

4

REPRESENTAÇÃO DE INCERTEZAS

A capacidade de modelagem em programação estocástica está relacionada à capacidade de se modelar adequadamente as incertezas envolvidas nos problemas, que são incorporadas aos modelos através de distribuições discretas e expressas por um número finito de realizações de cenários (ou por uma representação em árvore).

A geração de cenários é uma etapa importante no desenvolvimento de modelos de programação estocástica, e esse processo envolve duas questões principais e antagônicas:

- O número de cenários usados deve ser pequeno o suficiente para que o modelo seja computacionalmente viável;
- O número de cenários deve ser grande o suficiente de tal forma que represente adequadamente as distribuições das variáveis aleatórias utilizadas no modelo.

Diferentes métodos de geração de cenários são aplicados em programação estocástica, e como mencionam Kaut e Wallace em [65], a comparação entre eles deve ser focada em suas performances, e não somente em suas propriedades teóricas. Alguns métodos, por exemplo, aproximam perfeitamente a distribuição quando o número de cenários tende para infinito, mas não necessariamente têm um desempenho bom quando se usam um número finito de realizações. Outros métodos, no entanto, apesar de não garantirem convergência para a distribuição real, têm uma performance boa quando aplicados em problemas reais.

4.1

Tipos de Métodos de Geração de Cenários

Diferentes métodos de geração de cenários encontrados na literatura são discutidos nos trabalhos [57] e [58]. Nas próximas seções são introduzidos os principais deles, terminando com o método aplicado à geração de cenários para o modelo desta tese.

4.1.1

Correspondência de Propriedades Estatísticas (*Property Matching*)

Os métodos de geração de cenários que recebem este nome têm em seu princípio geral a determinação de algumas propriedades estatísticas dos dados (os primeiros momentos, por exemplo), e a partir delas buscar criar cenários que as reproduzam. A geração de cenários é então corresponde a procedimentos heurísticos ou de otimização. A vantagem destes métodos é a não necessidade de se conhecer a função que descreve a distribuição de probabilidade da variável aleatória, mas apenas algumas de suas características estatísticas.

O método mais usado tem o nome de *Moment Matching*, como os apresentados em [41] e [64]. Em [41], por exemplo, os autores propuseram um algoritmo que produz cenários com uma distribuição conjunta e discreta onde quatro primeiros momentos (média, variância e graus de assimetria e curtose) e correlações do conjunto de dados são reproduzidos.

Outros métodos desta classe incluem os baseados em regressão e em análise de componentes principais.

4.1.2

Métodos de Simulação por Amostragem (*Sampling*)

Os modelos de amostragem são utilizados para casos em que se conhece a distribuição de probabilidade da variável aleatória que representa a incerteza do modelo. Como o próprio nome diz, estes métodos consistem em construir cenários através de amostras aleatórias obtidas das distribuições reais, de forma que para cada valor x obtido é associada uma probabilidade $p(x)$. Nesta classe os métodos mais aplicados são *Internal Sampling* [59], Monte Carlo [60] e quasi-Monte Carlo *Sampling* [61]. Na literatura recente, o uso dessas técnicas em problemas de otimização estocástica é conhecido por *Sampling Average Approximation (SAA)* [82][83][84].

4.1.3

Modelos Econométricos e Séries Temporais

Um caso particular de métodos de simulação, bastante empregado na geração de cenários de preços de energia, são os baseados na análise de séries temporais para a estimação modelos estocásticos. Estes métodos visam determinar uma distribuição condicional dos cenários que serão simulados a partir dos diversos métodos anteriormente citados.

A idéia principal reside no uso da informação obtida a partir dos valores passados da própria série para inferir sobre o seu comportamento futuro. Dado um conjunto de dados históricos, estes modelos tentam capturar as características empíricas relevantes da série, descrevê-las, e a partir daí gerar cenários. Alguns modelos vastamente estudados e utilizados para esse fim são os modelos ARMA (e seus derivados), desenvolvidos por Box & Jenkins[66].

O modelo aplicado nesta tese se inclui nesta classe, como será mostrado na seção 4.2.

4.1.4

Outros Métodos

Outros modelos de geração de cenários incluem os baseados em redes neurais e clusterização [62][63]. Alguns métodos também utilizam uma combinação de mais de uma técnica, e por este motivo são chamados de métodos híbridos.

4.2

O Modelo Aplicado nesta Tese

Os preços *spot* horários constituem uma série temporal caracterizada pela presença de sazonalidade e variações estocásticas nos níveis. Muitos autores têm aplicado técnicas de geração de cenários baseadas em análise de séries temporais, mais precisamente em modelos ARMA e seus derivados [39][40][43].

Tradicionalmente assume-se que os preços *spot*, ou suas variações, são regidos por modelos em que os ruídos seguem uma distribuição normal com

variância constante, ignorando o fato de que a volatilidade dos preços também varia no decorrer do tempo. Alguns estudos (por exemplo, [46]) têm sido feitos considerando esta possível dinâmica.

O modelo aplicado nesta tese se baseou, portanto, principalmente no trabalho [46], em que o autor aplicou um modelo GARCH que, como será mostrado, assume que a variância da série de preços *spot* não se mantém constante no decorrer do tempo.

Para a análise das séries de preços *spot* e Elbas usou-se o logaritmo dos valores, que é um artifício bastante utilizado para estabilizar a variância e evitar que sejam gerados cenários com preços negativos. A correlação das séries de preços *spot* e Elbas após aplicar o logaritmo se torna ainda maior, igual a 0,90. A Figura 4.1 abaixo mostra o gráfico do logaritmo das séries históricas de preços *spot* na Suécia (linha preta) e preços do mercado Elbas (linha cinza). A Figura 4.2 mostra a regressão entre estas mesmas séries.

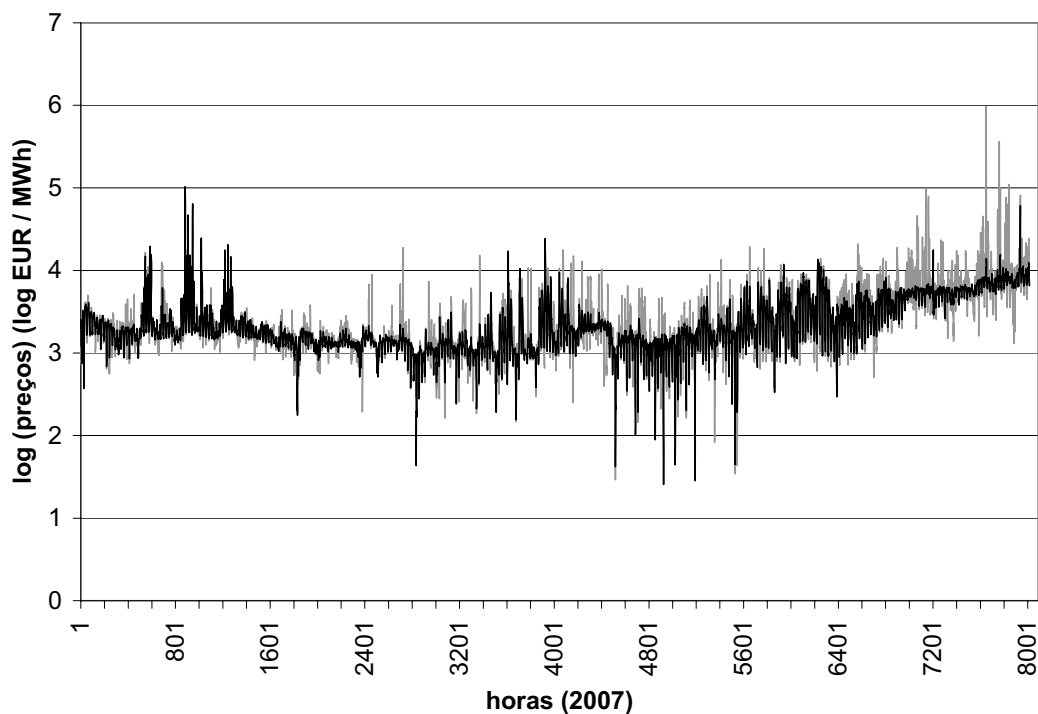


Figura 4.1 – Histórico do logaritmo dos preços *spot* da Suécia (linha preta) e preços Elbas (linha cinza) (1º de janeiro a 30 de novembro de 2007)

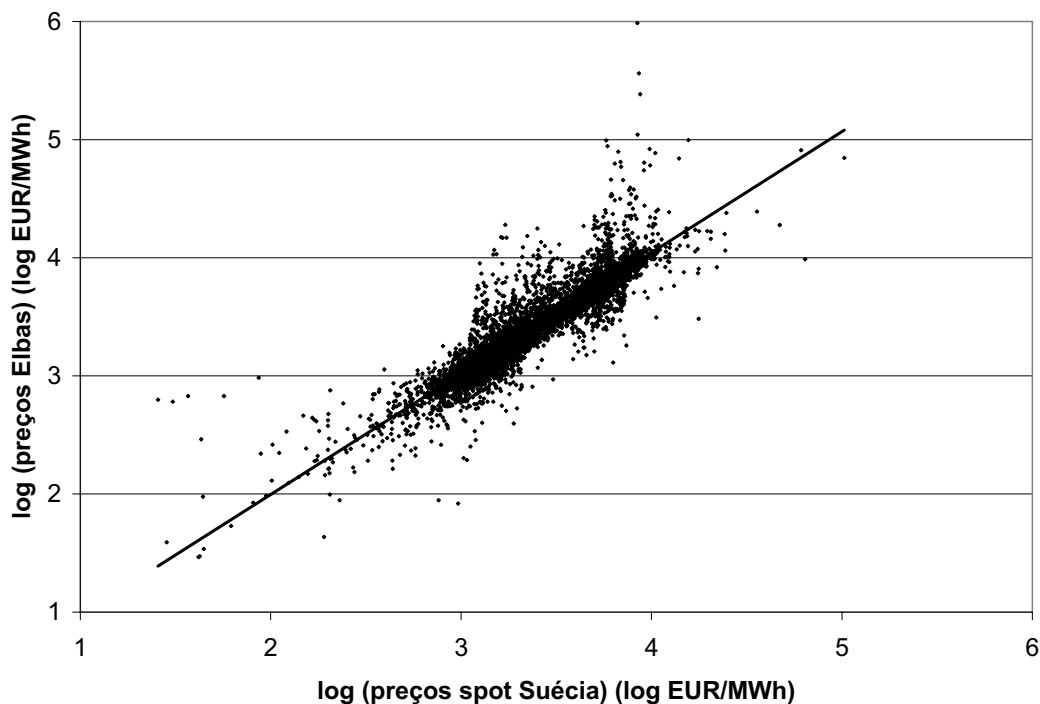


Figura 4.2 – Regressão entre o logaritmo dos preços *spot* e preços Elbas (2007)

A alta correlação entre os preços praticados no mercado Elbas e os preços *spot* é uma característica que deve ser capturada pelo modelo de geração de cenários. Assim sendo, dois modelos foram desenvolvidos nesta tese:

- Preços *Spot*: Modelo ARMA-GARCH, ou seja, um processo ARMA com os ruídos regidos por um processo GARCH;
- Preços Elbas: Modelo ARMAX-GARCH. O modelo ARMAX dos preços Elbas gera cenários condicionados a cada cenário de preço *spot*, utilizando-os como variáveis explicativas.

Ambos os modelos são explicados em detalhes nas próximas seções.

4.2.1

Modelos ARMA

A forma geral de um processo ARMA(p,q) é dado por [94]:

$$y_t = C + \sum_{i=1}^p \phi_i y_{t-i} + \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \varepsilon_t \quad (4.1)$$

ou,

$$\Phi(B)y_t = C + \Theta(B)\varepsilon_t, t \in \mathbb{Z} \quad (4.2)$$

onde $\Phi(B)$ e $\Theta(B)$ são polinômios da forma $\Phi(B) = 1 - \sum_{i=1}^p \phi_i B^i$ e

$\Theta(B) = 1 + \sum_{j=1}^q \theta_j B^j$ e B denota operador de atraso, i.e., $B^i y_t = y_{t-i}$. As

inovações $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ são regidas por um processo ruído branco, i.e., $\varepsilon_t, t \in \mathbb{Z}$ são variáveis aleatórias independentes e normalmente distribuídas com média zero e variância constante, ou, $\mathbb{E}(\varepsilon_t) = 0$ e $\mathbb{E}(\varepsilon_t^2) = \sigma^2$, para todo t , além de não terem correlação temporal, ou seja, $\mathbb{E}(\varepsilon_t \varepsilon_s) = 0$, para todo $s \neq t$.

4.2.2

Modelos ARMAX

A forma geral de um processo ARMAX(p,q) é dada por [94]:

$$y_t = C + \sum_{i=1}^p \phi_i y_{t-i} + \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \xi X_t + \varepsilon_t \quad (4.3)$$

A única diferença deste modelo para o ARMA, descrito anteriormente, é a presença da variável explicativa X_t e seu respectivo coeficiente de regressão ξ ¹⁰.

4.2.3

Modelos GARCH

A classe de modelos auto-regressivos e heterocedásticos de estimação de variância teve origem na observação do fenômeno de *clustering* em séries de retornos de ativos financeiros, em que altos retornos tendem a ser seguidos por outros retornos elevados, fazendo com que a variância da série apresente grande correlação serial. O modelo ARCH (*Autoregressive Conditional Heterocedasticity*) foi introduzido por Engle em 1982 [68], e uma extensão dele, chamada GARCH (*General Autoregressive Conditional Heterocedasticity*), foi proposta por Bollerslev em 1986 [69].

¹⁰ X_t pode representar um vetor de variáveis explicativas, com ξ sendo um vetor com os coeficientes de regressão referentes a cada variável (ou, posição) do vetor X_t .

Os modelos GARCH assumem que a variância condicional das inovações são serialmente correlacionadas, ou seja, não é necessário assumir que a variância $\mathbb{E}(\varepsilon_t^2) = \sigma^2$ se mantém constante no decorrer do tempo. Esta heterocedasticidade muitas vezes melhora significativamente a aderência do modelo à série de dados.

A forma geral do modelo GARCH(p,q) para a variância condicional das inovações, $\sigma_t^2 = \text{Var}(\varepsilon_t | \psi_{t-1})$, onde ψ_{t-1} é o conjunto de todas as informações disponíveis até a etapa t-1, é formulado como:

$$\varepsilon_t = z_t \sigma_t, \text{ onde } z_t \sim iid N(0, 1) \quad (4.4)$$

$$\sigma_t^2 = K + \sum_{i=1}^p \alpha_i \sigma_{t-i}^2 + \sum_{j=1}^q \beta_j \varepsilon_{t-j}^2 \quad (4.5)$$

O modelo GARCH considera haver uma dependência temporal na variância condicional, ou seja, a dinâmica de ε_t^2 pode ser explicada através dos choques de volatilidades passadas ε_{t-i}^2 . Note que $\frac{\varepsilon_t}{\sigma_t}$, que é igual a z_t , é um ruído branco com media 0 e variância igual a 1, e essa hipótese é testada na validação do modelo ajustado.

4.2.4

Geração de cenários de preços Spot: modelo ARMA-GARCH

O primeiro modelo de geração de cenários desenvolvido nesta tese foi para os preços *spot* da Noruega, mais especificamente da área NO1¹¹, onde se localizam as hidrelétricas do caso de estudo.

Os dados históricos utilizados correspondem aos preços *spot* horários de 1º de janeiro a 30 de novembro de 2007, e foram disponibilizados pelo Nord Pool. O modelo proposto para os preços *spot* consiste de um processo ARMA com erros (inovações) regidos por um processo GARCH.

O desenvolvimento e ajuste do modelo proposto é dividido nas seguintes etapas:

1. Especificação do modelo a partir dos dados históricos;
2. Estimação dos parâmetros do modelo;

¹¹ Conforme explicado na seção 0, devido a restrições de transmissão, o Nord Pool pode ser dividido em áreas com diferentes preços *spot*.

3. Validação do modelo;
4. Simulação para a geração de cenários

Cada uma destas etapas é explicada a seguir:

1) Especificação do modelo a partir dos dados históricos

A especificação do modelo a partir dos dados históricos começa com uma cuidadosa análise das principais características da série temporal, onde são observadas a média, variância, além dos comportamentos sazonais dos preços em uma base diária e semanal.

Antes da criação do primeiro modelo a ser testado, a série de preços é transformada em uma série de retornos através da aplicação do logaritmo e da primeira diferença, i.e., a nova série a ser analisada passa a ser $y_t = \log \frac{\rho_{t+1}}{\rho_t} = \log \rho_{t+1} - \log \rho_t$. Este procedimento fez com que a série se tornasse estacionária e é uma técnica bastante utilizada para analisar séries de preços, pois, como mencionado anteriormente, além de normalmente estabilizar sua variância, evita que sejam gerados cenários com preços negativos.

Partindo do modelo geral ARMA-GARCH (4.1)-(4.4)-(4.5), a estrutura dos polinômios que compõe o modelo é determinada através da investigação das autocorrelações e autocorrelações parciais da série.

Com sucessivas tentativas, refinamentos são feitos através da análise das autocorrelações e autocorrelações parciais dos resíduos.

2) Depois de identificado um modelo, a estimação dos parâmetros é feita através da maximização da função de verossimilhança.

3) Nesta etapa é feita a validação do modelo ajustado. A hipótese de que os resíduos padronizados (ruídos ajustados divididos pelos desvios padrões condicionais) são um ruído branco é testada através do estudo de suas autocorrelações e autocorrelações parciais, assim como através da estatística de Ljung-Box (teste de não correlação, ou aleatoriedade) e do teste de hipótese de Engle's (teste de presença de volatilidade condicional) para diferentes intervalos de tempo (*lags*). O AIC/BIC (*Akaike and Bayesian information criteria*) auxilia na comparação entre diferentes modelos, penalizando o excesso de parâmetros.

Quando os parâmetros calculados têm significância estatística e os testes são validados o modelo está pronto para ser usado para simular cenários. Caso contrário, o modelo é refinado voltando-se à etapa 1.

4) A simulação Monte Carlo é usada para a geração de cenários de preços horários, feita a partir de sorteios de uma variável normal independente e identicamente distribuída $z_t, t = 1, \dots, T$.

Os passos de 1 a 4 são todos realizados com o auxílio da toolbox GARCH do Matlab [85].

O modelo final identificado através do procedimento descrito é:

$$\begin{aligned} & (1 - \phi_1 B^1 - \phi_2 B^2 - \phi_4 B^4 - \phi_{23} B^{23} - \phi_{24} B^{24} - \phi_{48} B^{48} - \phi_{72} B^{72} \\ & - \phi_{120} B^{120} - \phi_{144} B^{144} - \phi_{167} B^{167} - \phi_{168} B^{168} - \phi_{336} B^{336} - \phi_{504} B^{504}) \\ & \times (1 - B) \log \rho_t = C + \varepsilon_t \end{aligned} \quad (4.6)$$

onde ε_t é um modelo GARCH(25), tal que (4.4)-(4.5) são:

$$\varepsilon_t = z_t \sigma_t, \text{ onde } z_t \sim iid N(0, 1) \quad (4.7)$$

$$\sigma_t^2 = K + \sum_{i=1}^4 \alpha_i \sigma_{t-i}^2 + \sum_{j=22}^{25} \alpha_j \sigma_{t-j}^2 \quad (4.8)$$

Vale observar que, como o dia tem 24 horas, o modelo ajustado além de capturar a tendência dos preços das horas mais recentes ($i=1, i=2$ e $i=4$), ele também considera a influência do preço referente à mesma hora do dia anterior ($i=24$), à mesma hora de dois dias atrás ($i=48$), à mesma hora de uma semana atrás ($i=168$), etc.

A tabela a seguir mostra os valores dos parâmetros simulados.

Tabela 4.1 – Parâmetros estimados para o modelo de geração de cenários de preços *spot* da área NO1

Parâmetro	C	ϕ_1	ϕ_2	ϕ_4	ϕ_{23}	ϕ_{24}	ϕ_{48}
Estimativa	3.00e-04	0.1821	-0.0608	-0.0462	0.0614	0.2125	0.0284
Parâmetro	ϕ_{72}	ϕ_{120}	ϕ_{144}	ϕ_{167}	ϕ_{168}	ϕ_{336}	ϕ_{504}
Estimativa	0.0492	0.0326	0.0480	0.0484	0.1466	0.0121	0.0354
Parâmetro	K	α_1	α_2	α_3	α_4	α_{22}	α_{23}
Estimativa	4.44e-005	0.3529	0.0737	0.0601	0.0138	0.0488	0.1898
Parâmetro	α_{24}	α_{25}					
Estimativa	0.1954	0.0655					

Para ilustrar a geração de cenários a figura 4.3 abaixo mostra 20 cenários de preços *spot* gerados através da simulação Monte Carlo do modelo descrito acima:

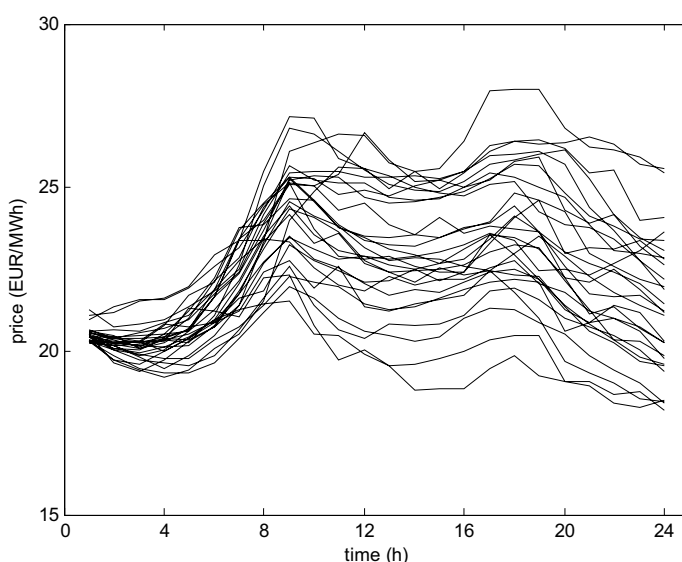


Figura 4.3 – 20 cenários de preços spot da área NO1

4.2.5

Geração de cenários de preços Elbas: modelo ARMAX-GARCH

Como foi mostrado na seção 3.2.1, os preços praticados no mercado Elbas têm uma correlação alta com os preços *spot*, e essa característica deve ser capturada pelos modelos de geração de cenários. Assim sendo, o modelo para os preços Elbas desenvolvido nesta tese corresponde a um ARMAX-GARCH, com

os preços Elbas condicionados aos de preços *spot*. O modelo então serve para gerar cenários condicionados a cada cenário de preço spot, utilizando-os como variáveis explicativas.

Dado que a Noruega ainda não havia entrado no Elbas quando este estudo foi feito, o modelo ARMAX foi ajustado utilizando-se dados históricos de preços Elbas e preços spot da Suécia como variáveis explicativas. Feito isso, para gerar cenários de preços Elbas para agentes na área NO1 utilizamos este mesmo modelo e cenários de preços spot da área NO1 como variáveis explicativas.

Vale ressaltar, novamente, que o sistema computacional onde são realizadas transações do Elbas automaticamente controla possíveis restrições de capacidade das linhas de transmissão. Se duas áreas não possuem capacidade de transmissão entre elas, dois agentes, um em cada uma delas, não poderão transacionar energia. Portanto, os preços do Elbas observados por cada agente são correlacionados com os preços *spot* de sua própria área, e isso foi usado para supor que a Noruega e Suécia possuem a mesma relação entre seus preços *spot* e os preços do Elbas.

Os dados utilizados foram obtidos do Nord Pool, e corresponderam a preços spot da Suécia e preços Elbas¹² de 1º de janeiro a 30 de novembro de 2007, correspondendo a um histórico com 334 valores.

O procedimento para se ajustar o modelo para geração de cenários de preços Elbas segue o mesmos 3 primeiros passos usados da seção anterior. Para estabilizar a variância e evitar cenários com preços negativos, os preços Elbas também foram transformados em uma série de retornos, da mesma maneira feita com os preços *spot*.

A diferença em relação ao procedimento da geração de cenários de preços *spot* está apenas no passo 4. Neste caso, temos que considerar os cenários de preços *spot* como variáveis explicativas para a geração dos cenários de preços Elbas.

O modelo final identificado através do procedimento descrito é:

$$(1 - \phi_1 B^1 - \phi_4 B^4) \times (1 - B) \log \varphi_t = C + \xi_{h(t)} X_t + \varepsilon_t \quad (4.9)$$

¹² Média dos preços das negociações no mercado Elbas durante cada hora.

onde X_t é a variável explicativa, ou seja, a série de retornos do logaritmo dos preços *spot*, e ε_t é um processo GARCH(24,4), tal que (4.4)-(4.5) são:

$$\varepsilon_t = z_t \sigma_t, \text{ onde } z_t \sim iid N(0, 1) \quad (4.10)$$

$$\sigma_t^2 = K + \sum_{i=23}^{24} \alpha_i \sigma_{t-i}^2 + \sum_{j=1}^4 \beta_j \varepsilon_{t-j}^2 \quad (4.11)$$

ξ 's são os coeficientes de regressão indexados por $h(t)$, uma função que retorna uma das 24 horas do dia correspondentes ao período t . Por exemplo, $h(1)=1$, $h(2)=2, \dots, h(24)=24$, $h(25)=1$, $h(26)=2, \dots, h(48)=24$, e assim por diante. portanto, ξ_k , $k = \{1, \dots, 24\}$, são os coeficiente estimados para cada uma das 24 horas do dia¹³. Os valores estimados se encontram na tabela 4.2 a seguir:

Tabela 4.2 – Parâmetros estimados para o modelo de geração de cenários de preços Elbas

Parâmetro	C	ϕ_1	ϕ_4	ξ_1	ξ_2	ξ_3	ξ_4
Estimativa	-0.0814	0.5653	0.0206	0.4221	0.4200	0.4232	0.4274
Parâmetro	ξ_5	ξ_6	ξ_7	ξ_8	ξ_9	ξ_{10}	ξ_{11}
Estimativa	0.4346	0.4516	0.4635	0.4720	0.4644	0.4479	0.4477
Parâmetro	ξ_{12}	ξ_{13}	ξ_{14}	ξ_{15}	ξ_{16}	ξ_{17}	ξ_{18}
Estimativa	0.4529	0.4309	0.4332	0.4358	0.4310	0.4364	0.4451
Parâmetro	ξ_{19}	ξ_{20}	ξ_{21}	ξ_{22}	ξ_{23}	ξ_{24}	K
Estimativa	0.44292	0.4421	0.44249	0.4280	0.4337	0.4196	3.67e-04
Parâmetro	α_{23}	α_{24}	β_1	β_2	β_3	β_4	
Estimativa	0.2676	0.3624	0.2388	0.0941	0.0189	0.0182	

Para ilustrar a geração de cenários a figura 4.4 abaixo mostra 10 cenários de preços Elbas para 3 cenários de preços *spot* diferentes (linhas grossas nos gráficos) através da simulação Monte Carlo do modelo descrito acima:

¹³ Como o modelo geral ARMAX (4.3), poderíamos ter usado apenas um coeficiente ξ para todas as horas, e neste caso um número menor de parâmetros precisaria ser estimado. Esta possibilidade foi analisada e os modelos ajustados com menos parâmetros foram comparados através do AIC/BIC Criteria, mas os resultados foram piores do que o dos modelos mais flexíveis.

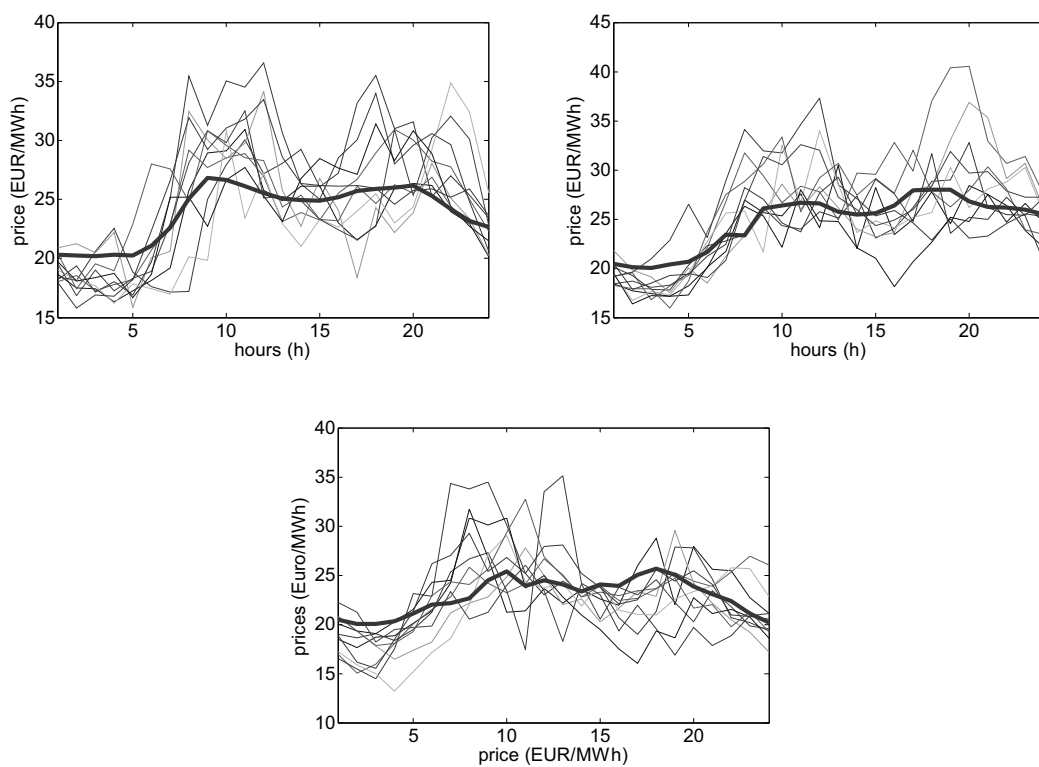


Figura 4.4 – 10 cenários de preços Elbas gerados para 3 cenários de preços *spot* diferentes da área NO1.