

Dynamiczne Modele Ekonometryczne
X Ogólnopolskie Seminarium Naukowe, 4–6 września 2007 w Toruniu
Katedra Ekonometrii i Statystyki, Uniwersytet Mikołaja Kopernika w Toruniu

Dominik Śliwicki
Uniwersytet Mikołaja Kopernika w Toruniu

Estymacja jądrowa w dynamicznych modelach regresji

1. Wprowadzenie

Modelowanie ekonometryczne w znacznej mierze opiera się na metodach parametrycznych, zakładających określony charakter zależności między badanymi zmiennymi (np. liniowa). Często charakter tych zależności nie jest znany bądź trudny do zidentyfikowania, co powoduje konieczność wprowadzenia do analizy innych metod, które nie wymagają tak restrykcyjnych założeń. Alternatywą metod parametrycznych są metody określane nieparametrycznymi a w szczególności estymatory jądrowe, które zostały zaproponowane na przełomie lat pięćdziesiątych i sześćdziesiątych niezależnie przez Rosenblatta (1956) i Parzena (1962) a ich zasadnicza koncepcja wywodzi się z problemu estymacji funkcji gęstości rozkładu prawdopodobieństwa. Od tamtej pory nastąpił ich dynamiczny rozwój, pojawiło się wiele opracowań na ich temat oraz znaleziono obszary zastosowania w ekonomii, medycynie, biologii, psychologii, sterowaniu optymalnym. Jedynie kwestia uwzględnienia w modelach regresji nieparametrycznej opóźnień czasowych jest stosunkowo młodym obszarem zainteresowania badaczy i niewiele jest w literaturze opracowań poświęconych temu problemowi (Phillips, Park, 1998).

Układ artykułu jest następujący. W części drugiej przedstawiono jądrowe estymatory regresji tj. estymator lokalnie stały Nadaraya-Watsona oraz lokalnie wielomianowy estymator jądrowy regresji. Część trzecia poświęcona jest dynamizacji jądrowych estymatorów regresji. W części czwartej pokazano wyniki eksperymentów symulacyjnych i porównano je z modelem autoregresyjnym. Część piąta zawiera wyniki analizy empirycznej. Artykuł kończy podsumowanie.

2. Jądrowe estymatory funkcji regresji

Niech dwie zmienne losowe X i Y o realizacjach (x_t, y_t) będą powiązane relacją:

$$y_t = m(x_t) + \varepsilon_t, \quad (1)$$

gdzie: ε_t – biały szum o własnościach:

$$E(\varepsilon_t) = 0, \quad (2)$$

$$\text{cov}[m(x_t), \varepsilon_t] = 0, \quad (3)$$

$$\varepsilon_t \sim iid, \quad (4)$$

$$t = 1, 2, \dots, n, \quad (5)$$

$m(\cdot)$ – gładka ale o nieznannej postaci funkcja.

Funkcja $m(x)$ opisuje warunkową wartość oczekiwaną zmiennej losowej Y względem wartości zmiennej losowej X :

$$m(x_t) = E(Y | X = x_t) \quad (6)$$

Podstawowa postać jądrowego estymatora regresji została zaproponowana przez Nadaraya (1964) oraz Watsona (1964):

$$m(x) = \frac{\sum_{t=1}^n k\left(\frac{x-x_t}{h}\right) y_t}{\sum_{t=1}^n k\left(\frac{x-x_t}{h}\right)}, \quad (7)$$

gdzie: h – parametr wygładzania, (inne nazwy to szerokość pasma oraz szerokość okna); ściśle rzecz biorąc h jest pewną dodatnią funkcją zależną od rozmiaru próby n , $h = h(n)$ która zbiega do zera jeżeli rozmiar próby $n \rightarrow \infty$, czyli:

$$\lim_{n \rightarrow \infty} h(n) = 0 \quad (8)$$

$$\text{oraz } \lim_{n \rightarrow \infty} nh(n) = \infty, \quad (9)$$

$k(z)$ – borelowska funkcja określana mianem jądra – jest to funkcja gęstości rozkładu prawdopodobieństwa zmiennej losowej spełniająca warunki:

$$\int k(z) dz = 1 \quad \text{– pole powierzchni pod krzywą równe 1,}$$

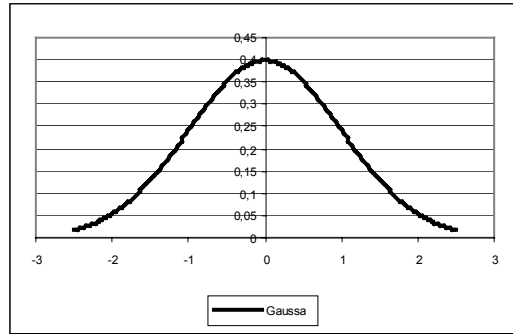
$k(z) = k(-z)$ – funkcja symetryczna względem zera,

$k(0) \geq k(z)$ – funkcja posiadająca w punkcie $z = 0$ słabe maksimum globalne.

Jako funkcję jądrową można zastosować między innymi:

– jądro rozkładu normalnego Wykres 1. Funkcja gęstości rozkładu normalnego

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}. \quad (10)$$



Źródło: opracowanie własne.

Funkcja jądrowa jest funkcją wag. Wagi są większe dla tych wartości y_t , dla których wartości x_t są bliżej x oraz mniejsze dla tych wartości y_t , dla których x_t są dalej od x . Aby zwiększyć wrażliwość na pomiar odległości $x - x_t$, przeskalowuje się jądro używając zmiennej h .

Wybór wartości parametru wygładzania ma zasadnicze znaczenie dla kształtu linii regresji. Istnieje kilka metod wyznaczania parametru wygładzania na podstawie próby. Optymalną wartość można uzyskać poprzez minimalizację wartości oczekiwanej kwadratów reszt:

$$E[\hat{m}(x) - m(x)]^2 = \text{var}(\hat{m}(x)) + [\text{bias}(\hat{m}(x))]^2 \rightarrow \min, \quad (11)$$

gdzie: $\text{var}(\hat{m}(x)) = \frac{\sigma_\varepsilon^2}{2nh} \int k(u)^2 du$ – wariancja,

$\text{bias}(\hat{m}(x)) = h^2 m''(x) \int k(u)u^2 du$ – obciążenie,

σ_ε^2 – wariancja resztowa.

Optymalna wartość $E(\cdot)$ zależy od odpowiedniej kombinacji pomiędzy obciążeniem oraz wariancją.

Silverman (1986) proponuje używanie wzoru:

$$h = 0,9 \min\left(s, \frac{(\hat{q}_3 - \hat{q}_1)}{1,349}\right) n^{-1/5}, \quad (12)$$

gdzie: s – odchylenie standardowe zmiennej x_t , $\hat{q}_3 - \hat{q}_1$ – rozstęp kwartylowy.

Inną metodą jest technika zwana *cross-validation*. Jedną z pierwszych obserwacji (x_j, y_j) jest usuwana, a pozostałe $n-1$ obserwacji są używane do wyznaczenia wygładzenia w punkcie x_j :

$$\hat{m}_j(x_j) = \frac{1}{n-1} \sum_{t \neq j} w_t(x_j) y_t, \quad (13)$$

$$\text{gdzie: } w_t(x) = \frac{k\left(\frac{x_t - x}{h}\right)}{\frac{1}{n} \sum_{t=1}^n k\left(\frac{x_t - x}{h}\right)} \quad (14)$$

otrzymana wartość jest szacunkiem y_j ; następnie powtarza się krok poprzedni dla $j = 1, 2, \dots, n$ oraz określa funkcję postaci:

$$CV(h) = \frac{1}{n} \sum_{j=1}^n [y_j - \hat{m}_j(x_j)]^2 W(x_j), \quad (15)$$

gdzie: $W(\cdot)$ – nieujemna funkcja wag, która może być użyta do obniżenia wag obserwacji granicznych. Spadek wag wyznaczonych dla punktów granicznych jest potrzebny ponieważ, te punkty zwykle mają mniej sąsiednich obserwacji. Funkcja $CV(h)$ jest nazywana *cross – validation function* ponieważ, uaktualnia zdolność wygładzenia do przewidywania $\{y_t\}_{t=1}^n$. Wybiera się szerokość przedziału h , która minimalizuje funkcję (15).

W roku 1977 Stone zaproponował koncepcję lokalnie wielomianowego estymatora jądrowego funkcji regresji. Ogólnie polega ona na dopasowaniu lokalnie, czyli w pewnym punkcie x , wielomianu stopnia p :

$$\beta_0 + \beta_1(\cdot - x) + \dots + \beta_p(\cdot - x)^p, \quad (16)$$

do próby $(x_t; y_t)$, wykorzystując ważoną metodę najmniejszych kwadratów z jądrowymi wagami $\frac{1}{h} k\left(\frac{x_t - x}{h}\right)$. Estymator $m(x)$ wyraża się wzorem:

$$\hat{m}(x; p, h) = \mathbf{e}_1^T (\mathbf{X}_x^T \mathbf{W}_x \mathbf{X}_x)^{-1} \mathbf{X}_x^T \mathbf{W}_x \mathbf{Y}, \quad (17)$$

gdzie: \mathbf{e}_1 – wektor jednostkowy z wartością 1 na pierwszej współrzędnej oraz 0 pozostałych,

\mathbf{W}_x – diagonalna macierz wag postaci:

$$\mathbf{W}_x = \text{diag} \left\{ \frac{1}{h} k\left(\frac{x_1 - x}{h}\right), \dots, \frac{1}{h} k\left(\frac{x_n - x}{h}\right) \right\}, \quad (18)$$

\mathbf{X}_x – macierz postaci,

$$\mathbf{X}_x = \begin{bmatrix} 1 & x_1 - x & \cdots & (x_1 - x)^p \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_n - x & \cdots & (x_n - x)^p \end{bmatrix}. \quad (19)$$

Wyrażenie (17) nazywa się lokalnie wielomianowym estymatorem jądrowym (*local polynomial kernel estimator*). Dla $p = 1$ redukuje się do lokalnie liniowego estymatora jądrowego, a dla $p = 0$ estymatora Nadaraya – Watsona.

Gdy liczba zmiennych objaśniających d jest większa niż jeden, wzór na jądrowy estymator regresji może być przedstawiony w dwóch postaciach:

- gdy do jego konstrukcji zostało zastosowane jądro produktowe:

$$K(\mathbf{z}) = \prod_{i=1}^d k(z_i), \quad (20)$$

$K(\mathbf{z})$ – d wymiarowa funkcja gęstości rozkładu prawdopodobieństwa,

$$m(x_1, x_2, \dots, x_d) = \frac{\sum_{t=1}^n k\left(\frac{x_1 - x_{1t}}{h_1}\right) k\left(\frac{x_2 - x_{2t}}{h_2}\right) \dots k\left(\frac{x_d - x_{dt}}{h_d}\right) y_t}{\sum_{t=1}^n k\left(\frac{x_1 - x_{1t}}{h_1}\right) k\left(\frac{x_2 - x_{2t}}{h_2}\right) \dots k\left(\frac{x_d - x_{dt}}{h_d}\right)}. \quad (21)$$

- gdy do jego konstrukcji zostało zastosowane jądro radialne:

$$K(\mathbf{z}) = c_{k,d} k\{(\mathbf{z}^T \mathbf{z})^{1/2}\}, \quad (22)$$

$$c_{k,d}^{-1} = \int k\{(\mathbf{z}^T \mathbf{z})^{1/2}\} d\mathbf{z}, \quad (23)$$

$$m(x_1, x_2, \dots, x_d) = \frac{\sum_{t=1}^n k \sqrt{\left(\frac{x_1 - x_{1t}}{h_1}\right)^2 + \left(\frac{x_2 - x_{2t}}{h_2}\right)^2 + \dots + \left(\frac{x_d - x_{dt}}{h_d}\right)^2} y_t}{\sum_{t=1}^n k \sqrt{\left(\frac{x_1 - x_{1t}}{h_1}\right)^2 + \left(\frac{x_2 - x_{2t}}{h_2}\right)^2 + \dots + \left(\frac{x_d - x_{dt}}{h_d}\right)^2}}. \quad (24)$$

W zapisie macierzowym, uogólnienie na przypadek regresji wielowymiarowej jest analogiczne do lokalnie liniowego estymatora jądrowego dla przypadku jednowymiarowego, postaci:

$$\hat{m}(\mathbf{x}; \mathbf{1}, \mathbf{H}) = \mathbf{e}_1^T (\mathbf{X}_x^T \mathbf{W}_x \mathbf{X}_x)^{-1} \mathbf{X}_x^T \mathbf{W}_x \mathbf{Y}, \quad (25)$$

$$\text{gdzie: } \mathbf{X}_x = \begin{bmatrix} 1 & (\mathbf{X}_1 - \mathbf{x})^T \\ \vdots & \vdots \\ 1 & (\mathbf{X}_n - \mathbf{x})^T \end{bmatrix}, \quad (26)$$

$$\mathbf{W}_x = \text{diag}\{K_H(\mathbf{X}_1 - \mathbf{x}), \dots, K_H(\mathbf{X}_n - \mathbf{x})\}, \quad (27)$$

$\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$ – d – wymiarowe zmienne losowe,

$$\mathbf{X}_i = (X_{i1}, X_{i2}, \dots, X_{id})^T, \quad (28)$$

$$\mathbf{x} = (x_1, x_2, \dots, x_d)^T, \quad (29)$$

$$\mathbf{H} \text{ – symetryczna, } d \times d \text{ macierz parametrów wygładzania,}$$

$$K_H(\mathbf{z}) = |\mathbf{H}|^{-1/2} K(\mathbf{H}^{-1/2}\mathbf{z}) \text{ – wagi.} \quad (30)$$

3. Dynamizacja jądrowych estymatorów regresji

Dynamizacja jądrowego estymatora regresji Nadaraya-Watsona polega na podstawieniu: $x_t = y_{t-l}$:

$$m(x) = \frac{\sum_{t=2}^n k\left(\frac{x - y_{t-1}}{h}\right) y_t}{\sum_{t=2}^n k\left(\frac{x - y_{t-1}}{h}\right)}. \quad (31)$$

W przypadku estymatora lokalnie wielomianowego, dynamizacja polega na dopasowaniu lokalnie do próby $(y_{t-1}; y_t)$, wielomianu stopnia p postaci:

$$\beta_0 + \beta_1(y_{t-1} - x) + \dots + \beta_p(y_{t-1} - x)^p. \quad (32)$$

Gdy rozważany jest jądrowy estymator regresji wielowymiarowej, możliwe jest uwzględnienie wyższych rzędów opóźnień. Podstawiając $x_{jt} = y_{t-j}$ gdzie $j = 1, 2, \dots, d$ otrzymuje się jądrowy estymator regresji wielowymiarowej, w którym w roli zmiennych objaśniających są opóźnione wartości zmiennej objaśnianej. Wykorzystując jądro produktowe można wyrazić go wzorem:

$$m(x_1, x_2, \dots, x_d) = \frac{\sum_{t=d+1}^n k\left(\frac{x_1 - y_{t-1}}{h_1}\right) k\left(\frac{x_2 - y_{t-2}}{h_2}\right) \dots k\left(\frac{x_d - y_{t-d}}{h_d}\right) y_t}{\sum_{t=d+1}^n k\left(\frac{x_1 - y_{t-1}}{h_1}\right) k\left(\frac{x_2 - y_{t-2}}{h_2}\right) \dots k\left(\frac{x_d - y_{t-d}}{h_d}\right)}. \quad (33)$$

4. Analiza symulacyjna

Wygenerowano szeregi po 1000 obserwacji według schematu:

$$y_t = \alpha y_{t-1} + \varepsilon_t, \quad (34)$$

gdzie: $\alpha \in \{0.2; 0.4; 0.5; 0.6; 0.8; 1\}$ oraz $E(\varepsilon_t^2) \in \{1; 10\}$.

Na podstawie wygenerowanych szeregów szacowano regresje jądrowe estymatorem lokalnie stałym, lokalnie liniowym oraz lokalnie wielomianowym stopnia p . Następnie wyznaczono średnie kwadratowe błędy szacunku estymatorów. Tabela 1. zawiera średnie kwadratowe błędy szacunku wyznaczone na podstawie wyników zastosowania estymatorów jądrowych do szacowania regresji, w której $\varepsilon_t \sim N(0; 1)$, a w tabeli 2. $\varepsilon_t \sim N(0; 10)$. Ostatnie wiersze tabel 1. i 2. zawierają średnie błędy kwadratowe modeli AR(1).

Tabela 1. Wartości średnich błędów kwadratowych dla regresji, w której $\varepsilon_t \sim N(0;1)$

p	$\alpha = 0.2$	$\alpha = 0.4$	$\alpha = 0.5$	$\alpha = 0.6$	$\alpha = 0.8$	$\alpha = 1$
0	1.019430	1.021337	1.022436	1.024786	1.046594	1.633939
1	1.019430	1.021337	1.022436	1.024786	1.046594	1.633939
2	1.015097	1.014333	1.014255	1.013261	1.019355	1.016671
3	1.014818	1.014045	1.013430	1.011739	1.018472	1.016636
4	1.013551	1.013074	1.012525	1.011457	1.016665	1.016754
5	1.013823	1.012499	1.011395	1.010105	1.014810	1.016238
6	1.013425	1.012584	1.011004	1.010927	1.015060	1.014149
7	1.012580	1.012567	1.010389	1.009317	1.017462	1.014555
8	1.012997	1.010824	1.010747	1.010444	1.016813	1.014007
s_{AR}	1.021764	1.021734	1.021724	1.021721	1.021796	1.021829

Źródło: obliczenia własne.

Tabela 2. Wartości średnich błędów kwadratowych dla regresji, w której $\varepsilon_t \sim N(0;10)$

p	$\alpha = 0.2$	$\alpha = 0.4$	$\alpha = 0.5$	$\alpha = 0.6$	$\alpha = 0.8$	$\alpha = 1$
0	3.213288	3.219307	3.224679	3.231701	3.271823	4.454634
1	3.213288	3.219307	3.224679	3.231701	3.271823	4.454634
2	3.202904	3.196437	3.201372	3.204280	3.203793	3.202363
3	3.197338	3.190224	3.196610	3.201291	3.200701	3.197956
4	3.187706	3.183941	3.191048	3.199351	3.198205	3.193514
5	3.182985	3.179918	3.190115	3.198300	3.198696	3.194257
6	3.177119	3.178370	3.190773	3.198794	3.201155	3.192298
7	3.176087	3.185525	3.195887	3.198639	3.199071	3.200006
8	3.173461	3.190524	3.197036	3.200109	3.202831	3.199586
s_{AR}	3.221227	3.221260	3.221307	3.221335	3.221320	3.220417

Źródło: obliczenia własne.

Przyjmując za kryterium porównań średni błąd kwadratowy, wyniki analizy symulacyjnej wskazują na lepsze dopasowanie regresji jądrowej do danych empirycznych w porównaniu z modelami AR(1). W tabelach wyróżniono minimalne wartości średnich błędów kwadratowych – w każdym przypadku są one niższe od średnich błędów kwadratowych modeli autoregresyjnych.

5. Analiza empiryczna

Analizie empirycznej został poddany szereg czasowy stóp zwrotu z cen akcji spółki Swarzędz z okresu 07. 09. 1999 – 22. 05. 2007. Analiza struktury pozwoliła stwierdzić, iż jest to proces AR(1). Oszacowany model ma postać:

$$r_t = 0.0684r_{t-1} + u_t. \quad (35)$$

Tabela 3. zawiera wartości średnich błędów kwadratowych regresji jądrowych oraz modelu AR(1). p – stopień wielomianu określonego wzorem (32).

Tabela 3. Wartości średnich błędów kwadratowych regresji jądrowych oraz modelu (35)

p	0	1	2	3	4
s(u)	0.039069	0.039069	0.038847	0.038786	0.038789
p	5	6	7	8	AR(1)
s(u)	0.038726	0.038699	0.038687	0.038693	0.039249

Źródło: obliczenia własne.

Wyniki analizy wskazują na lepsze dopasowanie regresji jądrowej do danych empirycznych w porównaniu z modelem AR(1) ponieważ, wartości średniego błędu kwadratowego regresji jądrowych są mniejsze od błędu modelu AR(1).

6. Zakończenie

W artykule zaprezentowano koncepcję jądrowych estymatorów regresji w ujęciu statycznym i dynamicznym. Na podstawie pokazanych przykładów symulacyjnych oraz analizy empirycznej można stwierdzić, iż jądrowe estymatory regresji w wersji dynamicznej dobrze dopasowują się do próby, dając mniejsze wartości średniego błędu kwadratowego w porównaniu z modelem parametrycznym.

Literatura

- Nadaraya, E. A. (1964), On Estimating Regression, *Theory Probab. Appl.*, 10, 186–190.
- Parzen, E. (1962), On Estimation of a Probability Density Function and Mode, *The Annals of Mathematical Statistics*, 33, 1065–1076.
- Phillips, P. C., Park, J. Y. (1998), *Nonstationary Density Estimation and Kernel Autoregression*, Cowles Foundation for Research in Economics, Yale University, Economics Division, Seoul National University.
- Silverman, B. W. (1986), *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London.
- Stone, C. J. (1977), Consistent Nonparametric Regression, *The Annals of Mathematical Statistics*, 12, 1285 – 1297.
- Wand, M. P., Jones, M. C. (1995), *Kernel Smoothing*, Chapman and Hall, London.
- Watson, G. S. (1964), Smooth Regression Analysis, *Sankhya Ser. A*, 21, 101–116.