

4. METODOLOGIA

Após a obtenção do lucro do capital e do valor do capital de cada uma das pequenas empresas brasileiras, como descrito acima, estima-se a curva que relaciona estas duas variáveis. Esta estimação baseia-se em um método não-paramétrico. O objetivo de uma regressão não-paramétrica é encontrar uma relação entre as variáveis sem impor nenhum formato. As técnicas geralmente neste método se baseiam na estimação de densidades *kernel*. Portanto, faz-se necessário apresentar a estimação de densidades *kernel* antes de introduzir a metodologia utilizada neste capítulo.

4.1. Estimação de densidades *kernel*¹

A função de densidade *kernel* pode ser representada por:

$$K_h(x) = \frac{1}{h} K\left(\frac{x}{h}\right)$$

onde h é o parâmetro que regula o grau de suavidade de uma densidade *kernel*. Este parâmetro é denominado de janela.

O estimador de uma função de densidade *kernel* é dado por:

$$\hat{f}_h(x) = \frac{1}{n} \sum_{i=1}^n K_h(x - X_i) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right)$$

Este estimador representa a média das funções *kernel* nas observações X_i .

Como a *kernel* é uma função de densidade, então o seu estimador também o é, de modo que:

$$\int K(x) dx \Rightarrow \int \hat{f}_h(x) dx = 1$$

A função de densidade *kernel* é estimada baseando-se em dois parâmetros, na janela (h) e na função *kernel* (K) escolhida. Se uma janela h e uma função *kernel* são corretamente especificadas, então a correspondente função de densidade *kernel* estimada é única para aquele conjunto de dados.

¹ A estimação da função de densidade *kernel* da principal variável utilizada ao longo deste trabalho, lucro antes da retirada dos custos de capital, encontra-se no anexo 2.

A função de densidade *kernel* estimada é assintoticamente não-viesada quando h converge para zero. Mas, como geralmente trabalha-se com uma janela diferente de zero, tem-se um estimador viesado. O viés do estimador é uma função da janela escolhida, sendo menor quando a janela escolhida for pequena. A variância da função de densidade *kernel* também dependerá do tamanho da janela escolhida. Quanto maior for h , menor será a variância. Logo, aumentar o tamanho da janela, diminui a variância, mas aumenta o viés.

Assim, deve-se escolher um tamanho de janela que esteja comprometido com os dois efeitos. Uma das regras para escolha de h é escolher aquele que minimiza o erro quadrático médio, que é dado por:

$$\hat{h}_0 = 1.06 \min\left(\hat{\sigma}, \frac{\hat{R}}{1.34}\right) n^{-\frac{1}{5}}$$

onde $\hat{\sigma}$ representa o estimador para a variância de x ; \hat{R} a extensão do interquartil de x , dado por: $\hat{R} = X_{[0.75n]} - X_{[0.25n]}$; e n o número de observações.

Para a estimação de uma densidade *kernel*, além da escolha de h , é necessário escolher o K . A escolha de K também pode ser baseada na minimização do erro quadrático médio da estimação. As funções *kernel* conhecidas como “Epanechnikov” minimizam o erro quadrático médio, sendo conhecidas como funções *kernel* ótimas de segunda ordem. Estas funções possuem a seguinte representação:

$$k(u) = \frac{3}{4}(1-u^2)I(|u| \leq 1)$$

Como afirma Härdle (1990), na prática, a escolha da *kernel* é algo irrelevante para a precisão da estimação, sendo a escolha da janela o mais importante.

A partir da estimação de uma função densidade pode-se calcular alguns funcionais de uma densidade *kernel*, como por exemplo a sua derivada. Um dos métodos utilizados para estimar a derivada de uma densidade *kernel* é obter a derivada do estimador desta função de densidade *kernel*, isto é, se:

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right)$$

é o estimador da função de densidade *kernel*, denominada de \hat{f} , então:

$$\hat{f}_h^{(p)}(x) = \frac{1}{nh^{p+1}} \sum_{i=1}^n K^{(p)}\left(\frac{x - X_i}{h}\right)$$

pode ser escolhido como estimador de $f^{(p)}(x)$.

4.2. Regressão não-paramétrica utilizando *kernel*

Dado n observações independentes $\{(X_i, Y_i)\}_{i=1, \dots, n}$, a relação encontrada através de uma regressão é do tipo:

$$Y_i = m(X_i) + \xi_i,$$

sendo $E(\xi_i / X_i = x) = 0$ e $Var(\xi_i / X_i = x) = \sigma^2(x)$.

Geralmente, procura-se obter a dependência média de Y em torno de $X=x$, tendo-se que estimar a curva da média condicional dada por:

$$m(x) = E(y / X = x) = \frac{\int y f(x, y) dy}{f(x)},$$

onde $f(x)$ representa a densidade marginal de x ; e $f(x,y)$ representa a densidade conjunta de (X,Y) .

Se a distribuição de $f(x,y)$ for normal com média zero, a curva de regressão será linear e teremos: $m(x) = \rho x$, com $\rho x = Cov(X, Y)$.

A idéia da regressão não-paramétrica é ponderar a variável resposta Y em certa vizinhança de x , de modo que:

$$\hat{m}_h = n^{-1} \sum_{i=1}^n W_h(x; X_1, \dots, X_n) Y_i,$$

onde $W_h(\cdot)$ representa a função peso que é dada pela distância de X_i a x . A função peso irá depender do parâmetro de suavização dos dados e da amostra de variáveis explicativas.

Todos os métodos de regressão não-paramétrica ponderam a média da variável Y_i , onde os pesos $W_h(\cdot)$ dependem da técnica utilizada e da distância entre x e X_i que é ajustada pelo parâmetro de suavização h .

As duas técnicas mais conhecidas para estimar regressões não-paramétricas são: estimação de Nadaraya-Watson e a estimação k-NN. Estas técnicas são descritas a seguir.

Estimador Nadaraya-Watson

Dado observações independentes e identicamente distribuídas das variáveis aleatórias $\{(X_i, Y_i)\}_{i=1, \dots, n}$, o objetivo desta regressão é encontrar a média condicional de Y dado $X=x$, que é representada por:

$$m(x) = E(y / X = x) = \frac{\int y f(x, y) dy}{f(x)}$$

A densidade de x presente no denominador é obtida usando-se a estimação de uma função de densidade *kernel* e a densidade conjunta do denominador é obtida através de uma *kernel* multiplicativa.

$$\begin{aligned} \hat{f}_{h_1, h_2}(X, Y) &= n^{-1} \sum_{i=1}^n K_{h_1}(x - X_i) K_{h_2}(y - Y_i) \\ \int y \hat{f}_{h_1, h_2}(X, Y) dy &= n^{-1} \sum_{i=1}^n K_{h_1}(x - X_i) \int y K_{h_2}(y - Y_i) dy \\ &= n^{-1} \sum_{i=1}^n K_{h_1}(x - X_i) \int \frac{y}{h_2} K\left(\frac{y - Y_i}{h_2}\right) dy \\ &= n^{-1} \sum_{i=1}^n K_{h_1}(x - X_i) \int (sh_2 + Y_i) K(s) ds \\ &= n^{-1} \sum_{i=1}^n K_{h_1}(x - X_i) Y_i \end{aligned}$$

Assim, o estimador de $m(x)$ é dado por:

$$\hat{m}_h(x) = \frac{n^{-1} \sum_{i=1}^n K_h(x - X_i) Y_i}{n^{-1} \sum_{j=1}^n K_h(x - X_j)}$$

Os pesos neste caso serão dados por:

$$W_{h_i}(x) = \frac{h^{-1} K\left(\frac{x - X_i}{h}\right)}{\hat{f}_h(x)}$$

Os pesos dependem de toda a amostra $\{X_j\}_{j=1,\dots,n}$ e as observações Y_i irão obter mais peso nas áreas onde as observações X_i são esparsas.

Como ocorre na estimação de uma função de densidade *kernel*, a janela irá determinar o nível de suavização da estimação, de modo que ao diminuir a janela obtém-se uma estimação menos suave.

Apesar de \hat{m}_h ser um estimador viesado quando h é diferente de zero, ele é um estimador consistente para a curva de regressão $m(x)$ quando h tende para zero e nh tende para infinito².

Estimador “k-Nearest-Neighbor”(k-NN)

A estimação k-NN é definida como uma média ponderada da variável dependente em uma vizinhança. Esta vizinhança é definida através das variáveis X_i que constituem os k mais próximos do ponto x .

O estimador de $m(x)$ neste caso será dado por:

$$\hat{m}_h(x) = n^{-1} \sum_{i=1}^n W_{k_i}(x) Y_i,$$

onde a sequência de pesos $\{W_{k_i}(x)\}_{i=1}^n$ é definida em um conjunto de índices:

$$J_x = \{i/X_i \text{ é uma das } k \text{ observações próximas de } x\}$$

O conjunto destas observações na vizinhança de x define a sequência de pesos k-NN:

$$W_{k_i}(x) = \begin{cases} \frac{n}{k}, & \text{se } i \in J_x \\ 0, & \text{c.c} \end{cases}$$

Nesta estimação k-NN, o parâmetro k regula o grau de suavização da estimação, assim, aumentos em k levam a estimação a se tornar mais suave. Como

² Para maiores informações sobre as estatísticas de uma estimação Nadaraya-Watson ver HÄRDLE W.,1991, p. 53.

ocorria na estimação de funções de densidade *kernel*, o viés é crescente e a variância decrescente no parâmetro de suavização k .

A estimação k -NN pode ser entendida como uma média de *kernel* uniformes, $k(u) = \frac{1}{2}I(|u| \leq 1)$, com uma janela variável $h=R(k)$, que seria a distância entre x e a observação X_i mais próxima de x em dada vizinhança. Em analogia a estimação de Nadaraya-Watson, tem-se que o estimador de k -NN poderá ser representado por:

$$\hat{m}_k(x) = \frac{\sum_{i=1}^n K\left(\frac{x - X_i}{R}\right) Y_i}{\sum_{j=1}^n K\left(\frac{x - X_j}{R}\right)},$$

sendo o peso dado por:

$$W_{k_i}(x) = \frac{K_R(x - X_i)}{\hat{f}_R(x)}$$

Pode-se obter um estimador k -NN que seja equivalente a um estimador usando *kernel*, quando se utiliza a janela que minimiza o erro quadrático médio, ao aplicar a seguinte relação: $k=2nhf(x)$, onde n é o número de observações na amostra. Usando este valor para k , tem-se o mesmo erro quadrático médio assintótico para a estimação k -NN e a suavização por *kernel*. Assim, a regressão usando a suavização por *kernel* é equivalente a regressão usando a suavização pelo estimador k -NN³.

Como ambos os métodos apresentados fornecem estimadores consistentes para $m(x)$ e a amostra utilizada na estimação é relativamente grande, poderia-se empregar ambos para estimar a curva desejada. A partir desta curva, é possível calcular o lucro médio do capital e o lucratividade marginal do capital para as pequenas empresas brasileiras. De fato, as estimações realizadas através do método de suavização por *kernel*, o método de Nadaraya-Watson, e com o método

³ Para uma prova completa de como os dois métodos são equivalentes sob certas condições ver HÄRDLE W., 1991, p. 80.

k-NN apresentam resultados semelhantes⁴. Neste trabalho, apresenta-se apenas os gráficos realizados através do método k-NN.

Para se proceder a estimação da curva que relaciona o lucro do capital ao valor do capital utilizado pelas pequenas empresas brasileiras, foi necessário primeiramente eliminar as firmas que empregam capital nulo e aquelas que empregam um valor muito alto de capital, quando comparada as demais, pois estas causam distorção no resultado. Além disso, estas firmas que não possuem capital despertam pouco interesse para o objetivo deste trabalho, pois o capital pode não ser um insumo da sua função de produção. Assim, os dados utilizados na estimação referem-se a firmas que empregam valor de capital entre R\$1,00 e R\$100.000,00 e que na sua maioria possuem lucro positivo.

Para realizar as estimações através do método k-NN foi necessário também escolher o parâmetro de suavização, denominado de k. Como foi visto, a variância da curva estimada depende deste parâmetro, diminuindo a medida que aumenta-se o valor de k. Para obter uma curva que possa ser interpretada, é necessário escolher um parâmetro de suavização que diminua as flutuações desta curva, permitindo a análise de seu formato. Assim, estima-se a curva que relaciona o lucro do capital ao valor do capital utilizado pelas pequenas empresas brasileiras impondo diferentes valores de k. O parâmetro de suavização variou de 100 a 1000, em intervalos de 50 e acima de 1000. Percebe-se que quando o valor deste parâmetro é menor que 1000 a curva de lucro apresentava muitas flutuações. A partir de k igual a 1000, a curva é suave e mantém determinado formato.

A partir da curva estimada para o lucro do capital em relação ao valor do capital empregado pelas firmas são obtidas as curvas do lucro médio do capital e da lucratividade marginal do capital. A curva de lucro médio foi obtida dividindo-se o valor do lucro estimado pelo seu correspondente valor do capital. A curva da lucratividade marginal foi obtida calculando-se a derivada ao longo da curva estimada. Estas curvas podem indicar a existência de imperfeições no mercado de crédito ou de heterogeneidade produtiva entre as pequenas brasileiras, a partir de sua interpretação, com base no instrumental teórico construído neste trabalho.

⁴ O fato dos resultados com o método K-NN e Nadaraya-Watson serem praticamente os mesmos indica a robustez dos resultados apresentados neste trabalho.