Generating joint scenarios of renewable energy sources: the case for non-Gaussian models with time varying parameters

Henrique Hoeltgebaum, Cristiano Fernandes and Alexandre Street *Member IEEE*

***Abstract*—The development of medium- and long-term studies for power-system planning under uncertainty of renewable generation is one of the key challenges faced by planners and market players in most of the power system worldwide. The characterization wind power generation (WPG) stochastic processes to devise time- and spatial-dependent scenarios, based on simulation procedures, for one to few years horizon is a difficult task. Multiple regimes and non Gaussian distributions are one of the main issues faced in both centralized and private planning/investment studies. In this paper a new methodology to simulate long-term joint scenarios for time series of several wind power plants generation is presented. The proposed framework is derived based on a new class of time-series model with time-varying parameters and arbitrary non Gaussian distribution, known as Generalized Auto Regressive Score models (GAS). GAS models are conceived based on a user-defined marginal probability density function (PDF), providing a relevant flexibility and control of the simulated scenarios in long-run chronological Monte Carlo simulation procedures. In this work, we study the case of a Beta PDF and spacial dependence is captured by a t-Student copula with time varying correlation matrix. Case studies based on the Brazilian power system show that the proposed methodology is capable to address relevant issues that arise in long-term simulation studies.**

***Index Terms*—Score driven models, wind power, probabilistic forecasting, time varying parameters, dynamic copula.**

I. Introduction

- introduzir contribuiceos.
- melhorar a parte da copula.
- introducao da parada fuzzy nao Ã© muito besta?.

RENEWABLE energy expansion has been growing worldwide, mainly in response to governmental incentive for reducing greenhouse gas emissions. In particular, wind-power generation (WPG) is one of the largest sources of renewable energy, and according to the International Energy Agency, will respond to 18% of global power by 2050 [1]. However the uncertainty associated with its non dispatchable nature may jeopardize reliability of electricity supply. In attempting to minimize this kind of risk it is highly desirable to produce reliable forecasts and scenarios for WPG time series. The importance of such scenarios emerges in many instance, as for example: (i) energy trading, (ii) unit commitment, (iii) grid expansion planning, and (iv) investment decisions (see ([2]–[5]) and references therein).

Much research has been devoted to devise short-term forecast methods for wind speed and WPG time series (for instance, see [6]–[19] and [20] for a survey on probabilistic forecasting of wind power generation). In many applications, the modeling of these time series is conducted using conventional ARMA models with seasonal lags, or SARIMA models, under a Gaussian distribution ([21], [22]). In its original formulation, SARIMA models present a fixed scale parameter. To add extra flexibility to these models [23] proposes an ARIMA model with GARCH effect, allowing the conditional variance of wind power distribution to vary over time. Also, to provide a description of the spatial correlation of wind-speed time series, [11] proposed an Autoregressive Fractionally Integrated-GARCH model (ARFIMA-GARCH). In these and other models, conditional on the past, the distribution of the response variable is assumed Gaussian.

The non-Gaussian nature of WPG time series is well reported in the literature (see [13], [24]–[26], and references therein). Notwithstanding, only very limited classes of non-Gaussian time-series models were available for forecasting WPG. The state of the art literature on short-term probabilistic forecast of WPG mainly rely on nonparametric models. For instance, in [13] and [26], nonparametric approaches were devised to forecast conditional distributions. While in [13] kernel density estimators was employed to build a density forecast, in [26] a quantile-regression-based model was used. Despite the computational virtues found in estimation for non-parametric methods, the development of new parametric models capable to properly characterize the full non-Gaussian distribution for WPG time series is a relevant research theme, which did not received too much attention so far. Relevant applications arise from risk assessment in both medium- and long-term planning/investment studies, which mainly rely on few data points to characterize conditional estimates of extreme quantiles [5], [21], [27]–[29] , Outras referencias que nao as nossas.

In [30], a General Autoregressive Score (GAS) model was proposed to derive and estimate time series models with any non-Gaussian conditional distribution, either discrete or continuous, univariate or multivariate. In such paper, it has been shown that several well-known time-series models from the

This work was partially supported by the R&D project ANEEL PD-7625-0001/2013.

Henrique Hoeltgebaum, Cristiano Fernandes and Alexandre Street are with the Electrical Engineering Department, Pontifical Catholic University of Rio de Janeiro (PUC-Rio), Rio de Janeiro, RJ, Brazil (e-mail: hhhelfer@ele.puc-rio.br; cris@ele.puc-rio.br; street@ele.puc-rio.br).

econometric literature are a particular case of GAS models[1]. More specifically, a GAS model is built based on a user-defined conditional density function whose parameters follow a data-driven dynamic equation, generally on the form of an autoregressive mechanism. Then, by construction, time-varying parameters can be accommodated according to an updating mechanism that uses the score as its driving force. The use of the score function for updating time-varying parameters is intuitive given that it is defined as the steepest ascent direction (gradient) for improving the local fit of the model in terms of the likelihood. In such an updating mechanism, information from the whole density is used to track the evolution of time varying parameters.

Several applications from finance to oil using GAS models have been published in the recent econometric literature. We refer the interested reader to the main on-line repository on GAS papers [34]. The virtue of GAS models to address non-linear and non-Gaussian time series models recognized in the recent literature notwithstanding, to the best of the authors' knowledge, there is no publication in the power systems literature using GAS models. The consideration of a seasonal structure in such models was also not addressed in the literature so far, despite being a relevant structure shared by a wide range of time series such as WPG (see [21], [22], [35], [36] and references therein). Additionally, given the empirical evidence of positive correlation between WPG time series, it is crucial to incorporate such information when simulating joint scenarios. More commonly, such dependence is usually captured by the covariance matrix Σ [37]. For example, [38] uses a multivariate ARMA model and employed the variance-covariance matrix to characterize the spatial-temporal correlation of wind speed distribution. Under the multivariate assumption, a Vector Autoregressive Moving Average-GARCH model was employed to describe the bivariate wind vector time-series [24].

In our work, the dependence structure among the WPG time series of different wind plants is captured through a elliptical copula with time-varying correlation matrix [39]. Copula-based models are very flexible to construct multivariate distributions, since they allow the specification of the models for the marginals distributions (in our case the beta GAS models) separately from the dependence structure given by the copula. Details on estimation and inference for copulas applied to time series can be seen in [40]–[45]. The adoption of a time-varying correlation matrix for the copula is justified by the fact that correlation among wind plants time series are not constant over time due to the fact that wind have different regimes depending on seasons. In our set up the correlation matrix will also evolve in time according to an appropriate GAS mechanism, which will makes it easier to capture and correct for outlier effects, as presented in [39].

The objective of this paper is to propose a new methodology to address non-Gaussian WPG multivariate time-series through GAS models. We also propose a time-varying copula-based approach to model spatial dependencies in a two-step procedure. Therefore, the contribution of this paper are threefold:

- A new model capable to address non-Gaussian seasonal wind power generation time series. The proposed model allows the simulation of long-run scenarios widely demanded in many power system applications, such as planning and risk assessment studies.
- Spatial dependency is derived from a time-varying copula model with correlation matrix guided by second-step GAS procedure.
- , O QUE MAIS??.

The rest of this paper is organized as follows. In section II, the proposed GAS model framework with seasonal dynamics in the updating mechanism is presented. Estimation, diagnoses and forecast procedures are also discussed in this section. In section III, an elliptical-copula model is devised based on [39] to model the time varying correlation matrix through a second-step procedure also following a GAS mechanism. Finally, in section IV, a case study using realistic data from the Brazilian power system is presented to illustrate the application of the proposed framework and in section V, final conclusions are addressed.

II. Score driven models

Our time series data comprises of monthly measures of wind capacity factors (CF), which can only assume values in the range [0,100), indicating at the outset, the non adequacy of models with Gaussian distribution. In addition, such time series display a strong seasonal pattern derived from the inter annual variation of wind speed. These data features suggest that a proper model for CF series has to be non Gaussian with parameters that change in time, so that short term and seasonal variations can be properly captured. Firstly, one should choose an adequate density to each of the wind CF time series. After trying several non Gaussian densities the best candidate was the beta. More precisely, in our application the CF time series y_t is properly described by a beta density as follows,

$$p(y_{it}|f_{it}, \mathcal{F}_{t-1}; \theta) \sim Beta\left(\beta_{it}, \alpha_i\right), \ \forall \ i \in I \text{ and } t \in T, \quad (1)$$

where y_{it} is the wind CF of the wind plant i at time t, $\beta_{it} = h(f_{it})$, $\alpha_i > 0$, with $h(\cdot)$ to be chosen.

To fully specify a GAS model one has to choose which parameters of the distribution will evolve in time and those that will be fixed. The time varying parameters will then follow a GAS(p,q) updating mechanism. Considering the seasonal pattern of CF time series the appropriate form for this mechanism will be given by:

$$\begin{aligned} f_{i,t+1} = \omega_i &+ A_{i,1}s_{i,t} + A_{i,2}s_{i,t-1} + A_{i,3}s_{i,t-2} \\ &+ A_{i,11}s_{i,t-10} + A_{i,12}s_{i,t-11} \\ &+ B_{i,1}f_{i,t} + B_{i,2}f_{i,t-1} + B_{i,3}f_{i,t-2} \\ &+ B_{i,11}f_{ti,-10} + B_{i,12}f_{i,t-11} \end{aligned} \quad (2)$$

where $f_{i,t}$ is the time varying parameter of the beta density of the wind plant i at time t, $s_{i,t}$ is the score of the beta density of the wind plant i at time t and the ω's, A's and B's are fixed parameters which will be estimated by maximum likelihood.

[1] For instance, GARCH models that deal with heavy tailed distributions in finance [31], Autoregressive Conditional Duration models [32] to tackle asymmetric distributions in finance applications such as time duration, and the Poisson count models of Davis [33] are particular cases of GAS models.

To complete the description of the the updating mechanism of GAS models it is necessary to define s_t in Equation (2). The authors in [30] propose the following scheme:

$$s_t = \mathcal{I}_{t|t-1}^{-d} \cdot \nabla_t \tag{3}$$

$$\nabla_t = \frac{\partial \ln p(y_t|f_t, \mathcal{F}_{t-1}; \theta)}{\partial f_t} \tag{4}$$

$$\mathcal{I}_{t|t-1} = E_{t|t-1}[\nabla_t^T \nabla_t], \tag{5}$$

where $p(y_t|f_t, \mathcal{F}_{t-1}; \theta)$ is the observation density function, $\mathcal{F}_{t-1}$ collects all relevant information up to time $t-1$, and $\mathcal{I}_{t|t-1}$ is a scaling matrix with appropriate dimensions. Different choices of d results in different GAS models: $d = 1$ means that it will be use the inverse of Fisher Information matrix, $d = 1/2$ the pseudo-inverse of Fisher Information matrix $d = 0$ the identity matrix. More details can be found in [30]. In practice, the choice of d has shown to be an empirical question: for a given model one chooses the value of d which produces best diagnostics and forecasting.

A. Parametrization

Having in mind that the first shape parameter of the beta density, $\beta_t > 0$, it is natural to choose $\beta_t = exp\{f_t\}$. From this, it follows that Equations (1)-(5) should also be re-parametrized with regard to $\ln(\beta_t)$. Considering a monotonically increasing mapping function $h(\cdot)$, then $\tilde{f}_t = h(f_t)$. Taking $\dot{h}_t = \partial h(f_t)/f_t^{'}$, the beta GAS specification updating mechanism, will be

$$\begin{aligned} \dot{h}_t &= \frac{1}{\beta_t} \\ \tilde{\nabla}_t &= \beta_t \left\{ \ln y_t - [\ln k + \psi_1(\beta_t) - \psi_1(\beta_t + \alpha)] \right\} \qquad (6) \\ \tilde{\mathcal{I}}_{t|t-1} &= \left\{ \beta_t^2 [\psi_2(\beta_t + \alpha) - \psi_2(\beta)] \right\}^{-1} \\ \tilde{s}_t &= \frac{\ln y_t - [\ln k + \psi_1(\beta_t) - \psi_1(\beta_t + \alpha)]}{\beta_t^{1-2d}[\psi_2(\beta_t) - \psi_2(\beta_t + \alpha)]^{1-d}}, \qquad (7) \end{aligned}$$

where ψ_i stands for the i-th order derivative of the logarithm of the gamma function.

B. Initial values

In practice, to initialize recursion (2) at $t = 1$, it is necessary to have lagged values of the time varying parameter f_t. As s_t is a direct function of f_t, it is straightforward to calculate its lagged values. Then we will concentrate on deriving lagged values for f_t. Denote the initial time varying parameters values as $f_{i,1-q} = \{f_{i,0}, f_{i,-1}, ..., f_{i,1-q}\}$, where q denotes the persistence of the value f_t over time, as shown by equation (2). In our particular case, the wind CF time series is split into 12 monthly time series, i.e., $\{Y_{Jan}^{(i)}, ..., Y_{Dec}^{(i)}\}$. For each of these time series $\{Y_m^{(i)}\}_{m=1}^{12}$, estimate via maximum likelihood the static shape parameters of a beta density, α_i and β_i. Then use these estimates to initialize Equation (2), having in mind the new parametrization, $f_{i,t} = \ln(\beta_{i,t})$. Note that using this methodology, the first twelve observations of the time series are removed in order to estimate the initial values of the time varying parameter of the beta density. As a result, the maximum likelihood and the residuals of the beta GAS(p,q) model are only computed after the thirteenth observation.

C. Diagnostics

In GAS models diagnostics can be evaluated using an appropriate type of residuals for non linear/ non Gaussian time series models known as quantile residuals. As described in [46], the observed quantile residual is given by

$$r_{t,\hat{\theta}} = \Phi^{-1}(F(y_t|f_t, \mathcal{F}_t, \hat{\theta})). \tag{8}$$

where $F(\cdot)$ is the CDF associated with $p(y_t|f_t, \mathcal{F}_{t-1}; \theta)$ and Φ^{-1} is the quantile of a standard Gaussian distribution. Under correct model specification these residuals should be normally distributed and show no dependence. These can be checked, for example, using the Jarque Bera test for normality, and the Ljung Box test for absence of serial correlation and Ljung Box (on squared residuals) to test for absence of ARCH effect.

D. Univariate Forecasting

The k steps ahead distribution conditional on observations up to time t, $p(y_{t+k}|\mathcal{F}_t)$, is only analytical for $k = 1$, when it coincides with the chosen probability model. However, for $k > 1$ it has to be evaluated using Monte Carlo simulation.

III. Dependence model

In the previously section, non Gaussian time series models were proposed to model each of the wind CF time series. To derive a proper mechanism for joint scenario generation of CF's time series it is also necessary to capture the observed dependence among the CF's time series. For this conditional copula models introduced by [45] are used.

More specifically, using the framework developed in [42], it will be possible to allow the correlations between pairs of CF's time series to change in time according to a GAS mechanism that will be described in the sequel. As it will be seen, it is straightforward to combine marginal densities as given by GAS models with copulas that also have time varying parameters.

Considering each of the time series models y_{it} estimated individually using a GAS model, u_{it} denotes the probability integral transform (PIT) variable obtained from the marginal beta GAS(p,q) model described in section II. That is

$$u_{it} = F_{beta}(y_{it}|f_{i,t}, \mathcal{F}_{i,t}, \hat{\theta}_i) \ \forall \ i \in I \text{ and } t = 13, \ldots, T \tag{9}$$

where $\hat{\theta}_i = \{\hat{\alpha}_i, \hat{\omega}_i, \hat{A}_i's, \hat{B}_i's\} \ \forall \ i = 1, 2, 3$ is the estimated vector of fixed parameters from the univariate beta models. The driving mechanism presented in [39], was derived for a multivariate Student t, being the Gaussian density a particular case of this when $\nu^{-1} \rightarrow 0$. Such density then acts on the $T \times I$ vector $\tilde{y}_t^{'} = [F_\nu^{-1}(u_{1t}), ..., F_\nu^{-1}(u_{it})]^{'}$, where $F_\nu^{-1}(\cdot)$ is the univariate inverse Student t distribution with ν degrees of freedom. The updating of the correlation matrix through a GAS mechanism follows a similar idea from the dynamical conditional correlation (DCC) model proposed by [47], and commonly used in Finance. The updating mechanism of the time-varying parameter vector, proposed for this particular density in [39], is $f_t = vech(Q_t)$, where Q_t is a symmetric positive definite matrix, carrying all the information regarding the correlation among the wind plants CF's time series. For details, see [39]. , Melhorar essa parte do detalhamento das copulas com GAS

A. *Estimation of the copula parameters*

Estimation of the copula parameters are done using multi-stage optimization approach known as Inference for Margins (IFM) ([48], [49]). In IFM estimation, the marginal parameters are first estimated, and then conditional on these estimates, copula parameters are estimated via maximum likelihood. In IFM estimation the unknown parameter vector is split in two sets, one containing the marginal densities parameters and the other the copula parameters. A suitable justification for such approach is presented in [41] and shall not be discussed here.

IV. CASE STUDY - SIMULATING LONG-TERM JOINT SCENARIOS

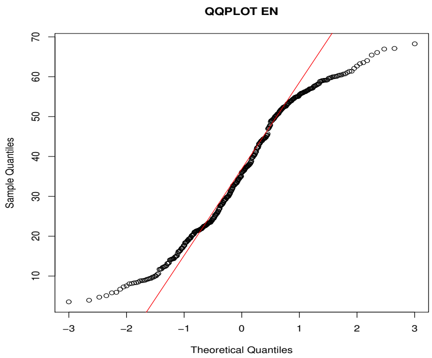

Fig. 1. QQ-plot of EN monthly capacity factor time series ranging from January 1981 to December 2011.

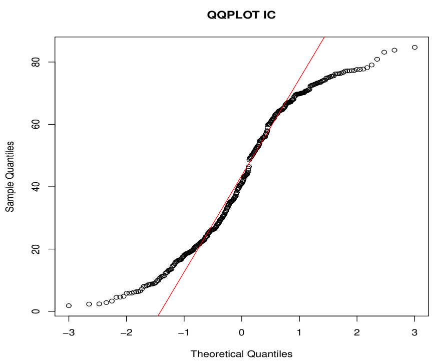

Fig. 2. QQ-plot of IC monthly capacity factor time series ranging from January 1981 to December 2011.

Our time series data comprises monthly wind CF, from January 1981 to December 2011, measured at three wind plants located in the Brazilian Northeast, namely: Rio do Fogo (RF), Icaraizinho (IC) and Enacel (EN). The last year of these time series were kept for out of sample forecasting evaluation.

Firstly, parameters of the beta GAS(p,q) models are estimated via maximum likelihood. To accommodate outliers, dummy variables (d_t) were used where appropriate (for details see Table I). In this study, the same lag structure was used for the three wind plants time series, as presented in Equation (2). After removing the first and last years of data for initial values estimation and forecasting evaluation, maximization of the likelihood function resulted in the estimated parameters presented in Table I.

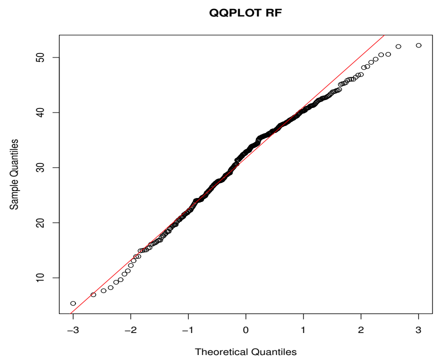

Fig. 3. QQ-plot of RF monthly capacity factor time series ranging from January 1981 to December 2011.

TABLE I
MAXIMUM LIKELIHOOD ESTIMATION OF THE BETA GAS(12,12) MODEL APPLIED TO THREE BRAZILIAN WIND PLANTS.

Parameter	Rio do Fogo (RF)		Icaraizinho (IC)		Enacel (EN)	
	Estim.	S.E.	Estim.	S.E.	Estim.	S.E.
A_1	0.551	0.006	0.539	0.071	0.540	0.060
A_2	0.231	0.004	-0.226	0.066	0.498	0.073
A_3	-0.147	0.007	0.352	0.070	0.041	0.053
A_{11}	0.008	0.007	0.219	0.038	0.169	0.034
A_{12}	0.110	0.004	-0.176	0.043	0.228	0.031
B_1	-0.080	0.002	1.058	0.055	-0.164	0.039
B_2	0.441	0.006	-0.818	0.043	-0.186	0.036
B_3	-0.369	0.011	0.156	0.031	-0.008	0.016
B_{11}	0.207	0.013	0.227	0.016	-0.004	0.005
B_{12}	0.325	0.011	0.121	0.037	0.721	0.035
ω	3.139	0.019	2.571	0.260	5.485	0.780
α	9.408	0.042	10.244	0.791	9.293	0.515
$d_{t=21}^{(Set,1983)}$	-	-	2.320	0.287	1.447	0.247
$d_{t=22}^{(Oct,1983)}$	-	-	2.021	0.309	3.610	0.481
$d_{t=29}^{(May,1984)}$	-	-	-0.031	0.641	-	-
$d_{t=129}^{(Sep,1992)}$	-	-	3.349	0.317	1.675	0.386
$d_{t=130}^{(Oct,1992)}$	-	-	-	-	1.189	0.345
$d_{t=165}^{(Sep,1995)}$	-	-	1.460	0.264	-	-
$d_{t=241}^{(Jan,2002)}$	-	-	-	-	-1.527	0.582
$d_{t=308}^{(Aug,2007)}$	1.771	0.346	-	-	-	-
$d_{t=309}^{(Sep,2007)}$	2.384	0.369	-	-	-	-
$d_{t=345}^{(Sep,2010)}$	3.004	0.277	-	-	-	-
$d_{t=346}^{(Oct,2010)}$	-	-	-1.462	0.350	-	-

Finally, to investigate correct model specification Table II report the results for the Jarque-Bera test for normality (under the correct specification of the beta density, the residuals should be normally distributed), Ljung-Box test for absence of autocorrelation and Ljung-Box on the squared residuals to check for ARCH effect (both were conducted using 30 lags).

TABLE II
P-VALUES OF STANDARD DIAGNOSTIC TESTS.

Test	**Rio do Fogo**	**Icaraizinho**	**Enacel**
Normality	0.242	0.582	0.386
Autocorrelation	0.174	0.854	0.127
Arch effect	0.703	0.479	0.175

From Table II, it can be concluded that the proposed beta

GAS(12,12) models are adequate to describe the three time series of CF's. For forecasting evaluation the prediction of the GAS models were compared to those obtained from the fitting of a SARIMA model to the logarithm of the CF time series. To obtain the prediction for the original CF's series these values have to be exponentiated. As it is known, exponentiation of Gaussian values can result in unrealistically large values for the original variable. This will not be the case with GAS models since from the outset the chosen probability model is appropriate to the support of the response variable. To provide some evidence on the distortion brought by the anti log transformation, as an evidence, we fitted a Gaussian SARIMA model to the log of the three CF time series wind plants consired here (EN,IC,RF), and simulated 2000 scenarios 12 steps ahead. Then we compared the empirical quantiles obtained from the simulated CF series to the quantiles of the original CF series (both in the original scale) $Q^{(\alpha\%)}\,\forall\,\alpha \in \{0.05, 0.10, 0.5, 0.9, 0.95\}$ which are depicted in Figure 5. As a result, the simulated quantiles are far beyond the historical quantiles of the CF time series, indicating the distortions brought by fitting the model to the log of the series.

The k steps ahead distribution conditional on observations up to time t, $p(y_{t+k}|\mathcal{F}_t)$, used to obtain out of sample forecasts, has no analytical form in GAS models, and therefore is estimated using Monte Carlo simulation. We produced 2000 scenarios for each of the k ahead forecasting horizon $k = 1, 2, ..., 12$. Goodness of fit results as shown in Table III indicates that our model seems competitive when compared to the benchmark SARIMA model.

TABLE III
FORECASTING EVALUATION BETWEEN BETA GAS(12,12) AND SARIMA MODEL.

Modelo	Medidas	RF	IC	EN
beta GAS(12,12)	RMSE	5.639	6.181	8.736
	MAE	5.209	5.021	7.346
	Pseudo R^2	0.834	0.941	0.875
SARIMA	RMSE	7.566	10.668	9.737
	MAE	6.263	7.675	8.222
	Pseudo R^2	0.809	0.831	0.859

The main contribution of the proposed framework here is not to improve the conditional mean forecast. The methodology is devote to improve the modelling of the join density of wind CF, providing accurate scenarios that take into account the dependency structure among different geographical areas.

In the following we will provide details on the multivariate modeling of the three CF time series using the dynamic copula approach previously described in Section III. After estimating the fixed parameter vector $\hat{\theta}_i$ of each marginal $i \in I$, PIT variables u_{it} are generated using the whole set of information up to time t, $\mathcal{F}_t$. Then, in order to estimate the copula parameters, the multi-stage optimization approach, IFM is used. The degrees of freedom of the copula model was estimated using the profile likelihood process (see [50]), resulting in $\hat{\nu} = 340$ is the value that maximizes the log likelihood, which means that the proposed elliptical copula reduces itself to a time-varying Gaussian copula.

The model procedure involving the elliptical copula presented in Section III was applied to the $T \times I$ vector $\tilde{y}_t^{'} = [F_\nu^{-1}(u_{1t}), F_\nu^{-1}(u_{2t}), F_\nu^{-1}(u_{3t})]^{'}$. From now on, this will be referred as t-GAS(p,q). In our application, the updating mechanism of the copula parameter with $p = 1$ and $q = 1$ was shown satisfactory. Table IV. presents the results for the estimation of the parameters for the t-GAS(1,1) model.

TABLE IV
MAXIMUM LIKELIHOOD ESTIMATION OF THE T-GAS(1,1) MODEL APPLIED TO THREE BRAZILIAN WIND PLANTS.

Parameter	**t-GAS**	
	Estim.	S.E.
ω	1.589	0.063
A	0.354	0.017
B	0.425	0.083

From Table IV, it follows that all parameters are statistically significant at 5% or less which implies that the correlation of the elliptical copula is indeed time-varying. The value of the parameter A, in the copula model, reinforce the fact that the association between these three wind plants might not be static over the analyzed period. It is important to mention that such feature will be contemplated when simulating values from the joint density through our proposed methodology.

Now, in order to introduce the dependence captured by the elliptical copula model in the joint scenarios of wind CF, the following steps should follow. First forecast the PIT variables 12 steps ahead using the t-GAS (1,1) model, producing m path 12 steps ahead of $\{\mathbf{u}_{it}^m\}_{t=T+1}^{T+12}$. With such values in hands follow the steps in the sequel.

1) Call $m = 1$, i.e., $\{\mathbf{u}_t^{(1)}\}_{t=t+1}^{t+12}$, then update Σ_t with $\hat{\theta}^{cop}$;
2) Use a random number generator to sample from a Student t density with the correlation matrix updated in the aforementioned step and the estimated degrees of freedom, $\hat{\nu} = 340$;
3) Apply the conditional sampling technique (see [40]), in order to generate conditioned pairs of PIT variables, i.e., $\{\mathbf{u}_t^{*(1)}\}_{t=t+1}^{t+k}$;
4) Since it is a fully parametric copula model, marginal densities are assumed to be are known. Calculate the quantile of a beta density function evaluated at $\{\mathbf{u}_t^{*(1)}\}_{t=t+1}^{t+k}$, where the first shape parameter $\beta_{i,t} = exp\{f_{i,t}\}$ was already calculated to generate the forecasting set S;
5) Repeat the above steps until $m = 2000$.

By following this procedure, one obtains 2000 replicates for each of the k steps ahead forecasting, $k = 1, 2, ..., 12$, for CF values. Denote this model beta *t-GAS*. Finally, to evaluate the quality of the simulated values, we propose the same comparison between the empirical quantiles obtained from the simulation with those obtained from the real data. In order to accomplish that, as done in the case of initial values for the beta GAS(p,q) model, first we segment the wind CF time series of wind plant $i \in I$ into 12 monthly time series, i.e., $\{Y_{Jan}^{(i)}, ..., Y_{Dez}^{(i)}\}$. The same is applied to the simulated sets from model beta *t-GAS*. The quantiles $Q^{(\alpha\%)}\,\forall\,\alpha \in \{0.05, 0.10, 0.5, 0.9, 0.95\}$ of each set were calculated and plotted in Figure 4, where the lines are the monthly quantiles of the simulated set from one of the two models, while the geometric figures represent the monthly quantiles of the real data set. We also display in Figure 5 the same analysis using SARIMA model with log transformation. Note that the quantiles from the simulated sets using our beta *t-GAS* model are very close to the ones estimated from the real data set. These findings suggests that the simulated values produced by the beta *t-GAS* model seem to capture the correct underlying dependency structure presented in the wind CF time series. To reinforce our findings, we also implemented the conditional coverage Value at Risk exceedances test of Cristoffersen [51] to check the coverage level of each predictive density against real data in Table V. Non significant p-values reinforce the idea that the proposed model is not adequate to fit the CF time series. As expected, p-values of Cristoffersen test applied to SARIMA predictive densities are all significant, considering a coverage level of 95%. In contrast, predictive densities produced by beta *t-GAS* model are all non-significant, using the same level of coverage.

TABLE V
P-VALUES CONDITIONAL COVERAGE VALUE AT RISK EXCEEDANCES TEST OF CRISTOFFERSEN CONSIDERING A COVERAGE LEVEL OF 95%.

Coverage level	Rio do Fogo	Icaraizinho	Enacel
t-GAS model	0.306	0.421	0.305
SARIMA	0.034	0.035	0.011

As already argued, such scenarios can benefit several areas. One of them is comercialization. For instance, in order to quantify the investor's risk at the Brazilian Free Trade Environment (FTE), first it is necessary to transform the capacity factors (CF_t) into energy generation in (%) of firm energy certificate (G_t). This is done by using the following formula

$$\tilde{G}_{i,t} = \frac{CF_{i,t}}{100} \cdot \frac{PoW_i}{FEC_i}, \tag{10}$$

where PoW_i and FEC_i are respectively the power and the firm energy certificate of the wind plant $i \in I$. Now, define $S = \{1, \ldots, 2000\}$ are the scenarios generated independently from the Betas densities; $C = 1, \ldots, 2000$ are the scenarios obtained using the dependence structure as given by the Student t copula and $T = \{1, \ldots, 12\}$ is the forecasting horizon for the risk evaluation for the year of 2011. At the most common contracts in the FTE, the so called quantity contracts, agents freely set a price of a bilateral contract where the seller compromises itself in delivering a fixed amount of energy Q, per period, in exchange of a fixed payoff (also per period). Still, in this setting, the contract represents only a financial instrument and does not impose a physical energy delivery obligation by the part of the seller. The differences between the energy due and the amount effectively generated are then settled at the spot market (by the spot price) at each period. In this sense, with $\tilde{G}_{t,i,z}$ and setting $z \in Z$, where $Z = \{S, C\}$ is the index that points out the aforementioned sets, it is possible to generate the distribution for a one year income obtained from a contract negotiated at the FTE. This is done by substituting $\tilde{G}_{t,i,z}$ as given in Equation (10), in the following equation:

$$\tilde{R}_{t,i,z} = PQh_t + (\tilde{G}_{t,i,z} - Q)\tilde{\pi}_{t,z} h_t \text{ for } t \in T, \tag{11}$$

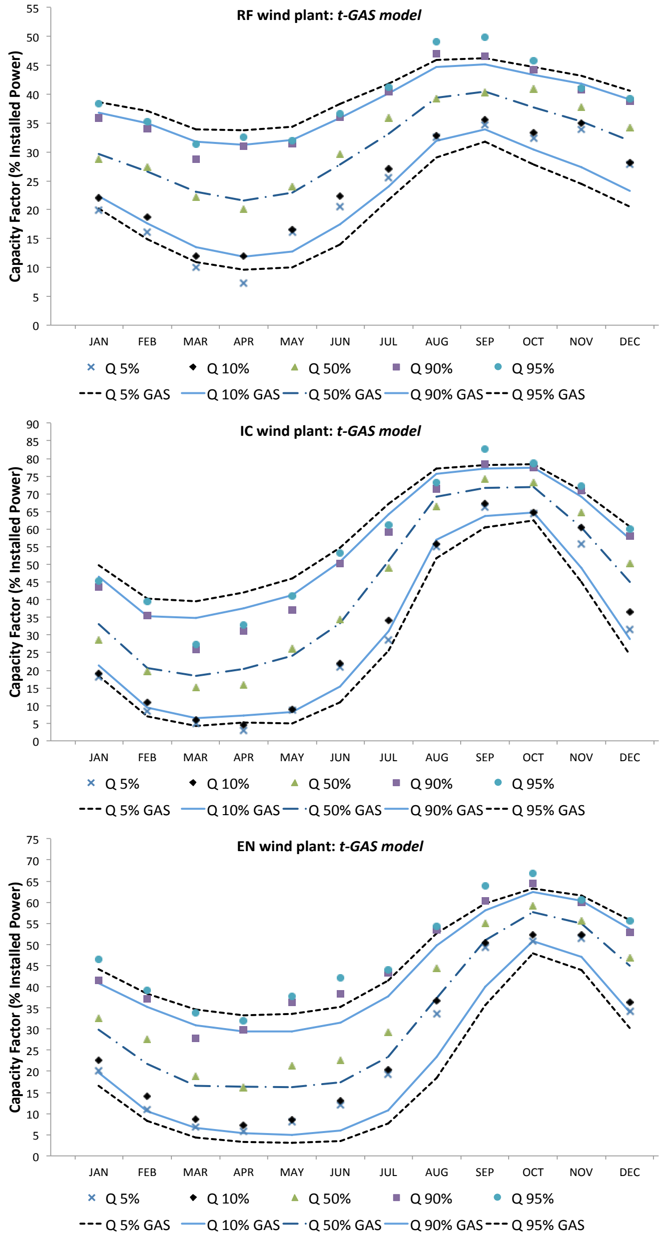

Fig. 4. Scenario evaluation of the year of 2011 through beta *t-GAS* model against the real data set.

where $\tilde{\pi}_{t,z}$ denotes the spot price at configuration z at each period of the contract's horizon, h_t is the total number of hours of period t; Q is the amount of energy sold through the bilateral contract, in avg-MW and, P is the price that defines the contract's fixed payoff parcel at each period. The latter two quantities are fixed at $P = 100$ R\$/MWh and $Q = 1 AvgMW$. The net present value (NPV) of each contract is then given by

$$\tilde{R}_{i,z} = \sum_{t \in T} \frac{\tilde{R}_{t,i,z}}{(1+K)^t} \text{ for } i \in I \tag{12}$$

where K is the interest rate, wich is fixed in zero. Now to generate the distribution of the NPV of the portfolio with contracts attached to those three wind plants, it is only necessary to substitute in Equation (11) the variable $\tilde{G}_{t,i,z}$

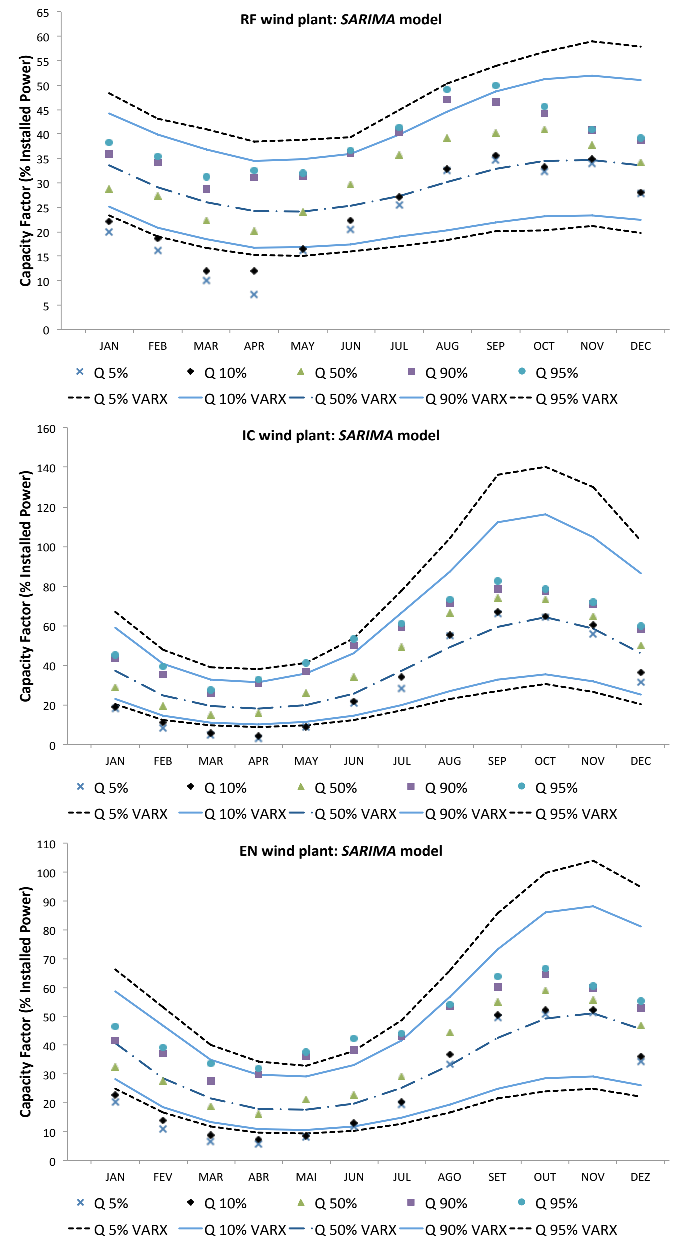

Fig. 5. Scenario evaluation of the year of 2011 through SARIMA model against the real data set.

for $\sum_{t \in T} \sum_{i \in I} \tilde{G}_{t,i,z}$. After that, to obtain the distribution of the cash-flow generated by this portfolio, one just adds their values, as given by Equation (11), resulting in

$$\tilde{R}_z^{Port} = \sum_{t \in T} \sum_{i \in I} \tilde{R}_{t,i,z}. \tag{13}$$

The CVaR estimated using the simulated distribution of the cash flows associated with this portfolio are presented in Table VI, where "dependent" refers to those scenarios generated through the Student t copula, and "independent",refers to simulations obtained directly from each of the beta marginal models.

We focus our attention on the CVaR of $\tilde{R}_z^{Port}$. Such a figure indicates that by considering dependence between the CF's of the different wind plants, one obtains an increase of 12% on

TABLE VI
CVAR FOR THE DISTRIBUTION OF CASH FLOWS FOR DIFFERENT PORTFOLIOS.

$CVaR^{95\%}$	Dependent	Independent
$\tilde{R}_{RF,z}$	-967.06	-1,051.00
$\tilde{R}_{IC,z}$	490.31	316.79
$\tilde{R}_{EN,z}$	622.79	467.27
$\tilde{R}_z^{Port}$	1,224.39	1,093.95

this risk measure compared to the risk evaluated by not taking into account dependence. In the present context this means that an investor would experience a 12% increase on his/her income when investing in such portfolio. Also the CVaR from the wind plant IC, $\tilde{R}_{2,v}$, presents the largest difference when comparing dependent with independent scenarios, 54.7%.

V. CONCLUSION

Our methodology for simulate scenarios from a joint density of wind capacity factor seems fruitful and have several benefits when comparing to the models presented in the literature so far. First, we propose a non Gaussian model to filter wind CF dynamic. By doing such, one does not need to transform variables and work with the log of the observed data point, our framework is taylor made for the range of values that CF's time series can assume. In addition, the simulated values will be always inside physical limits of production, i.e., it will never assume values bigger than the maximum production of wind plant i.

The results presented in this work indicates that our methodology figures as an alternative to simulate joint scenarios of renewable energy sources, where the simulation can be done only by one step.

REFERENCES

[1] International energy agency. [Online]. Available: https://www.iea.org/newsroom/news/2013/october/wind-power-seen-generating-up-to-18-of-global-power-by-2050.html

[2] A. Moreira, D. Pozo, A. Street, and E. Sauma, "Reliable renewable generation and transmission expansion planning: Co-optimizing system's resources for meeting renewable targets," *IEEE Transactions on Power Systems*, vol. PP, no. 99, pp. 1–1, 2016.

[3] R. Jabr, "Robust transmission network expansion planning with uncertain renewable generation and loads," *IEEE Transactions on Power Systems*, vol. 28, no. 4, pp. 4558–4567, 2013.

[4] C. Zhao and Y. Guan, "Data-driven stochastic unit commitment for integrating wind generation," *IEEE Transactions on Power Systems*, vol. 31, no. 4, pp. 2587–2596, July 2016.

[5] A. C. Passos, A. Street, and L. A. Barroso, "A dynamic real option-based investment model for renewable energy portfolios," *IEEE Transactions on Power Systems*, vol. 32, no. 2, pp. 883–895, March 2017.

[6] K. Conradsen, L. Nielsen, and L. Prahm, "Review of weibull statistics for estimation of wind speed distributions," *Journal of Climate and Applied Meteorology*, vol. 23, no. 8, pp. 1173–1183, 1984.

[7] B. G. Brown, R. W. Katz, and A. H. Murphy, "Time series models to simulate and forecast wind speed and wind power," *Journal of climate and applied meteorology*, vol. 23, no. 8, pp. 1184–1195, 1984.

[8] A. J. Conejo, J. Contreras, R. Espínola, and M. A. Plazas, "Forecasting electricity prices for a day-ahead pool-based electric energy market," *International Journal of Forecasting*, vol. 21, no. 3, pp. 435–462, 2005.

[9] H. A. Nielsen, H. Madsen, and T. S. Nielsen, "Using quantile regression to extend an existing wind power forecasting system with probabilistic forecasts," *Wind Energy*, vol. 9, no. 1-2, pp. 95–108, 2006.

[10] P. Pinson, C. Chevallier, and G. N. Kariniotakis, "Trading wind generation from short-term probabilistic forecasts of wind power," *IEEE Transactions on Power Systems*, vol. 22, no. 3, pp. 1148–1156, 2007.

[11] J. W. Taylor, P. E. McSharry, and R. Buizza, "Wind power density forecasting using ensemble predictions and time series models," *Energy Conversion, IEEE Transactions on*, vol. 24, no. 3, pp. 775–782, 2009.

[12] X. Wang, P. Guo, and X. Huang, "A review of wind power forecasting models," *Energy procedia*, vol. 12, pp. 770–778, 2011.

[13] R. J. Bessa, V. Miranda, A. Botterud, J. Wang, and M. Constantinescu, "Time adaptive conditional kernel density estimation for wind power forecasting," *Sustainable Energy, IEEE Transactions on*, vol. 3, no. 4, pp. 660–669, 2012.

[14] G. Sideratos and N. D. Hatziargyriou, "Probabilistic wind power forecasting using radial basis function neural networks," *Power Systems, IEEE Transactions on*, vol. 27, no. 4, pp. 1788–1796, 2012.

[15] C. Wan, Y. Song, Z. Xu, G. Yang, and A. H. Nielsen, "Probabilistic wind power forecasting with hybrid artificial neural networks," *Electric Power Components and Systems*, vol. 44, no. 15, pp. 1656–1668, 2016.

[16] C. Gallego-Castillo, R. Bessa, L. Cavalcante, and O. Lopez-Garcia, "On-line quantile regression in the rkhs (reproducing kernel hilbert space) for operational probabilistic forecasting of wind power," *Energy*, vol. 113, pp. 355–365, 2016.

[17] E. Xydas, M. Qadrdan, C. Marmaras, L. Cipcigan, N. Jenkins, and H. Ameli, "Probabilistic wind power forecasting and its application in the scheduling of gas-fired generators," *Applied Energy*, vol. 192, pp. 382–394, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S030626191631443X

[18] L. Cavalcante, R. J. Bessa, M. Reis, and J. Browell, "Lasso vector autoregression structures for very short-term wind power forecasting," *Wind Energy*, vol. 20, no. 4, pp. 657–675, 2017.

[19] H.-z. Wang, G.-q. Li, G.-b. Wang, J.-c. Peng, H. Jiang, and Y.-t. Liu, "Deep learning based ensemble approach for probabilistic wind power forecasting," *Applied Energy*, vol. 188, pp. 56–70, 2017.

[20] Y. Zhang, J. Wang, and X. Wang, "Review on probabilistic forecasting of wind power generation," *Renewable and Sustainable Energy Reviews*, vol. 32, pp. 255–270, 2014.

[21] A. Street, D. A. Lima, A. Veiga, B. Fanzeres, L. Freire, and B. Amaral, "Fostering Wind Power Penetration into the Brazilian Forward-Contract Market," *IEEE PES General Meeting 2011*, pp. 1–8, Jul. 2011.

[22] M. Souto, A. Moreira, A. Veiga, A. Street, J. D. Garcia, and C. Epprecht, "A high-dimensional varx model to simulate monthly renewable energy supply," in *Power Systems Computation Conference (PSCC), 2014*. IEEE, 2014, pp. 1–7.

[23] A. Lau and P. McSharry, "Approaches for multi-step density forecasts with application to aggregated wind power," *The Annals of Applied Statistics*, pp. 1311–1341, 2010.

[24] J. Jeon and J. W. Taylor, "Using conditional kernel density estimation for wind power density forecasting," *Journal of the American Statistical Association*, vol. 107, no. 497, pp. 66–79, 2012.

[25] J. W. Taylor and J. Jeon, "Forecasting wind power quantiles using conditional kernel estimation," *Renewable Energy*, vol. 80, pp. 370–379, 2015.

[26] C. Wan, J. Lin, J. Wang, Y. Song, and Z. Y. Dong, "Direct quantile regression for nonparametric probabilistic forecasting of wind power generation," *IEEE Transactions on Power Systems*, vol. 32, no. 4, pp. 2767–2778, July 2017.

[27] A. Shapiro, W. Tekaya, J. P. da Costa, and M. P. Soares, "Risk neutral and risk averse stochastic dual dynamic programming method," *European Journal of Operational Research*, vol. 224, no. 2, pp. 375 – 391, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0377221712006455

[28] B. Fanzeres, A. Street, and L. A. Barroso, "Contracting strategies for renewable generators: A hybrid stochastic and robust optimization approach," *IEEE Transactions on Power Systems*, vol. 30, no. 4, pp. 1825–1837, July 2015.

[29] G. Kariniotakis, *Renewable Energy Forecasting: From Models to Applications*, ser. Woodhead Publishing Series in Energy. Elsevier Science, 2017.

[30] D. Creal, S. J. Koopman, and A. Lucas, "Generalized autoregressive score models with applications," *Journal of Applied Econometrics*, vol. 28, no. 5, pp. 777–795, 2013.

[31] R. F. Engle and T. Bollerslev, "Modelling the persistence of conditional variances," *Econometric reviews*, vol. 5, no. 1, pp. 1–50, 1986.

[32] R. F. Engle and J. R. Russell, "Autoregressive conditional duration: a new model for irregularly spaced transaction data," *Econometrica*, pp. 1127–1162, 1998.

[33] R. A. Davis, W. T. Dunsmuir, and S. B. Streett, "Observation-driven models for poisson counts," *Biometrika*, pp. 777–790, 2003.

[34] Gas on-line repository. [Online]. Available: http://www.gasmodel.com/

[35] F.-M. Tseng and G.-H. Tzeng, "A fuzzy seasonal arima model for forecasting," *Fuzzy Sets and Systems*, vol. 126, no. 3, pp. 367–376, 2002.

[36] M. Lei, L. Shiyan, J. Chuanwen, L. Hongling, and Z. Yan, "A review on the forecasting of wind speed and generated power," *Renewable and Sustainable Energy Reviews*, vol. 13, no. 4, pp. 915–920, 2009.

[37] P. Pinson, G. Papaefthymiou, B. Klockl, H. A. Nielsen, and H. Madsen, "From probabilistic forecasts to statistical scenarios of short-term wind power production," *Wind energy*, vol. 12, no. 1, pp. 51–62, 2009.

[38] J. M. Morales, R. Minguez, and A. J. Conejo, "A methodology to generate statistically dependent wind speed scenarios," *Applied Energy*, vol. 87, no. 3, pp. 843–855, 2010.

[39] D. Creal, S. J. Koopman, and A. Lucas, "A dynamic multivariate heavy-tailed model for time-varying volatilities and correlations," *Journal of Business & Economic Statistics*, vol. 29, no. 4, pp. 552–563, 2011.

[40] U. Cherubini, E. Luciano, and W. Vecchiato, *Copula methods in finance*. John Wiley & Sons, 2004.

[41] H. Joe, "Asymptotic efficiency of the two-stage estimation method for copula-based models," *Journal of Multivariate Analysis*, vol. 94, no. 2, pp. 401–419, 2005.

[42] A. J. Patton, "Copula methods for forecasting multivariate time series," *Handbook of economic forecasting*, vol. 2, pp. 899–960, 2012.

[43] ——, "Copula–based models for financial time series," in *Handbook of financial time series*. Springer, 2009, pp. 767–785.

[44] ——, "Modelling asymmetric exchange rate dependence*," *International economic review*, vol. 47, no. 2, pp. 527–556, 2006.

[45] ——, "Applications of copula theory in financial econometrics," Ph.D. dissertation, University of California, San Diego, 2002.

[46] P. K. Dunn and G. K. Smyth, "Randomized quantile residuals," *Journal of Computational and Graphical Statistics*, vol. 5, no. 3, pp. 236–244, 1996.

[47] R. Engle, "Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models," *Journal of Business & Economic Statistics*, vol. 20, no. 3, pp. 339–350, 2002.

[48] J. J. Xu, "Statistical modelling and inference for multivariate and longitudinal discrete response data," Ph.D. dissertation, University of British Columbia, 1996.

[49] H. Joe, *Multivariate models and multivariate dependence concepts*. CRC Press, 1997.

[50] S. A. Murphy and A. W. Van der Vaart, "On profile likelihood," *Journal of the American Statistical Association*, vol. 95, no. 450, pp. 449–465, 2000.

[51] P. F. Christoffersen, "Evaluating interval forecasts," *International economic review*, pp. 841–862, 1998.