

3 Metodologia

3.1. Seleção da série

Para conseguir gerar previsão da variável da velocidade do vento por métodos estatísticos, se faz necessária uma boa base de dados para que seja possível encontrar o modelo de série temporal mais aderente à série de dados e ao mesmo tempo mais capaz, no momento, de efetuar a previsão para períodos futuros. Os dados foram conseguidos pela combinação de dois expedientes. O complexo possuía algumas torres com anemômetros instalados e aquele com maior quantidade e qualidade de dados foi utilizado. Trata-se da torre de nome TA-8503, localizada em posição estratégica. Seus dados apesar de confiáveis eram insuficientes para a criação de um modelo robusto, o que obrigou a que se fizesse uma busca mais ampla por esta informação. O projeto da NASA, de nome MERRA¹, foi escolhido para agregar informação à série anemométrica. Este projeto possui dados de reanálise de locais de todo o mundo e, através de modelos, infere médias de velocidade de vento para cada coordenada do globo. A série do MERRA possui base horária e é calculada a 50 metros de altura. A latitude e longitude escolhidas foram 39.999 e -3.500, que correspondem às coordenadas exatas de onde se encontra o parque. O período inicial de avaliação da série MERRA foi entre 30/04/1979 a 29/04/2014, ajustada através do perfil vertical do vento para 90m de altura.

Chegou-se à série final, uma combinação das duas descritas, ocorrendo então uma concatenação de ambas com alguns ajustes, tendo por resultado final uma única fonte de informação com periodicidade mensal, que se inicia em janeiro de 1990 e terminando em meados de 2014.

¹Modern Era Retrospective analysis for Research and Applications

3.2. Seleção de modelo de previsão

Como afirmado anteriormente, a ideia que norteia a seleção de modelos econométricos é o princípio da parcimônia. Quando se estima um modelo utilizando variáveis defasadas, algumas observações são perdidas. Para conseguir analisar adequadamente modelos alternativos o número de observações 'utilizáveis' deve ser fixado. Caso contrário, haveria comparação de modelos em amostras diferentes. Há vários modelos de seleção por critérios que lidam com o *trade-off* da redução da soma dos resíduos ao quadrado (comparando adicionar mais *lags* com um modelo mais parcimonioso). Um dos mais amplamente utilizados e o mais indicado para séries temporais e largas amostras (BISGAARD; KULAHCI, 2011) é o critério de informação bayesiano (BIC ou *Bayesian Information Criterion*). O BIC pode ser descrito pela seguinte forma:

$$BIC_p = -2 \log(L_p) + [(p + 1) + 1] \log(n) \quad (22)$$

Onde:

L_p é a função de máxima verossimilhança do modelo, p é o número de variáveis explicativas consideradas no modelo e n é o tamanho da amostra.

Ao utilizar o BIC como critério de seleção de modelo, com ou sem variáveis exógenas no modelo e considerando inclusive modelos que incluem sazonalidade, tais como os modelos da família SARMA² o modelo com o menor critério de informação existente foi um ARX(2,0). Neste caso, o componente de média móvel (MA) é zerado e não aparece na equação do modelo. A equação (23) mostra com rigidez matemática como se comporta o modelo.

$$G_t = c + \phi_1 \times (G_{t-1} - d_{t-1}) + \phi_2 \times (G_{t-2} - d_{t-2}) + d_t + \varepsilon_t \quad (23)$$

O modelo ARX consiste em combinar um modelo da família ARMA com variáveis exógenas. No caso específico, *dummies* para capturar a sazonalidade da velocidade do vento inerente aos dados mensais encontrados. Estas estão representadas pelas variáveis d da equação.

² Seasonal Autorregressive Moving Average

Outro ponto importante focado ao se escolher o modelo foram as medidas de erro de previsão. Duas foram utilizadas para se chegar à decisão do melhor modelo a ser escolhido, o MAE³ e o RMSE⁴. O primeiro consiste em calcular o valor absoluto de cada um dos erros do modelo e, em seguida, fazer a média. Quanto menor o número melhor é o modelo. O segundo consiste em elevar ao quadrado cada um dos erros do modelo, tirar a média e, em seguida, calcular a raiz quadrada do resultado. Este último método penaliza o erro grande por calcular o quadrado deste.

3.3. Otimização

O objetivo desejado para este trabalho é a maximização do lucro de uma empresa atuando no ramo de energia eólica no contexto do mercado brasileiro de energia. Portanto, nada mais trivial do que a função objetivo escolhida para o problema seja uma maximização de lucros. O *framework* escolhido para o problema foi o de M-CVaR, ou seja, uma otimização da média dos lucros com restrições de CVaR. No problema (24)) é possível observar a função objetivo e as restrições de CVaR.

$$\text{Max } \frac{1}{s} \sum_{s=1}^s \Pi_s$$

sujeito a:

$$P^c Q^c - P^T Q^T + PLD \times (G_s + Q^T - Q^c) = \Pi_s \quad \forall s$$

$$z - \frac{1}{s} \sum_{s=1}^s \frac{\delta_s}{\alpha} \geq CVaR \quad (24)$$

$$\delta_s \geq z - \Pi_s \quad \forall s$$

$$\delta_s \geq 0 \quad \forall s$$

³ Mean absolut error

⁴ Root mean square error

Onde s é o número de cenários⁵ simulados a serem computados, Π_s é o lucro em cada um dos s cenários. G_s é a potência gerada por uma torre em cada cenário em MW médio, $P^T Q^T$ é o valor da quantidade comprada (ou vendida) de energia no mercado a termo no momento da decisão, $P^c Q^c$ o valor da quantidade comprada (ou vendida) de energia a preço de contrato e $PLD \times (G_s + Q^T - Q^c)$ é o valor de “ajuste” no período futuro calculado a preço de liquidação. A variável z é um escalar que representa o VaR, δ_s é o valor da diferença entre o que é considerado um cenário ‘ruim’ e o resultado obtido no cenário s . E, por fim, a variável α representa a porcentagem de cenários ruins admitidos (significância) e CVaR é o valor numérico escolhido para o CVaR.

A intuição para as restrições é a de que o valor gerado a preço de contrato menos (ou mais) o valor gerado a partir de compras (ou vendas) de energia no mercado a termo e mais (ou menos) os valores do ajuste a preço de liquidação devem ser iguais ao lucro da empresa.

Ademais, o valor de z determina o que é considerado um ‘cenário ruim’, então a diferença da média dos cenários ‘ruins’ para z constitui o valor de CVaR e está descrita na segunda restrição. A terceira restrição é ativada quando a função objetivo se encontra em um cenário denominado ‘ruim’, nestes cenários a quarta restrição não está ativada. Analogamente, nos cenários ‘bons’ a terceira restrição não está ativada e a quarta aparece para cumprir o papel de impedir que os cenários positivos interfiram na média do cálculo do CVaR.

Para uma melhor visualização, segue na **Erro! Fonte de referência não encontrada.** uma visualização gráfica para o que foi descrito. O valor z que divide os cenários bons dos ruins aparece como “-VaR” na figura.

⁵Analogamente ao exemplo dos portfólios na seção ‘problema de otimização’ do referencial teórico, no parque eólico são criados cenários de vento e, por conseguinte, geração de energia. A estes cenários são atribuídos valores a partir de preços, também simulados, para que então o gerador tome a decisão de compra ou venda no mercado a termo de energia. Mais especificamente, foram efetuadas mil simulações de geração de energia e se detinha mil valores de PLD para cada um dos meses abordados. Cada um dos valores de geração simulados foi pareado de maneira aleatória com um dos mil PLDs simulados para aquele mês.

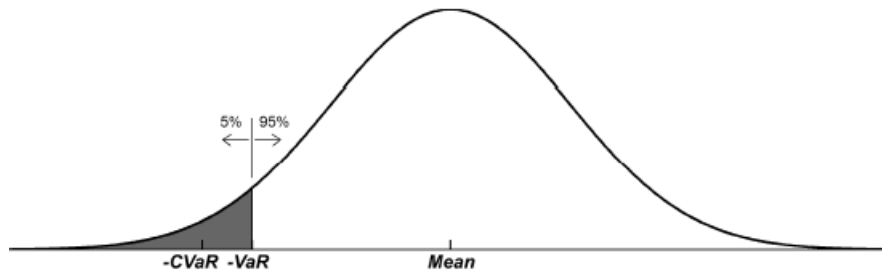


Figura 4 – VaR e CVaR
Fonte: Própria

A implementação do modelo apresentado na equação (24)) pode ser vista na Figura 5 no apêndice. A linguagem escolhida para tal foi o Julia, uma linguagem de alto nível, alta performance, tipagem dinâmica e voltada para programação técnica e científica.

3.4. Passos metodológicos

Os passos seguidos para a realização deste trabalho foram pensados para um ambiente específico e muito peculiar. A ordem em que se empregou o conhecimento foi de primeiro realizar uma modelagem que fosse capaz de capturar a velocidade do vento transformando-a em geração de energia para que, em seguida, fosse possível realizar simulações que servissem para uma otimização financeira.

Após leitura de referencial teórico dedicado ao tema de velocidade do vento, comportamento do mesmo na região onde se encontra o parque eólico, reconhecimento das séries de vento disponíveis e de modelos de previsão de vento, foi decidido que o modelo utilizado para previsão e simulação de vento neste trabalho seguiria o referencial econométrico de modelos ARIMA. Assim sendo, tanto a série de reanálise quanto a medida pelo anemômetro local receberam uma série de adequações relativas às informações disponíveis. Encontrou-se o valor do perfil vertical do vento do local e este foi utilizado para igualar a velocidade de ambas as séries para um mesmo patamar de altura. Isto ocorreu, pois a série de reanálise estava disponível para ventos até 50 metros e as torres do parque se encontram a 90 metros. Em seguida, com a verificação de que as séries eram muito próximas, foi realizada uma concatenação destas. A ideia norteadora desta ação é a de, gradativamente, substituir a série de reanálise pelos dados reais, medidos *in loco*, de forma a não ter mais dados ‘não medidos’ sendo

utilizados. Com a mudança gradativa dos dados se fará necessária nova estimação dos parâmetros do modelo no futuro. Outro ponto importante a ser abordado é que houve uma transformação em médias mensais com os dados utilizados.

O modelo estimado é da família ARIMA, como afirmado anteriormente, e a maneira empregada para a escolha do modelo foi a metodologia de seleção Box-Jenkins. Ao abordar o problema de encontrar um modelo econométrico por esta metodologia tem-se a certeza de que este não será superparametrizado. Ademais, modelos estatísticos têm como ponto positivo a possibilidade de criação de intervalos de confiança e do pleno entendimento do que se passa ‘dentro’ do modelo, não sendo, ‘caixas-pretas’ como alguns modelos de aprendizado por máquina.

A seguir, identificou-se a necessidade de uma transformação das séries de vento em séries de potência. Isto porque a curva de potência das máquinas não é linear, o que significa que médias de velocidade de vento iguais podem gerar potências diferentes ao serem transformadas em eletricidade pelos aerogeradores. Portanto, a série foi calculada levando-se em consideração este fator e o modelo foi parametrizado para poder lidar com estes dados. A partir deste modelo efetuaram-se testes para tentar descobrir se este era adequado no campo estatístico.

Ao observar que o modelo era correto, partiu-se para a parte das simulações. Utilizando um pacote computacional da linguagem R foi possível efetuar simulações de Monte Carlo de forma a guardar cada um dos resultados destas simulações para utilizá-las posteriormente na otimização financeira. A simulação ocorre sorteando-se valores aleatórios para o erro do modelo, seguindo o padrão de comportamento de uma distribuição normal.

Então passou-se para a otimização financeira. A modelagem financeira ocorreu de forma a ter como objetivo a maximização do lucro da empresa controladora do parque eólico para vinte e oito torres. O resultado pode ser facilmente extrapolado para um número maior de torres, se necessário, ou para outro parque inteiramente diferente. Isto seria possível apenas com dados de vento suficientemente longos desta outra localidade. O detentor do parque visa maximizar seu lucro tendo em vista as regras vigentes do mercado de energia brasileiro. Isto quer dizer que sua atuação é sempre visando, no mínimo, o mês

seguinte e que este não pode não cumprir o contrato que detém para atuação no mercado regulado (ACR).

A função objetivo é, portanto, uma maximização do lucro mensal respeitando-se a regra do cumprimento do contrato e com dois arcabouços de restrições, de VaR e de CVaR. O primeiro admite um percentual de perdas estipulado. Por exemplo, em todos os cenários gerados pela simulação e *imputados* na otimização, apenas um percentual α pode vir a gerar cenários negativos. O limite do que é considerado um cenário negativo e os valores de α são previamente determinados pelo gerador de energia e podem ser alterados. O segundo arcabouço, de CVaR, apresenta uma restrição melhor e maior, no sentido de dificultar mais a movimentação por compra ou venda de energia no mercado livre. O CVaR age determinando qual a média desejada das perdas no caso de se estar em um dos α eventos ruins. Assim sendo, o operador fica limitado em possíveis alavancagens que gerariam cenários negativos que fossem de tal magnitude que levassem a empresa à falência.