to March 2012, the authors recursively estimate the models and evaluate tail risk measures for the portfolio's profit-and-loss distribution for long and short positions at various horizons and confidence levels. Both in-sample and out-of-sample analyses applied to different energy portfolios show the importance of heavy tails and positive asymmetry in the distribution of energy risk factors. Thus, tail risk measures for energy portfolios based on standard methods (e.g. normality, constant covariance matrix) and on models with exponential tail decay underestimate actual tail risk, especially for short positions and short time horizons.

1. Introduction

The growth of energy markets has been sustained by continued deregulation processes, which have encouraged the separation of the formerly integrated value chains. This process has increased market risk exposures at every stage of the chain, including the purchase and sale of fuels, electricity generation, and obtaining gas or electricity for retail supply. In addition to the physical resource holders, financial players, such as banks and hedge funds, increasingly participate in energy markets to satisfy their customers' demands to gain or hedge energy risk exposure, as well as to trade on their own behalf. In this context, energy-related companies and financial players experience greater exposures to energy price risk, which has particular characteristics that make it different from other market risks and requires clearer explication.

In this article, we analyze the energy price risks from a multivariate perspective. We study the aggregate tail risk of different linear energy portfolios using an asset-level approach. Accordingly, we can propose a multivariate model for the vector of energy risk factors; using the portfolio exposures to each factor, we in turn can calculate the aggregate tail behavior of the portfolio. Next, we compute the corresponding portfolio risk measures and evaluate the extent to which the tail pattern of the model is important in practice.

With this asset-level approach, we can capture the entire structure of energy risk factors in a portfolio and their multivariate behavior. In particular, we seek to include all the stylized features of deregulated energy markets, specifically: large volatility, price spikes, time-varying correlation, dependence in the extremes, and mean-reversion patterns (e.g., Benth et al., 2008, Cartea and Figueroa, 2005, Escribano et al., 2011, Huisman and Mahieu, 2003, Pirrong, 2012, and Routledge et al., 2001).

For this purpose, we consider an econometric specification with time-varying conditional means, volatilities, and correlations, in which the innovation vector follows a multivariate generalized hyperbolic (GH) distribution.[2] In this way, we extend previous theoretical and empirical studies that employ some distributions of the GH family to the risk analysis of energy assets (see Benth and Šaltytė Benth, 2004,

☆ The authors are grateful to Álvaro Cartea, Wenbo Hu, Pablo Villaplana, and the seminar participants at the 5th AEEE-IAEE Conference on Energy Economics and Universidad Carlos III, for their helpful comments and suggestions. M. Moreno acknowledges the financial support provided by grants P08-SEJ-03917 and JCCM PPII11-0290-0305.
* Corresponding author at: Department of Business, Universidad Carlos III de Madrid, C/ Madrid, 126 28903 Getafe, Madrid, Spain.
E-mail addresses: cugonzal@emp.uc3m.es (C. González-Pedraz), manuel.moreno@uclm.es (M. Moreno), ypenya@eco.uc3m.es (J.I. Peña).
1 Tel.: +34 91 624 9772.

2 With different parametrizations, this family of distributions has been applied in financial modeling of univariate and multivariate problems; see for example Aas and Haff (2006), Bingham and Kiesel (2002), Hu and Kercheval (2010), Mcneil et al. (2005), and Prause (1999).

http://dx.doi.org/10.1016/j.eneco.2014.05.004
0140-9883/

Boerger et al., 2009, Eberlein and Stahl, 2003, Giot and Laurent, 2003, and Weron, 2006).

We apply our multivariate GH specification to model a vector of crude oil, natural gas, coal, and electricity returns. Using daily data from August 2005 to March 2012, we evaluate the tail risk of the profit-and-loss (P&L) distribution of different energy portfolios.

We address the analysis of the aggregate tail risk by calculating two risk measures, the value at risk (VaR) and the expected shortfall (ES), for long and short trading positions in the energy portfolios, and for different day horizons and confidence levels. Whereas most equity risk studies have focused on the left tail of long positions, the presence of positive jumps in the data generating process of energy commodities suggests that the analysis of the right tail of the possibly asymmetric P&L distribution (short positions) could be relevant for traders who are worried about increases in energy prices.

Finally, using different backtest procedures, we monitor, for the out-of-sample period, the performance of the risk measure estimates that correspond to the GH models. Specifically, we propose a new backtest procedure that employs the superior predictive ability (SPA) test of Hansen (2005), along with a metric function based on the ES backtesting measures of Embrechts et al. (2005), to compare the performance of the whole set of alternative models.

The main empirical results of the article are as follows. First, we find evidence of time-varying evolution in the conditional correlations between energy markets. In addition, results for the recursive estimations over the out-of-sample period reveal the presence of fat tails and positive skewness in the multivariate distribution of energy risk factors. The out-of-sample evaluation of the risk measure forecasts favors these findings. In summary, results show that tail risk measures for energy portfolios based on standard methods (e.g. normality, constant covariance matrix) strongly underestimate actual tail risk. The heavy-tail models behave much better than alternative versions, with regard to the tail risk of short positions and short time horizons. According to the results of the SPA backtest, models with exponential tail decay yield inferior tail estimates for short portfolio positions compared with models with polynomial tail decay, especially for the far tail ($\alpha < 1\%$) of utility portfolios at short horizons.

The extent of the underestimations of the tail risk of the portfolio loss distribution depends on whether we are analyzing short or long positions in the energy portfolio, the type of portfolio, the horizon, and how far out in the tail the risk is being analyzed.

In Section 2, we begin by presenting the energy portfolios and their profit-and-loss function. Section 3 introduces the econometric specifications for the energy returns vector. With Section 4, we characterize the conditional risk measures. Next, in Section 5, we provide a description of the data and the results from the empirical estimations. After we report the results of the risk measure forecast and analyze their out-of-sample performance, in Section 6; we conclude in Section 7. A final appendix provides some technical details.

2. Portfolios of energy commodities

We approximate a given exposure to energy price risk using a corresponding portfolio of energy futures. Thus, changes in the energy price risk factors can be mapped linearly to changes in the value of the energy futures portfolio.[3] For example, a linear portfolio could represent directly the energy futures positions of an institutional investor or the energy price exposure of an electricity producer with fuel-fired power plants. In this article, we consider four energy commodities: crude oil, natural gas, coal, and electricity, identified by subscripts i equal to 1, 2, 3, and 4, respectively. These four commodities substantially represent any general exposure to energy price risk.[4]

Let $F_{i,t}$ denote the settlement price at time t of an energy futures contract i. We then assume that the value W_t of a given energy portfolio is determined by a linear combination of futures $F_{i,t}$, such that

$$W_t = \sum_{i=1}^{4} q_{i,t} F_{i,t} = \sum_{i=1}^{4} w_{i,t} = \boldsymbol{w}_t^{\prime} \iota_4, \tag{1}$$

where the quantities $q_{i,t}$ define the size and sign of the exposure to the energy commodity i; $w_t = (w_{1,t},\ldots,w_{4,t})'(q_{1,t}F_{1,t},\ldots,q_{4,t}F_{4,t})'$ defines the portfolio weights in dollars of each energy commodity; and ι_4 is a 4×1 vector of ones.

Thus, the h-period return (in dollars) on an energy portfolio at time t is given by

$$\Delta W_t(h) = W_t - W_{t-h} = R_t(h) W_{t-h} = w_{t-h}'(exp(\boldsymbol{r}_t(h)) - \iota_4), \tag{2}$$

where $R_t(h)$ is the h-period net return on the portfolio, $\boldsymbol{r}_t(h) = \sum_{k=0}^{h-1} r_{t-k}$ is the 4×1 vector of h-period log-returns, and $\boldsymbol{r}_t$ is the 4×1 one-period log-return vector at time t with the i-th component $r_{i,t} = log(F_{i,t}/F_{i,t-1})$. The energy log-returns $r_{i,t}$ constitute the vector of risk factors.

According to Eq. (2), the portfolio's profit-and-loss (P&L) distribution is determined by the multivariate density model of energy risk factors $\boldsymbol{r}_t$ and the positions in the energy commodities $\boldsymbol{w}_t$. The multivariate model of risk factors describes the joint behavior of the four commodities. The different positions in energy futures define the size and sign of the exposure to each energy commodity, mapping the multivariate model onto a specific P&L distribution. In this article, we consider four predefined portfolios: two related to power utilities and two others more related to financial players.

As representative portfolios of electricity producers, we include the energy portfolio of a utility with a diversified mix of generation that operates in the Pennsylvania–Jersey–Maryland (PJM) Interconnection and a linear portfolio corresponding to a gas-fired power plant. In addition, we account for two typical portfolios of financial players that seek exposure to energy commodities: an equally weighted portfolio and the minimum variance portfolio. In Appendix A, we provide more details about the construction of these portfolios.

3. A dynamic multivariate GH model for energy returns

Using $\boldsymbol{r}_t = (r_{1,t}, \ldots, r_{4,t})'$ as the vector of the four energy assets log-returns at time t, we assume that the data generating process for $\{\boldsymbol{r}_t : t = 1, \ldots, T\}$ is given by

$$\boldsymbol{r}_t = m_t + \varepsilon_t, \text{ and} \tag{3}$$

$$\varepsilon_t = \boldsymbol{H}_t^{1/2} \boldsymbol{x}_t, \tag{4}$$

where m_t and ε_t are the 4×1 vectors of conditional means and unexpected returns, respectively; $\boldsymbol{H}_t^{1/2}$ is the 4×4 Cholesky factor of the time-varying 4×4 covariance matrix $\boldsymbol{H}_t$, such that $\boldsymbol{H}_t = \boldsymbol{H}_t^{1/2}(H_t^{1/2})'$; and $\boldsymbol{x}_t$ is the 4×1 vector of independent innovations, which follows a four-variate generalized hyperbolic (GH) distribution with zero mean and an identity covariance matrix.

To capture the possible presence of serial correlation in energy returns, we consider a diagonal vector autoregressive (VAR) process

[3] A portfolio of futures contracts can also be considered a first-order approximation of more general energy asset portfolios with non-linear payoffs (Cartea and González-Pedraz, 2012; Tseng and Barz, 2002).

[4] Oil constitutes 33% of the world's total primary energy supply, followed by coal with a share of 27% and natural gas with 21%. In addition, coal, natural gas, and oil represent 41%, 21%, and 5%, respectively, of the world's electricity generation (International Energy Agency, 2011).

with up to 5 lags for the vector of returns.[5] Thus, the conditional mean vector is given by

$$\boldsymbol{m}_t = \mathbb{E}(\boldsymbol{r}_t|\mathcal{F}_{t-1}) = \boldsymbol{m}_0 + \sum_{j=1}^{5} \boldsymbol{\Phi}_j \boldsymbol{r}_{t-j}, \tag{5}$$

where $\boldsymbol{m}_0$ is a constant 4×1 vector, $\boldsymbol{\Phi}_j$ is a 4×4 diagonal matrix, and the expectation is conditional on the history of the process up to time $t-1$, that is, $\mathcal{F}_{t-1} = \sigma(\{\boldsymbol{r}_s : s \le t-1\})$.

At the same time, the conditional covariance matrix $\boldsymbol{H}_t$ can be decomposed as follows

$$\boldsymbol{H}_t = \mathrm{cov}(\boldsymbol{r}_t|\mathcal{F}_{t-1}) = \boldsymbol{D}_t \boldsymbol{P}_t \boldsymbol{D}_t, \tag{6}$$

where $\boldsymbol{D}_t$ is the 4×4 diagonal matrix composed by the conditional volatilities of $\boldsymbol{r}_t$, and $\boldsymbol{P}_t$ is the 4×4 conditional correlation matrix. We want to capture possible persistence and asymmetry in conditional variances and correlations. For that purpose, we assume univariate asymmetric GARCH(1,1) processes for the conditional variances and a modified version of the asymmetric dynamic conditional correlation (ADCC) model of Cappiello et al. (2006) for the time-varying correlation matrix. That is, the elements of the diagonal volatility matrix $\boldsymbol{D}_t$, $\sqrt{h_{i,t}}$, satisfy

$$h_{i,t} = \alpha_{0,i} + \alpha_{1,i}\varepsilon_{i,t-1} + \alpha_{1,i}^{-}\varepsilon_{i,t-1}\mathbb{1}_{\{\varepsilon_{i,t-1}\le 0\}} + \alpha_{2,i}h_{i,t-1}, \quad i = 1,\dots,4, \tag{7}$$

where $\alpha_{0,i} > 0$, $\alpha_{1,i} \ge 0$, $\alpha_{1,i} + \alpha_{1,i}^{-} \ge 0$, $\alpha_{2,i} > 0$, and $\alpha_{1,i} + \alpha_{1,i}^{-}/2 + \alpha_{2,i} < 1$, which guarantees that the process is positive and covariance stationary. The dynamics of the correlation matrix in our version of the ADCC model is given by

$$\boldsymbol{P}_t = \mathrm{diag}(\boldsymbol{Q}_t)^{-1/2}\boldsymbol{Q}_t \mathrm{diag}(\boldsymbol{Q}_t)^{-1/2}, \tag{8}$$

$$\boldsymbol{Q}_t = \left[(1-\delta_1-\delta_2)\overline{\boldsymbol{Q}} - \delta_1^{+}\overline{\boldsymbol{N}}\right] + \delta_1 \boldsymbol{u}_{t-1}\boldsymbol{u}_{t-1}' + \delta_1^{+}\boldsymbol{n}_{t-1}\boldsymbol{n}_{t-1}' + \delta_2 \boldsymbol{Q}_{t-1}, \tag{9}$$

where δ_1, δ_1^{+}, $\delta_2 \ge 0$; $\boldsymbol{u}_t = \boldsymbol{D}_t^{-1}\varepsilon_t$ is the 4×1 vector of standardized residuals; $\boldsymbol{n}_t = \boldsymbol{u}_t \mathbb{1}_{\{\boldsymbol{u}_t \le 0\}}$; and $\overline{\boldsymbol{Q}} = \mathbb{E}(\boldsymbol{u}_{t-1}\boldsymbol{u}_{t-1}')$ and $\overline{\boldsymbol{N}} = \mathbb{E}(\boldsymbol{n}_{t-1}\boldsymbol{n}_{t-1}')$ are the unconditional covariance matrices of $\boldsymbol{u}_t$ and $\boldsymbol{n}_t$. A sufficient condition for $\boldsymbol{Q}_t$ to be positive definite is that $\delta_1 + \delta_2 + \overline{\eta}\delta_1^{+} < 1$, where $\overline{\eta}$ is the maximum eigenvalue of $\overline{\boldsymbol{Q}}^{-1/2}\overline{\boldsymbol{N}}\,\overline{\boldsymbol{Q}}^{-1/2}$. With this specification, we investigate, in the conditional correlation, the presence of asymmetric responses to positive shocks.

Motivated by the presence of jumps and spikes in energy prices, we employ multivariate GH distributions to model the conditional distribution of the vector of innovations $\boldsymbol{x}_t$. These GH distributions are flexible enough to accommodate different tail behaviors and types of asymmetry (e.g., thin or heavy tails, symmetric or positive/negative skewness). The GH family can be obtained using the following normal mean–variance mixture representation (see Mcneil et al., 2005):

$$\boldsymbol{x}_t \overset{\text{dist.}}{=} \boldsymbol{\mu} + \omega_t\boldsymbol{\gamma} + \omega_t^{1/2}\boldsymbol{A}\boldsymbol{z}_t, \quad \boldsymbol{z}_t \sim \mathrm{N}_4(\boldsymbol{0}, \boldsymbol{I}_4), \tag{10}$$

where $\boldsymbol{\mu}$ and $\boldsymbol{\gamma}$ are the 4×1 location and skewness parameter vectors, respectively, and $\boldsymbol{\Sigma} = \boldsymbol{A}\boldsymbol{A}'$ is the 4×4 dispersion matrix. The random vector $\boldsymbol{z}_t$ follows a four-variate Gaussian distribution with zero mean and identity covariance, and $\omega_t \ge 0$ is a non-negative random variable independent of $\boldsymbol{z}_t$. The mixing random variable ω_t can be understood as a shock that affects the covariance of energy assets, due to the arrival of new information in the markets (e.g., shortages in future supply, unexpected increases in demand).[6]

In the case of GH distributions, the mixing random variable ω_t follows a generalized inverse Gaussian (GIG) distribution, $\omega_t \sim N^{-}(\lambda, \chi, \psi)$. The very flexible GIG distribution includes as special boundary cases the gamma and inverse gamma distributions. Thus, for certain values of the parameters λ, χ, and ψ and the skewness vector $\boldsymbol{\gamma}$, we can obtain different cases of GH distributions. We consider five particular cases of multivariate GH distributions: the normal inverse Gaussian (NIG) distribution (i.e., ω_t follows an inverse Gaussian distribution, which corresponds to the case $\lambda = -1/2$), the variance-gamma (VG) distribution (i.e., ω_t follows a gamma distribution, corresponding to the parameters $\lambda > 0$ and $\chi = 0$), the skewed t (skT) distribution (i.e., ω_t follows an inverse gamma distribution, corresponding to the case in which $\lambda < 0$, $\chi = -2\lambda$, and $\psi = 0$), Student's t (T) distribution (with same mixing distribution as the skewed t but with $\boldsymbol{\gamma} = 0$), and the Gaussian (G) distribution (for which $\boldsymbol{\gamma} = 0$ and $\omega_t = 1$). In Appendix B, we provide more details about the density functions of these distributions.

Using the normal mean–variance specification of Eq. (10), we can compute the mean and covariance values for the vector of innovations $\boldsymbol{x}_t$ for each GH distribution, provided that the mean and variance of ω_t exist and are finite. If we assume that $\boldsymbol{x}_t$ has zero mean and unit covariance, $\mathbb{E}(\boldsymbol{x}_t) = 0$ and $\mathrm{cov}(\boldsymbol{x}_t) = \boldsymbol{I}_4$, then the location vector and dispersion matrix of the conditional distribution must satisfy the following conditions:

$$\boldsymbol{\mu} = -\mathbb{E}(\omega_t)\boldsymbol{\gamma}, \text{ and} \tag{11}$$

$$\boldsymbol{\Sigma} = \left(\boldsymbol{I}_4 - \mathrm{var}(\omega_t)\gamma\gamma'\right)/\mathbb{E}(\omega_t). \tag{12}$$

The proposed GH distributions have the advantages of exhibiting different tail patterns (Bibby and Sørensen, 2003). On the one hand, the tails of the NIG and VG distributions decay exponentially, such that their probability density functions behave, when $\boldsymbol{x}_t \to \pm\infty$, proportionally to an exponential function. This pattern is intermediate between the behavior of the Gaussian distribution, which decays more rapidly, and other, more extreme, polynomial decays. For this reason, NIG and VG distributions are sometimes referred to as semi-heavy tailed. Furthermore, when the i-th element of the asymmetry vector $\boldsymbol{\gamma}$ differs from zero (i.e., $\gamma_i \ne 0$), the two tails of the i-th innovation $x_{i,t}$ behave differently (the right tail is heavier when $\gamma_i > 0$, whereas the left tail is heavier when $\gamma_i < 0$). The tails of the T distribution instead are symmetric and behave as polynomials, such that they decay slower than those of the NIG and VG distributions. Finally, the skT distribution offers the special property of possessing, for each component of the vector of innovations, one heavy (polynomial decay) and one semi-heavy (exponential decay) tail. Thus, when $\gamma_i > 0$, the right tail ($x_{i,t} \to \infty$) is the heaviest, while the left tail ($x_{i,t} \to -\infty$) decays exponentially; these roles switch for $\gamma_i < 0$.

4. Conditional risk measures

Measuring conditional risk is a natural and direct way to analyze the tail behavior of energy portfolios. In our approach, we study both long and short positions in the energy portfolios. Thus, we focus on the two tails of the P&L distribution. For short positions, the portfolio holder loses money when the portfolio value increases, so we attend to the right side of the distribution. For long positions, we focus on the left tail.

We consider two measures of risk: the value at risk (VaR) and the expected shortfall (ES). The VaR is widely used in the financial industry to monitor risk exposures for regulatory purposes and to establish trading constraints in investment decisions. In the energy industry, especially for producers, VaR is becoming more popular, with increasing relevance for corporate decisions. For example, VaR provides insights to determine hedging policies or, in the case of utilities, to obtain an optimal selection in the generation mix. Formally, for a certain horizon h and confidence level α, the VaR is defined as the α-quantile of the

[5] In a previous empirical analysis, we determine the number of lags in the model, first, taking into account those lags that are significant and help to reduce the presence of autocorrelation in the asset returns; then, considering the values of the Bayesian Information Criteria to select among the different alternative models.

[6] Conditioned on ω_t, the vector of innovations $\boldsymbol{x}_t$ is normally distributed, such that $\boldsymbol{x}_t|\omega_t \sim \mathrm{N}_4(\boldsymbol{\mu} + \omega_t\boldsymbol{\gamma}, w_t\boldsymbol{\Sigma})$.

conditional distribution of portfolio changes $\Delta W_t(h)$. That is, the probability of incurring losses greater than a certain threshold value, called the VaR, is equal to α:

$$P(\Delta W_t(h) \leq \mathrm{VaR}_t(\alpha, h) | \mathcal{F}_{t-1}) = \alpha. \tag{13}$$

Despite its widespread use, the VaR also has been subject to substantial criticism, particularly because diversification does not always reduce risk when it is measured by VaR. In addition, the VaR ignores important information related to the tails of the loss distribution beyond the α-quantile, disregarding the risk of extreme losses. In contrast, ES measures cope well with such shortcomings and describe tail risk better (Artzner et al., 1999). The ES is defined as the expected loss, conditional on the loss exceeding the VaR over a certain horizon h,

$$\mathrm{ES}_t(\alpha, h) = \mathbb{E}[(\Delta W_t(h) | \Delta W_t(h) \leq \mathrm{VaR}_t(\alpha, h)) | \mathcal{F}_{t-1}], \tag{14}$$

that is, the mean portfolio loss in the α% of worst cases.

We describe our approach to compute VaR and ES under the dynamic econometric models proposed in the previous section for the vector of energy risk factors $\boldsymbol{r}_t$. Thus, we begin by modeling the joint distribution of energy returns $\boldsymbol{r}_t$, then we aggregate these results for each portfolio according to its exposures to each commodity. That is, we employ an asset-level approach to measure the tail risk of the energy portfolios.

To aggregate the risk factors, it is convenient to represent the portfolio's P&L as a linear function of the individual energy log-returns. We therefore approximate the changes in the portfolio value defined in Eq. (2) as

$$\Delta W_t(h) \approx \boldsymbol{w}'_{t-h}\boldsymbol{r}_t(h) = \sum_{i=1}^{4} q_{i,t-h} F_{i,t-h} r_{i,t}(h). \tag{15}$$

Because the GH distributions are closed under linear transformations (see Mcneil et al., 2005, their Proposition 3.13), when we aggregate the energy risk factors in a given portfolio, the linearized P&L distribution still belongs to the same class of GH distributions as does the vector of risk factors.

For the conditional mean $\boldsymbol{m}_t$ and covariance $\boldsymbol{H}_t$, and taking into account that the vector of innovations follows a four-variate GH distribution with mixing variable parameters (λ, χ, and ψ) and location, dispersion, and asymmetry parameters $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, and $\boldsymbol{\gamma}$, the linearized portfolio P&L in Eq. (15) follows a univariate GH distribution with parameters (λ, χ, and ψ) unaltered and with location, dispersion, and asymmetry parameters given by:

$$\begin{aligned} \mu_P &= \boldsymbol{w}'_{t-h}\left(\boldsymbol{m}_t + \boldsymbol{H}_t^{1/2}\boldsymbol{\mu}\right), \\ \Sigma_P &= \boldsymbol{w}'_{t-h}\left(\boldsymbol{H}_t^{1/2}\boldsymbol{\Sigma}\left(\boldsymbol{H}_t^{1/2}\right)'\right)\boldsymbol{w}_{t-h}, \quad \text{and} \\ \gamma_P &= \boldsymbol{w}'_{t-h}\left(\boldsymbol{H}_t^{1/2}\gamma\right). \end{aligned} \tag{16}$$

Thus, we obtain the parametric distribution function for the portfolio P&L. In turn, we can analyze the tail behavior and term structure of risk directly for each portfolio P&L distribution.

Finally, we adopt two alternative numerical implementations for calculating the risk measures. We can compute VaR and ES under the GH model by solving the integrals implicit in Eqs. (13) and (14) numerically for the portfolio P&L distribution. Alternatively, we can apply Monte Carlo simulations, which are usually more effective and preferred in this context. The latter procedure to characterize the portfolio risk measures is as follows:

1. Using our dynamic specification, defined by the conditional mean, variance, and correlation Eqs. (5), (7), and (8), we simulate the $t+\tau$ vector of returns, given Eqs. (3)–(4), as

$$\boldsymbol{r}_{t+\tau}^{(n)} = \boldsymbol{\mu}_{t+\tau} + \boldsymbol{H}_{t+\tau}^{1/2} x_{t+\tau}^{(n)}, \quad n = 1, \ldots, N,$$

where the vector of innovations $\{\boldsymbol{x}_{t+\tau}^{(n)}\}_{n=1}^{N}$ is drawn from the appropriate GH distribution, according to the normal mixture representation in Eq. (10). Then N is the number of Monte Carlo simulations, which we set to 100,000. The h-period returns generated by the n-th simulation at time t are given by $\boldsymbol{r}_t^{(n)}(h) = \sum_{\tau=1}^{h} \boldsymbol{r}_{t+\tau}^{(n)}$.
2. We build the simulated h-period portfolio P&L, $\{\Delta W_t^{(n)}(h)\}_{n=1}^{N}$, using the energy portfolio weights at time t, $\boldsymbol{w}_t$ (see Eq. (15)). That is, $\Delta W_t^{(n)}(h) \approx \boldsymbol{w}_t \boldsymbol{r}_t^{(n)}(h)$. Alternatively, instead of implementing Steps 1 and 2, we can simulate changes in the value of the energy portfolios using the parameters of the P&L distribution that result from Eq. (16).
3. Finally, we calculate the α percentile $\widehat{\mathrm{VaR}}_t(\alpha, h)$ and expected shortfall $\widehat{\mathrm{ES}}_t(\alpha, h)$ for the simulated distribution of the h-period portfolio value changes $\{\Delta W_t^{(n)}(h)\}_{n=1}^{N}$:

$$N^{-1}\sum_{n=1}^{N} \hat{I}_t^{(n)}(\alpha, h) = \alpha \quad \mapsto \quad \widehat{\mathrm{VaR}}_t(\alpha, h), \quad \text{and}$$
$$\left(\sum_{n=1}^{N} \hat{I}_t^{(n)}(\alpha, h)\right)^{-1} \sum_{n=1}^{N} \hat{I}_t^{(n)}(\alpha, h)\Delta W_t^{(n)}(h) \quad \mapsto \quad \widehat{\mathrm{ES}}_t(\alpha, h),$$

where $\hat{I}_t^{(n)}(\alpha, h) = \mathbb{1}_{\left\{\Delta W_t^{(n)}(h) \leq \mathrm{VaR}_t(\alpha, h)\right\}}$.

5. Data description and model estimation

In this section, we present the data we used to build the energy portfolios, then report on the parameter estimates of the GH models presented in the previous sections and some in-sample analyses of the results.

5.1. Description of the energy futures data

Our energy portfolios consist of energy commodity futures for crude oil, natural gas, coal, and electricity. These four commodities effectively represent a wide range of exposures to energy price risk. In all cases, we employ daily series of one-month ahead monthly futures contracts traded on the New York Mercantile Exchange (NYMEX), which are the most liquid contracts for the four energy commodities analyzed.[7] For oil, we consider the light sweet crude oil futures contract, which is quoted in dollars per U.S. barrel. For natural gas, we use the futures contract for the delivery location at the Henry hub in Louisiana, which trades in dollars per million British thermal units (mmBtu) and represents the benchmark for gas prices in the United States. For coal, we employ futures written on the Central Appalachian bituminous coal, quoted in dollars per U.S. tons. Finally, for electricity, we use the PJM monthly peak electricity futures traded in dollars per MWh.

The full-sample period runs more than six years from August 2005 to March 2012, and includes 1640 daily observations. We consider all data since the launch of the PJM electricity futures in the NYMEX (April 2003) until the day of the analysis (March 2012), but we drop the first observations (from April 2003 to August 2005), for which liquidity of electricity and coal futures was very scarce. To avoid in-sample overfitting and spurious findings, we reserve the last two years of data, from March 2010 to March 2012 (504 observations), for the out-of-sample investigation of the tail risk. The data came from the database of Thomson Reuters Datastream.

Table 1 reports some summary statistics of the daily returns for crude oil, natural gas, coal, and electricity. Fig. 1 presents the relative prices and quantile–quantile plots for the four energy commodities from August 2005 to March 2012. The table shows that no energy return shows any significant trend over either period: The means are small compared with the standard deviations of each series. Electricity has

[7] We have built the time series of the one-month ahead futures rolling over the front-month contract according to the NYMEX trading termination scheme for each futures contract (usually, between 8 and 1 business days before delivery).

Table 1
Descriptive statistics for energy returns.
This table reports sample statistics for the daily returns for crude oil, natural gas, coal, and electricity futures. The full sample period ranges from August 2005 to March 2012 and includes 1,640 observations. The in-sample period runs from August 2005 to March 2010 (1136 observations), and the out-of-sample period from March 2010 to March 2012 (504 observations). The Mean, *Std. Dev.*, *Min.*, Max., VaR 5%, and ES 5% are expressed in daily percentages. JB is the Jarque–Bera normality test statistic. Q(m) and LM(m) are the Ljung–Box and the Lagrange-Multiplier statistics, conducted using m lags to test for the presence of serial correlation in returns and squared returns, respectively. The p-values are reported in parentheses.

	10/08/2005–13/03/2012				10/08/2005–15/03/2010				15/03/2010–13/03/2012			
	Oil	Gas	Coal	Elec.	Oil	Gas	Coal	Elec.	Oil	Gas	Coal	Elec.
Mean	0.024	−0.079	0.001	0.068	0.027	−0.074	0.010	0.016	0.015	−0.090	−0.020	0.184
	(0.669)	(0.297)	(0.984)	(0.648)	(0.704)	(0.451)	(0.853)	(0.924)	(0.845)	(0.411)	(0.677)	(0.533)
Std. Dev.	2.246	3.062	1.632	6.029	2.430	3.295	1.825	5.739	1.764	2.462	1.078	6.642
Min.	−13.07	−10.78	−10.77	−36.60	−13.07	−10.78	−10.77	−26.82	−9.038	−8.057	−4.620	−36.60
Max.	13.34	26.77	11.12	47.64	13.34	26.77	11.12	47.64	5.164	16.69	2.984	44.55
Skew.	−0.013	1.194	−0.221	0.945	0.080	1.241	−0.201	0.975	−0.573	0.766	−0.405	0.872
	(0.831)	(0.000)	(0.000)	(0.000)	(0.27)	(0.000)	(0.006)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Kurt.	8.479	11.41	12.69	14.04	8.295	11.20	11.48	12.51	5.223	7.920	4.483	15.34
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
JB	2052	5223	6425	8576	1328	3476	3411	4464	131.3	558.6	60.00	3261
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
VaR	3.425	4.638	2.288	8.035	3.619	4.865	2.511	8.204	2.975	4.088	1.836	7.502
ES	5.323	6.260	4.116	13.84	5.693	6.661	4.697	13.32	4.414	5.114	2.666	14.99
Q(5)	4.086	13.14	74.21	18.89	3.447	10.43	48.77	13.58	6.713	3.919	37.04	8.135
	(0.537)	(0.022)	(0.000)	(0.002)	(0.631)	(0.064)	(0.000)	(0.019)	(0.243)	(0.561)	(0.000)	(0.149)
LM(5)	283.2	17.52	192.5	14.64	204.8	11.43	126.4	6.361	14.24	3.620	9.555	8.745
	(0.000)	(0.004)	(0.000)	(0.012)	(0.000)	(0.043)	(0.000)	(0.273)	(0.014)	(0.605)	(0.089)	(0.12)
corr(oil,j)	1.000	0.250	0.325	0.076	1.000	0.282	0.317	0.094	1.000	0.119	0.371	0.031
		(0.000)	(0.000)	(0.002)		(0.000)	(0.000)	(0.001)		(0.007)	(0.000)	(0.487)
corr(gas,j)		1.000	0.211	0.102		1.000	0.193	0.095		1.000	0.310	0.126
			(0.000)	(0.000)			(0.000)	(0.001)			(0.000)	(0.005)
corr(coal,j)			1.000	0.035			1.000	0.017			1.000	0.099
				(0.161)				(0.561)				(0.027)

the highest volatility, kurtosis, and risk measures over the entire sample. It also shows extreme positive and negative daily moves, some larger than 30%. Oil, gas, and coal indicate high risk, though they also experience a decrease in volatility and tail risk during the 2010–2012 period compared with the 2005–2010 period.

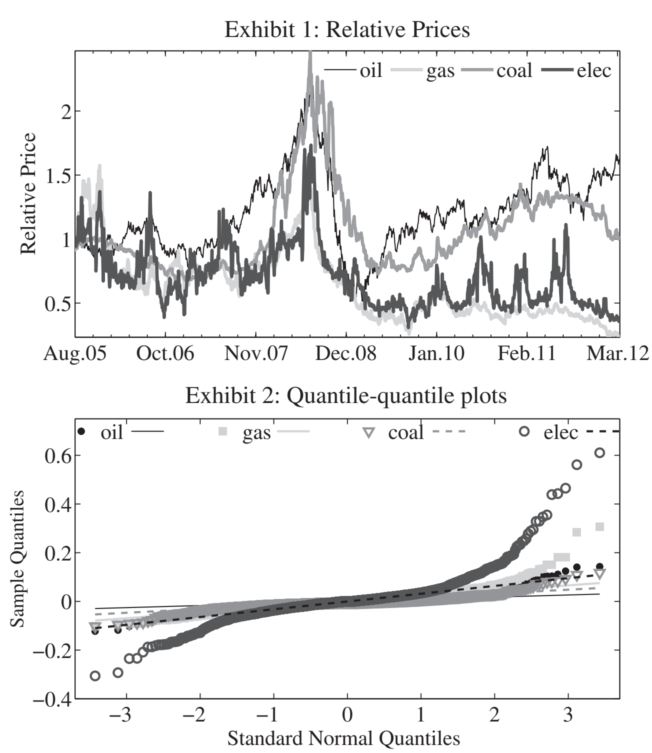

Fig. 1. Relative prices and QQ-plots for oil, gas, coal, and electricity. Exhibit 1 shows the price series of the energy futures from August 2005 to March 2012 (full-sample period). Exhibit 2 presents the sample quantiles of the daily returns for the four energy commodities; dashed lines represent the quantiles of a standard normal distribution.

We observe non-negligible skewness across commodity returns. Electricity and natural gas returns exhibit significantly positive skewness for all periods, suggesting that positive moves are more frequent than negative ones in these markets. The qq-plots in Fig. 1 confirm this evidence. In contrast, coal returns show negative asymmetry, and oil returns exhibit significantly negative skewness only for the last period of the sample. Taking into account these skewness measures, as well as the Jarque–Bera (JB) statistics, we reject unconditional normality in favor of the presence of heavy tails and asymmetry. We also test serial correlation in returns and squared returns using the Ljung–Box (Q) and Lagrange Multiplier (LM) statistics, respectively. The p-values in parentheses indicate that all commodities, except oil, display autocorrelation in returns, and strong evidence of temporal variation in the second moment of energy returns.

Looking at the correlation coefficients, we notice that linear dependence between energy commodities is positive and significant in general, but it varies among sample periods. The correlation coefficients between fuels (oil, gas, and coal), which to some extent represent substitute goods, range from 19% to 32% over the 2005–2010 period. In the 2010–2012 period, the correlations of oil and gas with coal increase to greater than 31%, whereas the correlation between oil and gas decreases from 28% to 12%. The linear dependence between electricity and fuels is less than 10% and only significant for oil and natural gas during 2005–2010. In the last period, correlation with coal increases to 10%.[8]

In addition to testing the univariate normality of energy returns, we conduct Mardia's test of multivariate normality (not reported here). This test is based on multivariate measures of skewness and kurtosis, defined on the bilinear form $D_{tt'} = (\boldsymbol{r}_t - \overline{\boldsymbol{r}})'\overline{\boldsymbol{H}}^{-1}(\boldsymbol{r}_{t'} - \overline{\boldsymbol{r}})$, where $\overline{\boldsymbol{r}}$ and $\overline{\boldsymbol{H}}$ are the sample mean and covariance of returns, respectively. The large values that we obtain for the test statistics, corresponding to multivariate skewness and kurtosis measures, reject the null hypothesis of joint

[8] The electricity futures are written on peak-load power, such that we should expect a higher correlation with peak-load fuels, such as natural gas, than with coal, which is usually a source for generating intermediate- and base-load electricity.

normality of energy returns (p-values are less than 0.001 for all the sample periods).

5.2. Estimation results for the multivariate model

In this section, we present in two stages the estimation results of the multivariate GH models for the energy returns. The large dimension of the model prompts us to use a sequential approach to estimate the set of parameters. In the first stage, we carry out the quasi-maximum likelihood (QML) estimation of the dynamic regression model for the conditional mean and covariance, defined in Eqs. (5), (7), and (8). In the second stage, we obtain the ML parameter estimates of the different multivariate GH conditional distributions. This latter estimation can be implemented using a variant of the expectation–maximization (EM) procedure presented by Mcneil et al. (2005), which relies on the normal-mixture representation for GH distributions from Eq. (10). In Appendix C, we describe the log-likelihood functions that correspond to both stages and the optimization algorithms.

We perform the first estimation for the period that ranges from August 2005 to March 2010, formed by 1136 daily observations. In turn, we recursively reestimate the models throughout the out-of-sample period, using a rolling window of constant size ($T = 1136$), such that we obtain a full sequence of parameter estimates for each point in the out-of-sample period (504 re-estimations total). For each point in the sequence of parameter estimates, we can compute the h-period forecast density that we later use to calculate the tail risk of the energy portfolios.

5.2.1. QML estimates of the conditional moments

To display the outputs of the dynamics specification, in Table 2 we present the QML parameter estimates, robust p-values, and residual summary statistics of the conditional mean and covariance equations for the first estimation window (August 2005 to March 2010). We include in each mean equation only those autoregressive lags that have significant coefficients or cause the presence of autocorrelation to become statistically insignificant in the resulting residuals. The coefficients of lagged returns are negative, except for coal, which has significantly positive autoregressive parameters. For natural gas and electricity, some autoregressive lags are also notably significant.

We observe different patterns in the variance equation, especially with respect to the leverage effect. For crude oil and natural gas, the parameter α_1^-, corresponding to the leverage effect, is positive and significant, which suggests that negative shocks have a stronger effect on variance than do positive ones. Coal and electricity do not indicate any such asymmetry in terms of the response of volatility to negative moves, which suggests that positive shocks could have more impact on variance. Volatility persistence, measured as $\alpha_1 + \alpha_1^-/2 + \alpha_2$, also is very large (>0.95) for fossil fuels but smaller for electricity variance (around 0.50). According to the Ljung–Box (Q) and Lagrange multiplier (LM) statistics in Table 2, using this mean and variance specification, we can greatly reduce the presence of temporal dependence in the residuals and squared residuals.

When we consider the time-varying evolution of the correlation matrix for the vector of four energy returns, we find that dependence dynamics are strongly persistent ($\delta_2 = 0.89$), the correlation increases when energy assets are affected by shocks of the same sign (δ_1 is small but significantly positive), and the asymmetry effect in the ADCC model is rather insignificant ($\delta_1^+ < 0.01$). However, when we consider the dynamics of the correlation between pairs of energy returns, we obtain significant, positive estimates of the asymmetry parameter δ_1^+. In particular, for natural gas and electricity, we find that $\delta_1^+ = 0.099 > \delta_1 = 0.043$, suggesting that the correlation between gas and electricity increases more after a positive co-movement ("co-boom") than after equally large negative co-movements. Although not previously documented in this context, such positive asymmetry in correlation seems sound from an economic perspective, because an increase in gas prices has a strong positive impact on the generation costs of peak-load electricity. We study this relationship between gas and electricity in depth when we analyze the tail risk of the gas-fired power plant portfolio.

When we update our parameter estimates at each point of the out-of-sample period, we obtain a time-series of the conditional moments from March 2010 to March 2012. Thus in Fig. 2, we plot the 10-day-ahead forecasts of the conditional volatilities and correlations over the out-of-sample period. We observe several interesting features from a risk analysis perspective. As expected, electricity is highly volatile, with a strong presence of spikes, some greater than 50%. In addition, electricity volatility peaks revert to the mean quicker than those of other energy volatilities. Correlations are not very high during this period, especially between electricity and fuels. The correlation between oil and gas also decreases substantially, and only the correlations between coal and oil and between coal and gas exhibit a significant increase throughout the out-of-sample period.

Table 2
In-sample QML parameter estimates of the conditional model.
This table reports QML parameter estimates and residual summary statistics for the conditional mean and variance equations. The p-values, presented in parentheses, are computed using robust standard errors. Q(m) and LM(m) are the Ljung–Box and the Lagrange-Multiplier statistics, conducted using m lags to test for the presence of serial correlation in residuals and squared residuals.

Panel A: univariate dynamics, GJR(1,1,1)

	Crude oil	Natural gas	Coal	Electricity
Mean equation				
$m_0 \times 10^2$	0.060	−0.067	0.008	−0.242
	(0.276)	(0.420)	(0.831)	(0.123)
ϕ_1	−0.033	−0.038	0.105	−0.042
	(0.285)	(0.204)	(0.002)	(0.457)
ϕ_2			0.092	
			(0.007)	
ϕ_3				−0.077
				(0.016)
ϕ_4				−0.064
				(0.019)
ϕ_5		−0.073		
		(0.043)		
Variance equation				
$\alpha_0 \times 10^3$	0.005	0.008	0.002	1.941
	(0.025)	(0.060)	(0.001)	(0.000)
α_1	0.027	0.024	0.056	0.394
	(0.046)	(0.045)	(0.000)	(0.000)
α_1^-	0.060	0.076	0.007	0.011
	(0.002)	(0.022)	(0.612)	(0.914)
α_2	0.933	0.936	0.935	0.092
	(0.000)	(0.000)	(0.000)	(0.273)
Statistics for residuals				
Q(5)	7.394	1.281	4.024	7.185
	(0.193)	(0.937)	(0.546)	(0.207)
LM(5)	14.158	8.596	6.272	2.817
	(0.105)	(0.126)	(0.281)	(0.728)

Panel B: correlation dynamics, ADCC(1,1,1) and DCC(1,1)

	Oil–Gas–Coal–Elec		Gas–Elec	
	ADCC	DCC	ADCC	DCC
δ_1	0.017	0.020	0.043	0.071
	(0.042)	(0.022)	(0.120)	(0.006)
δ_1^+	0.008		0.099	
	(0.387)		(0.064)	
δ_2	0.894	0.893	0.663	0.670
	(0.000)	(0.000)	(0.000)	(0.000)

5.2.2. Multivariate GH conditional distributions

We now fit the various conditional distributions using the vector of innovations obtained from the previous QML estimation. As we did

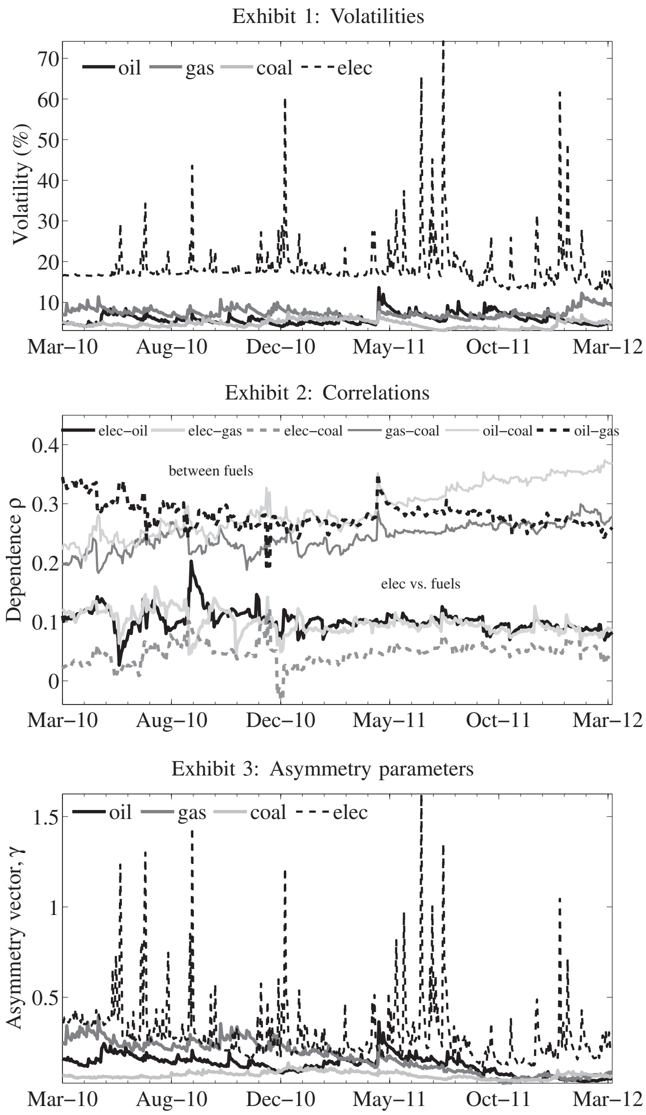

Fig. 2. Volatility, correlation, and asymmetry parameter estimates. Exhibits 1 and 2 show the 10-day ahead forecasts of the conditional volatilities and conditional correlations over the out-of-sample period (March 2010 to March 2012). Exhibit 3 shows the asymmetry parameter vector (γ) for the multivariate VG distribution for each rolling estimation over the out-of-sample period. We employ a rolling window of constant size (T = 1,136) to obtain the sequence of parameter estimates (504 re-estimations total).

Table 3
In-sample parameter estimates of the multivariate distributions.
This table presents the parameter estimates for the multivariate conditional distribution of the vector of innovations. We consider two elliptically symmetric distributions, the multivariate Gaussian (G) and Student's t (T) distributions, and three GH distributions: the multivariate skewed t (skT), variance-gamma (VG), and normal inverse Gaussian (NIG) distributions. The standard errors of these estimates are computed using a stationary bootstrap (500 samples), and their corresponding p-values appear in parentheses. The log-L, AIC, and BIC are the values at the optimum of the log-likelihood function and the Akaike and Bayesian information criteria, respectively.

	G	T	skT	VG	NIG
Parameters of the mixing r.v. W					
λ		−2.078	−2.084	1.295	−0.500
		(0.005)	(0.002)	(0.000)	
χ		4.157	4.168	0.000	0.757
		(0.002)	(0.001)		(0.001)
ψ		0.000	0.000	2.588	0.637
				(0.003)	(0.000)
Asymmetry vector γ					
γ(oil)			0.026	0.102	0.058
			(0.117)	(0.020)	(0.042)
γ(gas)			0.037	0.106	0.078
			(0.049)	(0.018)	(0.021)
γ(coal)			0.018	0.050	0.041
			(0.263)	(0.034)	(0.050)
γ(elec)			0.015	0.078	0.039
			(0.332)	(0.028)	(0.190)
Dispersion matrix Σ					
ρ(oil, gas)	0.347	0.367	0.365	0.360	0.364
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
ρ(oil, coal)	0.218	0.245	0.245	0.240	0.243
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
ρ(oil, elec)	0.109	0.118	0.117	0.119	0.118
	(0.001)	(0.000)	(0.001)	(0.000)	(0.000)
ρ(gas, coal)	0.213	0.247	0.246	0.235	0.244
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
ρ(gas, elec)	0.118	0.222	0.220	0.207	0.219
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
ρ(coal, elec)	0.032	0.064	0.063	0.058	0.063
	(0.105)	(0.071)	(0.077)	(0.079)	(0.075)
Information criteria					
log-L	−5.555	−5.221	−5.220	**−5.186**	−5.200
AIC	11.134	10.469	10.473	**10.406**	10.433
BIC	11.196	10.535	10.557	**10.491**	10.517

Bold numbers indicate the model that gives the largest value of the log-likelihood and the smallest values of Akaike and Bayesian information criteria.

previously, we repeat the multivariate distribution estimation for each point in the out-of-sample period. In Table 3, we present, for the first estimation period (August 2005 to March 2010), the ML parameter estimates and bootstrapped p-values of the various GH distributions. The in-sample results offer strong evidence against multivariate normality, as we expected.

First, the shape parameter estimates of the mixing distributions (λ, χ, and ψ) point to the presence of fat tails in the different GH models. In particular, for the T and skT distributions, the small value of parameter χ (around 5) indicates the existence of jumps and tail dependence. Similar arguments apply to the VG and NIG parameter estimates.

Second, the asymmetry parameter estimates $\boldsymbol{\gamma}$ for the three skewed GH distributions (skT, VG, and NIG) are positive for all vector components, suggesting positive skewness in the multivariate conditional distribution of daily energy returns. Only the asymmetry parameter for natural gas is statistically significant for the three GH models, whereas the parameter for electricity is significant just for the VG case.

Third, looking at the log-likelihood values (log-L), together with the Akaike and Bayesian information criteria (AIC and BIC in Table 3),[9] to compare the accuracy of the alternative conditional distributions, we find that the VG model is preferable over other GH alternatives for this estimation window.

We reach similar results when we reestimate the GH conditional distributions throughout the out-of-sample period. Specifically, we use a rolling window of size T = 1136, for a total of 504 re-estimations. For example, Exhibit 3 of Fig. 2 displays the evolution of the asymmetry parameter estimates ($\boldsymbol{\gamma}$) for the VG model from March 2010 to March 2012 (similar patterns are obtained for the other skewed GH distributions). For all assets, the parameter estimates are always positive. In particular, the asymmetry parameter of electricity presents large positive spikes that seem to revert toward a mean value. For gas, we observe a downward trend over the sample.

We also compute the BIC values (not reported here) corresponding to the recursive models' estimations. The sequence of re-estimation

[9] The selection criteria shown in Table 3 are given by the following: $log\text{-}L = \log\mathcal{L}_2(\hat{\theta})/T$, AIC = $-2\text{log-L} + 2\,k/T$, and BIC = $-2\text{log-L} + k\log(T)/T$, where k is the number of parameters in each model, and T is the number of observations. Greater log-L, and lower AIC and BIC, values are preferred.

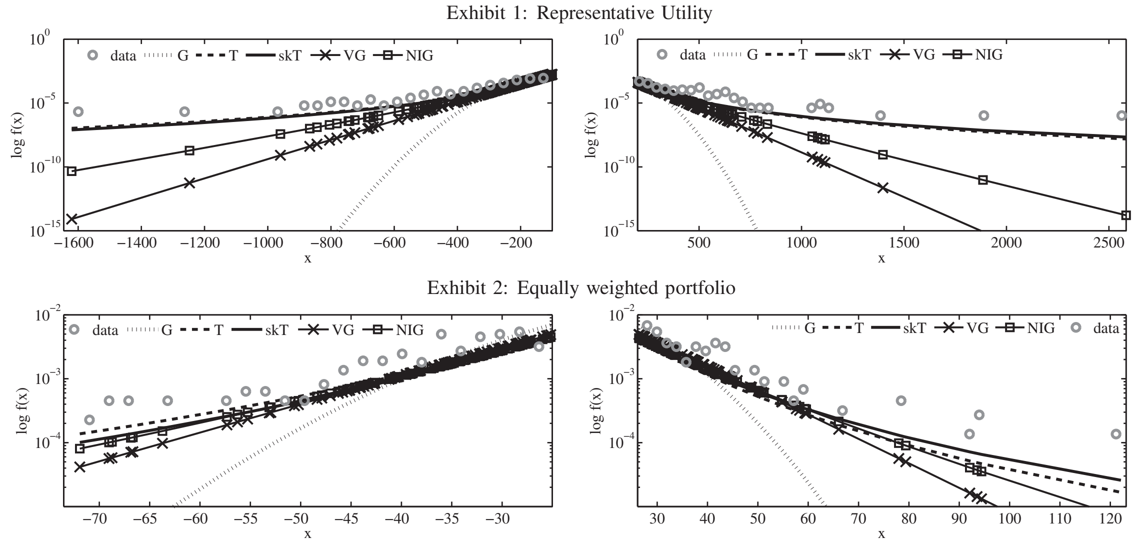

Fig. 3. Tail plots of energy portfolios. This figure shows the left and right tail plots for two of our energy portfolios: the representative PJM utility portfolio and the equally weighted portfolio. The initial exposures, in dollars, of the utility portfolio are around US$2700 for the fuels (oil, gas, and coal) and US$5016 for 57 MWh of electricity. For the equally weighted portfolio, we consider an arbitrary initial investment of US$1000. The circles represent the logarithm of the empirical probability density of innovations. These figures present the estimated probability density functions of the multivariate Gaussian, T, skewed T, VG, and NIG distributions.

results favors the hypothesis of fat tails and positive skewness for the multivariate distribution of energy risk factors over the whole out-of-sample period. Even for the most recent estimations (i.e., the last rolling window ranges from September 2007 to March 2012), for which electricity and coal futures are less influenced by the lack of trading volume, the results suggest the presence of extreme realizations in the conditional distribution.

The front-month futures contracts may exhibit excess volatility near delivery, precisely, at the time that these contracts are rolled over. To check if some of the tail risk we find in the estimation results are due to these rollover effects near maturity, we also estimate our models for the time series of two-month ahead futures contracts.[10] To build the generic time series of two-month ahead futures, we roll over the corresponding contract at least 22 business days before delivery. Therefore, using this time series, we limit the effects of rollover near maturity. The drawback of using two-month ahead contracts is that these futures are less liquid than the nearest to maturity contract, specially for coal and electricity.

The results show similar estimates for the asymmetry parameters and for the parameters of the mixing random variable W, which governs the degree of tail risk in the multivariate distribution. For example, for the VG model, the estimates of the parameters λ and ψ are 0.98 and 2.02, and the estimates of the asymmetry parameters are 0.045, 0.085, 0.055, and 0.067, for the oil, gas, coal, and electricity returns, respectively. All of them are significant at least at the 10% confidence level. Furthermore, the estimates of parameter χ for the T and skT models are around 4, showing the presence of extreme realizations in the conditional distribution.[11]

5.2.3. Left and right tails of the energy portfolios

To analyze the tails of the returns distribution of energy portfolios, we consider the following examples: a utility with different generation units, a gas-fired power plant, and equally weighted and minimum variance portfolios. Using the multivariate GH models previously estimated and knowing the portfolio weights, we can obtain a fitted distribution of portfolio returns for each GH model. Then, we compare the in-sample tail fit of the estimated models graphically, by plotting the estimated logarithmic density functions and the empirical log-density function of the portfolio.

The results of the multivariate estimations have shown the presence of tail asymmetry in the multivariate density functions. Now, looking at the tails of the different energy portfolios, we can analyze the joint effect of this asymmetry and the portfolio weights on the aggregate tail behavior (see Eq. (16)).

To focus on the aggregate tail risk behavior, we display in Fig. 3 enlarged sections of the left and right tails of the energy portfolios. The circles represent the empirical probability density of portfolio return innovations. The left panel of each exhibit presents the left tail of a long position in the portfolio, and the right panel is the corresponding right tail. The distributions of portfolio returns show positive skewness and fat tails. We find that the Gaussian model (dotted line) clearly underestimates the extent of both tails, that is, the probability of extreme realizations.

Although multivariate VG and NIG models have the largest log-likelihood values and the lowest BIC values (see Table 3), the T and skT models better estimate the aggregate tail risk behavior, according to the plots in Fig. 3. The slower tail decay of T and skT distributions (solid and dashed lines, respectively) causes them to outperform the tail fit of the VG and NIG models (marked with crosses and squares, respectively), especially for the right tail, which corresponds to losses of a short position in the energy portfolio.

We also observe slight differences between the tail fits of the T and skT distributions, partially due to the asymmetric tail behavior of the skT distribution. Exhibits for equally weighted and minimum variance portfolios show that the left tail of the estimated skT distribution is above the T distribution, whereas the right tail is below it.

Therefore, the extent of underestimation of the tail risk of the portfolio loss distribution strongly depends on whether we are analyzing the short or long position in the energy portfolio. In the next section, we further consider the aggregate tail behavior of the energy portfolios' loss distribution, looking at the out-of-sample performance of the VaR and ES measures.

6. Risk measures and out-of-sample performance

We apply the procedure explained in Section 4 to estimate the conditional risk measures of the energy portfolios for different multivariate

[10] We thank a referee for pointing out this issue.

[11] We do not report all the estimates in the interest of brevity. They are available upon request. For the in-sample window, we obtain BIC values equal to 10.90, 9.32, 9.34, 9.43, and 9.35 for the G, T, skT, VG, and NIG models, respectively. Therefore, in this sample, T and skT models are preferred.

GH distributions. See Appendix A for a detailed description of the four energy portfolios considered in the analysis. We compute VaR and ES over the out-of-sample period for different horizons and confidence levels, characterizing the term structure of these risk measures for each GH model. Then, we test the out-of-sample performance of the forecast risk measures, assessing the relative ability of the various multivariate models at hand.

6.1. Forecast VaR and ES

The multivariate GH models are estimated on the energy returns up to time t, and then we calculate, for a given confidence level, the out-of-sample h-horizon VaR and ES forecasts (i.e., the risk measures for the period $[t+1, t+h]$). In addition to our GH models, we also calculate the portfolio risk measures using several approaches: a traditional variance–covariance method with multivariate unconditional Gaussian distribution (VC); the Riskmetrics procedure or equally weighted exponential average model (EWMA), as first introduced by J.P. Morgan; the multivariate Gaussian GARCH with constant conditional correlation (CCC); and the non-parametric historical simulation method (HS).

Using the various multivariate approaches, we calculate the conditional risk measures (VaR and ES) of the four energy portfolios for horizons extending from 1 to 22 days. By way of a sensitivity exercise to the cut-off point selection, we also consider in our analyses different confidence levels α: 0.1%, 0.5%, 1%, and 5%. Thus, we can analyze the possible bias in the risk measure estimates due to the fixing of the confidence level. As an example, Fig. 4 shows the average ES of the equally weighted portfolio for different confidence levels and horizons.

The smallest ES estimates are generally produced by the EWMA or Gaussian models (CCC and VC). The largest ES estimates among the GH models correspond to the distributions with polynomial (slower) tail decays, that is, to the T and skT distributions. The asymmetric pattern of the skT model produces slightly larger tail risk estimates than the T model for the long positions of the energy portfolios, especially for the equally weighted and minimum variance portfolios. The ordering of the ES estimates across GH methods is invariant to the forecast horizon. In general, the tail risk estimates of the nonparametric HS are close to those of the T and skT models for the larger cut-off points (e.g. 1%). The estimation windows characterized by turbulent periods of fuel returns are responsible for these large risk measure estimates of the HS approach; as we observe, only heavy-tailed distributions are able to produce similar tail risk patterns.

In practice, our interest lies in comparing (backtest) the h-horizon risk measure forecasts for long and short positions with the actual portfolio losses during the two-year out-of-sample period, from March 2010 to March 2012. As a result of this comparison and following the notation of Section 4, we can compute the indicator variables $\hat{I}_t(\alpha, h) = \mathbb{1}_{\left\{\Delta W_t(h) \le \widehat{\text{VaR}}_t(\alpha,h)\right\}}$, which signal the violations of the risk measure $\widehat{\text{VaR}}_t(\alpha, h)$ (or $\widehat{\text{ES}}_t(\alpha, h)$) of the portfolio loss distribution. In the next subsection, we use the processes $I_t(\alpha, h)$ to obtain their corresponding failure rates and develop tests of the out-of-sample performance of the estimated risk measures.

In the following backtest results, we show mainly results for the 1% and 5% cut-off points. Similar conclusions to those of the 1% confidence level can be inferred for the 0.5% confidence level. For the 0.1% cut-off and for long horizons, such as 22-day or 10-day horizons, there are few observations in the tail to infer strong conclusions.

As an example of the series of forecast risk measures that we obtain, we present in Fig. 5 the 1% 1-day VaR over the out-of-sample period according to three different approaches: the EWMA, VG, and skT models. In the lower side of the figure, we draw the VaR violations of each model, corresponding to the long position in the portfolio. In the upper side, we mark the VaR violations for the short position. In general, there are more violations across models for short positions than for long ones, in support of the positive asymmetry of the P&L distribution of the energy portfolios. We also observe that the VG and skT estimates

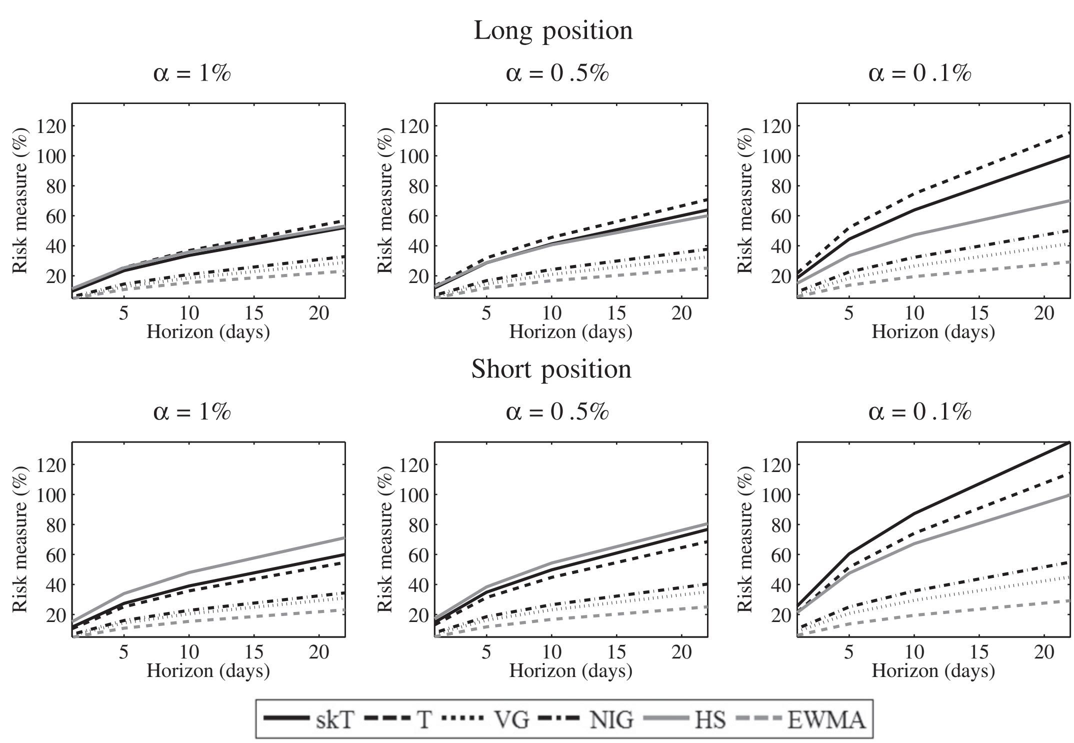

Fig. 4. Term structure of risk: ES. This figure shows the average expected shortfall (in percentage) of the equally weighted portfolio for different day horizons and tail cut-offs (α). We use the convention of reporting risk measures as positive numbers representing a loss.

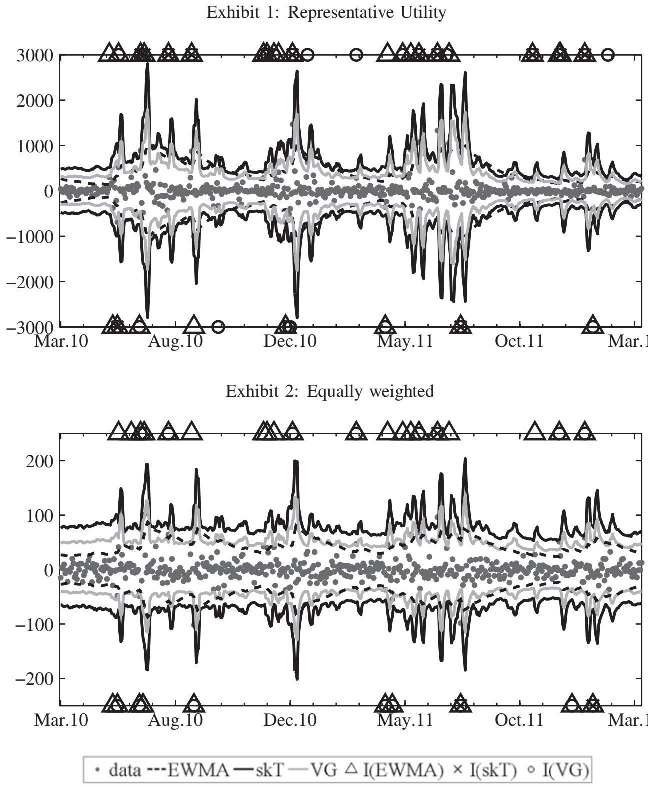

Fig. 5. 1-day risk measures and violations. This figure presents the 1% 1-day VaR for three different approaches over the out-of-sample period. Their corresponding violations are also reported. Triangles, circles, and x-marks denote violations of the EWMA, conditional VG, and conditional skewed T models, respectively. The initial exposures, in dollars, of the utility portfolio are around US$2700 for the fuels (oil, gas, and coal) and US$5016 for 57 MWh of electricity (see Appendix A for more details about the portfolios). For the equally weighted portfolio, we consider an arbitrary initial investment of US$1000.

(grey and black lines, respectively) respond more quickly to changing volatility than does the EWMA estimate (dashed line), which tends to be violated several times in a row during more turbulent periods (violations of the EWMA risk measure are marked with triangles). In addition, the VaR violations of the skT estimate (crosses) are fewer than those of the VG estimate (circles), suggesting again the importance of modeling the presence of heavy tails to produce conservative tail risk measures.

In the next subsection, we test if these differences are statistically and economically significant.

6.2. Tests of out-of-sample performance

In the previous subsection, we estimated, from March 2010 to March 2012, the risk measures at time t for the distribution of losses in the next h-day period $[t+1, t+h]$. Now, we backtest the risk measure estimates of the different models over time and compare their out-of-sample relative performance. Thus we can assess the differences in tail risk patterns, controlling for over-fitting and other spurious findings. In this subsection, we consider various backtests for the VaR and ES forecasts. In particular, to monitor the performance of tail risk estimates, we implement a new backtest that builds on the reality check tests of Hansen (2005) and White (2000) with a loss function based on the ES.

Using the indicator variable $\hat{I}_t(\alpha, h)$, we obtain the number of VaR violations for a given confidence level α over the tested period, as well as the proportion of losses beyond that VaR estimate. By definition, the probability of incurring a VaR violation for a successful model is α. Therefore, the estimated indicator variable $\hat{I}_t(\alpha, h)$ should behave similar to realizations of a Bernoulli random variable with probability α. We first test the unconditional coverage, to check if the number of violations is correct on average, using the log-likelihood ratio (LR) statistic proposed by Christoffersen (1998) and Kupiec (1995).[12]

For the five GH models and the four benchmark approaches, we report in Table 4 the percentage of VaR violations and the unconditional coverage LR p-values corresponding to (long and short) 1-day and 10-day VaR estimates at a 1% confidence level. The results show that the more traditional parametric approaches, such as VC, EWMA, CCC, and Gaussian-DCC models, tend to underestimate the VaR, especially for short positions and for utility portfolios. The non-parametric HS produces better coverage probabilities for these cases. The VaR estimates from the VG and NIG models are also too low for short positions, particularly at the 1-day horizon, whereas the T and skT models (i.e., those with polynomial tail decay) are generally better for these positions. For long positions, especially at the 10-day horizon, heavy-tailed models have very small and even zero coverage probabilities (i.e., there are no violations).

To backtest the success of our estimated ES, we follow the methods proposed by Embrechts et al. (2005) and Mcneil and Frey (2000). In both approaches, we evaluate the discrepancies between losses and ES estimates, defined by $\hat{D}_t(\alpha, h) = \Delta W_t(h) - \widehat{\mathrm{ES}}_t(\alpha, h)$, on those points at which the VaR estimate is violated, that is, when $\hat{I}_t(\alpha, h) = 1$. Ideally, we would expect that the variable $\left[\hat{D}_t(\alpha, h)\hat{I}_t(\alpha, h)\right]$ behaves similar to realizations of a distribution with mean close to zero.

Thus in Table 5, we present, for the 1-day and 10-day 1% ES estimates of the various multivariate models, the backtest measure V_1, defined as the conditional average of excesses $\hat{D}_t(\alpha, h)$ conditioned on the indicator variable $\hat{I}_t(\alpha, h)$, and the p-value (MF) corresponding to a bootstrap test that checks if ES values are systematically underestimated.[13] A good estimation of the ES leads to a non-negative and low value of the backtest measure V_1. We observe that ES exceedances are quite high for (conditional and unconditional) Gaussian models, especially for the two utility portfolios. Semi-heavy-tailed models (VG and NIG) behave reasonably well for long positions in the utility portfolios and for both short and long positions in the equally weighted and minimum variance portfolios, but they underestimate the 1-day tail risk of short positions in utility portfolios. The T and skT models produce average exceedances that are generally positive or at most slightly negative. These heavy-tailed models behave better than alternative versions with regard to the tail risk of short positions.

A more appropriate test would compare all models jointly, to determine whether the differences in the tail patterns are statistically significant. Thus, we next compare the performance of the tail risk estimates using the superior predictive ability (SPA) test of Hansen (2005), which requires the stationary bootstrap of Politis and Romano (1994) to be implemented. In our application of the SPA test, we generate $B =$ 10,000 bootstrap resamples. This test is designed to assess whether a particular model is significantly outperformed by others, while also controlling for the set of models being compared. Thus, over the out-of-sample (validation) period, we evaluate the risk measure forecasts using a prespecified loss function. The preferred model produces the smallest expected loss. Hansen (2005) proposes a procedure to approximate the distribution of t-statistics for the SPA test, and derives consistent estimates of the p-values. The null hypothesis that the benchmark

[12] Consider the following LR statistic: $\mathrm{LR} = 2log\left(\left(n_\alpha/L\right)^{n_\alpha}\left(1-n_\alpha/L\right)^{L-n_\alpha}\right) - 2log\left(\left(1-\alpha\right)^{n_\alpha}\alpha^{L-n_\alpha}\right)$, where $n_\alpha = \sum_{t=T+1}^{T+L}\hat{I}_t(\alpha, h)$ is the number of VaR violations over the out-of-sample (validation) period $(T+1, \ldots, T+L)$. Under the null hypothesis that expected proportion of violations should be α, the LR statistic should be asymptotically distributed as $\chi^2(1)$.

[13] The average of $\hat{D}_t(\alpha, h)$ conditioned on $\hat{I}_t(\alpha, h)$ is given by: $V_1 = \sum_{t=T+1}^{T+L}\hat{D}_t(\alpha, h)\hat{I}_t(\alpha, h)/\hat{n}(\alpha, h)$, where $\hat{n}(\alpha, h) = \sum_{t=T+1}^{T+L}\hat{I}_t(\alpha, h)$. We are more worried about the negative values of the differences $D_t(\alpha, h)$ than about the positive ones though, so we test the null hypothesis of a non-negative mean against the alternative of a mean less than zero. For that purpose, following Mcneil and Frey (2000), we apply a one-sided bootstrap test that makes no assumption about the violation residuals' distribution.

Table 4
Quantiles backtest: coverage measure.
This table reports the percentage of VaR violations (% viol.) and the *p*-values of the average coverage log-likelihood ratio (LR), corresponding to both 1-day and 10-day 1% VaR estimates over the out-of-sample period (March 2010 to March 2012) for different multivariate models.

	Representative utility portfolio								Equally weighted portfolio							
	Short position				Long position				Short position				Long position			
	$h=1$		$h=10$		$h=1$		$h=10$		$h=1$		$h=10$		$h=1$		$h=10$	
	% viol.	LR	% viol.	LR	% viol.	LR	% viol.	LR	% viol.	LR	% viol.	LR	% viol.	LR	% viol.	LR
VC	0.040	0.000	0.060	0.000	0.022	0.021	0.000	0.002	0.032	0.000	0.018	0.110	0.018	0.101	0.020	0.045
EWMA	0.042	0.000	0.067	0.000	0.016	0.222	0.004	0.130	0.040	0.000	0.022	0.019	0.020	0.050	0.010	0.982
CCC	0.046	0.000	0.067	0.000	0.020	0.050	0.000	0.002	0.038	0.000	0.018	0.101	0.018	0.110	0.014	0.383
G	0.046	0.000	0.067	0.000	0.020	0.050	0.000	0.002	0.036	0.000	0.020	0.045	0.018	0.110	0.014	0.383
T	0.022	0.021	0.008	0.657	0.004	0.121	0.000	0.002	0.008	0.629	0.000	0.002	0.004	0.121	0.000	0.002
skT	0.022	0.021	0.008	0.657	0.004	0.121	0.000	0.002	0.002	0.027	0.000	0.002	0.004	0.121	0.000	0.002
VG	0.040	0.000	0.042	0.000	0.018	0.110	0.000	0.002	0.018	0.110	0.002	0.030	0.016	0.222	0.012	0.646
NIG	0.038	0.000	0.034	0.000	0.012	0.677	0.000	0.002	0.018	0.110	0.002	0.030	0.010	0.986	0.004	0.130
HS	0.018	0.110	0.006	0.342	0.004	0.121	0.000	0.002	0.002	0.027	0.000	0.002	0.002	0.027	0.000	0.002
	Minimum variance portfolio								Gas-fired power plant							
	Short position				Long position				Short position				Long position			
	$h=1$		$h=10$		$h=1$		$h=10$		$h=1$		$h=10$		$h=1$		$h=10$	
	% viol.	LR	% viol.	LR	% viol.	LR	% viol.	LR	% viol.	LR	% viol.	LR	% viol.	LR	% viol.	LR
VC	0.034	0.000	0.012	0.646	0.020	0.050	0.016	0.206	0.038	0.000	0.053	0.000	0.026	0.003	0.000	0.002
EWMA	0.036	0.000	0.024	0.007	0.022	0.021	0.004	0.130	0.042	0.000	0.059	0.000	0.020	0.050	0.004	0.130
CCC	0.040	0.000	0.020	0.045	0.026	0.003	0.012	0.646	0.042	0.000	0.067	0.000	0.022	0.021	0.000	0.002
G	0.042	0.000	0.022	0.019	0.024	0.008	0.012	0.646	0.042	0.000	0.067	0.000	0.024	0.008	0.000	0.002
T	0.012	0.677	0.000	0.002	0.004	0.121	0.000	0.002	0.020	0.050	0.008	0.657	0.004	0.121	0.000	0.002
skT	0.010	0.986	0.000	0.002	0.004	0.121	0.000	0.002	0.020	0.050	0.008	0.657	0.004	0.121	0.000	0.002
VG	0.024	0.008	0.004	0.130	0.014	0.407	0.006	0.342	0.034	0.000	0.044	0.000	0.016	0.222	0.000	0.002
NIG	0.022	0.021	0.004	0.130	0.012	0.677	0.004	0.130	0.032	0.000	0.034	0.000	0.006	0.323	0.000	0.002
HS	0.006	0.323	0.000	0.002	0.004	0.121	0.000	0.002	0.016	0.222	0.004	0.130	0.002	0.027	0.000	0.002

Table 5
Expected shortfall (ES) backtests.
This table reports the Embrechts et al. (2005) backtest measure V_1 and the *p*-value of Mcneil and Frey's (2000) test (MF). Both are calculated for the 1-day and 10-day 1% ES forecasts of various multivariate models over the out-of-sample period.

	Representative utility portfolio								Equally weighted portfolio							
	Short position				Long position				Short position				Long position			
	$h=1$		$h=10$		$h=1$		$h=10$		$h=1$		$h=10$		$h=1$		$h=10$	
	V_1	MF	V_1	MF	V_1	MF	V_1	MF	V_1	MF	V_1	MF	V_1	MF	V_1	MF
VC	−349.2	0.000	−294.5	0.182	−137.7	0.090	0.000	1.000	−19.05	0.104	8.168	1.000	−23.95	0.160	4.869	1.000
EWMA	−214.2	0.000	−193.4	0.000	−175.1	0.000	−112.9	0.121	−11.18	0.000	−10.03	0.090	−14.23	0.000	6.873	1.000
CCC	−268.2	0.000	−232.0	0.230	−121.7	0.060	0.000	1.000	−11.57	0.010	8.544	1.000	−10.72	0.200	−2.352	1.000
G	−267.7	0.000	−229.2	0.200	−120.1	0.060	0.000	1.000	−12.88	0.020	8.388	0.980	−11.28	0.190	−2.143	1.000
T	−15.39	1.000	629.9	1.000	195.2	1.000	0.000	1.000	27.90	1.000	0.000	1.000	5.812	1.000	0.000	1.000
skT	−9.851	0.990	693.8	1.000	197.8	1.000	0.000	1.000	20.68	1.000	0.000	1.000	0.276	1.000	0.000	1.000
VG	−212.4	0.020	−27.83	0.990	−50.62	0.310	0.000	1.000	−8.834	0.950	29.64	1.000	−4.014	0.580	20.52	1.000
NIG	−182.9	0.100	62.24	1.000	−43.70	0.410	0.000	1.000	−1.254	1.000	49.66	1.000	−7.710	0.740	15.54	1.000
HS	−161.8	0.170	319.5	1.000	144.5	1.000	0.000	1.000	24.21	1.000	0.000	1.000	7.883	1.000	0.000	1.000
	Minimum variance portfolio								Gas-fired power plant							
	Short position				Long position				Short position				Long position			
	$h=1$		$h=10$		$h=1$		$h=10$		$h=1$		$h=10$		$h=1$		$h=10$	
	V_1	MF	V_1	MF	V_1	MF	V_1	MF	V_1	MF	V_1	MF	V_1	MF	V_1	MF
VC	−23.89	0.010	−2.360	0.679	−28.87	0.045	12.79	1.000	−652.3	0.000	−430.6	0.176	−224.0	0.107	0.000	1.000
EWMA	−18.14	0.000	−18.55	0.000	−16.54	0.000	16.30	1.000	−370.1	0.000	−331.0	0.090	−231.6	0.000	93.60	1.000
CCC	−17.62	0.000	−1.472	0.850	−9.401	0.140	2.233	1.000	−491.5	0.000	−351.3	0.160	−191.5	0.070	0.000	1.000
G	−16.60	0.010	1.001	0.870	−11.08	0.130	1.413	1.000	−486.4	0.000	−343.9	0.228	−162.3	0.160	0.000	1.000
T	24.60	1.000	0.000	1.000	3.043	1.000	0.000	1.000	117.1	1.000	677.9	1.000	548.1	1.000	0.000	1.000
skT	32.49	1.000	0.000	1.000	−3.146	1.000	0.000	1.000	5.656	1.000	586.7	1.000	511.5	1.000	0.000	1.000
VG	−10.22	0.400	28.47	1.000	−12.83	0.320	17.29	1.000	−453.2	0.020	−88.02	0.595	−50.40	0.479	0.000	1.000
NIG	−4.12	0.750	55.46	1.000	−9.026	0.540	32.40	1.000	−364.7	0.122	248.7	1.000	−231.8	0.188	0.000	1.000
HS	24.61	1.000	0.000	1.000	1.716	1.000	0.000	1.000	−242.7	0.198	451.9	1.000	218.1	1.000	0.000	1.000

Table 6
Superior predictive ability (SPA) backtests.
This table presents the results of SPA tests of 1% and 5% ES estimates for 1-day and 10-day portfolio returns over the out-of-sample period (March 2010 to March 2012) for different parametric multivariate models. Consistent p-values above 0.20 are highlighted in bold.

	Representative utility portfolio								Equally weighted portfolio							
	Short position				Long position				Short position				Long position			
	$h=1$		$h=10$		$h=1$		$h=10$		$h=1$		$h=10$		$h=1$		$h=10$	
	5%	1%	5%	1%	5%	1%	5%	1%	5%	1%	5%	1%	5%	1%	5%	1%
VC	0.064	0.010	0.013	0.018	0.047	0.049	0.117	0.153	0.014	0.023	0.108	0.151	0.132	0.133	0.124	0.146
EWMA	**0.202**	0.043	0.035	0.061	0.014	0.013	0.119	0.188	0.004	0.017	0.105	0.115	**0.256**	**0.217**	**0.376**	**0.250**
CCC	0.051	0.004	0.008	0.024	0.067	0.053	**0.579**	**0.426**	0.001	0.009	0.084	**0.247**	0.142	**0.464**	0.172	0.128
G	0.023	0.005	0.007	0.026	0.065	0.054	**0.577**	**0.432**	0.000	0.008	0.068	**0.249**	**0.180**	**0.429**	0.170	0.141
T	**1.000**	**0.571**	0.112	**1.000**	**1.000**	**1.000**	**0.749**	**0.435**	0.033	**0.693**	**0.915**	**0.746**	**0.670**	**1.000**	**0.885**	**0.773**
skT	**0.830**	**1.000**	**1.000**	**0.716**	**0.426**	**0.852**	**0.740**	**0.438**	**1.000**	**1.000**	**0.840**	**0.751**	**1.000**	**0.463**	**1.000**	**0.776**
VG	**0.251**	0.009	0.014	0.047	0.162	0.084	**0.510**	**0.434**	0.008	0.045	**0.822**	**0.715**	**0.667**	**0.708**	**0.197**	**0.546**
NIG	**0.551**	0.020	0.032	0.119	**0.270**	0.102	**0.499**	**0.424**	0.007	0.103	**1.000**	**0.723**	**0.814**	**0.331**	**0.297**	**0.608**
	Minimum variance portfolio								Gas-fired power plant							
	Short position				Long position				Short position				Long position			
	$h=1$		$h=10$		$h=1$		$h=10$		$h=1$		$h=10$		$h=1$		$h=10$	
	5%	1%	5%	1%	5%	1%	5%	1%	5%	1%	5%	1%	5%	1%	5%	1%
VC	0.012	0.016	0.138	**0.246**	0.062	0.152	0.129	0.158	0.028	0.012	0.014	0.021	0.022	0.066	0.144	0.142
EWMA	0.005	0.007	0.087	0.092	0.045	**0.449**	**0.455**	**0.510**	**0.201**	0.041	0.040	0.082	0.010	0.034	0.108	0.101
CCC	0.001	0.006	0.053	**0.272**	0.034	**0.789**	0.180	0.153	0.002	0.004	0.011	0.030	0.005	0.054	**0.618**	**0.422**
G	0.000	0.005	0.055	**0.276**	0.030	**0.794**	0.174	0.160	0.003	0.004	0.012	0.031	0.006	0.076	**0.617**	**0.415**
T	0.032	**0.219**	**1.000**	**0.708**	**1.000**	**1.000**	**0.840**	**0.713**	**0.516**	**1.000**	**1.000**	0.763	**1.000**	**0.683**	**0.676**	**0.420**
skT	**1.000**	**1.000**	**0.776**	**0.700**	**0.666**	**0.452**	**1.000**	**0.709**	**1.000**	**0.419**	0.158	1.000	**0.376**	**1.000**	**0.686**	**0.420**
VG	0.002	0.023	**0.619**	**0.735**	0.154	**0.627**	0.188	**0.657**	0.169	0.006	0.019	0.052	0.014	0.108	**0.502**	**0.428**
NIG	0.008	0.065	**0.726**	**0.747**	**0.313**	**0.709**	**0.417**	**0.583**	**0.263**	0.012	0.054	**0.261**	0.126	0.102	**0.476**	**0.431**

is the best forecasting model in terms of the loss function is rejected for small p-values (see Hansen, 2005 and Hansen and Lunde, 2005 for more details).

Because the ES measure describes overall tail risk behavior better, we decide here to focus on an ES-based loss function.[14] In particular, we consider a linear–linear (or "lin–lin") loss function $L\left(\widehat{\mathrm{ES}}_t(\alpha,h)\right)$ that penalizes events below the ES estimate more than those for which actual losses do not exceed the ES forecast, such that

$$L\left(\widehat{\mathrm{ES}}_t(\alpha,h)\right)=\left(\alpha-1_{\left\{V_1\left(\widehat{\mathrm{ES}}_t(\alpha,h)\right)<0\right\}}\right)V_1\left(\widehat{\mathrm{ES}}_t(\alpha,h)\right), \qquad (17)$$

where $V_1\left(\widehat{\mathrm{ES}}_t(\alpha,h)\right)$ represents the backtesting measure of Embrechts et al. (2005).

Employing the loss function proposed in Eq. (17), we report in Table 6 the consistent p-values corresponding to the SPA tests of the 1% and 5% ES estimates for 1-day and 10-day (long and short) portfolio losses. We include in the analysis the GH models and the parametric benchmark models. Any p-values greater than 0.20 are highlighted in bold, which indicates that the null hypothesis for the corresponding benchmark is not rejected. According to these results, we reject the claim that the (conditional and unconditional) Gaussian models, such as VC, EWMA, CCC, and G, perform as well as the best competing alternative model, with the possible exception of the long portfolio positions at 10-day horizons. The SPA tests support our previous findings, namely that models with exponential tail decay (i.e., VG and NIG models) yield inferior tail estimates for short portfolio positions, especially for the far tail ($\alpha=1\%$) of utility portfolios at the 1-day horizon. The p-values of the T and skT models are close to one for most portfolio positions, so we know that the polynomial tail decay is not outperformed by other tail patterns we have considered. In addition, for short positions, $\alpha=1\%$, and a 1-day horizon, these models are the only SPA benchmarks for which we do not reject the null hypothesis. In particular, at both confidence levels ($\alpha=5\%,1\%$), we cannot reject the skT model for any short or long position.

7. Conclusion

In this article, we have characterized the tail behavior of energy price risk using a dynamic multivariate model. We approximated exposure to energy price risk for physical and financial players using linear combinations (portfolios) of crude oil, natural gas, coal, and electricity futures. To model the stylized features of the vector of energy risk factors, we have proposed a flexible econometric specification with time-varying conditional mean, variance, and correlation, which accommodates the possible presence of serial dependence in returns, heteroskedasticity, and leverage effects. With respect to the conditional distribution, we considered the possibility that the vector of innovations may be generated by a multivariate GH distribution, which contains as particular cases some popular distributions, such as the NIG, the VG, the skewed t, Student's t, and the Gaussian distributions. With these distributions, we can model different dependence patterns (e.g., dependence in the extremes, positive or negative skewness) and tail decays (e.g., exponential vs. polynomial).

Our empirical application featured daily data from August 2005 to March 2012 related to energy futures traded in the NYMEX. We reserved the observations from March 2010 to March 2012 for our out-of-sample analysis. Thus, using the recursive estimates of the multivariate GH models, we calculated the conditional risk measures corresponding to four prespecified energy portfolios. Then, we evaluated the performance of those sequences of risk measure forecast over the out-of-sample period. Our in-sample and out-of-sample results showed the importance of fat tails and positive skewness in the multivariate distribution of energy risk factors. We also proposed comparing the tail risk estimates corresponding to the GH models and other more traditional procedures, by applying a test of superior predictive ability (SPA). Regarding the tail risk of short positions, our SPA backtest results confirmed that distributions with polynomial tail decay (heavy-tailed)

[14] We also take into account other loss functions, based on both VaR and ES forecasts, but we do not report them here.

outperformed alternative versions, especially for the utility portfolios. The distributions with exponential tail decays (Gaussian and semi-heavy-tailed) behaved reasonably well for long positions and longer horizons. Ultimately, the extent to which we underestimate the tail risk of the portfolio loss distribution depends on the portfolio weights of the different energy commodities, whether we are analyzing the short or long trading position, and the horizon and confidence level considered.

It is worth mentioning that many power firms in liberalized markets have two main lines of business: electricity generation and electricity distribution. These days, and given the chronic generation overcapacity afflicting many developed markets (United States, Europe) most firms tend to focus more on the distribution business which implies an aggregate short position in electricity. The evidence we present suggests that conventional market risk measures (Gaussian VaR and ES) severely underestimate market risk under these circumstances. This fact should be taken into account not only by the company's shareholders and creditors but also by market regulators and supervisors.

Some questions arise for further research. First, we did not consider the effects of parameter uncertainty in the calculations of the tail risk measures, and it would be interesting, albeit computationally intensive, to study the impacts on the results if we were to take such uncertainty into account. Second, we characterized the aggregate tail risk using prespecified energy price risk exposures, given by the portfolio weights, of various representative energy-market players. However, an advantage of our asset-level approach is that we can analyze the sensitivity of the tail risk measures to changes in the weights of the energy portfolio. Furthermore, we can use these multivariate approaches to determine how the risk measures we have analyzed might be used to construct an optimal energy portfolio. Finally, it would be interesting to compare the GH models with other parametric and semi-parametric multivariate approaches, such as those related to multivariate extreme value theories (e.g., Poon et al., 2004). We leave these questions for future analysis.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.eneco.2014.05.004.

References

Aas, K., Haff, I.H., 2006. The generalized hyperbolic skew Student's t-distribution. J. Financ. Econ. 4 (2), 275–309.

Artzner, P., Delbaen, F., Eber, J.M., Heath, D., 1999. Coherent risk measures. Math. Financ. 9 (3), 203–228.

Benth, F., Šaltytė Benth, J., 2004. The normal inverse Gaussian distribution and spot price modelling in energy markets. Int. J. Theor. Appl. Finance 7 (2), 177–192.

Benth, F., Šaltytė Benth, J., Koekebakker, S., 2008. Stochastic Modelling of Electricity and Related Markets. World Scientific Publishing, New York, NY.

Bibby, B., Sørensen, M., 2003. Hyperbolic processes in finance. In: Rachev, S. (Ed.), Handbook of Heavy Tailed Distributions in Finance. Elsevier/North-Holland, Amsterdam, pp. 211–248.

Bingham, N., Kiesel, R., 2002. Semi-parametric modelling in finance: theoretical foundations. Quant. Finance 2 (4), 241–250.

Boerger, R.H., Cartea, A., Kiesel, R., Schindlmayr, G., 2009. Cross-commodity analysis and applications to risk management. J. Futur. Mark. 29 (3), 197–217.

Cappiello, L., Engle, R., Sheppard, K., 2006. Asymmetric dynamics in the correlations of global equity and bond returns. J. Financ. Econ. 4 (4), 537–572.

Cartea, A., Figueroa, M.G., 2005. Pricing in electricity markets: a mean reverting jump diffusion model with seasonality. Appl. Math. Finance 12 (4), 313–335.

Cartea, Á., González-Pedraz, C., 2012. How much should we pay for interconnecting electricity markets? A real options approach. Energy Econ. 34 (1), 14–30.

Christoffersen, P.F., 1998. Evaluating interval forecasts. Int. Econ. Rev. 39 (4), 841–862.

Eberlein, E., Stahl, G., 2003. Both sides of the fence: a statistical and regulatory view of electricity risk. Energy Power Risk Manage. 8 (6), 32–36.

Embrechts, P., Kaufmann, R., Patie, P., 2005. Strategic long-term financial risks: single risk factors. Comput. Optim. Appl. 32 (1), 61–90.

Escribano, A., Peña, J.I., Villaplana, P., 2011. Modelling electricity prices: international evidence. Oxf. Bull. Econ. Stat. 73 (5), 622–650.

Giot, P., Laurent, S., 2003. Market risk in commodity markets: a VaR approach. Energy Econ. 25 (5), 435–457.

Hansen, P.R., 2005. A test for superior predictive ability. J. Bus. Econ. Stat. 23 (4), 365–380.

Hansen, P., Lunde, A., 2005. A forecast comparison of volatility models: does anything beat a GARCH(1, 1)? J. Appl. Econ. 20 (7), 873–889.

Hu, W., Kercheval, A.N., 2010. Portfolio optimization for t and skewed-t returns. Quant. Finance 10 (1), 91–105.

Huisman, R., Mahieu, R., 2003. Regime jumps in electricity prices. Energy Econ. 25 (5), 425–434.

International Energy Agency, 2011. Key World Energy Statistics. OECD, Paris.

Kupiec, P., 1995. Techniques for verifying the accuracy of risk measurement models. J. Deriv. 3, 73–84.

Mcneil, A.J., Frey, R., 2000. Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. J. Empir. Finance 7 (3), 271–300.

Mcneil, A.J., Frey, R., Embrechts, P., 2005. Quantitative Risk Management: Concepts, Techniques, and Tools. Princeton University Press, Princeton, NJ.

Pirrong, C., 2012. Commodity Price Dynamics: a Structural Approach. Cambridge University Press, New York, NY.

Politis, D., Romano, J., 1994. The stationary bootstrap. J. Am. Stat. Assoc. 89 (428), 1303–1313.

Poon, S.H., Rockinger, M., Tawn, J., 2004. Extreme value dependence in financial markets: diagnostics, models, and financial implications. Rev. Financ. Stud. 17 (2), 581–610.

Prause, K., 1999. The Generalized Hyperbolic Model: Estimation, Financial Derivatives, and Risk Measures. (Doctoral Thesis) University of Freiburg.

Routledge, B., Seppi, D., Spatt, C., 2001. The spark spread: an equilibrium model of cross-commodity price relationships in electricity. Working Paper. Carnegie Mellon University.

Tseng, C., Barz, G., 2002. Short-term generation asset valuation: a real options approach. Oper. Res. 50 (2), 297–310.

Weron, R., 2006. Modeling and Forecasting Electricity Loads and Prices: a Statistical Approach. Wiley, Chichester.

White, H., 2000. A reality check for data snooping. Econometrica 68 (5), 1097–1126.