



**Universidade do Estado do Rio de Janeiro**

Centro de Ciências Sociais

Faculdade de Ciências Econômicas

Roberto Ultra de Almeida e Silva

**Avaliação Econômica de Projetos Exploratórios na Camada do Pré-Sal:  
Uma Abordagem por Métodos de Aprendizagem Estatística**

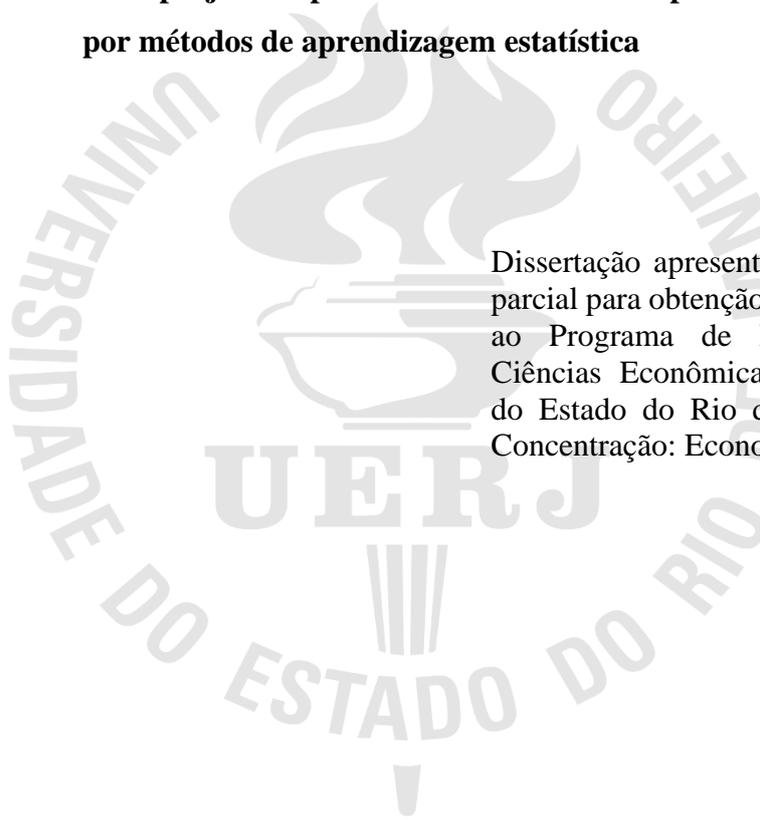
Rio de Janeiro

2020

Roberto Ultra de Almeida e Silva

**Avaliação econômica de projetos exploratórios na camada do pré-sal: uma abordagem  
por métodos de aprendizagem estatística**

Dissertação apresentada, como requisito parcial para obtenção do título de Mestre, ao Programa de Pós-Graduação em Ciências Econômicas, da Universidade do Estado do Rio de Janeiro. Área de Concentração: Economia Aplicada.



Orientador: Prof. Dr. Fernando Antonio Lucena Aiube

Rio de Janeiro

2020

CATALOGAÇÃO NA FONTE  
UERJ/REDE SIRIUS/BIBLIOTECA CCS/B

S586 Silva, Roberto Ultra de Almeida e.  
Avaliação econômica de projetos exploratórios na camada do pré-sal : uma abordagem por métodos de aprendizagem estatística / Roberto Ultra de Almeida e Silva.– 2020.  
108 f.  
  
Orientador: Fernando Antonio Lucena Aiube.  
Dissertação (Mestrado) – Universidade do Estado do Rio de Janeiro, Faculdade de Ciências Econômicas.  
Bibliografia: f.96-101.  
  
1. Petróleo – Aspectos econômicos – Brasil – Teses. 2. Pré-sal – Aspectos econômicos Brasil – Teses. 3. Projetos – Avaliação – Aspectos econômicos – Teses. I. Aiube, Fernando Antônio Lucena. II. Universidade do Estado do Rio de Janeiro. Faculdade de Ciências Econômicas. III. Título.  
  
CDU 338:553.92(81)

Bibliotecária: Regina Souza do Patrocinio CRB7/4954

Autorizo, apenas para fins acadêmicos e científicos, a reprodução total ou parcial desta dissertação, desde que citada a fonte.

\_\_\_\_\_  
Assinatura

\_\_\_\_\_  
Data

Roberto Ultra de Almeida e Silva

**Avaliação econômica de projetos exploratórios na camada do pré-sal: uma abordagem  
por métodos de aprendizagem estatística**

Dissertação apresentada como requisito parcial para obtenção do título de Mestre, ao Programa de Pós-Graduação em Ciências Econômicas, da Universidade do Estado do Rio de Janeiro. Área de Concentração: Economia Aplicada.

Aprovado em 18 de agosto de 2020.

Banca Examinadora:

---

Prof.Dr. Fernando Antonio Lucena Aiube (Orientador)  
Faculdade de Ciências Econômicas - UERJ

---

Prof.Dr. Salvador José Chrispim  
Petróleo Brasileiro S.A.

---

Prof. Dr. Boris Asrilhant  
Universidade Federal do Rio de Janeiro

---

Prof.Dr. Ricardo de Melo e Silva Accioly  
Instituto de Matemática e Estatística - UERJ

---

Prof.Dr. Edison Americo Huarsaya Tito  
Pontifícia Universidade Católica do Rio de Janeiro

---

Prof. Dr. Andrea Ugolini  
Faculdade de Ciências Econômicas - UERJ

Rio de Janeiro

2020

## AGRADECIMENTOS

Agradeço primeiramente a Deus por ter me dado forças e condições para que eu fizesse o mestrado. Agradeço aos meus pais, Marisa e Ricardo, pelo apoio em todos os momentos da minha vida, pela dedicação, ensinamentos e por possibilitarem que eu me aperfeiçoe cada vez mais. À minha futura esposa, Laila, por estar ao meu lado sempre dando força para terminar as minhas iniciativas. Também agradeço à toda a minha família, em especial meu irmão Ricardo e minha afilhada Sophia que estão sempre torcendo por mim.

Também agradeço aos professores da minha graduação na FEN e aos do PPGCE, destaco Carlos Alberto, Rafael Pinho e Fernando Aiube. Os primeiros por terem me incentivado a entrar no mestrado e o terceiro por ter me orientado como nenhuma outra pessoa faria.

Agradeço também aos profissionais que eu tive o prazer de trabalhar. Primeiramente do meu estágio no BNDES, onde eu tive a certeza de que gostaria de seguir esse caminho, e aos profissionais da Petrobras. Destaco meus chefes Cleomar e Suzanna pela compreensão e pelo apoio para terminar o mestrado.

Ainda agradeço aos meus professores e amigos da ECO – Escola de Educação Comunitária, que desde o C.A. me deram a base acadêmica para poder estar concluindo mais essa etapa.

Também agradeço à CAPES, pela bolsa de mestrado recebida enquanto eu não era convocado no concurso, à UERJ por resistir em tempos difíceis como foram os últimos 8 anos nos quais eu estive como discente nessa instituição, à IHS Markit pela liberação dos dados utilizados nessa dissertação e à Petrobras por todo conhecimento que estou adquirindo e pela importância da sua cadeia produtiva para o país. Os dados utilizados nesta dissertação foram obtidos com a empresa IHS Markit, empresa provedora global de informações de diversas indústrias. Tanto a IHS Markit quanto a Petrobras autorizaram o uso das informações da base de dados desta dissertação. Serão apresentadas apenas as principais estatísticas dos dados, sem revelar nomes dos projetos. Além disso, as empresas não tem responsabilidade pelo conteúdo e resultados da dissertação.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

Inclui conteúdo fornecido pela IHS Markit; Copyright © IHS Markit, 2020. Todos os direitos reservados.

## RESUMO

SILVA, R. U. de A. *Avaliação econômica de projetos exploratórios na camada do pré-sal: uma abordagem por métodos de aprendizagem estatística*. 108f. Dissertação (Mestrado em Ciências Econômicas) – Faculdade de Ciências Econômicas, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2020.

Esta dissertação utiliza métodos de aprendizagem estatística, para estimar custos operacionais e de investimento necessários para a avaliação econômica de prospectos exploratórios petrolíferos. Mediante as estimações do OPEX (*operational expenditure*) e do CAPEX (*capital expenditure*) serão preparados modelos de fluxo de caixa baseados nos regimes fiscais de partilha de produção e de concessão. De posse desses resultados, também será feita uma comparação desta avaliação com resultados oriundos de outros trabalhos que também realizaram a avaliação econômica de projetos de produção de petróleo.

Palavras-chave: Avaliação econômica. Incerteza exploratória. Aprendizagem estatística.

## ABSTRACT

SILVA, R. U. de A. *Economic evaluation of exploratory projects in the pre-salt layer: an approach using statistical learning methods*. 108f. Dissertação (Mestrado em Ciências Econômicas) – Faculdade de Ciências Econômicas, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2020.

This dissertation uses statistical learning methods to estimate operating and capital expenses used in the economic evaluation of oil exploration prospects. Based on the estimates of OPEX (operating expenses) and CAPEX (capital expenses), cash flow models will be prepared based on the production sharing contract and concession contract. With these results, a comparison of this evaluation will also be made with results from other studies that also carried out the economic evaluation of oil production projects.

Keywords: Economic evaluation. Exploratory uncertainty. Statistical learning.

## LISTA DE FIGURAS

Figura 1 - Exemplos de curva de produção .....	15
Figura 2 - A curva superior representa o MSE de teste e a inferior o MSE de treinamento ....	28
Figura 3 - Exemplo de comportamento dos coeficientes em função de $\lambda$ no ridge regression.....	30
0	
Figura 4 - Previsões do preço do petróleo Brent. ....	3939
Figura 5 - MSE em função de $\lambda$ no ridge regression da equação 40 .....	52
Figura 6 - Coeficientes em função de $\lambda$ no lasso da equação 43. ....	54
Figura 7 - Efeitos parciais das variáveis explicativas no CAPEX usando smoothing splines. ....	600
Figura 8 - Efeitos parciais do $V_{total}$ e da $LDA$ no CAPEX no melhor ajuste utilizando <i>smoothing splines</i> .....	611
Figura 9 - Efeitos parciais das variáveis explicativas no OPEX usando smoothing splines. .	622
Figura 10 - Efeitos parciais do $V_{total}$ e da <i>area</i> no OPEX no melhor ajuste utilizando <i>smoothing splines</i> . ....	633
Figura 11 - Variação do MSE em função do parâmetro de suavização ( $\alpha$ ) da regressão local ao tentar explicar o CAPEX em função do $Vóleo$ . ....	655
Figura 12 - Regressão local do CAPEX em função do $Vóleo$ utilizando $\alpha = 1,2$ . ....	655
Figura 13 - Efeitos parciais da regressão local do CAPEX em função do $Vóleo$ e da $LDA$ utilizando $\alpha = 1$ . ....	666
Figura 14 - Regressão local do OPEX em função da <i>area</i> utilizando $\alpha = 1$ .....	667
Figura 15 - Quatro projeções exemplificativas para o preço do petróleo.....	700
Figura 16 - Localização do prospecto Nordeste de Tupi.....	79
Figura 17 – Histograma do volume recuperável esperado. ....	80
Figura 18 - Possibilidades de curva de produção considerando a incerteza volumétrica.....	81

Figura 19 - Histograma do CAPEX esperado .....	822
Figura 20 - Histograma do OPEX esperado.....	833
Figura 21 - Histograma do preço do petróleo no ano do pico da produção .....	844
Figura 22 - Preço médio do petróleo por ano.....	855
Figura 23 - Incerteza no fluxo de caixa no regime de partilha de produção .....	866
Figura 24 - Incerteza no fluxo de caixa no regime de concessão.....	866
Figura 25 - Histograma do VPL esperado no regime de partilha de produção .....	87
Figura 26 - Histograma do VPL esperado no regime de concessão.....	88
Figura 27 - Histograma da participação governamental esperada no regime de partilha.....	900
Figura 28 - Histograma da participação governamental esperada no regime de concessão...	911
Figura 29 - Modelo de aquisição sísmica terrestre .....	1033
Figura 30 - Representação de uma aquisição sísmica marítima. (1) Air gun. (2) camada de rocha. (3) Receptores (hidrofones) (4) navio de aquisição sísmica.....	1044
Figura 31 - Trapa estrutural na esquerda e estratigráfica na direita .....	1055
Figura 32 - Exemplo esquemático do volume de uma acumulação .....	1066
Figura 33 - Exemplo de distribuição de probabilidade acumulada do volume recuperável de um prospecto exploratório .....	10808

## LISTA DE TABELAS

Tabela 1 - Estatística descritiva dos dados. ....	422
Tabela 2 - Correlação de Pearson entre o CAPEX e as variáveis independentes. ....	433
Tabela 3 - Correlação de Pearson entre o OPEX e as variáveis independentes. ....	444
Tabela 4 - Correlação de Pearson entre as variáveis independentes. ....	455
Tabela 5 - VIF para cada variável independente em cada regressão. ....	455
Tabela 6 - Resumo dos modelos lineares. ....	57
Tabela 7 - Resumo dos modelos flexíveis. ....	68
Tabela 8 - Cálculo de participação especial. ....	74
Tabela 9 - Distribuição dos investimentos em CAPEX. ....	766
Tabela 10 - Cálculo do fluxo de caixa anual para o regime de concessão ....	766
Tabela 11 - Cálculo do fluxo de caixa anual no regime de partilha de produção ....	77
Tabela 12 - Estatística descritiva do Brent no ano do pico da produção.....	844
Tabela 13 - Estatística descritiva do VPL no regime de partilha de produção.....	88
Tabela 14 - Estatística descritiva do VPL no regime de concessão ....	89
Tabela 15 - Comparação de valores variando a metodologia para previsão do CAPEX e do OPEX nos dois regimes fiscais avaliados em MM US\$. ....	922

## LISTA DE ABREVIATURAS E SIGLAS

ANP	Agência Nacional do Petróleo, Gás Natural e Biocombustíveis
ARCH	<i>Autoregressive conditional heteroscedasticity</i>
CAPEX	Custos de investimento
CSLL	Contribuição Social Sobre o Lucro Líquido
FPSO	<i>Floating, Production, Storage and Offloading</i>
FR	Fator de Recuperação
GAM	Modelos Aditivos Generalizados
GBM	Movimento Browniano Geométrico
IGP-DI	Índice Geral de Preços – Disponibilidade Interna
IHS	<i>Inverted Hockey Stick</i>
IR	Imposto de Renda
LAIR	Lucro Antes do Imposto de Renda
LDA	Lâmina d'água
LOOCV	<i>Leave-one-out cross-validation</i>
MQO	Mínimos Quadrados Ordinários
MSE	Erro Quadrático Médio
NP	Produção total do poço
OPEX	Custos operacionais
P&D	Pesquisa e Desenvolvimento
PECE	<i>Price-effect-cost-escalation</i>
PMBOK	<i>Project Management Body of Knowledge</i>
SGS	Simulação Gaussiana Sequencial
SPEE	Sociedade de Engenheiros de Avaliação de Petróleo
TIR	Taxa Interna de Retorno
TMA	Taxa Mínima de Atratividade
UEP	Unidade Estacionária de Produção
VIF	Fator de Inflação de Variância
VOIP	Volume de Óleo <i>In Place</i>
VPL	Valor Presente Líquido
WTI	<i>West Texas Intermediate</i>

## SUMÁRIO

INTRODUÇÃO .....	11
<b>1. REVISÃO DA LITERATURA .....</b>	<b>14</b>
1.1 Avaliação econômica de prospectos exploratórios .....	14
1.2 Métodos estatísticos e modelo de preços .....	24
<b>2. METODOLOGIA E DADOS .....</b>	<b>40</b>
2.1 Metodologia .....	40
2.2 Dados .....	411
<b>3. MODELOS .....</b>	<b>46</b>
<b>3.1. Modelos lineares .....</b>	<b>46</b>
3.1.1 <u>Stepwise</u> .....	50
3.1.2 <u>Ridge regression</u> .....	511
3.1.3 <u>Lasso</u> .....	54
<b>3.2 Modelos flexíveis .....</b>	<b>57</b>
3.2.1 <u>Smoothing splines</u> .....	58
3.2.2 <u>Regressão local</u> .....	64
<b>3.3 Simulação preço do petróleo.....</b>	<b>68</b>
<b>3.4 Simulação curva de produção .....</b>	<b>71</b>
<b>3.5 Premissas gerais .....</b>	<b>722</b>
<b>4. RESULTADOS .....</b>	<b>78</b>
4.1 Aplicação no projeto base .....	78
4.2 Indicadores econômicos e comparação entre regimes fiscais .....	85
4.3 Comparação com a previsão dos custos utilizando regressão linear .....	91
CONCLUSÃO .....	94
REFERÊNCIAS .....	96
APÊNDICE – Avaliação de prospectos exploratórios .....	102

## INTRODUÇÃO

### Contexto e relevância da pesquisa

Desde meados do século XX, uma das principais atividades econômicas no mundo é a exploração de petróleo e gás natural. Após o nascimento dessa indústria, seus níveis de relevância, competitividade e lucratividade foram crescentes. Isso é comprovado pela lista das maiores empresas de capital aberto no mundo, divulgada pela Forbes (2019), na qual a indústria de petróleo se colocou na terceira posição em quantidade de empresas, perdendo apenas para o setor financeiro e o setor de construção. O Brasil também se destaca nesta indústria, com a Petrobras se posicionando constantemente nas primeiras posições na lista de maiores empresas de capital aberto do país.

Neste sentido, é primordial para que a indústria continue pulsante, que as avaliações econômicas dos projetos sejam feitas da forma mais verossímil e representativa possível, levando em consideração todas as variáveis desta atividade. Com um bom planejamento e esforço, é possível reduzir bastante os imprevistos durante a vida de um projeto de desenvolvimento da produção de um campo já descoberto. As ferramentas mais recentes, auxiliadas pelo *Project Management Body of Knowledge* (PMBOK), contribuem para uma gestão de projetos cada vez mais eficiente e estruturada.

Porém, a exploração de petróleo possui uma especificidade relacionada ao descobrimento dos recursos. Nem sempre se encontra petróleo ao perfurar uma locação. Uma evidência disso é que apenas cerca de 30% dos poços perfurados localizam hidrocarbonetos em volumes viáveis para aproveitamento, conforme Kunjan (2016). Além disso, mesmo que haja uma descoberta, há uma incerteza relacionada a quantidade de petróleo no reservatório. Isso só fica claro após a perfuração de uma quantidade suficiente de poços de delimitação.

As avaliações de oportunidades exploratórias (possíveis acumulações de hidrocarboneto) envolvem principalmente duas fontes de incerteza: incerteza da existência ou não de petróleo e a incerteza com relação ao volume recuperável da jazida. Estas incertezas se aliam a outras, como preços futuros do petróleo e produtividade da jazida.

Dessa forma, uma avaliação econômica pontual e determinística, considerando valores médios destas variáveis traz uma resposta que não representa satisfatoriamente os riscos da

atividade. Ela negligencia alguns cenários e acaba levando a uma tomada de decisão que foi suportada por uma avaliação que não abordou todos os aspectos daquele projeto.

Este trabalho irá estimar modelos que consideram a incerteza destes parâmetros na avaliação de economicidade de projetos exploratórios do pré-sal. Serão comparadas metodologias da área de aprendizagem estatística com diferentes níveis de flexibilidade, para a estimação dos custos de investimentos (CAPEX) e operacionais (OPEX), considerando fenômenos como o *trade-off* entre viés e variância e o *overfitting*.

Espera-se chegar a modelos, flexíveis ou não, que possuam uma forma funcional que melhor represente os dados de CAPEX e OPEX. Mais ainda, que possibilitem realizar a avaliação econômica considerando as incertezas nos volumes esperados. Ou seja, cujo resultado traduza as incertezas exploratórias em incertezas relacionadas aos indicadores de economicidade para suporte ao processo decisório.

### **Objetivos**

Este trabalho se propõe a atingir três objetivos:

1. Estimar modelos que expliquem a variação dos parâmetros utilizados na avaliação econômica de projetos exploratórios, utilizando metodologias com diferentes graus de flexibilidade, da disciplina de aprendizagem estatística (*statistical learning*);
2. Gerar uma distribuição de valor presente líquido (VPL) a partir das variáveis obtidas, que considere as incertezas das oportunidades exploratórias, através de simulação usando os modelos estimados no item anterior;
3. Comparar com trabalhos de outros autores os resultados obtidos sob dois regimes fiscais: concessão e partilha de produção.

A metodologia proposta nesta dissertação possui limitações, principalmente associadas à estimação dos parâmetros para utilização na avaliação econômica e ao seu poder de explicação das variáveis dependentes.

Além disso, uma avaliação convencional de um projeto de exploração de petróleo, realizada por um especialista, demanda um longo trabalho. De maneira geral, envolve estudos de reservatórios, estimativas de custos das diversas instalações (sub-superfície, submarinas, plataformas de produção), escoamento da produção, dentre outros. O modelo econométrico aqui proposto certamente não poderá captar todas as disciplinas e o conhecimento de um avaliador, nem dará uma resposta tão precisa quanto a avaliação dele. Porém, é preciso frisar que este não

é o objetivo do trabalho, e sim estimar valores para as variáveis com base em informações disponíveis para dar uma resposta mais rápida das possibilidades da oportunidade exploratória.

### **Estrutura do trabalho**

Além desta Introdução, na qual estão sendo abordados o contexto e os objetivos da dissertação, este trabalho contém 5 capítulos.

No Capítulo 1 é feita uma revisão da literatura, na qual são apresentados os principais trabalhos que discutem temas relacionados com a dissertação e que possam contribuir para o entendimento da mesma. Também são revistas as metodologias de estimação que serão utilizadas no trabalho.

O Capítulo 2 dedicar-se-á à análise dos dados, explorando suas relações e propriedades.

O Capítulo 3 apresenta os modelos utilizados para a obtenção das variáveis necessárias para o cálculo do VPL e os resultados da estimação dos modelos. Além disso, também será detalhada a forma do cálculo do VPL e as premissas utilizadas para construção do fluxo de caixa.

O Capítulo 4 será dedicado ao cálculo do VPL e à análise dos resultados encontrados comparando-os com um caso real.

No quinto e último Capítulo são apresentadas as conclusões do trabalho e as considerações finais.

## 1. REVISÃO DA LITERATURA

Esta revisão da literatura será dividida em duas partes. A primeira tratará das especificidades da etapa de avaliação econômica de oportunidades exploratórias, feita da forma tradicional e determinística, incluindo as características dos regimes de partilha de produção e de concessão e suas respectivas tributações. E a segunda parte será dedicada à revisão dos métodos estatísticos que serão utilizados neste estudo e à exemplificação de aplicações no contexto deste trabalho.

Além disso, para fins de contextualização, o apêndice A se dedica à explanação da metodologia de avaliação de oportunidades exploratórias da indústria de óleo e gás e porque a incerteza destaca-se como um fato tão relevante nesta etapa da cadeia petrolífera.

### 1.1 Avaliação econômica de prospectos exploratórios

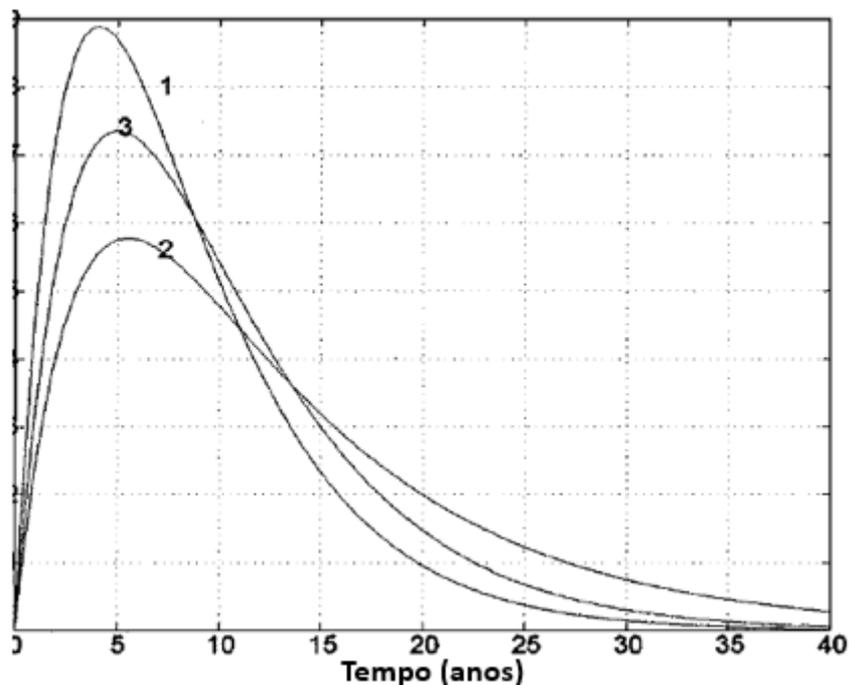
Encontrar e caracterizar uma oportunidade exploratória não é uma tarefa trivial. Da mesma forma, as etapas para realizar uma avaliação econômica completa também possuem uma alta complexidade.

Em um primeiro momento, é preciso estimar o volume e a curva de produção daquela acumulação que foi identificada. Ou seja, dada a média do volume recuperável obtida a partir da distribuição de probabilidade, como este volume será recuperado ao longo da vida desta jazida.

Em geral, esta curva de produção segue o formato da Figura 1, com um rápido crescimento, também chamado de *ramp-up*, até atingir um pico de produção logo no começo de sua vida, seguido por um declínio exponencial que permanece até que os custos operacionais superem a receita do projeto.

A Figura 1 ilustra alguns diferentes formatos possíveis de curvas de produção para a recuperação de um mesmo volume de óleo. A tendência é que a curva siga estes modelos, variando a rapidez com que se chega ao pico e a duração do projeto.

Figura 1 – Exemplos de curva de produção



Fonte: Feygin e Ryzhik, 2001.

Li e Horne (2005) sugerem 4 modelos para realizar a estimação dessa curva: hiperbólico, exponencial, harmônico e Li-Horne, este último modelado em Li e Horne (2003). As informações obtidas com as técnicas acima foram comparadas com dados de reservatórios de diferentes permeabilidades.

Segundo este artigo, o método exponencial tende a subestimar as taxas de produção e o harmônico tende a superestimar estes valores. O modelo hiperbólico é um valor médio entre os dois, conforme será visto abaixo. E a metodologia Li-Horne obteve o melhor desempenho no artigo mencionado, porém ele necessita de mais dados de entrada, como permeabilidade relativa e pressão capilar do reservatório.

No modelo hiperbólico, a vazão de produção de cada ano depende da vazão inicial, do tempo decorrido desde o início da produção e de duas constantes:

$$q(t) = \frac{q_i}{(1 + b * D_I * t)^{1/b}} \quad (1)$$

onde,  $q_i$  é a vazão inicial,  $t$  é o tempo decorrido desde o início da produção, e  $b$ , e  $D_I$  são constantes.

O modelo exponencial é um caso especial do que foi tratado anteriormente, no qual a constante  $b$  é igual a zero:

$$q(t) = \frac{q_i}{e^{D_I * t}} \quad (2)$$

E o harmônico é o caso em que a constante  $b$  é igual a 1:

$$q(t) = \frac{q_i}{(1 + D_I * t)} \quad (3)$$

O modelo Li-Horne parte da seguinte equação:

$$q(t) = \alpha_0 \frac{1}{R(t)} - b_0 \quad (4)$$

onde  $\alpha_0$  é uma função obtida a partir da área, espessura, saturação de água e da pressão capilar, e  $b_0$  é função da diferença de densidade entre a fase água e fase óleo e da permeabilidade relativa do reservatório.

Conforme a fórmula de Li-Horne tenta traduzir, este formato característico das curvas de produção resulta da diferença das condições de pressão, saturação de água, dentre outros fatores, ao longo da vida do projeto e de acordo com a exploração dos fluidos.

Segundo a sociedade de engenheiros de avaliação de petróleo (SPEE, 2002) que divulga boas práticas de avaliação de projetos de exploração e produção de petróleo, existem duas técnicas indicadas para estimar a curva de produção, o método exponencial e o hiperbólico. Logo, um desses dois métodos será escolhido para ser utilizado nesta dissertação. No tópico 3.4 serão detalhadas a metodologia e as premissas.

Em conjunto com a modelagem da curva de produção, é preciso definir o tipo e dimensionar a capacidade da unidade estacionária de produção (UEP). São definidos também a quantidade de poços produtores e injetores e seus custos de perfuração e completção, quantidade de sondas para perfuração, a infraestrutura submarina e a infraestrutura de interligação e de escoamento. Estes são os grandes investimentos que compõem o CAPEX (capital expenditure) para o desenvolvimento de um campo.

A capacidade da UEP deve ser suficiente para suportar o pico da produção diária presente na curva de produção. A expectativa de produção deve ser condizente com a

quantidade de poços produtores e injetores. Além disso, a quantidade de poços injetores está relacionada às condições do reservatório. Por exemplo, em geral reservatórios com baixa pressão devem ter uma quantidade maior de poços injetores, para estimular o fluxo de hidrocarboneto em direção ao poço produtor. Reservatórios com alta concentração de CO<sub>2</sub> também utilizam uma quantidade maior de injetores. Além disso, a decisão de aproveitar economicamente o gás produzido também afeta a decisão de escoar esse gás ou reinjetá-lo.

Os custos de perfuração e de completação dependem principalmente da profundidade em que está o objetivo a ser alcançado e da litologia, o que está relacionado também com o tempo para perfuração do poço e consequente taxa de aluguel da sonda. A quantidade de sondas de perfuração será função da distribuição dos poços que irão drenar o campo. Os barcos de lançamento de linhas para conexão dos poços à UEP são responsáveis por relevantes custos de investimento. As definições dos tempos de perfuração, completação e interligação permitirão definir mais apropriadamente o cronograma de entrada de poços e consequentemente a curva de produção.

Sobre a infraestrutura de escoamento, é preciso avaliar a proximidade com algum gasoduto já existente. Caso não exista, é preciso estimar o custo de construção. Também deve-se estimar o custo com navios aliviadores, que escoam a produção até a costa, quando do uso de um FPSO (*Floating, Production, Storage and Offloading*).

Suslick et al (2003), abordaram o desenvolvimento de reservatórios em águas ultraprofundas com presença de óleo pesado ao se aplicar a possibilidade de melhorias tecnológicas para recuperação destes recursos. Eles expressaram o CAPEX em função do volume de óleo, assim os valores ficaram entre 3,50 US\$/bbl e 7,90 US\$/bbl. Este artigo possui a semelhança com a presente dissertação por tratar de campos localizados em águas ultraprofundas no Brasil. No próximo tópico deste capítulo serão apresentadas outras estimativas para o CAPEX de um projeto de produção.

Também é preciso estimar o OPEX, que são os custos de manter a operação em andamento. Em geral, os maiores custos são representados pelos custos de operação das UEPs e dos poços e custos de *workover*, que são intervenções com foco no reparo de danos e manutenção (Soares, 2017).

Os custos operacionais podem depender de fatores diferentes dos utilizados para o CAPEX, como volume produzido, tempo de produção e custo de pessoal. No trabalho de Lucchesi (2019) este custo ficou entre 8 US\$/bbl e 12 US\$/bbl. Já Suslick et al (2003) fez uma

estimativa entre 2,00 US\$/bbl e 3,78 US\$/bbl. Essa diferença pode estar relacionada às diferentes amostras e realidades de produção, além da diferença temporal entre os trabalhos.

É importante destacar que os dois custos sofrem efeito de mercado e estão sujeitos à inflação, portanto pode haver uma diferença nessa estimativa dependendo da época em que foram calculados. Este é um dos fatores que explica a diferença entre os valores encontrados por diferentes artigos.

Segundo Surovtsev e Sungurov (2017), há uma convenção na engenharia de custos de que estimar o OPEX anual como 4% do CAPEX total não é uma aproximação ruim, porém não foi encontrada base disso na literatura. Para testar essa hipótese, os autores dividiram o OPEX total pelo CAPEX total da amostra disponível e obtiveram uma distribuição log-normal com média 3,4%, P90 de 1% e P10 de 6,6%. Assim, pelos dados disponíveis, esta pareceu uma boa aproximação para o OPEX.

Verre, Guibileo e Cadegiani (2009) citam que esta aproximação é feita frequentemente, sem mencionar o percentual utilizado. Porém, afirmam que esta forma de estimação não possui acurácia e não é relacionada com as especificidades do projeto. Assim, eles propõem uma estimativa com base na metodologia de custeio ABC, chamada também de custeio baseado em atividades. Esta forma de cálculo sugere que o avaliador faça a estimativa para cada tipo de atividade, separadas em 3 tipos no artigo: custos de operação, gastos com serviços e gastos com administração. Para colocar uma incerteza no valor total do OPEX, eles sugerem que sejam colocados valores mínimos e máximos em cada um dos componentes do OPEX para posteriormente, por meio de uma distribuição de Monte Carlo obter a distribuição de probabilidades deste item.

Além das questões comentadas nos parágrafos anteriores, é preciso alinhar de forma adequada as premissas econômicas e de produção antes de realizar as avaliações, como data de início da produção, taxa de desconto e distribuição dos investimentos em infraestrutura no tempo.

Diversos trabalhos abordam a avaliação econômica de projetos da indústria de petróleo. Gustavson (2000) afirma que o fluxo de caixa descontado depende de quatro parâmetros principais: volume produzido, preços do óleo, custos operacionais e taxa de desconto. Todas essas variáveis são incertas no momento de perfurar uma oportunidade exploratória. Neste trabalho serão consideradas todas estas incertezas, com exceção da taxa mínima de atratividade, que será fixa.

Lucchesi (2019) fez uma análise de sensibilidade buscando identificar como quatro fatores afetavam a rentabilidade de um projeto. Assim, ao testar variações nos preços, CAPEX, OPEX e mudanças de até 20% no volume recuperável, chegou-se à conclusão de que, pelo menos para o regime de partilha brasileiro, os preços são os que mais impactam a lucratividade, seguidos do volume e depois do CAPEX.

Como já citado na primeira parte desta revisão da literatura, o volume segue uma distribuição de probabilidade. Neste trabalho não haverá necessidade de estimar essa distribuição. Será considerado que todo o trabalho de incerteza nos parâmetros volumétricos, discutidos no apêndice A, foi feito anteriormente por um profissional de geologia capacitado para tal e resultou em uma distribuição cujos percentis serão utilizados para reconstruir a curva de distribuição acumulada e realizar as iterações em cima de seus valores, conforme será feito no Capítulo 4.

Conforme Harzer et al (2014), há várias formas da empresa determinar a taxa mínima de atratividade (TMA) que será utilizada para calcular o VPL dos seus projetos. Segundo os autores, a metodologia clássica indica a utilização de um *spread* em relação à taxa livre de risco. E ele deve ser proporcional ao risco deste projeto. A TMA utilizada nesta dissertação será definida no Capítulo 3.

Em relação aos preços do petróleo e gás, observa-se uma grande oscilação de acordo com a oferta e a demanda, fatores geopolíticos, demográficos e de governo (Moncrieff, 2008). Em geral, são utilizados dois preços de referência internacional, o petróleo *West Texas Intermediate* (WTI) e o Brent. O primeiro é negociado na bolsa de Nova Iorque e tem como referência o petróleo extraído do Golfo do México. Já o segundo é negociado na bolsa de Londres e tem como referência o produto do Mar do Norte. No Brasil, geralmente o preço utilizado como referência é o petróleo tipo Brent. A previsão do preço neste trabalho está associado à dinâmica de modelos baseados em processos estocásticos que serão apresentados na segunda parte desta revisão da literatura.

Outra variável importante para o cálculo do valor dos projetos é a participação governamental. Ou seja, o valor pago em impostos para o governo do país onde está localizado o campo. Neste ponto, será feita uma revisão da literatura tratando sobre os dois regimes fiscais em vigor no Brasil: partilha de produção e concessão.

A indústria de petróleo tem algumas características que impactam sua atividade como alto risco, custo fixo elevado, grande barreira a entrada, frequência da ocorrência de custos

afundados, incerteza com relação aos preços, e à viabilidade econômica dos campos. Em outras palavras é uma indústria de capital intensivo e com grande incerteza envolvida no processo decisório.

Em linha com essas características e considerando o artigo de Tirole (2015), a exploração do petróleo é um caso no qual a ferramenta ideal para maximizar o valor do governo é o leilão do direito de exploração de cada bloco exploratório, dando o monopólio da exploração daquela área à uma empresa ou grupo de empresas consorciadas. Dessa forma, há os benefícios da concorrência no momento da disputa pelo campo do petróleo entre as empresas, já que não seria eficiente ter várias empresas operando o mesmo campo dados os altos custos fixos.

Até 1995, apenas a Petrobras podia explorar o petróleo no Brasil, a partir de então, com uma emenda constitucional, foi criada a possibilidade do regime de concessão. Na mesma emenda foi criada a Agência Nacional do Petróleo, Gás Natural e B combustíveis (ANP), com o objetivo de ser uma agência independente com autonomia de decisão, nos moldes de outros países, como Reino Unido. Neste modelo, a propriedade do petróleo extraído é do concessionário. E ele remunera ao governo, além dos tributos convencionais incidentes sobre o lucro, imposto de renda (IR) e contribuição social sobre o lucro líquido (CSLL), outras quatro remunerações:

- *Bônus de assinatura*: Seu valor mínimo é definido no edital e o valor da proposta vencedora é recebido no início do contrato em parcela única. Em geral, bacias que têm maior probabilidade de ter um prospecto de qualidade e com bom retorno têm bônus de assinatura maiores, e quando há maior incerteza, o bônus é menor.

- *Royalties*: São pagos mensalmente com base na receita bruta. Garantem um ganho ao Estado, independentemente de o campo ser lucrativo ou não para a empresa. Internacionalmente varia entre 2% e 30%, mas usualmente fica entre 5% e 10% da receita bruta. No Brasil, a alíquota é de 10% para este regime fiscal, com algumas exceções que podem ser definidas pela ANP. O fato de os *royalties* incidirem sobre a receita bruta, ou seja, sem descontar os custos, estimula a empresa a reduzir seus custos. Basta que o campo esteja em operação para que o governo receba os *royalties*.

- *Participação especial*: possui alíquota entre 10% e 40% cobradas sobre a receita líquida da produção (deduzidos custos) da concessionária, devidos trimestralmente, quando esta fica muito alta. Essas taxas aumentam progressivamente com a receita líquida. Os interesses do governo e da empresa convergem neste caso, pois ambos querem aumentar a receita líquida.

Porém, como os custos são deduzidos, é gerado um incentivo para a empresa inflar seus custos, reduzindo a base da tributação.

- Taxa por ocupação ou retenção de área: É uma taxa, calculada pelo tamanho do bloco concedido. No edital é fixado um valor por km<sup>2</sup>, entre R\$10,00 e R\$5000,00, de acordo com a localização e as características geológicas do campo, que é pago anualmente. Essa taxa é utilizada em diversos países e o Brasil cobra uma das menores.

Já o regime de partilha foi instituído em 2010, após a descoberta das acumulações no pré-sal. Amaral (2014) diz que geralmente este regime é adotado juntamente com um discurso mais nacionalista, de maior presença do Estado. Tendo muito a ver com a situação política de cada país no momento da criação da legislação. Além do caráter nacional, busca-se aumentar a renda da União nestes campos de baixo risco e alto potencial.

Neste modelo, o petróleo continua sendo de propriedade da União, e esta cede para a operadora a quantidade em óleo referente ao custo da empresa e divide o excedente em óleo. Vence a licitação a empresa que oferecer em favor do Estado a maior quantidade de excedente em óleo. E a Petrobras tem o direito de integrar todos os consórcios com participação de até 30%, antes da licitação, apenas manifestando seu interesse. Se for de sua vontade, ela pode participar da licitação para aumentar sua participação.

O custo em óleo é o valor do custo para exploração e produção dado em óleo. O Estado pode fixar um valor máximo para essa recuperação de custo. O que sobra, é o chamado óleo lucro (*profit oil*), que é dividido entre a união e o contratante de acordo com a oferta ganhadora de licitação. Essa percentual pode variar de acordo com o preço do petróleo e com a produtividade dos poços, conforme tabela presente no edital.

Há o pagamento de *royalties* no percentual de 15% e de bônus de assinatura (já fixado no contrato), que não são critérios para vencer o leilão. O único critério é o maior valor de excedente em óleo.

Neste regime, o Estado deve conhecer muito bem as características de cada campo e do setor como um todo, principalmente para definir o excedente em óleo mínimo, o valor máximo da recuperação do custo em óleo no edital e para fiscalizar a produção de modo a verificar o preço praticado e quantidade produzida (Gomes, 2009). Caso contrário, as empresas podem inflar os custos e lucrar com um valor monetário que poderia estar fluindo para o Estado, ou então subfaturar o preço ou a produtividade para que o excedente em óleo para o governo seja menor.

Amaral (2014) faz uma analogia entre as participações especiais da concessão e o excedente em óleo da partilha. Ambos incidem apenas quando os custos de explorar o campo já foram deduzidos. Além disso, um aumento da produtividade e do preço do petróleo podem representar um aumento na fatia do governo.

É recorrente que as comparações de fração do governo indiquem uma maior parcela governamental no regime de partilha, conforme Gobetti e Serra (2011), Damé (2011) e Marques (2015).

De posse dos dados anteriores, é possível realizar o cálculo do VPL. Alguns artigos abordam de forma diferente o VPL a ser tratado.

O artigo de Lima, Suslick e Bordieri (2005) traz uma discussão sobre como estimar a volatilidade da produção de óleo, relacionando esta volatilidade com a volatilidade no preço do óleo e como isso pode impactar no VPL da oportunidade analisada. Foram analisados doze projetos, e eles chegaram à conclusão de que um aumento na volatilidade do preço do petróleo provoca um aumento ainda maior na volatilidade do projeto.

Lima, Suslick e Bordieri (2005) consideraram a seguinte formulação para o VPL:

$$VPL = \sum_{t=0}^T \frac{(P_{ano} - OPEX_{barril\ ano}) * Vol_{ano} * (1 - PG)}{(1 + i)^t} - \sum_{t=0}^N \frac{I_t}{(1 + i)^t} \quad (5)$$

onde,

$P_{ano}$  = Média anual do preço do óleo

$OPEX_{barril\ ano}$  = Média anual do custo operacional por barril

$Vol_{ano}$  = Volume de óleo produzido por ano

$I_t$  = CAPEX por ano

$i$  = Custo de capital do investimento,

$PG$  = Participação do governo em percentual

$t$  = Vida do projeto em anos

É importante destacar que nesta fórmula de VPL há uma simplificação grande ao inserir a participação do governo apenas como um percentual, sem discriminar quanto dessa participação vem dos impostos sobre a receita bruta ( $P_{ano} * Vol_{ano}$ ) e quanto vem dos impostos

sobre o lucro. Ou seja, esta fórmula não discrimina as características tributárias de cada regime fiscal, é mais genérica.

Blake e Roberts (2006) compilaram algumas fórmulas para o cálculo do VPL em diferentes regimes fiscais. Por exemplo, para o regime de concessão:

$$NCF = P_{ano} * Vol_{ano} - CAPEX_{ano} - OPEX_{ano} - TAX - ROY - BONUS \quad (6)$$

onde,

NCF = Fluxo de caixa líquido

$OPEX_{ano}$  = Média anual do custo operacional (OPEX)

ROY = *Royalties* pagos no ano

TAX = Demais taxas do governo

BONUS = Bônus de assinatura

E para o regime de partilha:

$$NCF = P_{ano} * Vol_{ano} - CAPEX_{ano} - OPEX_{ano} - ROY - BONUS - POG - TAX \quad (7)$$

onde,

POG = parcela do excedente em óleo pago para o governo

Estes dois modelos anteriores são mais completos ao separar as participações do governo de acordo com cada regime fiscal. Porém, não consideram a depreciação e para obter o VPL ainda seria preciso inserir o fator de desconto pela taxa mínima de atratividade.

Marques (2015) utiliza uma equação mais adequada para cálculo do fluxo de caixa de cada ano. Ela considera o abatimento da depreciação para fins de cálculo do lucro tributável, além de também discriminar os diferentes tributos dos regimes.

Para concessão, a fórmula utilizada é a seguinte:

$$NCF = (R - ROY - PIS - PE - OPEX - NDep - Dep_{equip}) * (1 - T) + Dep_{equip} \quad (8)$$

E para o regime de partilha:

$$NCF = (R - ROY - PIS - GS - OPEX - NDep - Dep_{equip}) * (1 - T) + Dep_{equip} \quad (9)$$

onde,

R = Receita

PIS = Taxa de contribuição social

PE = Participação especial

*NDep* = Investimentos não depreciáveis

*Dep<sub>equip</sub>* = Depreciação dos equipamentos

T = Taxa de imposto de renda

GS = Excedente em óleo do governo

No Capítulo 3 deste trabalho será visto que estas fórmulas se aproximam da metodologia utilizada nesta dissertação.

## 1.2 Métodos estatísticos e modelo de preços

Esta seção da dissertação dedicar-se-á ao detalhamento dos métodos estatísticos que podem ser utilizados para estimar o CAPEX e o OPEX.

Segundo Jansen (2018), existem dois objetivos ao tentar estimar uma função  $f$  que explique uma variável dependente em função de outra(s). O primeiro é a previsão. Frequentemente é útil prever com certa acurácia o comportamento de determinada variável no futuro. Essa acurácia depende do erro redutível, ou seja, aquele que pode ser reduzido ao se escolher a ferramenta estatística mais adequada e do erro irredutível que é representado pela parcela do erro da estimativa que não é possível alterar/reduzir pois está relacionado à natureza dos dados, quantidade de variáveis independentes disponíveis para análise, dentre outros fatores.

O segundo objetivo é a inferência, ou seja, identificar como uma variável dependente se comporta em relação às demais e tomar conclusões a partir disso. Quais as variáveis que mais impactam e qual a intensidade desta influência são questões de interesse em muitos casos.

Geralmente o modelo a ser escolhido depende do objetivo principal a ser atingido. Modelos mais simples como regressão linear são mais fáceis de realizar inferências, devido a sua maior possibilidade de interpretação. Porém, em geral, não possuem uma capacidade de previsão tão alta quanto os modelos mais flexíveis.

Também é possível testar mais de um modelo para estimar a relação entre os dados, porém é preciso utilizar critérios para escolha. Um critério tido como adequado na literatura, conforme defendido por Wallach e Goffinet (1989), Wu, McAuley e Harris (2011), Hocking (1976), dentre outros, para aferir a qualidade do ajuste da previsão fora da amostra é o erro quadrático médio (MSE) expresso por:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2, \quad (10)$$

onde  $n$  é o número de observações fora da amostra (*test set*),  $y_i$  é o valor observado da variável dependente e  $\hat{y}_i$  é o valor estimado da variável dependente.

O MSE diz o quão distante estão os valores previstos das observações, ou seja, é uma medida da acurácia de previsão do modelo e fornece uma estrutura lógica para comparar modelos (Wallach e Goffinet, 1989).

Para medir o MSE o ideal é que sejam separados dois conjuntos de dados. O primeiro no qual o ajuste será feito de forma a minimizar o MSE, chamado de dados de treinamento (*training set*), e o segundo (que não foi utilizado para estimar a função) no qual este ajuste será testado (*test set*), ou fora da amostra. Dessa forma, o desejável é que o erro no conjunto de dados de teste (MSE de teste) seja o menor possível. Utilizar os próprios dados de treinamento para medir o MSE subestima seu verdadeiro valor, pois os dados já foram utilizados para estimar o modelo.

Modelos mais flexíveis requerem uma estimação de maior número de parâmetros. Tais modelos podem gerar o que se denomina de *overfitting*. Este fenômeno significa que o modelo está buscando seguir o termo de erro muito proximamente. Por isso, é importante realizar testes

fora dos dados de treinamento, pois apesar de o erro estar diminuindo nestes dados, é possível que ao aplicar o modelo aos dados de teste (ou fora da amostra) obtenha-se uma previsão de valores com maior MSE.

Segundo Hawkins (2004), o *overfitting* decorre da falta de parcimônia do avaliador, que pode incluir variáveis em excesso ou realizar transformações irrelevantes nas mesmas ou também decorre do uso de modelos mais flexíveis do que o necessário para explicar o comportamento de um componente. É essencial a preocupação com o *overfitting* e a realização de medidas que favoreçam a sua identificação.

Há outras formas de medir o MSE de teste sem ser simplesmente separando dois conjuntos de dados e utilizando um deles como teste. Tratam-se dos métodos de reamostragem, dos quais dois serão lembrados nesta revisão da literatura. Esses métodos são muito úteis quando há uma quantidade pequena de dados disponíveis, como é o caso desta dissertação.

No *Leave-one-out cross-validation* (LOOCV), o modelo de ajuste é feito com  $n-1$  observações (onde  $n$  é o número de observações disponíveis para a análise). E a observação restante é utilizada para calcular o erro quadrático. Isso é repetido  $n$  vezes. Assim, é possível calcular o MSE como a média dos  $n$  erros quadráticos encontrados. Os pontos positivos desta metodologia são que o viés é menor, já que o ajuste é feito quase que na totalidade dos dados, e a aleatoriedade é menor do que ao separar apenas um conjunto de teste (James et al, 2013).

$$MSE \text{ de teste} = \frac{1}{n} * \sum_{i=1}^n MSE_i \quad (11)$$

No *K-fold cross-validation* a lógica é parecida com o LOOCV. Nele, ao invés de separar apenas uma observação da amostra, são separados  $k$  grupos. Por exemplo, no *5-fold cross-validation* o MSE é calculado em um dos 5 grupos a partir do ajuste do modelo nas observações dos outros 4 grupos. Esse processo é repetido 5 vezes e o MSE é a média dos 5 MSE calculados. Portanto, o LOOCV nada mais é do que um caso específico do *k-fold cross-validation* quando  $k$  é igual ao número de observações.

Comparando os dois modelos, o segundo tem uma vantagem computacional em relação ao primeiro. Quando o número de observações é muito grande, o LOOCV demanda maior capacidade de processamento. Porém, outro ponto importante na escolha da metodologia é o *trade-off* entre viés e variância. O LOOCV tem um viés bem baixo, já que o ajuste é feito em

uma quantidade maior de dados. Porém, este método acaba tendo uma variância maior do que no *k-fold cross-validation*. Dessa forma, diversos artigos, como James et al (2013), Rodriguez, Perez e Lozano (2010), e Anguita et al (2009), sugerem que valores de  $k$  iguais a 5 ou 10 tendem a não sofrer nem grande viés nem alta variância.

Continuando no tópico do *trade-off* entre viés e variância, este aspecto também surge no momento da escolha do modelo a ser utilizado para a regressão. Conforme a fórmula abaixo, o valor esperado do MSE de teste é composto pela variância da função (primeiro termo), pelo viés desta função (segundo termo) e pela variância do termo de erro (terceiro termo):

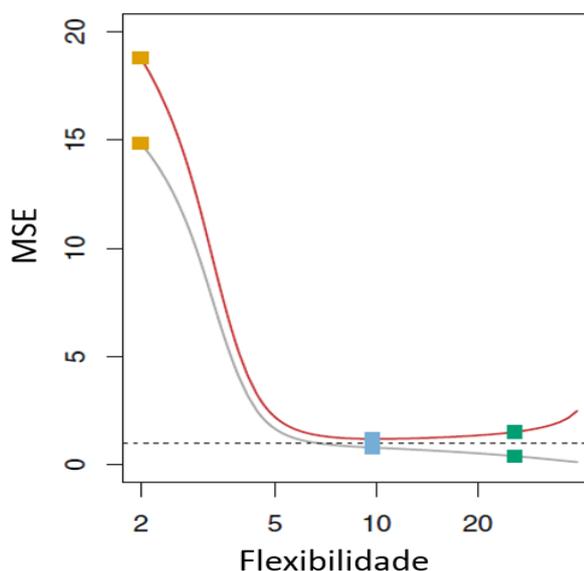
$$E(y_0 - \hat{y}_l)^2 = Var(\hat{y}_l) + [vies(\hat{y}_l)]^2 + Var(\epsilon) \quad (12)$$

onde  $\hat{y}_l = \hat{f}(x_0)$ . A variância do termo de erro (última parcela do segundo membro da equação) representa o erro irreduzível. A variância de um modelo é maior quando a alteração dos dados de treino gera uma variação grande na estimação de  $\hat{y}_l = \hat{f}(x_0)$ . Em geral, modelos flexíveis possuem uma variância maior. Uma variância alta não é desejável, pois um modelo que permite muita variação, dependendo da escolha das observações, não será robusto e estará mais sujeito ao *overfitting* (Briscoe e Feldman, 2010).

Já o viés de um método estatístico tende a ser maior em modelos mais rígidos, pois eles se adaptam pouco aos dados. Por exemplo, em relações funcionais não-lineares, uma regressão linear nem sempre vai conseguir representar a relação entre as variáveis. Um modelo mais flexível poderia se ajustar melhor aos dados e reduzir o viés encontrado.

Portanto, é preciso escolher um modelo que minimize o MSE de teste considerando esse *trade-off* entre variância e viés dos modelos mais ou menos flexíveis. Segundo James et al (2013), ao aumentar a flexibilidade de um modelo o viés reduz em uma taxa maior do que o aumento da variância até certo ponto. A partir de então o aumento da flexibilidade produz pouco impacto no viés e aumenta drasticamente a variância. É preciso buscar o ponto ótimo desta relação. Isso pode ser visto na Figura 2. Além disso, esta figura também mostra que o MSE de treinamento mantém a queda à medida que a flexibilidade aumenta. Já o MSE de teste (que pode ser estimado por *cross-validation*) não sofre deste problema, por isso é o mais adequado para a escolha do nível de flexibilidade do modelo.

Figura 2 – A curva superior representa o MSE de teste e a inferior o MSE de treinamento.



Fonte: James et al, 2013.

É importante ressaltar que dificilmente o viés e a variância podem ser medidos pela ausência de conhecimento da real função dos dados.

Entrando nos métodos propriamente ditos, a ferramenta mais utilizada para tentar encontrar a relação funcional entre os dados é a regressão linear. Quando se deseja explicar uma variável com base em outras, é utilizada a regressão linear múltipla expressa por:

$$y = \beta_0 + \beta_1 * x_1 + \beta_2 * x_2 + \dots + \beta_p * x_p + \epsilon \quad (13)$$

onde,  $y$  é a variável dependente,  $x$  representa as variáveis independentes,  $p$  é quantidade de regressores,  $\beta$  são os coeficientes da regressão e  $\epsilon$  o termo de erro.

Os valores de  $\beta$  podem ser traduzidos como o quanto  $y$  varia quando uma variável  $x$  varia (mantendo as outras constantes). Ou seja, representam o quanto as variáveis explicativas impactam no que se está querendo prever. E o sinal de  $\beta$  diz qual é a relação dessas variáveis (Montgomery, Peck e Vining, 2012).

A metodologia mais utilizada para estimação dos parâmetros é o de mínimos quadrados ordinários (MQO). Que busca obter os coeficientes que minimizam a expressão:

$$\sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^p \beta_j * x_{ij})^2 \quad (14)$$

onde  $n$  é número de observações.

Uma vez esclarecidos os conceitos do modelo, também é importante falar sobre a escolha das variáveis que irão compor o modelo. Afinal, raramente todas as variáveis disponíveis para a análise se relacionam com a variável dependente e um modelo com menos variáveis facilita a interpretação. Dessa forma, uma metodologia possível é realizar testes com todas as possibilidades de combinação de variáveis e ver qual tem o menor MSE de teste. Porém, dependendo da quantidade de variáveis preditoras fica muito custoso realizar esta análise já que o número de possíveis combinações pode ser muito alto. Assim, algumas metodologias que simplificam este problema surgem para auxiliar esta escolha.

A metodologia *forward stepwise* é iterativa e busca encontrar soluções sem ter que testar todas as possíveis combinações de variáveis preditoras. Ela começa com um modelo sem nenhuma variável independente e a cada iteração adiciona a variável que promove a maior melhoria no ajuste do modelo no sentido de aumento do  $R^2$ . A principal vantagem do *stepwise* é computacional, além da possibilidade de tornar viável essa escolha de variáveis. Porém, nem sempre o modelo escolhido será realmente o melhor dentre os possíveis, conforme mostra Wang (2016).

Isso ocorre, pois a variável a ser incluída na primeira iteração pode ser a que aumente mais o  $R^2$ . Assim, as iterações seguintes necessariamente irão contê-la. Porém, é possível que esta variável não esteja no modelo que fornece o melhor ajuste dos dados.

Existe também o *backward stepwise* que consiste na mesma lógica, porém ele inicia as iterações com todas as variáveis no modelo e vai retirando a que menos contribui para a previsão. Possui a mesma vantagem e desvantagem do anterior.

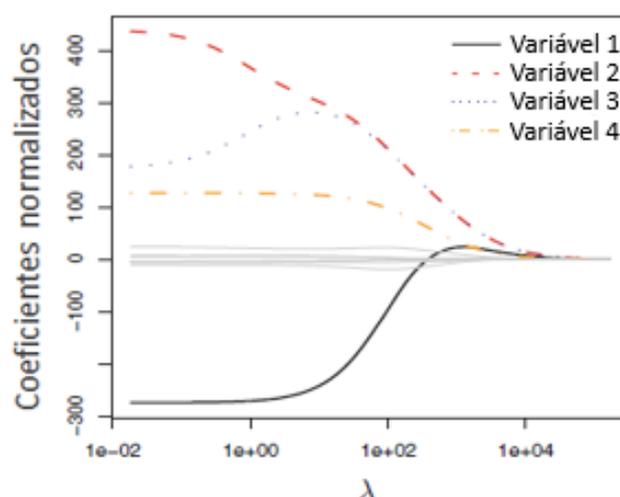
Outras metodologias para buscar um modelo com as melhores variáveis utilizam o encolhimento dos parâmetros como ferramenta. São elas: *ridge regression* e *lasso (least absolute shrinkage and selection operator)*.

No *ridge regression*, ao invés de minimizar a soma dos erros ao quadrado, da mesma forma que o MQO, minimiza-se a seguinte expressão:

$$\sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^i \beta_j * x_{ij})^2 + \lambda * \sum_{j=1}^p \beta_j^2 \quad (15)$$

Ou seja, além da expressão do MQO, ele multiplica a soma dos quadrados dos coeficientes por um valor  $\lambda$ . Ohishi, Yanagihara e Fujikoshi (2020), enfatizam que esta parcela da fórmula se torna um incentivo para reduzir o valor dos coeficientes em direção à zero. Esse incentivo depende do valor escolhido para  $\lambda$ . Se este valor for igual a zero, a regressão e os coeficientes serão iguais ao MQO. Se o valor tender ao infinito, e os coeficientes estimados serão próximos de zero, conforme é possível observar na Figura 3.

Figura 3 - Exemplo de comportamento dos coeficientes em função de  $\lambda$  no ridge regression.



Fonte: James et al, 2013.

No MQO uma mudança de escala nos parâmetros não produz diferença no resultado da regressão. Porém, no *ridge regression* é preciso observar que uma multiplicação nas variáveis independentes irá alterar significativamente os coeficientes estimados, devido ao último termo que está presente no valor que se deseja minimizar. Portanto, é essencial regularizar as variáveis independentes antes de aplicar esta técnica.

Para explicar as vantagens desta regressão em relação ao MQO, é preciso retornar ao conceito de *trade-off* entre viés e variância. Conforme foi dito, modelos mais flexíveis levam à uma maior variância. Conforme o valor de  $\lambda$  aumenta, o modelo fica mais distante do MQO e sua flexibilidade vai diminuindo (James et al, 2013). Assim, valores mais altos de  $\lambda$  levam à uma variância menor e um viés maior. Dessa forma, aumentando-se  $\lambda$  até um certo valor, há

uma queda grande na variância com pouco incremento no viés. Assim, o MSE reduz. Esta é a grande vantagem em relação ao MQO.

Uma característica do *ridge regression* é que nenhum dos coeficientes será igual a zero. Ou seja, ele não elimina do modelo as variáveis que têm pouca capacidade de explicação, como faz o método *stepwise*, por exemplo. Outra metodologia que envolve o encolhimento dos coeficientes, mas que permite que alguns sejam iguais a zero é o *lasso*.

Esta técnica minimiza a seguinte expressão:

$$\sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^p \beta_j * x_{ij})^2 + \lambda * \sum_{j=1}^p |\beta_j| \quad (16)$$

Assim, da mesma forma que o *ridge regression*, se  $\lambda$  for igual a zero, a equação se reduz à regressão por MQO, porém na medida que  $\lambda$  aumenta, o segundo termo vai ganhando mais importância. No caso do *lasso*, para maiores valores de  $\lambda$ , os coeficientes dos parâmetros que possuem menor poder de explicação da variável dependente são zerados. Quando  $\lambda$  tende ao infinito, todos os coeficientes serão zerados. Isso confere uma vantagem em relação ao *ridge regression*, já que permite uma maior facilidade de interpretação dos resultados (HANS, 2009).

Assim como no *ridge regression*, é preciso identificar o valor de  $\lambda$  que gera o menor MSE. Isso é feito por meio da estimação do MSE no conjunto de teste utilizando *cross-validation* para diversos valores de  $\lambda$ .

É importante esclarecer que não é possível garantir que uma metodologia será sempre melhor do que a outra. O ideal é realizar o ajuste com as duas técnicas e verificar qual produz o menor MSE utilizando *cross-validation*. Em geral, a literatura afirma que o *lasso* produz melhores resultados quando a variável dependente é explicada principalmente por poucos parâmetros e o *ridge regression* funciona melhor quando a variável dependente depende de muitos parâmetros, com pouca predominância de um sobre o outro (TIBSHIRANI, 1996).

No âmbito da regressão linear, alguns trabalhos fizeram estimação de variáveis tratadas nesta dissertação. Porém, nenhum deles explicitou qual metodologia foi utilizada para escolha das variáveis independentes.

Karlik (1991) listou alguns exemplos de parâmetros que poderiam ter correlação com o custo das instalações, entre eles: lâmina d'água, número de poços, tamanho do reservatório, área do campo e tempo de vida do projeto. Além disso, ao fazer a regressão utilizando o volume total de reservas como variável independente encontrou boa capacidade de explicação.

Mesz (2018) obteve, usando regressão linear, uma relação utilizando o volume de óleo como variável independente resultando em um  $R^2$  de 89,49%.

$$\widehat{CAPEX} = 7,1229 * Vol + 4956,6 \quad (17)$$

onde, Vol é o volume de óleo recuperável.

Apesar do  $R^2$  ter um valor relevante, esta regressão foi feita com um pequeno número de observações (10 apenas) e utilizou apenas um regressor. Portanto, é possível encontrar uma equação que preveja de forma mais precisa o CAPEX caso haja um maior número de observações e de variáveis candidatas a preditoras. Além disso, não foi disponibilizado o valor do MSE de teste.

Apesar de ter projetos de realidade diferente dos que serão abordados nesta dissertação, o trabalho de Kuznetsov et al (2011) trouxe uma metodologia para a obtenção do CAPEX a partir de parâmetros como volume e lâmina d'água. Estes parâmetros também serão testados nos métodos estatísticos aqui propostos.

Utilizando uma amostra de 73 campos localizados na Austrália, Surovtsev e Sungurov (2017) obtiveram uma relação linear entre o CAPEX e o volume no pico da curva de produção do campo. Esta premissa é válida, dado que a maior parcela do CAPEX geralmente é representada pela UEP que será construída ou afretada e a sua capacidade é dimensionada geralmente pelo pico da produção.

Jablonowski e Strachan (2008) estimaram o custo de investimento em dois tipos de plataformas flutuantes, a saber: Spar e TLP (plataforma de pernas atirantadas). Para a primeira, o pico de produção de óleo e a quantidade de *slots* para poços foram significativos ao tentar obter a regressão linear do custo total. Já para a segunda, além destas duas variáveis, o pico de produção de gás também foi significativo, obtendo  $R^2$  de 88% e 95%, respectivamente. Neste estudo, a lâmina d'água não obteve significância estatística para entrar nas regressões, o que não era esperado pelos autores. Porém, eles observaram que o problema pode estar na base de dados que se concentra muito em águas rasas.

Além da estimação do CAPEX, diversos trabalhos também buscaram obter o OPEX em função de outros parâmetros.

Mesz (2018), por exemplo, separou a estimação em custos fixos e variáveis. Para os custos fixos, foi estimada a relação custos fixos por ano de produção e obtida a seguinte regressão linear com  $R^2$  de 87,94%:

$$\frac{\widehat{CF}}{ano} = 0,0909 * Vol + 104,46 \quad (18)$$

Já os custos variáveis, foram obtidos sob a forma de custo por barril. Ou seja, utilizou-se uma regressão entre o custo variável dividido pela quantidade de barris de óleo equivalente produzidos por dia e o volume total. Assim, a equação com R<sup>2</sup> de 61.74% ficou da seguinte forma:

$$\frac{\widehat{CV}}{boepd} = 10^7 * Vol^2 - 0,0017 * Vol + 10,095 \quad (19)$$

Da mesma forma que ocorreu com o CAPEX, o baixo número de observações e de variáveis preditoras pode ter prejudicado a estimação do modelo.

Além das estimativas por meio da regressão linear, este trabalho, diferentemente dos já citados, utilizará metodologias da disciplina de Aprendizagem Estatística que permitem maior flexibilidade no modelo para previsão das variáveis dependentes. A expectativa é que se encontre um ajuste com um erro quadrático médio menor do que os modelos mais rígidos. Abaixo serão apresentadas estas metodologias.

Uma delas começa com a descrição da regressão *splines*. James et al (2013) faz a explicação a partir da escolha de um modelo polinomial de grau 3 para explicar determinada variável dependente, e dividindo a regressão em dois intervalos:

$$y_i = \begin{cases} \beta_{01} + \beta_{11} * x_i + \beta_{21} * x_i^2 + \beta_{31} * x_i^3 + \epsilon_i, & se x_i < c \\ \beta_{02} + \beta_{12} * x_i + \beta_{22} * x_i^2 + \beta_{32} * x_i^3 + \epsilon_i, & se x_i \geq c \end{cases} \quad (20)$$

O ponto  $c$  no qual a equação muda é denominado de nó. Quanto mais nós, mais flexível é o modelo.

A descontinuidade é algo que muitas vezes se deseja evitar. Assim, é possível ajustar um modelo que esteja sujeito à restrição de que a curva resultante seja contínua. Porém, esta curva ainda terá um formato não muito usual na junção, fazendo com que surja um comportamento em “v” no encontro dos nós.

É possível retirar este formato ao sujeitar o modelo a mais duas restrições: que a primeira e a segunda derivadas também sejam contínuas.

Mas como é possível realizar essas restrições na estimação? Utilizando um modelo base, onde  $k$  é o número de nós. Primeiramente, a regressão ficaria da seguinte forma:

$$y_i = \beta_0 + \beta_1 * b_1(x_i) + \beta_2 * b_2(x_i) + \dots + \beta_{k+3} * b_{k+3}(x_i) + \epsilon_i \quad (21)$$

Para representar um *spline* de grau 3, conforme tem sido exemplificado, é preciso utilizar uma função para truncar o encontro em cada um dos nós. Essa função truncada é definida como:

$$h(x, \xi) = (x - \xi)^3 = \begin{cases} (x - \xi)^3, & \text{se } x > \xi \\ 0, & \text{se } x \leq \xi \end{cases} \quad (22)$$

Assim, a equação *spline* fica da seguinte forma:

$$y_i = \beta_0 + \beta_1 * x_1 + \beta_2 * x_2^2 + \beta_3 * x_3^3 + \beta_4 * h(x, \xi_1) + \dots + \beta_{k+3} * h(x, \xi_k) + \epsilon_i \quad (23)$$

O que dá um total de  $k+4$  coeficientes para estimar.

Para melhorar ainda mais esse tipo de regressão, é adicionado uma última restrição: Que nos valores iniciais e finais da distribuição, ou seja, nas fronteiras, a função seja linear. Essa regressão é chamada de *natural splines*. Elas resultam em um modelo mais suavizado e mais estável nas bordas. Assim, geralmente representam melhor a realidade.

A quantidade de nós é uma informação importante para esse tipo de análise. Em geral, seria ideal colocar mais nós em locais onde as observações têm maior variabilidade para poder representar essa flexibilidade no modelo. Porém, é mais comum que os nós sejam igualmente distribuídos pelos dados. Em analogia às situações anteriores, o *cross validation* pode ser uma boa alternativa para definir a quantidade de nós a se utilizar.

Existe outro tipo de *spline*, chamado de *smoothing spline*. Esta metodologia também busca suavizar a curva obtida na regressão, porém minimizando a seguinte expressão:

$$\sum_{i=1}^n (y_i - g(x_i))^2 + \lambda \int g''(t)^2 dt \quad (24)$$

O primeiro termo já é conhecido. Já o segundo termo é um valor que tenta penalizar a variabilidade da função. Ou seja, a segunda derivada representa os declives ou aclives na equação. Se a função muda muito de direção e em um grau alto, a segunda derivada será maior. Outra característica dessa metodologia é que ela considera nós em cada  $x_i$ .

O valor de  $\lambda$  faz a calibração de quão suave a curva deve ser. Para valores mais altos, o segundo termo ganha mais peso, portanto maior a suavidade da curva resultante. Quando for igual a zero, seguirá exatamente os pontos observados através da interpolação.

Mais uma vez, o valor ótimo deste parâmetro pode ser escolhido por *cross validation*. Yuan (2006), confirma que este é o modo mais comum de estimar este parâmetro, mas alerta para a maior complexidade computacional.

Outra forma mais flexível de obter uma curva é a partir da regressão local. Nela, o ajuste é feito para cada ponto dos dados utilizando apenas os dados mais próximos desta observação. É possível definir pesos maiores para variáveis mais próximas ao dado ajustado. Além disso, este ajuste pode ser feito utilizando o modelo linear, ou polinomial (Loader, 1999).

A quantidade de dados a serem utilizados em cada ajuste é dada pela letra  $k$ . Com isso, o parâmetro que deve ser definido neste tipo de regressão é o  $\alpha$ , que é dado pela fórmula:

$$\alpha = \frac{k}{n} \quad (25)$$

Ou seja,  $s$  representa o percentual de dados que estão sendo utilizados para ajustar aquele modelo. Quanto menor o valor de  $\alpha$ , o modelo tende a ser mais curvo. Se  $\alpha$  for igual a 1, equivale a ajustar um modelo com todos os dados.

Todos os modelos descritos nos parágrafos anteriores foram exemplificados utilizando somente uma variável explicativa. Porém, é possível ajustar modelos flexíveis utilizando mais variáveis. Isso é feito por meio de um modelo aditivo generalizado (GAM).

Estes modelos seguem a mesma forma aditiva da regressão linear, porém permitem ajustes de funções não-lineares em cada uma das variáveis (Wood, 2017).

Logo, para obter sua forma, substitui-se os termos  $\beta_j * x_{ij}$  da fórmula de regressão linear múltipla pelo termo  $f_j(x_{ij})$ :

$$y_i = \beta_0 + \sum_{j=1}^p f_j(x_{ij}) + \epsilon_i \quad (26)$$

Assim, para cada variável independente, pode-se ter um tipo de função como *smoothing splines* ou regressão local. Portanto, cada variável pode seguir um modelo não-linear, enquanto o GAM como um todo segue a linearidade. Isso permite maior acurácia na predição, dado que

os ajustes podem ser melhores ao se utilizar modelo flexíveis. Também permite uma boa interpretação ao poder associar cada preditor diretamente com a variável dependente.

Nas referências consultadas, não foi encontrado nenhum modelo de estimação para o CAPEX e o OPEX por meio destes modelos mais flexíveis.

Para os preços do petróleo, há uma grande variedade de técnicas utilizadas para previsão/simulação. Muitas delas baseiam-se na descrição dos preços por processos estocásticos como o Movimento Browniano Geométrico (GBM) e de reversão à média.

Hahn e Dyer (2005) afirmaram que, ao supor que o preço segue um processo estocástico com distribuição log-normal, cuja variância cresce linearmente com o tempo, o GBM não consegue capturar o fenômeno de reversão à média evidenciado empiricamente para os preços do petróleo. Em seu artigo, eles apresentaram uma metodologia com enfoque discreto e a evolução dos preços seguindo uma árvore binomial, na qual os preços muito extremos são censurados e os preços convergem para os valores discretizados do processo Ornstein-Uhlenbeck.

Olsen et al (2005), questionou a robustez das avaliações econômicas que utilizam 3 cenários de preços (pessimista, moderado e otimista), procedimento muito comum para representar a incerteza deste fator. O autor analisou esta técnica e outras 4 (*Bootstrap*, *Inverted Hockey Stick* (IHS), histórica e Simulação Gaussiana Sequencial (SGS)) em 26 projetos para verificar como se comportava a incerteza associada a indicadores econômicos como VPL e taxa interna de retorno (TIR). A metodologia *Bootstrap* assume que a volatilidade dos preços passados será repetida no futuro. Dessa forma, múltiplas previsões de preços são geradas aleatoriamente ao amostrar variações de preços históricas. A vantagem deste método é que ele pode realizar milhares de amostragens e obter uma distribuição probabilística completa. A desvantagem é que esta distribuição pode conter valores tão extremos que nunca foram observados.

Akilu, Mcvay e Lee (2006) propuseram a metodologia IHS que busca unir a facilidade de cálculo dos métodos convencionais com a capacidade de reproduzir incertezas dos métodos estocásticos. Ela se baseou na possibilidade de alterações bruscas e rápidas nos preços do início da previsão (que provocam mais impactos econômicos). Com isso, considera valores máximos e mínimos históricos e as maiores subidas e descidas dos preços. Os autores posteriormente limitaram as subidas e descidas em 70% da diferença entre o preço de início da previsão e o

preço limite histórico. Assim, foi feita uma previsão otimista e uma pessimista. Para o cenário moderado foi considerado um valor constante a partir da observação mais recente.

A técnica histórica para previsão de preços foi exemplificada em Olsen et al (2005). A partir de um conjunto de dados históricos de preços entre os anos de 1974 e 2002, foram definidos 29 intervalos de 10 anos de duração. Esses intervalos são alternativas de previsão para os preços a partir da última observação.

A SGS, utilizada em Holmes (2006), gera vários cenários equiprováveis utilizando de ferramentas como distribuição normal padrão. Pela dificuldade de cálculo e excesso de rigor não será utilizado neste trabalho.

Em todos os 26 projetos a previsão convencional previu um intervalo mais estreito de valores dos indicadores em comparação com os outros 4 métodos. O que sugere que esta prática tradicional subestima a incerteza nos preços. Segundo os autores, a metodologia SGS é a mais rigorosa e com o melhor resultado, apesar de ser mais difícil de aplicar. Porém, a que possui melhor relação custo-benefício é a IHS.

O trabalho de Souza (2019), fez uma extensa revisão de metodologias ligadas a processos estocásticos com utilização de um ou mais fatores para descrição do preço de commodities. O trabalho de Schwartz e Smith (2000) é citado e descrito como um modelo dois fatores. O primeiro fator está relacionado as variações de curto prazo, que no caso de preços de petróleo pode ser uma crise política temporária ou uma variação momentânea dos estoques, por exemplo. O segundo fator, de longo prazo, está relacionado com um impacto permanente nos preços do petróleo, como avanços significativos em outras fontes de energia que possam substituir os derivados do petróleo, mudanças regulatórias ou mesmo descobertas/produção relacionadas ao uso de uma nova tecnologia que traga grande impacto na oferta mundial. Aiube e Samanez (2014), ressaltam um impacto de longo prazo na indústria de óleo e gás. Destacam o uso de uma nova tecnologia de fraturamento hidráulico em poços horizontais que ficou conhecida como shale oil/gas. A maior produtividade, assim obtida, trouxe um efeito negativo nos preços, resultante da maior oferta.

Como dito, o primeiro fator tem caráter temporário e tende a se dissipar, portanto reverte à média zero e foi modelado como um processo de reversão à média do tipo Ornstein-Uhlenbeck. Dessa forma, assim como outros trabalhos e evidências empíricas, o preço do petróleo possui um efeito de reversão à média no curto prazo.

O segundo componente é modelado como GBM. Neste fator existe uma variável que indica a tendência de longo prazo da indústria.

Assim, o logaritmo do preço à vista é obtido da seguinte forma:

$$\ln S_t = X_{1t} + X_{2t} \quad (27)$$

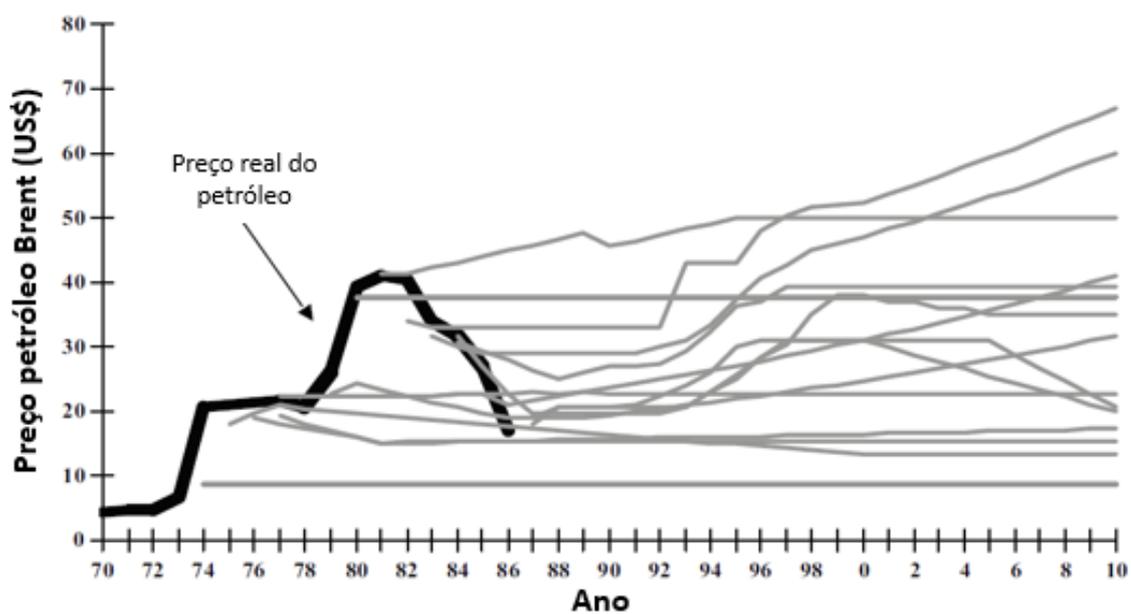
$$dX_{1t} = -kX_{1t}dt + \sigma_{X1}dB_{X1,t} \quad (28)$$

$$dX_{2t} = \mu dt + \sigma_{X2}dB_{X2,t} , \quad (29)$$

Onde  $S_t$  é o preço à vista,  $k$  é velocidade de reversão,  $\sigma_{X1}$  é a volatilidade do fator de curto prazo,  $\mu$  é a tendência de longo prazo,  $\sigma_{X2}$  é a volatilidade do fator de longo prazo e  $dB_{X1,t}$ ,  $dB_{X2,t}$  são incrementos do processo Browniano padrão para as duas variáveis. Além disso, as variáveis são correlacionadas por  $\rho_{X1X2}dt = dB_{X1,t} dB_{X2,t}$ .

Para finalizar a discussão sobre a estimação dos preços futuros, é importante mencionar uma imagem (Figura 4) presente em Tobin (2008) que mostra as projeções de preços futuros de uma grande empresa de petróleo da década de 70. A cada ano essas projeções são atualizadas, porém é possível perceber que raramente essas estimativas apontam para uma queda futura no preço, mas sim para um patamar constante ou crescente. Este fenômeno é conhecido com ancoragem. Ou seja, assume-se que o que está ocorrendo hoje continuará em vigor no futuro. Este é mais um motivo para utilizar processos estocásticos para estimar os preços futuros do óleo.

Figura 4 - Previsões do preço do petróleo Brent.



Fonte: Tobin, 2008.

O modelo de simulação do preço do óleo utilizado por esta dissertação será detalhado no Capítulo 3.

## 2. METODOLOGIA E DADOS

Este capítulo se dedicará a descrever a metodologia, caracterizar os dados utilizados nesta dissertação e apresentar as principais informações individuais e em conjunto dos mesmos.

### 2.1 Metodologia

Para o desenvolvimento desta dissertação foi necessário abranger conteúdos de diversas áreas de conhecimento, como economia, estatística, finanças e engenharia de petróleo. Dessa forma, é possível classificar este trabalho como interdisciplinar.

Pode-se classificar também como uma pesquisa aplicada, já que foi baseada em casos reais e houve uma aplicação prática dos conteúdos abordados.

Também é possível classificar a pesquisa como descritiva, já que busca relações entre as variáveis e estimar modelos de previsão. Além disso, este trabalho segue uma abordagem quantitativa.

A metodologia desta dissertação consiste no uso dos métodos descritos no Tópico 1.2 aplicados a uma base de dados específica para obter modelos de previsão de CAPEX e OPEX de projetos de produção de petróleo. Após esta aplicação, será escolhido o modelo com melhor capacidade de previsão, ou seja, o que obteve o menor MSE calculado a partir do *k-fold cross-validation* utilizando  $k = 5$  e  $k = 10$ , conforme discutido na Revisão da Literatura. Dessa forma, será examinado o valor de  $k$  mais adequado para utilização.

No tópico seguinte deste capítulo serão detalhados os dados e as variáveis explicativas disponíveis a serem utilizadas como candidatas nos modelos de previsão.

É importante ressaltar que os resultados obtidos não podem ser generalizados para outras bases de dados.

Após a aplicação e escolha do melhor modelo, ele será utilizado para a estimação dos valores de CAPEX e OPEX de um prospecto exploratório, que está descrito no Capítulo 4.

É importante ressaltar que por se tratar de uma oportunidade exploratória, serão simulados 10.000 possíveis valores de volumes a serem encontrados no reservatório. Da mesma forma serão simuladas 10.000 trajetórias de preços à vista utilizando o modelo presente no Tópico 3.3.

Dando continuidade à dissertação, será construído o fluxo de caixa do projeto nos regimes de concessão e de partilha de produção, conforme os modelos de fluxo de caixa do Tópico 3.5.

Dessa forma, serão obtidos 10.000 fluxos de caixa diferentes, resultando em uma distribuição de VPLs tanto para o regime de concessão quanto para o regime de partilha de produção.

## **2.2 Dados**

Estes dados foram obtidos na IHS Markit. Por questão de sigilo, os projetos referentes aos dados não serão revelados. O software de programação R será utilizado como meio para realizar todas as análises presentes nesta dissertação. Ele tem a vantagem de ser gratuito, possibilitar a aplicação de uma enorme variedade de ferramentas estatísticas, e possuir bastante conteúdo de ensino disponível na internet. Foram utilizadas as seguintes bibliotecas: ridge, mgcv, gamclass, boot, glmnet e mass.

O número de observações na amostra obtida é de 54 projetos de desenvolvimento da produção, todos eles localizados na camada do pré-sal.

Como já comentado, as variáveis dependentes que se quer prever são o CAPEX e o OPEX. E os parâmetros disponíveis candidatos a regressores são: Volume recuperável de óleo ( $V_{\text{óleo}}$ ), volume recuperável de gás ( $V_{\text{gás}}$ ), volume recuperável de hidrocarboneto ( $V_{\text{total}}$ ), profundidade do reservatório (Prof), soterramento do reservatório ( $Sot$ ), lâmina d'água ( $LDA$ ) e área do campo ( $area$ ).

Como são correlacionados, quando  $V_{\text{total}}$  for incluído na regressão,  $V_{\text{óleo}}$  e  $V_{\text{gás}}$  não poderão estar. Além disso, pela profundidade do reservatório ser a soma do soterramento com a lâmina d'água, ele não será utilizado como regressor.

Após a análise das distribuições dessas variáveis, três observações *outliers* foram identificadas por meio da técnica que utiliza a amplitude interquartílica para definir os limites superiores e inferiores (Moore, McCabe e Craig, 2009). O quadro abaixo mostra as principais estatísticas descritivas das observações que serão utilizadas neste trabalho. Todos os valores monetários estão expressos em termos reais e trazidos a valor presente para 2020 pela TMA de 10%, conforme será feito no restante do trabalho. Importante ressaltar que foi utilizado um fluxo real e que a TMA não tem inflação, ou seja, é uma taxa real. Esta tabela é baseada no conteúdo fornecido pela IHS Markit (Copyright © IHS Markit, 2020. Todos os direitos reservados).

Tabela 1 - Estatística descritiva dos dados

Estatística	CAPEX	OPEX	Vóleo	Vgás	Vtotal	Sot	LDA	Area
Unidade	MM US\$	MM US\$	MM bbl	MM boe	MM boe	M	M	Km <sup>2</sup>
Mínimo	336,90	654,90	49,99	1,72	53,44	444,00	105,00	6,36
Média	4166,60	5089,90	432,59	130,33	562,92	3083,00	1848,00	165,32
Mediana	3702,80	5601,20	489,00	120,85	625,14	3047,00	1964,00	143,76
Máximo	8968,40	8259,00	650,00	557,51	1007,51	5239,00	2709,00	602,90
Variância	4268910,00	5563670,00	36423,61	10674,07	74451,92	650125,00	274885,80	15502,05
Desvio Padrão	2066,13	2358,74	190,85	103,32	272,86	806,30	524,30	124,51

Fonte: O autor, 2020.

Por estas estatísticas já é possível realizar algumas observações. Em geral, o fluido predominante é de óleo, o que está condizente com as mais recentes publicações na ANP<sup>1</sup> em que se pode constatar que o óleo representa cerca de 80% de toda a produção no pré-sal.

<sup>1</sup> Agência Nacional do Petróleo, Gás Natural e Bicomustíveis. *Boletim da produção de Petróleo e gás natural*. Fev. 2020.

Como era de se esperar, estes projetos concentram-se mais em águas ultraprofundas, ou seja, acima de 1500 metros (Buckley e Uehara, 2017). Tornando mais desafiador o desenvolvimento da produção e exigindo volumes de investimentos maiores do que em campos em águas rasas.

Também é importante analisar as relações entre as variáveis dependentes e as variáveis explicativas uma a uma. A tabela abaixo mostra o coeficiente de correlação linear de Pearson entre estas variáveis e o CAPEX.

Tabela 2 - Correlação de Pearson entre o CAPEX e as variáveis independentes.

Variável	Correlação de Pearson
Vtotal	0,53
Vóleo	0,50
Vgás	0,47
Sot	0,26
LDA	0,50
Área	0,53

Fonte: O autor, 2020.

De acordo com a definição de Stevenson (1986), pode-se dizer que as relações entre as variáveis e o CAPEX são moderadas, com exceção do soterramento que tem uma relação mais fraca.

Parece que o CAPEX tem uma relação maior com os volumes, indicando que a regressão linear pode ser uma relação funcional satisfatória.

A variável *area* também apresentou correlação elevada. Trata-se de um resultado esperado já que para ter aproveitamento econômico de uma área maior, geralmente depende-se mais com linhas e demais gastos de infraestrutura submarina, por exemplo.

O soterramento apresentou uma correlação mais baixa. Já a lâmina d'água apresentou uma correlação maior, indicando que esta variável impacta mais nos custos de investimento destes projetos.

Fazendo a mesma análise para o OPEX, obtemos as seguintes relações:

Tabela 3 - Correlação de Pearson entre o OPEX e as variáveis independentes.

Variável	Correlação de Pearson
Vtotal	0,68
Vóleo	0,67
Vgás	0,55
Sot	0,11
LDA	0,28
Área	0,57

Fonte: O autor, 2020.

Assim como na análise das relações com o CAPEX, os volumes recuperáveis tiveram correlação mais elevada com o OPEX. Isso faz sentido, pois quanto maior a quantidade de fluido processado, maiores os gastos do dia a dia da operação e com serviços associados. Portanto, são fortes candidatos para explicar os valores de custos operacionais de cada projeto. Além destas duas variáveis, a área também está mais fortemente associada com o OPEX do que com o CAPEX.

Por outro lado, as duas medidas de profundidade tiveram uma correlação maior com os custos de investimento. Isto está alinhado com a teoria de avaliação de projetos. Estas variáveis impactam mais no momento do desenvolvimento do projeto, com o dimensionamento da quantidade de UEPs, comprimento das linhas e custos de poços. Porém, apesar disso e como visto nos gráficos, estas duas características dos projetos também tem certa relação com os custos operacionais, já que os custos de manutenção dos itens descritos anteriormente se estendem por todo o projeto.

Também é importante verificar o quanto as variáveis independentes são correlacionadas entre si, para evitar o problema de multicolinearidade. Caso este fenômeno ocorra, a estimação dos coeficientes será prejudicada, afetando sua precisão e sua variância, conforme demonstra Montgomery, Peck e Vining (2012).

Para uma análise preliminar, a observação da matriz de correlação entre as variáveis explicativas pode levar a algumas pistas sobre possíveis problemas de multicolinearidade. Conforme tabela 4, o  $V_{total}$  obviamente tem uma correlação alta com o  $V_{óleo}$  e o  $V_{gás}$ . Além disso, a *area* também se correlaciona de forma moderada com estas variáveis.

Tabela 4 - Correlação de Pearson entre as variáveis independentes.

	Vóleo	Vgás	Vtotal	Sot	LDA	Area
Vóleo	1,000	0,694	0,962	0,160	0,345	0,490
Vgás	0,694	1,000	0,864	0,234	0,376	0,715
Vtotal	0,962	0,864	1,000	0,200766	0,384	0,613
Sot	0,160	0,234	0,201	1,000	0,052	0,196
LDA	0,345	0,376	0,384	0,052	1,000	0,422
Area	0,490	0,715	0,613	0,196	0,422	1,000

Fonte: O autor, 2020.

Porém, para deixar esta análise mais objetiva será utilizada a metodologia de fator de inflação de variância (VIF). Este valor é estimado para cada variável contida na regressão. Conforme Montgomery, Peck e Vining (2012), este valor segue a fórmula abaixo, onde o R-quadrado é obtido a partir da regressão da variável independente em análise em função dos outros regressores.

$$VIF_j = \frac{1}{1 - R_j^2} \quad (30)$$

Portanto, para estimar estes valores para cada variável foram feitas três simulações. A primeira contendo os seis candidatos a regressores, a segunda contendo todas menos o  $V_{total}$  e a terceira contendo todas menos o  $V_{óleo}$  e o  $V_{gás}$ . A tabela abaixo resume os resultados

Tabela 5 - VIF para cada variável independente em cada regressão.

	Vóleo	Vgás	Vtotal	Sot	LDA	Area
Primeira regressão	21,20	17,30	0,02	91,31	78,27	0,07
Segunda regressão	1,96	3,04	-	1,06	1,26	2,18
Terceira Regressão	-	-	1,67	1,05	1,73	1,26

Fonte: O autor, 2020.

Quanto mais próximo de 1, menor a multicolinearidade dos regressores. A literatura admite que um valor acima de 10 pode trazer grandes problemas de estimação. Como era de se suspeitar, a primeira regressão, que incluiu todas as variáveis, teve vários problemas desse tipo. Já a segunda e a terceira apresentaram valores aceitáveis, ou seja, estas variáveis podem ser incluídas em conjunto no modelo. Destaca-se o fato de que a inserção apenas do  $V_{total}$  no lugar dos volumes separadamente resultou em valores mais satisfatórios.

### 3. MODELOS

A partir dos dados detalhados no capítulo anterior, serão testadas diversas metodologias que busquem explicar o CAPEX e OPEX, até que se escolha a que possui menor erro quadrático médio (MSE).

O primeiro modelo testado será o de forma funcional mais simples, ou seja, a regressão linear. A partir dele, serão utilizadas 3 metodologias para tentar otimizar o modelo linear base. Estas metodologias serão: *stepwise*, *ridge regression* e *lasso*. Todas elas foram apresentadas de forma mais detalhada no Capítulo 1.

Na segunda parte deste Capítulo, serão feitas estimativas utilizando metodologias mais flexíveis, que buscam obter maior aderência aos dados: regressão local e *smoothing splines*.

O tópico seguinte deste Capítulo se dedicará ao método para a simulação do preço do petróleo. E na quarta parte do capítulo será detalhada a simulação da curva de produção.

A última parte irá focar na consideração da participação governamental, na distribuição dos fluxos de caixa ao longo do tempo, e finalmente no cálculo do VPL.

#### 3.1. Modelos lineares

Seguindo a estrutura descrita no tópico anterior, serão feitas 2 regressões para o CAPEX. A primeira utilizando os volumes de óleo e gás, separadamente. A segunda utilizando o volume total devido ao problema de multicolinearidade descrito no Capítulo 2. O mesmo será feito para o OPEX.

Todas as equações estimadas respeitaram as hipóteses de normalidade dos resíduos e de homocedasticidade. Estes testes foram feitos respectivamente pelo teste Jarque-Bera, que conforme Jarque, Bera (1980), a hipótese nula é que a curtose e a assimetria dos resíduos seguem uma distribuição normal, e pelo teste ARCH (Autoregressive conditional heteroscedasticity), cuja hipótese nula é que não há heterocedasticidade nos resíduos.

A escolha do modelo com melhor poder de previsão será feita pelo MSE, conforme indicado por Wallach e Goffinet (1989) e Wu, McAuley e Harris (2011), por exemplo. Apesar de o  $R^2$  também ser muitas vezes utilizado para a escolha entre ajustes, é importante destacar que o MSE é mais adequado quando se trata de avaliar a capacidade de um modelo prever valores que não estão entre os dados utilizados para fazer o ajuste, logo ele é um melhor indicador do poder preditivo do modelo.

Abaixo estão os resultados para as duas regressões do CAPEX:

$$\widehat{CAPEX} = -1311,7 + 3,2 V_{\text{óleo}} - 2,0 V_{\text{gás}} + 0,4 Sot + 1,2 LDA + 5,0 area \quad (31)$$

$$P \text{ valor} = (0,2962) (0,0612) (0,6006) (0,1461) (0,0173) (0,0707)$$

$$R^2 = 0,3983$$

$$MSE = 2.783.681$$

$$\widehat{CAPEX} = -962,9 + 1,8 V_{\text{total}} + 0,4 Sot + 1,2 LDA + 3,7 area \quad (32)$$

$$P \text{ valor} = (0,4256) (0,0914) (0,1716) (0,0163) (0,1305)$$

$$R^2 = 0,3969$$

$$MSE = 2.877.697$$

O valor do  $R^2$  ajustado foi praticamente igual para os dois modelos. Outra observação importante é com relação ao sinal das variáveis. O esperado é que todos os sinais sejam positivos, de acordo com o que foi visto na análise dos dados e conforme a teoria de avaliação econômica deste tipo de projeto. Porém, isso ocorreu apenas no segundo modelo.

Com relação ao teste de significância dos coeficientes, o primeiro modelo apresentou mais variáveis significativas:  $LDA$ ,  $V_{\acute{o}leo}$  e  $area$  ao nível de 10%. Enquanto no segundo apenas o  $V_{total}$  e a  $LDA$  foram significantes.

Com relação ao MSE, foi utilizada a metodologia de *K-fold cross validation*, usando  $k$  igual a 10, conforme descrito na revisão da literatura. Considerando estes dois modelos com todas as variáveis incluídas, o primeiro modelo teve uma resposta melhor para a previsão do CAPEX.

A partir destes resultados, foram estimadas equações contendo apenas as variáveis significativas, inclusive sem o intercepto para o caso das duas regressões acima. O melhor resultado está no quadro abaixo.

$$\widehat{CAPEX} = 2,9 V_{\acute{o}leo} + 1,2 LDA + 4,6 area \quad (33)$$

$$P \text{ valor} = (0,0306) (0,0004) (0,0361)$$

$$R^2 = 0,8825$$

$$MSE = 2.705.727$$

Houve uma melhora substancial em comparação às primeiras equações. O  $R^2$  mais que duplicou e a estatística de erro também reduziu. Dessa forma, até o momento, a equação 33 é a melhor equação obtida para a previsão do CAPEX.

As regressões para o OPEX serão apresentadas a seguir:

$$\widehat{OPEX} = 1783,3 + 7,2 V_{\acute{o}leo} - 2,6 V_{g\acute{a}s} - 0,1 Sot - 0,2 LDA + 7,4 area \quad (34)$$

$$P \text{ valor} = (0,1793) (0,0001) (0,5218) (0,7366) (0,6874) (0,0116)$$

$$R^2 = 0,4876$$

$$MSE = 3.324.452$$

$$\widehat{OPEX} = 2439,8 + 4,7 V_{total} - 0,1 Sot - 0,2 LDA + 5,0 area \quad (35)$$

$$P \text{ valor} = (0,0658) (0,0002) (0,6229) (0,7252) (0,0581)$$

$$R^2 = 0,4592$$

$$MSE = 3.662.573$$

Pelos resultados, é possível observar que a primeira equação possui um maior ajuste aos dados, graças ao maior  $R^2$  e possui um menor MSE. Além disso, a significância dos coeficientes dos regressores foi mais satisfatória. Porém, assim como feito para o CAPEX, é possível obter um melhor resultado ao utilizar apenas as variáveis significativas

$$\widehat{OPEX} = 8,8 V_{\acute{o}leo} + 6,5 area \quad (36)$$

$$P \text{ valor} = (0,0001) (0,0054)$$

$$R^2 = 0,9059$$

$$MSE = 2.981.588$$

Houve uma grande melhora ao reduzir o número de variáveis. O  $R^2$  quase dobrou e o MSE reduziu mais de 10%. Assim, esta é a melhor equação para a estimativa do OPEX até o momento.

Da mesma forma que Mesz (2018), Lucchesi (2019) e Suslick et al (2003) fizeram, foi realizada uma tentativa de obter uma equação utilizando o OPEX por barril como variável dependente. O melhor  $R^2$  encontrado foi de 0,2111. Devido a isso, a problemas de significância dos coeficientes e a não aderência à hipótese de normalidade dos resíduos, conclui-se que não houve ganho ao utilizar o valor de OPEX por barril. Dessa forma, o restante desta dissertação concentrar-se-á apenas nos modelos que utilizam o OPEX total.

Para fins de curiosidade científica e comparação com a literatura que diz que o OPEX pode ser previsto em função do CAPEX, como Surovtsev e Sungurov (2017), e Verre, Guibileo e Cadegiani (2009), foi feita uma regressão utilizando os dados de CAPEX para explicar o OPEX, que resultou em um  $R^2$  de 0,9016 e significância no nível de 0,1%. Ou seja, foi possível comprovar que há uma relação bem próxima da linear entre estes dois custos.

### 3.1.1 Stepwise

Conforme apresentado na revisão da literatura, esta metodologia busca realizar uma escolha das variáveis a partir do seu poder de explicação da variável dependente. Ou seja, não há transformação da variável, nem redução dos coeficientes.

A metodologia foi realizada 4 vezes. Sendo que foi testada a utilização do volume recuperável total ou separado em óleo e gás. E em cada uma delas foi feita a tentativa de *forward* e *backward*.

Da mesma forma que ocorrerá no restante deste capítulo, a primeira análise será feita para o CAPEX.

Apenas em uma tentativa obteve-se redução das variáveis. Ao se utilizar o *stepwise backward* no modelo separado por óleo e gás, o  $V_{gás}$  foi excluído da regressão. O resultado foi o seguinte:

$$\widehat{CAPEX} = -1164,9 + 2,7 V_{\text{óleo}} + 0,4 Sot + 1,2 LDA + 4,2 area \quad (37)$$

$$P \text{ valor} = (0,3367) (0,0551) (0,1571) (0,0170) (0,0642)$$

$$R^2 = 0,4078$$

$$MSE = 2.753.388$$

O modelo apresentou uma melhoria em relação ao que utilizou todas as variáveis. Porém, esta metodologia não eliminou a variável *Sot* o que provocaria uma melhora no modelo. Além disso, como visto no tópico anterior, a exclusão do intercepto também produz uma melhor estimativa do CAPEX com base na amostra disponível.

Logo, o melhor modelo para a previsão do CAPEX ainda é o da equação 33 que possui apenas o  $V_{\text{óleo}}$ , *LDA* e a *area* como variáveis explicativas.

Fazendo a mesma análise para os modelos do OPEX, novamente apenas a metodologia *stepwise backward* realizou a seleção de variáveis.

$$\widehat{OPEX} = 1320,3 + 6,4 V_{\acute{o}leo} + 6,0 \text{ area} \quad (38)$$

$$P \text{ valor} = (0,0272) \quad (0,0001) \quad (0,0076)$$

$$R^2 = 0,5116$$

$$MSE = 2.816.740$$

O ponto positivo para esta metodologia é que ela retirou exatamente as variáveis que estavam com sinal negativo, ou seja, que estavam indo contra o que seria o esperado. Além disso, todas as variáveis obtiveram algum nível de significância.

Pela ótica do MSE, esta equação obteve o melhor resultado até agora, sendo a escolhida para explicar o OPEX dadas as metodologias já testadas nesta dissertação.

### 3.1.2 Ridge regression

Neste tópico será seguida a mesma ordem dos parágrafos anteriores. Porém, antes de iniciar as estimativas, é importante lembrar que no *ridge regression* também deve-se estimar o parâmetro  $\lambda$ . Para escolher este número, testa-se vários valores até encontrar aquele que minimiza o MSE.

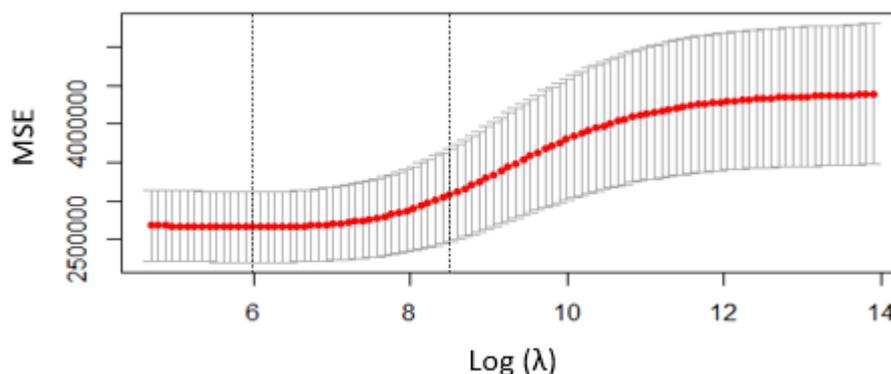
Também é essencial lembrar que, no limite, quanto este parâmetro for igual a zero, os coeficientes serão iguais aos do MQO, conforme encontrados na primeira parte deste capítulo. Por outro lado, quanto maior o  $\lambda$ , menores os coeficientes. Além disso, foi feita a regularização das variáveis preditoras, como mencionado na revisão da literatura.

Segundo Burn (2019), este é um dos modelos mais comumente utilizados na literatura de *machine learning*.

Conforme discutido na metodologia, o *ridge regression* provoca um encolhimento dos coeficientes de forma a tentar reduzir o MSE obtido com os coeficientes originais. Logo, será possível perceber uma redução destes valores em relação às estimativas anteriores. Além disso, este encolhimento também reduz a possibilidade de haver um *overfitting* no modelo. E como já comentado, o MSE será utilizado para comparar os modelos obtidos em cada um dos tópicos.

No primeiro modelo do CAPEX, que utiliza os volumes dos fluidos separadamente, o  $\lambda$  que chega ao MSE mínimo é igual a 401. Para o segundo modelo, a imagem abaixo ilustra como o MSE varia de acordo com o valor de  $\lambda$ , neste caso o a escolha deste parâmetro para chegar ao MSE mínimo é igual a 366 ( $e^{5,9}$ ).

Figura 5 - MSE em função de  $\lambda$  no ridge regression da equação 40



Fonte: O autor, 2020.

$$\widehat{CAPEX} = -571,1 + 2,4 V_{\text{óleo}} + 0,4 V_{\text{gás}} + 0,4 Sot + 1,0 LDA + 3,8 area \quad (39)$$

$$R^2 = 0,4497$$

$$MSE = 2.660.194$$

$$\lambda = 401$$

$$\widehat{CAPEX} = -455,6 + 1,7 V_{\text{total}} + 0,4 Sot + 1,1 LDA + 3,6 area \quad (40)$$

$$R^2 = 0,4418$$

$$MSE = 2.638.549$$

$$\lambda = 366$$

Como era de se esperar, houve uma redução nos coeficientes. Além disso, o *ridge regression* promoveu uma melhora na estatística do erro da previsão.

Ou seja, o *ridge regression* obteve uma equação melhor do que a escolha das variáveis pelo método *stepwise* e melhor do que a regressão linear que utilizou apenas as variáveis significativas. Assim, a equação 40 para previsão do CAPEX, obtida neste tópico, é a que obtém o melhor ajuste aos dados.

Foi testada a opção de realizar o *ridge regression* apenas com as variáveis com coeficientes significativos da regressão linear. Porém, não houve melhoria no termo de erro.

Agora, estimando as equações para o OPEX, chega-se nos seguintes resultados:

$$\widehat{OPEX} = 1905,7 + 6,3 V_{\text{óleo}} - 0,7 V_{\text{gás}} - 0,1 \text{Sot} - 0,1 \text{LDA} + 6,4 \text{area} \quad (41)$$

$$R^2 = 0,5200$$

$$MSE = 3.315.150$$

$$\lambda = 481$$

$$\widehat{OPEX} = 2386,3 + 3,9 V_{\text{total}} - 0,1 \text{Sot} - 0,1 \text{LDA} + 4,9 \text{area} \quad (42)$$

$$R^2 = 0,4949$$

$$MSE = 3.540.450$$

$$\lambda = 402$$

Os MSEs foram menores do que os obtidos na regressão linear com todos os regressores, porém foram maiores do que ao fazer a escolha das variáveis explicativas por meio do *stepwise backward*. Portanto, para o OPEX, o *ridge regression* não foi capaz de trazer uma melhoria.

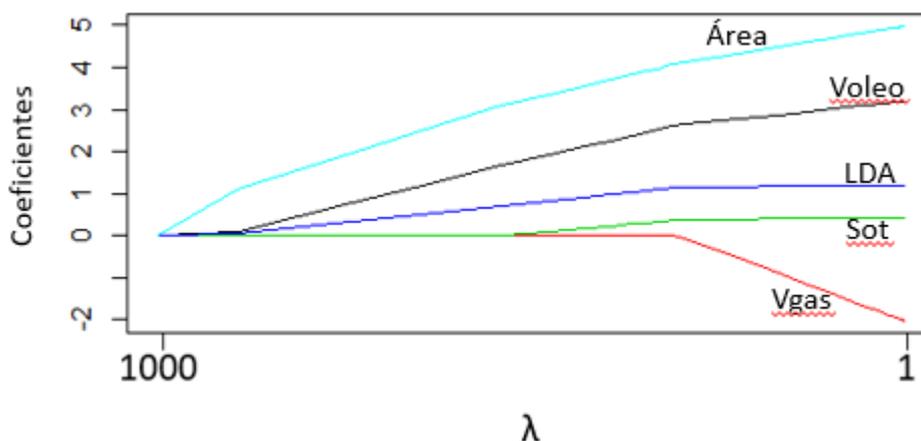
Assim como para o CAPEX, também não houve melhora ao realizar o ajuste apenas nas variáveis significativas.

### 3.1.3 Lasso

Na metodologia *lasso* também há uma redução nos coeficientes das variáveis independentes, porém, diferentemente do *ridge regression*, estes valores podem ser iguais a zero dependendo da escolha do valor de  $\lambda$ .

Iniciando novamente pelo primeiro modelo do CAPEX, o gráfico a seguir demonstra como os coeficientes variam de acordo com o valor de  $\lambda$ .

Figura 6 - Coeficientes em função de  $\lambda$  no lasso da equação 43.



Fonte: O autor, 2020.

Este gráfico é interessante para mostrar a ordem de entrada das variáveis no modelo. Para valores mais altos de  $\lambda$ , apenas a *area* entra como variável explicativa. Logo depois o  $V_{\acute{o}leo}$  e a *LDA* entram praticamente juntos. Após uma redução maior no parâmetro, o soterramento entra no modelo. E por último o  $V_{gás}$  entra para valores menores de  $\lambda$ . No limite, quando este valor se aproxima de zero, os coeficientes são iguais aos da regressão linear obtida no começo deste capítulo.

Os resultados para os dois modelos do CAPEX estão abaixo.

$$\widehat{CAPEX} = -692,6 + 2,5 V_{\acute{o}leo} + 0,3 Sot + 1,1 LDA + 4,0 area \quad (43)$$

$$R^2 = 0,4527$$

$$MSE = 2.637.826$$

$$\lambda = 67$$

$$\widehat{CAPEX} = -922,1 + 1,8 V_{total} + 0,4 Sot + 1,2 LDA + 3,7 area \quad (44)$$

$$R^2 = 0,4451$$

$$MSE = 2.662.459$$

$$\lambda = 6$$

É interessante perceber que para o primeiro modelo, o  $\lambda$  que minimiza o MSE retirou o  $V_{g\acute{a}s}$  do modelo ao atribuir o valor zero ao seu coeficiente. Foi a mesma variável excluída ao se utilizar a metodologia *stepwise backward*. E no segundo modelo nenhuma variável foi excluída.

Olhando para os valores de MSE obtidos e considerando todos os modelos estimados para o CAPEX até aqui, o primeiro modelo estimado para o *lasso* (equação 43) obteve o melhor MSE, representando uma leve melhoria em comparação ao *ridge regression*.

Também foi testada a hipótese de utilizar apenas as variáveis com coeficientes significativos da regressão linear, porém não houve melhora.

Para o OPEX, houve uma seleção maior das variáveis, conforme visto nos resultados abaixo.

$$\widehat{OPEX} = 1814,9 + 5,7 V_{\acute{o}leo} + 4,9 area \quad (45)$$

$$R^2 = 0,5209$$

$$MSE = 3.041.264$$

$$\lambda = 204$$

$$\widehat{OPEX} = 2324,4 + 3,9 V_{total} + 3,3 area \quad (46)$$

$$R^2 = 0,4820$$

$$MSE = 3.287.260$$

$$\lambda = 271$$

Como se pode observar, o *lasso* retirou a *LDA* e o *Sot* das duas regressões e retirou o  $V_{gás}$  da primeira. Ou seja, mais uma vez a escolha das variáveis foi igual ao do *stepwise backward*. A diferença entre elas é o encolhimento dos coeficientes das variáveis que ficaram no modelo.

Os MSEs destas duas regressões foram superiores aos obtidos nas metodologias anteriores, portanto não houve melhora para o modelo de previsão do OPEX. Também não foi obtida nenhuma evolução ao utilizar apenas as variáveis significativas da regressão linear. Dessa forma, o melhor modelo para previsão do OPEX continua sendo a equação 38.

Portanto, neste tópico foram estimadas 16 equações, todas elas supondo uma relação funcional linear para explicar o CAPEX e o OPEX. Dentre as diferenças das equações estão as variáveis utilizadas e a metodologia de ajuste. Foram testadas duas metodologias para escolha das variáveis independentes e duas metodologias que realizavam o encolhimento dos coeficientes.

A escolha da melhor equação foi feita com base no MSE. A tabela 6 mostra o resumo das equações encontradas nesta parte da dissertação e destaca o melhor modelo para explicação de cada um dos custos.

Tabela 6 - Resumo dos modelos lineares.

	Equação	Tópico	R <sup>2</sup>	MSE	Variáveis	Volume
CAPEX	31	3.1 Regressão Linear	39,83%	2.783.681	Todas	Vóleo + Vgás
	32	3.1 Regressão Linear	39,69%	2.877.697	Todas	Vtotal
	33	3.1 Regressão Linear	88,25%	2.705.727	Apenas significativas	Vóleo
	37	3.1.1 Stepwise	40,78%	2.753.388	Selecionadas	Vóleo
	39	3.1.2 Ridge Regression	44,97%	2.660.194	Todas	Vóleo + Vgás
	40	3.1.2 Ridge Regression	44,18%	2.638.549	Todas	Vtotal
	43	3.1.3 Lasso	45,27%	2.637.826	Selecionadas	Vóleo
	44	3.1.3 Lasso	44,51%	2.662.459	Selecionadas	Vtotal
OPEX	34	3.1 Regressão Linear	48,76%	3.324.452	Todas	Vóleo + Vgás
	35	3.1 Regressão Linear	45,92%	3.662.573	Todas	Vtotal
	36	3.1 Regressão Linear	90,59%	2.981.588	Apenas significativas	Vóleo
	38	3.1.1 Stepwise	51,16%	2.816.740	Selecionadas	Vóleo
	41	3.1.2 Ridge Regression	52,00%	3.315.150	Todas	Vóleo + Vgás
	42	3.1.2 Ridge Regression	49,49%	3.540.450	Todas	Vtotal
	45	3.1.3 Lasso	52,09%	3.041.264	Selecionadas	Vóleo + Vgás
	46	3.1.3 Lasso	48,20%	3.287.260	Selecionadas	Vtotal

Fonte: O autor, 2020.

Assim, acredita-se que foram praticamente esgotadas as chances de obter um MSE menor dos que foram encontrados, considerando uma relação funcional linear. Porém, um dos objetivos desta dissertação é realizar a estimação com modelos que se aproximam do *machine learning* e que tem um grau de flexibilidade maior em sua forma funcional. O tópico seguinte tratará de tentativas de obter ajustes com MSE menores dos que os já encontrados, utilizando as mesmas variáveis independentes.

### 3.2 Modelos flexíveis

Os modelos obtidos a partir de regressão linear trazem consigo vários benefícios, como a facilidade de interpretação, a possibilidade de fazer inferências e a simplicidade de uma relação que é facilmente entendida.

Nesta dissertação o ponto de partida foi esta forma funcional e a partir dela foram testadas diversas metodologias para melhorar o poder de explicação do modelo linear. Apesar de conseguir alguma melhoria, a relação funcional entre as variáveis continuará sendo linear.

Portanto, em busca de uma melhoria mais significativa, esta parte do Capítulo dedicará-se à utilização de ferramentas mais flexíveis e que se aproximam ainda mais da disciplina de *machine learning* para obter modelos que expliquem as variações no CAPEX e no OPEX de projetos de produção de petróleo.

Serão testados dois tipos de função para realizar o ajuste de forma satisfatória: *smoothing splines* e regressão local. Logo, esta parte da dissertação está dividida em dois tópicos.

Os modelos deste capítulo permitem uma flexibilidade bem maior em comparação com a regressão linear. Além disso, tem uma capacidade de inferência melhor do que metodologias mais avançadas de *machine learning*. Porém, têm um nível de complexidade de definição de parâmetros maior do que a regressão linear.

Por serem duas técnicas flexíveis que exigem a definição de um grau de suavização, serão feitas duas tentativas de ajuste. Uma de forma mais automática, com a escolha do parâmetro que minimiza o MSE e outra de forma mais empírica, observando o gráfico do ajuste obtido e verificando sua validade.

### 3.2.1 Smoothing splines

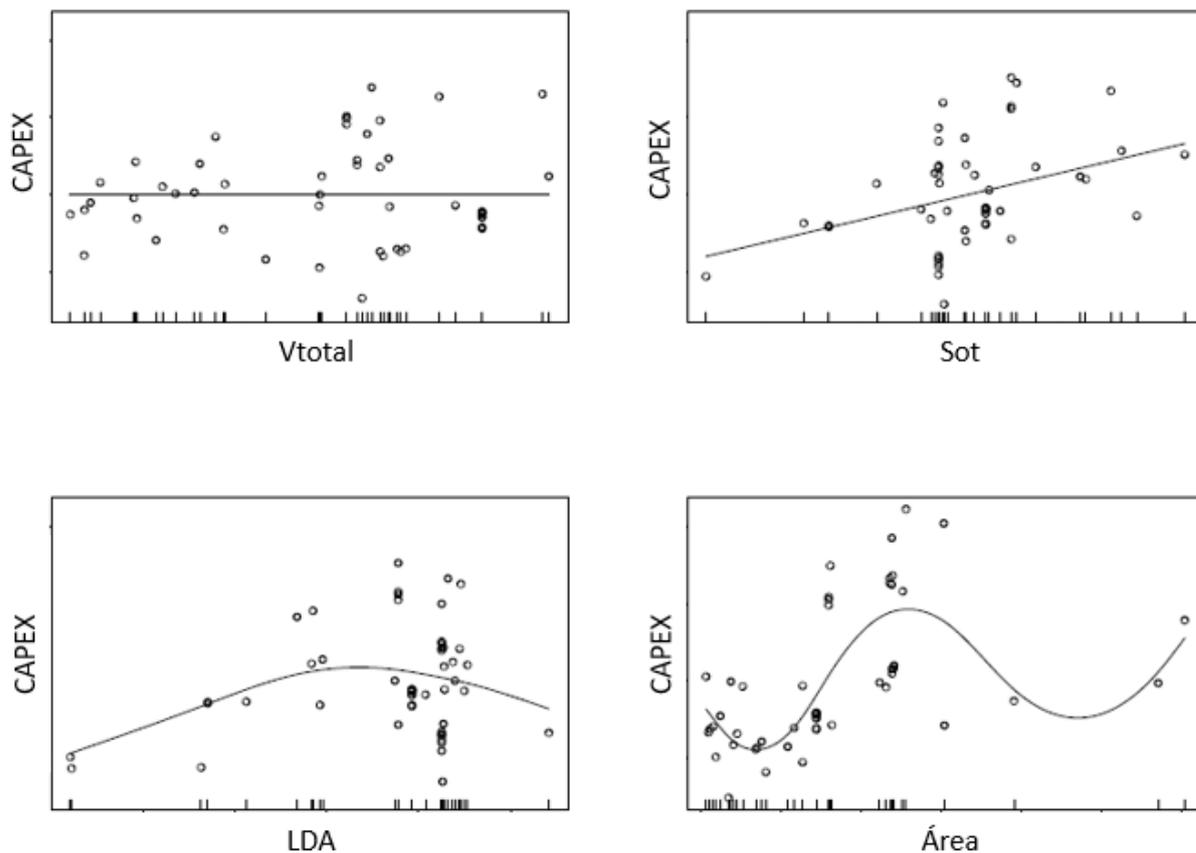
Como descrito na revisão da literatura, este método flexível tem a vantagem de considerar nós em todos os pontos da curva, ao contrário das *natural splines* nas quais é necessário definir quantos existirão e onde eles se localizarão. Dessa forma, as *smoothing splines* possuem um parâmetro a menos para ser escolhido no momento de caracterizar o modelo. Assim, passa-se a preocupação para o mais importante: o valor de  $\lambda$ , parâmetro que define quão flexível será a curva obtida e que está intimamente ligado ao conceito de *overfitting*.

Como tratado na revisão da literatura, valores altos aproximam-se da regressão linear e deixam o modelo menos curvilíneo. Já valores mais baixos de  $\lambda$  tendem a se ajustar melhor aos dados e, no limite, realizam praticamente uma interpolação entre as observações. Portanto, o perigo de se utilizar um  $\lambda$  muito baixo é cair no problema de *overfitting*.

O R oferece uma ferramenta, baseada na minimização do MSE utilizando *cross-validation*, para auxiliar na definição do lambda (Yuan, 2006). Porém, sempre com a ajuda de análises gráficas, será ponderada a possibilidade de utilizar valores diferentes a fim de evitar o ajuste muito próximo dos dados, o que poderia causar problemas na previsão de valores fora do conjunto de dados utilizados na estimação.

Começando na busca pela explicação do CAPEX, as primeiras tentativas de ajuste buscaram utilizar o máximo de variáveis possíveis. Na Figura 7, está o ajuste com todas as variáveis disponíveis, e pode ser visto o efeito parcial de cada uma delas no CAPEX. É possível perceber que a maioria das curvas ficaram diferentes do que se esperaria. O CAPEX praticamente não apresentou variação com o aumento do  $V_{total}$ , para valores altos de lâmina d'água essa relação ficou negativa, e para a área esta relação ficou com grandes variações e mudanças de inclinação, o que não condiz com a teoria de avaliação econômica de projetos de petróleo.

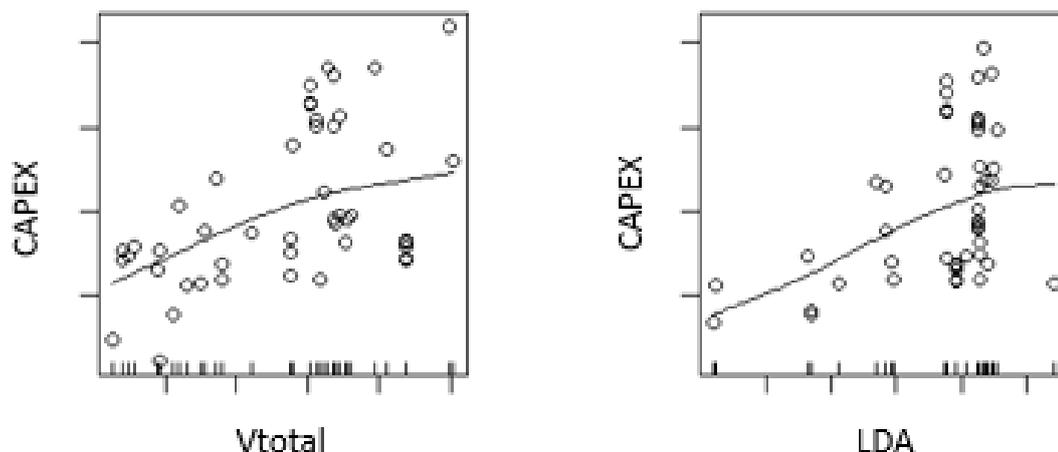
Figura 7 - Efeitos parciais das variáveis explicativas no CAPEX usando smoothing splines



Fonte: O autor, 2020.

A partir dos resultados acima, foram feitas diversas tentativas utilizando diferentes variáveis e valores de  $\lambda$  para obter uma relação que respeitasse a lógica do problema aqui tratado e que tivesse significância estatística. Ou seja, que as variáveis tivessem um comportamento crescente, ou no máximo constante e que o excesso de flexibilidade do modelo não gerasse curvas com muitas variações de direção. O melhor ajuste foi encontrado ao utilizar apenas as variáveis significativas obtidas na primeira parte deste capítulo, ou seja,  $V_{total}$  e  $LDA$ . Além disso, o valor de  $\lambda$  deste ajuste foi igual a 4. A visualização dos efeitos parciais destas duas variáveis está no gráfico a seguir.

Figura 8 - Efeitos parciais do  $V_{total}$  e da  $LDA$  no CAPEX no melhor ajuste utilizando *smoothing splines*



Fonte: O autor, 2020.

Assim, visualmente se observa uma relação com diferentes intensidades de crescimento entre estas duas características e o CAPEX. Para valores mais baixos, a inclinação da curva é maior e depois ela tende a se estabilizar e ter um acréscimo bem menor. Isso mostra que, possivelmente, para acumulações com grande volume recuperável e LDA mais profundas, as diferenças de CAPEX para desenvolver a produção são pequenas, porém, para volumes menores e LDA mais rasas, uma pequena variação nestas características pode provocar um impacto maior no custo de investimento.

No quadro abaixo estão apresentadas as estatísticas deste ajuste, mostrando uma significância nas variáveis suavizadas e no intercepto.

$$\widehat{CAPEX} = \text{Intercepto} + f(V_{total}) + f(LDA) \quad (47)$$

$$P \text{ valor} = (0,0001) \quad (0,0122) \quad (0,0378)$$

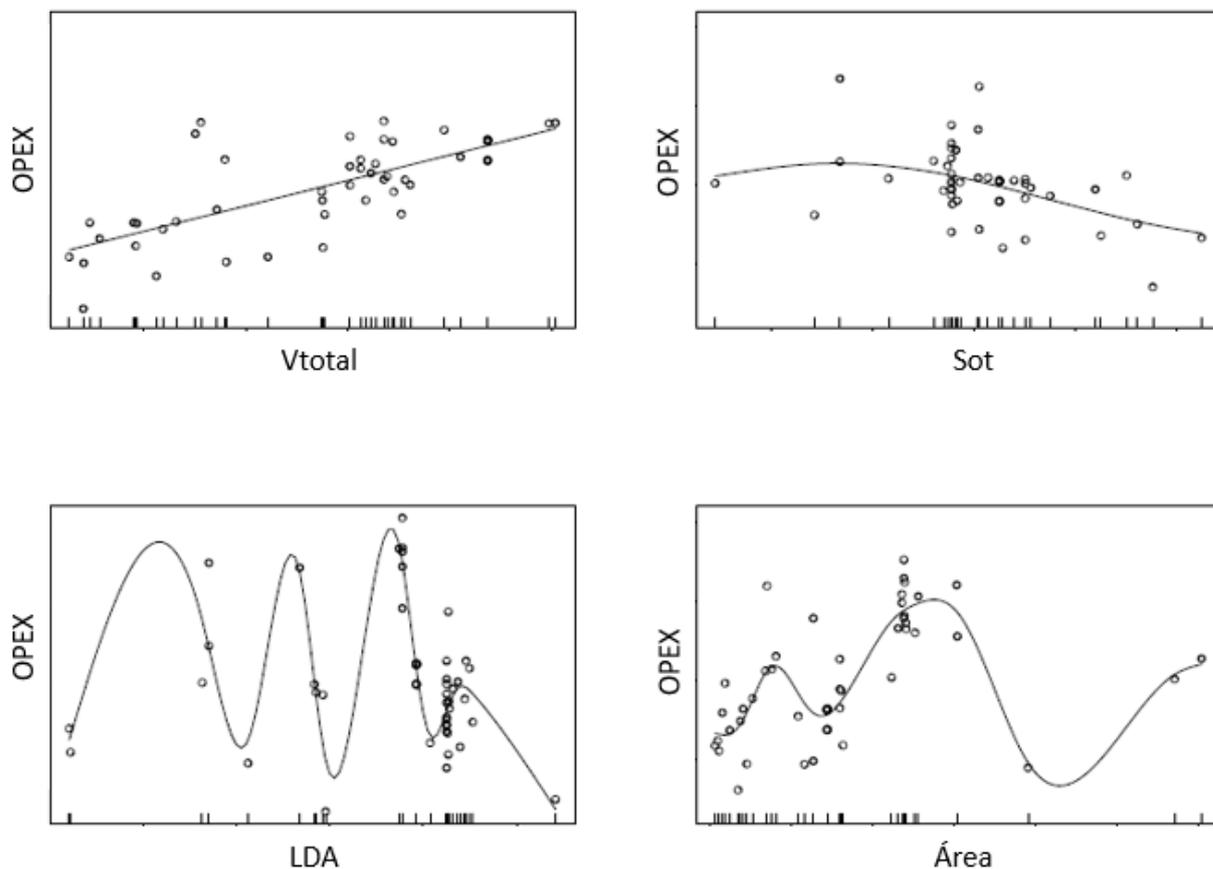
$$R^2 = 0,3900$$

$$MSE = 2.743.168$$

Pelos valores de  $R^2$  e MSE, este ajuste não conseguiu ser melhor do que os obtidos anteriormente por meio do *lasso*. Ou seja, a relação linear com regularização dos coeficientes e seleção das variáveis mostrou um melhor resultado. Logo, este modelo flexível para o CAPEX não será utilizado neste trabalho.

Foi feita a mesma análise para o OPEX. Iniciou-se com a tentativa de utilização de todas as variáveis disponíveis. Porém, conforme a Figura 9, também não foi possível obter um ajuste que respeitasse as expectativas deste trabalho.

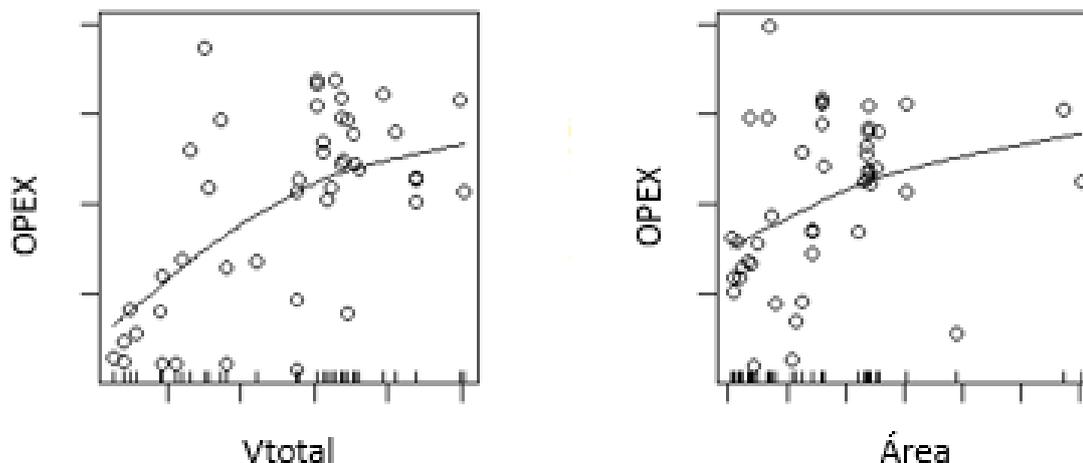
Figura 9 - Efeitos parciais das variáveis explicativas no OPEX usando smoothing splines.



Fonte: O autor, 2020.

Logo, novamente foram testadas diversas combinações de variáveis e de parâmetros de suavização de forma a obter um ajuste aceitável e que tivesse significância estatística. Este resultado está exposto na Figura 10.

Figura 10 - Efeitos parciais do  $V_{total}$  e da  $area$  no OPEX no melhor ajuste utilizando *smoothing splines*.



Fonte: O autor, 2020.

Da mesma forma que foi visto para os custos de investimento, o volume recuperável apresenta uma relação com os custos operacionais que possui maior inclinação nos menores volumes e depois tende a uma estabilização. Para a variável *area* é quase possível ver dois segmentos de reta distintos. O parâmetro de suavização utilizado para obter estas curvas foi 2,5.

$$\widehat{OPEX} = \text{Intercepto} + f(V_{total}) + f(\text{area}) \quad (48)$$

$$P \text{ valor} = (0,0001) \quad (0,0009) \quad (0,0919)$$

$$R^2 = 0,5410$$

$$MSE = 2.770.865$$

As duas variáveis foram significativas, sendo que a *area* foi apenas para um nível de 10%. Como se pode ver, o MSE foi o menor obtido até este momento da dissertação. Portanto o modelo flexível utilizando *splines* se mostrou melhor do que os modelos lineares para explicar a relação do OPEX com as variáveis disponíveis na amostra.

### 3.2.2 Regressão local

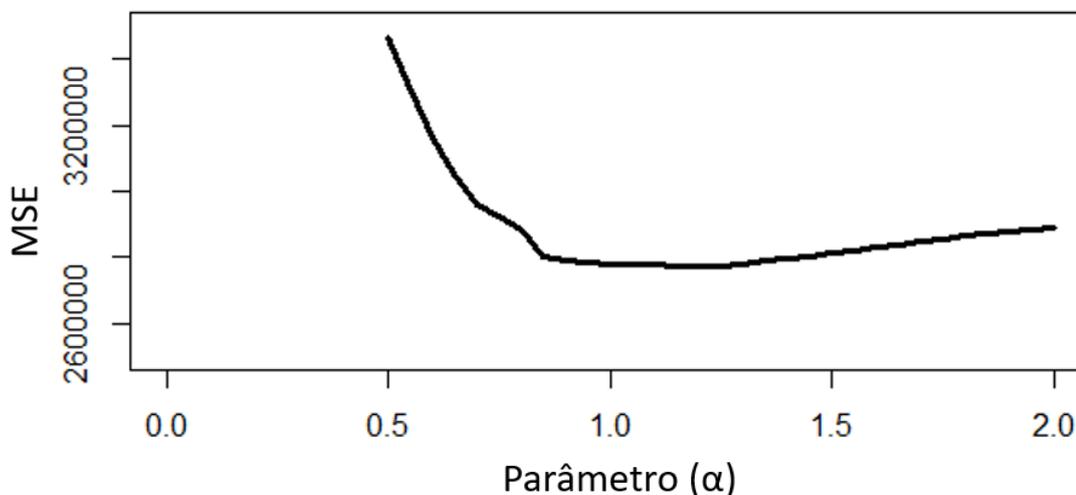
Este é outro método no qual é preciso definir um parâmetro de suavização da curva, chamado de  $\alpha$ . Quando este valor é menor do que 1, representa a proporção dos dados que são utilizados para realizar a regressão em cada ponto. Portanto, quanto menor o valor de  $\alpha$ , mais curvilíneo será o resultado. Ou seja, se  $\alpha = 0,6$ , para determinado ponto  $x$  serão usados 60% dos dados, considerando os mais próximos daquele ponto para realizar o ajuste. Além disso, o modelo fornece pesos maiores para as observações mais próximas. No caso de  $\alpha$  ser maior do que 1, todos os dados são utilizados para a regressão, mas os pesos deles se alteram, de forma a deixar a curva mais próxima de uma reta, conforme  $\alpha$  aumenta (Chambers e Hastie, 1991).

Segundo Catherine (2004) este tipo de método fica menos útil quando se utiliza mais do que 2 ou 3 variáveis independentes. Os principais motivos são que para níveis maiores de dimensão seriam necessários mais dados para obter um ajuste representativo (problema conhecido como a maldição da dimensionalidade) e problemas de visualização para verificar como a variável explicada varia em função das outras e se esta variação está de acordo com o que é esperado. Portanto, neste tópico serão estimadas regressões com no máximo duas variáveis explicativas.

Como feito para a definição dos parâmetros do *ridge regression* e do *lasso*, o *cross-validation* será utilizado como ferramenta. A título de exemplificação, a Figura 11 mostra como o MSE varia de acordo com a escolha do parâmetro no caso da regressão do CAPEX em função do  $V_{\text{óleo}}$ . É possível perceber que há um grande ganho ao aumentar o valor até 0,9 e depois há uma grande perda ao aumentar a partir de 1,3. Valores dentro deste intervalo resultam em um erro razoavelmente parecido. Portanto, por meio desta análise, o parâmetro que minimiza o MSE é igual a 1,2.

Porém, assim como no tópico anterior, o ajuste pode ficar muito flexível e distante do resultado esperado a partir do conhecimento da indústria. Logo, poderão ser feitos ajustes manuais no parâmetro de suavização para obter uma curva com melhor representatividade.

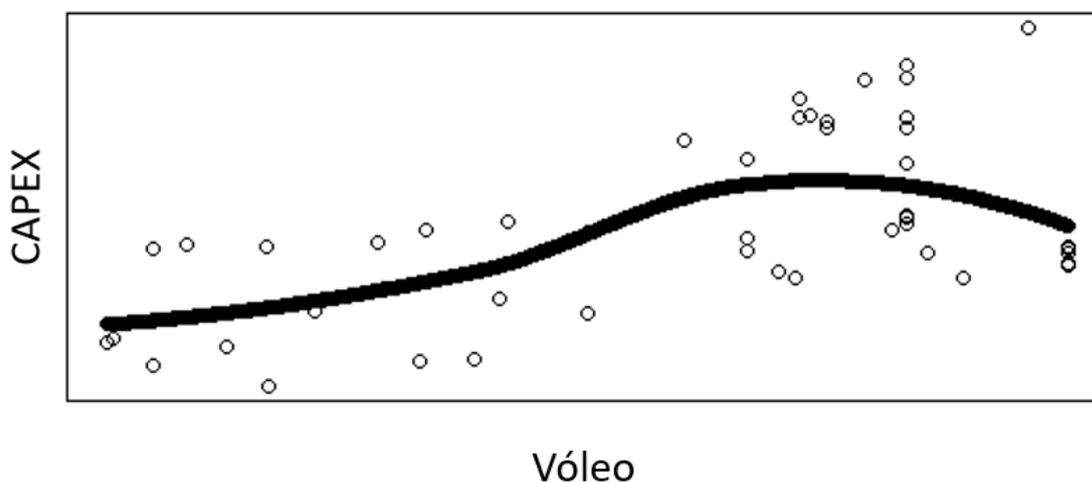
Figura 11 - Variação do MSE em função do parâmetro de suavização ( $\alpha$ ) da regressão local ao tentar explicar o CAPEX em função do  $V_{\text{óleo}}$ .



Fonte: O autor, 2020.

Portanto, a utilização de um  $\alpha$  igual a 1,2 leva ao resultado da Figura 12.

Figura 12 - Regressão local do CAPEX em função do  $V_{\text{óleo}}$  utilizando  $\alpha = 1,2$ .



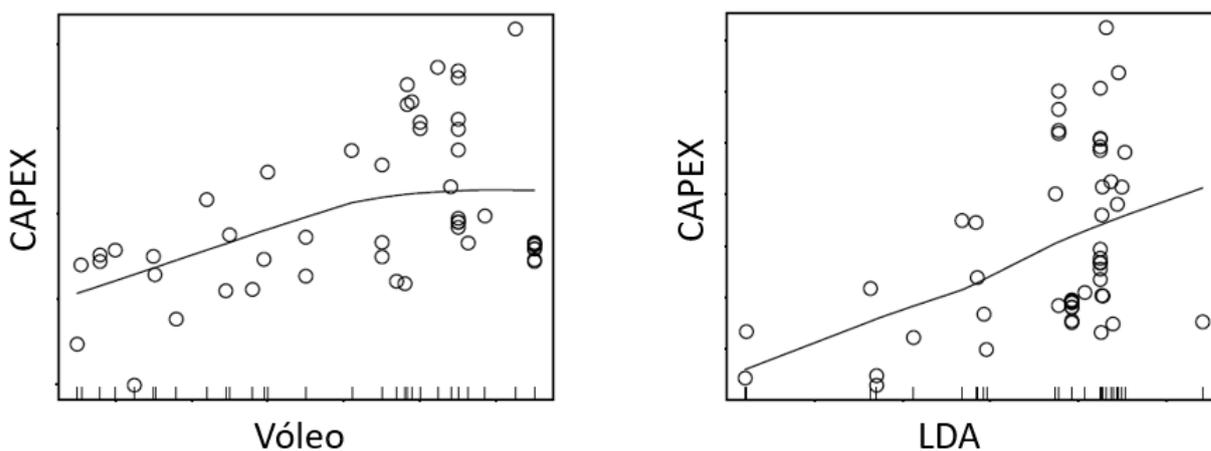
Fonte: O autor, 2020.

É possível ver que a curva acompanha de forma mais próxima as variações nos dados, com diferentes inclinações de acordo com o intervalo da curva. Porém, um ponto de atenção está nos volumes maiores. A partir de certo volume a curva fica descendente, o que não está de acordo com o que se espera. Logo, este ajuste foi considerado inadequado.

Assim, foram testados diversos valores para o parâmetro de suavização para as diferentes variáveis. Também foram testadas algumas regressões utilizando duas variáveis independentes.

Com isso, a regressão local que obteve o melhor ajuste foi a que utilizou o  $V_{\acute{o}leo}$  e a  $LDA$  para explicar o CAPEX utilizando o parâmetro igual a 1. Abaixo é possível ver o resultado deste ajuste.

Figura 13 - Efeitos parciais da regressão local do CAPEX em função do  $V_{\acute{o}leo}$  e da  $LDA$  utilizando  $\alpha = 1$ .



Fonte: O autor, 2020.

$$\widehat{CAPEX} = f(V_{\acute{o}leo}) + f(LDA) \quad (49)$$

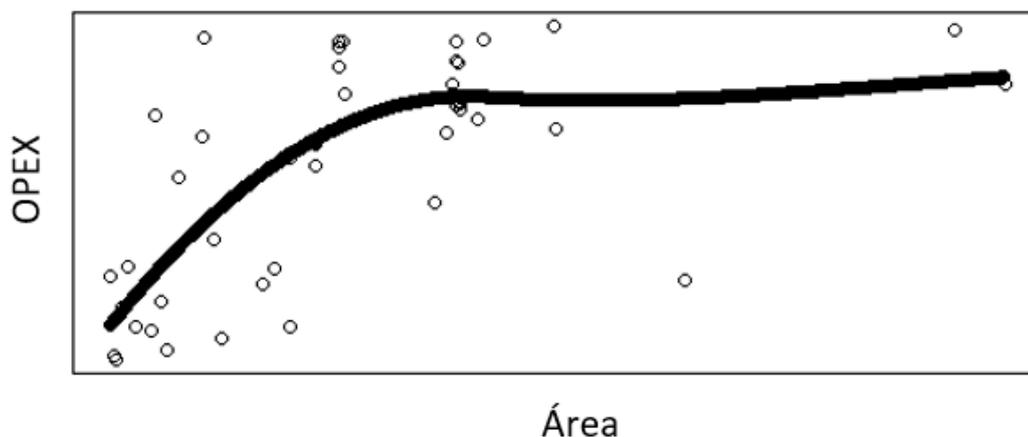
$$R^2 = 0,3895$$

$$MSE = 2.685.889$$

Nesta regressão o MSE ficou acima de ajustes feitos anteriormente nesta dissertação. Portanto, ela não é a mais adequada para explicar o CAPEX.

Para fazer o ajuste para o OPEX foram seguidos os mesmos passos. O melhor ajuste utilizou a *area* como variável independente e está ilustrado na Figura 14.

Figura 14 - Regressão local do OPEX em função da *area* utilizando  $\alpha = 1$



Fonte: O autor, 2020.

$$\widehat{OPEX} = f(\text{area}) \quad (50)$$

$$R^2 = 0,5248$$

$$MSE = 3.116.401$$

A flexibilidade deste ajuste parece adequada e o grau de inclinação da curva em seus diversos pontos parece estar acompanhando os dados, com destaque para a alta variação em projetos com áreas menores e depois uma tendência de estabilização nos custos operacionais para áreas maiores. O parâmetro de suavização utilizado foi igual a 1. Portanto, apesar de ter ficado mais bem ajustado do que se fosse utilizada uma relação linear, esta não foi a melhor regressão obtida nesta dissertação até aqui para explicar o OPEX.

A tabela 7 mostra um resumo dos melhores resultados obtidos por meio dos modelos flexíveis. É interessante perceber que os parâmetros de custo obtiveram o melhor ajuste com metodologias diferentes. O CAPEX foi mais bem previsto ao utilizar a regressão local com o  $V_{\text{óleo}}$  e a  $LDA$  como regressores, e o OPEX demonstrou melhor resultado ao ser previsto usando *smoothing splines* com o  $V_{\text{total}}$  e a *area* como variáveis explicativas.

Tabela 7 - Resumo dos modelos flexíveis

	Equação	Tópico	R <sup>2</sup>	MSE	Variáveis	Volume
CAPEX	47	3.2.1 Smoothing Splines	39,00%	2.743.168	Apenas significativas	Vtotal
	49	3.2.2 Regressão Local	38,95%	2.685.889	Vóleo e LDA	Vóleo
OPEX	48	3.2.1 Smoothing Splines	54,10%	2.770.865	Apenas significativas	Vtotal
	50	3.2.2 Regressão Local	52,48%	3.116.401	Área	-

Fonte: O autor, 2020.

Comparando a tabela 6 com a tabela 7, é possível ver que o menor MSE para o CAPEX foi obtido na equação 43, que utilizou um modelo linear (*lasso*) com  $V_{\text{óleo}}$ ,  $Sot$ ,  $LDA$  e  $area$  como variáveis explicativas. Apesar da regressão local também ter obtido um MSE parecido, o R<sup>2</sup> do *Lasso* foi o segundo melhor de todas as equações estimadas.

Ao contrário do CAPEX, o OPEX obteve o melhor resultado com um modelo flexível. A equação 48 da *smoothing splines* obteve o menor MSE de todas as equações estimadas para o OPEX. E apresentou o segundo melhor R<sup>2</sup> de todos os ajustes.

Independentemente do nível de flexibilidade, é possível observar pelos resultados que os modelos que utilizaram técnicas mais próximas do *machine learning* obtiveram melhores resultados. Isso mostra a importância do esforço de sair da regressão linear simples para técnicas mais avançadas de modelagem. Essa transição permite que haja uma capacidade de previsão e um resultado mais próximo do real, o que é de vital importância em um setor como de exploração de petróleo.

Portanto, para o cálculo do fluxo de caixa do projeto, o CAPEX seguirá um modelo linear obtido através do *lasso* e o OPEX seguirá um modelo flexível obtido por meio de *smoothing splines*.

### 3.3 Simulação preço do petróleo

Para a simulação dos preços do petróleo, foi utilizado o modelo de dois fatores de Schwartz e Smith (2000). Conforme detalhado na revisão da literatura, o primeiro fator está

relacionado às oscilações de curto prazo, como nível dos estoques, e o segundo fator está relacionado a efeitos de longo prazo, como desenvolvimento de tecnologias, novas descobertas e surgimento de energias alternativas.

Os parâmetros de volatilidade do curto e do longo prazo e a velocidade de reversão à média utilizados foram os obtidos por Souza (2019), todos significativos ao nível de 1%. Os valores iniciais dos fatores deste modelo foram ajustados para refletir o preço inicial de 45 US\$/Barril ( $X_{1_0} = 0,5$  e  $X_{2_0} = 3,31$ ). O trabalho de Souza (2019) não obteve uma tendência de longo prazo com significância estatística, portanto, pela atual realidade do mercado ela foi definida em um valor acima do obtido no trabalho citado. A correlação entre os componentes também foi retirada de Souza (2019) sendo significativa ao nível de 1%.

O resultado da incerteza nos preços do petróleo está apresentado no Capítulo 4, mais especificamente na Figura 21 e na tabela 12.

$$\ln St = X_{1t} + X_{2t} \quad (51)$$

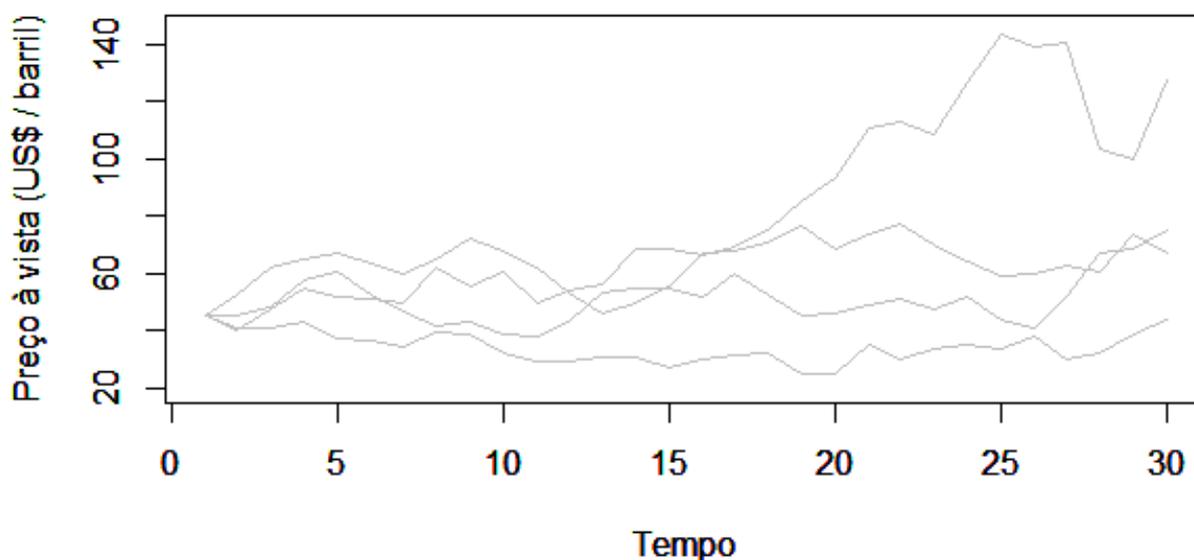
$$dX_{1t} = -kX_{1t}dt + \sigma_{X1}dB_{X1,t} \quad (52)$$

$$dX_{2t} = \mu dt + \sigma_{X2}dB_{X2,t} , \quad (53)$$

Onde,  $k = 0,84842$ ,  $\sigma_{X1} = 0,36258$ ,  $\mu = 0,6$ ,  $\sigma_{X2} = 0,32625$ . Além disso, as variáveis são correlacionadas por  $\rho_{X1X2} = 0,74247$ .

Assim, para cada volume simulado haverá uma simulação de preço do petróleo, totalizando 10000 simulações. A título de exemplificação, foram feitas 4 simulações de preços futuros cujas projeções estão apresentadas na Figura 15.

Figura 15 - Quatro projeções exemplificativas para o preço do petróleo



Fonte: O autor, 2020.

Conforme Schiozer, Lima e Suslick (2008), também é importante considerar a relação entre os preços do petróleo e os custos. Existem evidências, como demonstrado em Hastenreiter et al (2012) e Bradley e Wood (1994) que mostram que essas duas variáveis são correlacionadas. Logo, a não consideração desta relação pode levar a uma volatilidade superdimensionada do fluxo de caixa e a uma tomada de decisão equivocada. Ou seja, em cenários de preços altos os custos estariam subestimados e em cenários de preços baixos estariam superestimados, levando a VPLs mais extremos, o que não condiz com a realidade.

Na base de dados utilizada, o CAPEX não é sensível ao preço do óleo, portanto o CAPEX será independente do preço, conforme capítulo anterior. Apesar disso, é possível que outras bases de dados apresentem uma sensibilidade entre essas duas variáveis, dado que o custo de sondas de perfuração pode ter alguma variação de acordo com a oscilação do preço do petróleo, por exemplo, conforme encontrado por Toews e Naumov (2015).

Já para o OPEX foi identificado na base de dados disponível que existe a possibilidade de se encontrar uma relação. Com base nisso, todo o trabalho de estimação do OPEX foi refeito, considerando o preço do petróleo como uma variável independente. Porém, em todas as estimativas os resultados foram inferiores aos do capítulo anterior. O melhor resultado utilizando o preço do petróleo como variável explicativa foi o de *smoothing splines* em conjunto com o volume total. Esta estimativa obteve um  $R^2$  de 47,60% e um MSE de 3.392.839.

Será utilizada a metodologia baseada no fenômeno *price-effect-cost-escalation* (PECE) introduzido por Nystad (1981). Essa metodologia trata da estimação do percentual que os custos

variam em função da variação dos preços. Outros trabalhos utilizaram esta metodologia, como Schiozer, Lima e Suslick (2008) que afirmaram que a relação de causalidade é dos custos em função do preço e Macary, Razek e Goary (2000) que também defendem a importância do PECE para previsões de longo prazo.

Além deles, o trabalho de Hastenreiter et al (2012), utilizou esta mesma metodologia com base em dados de custos operacionais e preços entre 1994 e 2009 e encontrou um percentual de 46,6% nos campos offshore com  $R^2$  de 83,6%. Este valor será utilizado como referência nesta dissertação, ou seja, uma variação de 10% no preço leva a uma variação de 4,66% no OPEX.

Dentre os trabalhos anteriores, apenas Schiozer, Lima e Suslick (2008) falam em possibilidade de defasagem de tempo entre a alteração do preço e o reflexo nos custos operacionais. Apesar disso, a maior correlação encontrada por eles foi contemporânea, ou seja, sem nenhuma defasagem. Este trabalho adotará essa premissa.

### **3.4 Simulação curva de produção**

Para a curva de produção, será utilizado o modelo exponencial, conforme indicado pela SPEE (2002) e detalhado na revisão da literatura. As informações necessárias para a estimação da curva de produção são: Volume de óleo recuperável, vazão inicial ( $q_i$ ) do poço e produção total de cada poço (NP).

O processo se inicia pelas informações do poço tipo. Geralmente, o poço tipo de um projeto de produção é baseado em poços análogos de campos em produção, que possuem semelhanças com relação à produtividade do reservatório e características como porosidade, permeabilidade, viscosidade do fluido e teor de  $CO_2$ , por exemplo.

Para esta dissertação, conforme será visto no Capítulo 4, as informações de vazão inicial e produção total do poço já estarão definidas para o projeto. A partir destas duas informações,

é possível obter o parâmetro de declínio exponencial  $D_i$  que estima a vazão do poço para cada ano, conforme equação 2.

A partir do NP por poço e do volume recuperável, é possível definir a quantidade de poços necessários para explorar o petróleo.

E com relação à entrada em produção dos poços, por ser uma análise que deve se adequar a vários volumes e quantidades de poços, será feita uma simplificação colocando 1/6 dos poços em produção no primeiro ano, mais 1/2 no segundo ano e a partir do terceiro ano todos os poços entram em produção e atinge-se o pico da curva de produção. Isso garante um *ramp-up* adequado da produção do campo.

Portanto, é possível obter resultados próximos a realidade utilizando esses parâmetros e aproximações, dado que o volume recuperável será uma variável incerta e sujeita a uma ampla gama de valores.

### **3.5 Premissas gerais**

Além da estimação dos componentes tratados nos tópicos anteriores, um item essencial para o cálculo do valor de um projeto de produção de petróleo é a participação do governo neste projeto. Ou seja, é preciso estimar quanto que o governo vai receber de impostos e tributações. Para isso, serão feitas estimativas para dois regimes fiscais: Concessão e partilha de produção.

As principais características de tributação dos dois regimes já foram apresentadas na revisão da literatura, mas é importante destacar que, conforme Soares (2017), ambos os regimes podem ser equivalentes dependendo dos valores de bônus, excedente em óleo, participação especial e royalties definidas pelo regulador.

Partindo para as premissas, para o regime de concessão, como este trabalho não está fazendo avaliação para participação em leilões, o bônus de assinatura será considerado como

custo afundado. Além disso, por ter um valor irrisório, a taxa de retenção de área será desconsiderada da análise. Portanto, a participação governamental será representada pelo IR, CSLL, *royalties* e participação especial.

Como em todos os trabalhos citados, não será considerada a alíquota de 1% da receita bruta de pesquisa e desenvolvimento (P&D). Conforme contrato da 16ª rodada de concessão de blocos exploratórios, esta alíquota é devida apenas nos anos em que há pagamento de participação especial. Além disso, até 40% podem ser investimentos em pesquisas da própria empresa e os outros 60% devem ser em universidades e outras empresas brasileiras. Podendo haver compensação dos valores gastos anteriormente. Logo, considerando que a empresa já realiza esses investimentos em P&D normalmente, ela irá apenas compensar gastos já realizados, sem afetar o fluxo de caixa do projeto em si.

Dessa forma, o IR e o CSLL corresponderão a 34% do lucro tributável, os *royalties* serão 10% da receita bruta e a participação especial seguirá a tabela 8, conforme Resolução nº 12 de 2014 da ANP. Esta tabela se refere ao pagamento deste tributo em campos com lâmina d'água acima de 400 metros a partir do quarto ano de produção. Como simplificação, será utilizada a mesma tabela para o primeiro, segundo e terceiro ano. Importante destacar que essa simplificação tende a reduzir o VPL do projeto. As tabelas para os três primeiros anos de produção têm uma faixa de isenção maior e uma parcela de dedução também maior. Logo, esta dissertação está sendo conservadora no cálculo do VPL do projeto e assumindo uma participação especial ligeiramente maior do que a real.

Tabela 8 - Cálculo de participação especial.

Volume de Produção Fiscalizada no Período-Base (em milhares de metros cúbicos de petróleo equivalente)	Parcela a Deduzir da Receita Líquida no Período-Base (em reais)	Alíquota (em %)
Até 450	-	isento
Acima de 450 até 900	$450 \times RLP \div VPF$	10
Acima de 900 até 1350	$675 \times RLP \div VPF$	20
Acima de 1350 até 1.800	$900 \times RLP \div VPF$	30
Acima de 1.800 até 2.250	$360 \div 0,35 \times RLP \div VPF$	35
Acima de 2.250	$1.181,25 \times RLP \div VPF$	40

Fonte: Resolução nº 12 de 2014 da ANP.

Onde, RLP é a receita líquida da produção e VPF é o volume de produção fiscalizada em milhares de metros cúbicos.

Há a possibilidade do preço de referência da ANP para cálculo da RLP e consequentemente da participação ser ligeiramente inferior ao preço real praticado no mercado, o que beneficia o fluxo de caixa da empresa. O cálculo desse preço de tributação envolve a média mensal do petróleo Brent e um diferencial entre a qualidade do petróleo que se quer definir o preço e a qualidade do petróleo de referência (Brent). Portanto, cada campo possui um valor diferente de preço de venda. Poucos estudos abordam esta diferença de preços devido à dificuldade de se obter os dados reais do preço de venda. Uma boa base de dados, porém paga, é a da empresa PLATT's. Um estudo de Szklo, Machado e Morais (2007) utilizaram essa base de dados para mostrar essa diferença dos preços no campo de Marlim, que pode chegar a até 15%. Entretanto, pela falta de uma maior quantidade de dados, neste trabalho será adotada uma postura conservadora na qual o preço de referência para tributação será 5% inferior ao de venda. No caso de haver acesso a esta base de dados ou a outra que forneça o desconto do preço de referência de tributação para o preço real, é preferível utilizar este dado para calcular os royalties e a participação especial devido à sua maior exatidão.

Da mesma forma que Marques (2015), será usado um período de 10 anos para o cálculo de depreciação linear dos custos de investimentos. Assim, a partir do ano em que o houve o investimento do CAPEX, terão 10 anos de depreciação deste valor, com 10% em cada ano. Porém, diferentemente deste artigo, esta dissertação irá ajustar as parcelas da depreciação para

valores reais deflacionando no tempo de acordo com as projeções do Índice Geral de Preços – Disponibilidade Interna (IGP-DI) que constam no Sistema de Expectativas de Mercado do Banco Central do Brasil no dia 01/07/2020, que resultam em um valor médio anual de 3,93% ao ano.

Para o regime de partilha, as premissas serão as mesmas. Considerando a alteração do valor da alíquota de pagamento de royalties para 15% e a substituição da participação especial pelo excedente em óleo. Para definição do valor a ser utilizado neste parâmetro, foi feita a média dos valores vencedores das 6 rodadas de partilha de produção já ocorridas no Brasil. A média das ofertas vencedoras foi composta pelos 15 blocos arrematados desde 2013 e foi igual a 47,69%. Por simplificação, não será aplicada a tabela de variação desse excedente em função da vazão média do poço e do preço do Brent.

Da mesma forma que Marques (2015), a base para tributação do excedente em óleo é obtida a partir da seguinte equação:

$$BaseEO = RB - ROY - Rec.Custos \quad (54)$$

Além disso, conforme edital da 6ª rodada de partilha de produção, o limite para recuperação de custos será de 80%.

Para a consideração do OPEX no tempo foi considerada uma parcela fixa de 50% do OPEX total e outros 50% variando conforme a curva de produção. Para isto, foi considerado o artigo de Verre, Guibileo e Cadegiani (2009), que mostra que os dois principais componentes do OPEX são o custo de pessoal e o custo com serviços, como os de escoamento da produção.

Para a distribuição do CAPEX, foi feita a distribuição da tabela 9. Ela considera que uma parte do investimento em perfuração e completação ocorrerá durante e após o primeiro óleo para a entrada de novos poços. E que o abandono consumirá cerca de 3% do CAPEX descontado. Além disso, a curva de produção terá duração máxima de 27 anos, conforme ocorre nos contratos de concessão.

Tabela 9 - Distribuição dos investimentos em CAPEX

<b>Ano</b>	<b>Desembolso de CAPEX (%)</b>
PO - 3	10%
PO - 2	20%
PO - 1	35%
Primeiro óleo (PO)	15%
PO + 1	10%
PO + 2	7%
PO + 26	3%

Fonte: O autor, 2020.

Portanto, de posse de todas estas informações, é possível fazer a valoração de um projeto. Para cada ano, o fluxo de caixa será calculado da seguinte forma para o regime de concessão:

Tabela 10 - Cálculo do fluxo de caixa anual para o regime de concessão.

<b>Item</b>
Receita Bruta (RB)
(-) ROY
(-) OPEX
(-) Dep. CAPEX
(-) Participação especial (PE)
(=) LAIR
(-) IR/CSLL
(=) Lucro depois do IR
(-) CAPEX
(+) Dep. CAPEX
<b>(=) Fluxo de caixa (FLC)</b>

Fonte: O autor, 2020.

E da seguinte forma para o regime de partilha de produção:

Tabela 11 - Cálculo do fluxo de caixa anual no regime de partilha de produção

Item
Receita Bruta (RB)
(-) ROY
(-) OPEX
(-) Dep. CAPEX
(-) Excedente em óleo
(=) LAIR
(-) IR/CSLL
(=) Lucro depois do IR
(-) CAPEX
(+) Dep. CAPEX
<b>(=) Fluxo de caixa (FLC)</b>

Fonte: O autor, 2020.

Para cada um dos itens das duas tabelas a forma de cálculo já foi descrita de forma específica nessa dissertação.

Assim, a valoração da participação governamental seguiu mais a legislação e a realidade de regime fiscal do que os trabalhos de Blake e Roberts (2006) e Lima, Suslick e Bordieri (2005). Esta metodologia para o cálculo do fluxo de caixa é muito parecida com a utilizada por Marques (2015), ao considerar a depreciação dos equipamentos no fluxo de caixa.

O VPL deve ser calculado descontando todos os fluxos de caixa para valor presente. A TMA (taxa mínima de atratividade) deve ser tomada como o custo médio ponderado de capital da empresa. A título de exemplificação foi utilizada uma TMA de 10% ao ano.

$$VPL = \sum_1^n \frac{FLC}{(1 + TMA)^n} \quad (55)$$

O próximo capítulo utilizará um projeto de produção como base para comparação dos resultados obtidos através desta metodologia. Também serão feitas análises estimando a incerteza nos componentes do fluxo de caixa, de forma a representar adequadamente a realidade da exploração de petróleo, conforme discutido em outros trechos desta dissertação.

## **4. RESULTADOS**

Primeiramente é preciso definir qual projeto será utilizado como base e como vai ser feita a avaliação dos resultados. O artigo de Gaffney, Cline & Associates (2010), fez a avaliação econômica do volume médio de 10 prospectos e descobertas do pré-sal brasileiro. Para cada um deles há os dados necessários para a análise desta dissertação, ou seja, informações de volume recuperável, lâmina d'água, soterramento e área para cálculo do CAPEX e do OPEX e dados de vazão e produção por poço para estimação da curva de produção do projeto.

Foi escolhido o prospecto que tem uma distribuição de volume recuperável esperado similar aos volumes dos projetos utilizados nesta dissertação.

### **4.1 Aplicação no projeto base**

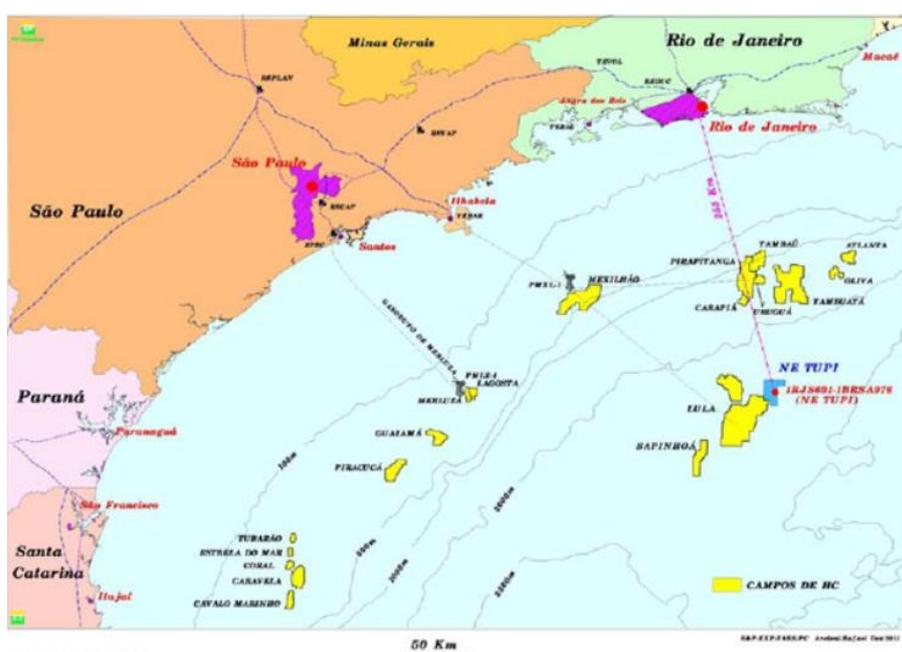
No âmbito do Gaffney, Cline & Associates (2010) o prospecto de Nordeste de Tupi se tratava de uma potencial acumulação próxima ao campo de Tupi sob o acordo da cessão onerosa. Mais tarde, em 2012 foi confirmada a descoberta de óleo de 26° API pela Petrobras através do poço 1-BRSA-976-RJS e em 2014 foi feita a declaração de comercialidade desta área.

Neste artigo o projeto foi descrito como localizado em lâmina d'água de aproximadamente 2100 metros, e a cerca de 250 km da cidade do Rio de Janeiro. Com uma

área de 60 km<sup>2</sup> e profundidade esperada de 5625 m até o *spill point* da estrutura. O início da sua produção está previsto para 2021<sup>2</sup>.

Importante ressaltar que para a finalidade deste trabalho, considera-se esta área ainda como um prospecto com grande incerteza no volume esperado. Ou seja, a mesma informação que existia na época do artigo. Porém, esta metodologia pode ser aplicada para projetos que realmente estejam nesta fase exploratória, desde que se tenha as informações sobre ele. Por exemplo, pode-se aplicar esta metodologia com projetos exploratórios atualmente existentes nas gerências de exploração das companhias de petróleo, desde que haja disponibilidade dos dados.

Figura 16 - Localização do prospecto Nordeste de Tupi.



Fonte: G1. *Petrobras confirma potencial na área Nordeste de Tupi*. 2012. Disponível em:

<<http://g1.globo.com/economia/noticia/2012/03/petrobras-confirma-potencial-na-area-nordeste-de-tupi-2.html>>. Acesso em: 20 Jun. 2020.

O volume inclui 3 previsões, sendo uma pessimista igual a 120 MM boe, uma moderada igual a 311 MM boe e outra otimista de 650 MM boe. Estes valores serão interpretados como

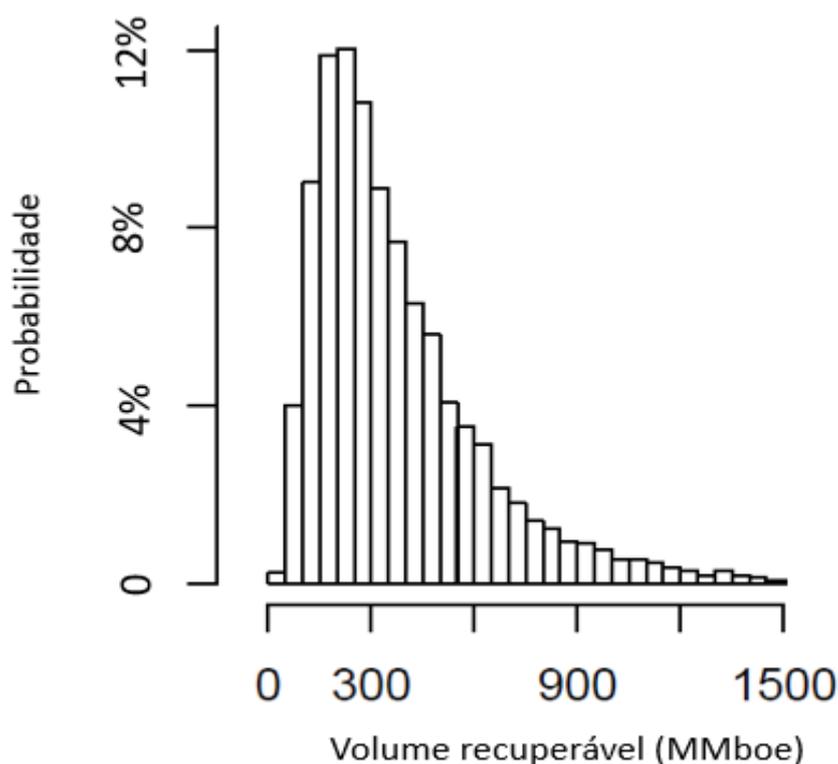
<sup>2</sup> Portos e Navios. *Módulos da FPSO de Sépia serão levados para integração na China*. Disponível em: <<https://www.portosenavios.com.br/noticias/ind-naval-e-offshore/modulos-da-fpso-de-sepia-serao-levados-para-integracao-na-china>>. Acesso em: 05 Jul. 2020.

P10, P50 e P90 na distribuição log-normal de volume recuperável, seguindo a metodologia explicitada no apêndice A.

Representando uma evolução em relação a alguns trabalhos citados nesta dissertação, foram feitas 10000 simulações para obter o fluxo de caixa esperado desta oportunidade exploratória. Como o tempo não foi um fator limitador para realização das simulações, não foi utilizada nenhuma técnica para melhoria de desempenho computacional da simulação de Monte Carlo.

A Figura 17 apresenta o histograma representando a distribuição log-normal do volume recuperável. Percebe-se que ele segue o formato que é esperado deste tipo de distribuição.

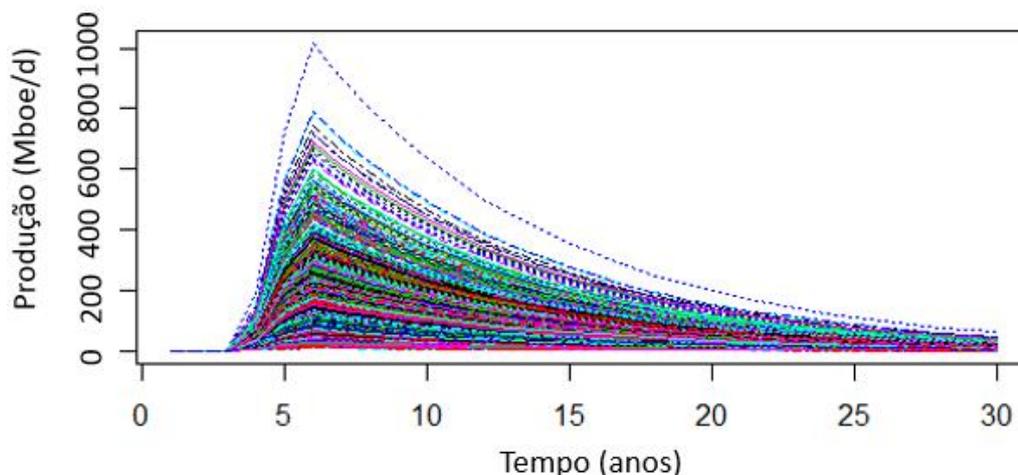
Figura 17 – Histograma do volume recuperável esperado.



Fonte: O autor, 2020.

Para a construção da curva de produção, a vazão de produção média esperada por poço é de 14000 boe/d. E a quantidade de poços produtores é igual a 7 no volume recuperável moderado (P50). Com estes parâmetros é possível construir a curva de produção esperada do projeto por meio de declínio exponencial, conforme indicado pela SPEE (2002).

Figura 18 - Possibilidades de curva de produção considerando a incerteza volumétrica.



Fonte: O autor, 2020.

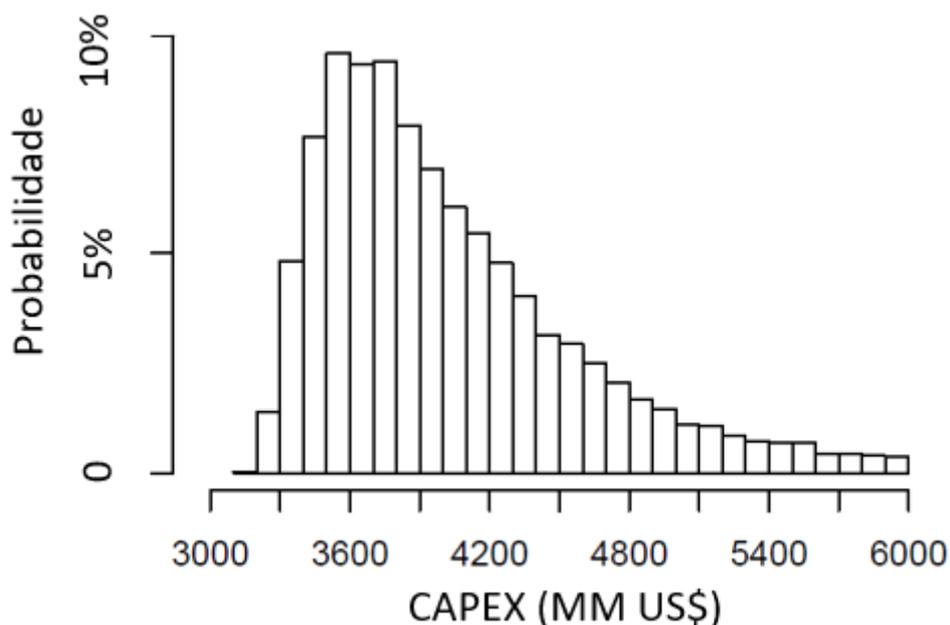
Assim, apesar de o volume médio resultar em uma curva de produção com pico esperado em 111 mil barris por dia, é possível perceber que este perfil pode ser muito diferente ao considerar a distribuição volumétrica como um todo, indo de 39 mil barris por dia no P10 até um valor de 206 mil barris por dia no P90. E, por se tratar de uma grande quantidade de iterações, obtém-se valores altamente improváveis como um projeto com curva de produção de quase 1 milhão de barris por dia no percentil 99,999%. A título de comparação, o FPSO contratado<sup>3</sup> para este campo, com início da produção prevista para 2021, tem capacidade para 180 M bbl/d. Ou seja, aparentemente o volume que posteriormente foi encontrado no reservatório se aproxima mais do P90 da distribuição.

O CAPEX esperado, considerando os parâmetros de lâmina d'água, soterramento, área e volume, obviamente, também varia muito. Podendo exigir um gasto de investimento variando de cerca de US\$ 3450 MM até US\$ 4900 MM em se tratando de P10 e P90, respectivamente.

---

<sup>3</sup> Portos e Navios. *Módulos da FPSO de Sépia serão levados para integração na China*. Disponível em: <<https://www.portosenavios.com.br/noticias/ind-naval-e-offshore/modulos-da-fps0-de-sepia-serao-levados-para-integracao-na-china>>. Acesso em: 05 Jul. 2020.

Figura 19 - Histograma do CAPEX esperado

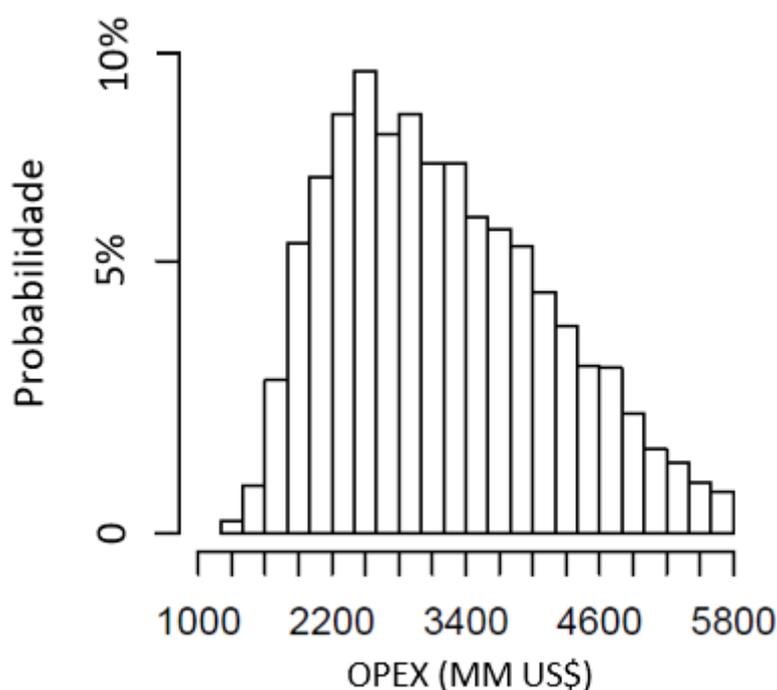


Fonte: O autor, 2020.

Esse tipo de análise possibilita um melhor planejamento de recursos para a atividade, ainda mais quando está associada a uma probabilidade, como é o caso. A empresa pode ter uma previsão probabilística do valor que terá que utilizar para fazer o desenvolvimento da produção daquela acumulação, em caso de descoberta. Neste prospecto, por exemplo, é possível dizer que há 19% de chance de necessitar de um investimento acima de 4500 MM US\$. No próximo tópico deste capítulo a distribuição desse CAPEX vai ficar mais visível ao mostrar o fluxo de caixa ao longo dos anos.

É possível fazer o mesmo tipo de análise com o OPEX, cujo valores de P10 e P90 ficam em cerca de US\$ 2000 MM US\$ e US\$ 4700 MM, respectivamente.

Figura 20 - Histograma do OPEX esperado.



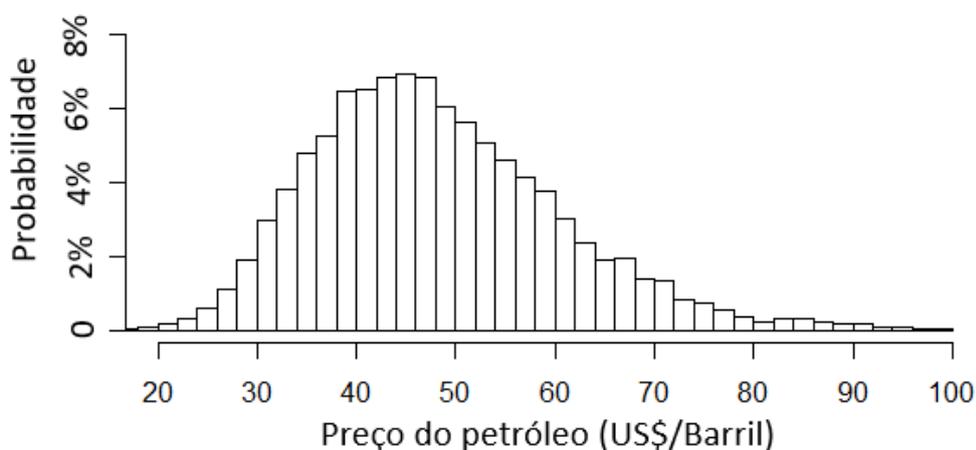
Fonte: O autor, 2020.

Com estes resultados, pode-se dizer que a probabilidade de os custos operacionais superarem US\$ 4500 MM é de 13%. Por se tratar de um projeto de produção, esse tipo de informação traz valor para um planejamento de longo prazo da empresa, mostrando a necessidade de caixa para fazer frente aos custos que surgirão. Destaca-se, que neste trabalho está sendo feita a análise de apenas um projeto. Porém, é possível fazer este tipo de avaliação com uma carteira de projetos, que possuam diferentes fatores de chance, para uma avaliação robusta do futuro da companhia.

É importante ressaltar que sem inserir a relação entre o OPEX e o preço do petróleo, estes custos operacionais estariam entre US\$ 2400 MM e US\$ 5300 MM. E, portanto, o projeto estaria subavaliado. Como mostra Schiozer, Lima e Suslick (2008), considerar a correlação entre essas duas informações pode inclusive alterar uma tomada de decisão.

O histograma abaixo mostra a incerteza no preço do petróleo no ano do pico da curva de produção (sexto ano do fluxo de caixa, terceiro ano de produção). Percebe-se que a maior parte dos valores se concentra entre 40 e 50 US\$/barril. Porém, outros cenários de preço estão representados nos resultados desta dissertação.

Figura 21 - Histograma do preço do petróleo no ano do pico da produção



Fonte: O autor, 2020.

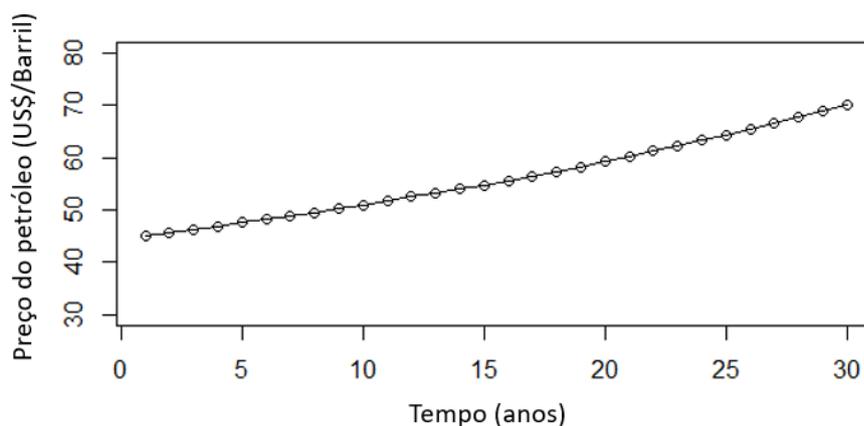
Tabela 12 - Estatística descritiva do Brent  
no ano do pico de produção

<b>Estatística</b>	<b>Brent (US\$/Barril)</b>
Mínimo	14,39
P25	39,27
Média	48,22
Mediana	46,66
P75	55,59
Máximo	127,61

Fonte: O autor, 2020.

E a Figura 22 representa o preço médio em cada ano, considerando as 10000 simulações. No ano de pico da produção o preço médio é de 48,22 US\$/barril, porém ele apresenta uma leve tendência de crescimento ao longo do tempo até chegar ao valor de 70 US\$/Barril no último ano do fluxo de caixa.

Figura 22 - Preço médio do petróleo por ano.



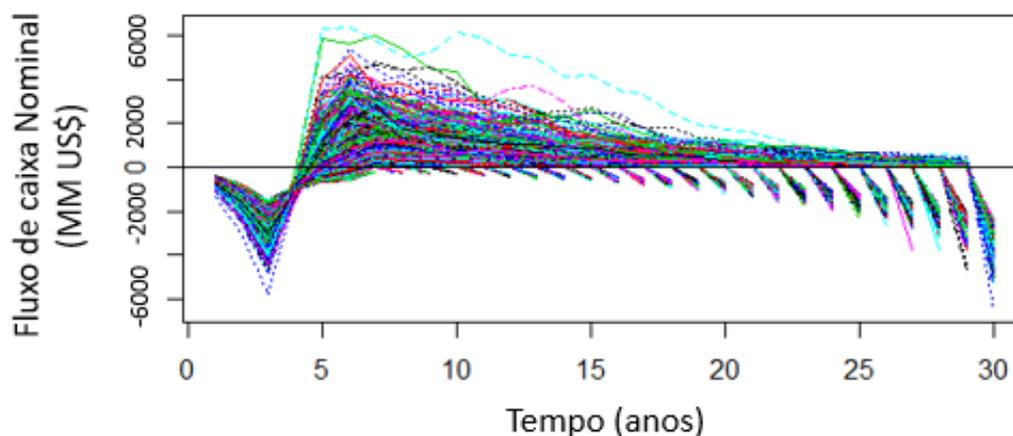
Fonte: O autor, 2020.

Após apresentar os parâmetros principais necessários para a avaliação econômica, o próximo tópico irá construir o fluxo de caixa nos dois regimes fiscais utilizados e obter os principais indicadores do projeto.

#### 4.2 Indicadores econômicos e comparação entre regimes fiscais

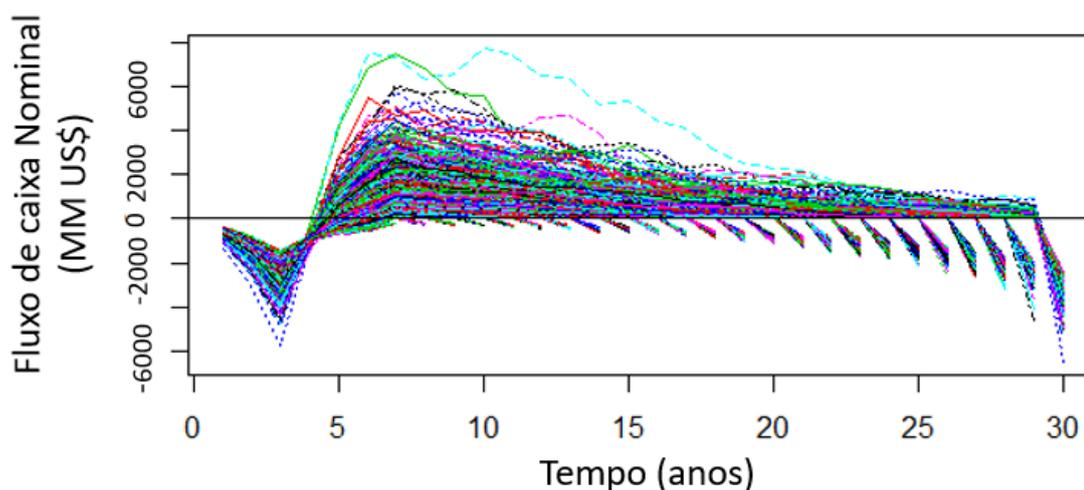
Neste tópico chega-se a outro objetivo do trabalho, que é obter uma distribuição de VPL para o projeto. Porém, antes de mostrar o VPL, os dois gráficos abaixo representam o fluxo de caixa esperado para este prospecto nos regimes de partilha de produção e de concessão, conforme premissas do capítulo 3.

Figura 23 - Incerteza no fluxo de caixa no regime de partilha de produção



Fonte: O autor, 2020

Figura 24 - Incerteza no fluxo de caixa no regime de concessão.



Fonte: O autor, 2020.

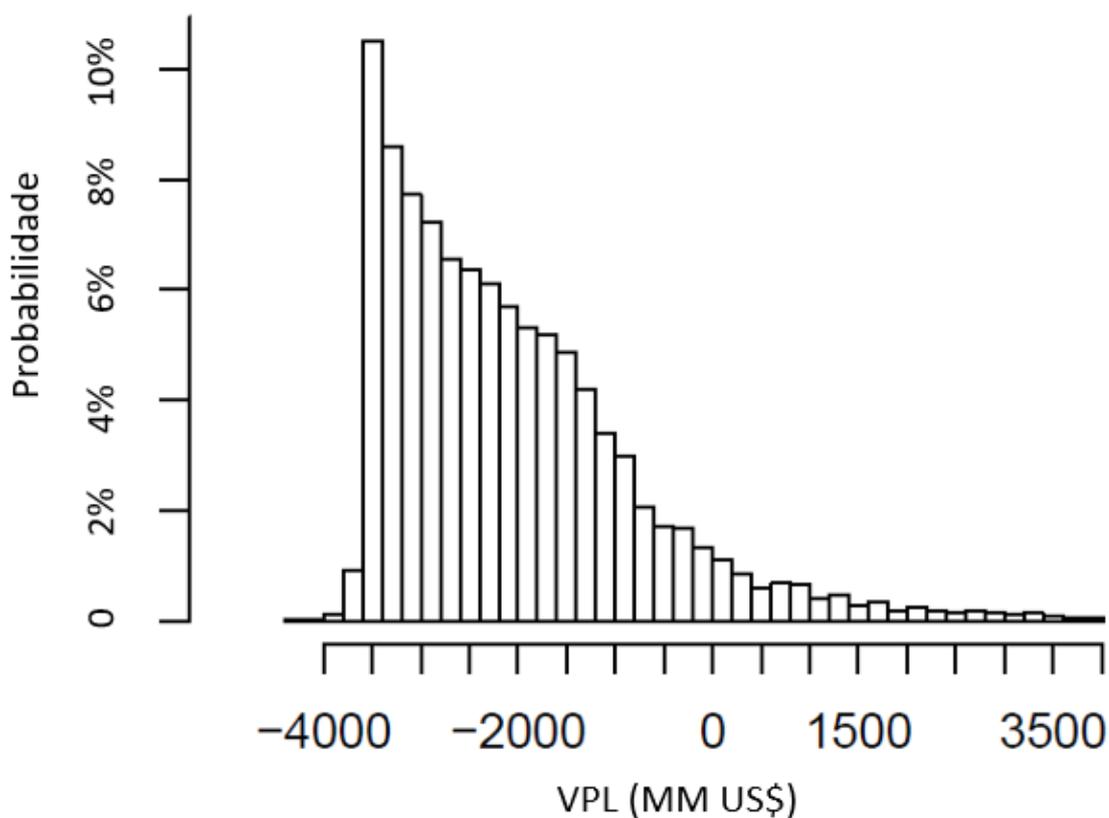
Importante ressaltar que ambas as análises tiveram como base as mesmas simulações de volume e preço do petróleo. Além disso, a estimativa dos custos também foi igual. Portanto, a única diferença é o regime fiscal.

Os valores negativos no final dos fluxos de caixa representam o abandono do projeto. É possível perceber pelas figuras que foi considerado o corte econômico no projeto. Ou seja, quando a receita começou a ficar menor do que os custos e encargos tributários, o projeto foi abandonado. Logo, quanto menor o volume da simulação, mais cedo ocorre o abandono. Esta consideração faz com que o VPL seja maximizado e represente melhor a tomada de decisão de um caso real.

Comparando as figuras com os dois regimes, pode-se observar que o regime de concessão possui fluxos de caixa em patamares levemente superiores ao regime de partilha. Isso indica uma possível maior rentabilidade devido à menor participação do governo.

O principal indicador de economicidade utilizado hoje em dia nas empresas para avaliação de projetos é o VPL, segundo Brealey, Myers e Allen (2013). E os gráficos abaixo mostram as possibilidades de resultado deste indicador dependendo do volume recuperável desta oportunidade exploratória e da premissa de preços.

Figura 25 - Histograma do VPL esperado no regime de partilha de produção



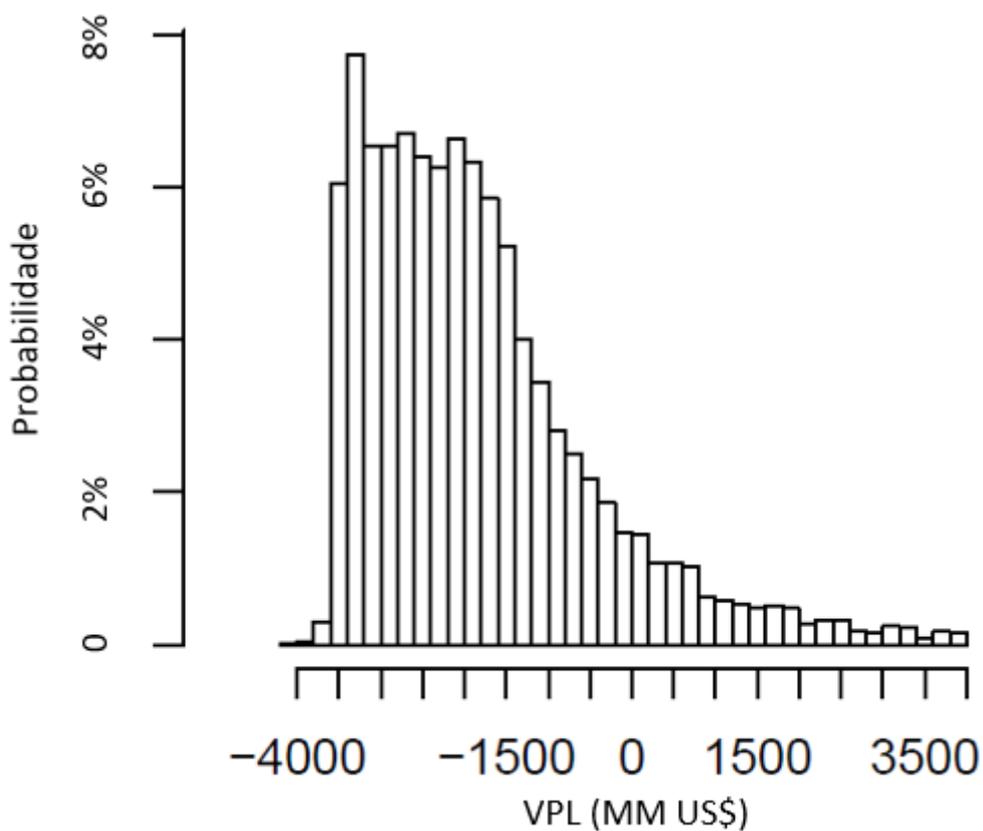
Fonte: O autor, 2020.

Tabela 13 - Estatística descritiva do VPL no regime de partilha de produção.

<b>Estatística</b>	<b>VPL (MM US\$)</b>
Mínimo	-4227
P25	-3075
Média	-2025
Mediana	-1336
P75	-1406
Máximo	19853

Fonte: O autor, 2020.

Figura 26 - Histograma do VPL esperado no regime de concessão.



Fonte : O autor, 2020.

Tabela 14 - Estatística descritiva do VPL no regime de concessão

<b>Estatística</b>	<b>VPL (MM US\$)</b>
Mínimo	-4138
P25	-2865
Média	-1744
Mediana	-2098
P75	-1179
Máximo	24578

Fonte: O autor, 2020.

No regime de partilha, o VPL se concentra entre US\$ -3423 MM e US\$ -344 MM em termos de P10 e P90. O valor médio indica um VPL de US\$ -2025 MM. E a probabilidade deste projeto ter um VPL positivo é de 7%.

No regime de concessão o VPL está entre US\$ -3315 MM e US\$ 142 MM considerando o P10 e o P90. Na média, este valor é de US\$ -1744 MM. E neste regime, a probabilidade de obter um VPL positivo neste projeto é de 11%.

Com isso, observa-se que nos valores médios deste trabalho o VPL no regime de concessão é maior do que no regime de partilha, mesmo resultado obtido por Marques (2015).

Ao longo desta dissertação foram descritas as premissas do trabalho que foram majoritariamente conservadoras, como a tributação da participação especial, a relação do preço de venda e o preço de tributação e a não consideração da recuperação de créditos fiscais de tributos indiretos. No próximo parágrafo será feita uma comparação com o trabalho base e será possível entender boa parte da diferença de VPL.

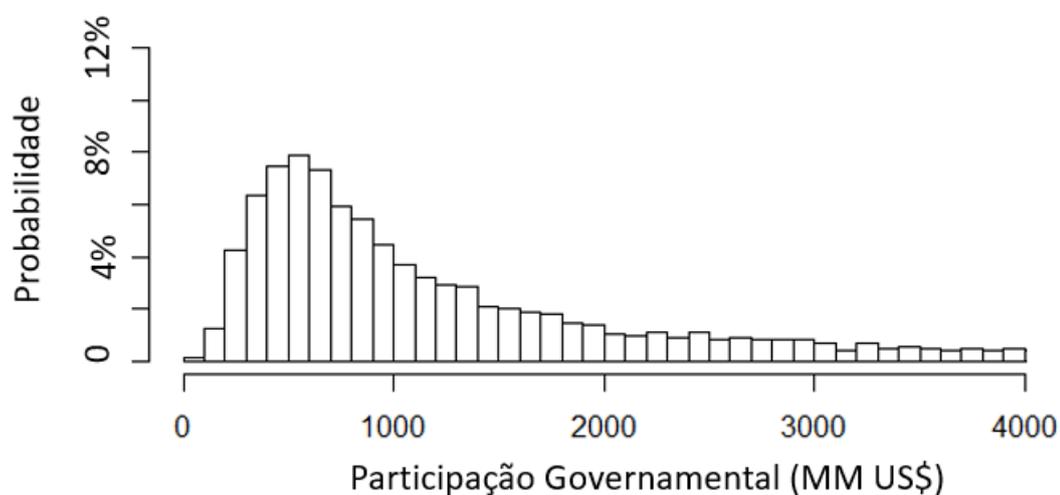
O trabalho de Gaffney, Cline & Associates (2010) estimou o VPL médio deste prospecto, com uma taxa de desconto de 10% ao ano, em US\$ 2931 MM, acima dos VPLs encontrados nesta dissertação. Isso se deve principalmente a dois fatores. O primeiro é a realidade de preços de mercado na época da realização de cada um dos trabalhos. O preço médio utilizado naquele artigo foi em torno de US\$ 110 por barril. Como se pode ver na Figura 15, bem acima do que foi utilizado neste trabalho. Se utilizada a mesma premissa de preço nesta dissertação obter-se-ia um VPL de US\$ 1407 MM no regime de concessão.

Além disso, a avaliação econômica de Gaffney, Cline & Associates (2010) foi feita sob o regime de cessão onerosa, que não tem pagamento de participação especial, por exemplo.

Segundo os próprios autores, isso pode levar a um valor cerca de 25% superior ao regime de concessão, que foi aqui utilizado.

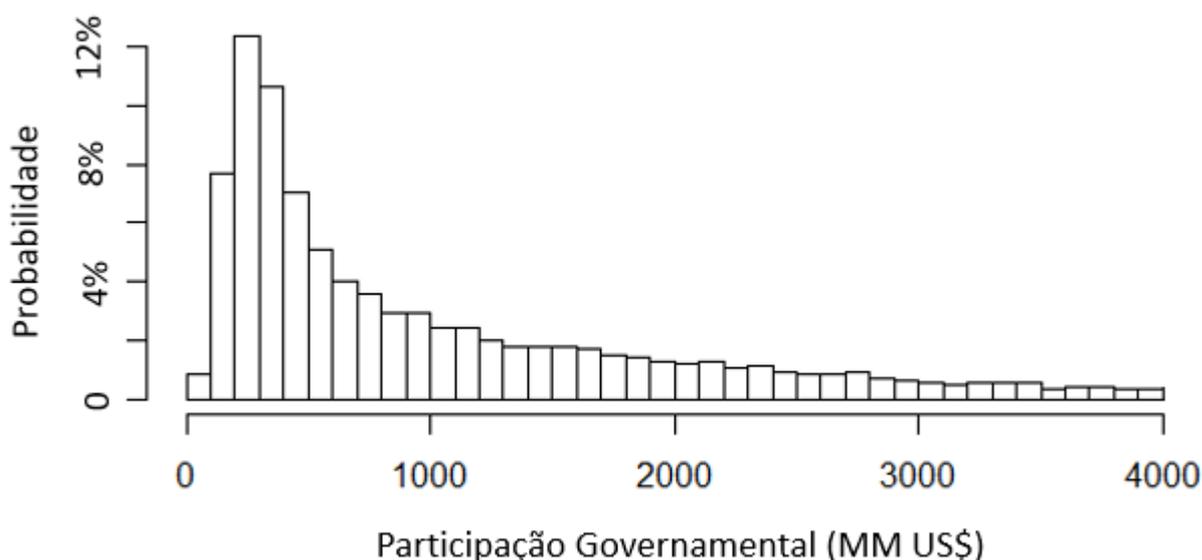
Voltando à comparação dos resultados nos dois regimes fiscais estimados nestes trabalhos, em termos de participação governamental, também há diferenças. No regime de partilha de produção, este valor é dado pela soma dos royalties (15% da receita bruta), imposto de renda e excedente em óleo. Já no regime de concessão, este valor é dado pela soma de royalties (10% da receita bruta), imposto de renda e participação especial. O detalhamento do cálculo destas contribuições está no capítulo 3. O histograma desta parcela em cada um dos modelos de tributação está nos gráficos abaixo.

Figura 27 - Histograma da participação governamental esperada no regime de partilha.



Fonte: O autor, 2020.

Figura 28 - Histograma da participação governamental esperada no regime de concessão



Fonte: O autor, 2020.

Portanto, no regime de partilha de produção a média da parcela governamental é igual a US\$ 1995 MM, 18% maior do que no regime de concessão (US\$ 1691 MM). Este resultado é similar ao encontrado pelos trabalhos de Marques (2015) e Soares (2017) nos quais a parcela do governo no regime de partilha se mostrou sempre maior, entre 5% e 35%, do que no regime de concessão. Logo, conforme Gobetti e Serra (2011), a tributação do regime de partilha permite que o governo se aproprie de uma parcela maior da riqueza do petróleo.

### 4.3 Comparação com a previsão dos custos utilizando regressão linear

Um dos objetivos deste trabalho foi obter modelos de previsão de CAPEX e OPEX que não se limitassem a uma simples regressão linear entre estes dois custos e as variáveis explicativas. Com isso, ao final do Tópico 3.2 foram selecionadas duas metodologias que melhor previam essas variáveis: lasso para o CAPEX e *smoothing splines* para o OPEX.

Este tópico irá comparar os resultados obtidos a partir destes dois modelos, que se encontram no tópico anterior, com os que seriam obtidos ao utilizar a regressão linear para previsão dos custos (equações 31 e 34).

Importante ressaltar que as simulações de preço e de volumes são iguais ao tópico anterior, variando apenas o modelo de previsão dos custos.

A tabela abaixo mostra, nos dois regimes fiscais analisados, a previsão média para o CAPEX e o OPEX deste projeto em cada metodologia e estatísticas do VPL para cada cenário feito.

Tabela 15 - Comparação de valores variando a metodologia para previsão do CAPEX e do OPEX nos dois regimes fiscais avaliados em MM US\$

	Concessão		Partilha		
	Regressão linear	Metodologia utilizada	Regressão linear	Metodologia utilizada	
Média CAPEX	4291	4091	4291	4091	
Média OPEX	3687	3248	3687	3248	
P25 VPL	-3028	-2865	-3232	-3075	
Média VPL	-2134	-1744	-2387	-2025	
P75 VPL	-1617	-1179	-1816	-1406	
Prob VPL > 0	6%	11%	3%	7%	
VPL com preço de Gaffney, Cline & Associates (2010)	1206	1407	169	319	2931
					Original

Fonte: O autor, 2020.

É possível perceber que a utilização de regressão linear para previsão dos custos envolvidos em um projeto de produção de petróleo levariam a valores superiores do que as metodologias que se aproximam do *machine learning* que foram utilizadas neste trabalho.

Também se observa na última linha da tabela que ao utilizar o mesmo preço e o mesmo volume que foi utilizado por Gaffney, Cline & Associates (2010) as metodologias escolhidas nesta dissertação tiveram um resultado mais próximo, em termos de VPL, do que ao utilizar a regressão linear. Conforme mencionado no tópico anterior a diferença de VPL encontrado nesta dissertação para o artigo também se deve aos diferentes regimes fiscais que foram utilizados.

Portanto, o VPL utilizando regressão linear é inferior ao VPL das outras análises em ambos os regimes fiscais. Ao se comparar estes valores com o projeto que foi utilizado como

base, fica evidente que o fluxo de caixa e os indicadores econômicos que utilizaram as metodologias lasso e *smoothing splines* para previsão do CAPEX e do OPEX foram melhores, pois ficaram mais próximos do que foi obtido no artigo de Gaffney, Cline & Associates (2010). Isso mostra a evolução que a utilização de métodos de previsão que se aproximam *machine learning* pode trazer.

## CONCLUSÃO

Esta dissertação se propôs a (1) estimar dois parâmetros necessários para a avaliação econômica de projetos exploratórios, CAPEX e OPEX, utilizando metodologias mais simples, como regressão linear, e outras com maior grau de flexibilidade que se aproximassem da aprendizagem estatística e permitissem um melhor ajuste aos dados; (2) representar a incerteza volumétrica em termos de incerteza econômica; e (3) comparar os resultados nos principais regimes de tributação vigentes no Brasil.

A utilização de distribuições nos indicadores financeiros e nas variáveis do fluxo de caixa no contexto da exploração de petróleo permite um melhor planejamento da financiabilidade da empresa, visto que é possível atribuir probabilidades para a necessidade de recursos financeiros ao invés de ficar sujeito à apenas uma previsão determinística obtida a partir da média da distribuição volumétrica. Também permite estimar o valor de uma empresa com melhor representatividade para ativos exploratórios, cujo possível resultado econômico não se restringe à média. A utilização de técnicas mais flexíveis para a estimação dos parâmetros necessários para a análise econômica busca seguir as tendências da análise de dados atual.

Para a estimação do CAPEX, a técnica que teve o melhor resultado em termos de MSE foi o *lasso* com volume de óleo, soterramento, lâmina d'água e área como variáveis independentes. Ou seja, todas as variáveis candidatas a regressoras entraram no modelo, com exceção do volume de gás. Isso indica que, para os dados analisados nesta dissertação, a presença de gás não possui impacto significativo no CAPEX. E para a estimação do OPEX, a metodologia com menor valor de MSE foi a *smoothing splines* com o volume total e área como variáveis explicativas.

A partir destas estimativas, foi possível traduzir a incerteza no volume a ser recuperado em valores de CAPEX e OPEX esperados. Ademais, a construção de fluxos de caixa no regime de partilha de produção e de concessão permitiu avaliar a distribuição do VPL, que teve 7% e 11% de chance de ser positivo para os respectivos regimes. Importante destacar que este resultado se aplica para o prospecto utilizado como base e sob as premissas desta dissertação,

incluindo as simulações de preço do petróleo que buscaram representar a realidade mais recente da indústria de petróleo. Segundo Lucchesi (2019), esta é a variável que mais impacta na rentabilidade do projeto.

Outro resultado obtido foi a observação de uma maior apropriação do governo no regime de partilha, conforme encontrado por outros trabalhos citados ao longo desta dissertação. Entretanto, é importante ressaltar que os regimes podem ser equivalentes dependendo dos valores de bônus, excedente em óleo, participação especial e royalties definidas pelo regulador.

Além disso, como sensibilidade, foi feita a estimativa do CAPEX e do OPEX utilizando regressão linear para comparar com os modelos que obtiveram os melhores resultados nessa dissertação (*lasso* para o CAPEX e *smoothing splines* para o OPEX). Ao comparar os valores desses custos e os VPLs obtidos, verificou-se que a utilização de uma regressão linear estaria subavaliando o projeto. Ou seja, os modelos que foram selecionados nesta dissertação demonstraram uma capacidade de previsão superior à regressão linear.

Por meio desses resultados foi possível observar a importância da utilização de métodos de previsão mais próximos do *machine learning*. A melhoria em relação a uma simples regressão linear foi significativa e pode alterar o resultado do processo decisório. Isso se destaca ainda mais na indústria de petróleo que é intensiva em capital e que envolve grandes incertezas.

Pesquisas futuras podem tentar aumentar o grau de flexibilidade na estimação dos parâmetros de custo, inclusive utilizando técnicas de aprendizado não supervisionado, nas quais o próprio algoritmo tenta encontrar padrões e relações entre as variáveis, sem nenhuma predeterminação de relação pelo responsável pela análise. Além disso, havendo a disponibilidade de se utilizar uma base de dados com uma quantidade maior de observações e de candidatos a variáveis independentes, provavelmente também haverá uma evolução na estimação destes parâmetros. Também pode-se utilizar uma base de dados com projetos mais homogêneos, ao contrário do que foi feito nesta dissertação e que possibilitou uma abrangência maior dos resultados. A separação do CAPEX em suas contas mais detalhadas como perfuração, instalações e escoamento também pode promover uma melhoria na estimação dos modelos, caso a base de dados tenha essa informação desagregada.

## REFERÊNCIAS

- AIUBE, F. A. L.; SAMANEZ, C. P. *On the Comparison of Schwartz and Smith's Two-and Three-Factor Models on Commodity Prices*. *Applied Economics*, Londres, v. 46, n. 30, p. 3736-3749, 2014.
- AKILU, N. Y.; MCVAY, D. A.; LEE, W. J. Inverted Hockey Stick Method Quantifies Price Uncertainty in Petroleum Investment Evaluation. *Journal of Petroleum science and technology*, v. 24, n. 2, p. 167-185, 2006.
- AMINZADEH, F.; DASGUPTA, S. N. *Geophysics for petroleum engineers*. 1º ed. Amsterdam: Elsevier Science Ltd., 2013. Cap. 3. Cap. 3: Fundamentals of petroleum geophysics.
- ANGUITA, D. et al. K-Fold Cross Validation for Error Rate Estimate in Support Vector Machines. In: *DMIN, 2009*, Genova. *University of Genova*. p.1-7.
- AMARAL, M. R.; *O Pré-sal no Regime de Partilha de Produção a Primeira Licitação e os Desafios de Implementação*. 2014. 114 f. Dissertação (Mestrado em Regulação e Gestão de Negócios) – Universidade de Brasília, Distrito Federal, 2014.
- BLAKE, A. J.; ROBERTS, M. C. Comparing Petroleum Fiscal Regimes Under Oil Price Uncertainty. *Resources Policy*, v. 31, n. 2, p. 95-105, 2006.
- BRADLEY, M. E.; Wood, A. R. O. Forecasting Oilfield Economic Performance. *Journal of Petroleum Technology*, v. 46, n. 11, p. 965-971, 1994.
- BREALEY, R. A.; MYERS, S.C.; ALLEN, F. *Princípios de Finanças Corporativas*. 12. ed. [S. l.]: AMGH, 2018. p.986.
- BRISCOE, E.; FELDMAN, J. Conceptual Complexity and The Bias/Variance Tradeoff. In: *International Journal of Cognitive Science*, 2010. p. 1-15.
- BUCKLEY, K.; UEHARA, R. Subsea Concept Alternatives for Brazilian Pre-Salt Fields. In: OFFSHORE TECHNOLOGY CONFERENCE BRASIL, 2017, Rio de Janeiro. *Offshore Technology Conference*, 2017. p. 1-19.
- BURN, R. *What Form of Cross-Validation Should You Use?* 2019. Disponível em: <https://towardsdatascience.com/what-form-of-cross-validation-should-you-use-76aaecc45c75>. Acesso em: 20 Jun. 2020.
- CAPEN, E. C. A Consistent Probabilistic Approach to Reserves Estimates. In: *SPE hydrocarbon economics and evaluation symposium*. Society of Petroleum Engineers, 1993.
- CARVALHO, R. F. C. *Migração Reversa No Tempo De Dados Sísmicos Blended*. 2012. 123 f. Dissertação (Mestrado em Engenharia Civil) – Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2012.
- CHAMBERS, J. M.; HASTIE, T. J. *Statistical Models in S*. 1. ed. Londres: Chapman & Hall/CRC, 1991. 626p.

- DAMÉ, O. M. *Regime De Partilha De Produção Brasileiro: Uma Abordagem De Teoria De Leilões*. 48 f. Tese (Doutorado) – Fundação Getulio Vargas Escola de Pós-Graduação em Economia, Rio de Janeiro, 2011.
- FEYGIN, M. V.; RYZHIK, V. M. Estimation of The Oil Reserves Requirement to Meet A Given Production Level—Mathematical Modeling. *Natural Resources Research*, v. 10, n. 1, p. 51-58, 2001.
- FORBES – *Global 2000: As Maiores Empresas De Capital Aberto Do Mundo Em 2019*. Disponível em: <https://forbes.com.br/listas/2019/05/global-2000-as-maiores-empresas-de-capital-aberto-do-mundo-em-2019/>. Acesso em: 20 Jun. 2020.
- GAFFNEY, Cline & Associates. Review and Evaluation of Ten Selected Discoveries and Prospects in the Pre-Salt Play of the Deepwater Santos Basin. *Brazil, ANP*, 2010.
- GOBETTI, S. W.; SERRA, R. V. Novo Marco Regulatório do Petróleo: Desafios na Transição do Regime de Concessão para o Regime de Partilha. XVI PRÊMIO TESOURO NACIONAL, v. 16, 2011.
- GOMES, C. J. V. O Marco Regulatório da Prospecção de Petróleo no Brasil: o Regime de Concessão e o Contrato de Partilha de Produção. *Centro de Estudos da Consultoria do Senado Federal*, 2009.
- GUSTAVSON, J. B. Valuation of non-US Oil and Gas Properties. *Journal of petroleum technology*, v. 52, n. 02, p. 56-61, 2000.
- HAHN, W. J.; DYER, J. S Incorporating Mean-Reverting Price Forecasts into Exploration and Production Project Valuation. In: SPE HYDROCARBON ECONOMICS AND EVALUATION SYMPOSIUM, 2005, Dallas. *Society of Petroleum Engineers*, 2005. p.1-9.
- HANS, C. Bayesian lasso regression. *Biometrika*, v. 96, n. 4, p. 835-845, 2009.
- HARZER, J. H. et al. Abordagem Probabilística do Indicador TMA/TIR para Avaliação do Risco Financeiro em Projetos de Investimentos. In: CONGRESSO BRASILEIRO DE CUSTOS, 21,2014, Natal. *Congresso Brasileiro de Custos*, 2014. p. 2-16.
- HASTENREITER, L.; GALVAO, M. Probabilistic Evaluation of Oil and Gas Reserves Considering the Effect of Change in the Oil Price. In: OFFSHORE TECHNOLOGY CONFERENCE BRASIL, 2012, Houston: *Offshore Technology Conference*, 2012. p. 1-11.
- HAWKINS, D. M. The Problem of Overfitting. *Journal of chemical information and computer sciences*, v. 44, n. 1, p. 1-12, 2004.
- HOCKING, R.R. A Biometrics Invited Paper. The analysis and selection of variables in linear regression. *Biometrics*, p. 1-49, 1976.
- HOLMES, J. C.; MENDJOGE, A. V.; MCVAY, D. A. Quantification of Oil Price Uncertainty in Economic Evaluation Using Sequential Gaussian Simulation. *Petroleum science and technology*, v. 24, n. 8, p. 943-962, 2006.

JABLONOWSKI, C. J.; STRACHA, A. Empirical Cost Models for TLPs and Spars. In: ANNUAL TECHNICAL CONFERENCE AND EXHIBITION, Denver, 2008. *Society of Petroleum Engineers*, 2008. p.1-10.

JAMES, G. et al. *An Introduction to Statistical Learning*. New York: springer, 2013.

JANSEN, S. *Hands-On Machine Learning for Algorithmic Trading: Design and Implement Investment Strategies Based on Smart Algorithms that Learn from Data Using Python*. Packt Publishing Ltd., 2018. E-book.

JARQUE, C. M.; BERA, A. K. Efficient Tests for Normality, Homoscedasticity and Serial Independence of Regression Residuals. *Economics Letters*, v. 6, n. 3, p. 255-259, 1980.

KARLIK, C. W. Parametric Estimating of Oil and Gas Production Facilities Capital Costs Worldwide. In: SPE HYDROCARBON ECONOMICS AND EVALUATION SYMPOSIUM, 1991, Dallas: *Society of Petroleum Engineers*, 1991. p.1-9.

KUNJAN, B. Exploration Chance of Success Predictions—Statistical Concepts and Realities. *ASEG Extended Abstracts*, v. 2016, n. 1, p. 1-8, 2016.

KUZNETSOV, M. et al. Capital Costs Estimation Method for Arctic Offshore Oil Projects. In: SPE ARCTIC AND EXTREME ENVIRONMENTS CONFERENCE AND EXHIBITION, Moscow, 2011. *Society of Petroleum Engineers*, 2011. p.1-5.

LI, K; HORNE, R. N. An Analytical Model for Production Decline-Curve Analysis in Naturally Fractured Reservoirs. *SPE Reservoir Evaluation and Engineering*, v. 8, n. 3, p. 197-204, 2005.

LI, K.; HORNER, R.N. Verification of Decline Curve Analysis Models for Production Prediction. In: SPE WESTERN REGIONAL MEETING, Irvine, 2005. *Society of Petroleum Engineers*, 2005. p.1-8.

LIMA, G. A.C; SUSLICK, S. B.; BORDIERI, C. A. A Quantitative Method for Estimation of Volatility of Oil Production Projects. In: SPE HYDROCARBON ECONOMICS AND EVALUATION SYMPOSIUM, Dallas, 2005. *Society of Petroleum Engineers*, 2005. p.1-12.

LOADER, C. *Local Regression and Likelihood*. 1 ed. [S. I.]: Murray Hill: Springer, 1999. 209p.

LOADER, C. Smoothing: Local Regression Techniques. In: *Handbook of Computational Statistics*. Springer, Berlin, Heidelberg, p. 571-596, 2012.

LUCCHESI, R. D. Main Factors Impacting Oil Projects Return: A Sensitivity Analysis. In: OFFSHORE TECHNOLOGY CONFERENCE, Houston, 2019. *Offshore Technology Conference*, 2019. p.1-7.

MACARY, S. M.; RAZEK, M. A.; GOHARY, H.E. Analysis of Oil Field Economic Performance. In: INTERNATIONAL OIL AND GAS CONFERENCE AND EXHIBITION IN CHINA, Beijing, 2000. *Society of Petroleum Engineers*, 2000. p.1-8.

MARQUES, L. M. The Fiscal System Influence on Oil Fields Development and Government Participation. In: SPE ANNUAL TECHNICAL CONFERENCE AND EXHIBITION, Houston, 2015. *Society of Petroleum Engineers*, 2015. p. 1-14.

MESZ, L. *Investimento em Prospectos Exploratórios sob Incertezas Técnica-Geológica e de Preços*. 2018. 90 f. Trabalho de Conclusão de Curso (Graduação em Ciências Econômicas) – Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2018.

MONCRIEFF, I. Crystallography: The Long-Term Price of Oil. *Journal of Petroleum Technology*, v. 60, n. 03, p. 46-50, 2008.

MONTGOMERY, D. C.; PECK, E. A.; VINING, G. G. *Introduction to Linear Regression Analysis*. 5. ed. John Wiley & Sons, Hoboken, 2012.

MOORE, D. S.; MCCABE, G. P. *Introduction to The Practice of Statistic*. 6.ed. New York: WHFreeman & Co., 2009.

NOVAES, F.; PORSANI, M. J. Processamento de Dados Sísmicos da Bacia de Campos Utilizando de Convolução Multicanal para a Supressão de Múltiplas. In: 10th INTERNATIONAL CONGRESS OF THE BRAZILIAN GEOPHYSICAL SOCIETY, 2007, Rio de Janeiro. 2007. SBGf - *Sociedade Brasileira de Geofísica*, 2007. p. 1-7.

NYSTAD, A. N. Economic Analysis of the North Sea Oil and Gas Region. *Journal of Petroleum Technology*, v. 33, n. 12, p. 2.515-2.527, 1981.

OHISHI, M.; YANAGIHARA, H.; FUJIKOSHI, Y. A Fast Algorithm for Optimizing Ridge Parameters in A Generalized Ridge Regression by Minimizing A Model Selection Criterion. *Journal of Statistical Planning and Inference*, v. 204, p. 187-205, 2020.

OLSEN, G.T.; FARIYIBI, F.L.; MCVAY D.A. Price Uncertainty Quantification Models Advance Project Economic Evaluations. In: SPE HYDROCARBON ECONOMICS AND EVALUATION SYMPOSIUM, 2005, Dallas. *Society of Petroleum Engineers*, 2005. p. 1-9.

OTIS, R. M.; SCHNEIDERMAN, N. A process for evaluating exploration prospects. *AAPG bulletin*, v. 81, n. 7, p. 1087-1109, 1997.

RODRIGUEZ, J. D.; PEREZ, A.; LOZANO, J. A. Sensitivity Analysis of K-Fold Cross Validation in Prediction Error Estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 32, n. 3, p. 569-575, 2010.

ROSE, P. R. *Risk Analysis and Management of Petroleum Exploration Ventures*. Tulsa, OK: American Association of Petroleum Geologists, 2001.

SCHIOZER, R. F.; LIMA, G. A. C.; SUSLICK, S. B. The Pitfalls of Capital Budgeting When Costs Correlate to Oil Price. *Journal of Canadian Petroleum Technology*, v. 47, n. 08, 2008.

SCHWARTZ, E.; SMITH, J. E. Short-Term Variations and Long-Term Dynamics in Commodity Prices. *Management Science*, v. 46, n. 7, p. 893-911, 2000.

SOARES, L. S. F. *Regimes Fiscais na Indústria do Petróleo: A Influência de Características Contratuais na Atratividade Econômica de Projetos de Exploração e Produção*. 2017. 168 f. Dissertação (Mestrado em Finanças e Economia Empresarial) – Fundação Getúlio Vargas, Rio de Janeiro, 2017.

SOUZA, C. G. C. *Prêmio de Risco de Mercado Variante no Tempo e Estrutura de Erro Autorregressiva para os Preços do Petróleo*. 2019. 71 f. Dissertação (Mestrado em Ciências Econômicas) – Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2019.

SPEE – Society of Petroleum Evaluation Engineers. *Recommended Evaluation Practice #6: Definition of Decline Curve Parameters*. [ S. I.]: SPEE, 7 p , 2002.

STEVENSON, W.J.; DE FARIAS, A. A. *Estatística Aplicada à Administração*. 1 ed. São Paulo: Harbra, 1986.

SUROVTSEV, D.; SUNGUROV, A. " Vaguely Right or Precisely Wrong?": Making Probabilistic Cost, Time, and Performance Estimates for Bluefields Appraisal. In: SPE RUSSIAN PETROLEUM TECHNOLOGY CONFERENCE AND EXHIBITION, 2017, *Society of Petroleum Engineers*, 2017. p. 1-9.

SUSLICK, S. B. et al. Forecasting the Development of Heavy-Oil reserves in Ultra-Deep Waters Using Technological Risk Models. In: SPE HYDROCARBON ECONOMICS AND EVALUATION SYMPOSIUM, 2003, Dallas. *Society of Petroleum Engineers*, 2003. p. 1-7.

SZKLO, A.; MACHADO, G.; MORAIS, A. O Caso Marlim: Descontos da Corrente de Exportação em Relação ao Brent e Políticas da Petrobras Para Reduzir Esta Perda. In: CONGRESSO BRASILEIRO DE PESQUISA E DESENVOLVIMENTO EM PETRÓLEO E GÁS, 4., 2007, Campinas. *PDPETRO*, 2007. p. 1-8.

TIBSHIRANI, R. Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, v. 58, n. 1, p. 267-288, 1996.

TIROLE, J. Market Failures and Public Policy. *American Economic Review*, v. 105, n. 6, p. 1665-82, 2015.

TOBIN, J. Biases, Pressures, and Problems in Energy Price Forecasting. *Journal of Petroleum Technology*, v. 60, n. 09, p. 24-27, 2008.

TOEWS, G.; NAUMOV, A. The Relationship Between Oil Price and Costs in The Oil Industry. *The Energy Journal*, v. 36, n. Adelman Special Issue, 2015.

VERRE, F.; GIUBILEO, A.; CADEGIANI, C. Asset Life-Cycle OPEX Modelling with Montecarlo Simulation to Reduce Uncertainties and to Improve Field Exploitation. In: ANNUAL TECHNICAL CONFERENCE AND EXHIBITION, 2009, Louisiana. *Society of Petroleum Engineers*, 2009. p. 1-6.

VUKELIC, M.; PESCARA, R.; FRANCO, A. Economic Evaluation of a Multi-Prospect Exploratory Block. In: SPE LATIN AMERICAN AND CARIBBEAN PETROLEUM ENGINEERING CONFERENCE, 2001, Buenos Aires. *Society of Petroleum Engineers*, 2001. p. 1-16.

WALLACH, D.; GOFFINET, B. Mean Squared Error of Prediction as A Criterion for Evaluating and Comparing System Models. *Ecological modelling*, v. 44, n. 3-4, p. 299-306, 1989.

WANG, K.; CHEN, Z. Stepwise Regression and All Possible Subsets Regression in Education. *Electronic International Journal of Education, Arts, and Science (EIJEAS)*, v. 2, 2016.

WOOD, S. N. *Generalized Additive Models: An Introduction With R*. 2. ed. CRC press, London, 2017.

WU, S.; MCAULEY, K. B.; HARRIS, T. J. Selection of Simplified Models: I. Analysis of Model-Selection Criteria Using Mean-Squared Error. *The Canadian Journal of Chemical Engineering*, v. 89, n. 1, p. 148-158, 2011.

YUAN, M. GACV for Quantile Smoothing Splines. *Computational Statistics & Data Analysis*, v. 50, n. 3, p. 813-829, 2006.

## APÊNDICE – Avaliação de prospectos exploratórios

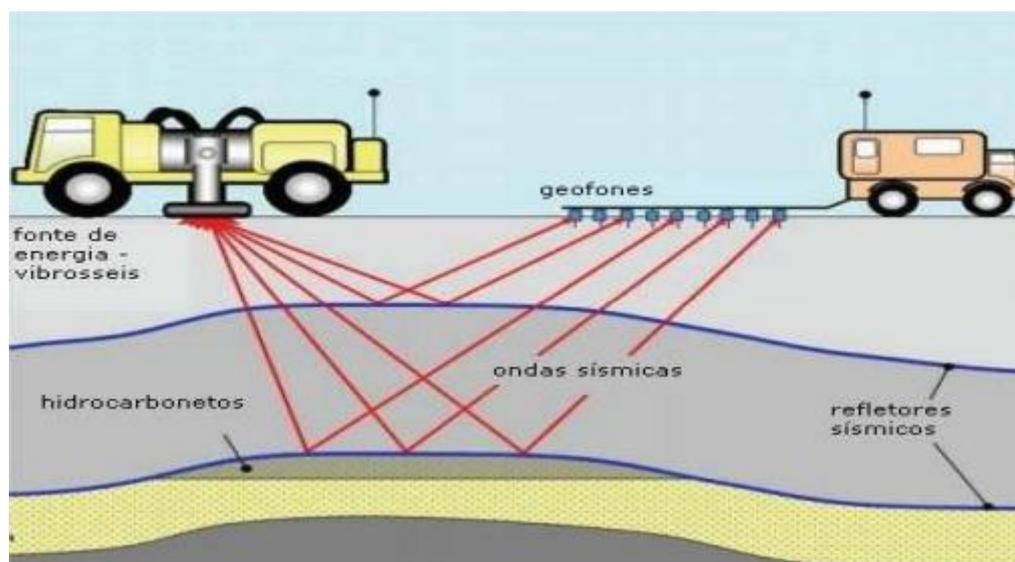
Para absorver o conteúdo desta dissertação e entender por que a consideração do risco é tão importante na exploração de petróleo, conforme preconiza Rose (2001), é preciso entender o processo de análise de uma possível acumulação que não é uma tarefa trivial e envolve muitos riscos.

Há diversas metodologias e ferramentas para se identificar possibilidades de acumulações de petróleo. Geralmente, a partir de métodos não-sísmicos, como gravimetria e magnetometria, geólogos e geofísicos começam a ter uma noção do que pode estar presente abaixo da superfície. Porém, são os métodos sísmicos (chamados de sísmicas de reflexão e refração) que proporcionam uma melhor representação do que se pode encontrar (Carvalho, 2012).

A forma de se mapear o subterrâneo varia de acordo com as características de cada área. Entretanto, é possível dividir em dois grandes grupos: Sísmica terrestre e sísmica marítima.

Em uma sísmica em terra, são instalados explosivos a cerca de 2 a 6 metros de profundidade (ou outro método de geração de vibração). Ao acioná-los, ondas irão se propagar para as camadas mais profundas, e serão refletidas parcialmente de volta, sendo captadas por receptores instalados na superfície. A velocidade de propagação dessas ondas fornece uma noção do tipo de rocha presente em cada profundidade e como elas estão dispostas, dessa forma, contribuindo para a análise das condições para se ter um sistema petrolífero.

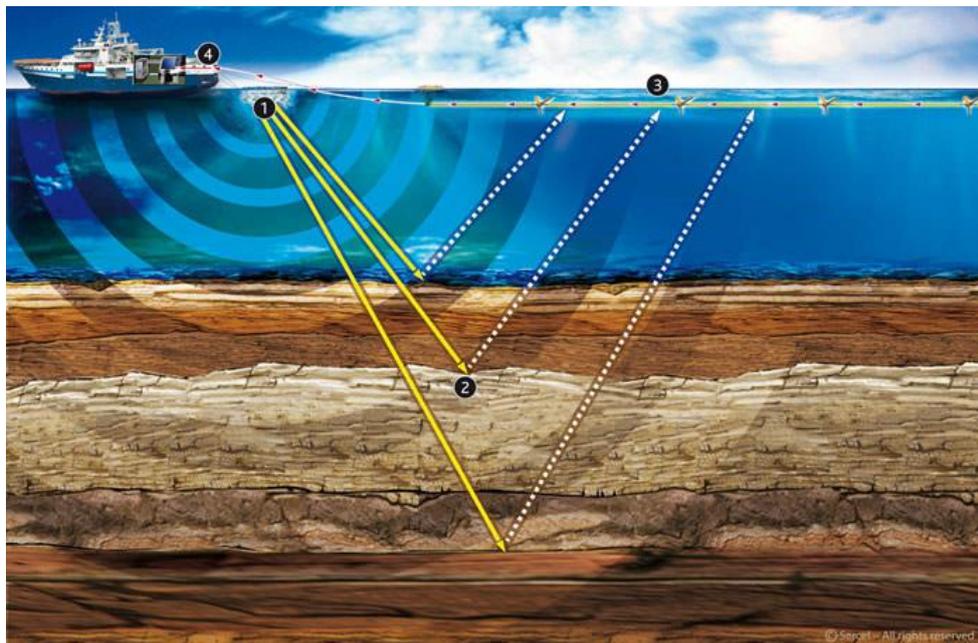
Figura 29 - Modelo de aquisição sísmica terrestre



Fonte: ACIEP. Disponível em: <<http://www.aciep.com/c-mo-se-hace-la-exploraci-n>> Acesso em: 05 Jul. 2020.

Em uma sísmica no mar, são utilizados navios equipados com air guns (canhões de ar comprimido) que liberam tiros de ar comprimido em direção ao leito marinho. Ao mesmo tempo, receptores de variação de pressão estão acoplados em cabos (streamers) amarrados ao navio e em seu rastro para recepção do sinal sísmico.

Figura 30 - Representação de uma aquisição sísmica marítima. (1) Air gun. (2) camada de rocha. (3) Receptores (hidrofonos) (4) navio de aquisição sísmica.



Fonte: Sercel. Disponível em: <[www.sercel.com/about/what-is-geophysics.aspx](http://www.sercel.com/about/what-is-geophysics.aspx)> Acesso em: 05 Jul. 2020.

Após a aquisição sísmica, é feito o processamento deste dado de modo a obter uma imagem de boa qualidade e retirando os ruídos presentes nas reflexões. (Novaes, 2007). Dessa forma, o dado estará pronto para ser trabalhado pelo intérprete responsável pelo mapeamento de estruturas naquela área.

Porém, por melhor que seja a qualidade da sísmica e das técnicas atuais, não é possível dizer com total certeza qual tipo de rocha está presente na subsuperfície, nem se há a presença de óleo.

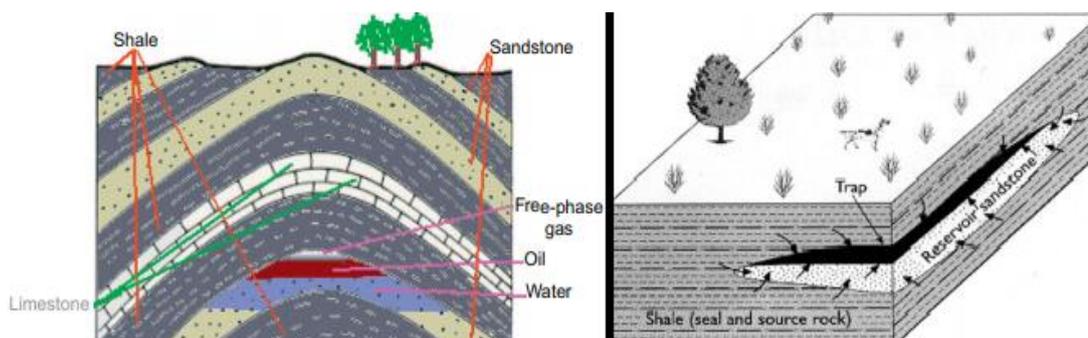
Para tratar este segundo risco e representar a probabilidade da presença de óleo, uma das metodologias mais utilizadas atualmente foi proposta por Otis e Schneidermann (1997). Ela apresenta quatro riscos geológicos na probabilidade de encontrar hidrocarboneto no objetivo. O primeiro risco é o de ter a presença de uma rocha geradora de óleo, sem uma rocha deste tipo não tem como encontrar hidrocarboneto.

O segundo risco trata da existência de uma rocha reservatório. Ou seja, uma rocha com características, dentre outras, de porosidade, que permita que o hidrocarboneto se acumule em seus poros.

O terceiro risco é o da existência de uma trapa, que também pode ser chamada de armadilha. Esta armadilha é necessária para o óleo não escapar da rocha reservatório e se

“espalhar”, sem ter uma acumulação concentrada em determinado ponto. O que impossibilita a extração econômica desse fluido. O risco de trapa incorpora o risco de geometria, ou seja, a armadilha deve ter uma estrutura que permita que o óleo não saia. E o risco de retenção, ou seja, o risco de ter uma rocha não porosa protegendo a trapa. Dois tipos clássicos de trapa, conforme Dasgupta e Aminzadeh (2013) estão presentes na Figura 31.

Figura 31 - Trapa estrutural na esquerda e estratigráfica na direita



Fonte: Dasgupta e Aminzadeh, 2013

Finalmente, o quarto risco é o de sincronismo e migração. Ou seja, a rocha geradora deve ter gerado o óleo no tempo certo e deve haver um caminho para que este óleo saia da rocha geradora e vá para a rocha reservatório.

A multiplicação destes quatro riscos é chamada de probabilidade de sucesso geológico. Ou seja, na ocorrência destes 4 elementos simultaneamente é possível dizer que há a presença de uma acumulação de petróleo.

Este mesmo artigo sugere uma metodologia para considerar a incerteza com relação a parâmetros que entram no cálculo do volume de óleo a ser encontrado (considerando a probabilidade de sucesso geológico). Estes parâmetros são:

- Área da acumulação: O dado sísmico não fornece uma certeza da área da acumulação. Ou seja, apesar de ele ser essencial para a identificação de oportunidades exploratórias, não se pode dizer com exatidão o tamanho da área desta oportunidade.

- Net pay: Altura da coluna de óleo a ser encontrado após a perfuração.

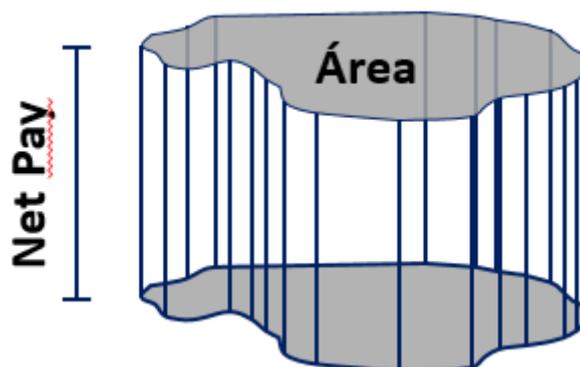
- Porosidade: Não é possível dizer através da sísmica qual o tipo de rocha e qual a porosidade exata que está presente. Uma porosidade maior propicia um acúmulo de hidrocarboneto maior.

- Saturação: O fluido não é puramente composto por hidrocarboneto aproveitável economicamente. Há um certo grau de mistura com outras substâncias como água e gás.

- Fator volume de formação de óleo: Ele é uma relação entre o volume de óleo no reservatório e o volume de óleo em condições de superfície.

Assim, o volume de óleo in place (VOIP) é dado pela multiplicação da área da acumulação pelo net pay, que resulta em um volume, conforme Figura 32. Como nem toda a rocha está preenchida por petróleo, este volume é multiplicado por uma porosidade. E, ainda há a multiplicação pela saturação, pois o petróleo vem misturado com outras substâncias e pelo fator volume de formação de óleo.

Figura 32 - Exemplo esquemático do volume de uma acumulação



Fonte: O autor, 2020.

Assim, a fórmula do VOIP é a seguinte:

$$VOIP_{\text{óleo}} = \frac{(A \times NP \times \Phi \times S_{hc})}{B_o} \quad (56)$$

onde:

$A$  = Área da acumulação

$NP$  = Net pay

$\Phi$  = Porosidade

$S_{hc}$  = Saturação

$B_o$  = Fator de formação de óleo

Porém, como estes cinco parâmetros são incertos, o artigo propõe que seja considerada uma distribuição probabilística em cada um deles. Esta distribuição, assim como a distribuição de volume resultante é, em geral, assumida como log-normal, assim como outros recursos naturais (CAPEN, 1993).

Dessa forma, para cada parâmetro são estimados pelos intérpretes valores de percentil 5% (P5) e 95% (P95) da distribuição e o volume é obtido a partir da multiplicação destes valores utilizando simulação de Monte Carlo. Essa técnica, consiste em milhares de iterações nas quais são multiplicados valores de cada uma das distribuições dos parâmetros, obtendo assim uma distribuição representativa da variável resultante, neste caso o VOIP.

Porém, é essencial dizer nesta dissertação que o VOIP não é o volume de óleo que será recuperado do reservatório e trazido para a superfície. Apenas uma fração deste volume (obtido pelo fator de recuperação) é passível de exploração pelas tecnologias atuais. Geralmente o fator de recuperação é influenciado por características como viscosidade do fluido, que impacta na necessidade de esforços para possibilitar seu escoamento, e permeabilidade do reservatório, pois apenas uma alta porosidade não garante que o petróleo irá fluir com facilidade pela rocha até o poço produtor.

O fator de recuperação (FR) também é estimado estatisticamente, da mesma forma que os demais parâmetros já apresentados.

Assim, a fórmula do volume recuperável é a seguinte:

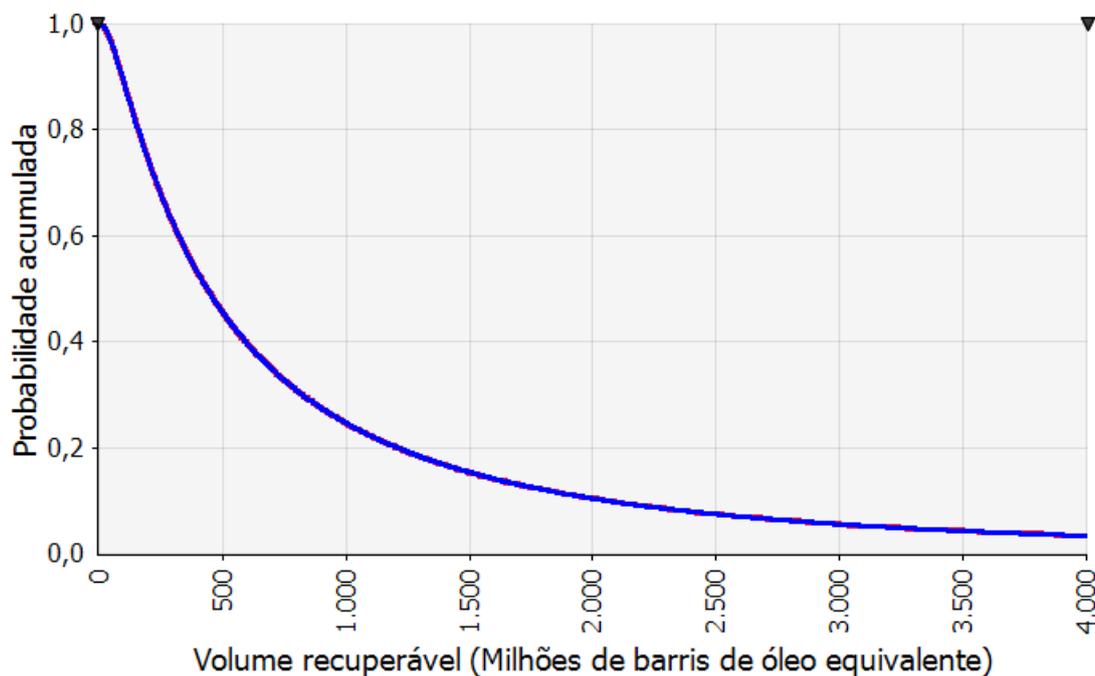
$$VRF_{\text{óleo}} = \frac{(A \times NP \times Phi \times Shc \times FR)}{Bo} \quad (2)$$

onde,

*FR* = Fator de recuperação (Volume recuperável dividido pelo volume in place).

Dessa forma, o volume de óleo a ser recuperado e aproveitado economicamente segue uma distribuição probabilística, conforme o exemplo da Figura 32. Neste caso, o volume médio da distribuição é de 897 milhões de barris de óleo equivalente (MM boe) e se encontra próximo ao percentil 30%.

Figura 33 - Exemplo de distribuição de probabilidade acumulada do volume recuperável de um prospecto exploratório



Fonte: O autor, 2020.

Neste momento se insere a avaliação econômica deste prospecto exploratório. Por se tratar de uma distribuição probabilística, é preciso escolher um ponto (volume de reservas) como referência para se realizar a avaliação econômica. Segundo Rose (2001), a análise na média da distribuição é a escolha adequada na maioria dos casos. Porém, para projetos mais importantes, a análise do fluxo de caixa descontado deveria ser feita também para o P90, P10 e P50 (mediana). Portanto, vemos que o ideal seria que a análise econômica fosse feita em todos os pontos, conforme propõe esta dissertação (com suas devidas limitações). Na ausência desta possibilidade, sugere-se a análise em 4 pontos da curva de distribuição probabilística.