3
Metodologia de Avaliação da Relação entre o Custo Operacional e o Preço do Óleo

Este capítulo tem como objetivo apresentar a metodologia que será empregada nesta pesquisa para avaliar a dependência entre duas importantes variáveis presentes no processo de estimativa de reservas: o custo operacional e o preço do óleo.

Na Seção 3.1 será apresentado o fenômeno *price-effect-cost-escalation* (*PECE*), apresentado por Nystad (1981) e utilizado no estudo de Bradley e Wood (1993). Este fenômeno é caracterizado pela mudança do comportamento dos custos em função da variação do preço do óleo.

A Seção 3.2 apresentará a metodologia de regressão linear simples que será utilizada para estudar a relação de dependência entre os custos operacionais e o preço do óleo. Será feita uma revisão do método dos Mínimos Quadrados Ordinários, da distribuição de probabilidade e variância dos coeficientes de regressão e do coeficiente de determinação. Serão apresentadas também as hipóteses do modelo clássico de regressão linear.

3.1
Efeito-Preço

O mais importante acelerador do custo operacional, segundo Bradley e Wood (1993), é o preço do óleo. A alteração desta variável econômica gera uma reestruturação da indústria, levando a mudanças no custo dos serviços. Nystad (1981) foi um dos primeiros autores a observar o impacto do preço do óleo nos custos. Este fenômeno foi batizado como *price-effect-cost-escalation* (*PECE*).

PECE foi definido como a resposta do custo em função de uma alteração no preço do óleo. Sendo assim,

$$\left(PECE\right)_t = \frac{\mu pc\left(PO_t - PO_i\right)}{\left(PO_i\right)} \qquad \text{(Eq. 1)}$$

sendo,

PO_i, o preço do óleo O no tempo inicial i,

PO_t, o preço do óleo O no tempo t,

μpc, o módulo $PECE$.

O estudo de Nystad (1981) teve como foco o efeito da mudança do preço do óleo nos custos de perfuração de poços. Já o trabalho de Bradley e Wood (1993) se utilizou do conceito apresentado por Nystad (1981) para estudar a relação entre o preço e os custos operacionais e de manutenção.

Contudo, estimar o módulo *PECE* para campos que já estão em operação é razoavelmente complexo. Segundo Bradley e Wood (1993), ao observar uma variação nos custos operacionais, é difícil saber se a causa está nas alterações das atividades do campo ou na mudança do preço do óleo.

Uma alternativa é analisar em conjunto dados históricos de índices de custos operacionais e índices de preços do óleo. Observar uma mudança no comportamento dos custos operacionais como função de uma alteração do preço pode ser uma opção para avaliar a relação de dependência destas duas variáveis.

Nesta pesquisa, a avaliação desta relação será feita por meio de regressão, assumindo que a relação entre estas duas variáveis é linear. Assim, na Seção 3.2 será apresentada a metodologia a ser utilizada para a análise de regressão linear.

3.2
Regressão Linear

A fim de investigar e modelar a relação entre o preço do óleo e os custos operacionais será utilizada a análise de regressão linear. Nesta pesquisa será aplicada a metodologia de Regressão Linear tal como utilizada em Gujarati e Porter (2011).

O índice de preço do óleo será chamado nesta pesquisa de variável explicativa, ou variável *X*. Já o índice de custo operacional será chamado de variável dependente, ou variável *Y*. Como índice, entende-se a razão entre o custo operacional, ou preço do óleo, no instante *t* e no instante inicial *i*.

Como se trata da dependência de uma variável em relação a uma única variável explicativa o tipo de análise a ser realizada denomina-se análise de regressão linear simples.

Os índices de custo operacional também são conhecidos como valores esperados condicionais, uma vez que dependem dos valores da variável índice de preço do óleo. Considerando uma relação linear entre a variável dependente e explicativa, é possível unir os valores esperados condicionais com uma linha reta. Assim, obtém-se a Linha de Regressão Populacional (LRP), ou a regressão de *Y* contra *X*.

3.2.1
Função de Regressão

Cada valor esperado condicional $E(Y/X_i)$ é uma função de X_i, em que X_i é um dado valor de *X*. Ou seja,

$$E(Y/X_i) = f(X_i) \qquad \text{(Eq. 2)}$$

$f(X_i)$ pode ser chamada de Função de Regressão Populacional (FRP). Como suposição, nesta pesquisa a FRP é uma função linear de X_i do tipo,

$$E(Y/X_i) = \beta_1 + \beta_2 X_i \qquad \text{(Eq. 3)}$$

em que β_1 e β_2 são parâmetros desconhecidos chamados de coeficientes de regressão. β_1 também é conhecido por intercepto e β_2 por coeficiente angular.

Na análise de regressão, tem-se por objetivo estimar os valores das incógnitas β_1 e β_2, com base nas observações de *X* e *Y*.

Contudo, é importante citar que em muitos casos trabalha-se com dados amostrais. Ou seja, os mesmos não representam o comportamento fidedigno de toda uma população. Portanto, ao invés de estimar a FRP, será estimada a Função de Regressão Amostral (FRA). Assim, as linhas de regressões aqui estimadas serão aproximações da linha de regressão populacional.

A equação de regressão para as amostras será, portanto, escrita como,

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i \qquad \text{(Eq. 4)}$$

em que,

$\hat{Y}_i$ é o estimador de $E(Y / X_i)$, $\hat{\beta}_1$ é o estimador de β_1 e $\hat{\beta}_2$ é o estimador de β_2.

Em termos estocásticos, a equação de regressão é expressa como,

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{u}_i \qquad \text{(Eq. 5)}$$

em que $\hat{u}_i$ denota o termo residual ou,

$$\hat{u}_i = Y_i - \hat{Y}_i$$

Assim, o passo inicial da análise de regressão linear é estimar a FRA da maneira mais precisa possível. Nesta pesquisa, o método a ser utilizado para estimar a FRA será o método dos Mínimos Quadrados Ordinários (MQO).

3.2.2
Método dos Mínimos Quadrados Ordinários

O método dos MQO adota o critério de que a FRA deve ser escolhida de tal forma que a soma dos quadrados dos resíduos seja a menor possível. Este critério é adotado já que os resíduos nada mais são do que a diferença entre os valores observados e os estimados de *Y*. Assim, tem-se que,

$$\sum \hat{u}_i^2 = \sum (Y_i - \hat{Y}_i)^2 = \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)^2 \qquad \text{(Eq. 6)}$$

Nota-se então que a soma dos quadrados dos resíduos são dependentes dos valores de $\hat{\beta}_1$ e $\hat{\beta}_2$. Assim, o princípio dos MQO é escolher os valores de $\hat{\beta}_1$ e $\hat{\beta}_2$ de tal forma que, para qualquer amostra, a soma dos quadrados dos resíduos seja a menor possível. Esta escolha é feita por um exercício de cálculo diferencial. Este processo de diferenciação resulta em duas equações chamadas de equações normais. Tais equações estão expressas a seguir:

$$\sum Y_i = n\hat{\beta}_1 + \hat{\beta}_2 \sum X_i \qquad \text{(Eq. 7)}$$

$$\sum Y_i X_i = \hat{\beta}_1 \sum X_i + \hat{\beta}_2 \sum X_i^2 \qquad \text{(Eq. 8)}$$

em que *n* é o tamanho da amostra.

Ao se resolver simultaneamente as duas equações normais obtêm-se os valores de $\hat{\beta}_1$ e $\hat{\beta}_2$ conforme as próximas equações.

$$\hat{\beta}_2 = \frac{\sum (X_i - \overline{X})(Y_i - \overline{Y})}{\sum (X_i - \overline{X})^2} = \frac{\sum x_i y_i}{x_i^2} \qquad \text{(Eq. 9)}$$

$$\hat{\beta}_1 = \frac{\sum X_i^2 \sum Y_i - \sum X_i \sum X_i Y_i}{n\sum X_i^2 - (\sum X_i)^2} = \overline{Y} - \hat{\beta}_2 \overline{X} \qquad \text{(Eq. 10)}$$

em que $\overline{X}$ e $\overline{Y}$ são as médias amostrais de X e Y, respectivamente. Além disso, fica definido que $x_i = \left(X_i - \overline{X}\right)$ e $y_i = \left(Y_i - \overline{Y}\right)$.

Gujarati e Porter (2011) mostraram que é necessário medir a precisão dos estimadores de $\hat{\beta}_1$ e $\hat{\beta}_2$. Uma forma é calcular a variância destes estimadores com base nas seguintes equações:

$$Var\left(\hat{\beta}_2\right) = \frac{\sigma^2}{\sum x_i^2} \qquad \text{(Eq. 11)}$$

$$Var\left(\hat{\beta}_1\right) = \frac{\sum X_i^2}{n\sum x_i^2}\sigma^2 \qquad \text{(Eq. 12)}$$

em que *Var* significa variância.

Os autores também mostraram que o valor de σ^2 pode ser estimado por:

$$\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{(n-2)} \qquad \text{(Eq. 13)}$$

em que $\hat{\sigma}^2$ é o estimador dos MQO de σ^2. A expressão *(n-2)* é conhecida por número de graus de liberdade, em que *n* é o tamanho da amostra. Já a expressão $\sum \hat{u}_i^2$ é conhecida por Soma dos Quadrados dos Resíduos (SQR).

3.2.3
Distribuição de Probabilidade dos Estimadores

Outra importante informação acerca de $\hat{\beta}_1$ e $\hat{\beta}_2$ é em relação à distribuição de probabilidade destes estimadores. Para se conhecer a distribuição, Gujarati e Porter (2011) mostraram que os coeficientes são funções lineares de Y_i, como mostra a exemplificação com $\hat{\beta}_2$ a seguir:

$$\hat{\beta}_2 = \sum k_i Y_i \qquad \text{(Eq. 14)}$$

em que $k_i = x_i \big/ \sum x_i^2$

Supondo que os valores de X são não-estocásticos, pode-se dizer então que $\hat{\beta}_2$ é uma função linear de Y_i, que é aleatória por hipótese.

Pode-se reescrever a expressão acima como,

$$\hat{\beta}_2 = \sum k_i \left(\beta_1 + \beta_2 X_i + u_i\right) \qquad \text{(Eq. 15)}$$

Como k_i, β_1, β_2 e X são fixos, conclui-se que $\hat{\beta}_2$ é função linear da variável aleatória u_i. Portanto, $\hat{\beta}_2$, assim como $\hat{\beta}_1$, tem a mesma distribuição de probabilidade de u_i.

Segundo Gujarati e Porter (2011), o método dos MQO não faz suposição quanto à distribuição de probabilidade de u_i. Porém, no contexto geral dos problemas de regressão, supõe-se que os u seguem uma distribuição normal.

A hipótese de normalidade para os u é justificada pelo Teorema Central do Limite (TCL). Os u representam a influência combinada de muitas variáveis que não foram incluídas no modelo. O TCL permite mostrar que se há um grande número de variáveis aleatórias independentes e identicamente distribuídas, a distribuição da soma destas variáveis tende à distribuição normal.

Desta forma, é razoável supor que os estimadores $\hat{\beta}_1$ e $\hat{\beta}_2$ seguem uma distribuição normal. Assim, $\hat{\beta}_1 \sim N\left(\beta_1, \sigma^2_{\hat{\beta}_1}\right)$ e $\hat{\beta}_2 \sim N\left(\beta_2, \sigma^2_{\hat{\beta}_2}\right)$.

3.2.4 Qualidade do Ajustamento da Regressão

Gujarati e Porter (2011) mostraram também que, depois de estimar os coeficientes de regressão, é necessário avaliar a qualidade do ajustamento da linha de regressão amostral aos dados. O coeficiente de determinação r^2 é uma medida de qualidade deste ajustamento. r^2 pode assumir valores entre 0 e 1, e quanto maior o valor assumido por este coeficiente, maior a parte da variação de Y que é explicada por X. Em outras palavras, r^2 mede o percentual da variação total de Y que pode ser explicado pelo modelo de regressão.

Assim,

$$r^2 = \frac{SQE}{SQT} \quad \text{(Eq. 16)}$$

em que, SQT é a soma dos quadrados totais e SQE é a soma dos quadrados explicados pela regressão. SQT e SQE podem ser calculados por:

$$SQT = \sum \left(Y_i - \overline{Y}\right)^2 \quad \text{(Eq. 17)}$$

$$SQE = \hat{\beta}_2 \sum x_i^2 \quad \text{(Eq. 18)}$$

A soma dos quadrados totais SQT pode ser escrita como,

$$SQT = SQE + SQR \quad \text{(Eq. 19)}$$

em que SQR é a soma dos quadrados dos resíduos como já citado anteriormente, ou variação inexplicada, e pode ser calculada como,

$$SQR = \sum \hat{u}_i^2 \quad \text{(Eq. 20)}$$

3.2.5
Hipóteses do Modelo Clássico de Regressão Linear

O modelo clássico de regressão linear parte de sete hipóteses, conforme descrito por Gujarati e Porter (2011). Tais hipóteses são:

Hipótese 1: o modelo de regressão é linear nos parâmetros, embora possa não ser linear nas variáveis.

Hipótese 2: os valores assumidos pela variável explicativa *X* podem ser fixos em amostras repetidas ou seus valores podem mudar de acordo com a variável dependente *Y*. No segundo caso deve ser assumido que as variáveis *X* e o termo de erro são independentes.

Hipótese 3: o valor médio ou esperado do termo de erro u_i é zero.

Hipótese 4: a variância do termo de erro é a mesma independente do valor de *X*.

Hipótese 5: não há auto-correlação entre os termos de erro.

Hipótese 6: o número de observações *n* deve ser maior que o número de parâmetros a serem estimados.

Hipótese 7: deve haver variabilidade dos valores de *X*, ou seja, os valores de *X* em uma amostra não devem ser os mesmos.

Nesta pesquisa, as hipóteses 1, 2, 3, 6 e 7 serão consideradas verdadeiras, seja por simplificação, seja pelas características dos dados que serão analisados.

Por simplificação entende-se que as hipóteses 1, 2 e 3 são atendidas. Sobre a Hipótese 1, o modelo de regressão será linear nos parâmetros e nas variáveis e não serão estudadas formas não lineares. Sobre a Hipótese 2, os valores assumidos por X, ou índice de preço do óleo, são fixos para todas as amostras. A Hipótese 3 assume que o valor esperado do termo de erro é zero. Esta hipótese implica que não há erro de especificação do modelo. Em outras palavras, outros fatores não incluídos no modelo não afetam sistematicamente o valor de Y. Nesta pesquisa, só será estudada a influência do preço do óleo no comportamento do custo operacional. A possibilidade de inclusão de outras variáveis não será estudada. Portanto, a Hipótese 3, por consideração, será verdadeira.

Pelas características dos dados, as Hipóteses 6 e 7 são satisfeitas. Sobre a Hipótese 6, o número de observações a serem analisadas será maior que o número de parâmetros a serem estimados, que no caso serão apenas dois, $\hat{\beta}_1$ e $\hat{\beta}_2$. Sobre a Hipótese 7, haverá variabilidade dos valores de X, ou do índice de preço do óleo.

As Hipóteses 4 e 5 também são conhecidas como Hipótese da Homocedasticidade e Hipótese da Ausência de Correlação dos Resíduos, respectivamente. Estas hipóteses não serão consideradas verdadeiras sem antes fazer a verificação por meio de testes específicos. Caso os testes indiquem que, para os conjuntos de dados, as Hipóteses 4 e 5 não são válidas, medidas corretivas serão efetuadas a fim de que se possa realizar a regressão linear de forma correta.

As metodologias de avaliação em relação à homocedasticidade e ausência de correlação dos resíduos estão descritas nos Apêndices I e II, respectivamente.

3.3
Considerações Finais

Neste capítulo foi apresentada a metodologia que será empregada para avaliar a dependência entre os custos operacionais de campos produtores de petróleo e o preço do óleo. Esta pesquisa tem como premissa que há uma relação linear entre estas duas variáveis e, portanto, será empregada a metodologia de regressão linear.

A fim de estimar a função de regressão linear amostral, será utilizado o método dos Mínimos Quadrados Ordinários. Com a aplicação deste método serão obtidas as estimativas do termo do intercepto, $\hat{\beta}_1$, e o coeficiente angular, $\hat{\beta}_2$, da função de regressão.

O coeficiente angular da regressão linear é uma importante variável nesta pesquisa. Como explicado anteriormente, *PECE* representa a resposta do custo em função de uma alteração no preço do óleo. Por sua vez, esta resposta depende do módulo *PECE*. Esta última variável representa a elasticidade entre o preço e o custo.

Ao se avaliar os dados históricos do incremento do preço do óleo e o impacto deste incremento no comportamento dos custos operacionais, a elasticidade entre estas duas variáveis poderá ser representada pelo coeficiente angular da linha de função de regressão amostral. Desta forma, o módulo *PECE* será o valor estimado para $\hat{\beta}_2$.

Este capítulo também demonstrou que a função de distribuição de probabilidade que caracteriza $\hat{\beta}_2$ é a distribuição normal, com valor esperado β_2 e variância $\sigma^2_{\hat{\beta}_2}$. O tipo de distribuição, o valor esperado e a variância do módulo *PECE* representam informações fundamentais que serão utilizadas nos próximos capítulos.