Volume: 09 Number: 02 Category: 05 Page: 20-28 ISSN No: 1303-6319 Cilt: 09 Sayı: 02 Kategori: 05

İLİŞKİLİ HATALARA SAHİP MODELLERDE SPLAYN VE ÇEKİRDEK REGRESYON KESTİRİCİLERİNİN PERFORMANSLARI

Serdar DEMİR* **Dursun AYDIN****

ÖZET

Parametrik olmayan regresyon kestiricilerinin performanslarının, ilişkili hatalara sahip modellerde ciddi biçimde azaldığı iyi bilinmektedir. Bu çalışmada, regresyon fonksiyonunun parametrik olmayan kestirimi için sıkça kullanılan kübik düzleştime splaynı ile Nadaraya-Watson çekirdek kestiricilerinin hatalı ilişkilere sahip modellerdeki performansları incelenmiştir. Bunun için bir benzetim çalışması gerçekleştirilerek, elde edilen kestirimlerin hata kareler ortalamaları karşılaştırılmıştır. Benzetim sonuçlarına göre, Nadaraya-Watson çekirdek kestiricisi küçük örneklemlerde iyi performans gösteriyorken, splayn kestiricisi büyük örneklemlerde iyi performans göstermektedir.

Anahtar Kelimeler: Çekirdek kestiricisi, İlişkili hatalar, Kübik splayn, Nadaraya-Watson.

1. GİRİŞ

Parametrik olmayan regresyon kestiricileri, bağımsız değişkene ilişkin gözlem değerleri x_i ($x_1<x_2<\cdots<x_n$) ve bağımlı değişkene ilişkin gözlem değerleri y_i olmak üzere,

$$Y_i = m(x_i) + \varepsilon_i \quad , (i = 1,2,\ldots,n) \tag{1}$$

biçimindeki bir regresyon modelinin bilinmeyen $m(x)$ fonksiyonunun kestirimi ile ilgilenmektedirler. Burada, ε_i hatalarının sıfır ortalamalı ve $\sigma^2\mathbf{V}$ varyans-kovaryans matrisi ile ilişkili hatalar oldukları varsayılmaktadır. **V** matrisi bir ağırlıklandırma matrisidir.

Parametrik olmayan regresyon kestirimleri için yaygın biçimde kullanılan kestiriciler, çekirdek (kernel) regresyon ve splayn (spline) düzleştirme kestiricilerdir (Hardle, 1991) (Eubank, 1999), (Green vd., 1994). Özellikle zaman serileri çözümlemelerinde sıkça karşılaşılan hataların ilişkili olması durumu, çekirdek ve splayn tipi kestiricilerin performanslarını ciddi biçimde etkilemektedir. Diggle ve Hutchinson (1989); Hurvich ve Zeger (1990); Khon vd. (1992); Wang (1998); Krivobokova ve Kauermann (2007); Liu (2009) ve Morton vd. (2009) ilişkili hataların varlığında splayn düzleştirme kestiricileri üzerinde çalışmalar yapmışlardır. Altman (1990); Chu ve Marron (1991); Hart (1991); Herrmann (1992); Rio (1996); Ray (1997); Opsomer vd. (2001); Carroll vd. (2002); Park vd. (2006); Kim vd. (2009); De Brabanter vd. (2010) ve Lee vd. (2010) hataların ilişkili olduğu durumlarda çekirdek tipi kestiricileri ve bant genişliği seçim yöntemlerini incelemişlerdir. Bu çalışmalar, genel olarak eşit aralıklı sabit tasarımlı model (örneğin, $x_i = i/n$) varsayımı yapılarak gerçekleştirilmiştir. İlişkili hatalara sahip modellerde, çekirdek ve splayn kestiricilerine ilişkin söz konusu bu

*Yrd. Doç. Dr., Muğla Sıtkı Kocaman Üniversitesi, Fen Fakültesi, İstatistik Bölümü, e-posta: serdardemir@mu.edu.tr

**Doç. Dr., Muğla Sıtkı Kocaman Üniversitesi, Fen Fakültesi, İstatistik Bölümü, e-posta: duaydin@mu.edu.tr

çalışmalarda kestiricilerin özellikleri kuramsal ve deneysel olarak ayrı ayrı incelenmekte, ancak doğrudan performansları karşılaştırılmamaktadır. Bunun nedeni, kuramsal karmaşıklık ve benzetimlerdeki işlem zorluklarıdır.

Bu çalışmada, rastgele tasarımlı model varsayımı altında Nadaraya-Watson çekirdek regresyon kestiricisi ile kübik splayn düzleştirme kestiricisinin performanslarının karşılaştırılması amaçlanmıştır. Kestiricilerin performansları, bir benzetim çalışması ile elde edilen verilere ilişkin kestirimlerin hata kareler ortalamalarına dayalı olarak karşılaştırılmaktadır. İncelenen kestiricilere ilişkin bilgiler bir sonraki bölümde sunulmaktadır. Üçüncü bölümde, kestiricilerin performanslarını karşılaştırmak amacıyla yapılan benzetim çalışmasının ayrıntıları ve sonuçları verilmektedir.

2. NADARAYA-WATSON ÇEKİRDEK VE KÜBİK SPLAYN KESTİRİCİLERİ

Çalışmanın bu kısmında, parametrik olmayan regresyon uygulamalarında en sık kullanılan Nadaraya-Watson çekirdek kestiricisi ve kübik splayn kestiricisi tanıtılmaktadır.

2.1 Nadaraya-Watson Çekirdek Regresyon Kestiricisi

Parametrik olmayan regresyon fonksiyonu kestirimi uygulamalarında sıkça kullanılan bir kestirici, Formül (2) ile verilen

$$\hat{m}_K(x) = \sum_{i=1}^{n} Y_i K[(x-X_i)/h] \Big/ \sum_{i=1}^{n} K[(x-X_i)/h] \tag{2}$$

Nadaraya-Watson (NW) çekirdek kestiricisidir (Nadaraya,1964); (Watson,1964) Düzleştirme parametresi olan h sabiti, kestirimin düzleştirme düzeyini kontrol etmekte ve bant genişliği olarak adlandırılmaktadır. Kestiricinin performansında önemli katkısı olan bant genişliğinin seçimi için çeşitli yöntemler mevcuttur. Başlıca seçim yöntemleri çapraz-geçerlilik, genelleştirilmiş çapraz-geçerlilik, ceza fonksiyonları, plug-in ve bootstrap yöntemleridir (Pagan vd.,1999). Hangi yöntemin daha iyi olduğu konusu tartışmalı bir konudur. Bu çalışmada, splayn kestiriciler için de sıkça kullanılan genelleştirilmiş çapraz-geçerlilik yöntemi ele alınmaktadır. Genelleştirilmiş çapraz-geçerlilik (GÇG) yöntemi ile,

$$GÇG(h) = \frac{\sum_{i=1}^{n}\left[Y_i - \hat{m}_K(X_i)\right]^2}{\left[1 - K(0) \Big/ \sum_{i=1}^{n} K\left(\frac{x_i - x_j}{h}\right)\right]^2} \tag{3}$$

biçimindeki fonksiyonu minimum yapan h bant genişliği optimal bant genişliği değeri olarak seçilir (Hardle,1991).

Formül 2'de, K(.) çekirdek fonksiyonu simetrik bir yoğunluk fonksiyonudur. Uygulamada sıkça kullanılan çekirdek fonksiyonları, formül 4 ile verilen Gaussian çekirdek fonksiyonu ve (5) formülü ile verilen Epanechnikov çekirdek fonksiyonudur.

$$K(u)=\exp(-u^2/2)/\sqrt{2\pi}\,, \qquad -\infty < u < \infty \tag{4}$$

$$K(u)=3(1-u^2)/4\,, \qquad |u| \le 1 \tag{5}$$

2.2 Kübik Splayn Düzleştirme Kestiricisi

Parametrik olmayan regresyon modelinde hataların ilişkili olması durumunda regresyon fonksiyonu $m(x)$'in kestirimi için Diggle ve Hutchinson (1989) tarafından kübik splayn düzleştirmeye dayalı bir yaklaşım önerilmiştir. Bilinen bir $\rho > 0$ parametresi için, $\mathbf{V} = [v_{ij}]$ iken, $v_{ij} = exp(-\rho\,|x_i - x_j|)$ otokorelasyon fonksiyonunu kullanan, eşit aralıklı tasarım noktaları (sabit tasarım) gerektirmeyen bu yöntemin splayn düzleştirmeye dayalı çözümü,

$$S[m(x)] = (\mathbf{Y}-\mathbf{M})^T\,\mathbf{V}^{-1}(\mathbf{Y}-\mathbf{M}) + \lambda\mathbf{M}^T\mathbf{CM} \tag{6}$$

biçimindeki cezalı kareler toplamı (penalized sum of squares) ölçütünün $m(x)$'e göre minimum yapılmasını amaçlamaktadır. Formül 6'nın sağ tarafındaki ilk terim verilerin uyumunu (goodness-of-fit), ikinci terim ise düzleştirmeyi (smoothness) tanımlayan ifadelerdir. Sabit bir değer olan λ ise, verilerin uyumu ile düzleştirme arasındaki dengeyi sağlamaktadır. Formül 6 ile bulunan $m(x)$ kestirimi, splayn düzleştirme kestirimi $\hat{m}_S(x)$ olarak tanımlanır. Eşitlikte, $\mathbf{Y}=\begin{bmatrix} y_1 & \cdots & y_n \end{bmatrix}^T$ ve $\mathbf{M} = \begin{bmatrix} m(x_1) & \cdots & m(x_n) \end{bmatrix}^T$ biçimindedir. $\mathbf{C}$ matrisi yarı-pozitif tanımlı bir ceza matrisidir ve $\mathbf{C} = \mathbf{QR}^{-1}\mathbf{Q}^T$ biçiminde elde edilir. Burada, $\mathbf{Q}$ matrisi elemanları, $h_i = x_{i+1} - x_i$ $(i = 1, 2,\ldots,n-1)$ olmak üzere, $i = 2, 3,\ldots,n$ ve $j = 2, 3,\ldots,n-1$ için

$$q_{ij} = \begin{cases} 1/h_{j-1}, & i = j-1 \text{ ise} \\ -1/h_{j-1} - 1/h_j, & i = j \text{ ise} \\ 0, & |i-j| \ge 2 \text{ ise} \\ 1/h_j, & i = j+1 \text{ ise} \end{cases}$$

biçiminde tanımlanan n×(n-2) boyutlu bir matristir. R matrisi elemanları, i=2,3,…,n ve j=2,3,…,n-1 için

$$r_{ij} = \begin{cases} h_{j-1}/6, & i = j-1 \text{ ise} \\ (-h_{j-1} - h_j)/3, & i = j \text{ ise} \\ 0, & |i-j| \ge 2 \text{ ise} \\ h_j/6 & i = j+1 \text{ ise} \end{cases}$$

biçiminde tanımlanan $(n-2)\times(n-2)$ boyutlu bir matristir (Green vd.,1994); (Eubank,1999). Diggle ve Hutchinson (1989), Formül 6'da $a_i = exp\{-\rho\,(x_{i+1} - x_i)\}$ alarak, $\mathbf{V}^{-1}$ yerine simetrik $\mathbf{W}$ bant matrisini kullanmıştır. $\mathbf{W}=[w_{ij}]$ matrisinin sıfırdan farklı olan elemanları aşağıdaki formüller yardımıyla hesaplanmaktadır (Diggle vd., 1989):

$$w_{11} = 1/(1-a_1^2)\,, \qquad i = 2,\ldots,n-1.$$

$$w_{ii} = 1 + a_{i-1}/(1-a_{i-1}^2) + a_i^2/(1-a_i^2)\,, \qquad i = 2,\ldots,n-1.$$

$$w_{nn} = 1/(1-a_n^2)\,, \qquad i = 1,\ldots,n.$$

$$w_{i+1,1} = w_{i,i+1} = -a/(1-a_1^2)\,, \qquad i = 1,\ldots,n.$$

Yukarıda verilen tanımlamalar ışığında, 6 formülünü minimum yapan kübik splayn düzleştirme kestiricisi

$$\hat{\mathbf{m}}_S = \begin{bmatrix} \hat{m}_S(X_1) \\ \vdots \\ \hat{m}_S(X_n) \end{bmatrix} = (\mathbf{W}^{-1} + \lambda \mathbf{C})^{-1} \mathbf{W}^{-1} \mathbf{Y} = \mathbf{S}_\lambda \mathbf{Y} \tag{7}$$

biçimindedir. Burada $\mathbf{S}_\lambda$ matrisi şapka ya da düzleştirme matrisi olarak adlandırılmaktadır.

Formül 6'daki düzleştirme parametresi λ bir denge unsuru olarak, kübik splayn kestiricisinin performansında önemli rol oynamaktadır. λ'nın küçük seçilmesi fazlasıyla pürüzlü bir tahmine, büyük seçilmesi ise fazlasıyla düz bir tahmine yol açabilmektedir. λ'nın belirlenmesine ilişkin literatürde pek çok yöntem (çapraz geçerlilik, genelleştirilmiş çapraz geçerlilik, Akaike bilgi ölçütü, Mallows'un Cp ölçütü, Plug-in vd.) mevcuttur. Kübik splayn kestiricisi için λ parametresinin seçiminde en çok tercih edilen yöntem, genelleştirilmiş çapraz-geçerlilik (GÇG) yöntemidir (Eubank,1999). Bu yöntem, hem etkin olması hem de özellikle kolay uygulanabilir olması nedeniyle ilgi çekmektedir. İlişkili hatalara sahip modeller için, Diggle ve Hutchinson (1989) tarafından uyarlanan,

$$GÇG(\lambda) = \left[\mathbf{Y} - \hat{\mathbf{m}}_S\right]^T \mathbf{W} \left[\mathbf{Y} - \hat{\mathbf{m}}_S\right] \Big/ \left\{ \mathrm{tr}\left[\mathbf{I} - \mathbf{S}_\lambda\right] \right\}^2 \tag{8}$$

biçimindeki genelleştirilmiş çapraz-geçerlilik fonksiyonu kullanılmaktadır. Bu fonksiyonu minimum yapan λ değeri, optimal düzleştirme parametresi değeri olarak seçilmektedir.

3. BENZETİM ÇALIŞMASI

İlişkili hatalara sahip 1 formülü ile verilen regresyon modellerinde bilinmeyen $m(x)$ fonksiyonun kestiriminde, Nadaraya-Watson çekirdek regresyon kestiricisi $(\widehat{m}_K)$ ile kübik splayn düzleştirme kestiricisinin) performanslarını karşılaştırmak için bir benzetim çalışması yapılmıştır. Bu benzetim çalışması için (Chu,1991); (Rio, 1996); (Goijier, 2002); (Kim, 2009); (Brabanter, 2010) ve (Lee, 2010) tarafından yapılan benzetim çalışmaları dikkate alınarak, $m(x)=x^3(1-x^3)$ regresyon fonksiyonu kullanılmış ve

$$y_i = x_i^3(1 - x_i^3) + \varepsilon_i \tag{9}$$

modeline uygun olarak veriler üretilmiştir. Burada, x_i'ler [0,1] aralığında tekbiçimli (uniform) dağılımdan rastgele çekilmiştir. Modelin ε_i hataları, $\varepsilon_0 \sim N(0,\ 0.0001/(1-\rho^2))$olmak üzere,

$$\varepsilon_i = \rho \varepsilon_{i-1} + u_i \quad ,\ i=1, 2, 3, \ldots, n$$

biçimindeki birinci dereceden otoregresif ($AR(1)$) bir süreçten üretilmiştir. Sürecin u_i hataları, sıfır ortalamalı ve 0.0001 varyanslı bir normal dağılımdan rastgele çekilmiştir.

Otokorelasyon katsayısı $\rho = 0.30, 0.60, 0.90$ değerlerinin herbiri için $n = 25, 50, 100$ büyüklüğünde rastgele örneklemler üretilmiştir. Benzetimler için tekrar sayısı $r = 500$ olarak alınmıştır.

Nadaraya-Watson kestirimleri için Gaussian çekirdek fonksiyonu kullanılmıştır. Bant genişliği h Formül 3 ile verilen genelleştirilmiş çapraz-geçerlilik yöntemiyle belirlenmiştir. Splayn kestirimleri için düzleştirme parametresi λ ise Formül 8 ile verilen genelleştirilmiş çapraz-geçerlilik yöntemiyle belirlenmiştir.

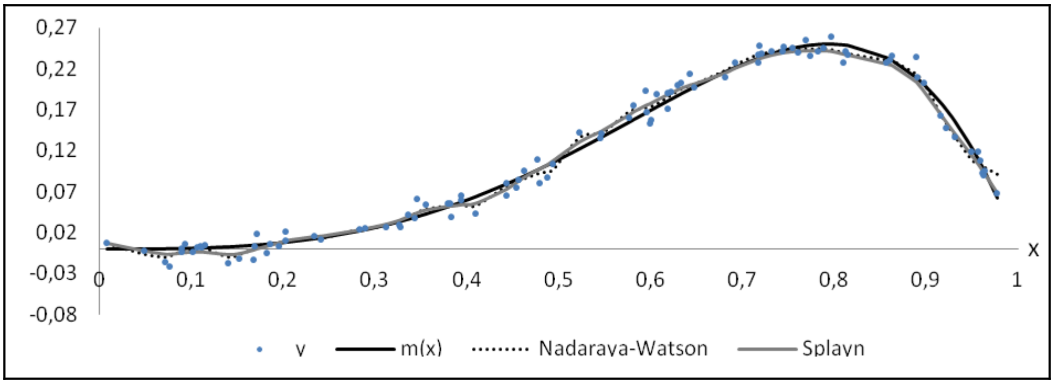
Her bir örneklem genişliği (25, 50, 100) ve her bir otokorelasyon katsayısı (0.30, 0.60, 0.90) için üretilen verilere ilişkin Nadaraya-Watson çekirdek kestirimleri ve splayn düzleştirme kestirimleri hesaplanmıştır. Daha sonra her bir kestirimin hata kareler ortalaması bulunmuştur. Bu işlem 500 kez tekrarlanmış ve 500 kestirim üzerinden Ortalama Hata Kareler Ortalamaları (OHKO= $\sum_{i=1}^{r} HKO_i / r$) bulunmuştur. Her bir duruma ilişkin kestirimlerin OHKO değerleri Tablo 1'de verilmektedir.

Tablo 1'den, ρ'nun her üç değeri için n=25 örneklem büyüklüğünde Nadaraya-Watson kestirimlerinin daha küçük hata kareler ortalamalarına sahip oldukları görülmektedir. Buna karşın, n=50 ve n=100 örneklem genişliklerinde ise splayn kestirimleri daha küçük hata kareler ortalamalarına sahip olmaktadırlar. Örneklem genişliği arttıkça, splayn kestiricisinin Nadaraya-Watson kestiricisine göre daha iyi performans gösterdiği söylenebilir. Ayrıca beklenildiği gibi, hataların ilişki düzeyleri arttıkça tüm örneklem büyüklüklerinde hem Nadaraya-Watson hem de splayn kestirimlerin hata kareler ortalamalarının büyüdüğü görülmektedir.

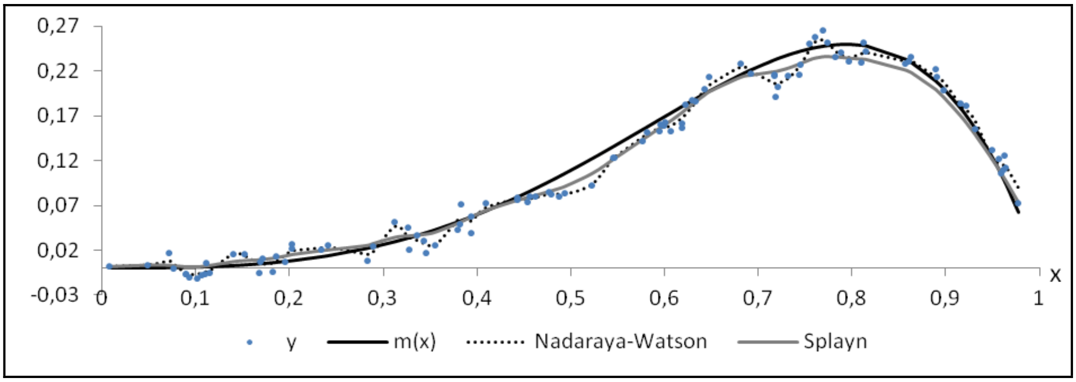
Tablo 1. Kestirimlerin ortalama hata kareler ortalamaları

ρ	n	OHKO($\widehat{m}_K$)	OHKO($\widehat{m}_S$)	OHKO($\widehat{m}_S$)/OHKO($\widehat{m}_K$)
	25	0.0000927	0.0001107	1.194681
0.30	50	0.0000684	0.0000507	0.741656
	100	0.0000499	0.0000366	0.732469
	25	0.0001611	0.0001756	1.089935
0.60	50	0.0001231	0.0001047	0.850226
	100	0.0001032	0.0000834	0.808648
	25	0.0007215	0.0007377	1.022485
0.90	50	0.0006614	0.0006062	0.916553
	100	0.0005553	0.0004495	0.809438

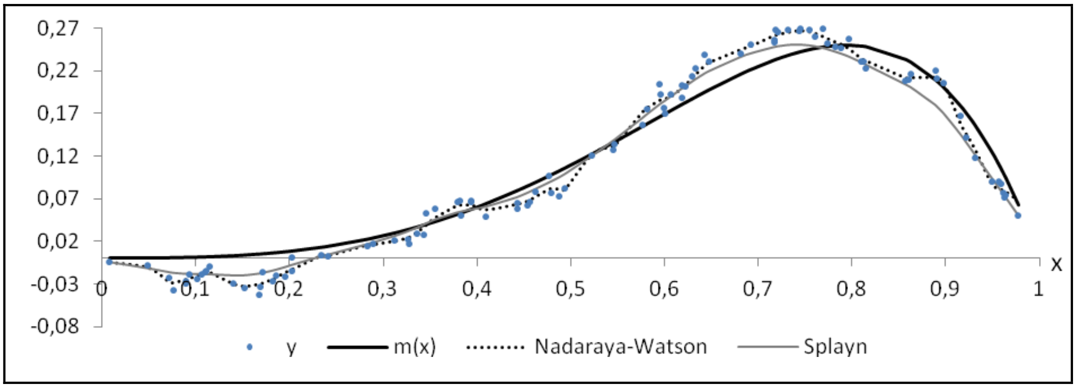
Rastgele seçilen n=100 büyüklüğündeki üç benzetim örneklemi için $m(x)$ fonksiyonunun Nadaraya-Watson çekirdek ve splayn düzleştirme kestirimlerinin grafikleri Şekil 1, Şekil 2 ve Şekil 3'de verilmektedir.



Şekil 1. ρ**=0.30 ve** n**=100 için splayn ve Nadaraya-Watson kestirimleri**



Şekil 2. ρ**=0.60 ve** n**=100 için splayn ve Nadaraya-Watson kestirimleri**



Şekil 3. ρ**=0.90 ve** n**=100 için splayn ve Nadaraya-Watson kestirimleri**

Şekil 1, Şekil 2 ve Şekil 3'ten, hatalar arasındaki ilişki arttıkça Nadaraya-Watson çekirdek ve splayn düzleştirme kestirimlerinin kötüleştiği görülmektedir. Bununla beraber, splayn kestirimlerinin Nadaraya-Watson kestirimlerine göre daha pürüzsüz eğriler ortaya çıkardığı söylenebilir.

4. SONUÇ VE TARTIŞMA

Bu çalışmada, hataları ilişkili olan rastgele tasarım modellerinde Nadaraya-Watson çekirdek kestiricisi ile kübik splayn düzleştirme kestiricisinin performansları, hata kareler ortalaması ölçütüne dayalı olarak karşılaştırılmıştır. Bu amaçla gerçekleştirilen benzetim sonuçlarına göre, küçük örneklemlerde Nadaraya-Watson çekirdek kestiricisinin, büyük örneklemlerde ise splayn kestiricisinin daha iyi performans gösterdiği söylenebilir. Ayrıca, beklenildiği gibi hatalardaki ilişki düzeyi arttıkça iki kestiricinin de performanslarının kötüleştiği görülmüştür. Bununla beraber, hataların ilişki düzeyinin artmasından, küçük örneklemlerde Nadaraya-Watson kestiricisi daha fazla etkilenirken, büyük örneklemlerde ise splayn kestiricisi daha fazla etkilenmektedir. Sonuç olarak, ilişkili hatalara sahip regresyon modellerinde, küçük örneklem durumunda Nadaraya-Watson çekirdek kestiricisinin kullanılması, büyük örneklem durumunda ise kübik splayn kestiricinin kullanılması önerilmektedir. Gelecekteki çalışmalarda, farklı bant genişliği ya da düzleştirme parametresi seçim yöntemleri bakımından karşılaştırılmalar yapılması planlanmaktadır.

5. KAYNAKLAR

Altman, N. S., 1990. Kernel Smoothing of Data with Correlated Errors, JASA, 85, 749-759.

Carroll, R. J., Linton, O., Mammen, E., Xiao, Z., 2002. More Efficient Kernel Estimation in Nonparametric Regression with Autocorrelated Errors, STICERD-Econometrics Paper Series, 2002/435.

Chu, C. K., Marron, J. S. 1991. Comparison of Two Bandwidth Selectors with Dependent Errors, Ann. Statist., 19, 1906-1918.

De Brabanter, K., De Brabanter, J., Suykens, J., De Moor, B., 2010. Kernel Regression with Correlated Errors. Proc. of the 11th International Symposium on Computer Applications in Biotechnology (CAB), Leuven, Belgium, Jul. 2010, 13-18.

De Gooijer, J. G., Gannoun, A., Larramendy, I., 2002. Nonparametric Regression with Serially Correlated Errors. Ann. Ins. Stat. Univ. Paris, 46, Fasc. 1-2.

Diggle, P. J., Hutchinson, M. F., 1989. On Spline Smoothing With Autocorrelated Errors. Australian Journal of Statistics. 31. 166-182.

Eubank, R., 1999. Nonparametric Regression and Spline Smoothing, New York, Dekker.

Green, D. J., Silverman, B. W., 1994. Nonparametric Regression and Generalized Linear Models, London, Chapman & Hall.

Hardle, W., 1991. Applied Nonparametric Regression, Cambridge University Press, Cambridge.

Hart, J. D., 1991. Kernel Regression Estimation with Time Series Errors, J. R. Statist. Soc. B, 53, 173-187.

Hermann, E., Gasser, T., Kneip, A., 1992. Choice of Bandwidth for Kernel Regression When Residuals are Correlated, Biometrika, 79, 783-795.

Hurvich, C. M., Zeger, S. L., 1990. A Frequency Domain Selection Criterion for Regression with Autocorrelated Errors, J. Am. Statist. Ass., 85, 705-714.

Kim, T. Y., Park, B. U., Moon, M. S., Kim, C., 2009. Using Bimodal Kernel Inference in Nonparametric Regression with Correlated Errors, J. Multivariate Analysis, 100, 1487-1497.

Kohn, R., Ansley, C. G., Wong, C. M., 1992. Nonparametric Spline Regression with Autoregressive Moving Average Errors, Biometrika, 79, 335-346.

Krivobokova, T., Kauermann, G., 2007. A Note on Penalized Spline Smoothing with Correlated Errors. Journal of the American Statistical Association, 102 (480), 1328-1337.

Lee, Y. K., Mamnen, E., Park, B. U., 2010. Bandwidth Selection for Kernel Regression with Correlated Errors, 44, 4, 327-340.

Liu Jun, M., 2009. Nonlinear Time Series Modeling Using Spline-Based Nonparametric Models, AMATH'09 Proceedings of the 15th American Conference on Applied Mathematics.

Morton, R., Kang, E. L., Henderson, B., 2009. Smoothing Splines for Trend Estimation and Prediction in Time Series, Environmetrics, 20, 249-259.

Nadaraya, E. A., 1964. On Estimating Regression, Theory Pb. Appl., Vol.10, 186-190.

Opsomer, J., Wang, Y., Yang, Y., 2001. Nonparametric Regression with Correlated Errors, Statist. Sci. 16, 134–153.

Pagan, A., Ullah A., 1999. Nonparametric Econometrics, Cambridge, Cambridge University Press.

Park, B. U., Lee, Y. K., Kim, T. Y., Park, C., 2006. A Simple Estimator of Error Correlation in Nonparametric Regression Models, Scand. J. Statist, 33, 451-462.

Ray, B. K., Tsay, R. S., 1997. Bandwidth Selection for Kernel Regression with Long-range Dependent Errors, Biometrika, 84(4), 791-802.

Rio, A. Q., 1996. Comparison of Bandwidth Selectors in Nonparametric Regression Under Dependence, Comput. Statist. Data Anal., 21, 563–580.

Wang, Y., 1998. Smoothing Spline Models with Correlated Random Errors, JASA, 93, 34–348.

Watson, G. S., 1964. Smooth Regression Analysis, Sankhya, Series A, 26, 359-372.

THE PERFORMANCES OF SPLINE AND KERNEL REGRESSION ESTIMATORS IN MODELS WITH CORRELATED ERRORS

ABSTRACT

It is well known that the performances of nonparametric regression estimators severely decrease in the models with correlated errors. In this paper, the performances of the cubic smoothing spline and the Nadaraya-Watson kernel estimators, which are often used for nonparametric estimation of regression function, are investigated in the models with correlated errors. For this purpose, a simulation study is performed and the mean square errors of the estimators are compared. The simulation results show that the Nadaraya-Watson kernel estimator performs well in small samples, while spline estimator performs well in large samples.

Keywords: Kernel estimator, Correlated errors, Cubic spline, Nadaraya-Watson.