

OTOREGRESİF MODELLERİN BAYES ANALİZİNİN HAVA KİRLİLİĞİ VERİLERİNE UYGULAMASI

Mehmet Ali CENGİZ*

Erol EĞRİOĞLU

ÖZET

Otoregresif (AR) modellerin istatistiksel analizi önemli bir çıkarım problemidir. Çoğu klasik yaklaşımlar, AR modelinin derecesinin belirlenmesinde ve parametrelerin tahmininde geniş olarak kullanılır. Bayesci yaklaşım da her iki amaç içinde kullanılabilir. Bu çalışmada AR modelinin derecesinin belirlenmesinde klasik yaklaşımlar kullanıldı ve Samsun bölgesindeki hava kirliliği verisi için gelecek değerleri tahmin etmede bilgi içermeyen önsellerin kullanımıyla Bayesci yaklaşım incelendi.

Anahtar Kelimeler: *Otoregresif Model, Bayesci Analiz*

1. GİRİŞ

Zaman serileri analizinde son yıllarda klasik çıkarım yöntemleri ile birlikte Bayes çıkarımı kullanılmaktadır. ARMA modellerinin Bayes analizi üzerine Zellner (1971), Box ve Jenkins (1976), Monahan (1983), Broemling ve Shaaraway (1988), Marriott ve Smith (1992), McCulloch ve Tsay (1994) ve Chib ve Greenberg (1994) çalışmaları mevcuttur. Bu çalışmalardan Zellner (1971), Box ve Jenkins (1976) ve Broemling ve Shaaraway (1988) de durağanlık ve tersine çevrilebilirlik şartlarının parametre uzayını kısıtlaması dikkate alınmamış diğerlerinde ise nümerik integrasyon teknikleri ile bu husus dikkate alınarak Bayes analizi gerçekleştirilmiştir. Bu çalışmalarda ayırıcı bir hususda kullanılan önsel dağılım içindir. Zellner (1971), Box ve Jenkins (1976) ve Broemling ve Shaaraway (1988) bilgi içermeyen önsel dağılımı kullanmış, diğerlerinde ise bilgi içeren önsel dağılım kullanılmıştır. Zellner (1971)'de birinci ve ikinci dereceden otoregresif modelin Bayes analizini önsel dağılımı uniform (düzgün) dağılım olarak ve yaklaşık olabilirlik fonksiyonunu kullanarak gerçekleştirmiştir. Box ve Jenkins (1976) çalışmasında ise Jeffreys'in önselini ve tam olabilirlik fonksiyonunu kullanarak tüm ARMA modellerinin Bayes analizini gerçekleştirmiştir. Monahan (1983) de ARMA modellerinin Bayes analizinde bilgi içermeyen önsellerin kullanılmasının gereksizliğini savunmuş ve modeldeki parametre sayısına bağlı bir bilgi içeren önsel seçerek ve nümerik integrasyon tekniklerini kullanarak analizi gerçekleştirmiştir. McCulloch ve Tsay (1994) ve Chib ve Greenberg (1994) ise nümerik integrasyon tekniği olarak Gibbs örneklemesini kullanarak ARMA modellerinin Bayes analizini gerçekleştirmiştir.

Bu çalışmada, ilk olarak Otoregresif modellerin Bayes Analizi verilmiştir. Farklı önseller için sonsal dağılımlar bulunmuştur. Elde edilen sonsal dağılımlar kükürtdioksit verileri için hesaplanmıştır.

2. OTOREGRESİF MODELLERİN BAYES ANALİZİ

Z_t , $EZ_t = 0$, $V(Z_t) = \sigma_z^2$ olan bir pür stokastik süreç olsun. Bu durumda p. dereceden bir otoregresif süreç

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + Z_t \quad (1)$$

dir. AR(p) modeli için olabilirlik fonksiyonunu $\sigma_z^2 V_p^{-1}$: $X' = (X_1, X_2, \dots, X_p)$ lerin varyans-kovaryans matrisi ve $\underline{\phi}^T = (\phi_1, \dots, \phi_p)$ olmak üzere

$$L(\underline{\phi} / X_n) = (2\pi\sigma_z^2)^{-n/2} |V_p|^{1/2} \exp\left[-\frac{1}{2\sigma_z^2} \tilde{Q}(\underline{\phi})\right] \quad (2)$$

dir. Burada $\tilde{Q}(\underline{\phi})$ ifadesi koşulsuz kareler toplamıdır ve V_{ij} 'ler V_p 'nin elemanları olmak üzere aşağıdaki formdadır.

$$\tilde{Q}(\underline{\phi}) = \sum_{i=1}^p \sum_{j=1}^p V_{ij} X_i X_j + Q(\underline{\phi})$$

Bayes analizinin aşağıdaki temel formülünü esas alarak üç önsel dağılıma göre sonsal dağılımları elde edebiliriz.

$$\text{Sonsal dağılım} \propto \text{Önsel dağılım} \times \text{Olabilirlik fonksiyonu} \quad (3)$$

Bilgi içermeyen önsel dağılımımız Jeffreys'in önseli ise $\underline{\phi}$ ve σ_z parametreleri için

$$P(\underline{\phi}, \sigma_z) \propto |I(\underline{\phi})|^{1/2} \cdot \sigma_z^{-1} \quad (4)$$

şeklinde elde edilir. (2) ve (4)'ün kullanımıyla

$$P_j(\underline{\phi}, \sigma_z / X) \propto \sigma_z^{-(n+1)} |I(\underline{\phi})|^{1/2} |V_p|^{1/2} \exp\left\{-\frac{\tilde{Q}(\underline{\phi})}{2\sigma_z^2}\right\} \quad (5)$$

elde ederiz. (5) ifadesini σ_z 'ye göre integrallersek $\underline{\phi}$ 'nin marjinal sonsal dağılımı bulunur.

$$P_j(\underline{\phi} / X) \propto |I(\underline{\phi})|^{1/2} |V_p|^{1/2} \{\tilde{Q}(\underline{\phi})\}^{-n/2}$$

Burada $|I(\underline{\phi})| = nV_p^{-1}$ sonucunu göz önüne alırsak $|I(\underline{\phi})|^{1/2} |V_p|^{1/2}$ ifadesini ihmal ederek otoregresif süreç için sonsal dağılımı yeniden yazabiliriz.

$$P_j(\underline{\phi} / X) \propto \{\tilde{Q}(\underline{\phi})\}^{-n/2} \quad (6)$$

Otoregresif sürecin şartsız kareler toplamı

$$\tilde{Q}(\underline{\phi}) = v s_z^2 + (\underline{\phi} - \hat{\underline{\phi}})' D_p (\underline{\phi} - \hat{\underline{\phi}})$$

dir. Burada $v = n - p$ ve $s_z^2 = \tilde{Q}(\underline{\phi}) / v$ dir. Ayrıca D_p matrisi elemanları, aşağıdaki eşitlikle hesaplanan simetrik kareler toplamlarından oluşur.

$$D_{ij} = D_{ji} = X_i X_j + X_{i+1} X_{j+1} + \dots + X_{n+1-j} X_{n+1-i}$$

Bu bilgiler ışığında D_p matrisi

$$D_p = \begin{bmatrix} D_{2,2} & \cdot & \cdot & \cdot & D_{2,p+1} \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ D_{2,p+1} & \cdot & \cdot & \cdot & D_{p+1,p+1} \end{bmatrix}$$

şeklindedir. Böylece sonsal dağılım

$$P(\underline{\phi} / X) \propto \left(1 + \frac{(\underline{\phi} - \hat{\underline{\phi}})' D_p (\underline{\phi} - \hat{\underline{\phi}})}{v S_z^2} \right)^{-n/2} \quad (7)$$

olur. Bu sonsal dağılım ise (n-p) serbestlik dereceli çok değişkenli Student-t dağılımı olduğu gözükmemektedir. (7) sonsal dağılımının özel bir formunu birinci dereceden otoregresif süreç için elde edersek aşağıdaki sonuçlara ulaşırız.

$$\frac{\phi_1 - \hat{\phi}_1}{S_{\hat{\phi}_1}} \sim t_{n-1}, \quad \hat{\phi}_1 = \frac{D_{12}}{D_{22}}, \quad S_{\hat{\phi}_1} = \left[\frac{1}{(n-1) D_{22}} \left\{ 1 - \frac{D_{12}^2}{D_{11} D_{22}} \right\} \right]^{1/2}$$

Ayrıca otoregresif süreç için $1 - \alpha$ H.P.D bölgesini tıpkı lineer modellerin Bayes analizinde olduğu gibi

$$(\underline{\phi} - \hat{\underline{\phi}})' D_p (\underline{\phi} - \hat{\underline{\phi}}) < p s_z^2 F_{\alpha, p, v}$$

dir. Jeffreys'in önseline göre elde edilen sonsal dağılım Student-t dağılımı olması nedeniyle gelecek tahminleri klasik analiz yöntemlerindeki tekniklerle elde edilebilir.

Şimdi de bilgi içermeyen önsel dağılımımız Improper önsel dağılım olsun. Improper önsel dağılım otoregresif süreç için aşağıdaki gibidir.

$$P(\underline{\phi}, \sigma_z) \propto |I(\underline{\phi})| \sigma_z^{-1} \quad (8)$$

Bu önsel dağılım ile elde edilen marjinal sonsal dağılım ise aşağıdaki gibidir.

$$P_I(\underline{\phi} / X) \propto |I(\underline{\phi})| |V_p|^{1/2} \left\{ \tilde{Q}(\underline{\phi}) \right\}^{-n/2} \quad (9)$$

Improper önsel dağılım ile elde edilen sonsal dağılım Student -t dağılımlı olmadığından gelecek tahminleri için nümerik integrasyon yöntemlerine ihtiyaç vardır. Bu durumda Jeffreys'in önselini tercih etmeye bir nedendir.

Son olarak da önsel dağılımımız uniform önsel olsun. Otoregresif süreç için uniform önsel dağılım aşağıdaki gibi olur.

$$P(\underline{\phi}, \sigma_z) \propto \sigma_z^{-1} \quad (10)$$

Uniform önsel dağılım ile elde edilen sonsal dağılım ise aşağıdaki gibi olur.

$$P_U(\underline{\phi} / X) \propto |V_p|^{1/2} \left\{ \tilde{Q}(\underline{\phi}) \right\}^{-n/2} \quad (11)$$

Uniform önsel dağılım ile elde edilen sonsal dağılımında Student-t dağılımlı olmaması gelecek tahminlerinin nümerik integrasyon teknikleri ile hesaplanmasına sebep olmaktadır. Dolayısıyla (11) sonsal dağılımını kullanmakta pratik değildir.

3. SAMSUN KÜKÜRTDİOKSİT VERİLERİNİN ANALİZİ

Devlet İstatistik Enstitüsünün Samsun ilinde il ve ilçe merkezlerine göre yaptığı ölçümlerden elde edilen havadaki kükürtdioksit ortalamaları Tablo.1'de ve bu verilerin plot grafiğini Şekil-1 de verildi. Seriyi durağan hale getirdikten sonra otokorelasyon ve kısmi otokorelasyon katsayıları elde edilerek grafikleri Şekil-2 ve Şekil-3 de verildi. Bu şekiller incelendi ve otokorelasyon katsayılarının grafiğinde 1,9 ve 12. gecikmelerde önemlilik ve kısmi otokorelasyon katsayılarının grafiğinde sadece birinci gecikmede önemlilik görüldüğünden AR(1) modelinin uygun olduğu sonucuna varıldı. Parametre tahmin sonuçları ve uyum iyiliği istatistikleri Tablo.2 de verildi. Bu sonuçlara göre Turbo Pascal dilinde yazdığımız programdan gelecek değerler elde edilerek Tablo.3 de , uyum tahminleri ile birlikte grafiği ise Şekil-8 de verildi.

Birinci dereceden otoregresif model için (6), (9) ve (11) sonsal dağılımları aşağıdaki gibidir.

$$P_J(\phi_1 / X) \propto \{\tilde{Q}(\phi_1)\}^{-n/2}$$

$$P_I(\phi_1 / X) \propto (1 - \phi_1^2)^{-1/2} \{\tilde{Q}(\phi_1)\}^{-n/2}$$

$$P_U(\phi_1 / X) \propto (1 - \phi_1^2)^{1/2} \{\tilde{Q}(\phi_1)\}^{-n/2}$$

Bu üç sonsal dağılım için gerekli hesaplamalar yapılarak grafikleri Şekil-4, Şekil-5, Şekil-6 ve Şekil-7 de verildi. Üç sonsal dağılım incelendiğinde hepsini maksimum değerini $\hat{\phi} = 0,3586$ noktasında aldığı görülmektedir.

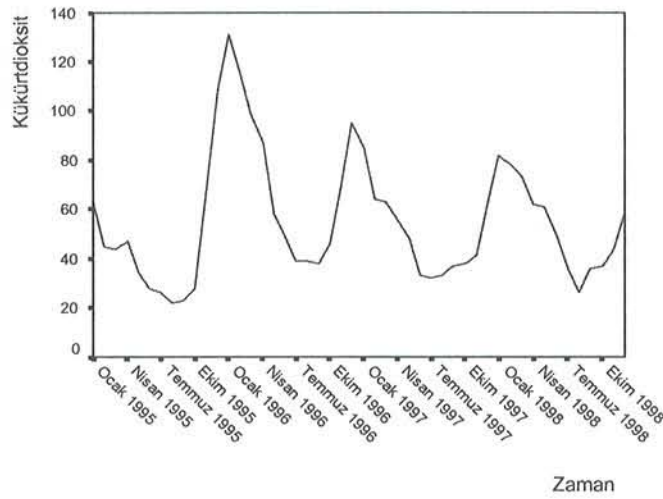
Jeffreys'in önseli ile elde edilen sonsal dağılımdan AR(1) modeli için %95 H.P.D bölgesi ise aşağıdaki gibi olur.

$$(\phi - 0.3586)(113950)(\phi - 0.3586) < 35289$$

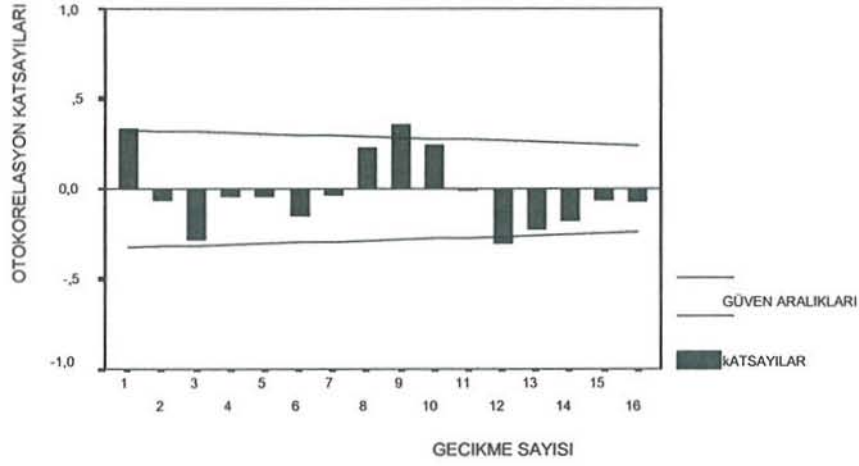
Burada $s_z^2 = 141.1561952$ dir. Herhangi bir parametre değerinin H.P.D bölgesi içinde olması için bu eşitsizliği sağlaması gerekmektedir. Örneğin 0.5 parametre değeri H.P.D. bölgesinde iken -0.5 parametre değeri H.P.D. bölgesinde değildir.

Tablo 1. 1995-1998 Yılları Samsun Kükürtdioksit Ortalamaları

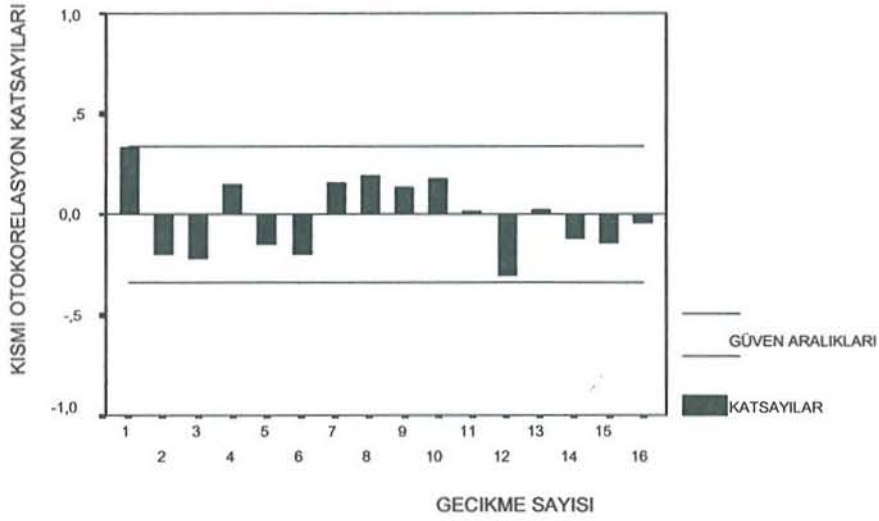
YILLAR	AYLAR	SO ₂	YILLAR	AYLAR	SO ₂
1995	1	63	1997	1	85
1995	2	45	1997	2	64
1995	3	44	1997	3	63
1995	4	47	1997	4	56
1995	5	34	1997	5	48
1995	6	28	1997	6	33
1995	7	26	1997	7	32
1995	8	22	1997	8	33
1995	9	23	1997	9	37
1995	10	28	1997	10	38
1995	11	65	1997	11	41
1995	12	109	1997	12	62
1996	1	131	1998	1	82
1996	2	116	1998	2	78
1996	3	99	1998	3	73
1996	4	87	1998	4	62
1996	5	58	1998	5	61
1996	6	49	1998	6	49
1996	7	39	1998	7	36
1996	8	39	1998	8	26
1996	9	38	1998	9	36
1996	10	46	1998	10	37
1996	11	69	1998	11	44
1996	12	95	1998	12	59



Şekil-1 1995-1998 Yılları Arasında Samsunda Yapılan Ölçümlerde Elde Edilen Kükürtdioksit Ortalamalarının Grafiği



Şekil-2 Durağanlaştırılmış Serinin Otokorelasyon Katsayılarının Grafiği



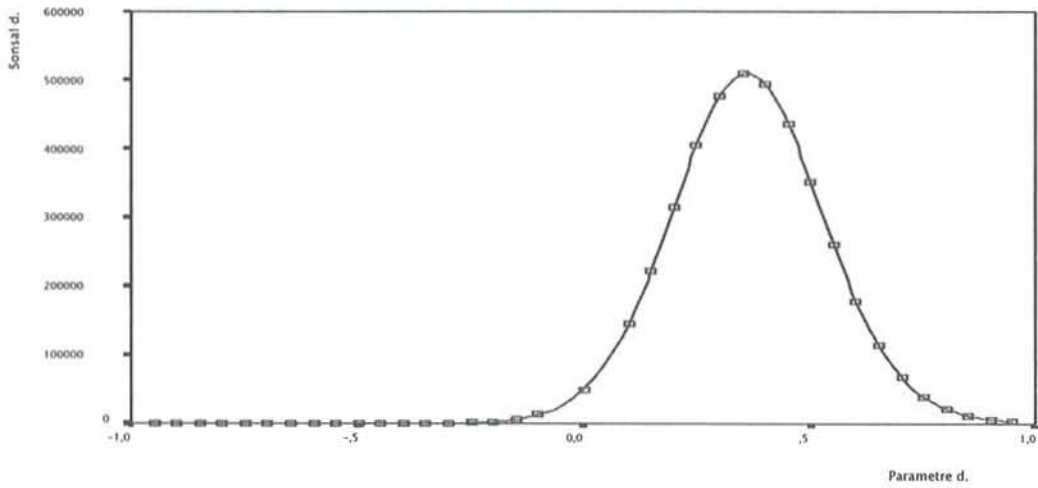
Şekil-3 Durağanlaştırılmış Serinin Kısmi Otokorelasyon Katsayılarının Grafiği

Tablo.2 Parametre Tahmin Sonuçları

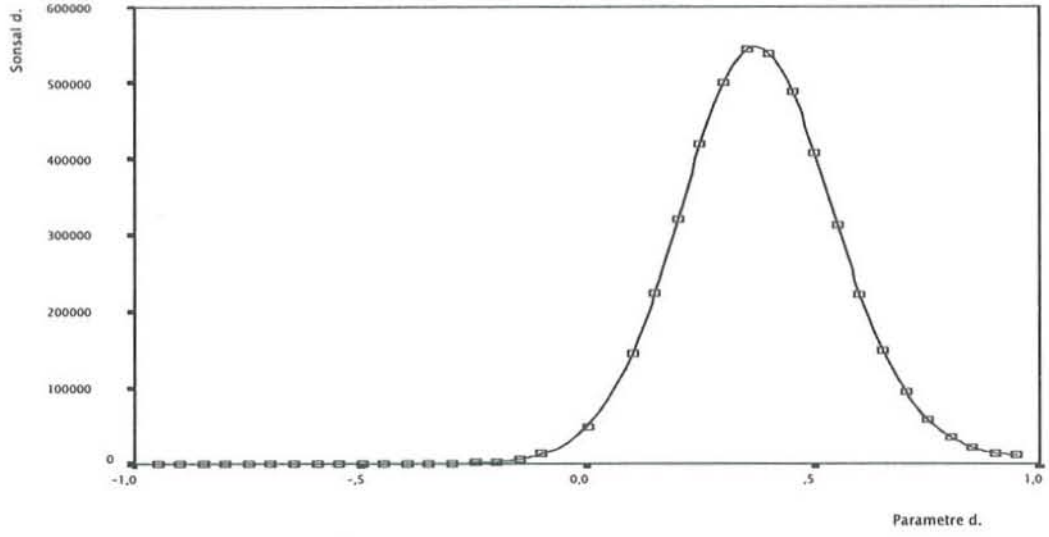
İstatistikler	Değerler
$\hat{\phi}_1$	0,3586
AIC (Akaike bilgi kriteri)	98,435
BIC (Bayesgil bilgi kriteri)	98,116
Q (Portmanteau uyum eksikliği testi)	14,42

Tablo.3 Gelecek Tahminleri

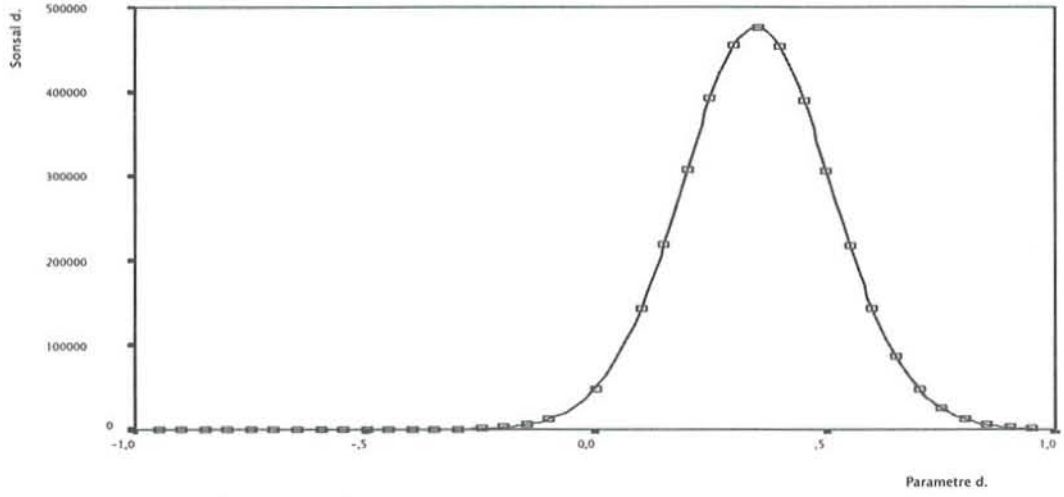
YILLAR	AYLAR	GELECEK TAHMİNLERİ	YILLAR	AYLAR	GELECEK TAHMİNLERİ
1999	1	76	2000	11	31
1999	2	72	2000	12	46
1999	3	66	2001	1	64
1999	4	55	2001	2	59
1999	5	54	2001	3	54
1999	6	42	2001	4	42
1999	7	29	2001	5	41
1999	8	19	2001	6	29
1999	9	29	2001	7	16
1999	10	30	2001	8	6
1999	11	37	2001	9	16
1999	12	52	2001	10	17
2000	1	70	2001	11	24
2000	2	65	2001	12	39
2000	3	60	2002	1	58
2000	4	49	2002	2	53
2000	5	48	2002	3	48
2000	6	36	2002	4	37
2000	7	23	2002	5	36
2000	8	13	2002	6	24
2000	9	23	2002	7	11
2000	10	24			



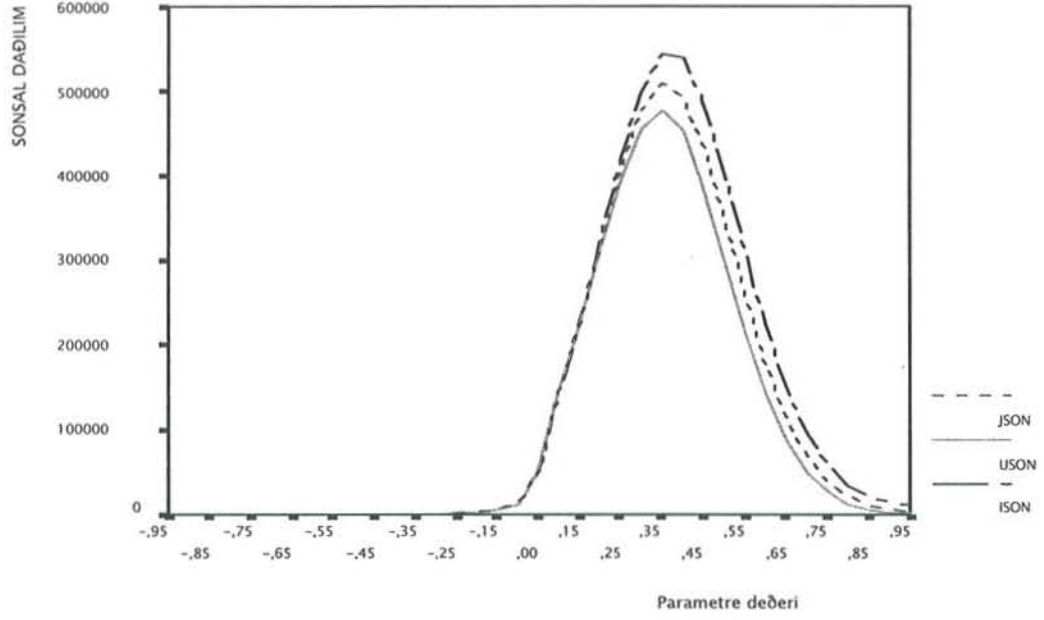
Şekil.4 Jeffreys'in Önsel dağılımı ile elde edilen sonsal dağılım



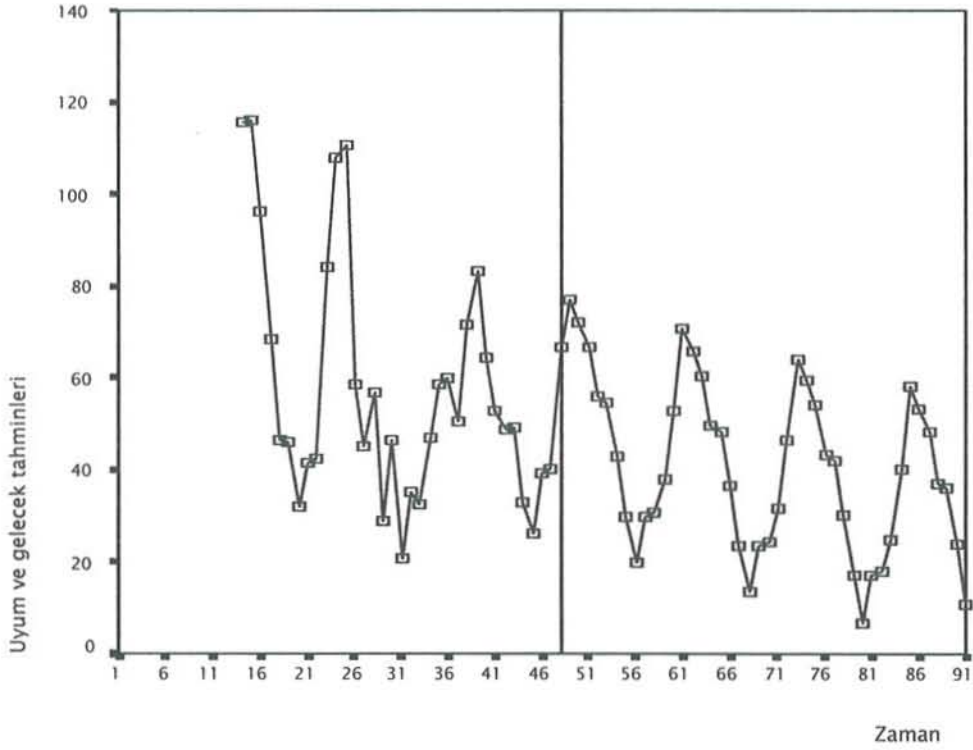
Şekil-5 Uniform Önsel dağılımı ile elde edilen sonsal dağılım



Şekil-6 İmproper Önsel dağılımı ile elde edilen sonsal dağılım



Şekil-7 Üç Sonsal Dağılımın Birlikte Grafiği



Şekil-8 Uyum Ve Gelecek Tahminlerinin Grafiği
(İlk 48 gözlem uyum tahmini, diğerleri ise gelecek tahminidir.)

4. SONUÇ VE ÖNERİLER

Yaptığımız çalışmada AR modellerinin Bayes analizi konusunda Box ve Jenkinsin çalışmaları genişletilerek Samsun da havadaki kükürtdioksit miktarlarını tahmin etmek için kullanıldı. Çalışma sonunda Samsun da havada bulunan kükürtdioksit miktarları ortalamalarının gelecek tahminleri elde edildi ve hava kirliliğinin gelecekte Samsun için ciddi bir tehdit olmayacağı öngörüldü.

Box ve Jenkins Bayes analizinde önsel dağılımı bilgi içermeyen Jeffreys'in önsel dağılımı olarak seçtiğini söylemiştik. Bizde İmproper önsel ve uniform önseli de analize kattık ve üç önsel dağılım için elde edilen üç sonsal dağılımın grafiksel incelemesini gerçekleştirdik. Üç sonsal dağılımında tepe değerini aynı parametre değeri için aldığı görüldü. Bu değer de eğrisel tahmin algoritmasından bulunan parametre değeri ile aynıdır. Fakat Box ve Jenkins sonsal dağılımın bir kısmını ihmal ederek Student-t dağılımı olduğunu göstermiştir. Dolayısıyla tahmin edici dağılımı kullanarak elde edilecek tahminler standart hesaplamalardaki tahminlere denktir. Ancak diğer iki sonsal dağılımın Student-t dağılımlı olmaması gelecek tahminlerinin elde edilmesinde tahmin edici dağılımın nümerik integrasyonlarla çözümlenmesi veya örnekleme yöntemlerinin kullanılmasını gerektirmektedir. Bu durumda Jeffreys'in önselinin kullanılmasının hesaplamalar yönünden kolaylık sağlayacağını açıkça göstermektedir. Fakat bu durum diğer iki sonsal dağılımdan elde edilecek sonuçların kötü olacağı manasına gelmemektedir.

Bu çalışmada Bayes çıkarımı parametre tahmininde ve gelecek değerlerin kestirimin de kullanıldı. Bu amaçla işlem kolaylığı ve gösterim amacıyla informatif olmayan önseller kullanıldı. İformatif yani bilgiye dayalı önseller kullanımıyla daha duyarlı analizler yapılabilir. Ayrıca model tespit aşamasında da Bayes çıkarımı kullanılabilir. Bunlar gelecekteki çalışmalarımızda bize ışık tutacak ve bizi yönlendirecek noktalardır.

KAYNAKLAR

BOX, G.E.P. AND JENKINS G.M. (1976), Time series analysis, forecasting and control, Holden Day.

BROEMLING, L. D. AND SHAARAWAY, S. , (1988), Time series: A Bayesian analysis in the time domain . In Bayesian analysis of time series and dynamic Models, Dekker , New York

CHIB, S. AND GREENBERG, E. (1994), Bayes inference in regressions models with ARMA (p,q) errors, Journal of Econometrics , 64 , 183-206.

MARRIOTT, J. M. AND SMITH , A. F. M. (1992), Reparametrization aspects of numerical Bayesian methods for autoregressive moving average models , Journal of Time Series Analysis , 13, 327-343 .

MCCULLOCH , R.E. AND TSAY, R. S. (1994), Bayesian analysis of autoregressive time series via the Gibbs sampler , Journal Of Time Series Analysis, 15,235-250.

MONAHAN , J. F. (1983), Fully Bayesian Analysis of ARMA time series models, Journal of Econometrics , 21, 307-331.

ZELLNER, A. (1971), An introduction to Bayesian inference in econometrics, John-Wiley& Sons, Inc.

BAYESIAN ANALYSIS OF AUTOREGRESSIVE MODELS WITH AN APPLICATION TO AIR POLLUTION DATA

ABSTRACT

Statistical analysis of autoregressive (AR) models is an important inference problem. Most statistical approaches are widely used in determining the order of the AR model and prediction the parameters. The Bayesian approach can be used for both aim as well. In this study we use the classical approach in determining the order of the AR model and investigate the Bayesian approach with different non-informative priors in predicting future values of the air pollution data for Samsun region.

Key Words: Autoregressive Model, Bayesian Analysis