

T.C.
DOKUZ EYLÜL ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
EKONOMETRİ ANABİLİM DALI
DOKTORA TEZİ

**VERİ DAĞILIMININ EN YAKIN BULANIK
GÖSTERİMİNE DAYALI ZAMAN SERİSİ
ETİKETLENDİRMESİ**

Sinem PEKER

Danışman
Prof. Dr. Efendi NASİBOĞLU

2010

Yemin Metni

Doktora Tezi olarak sunduđum **Veri Dađılımının En Yakın Bulanık Gsterimine Dayalı Zaman Serisi Etiketlendirmesi**” adlı alıřmanın, tarafımdan, bilimsel ahlak ve geleneklere aykırı dřecek bir yardıma bařvurmaksızın yazıldıđını ve yararlandıđım eserlerin kaynakada gsterilenlerden oluřtuđunu, bunlara atıf yapılarak yararlanılmıř olduđunu belirtir ve bunu onurumla dođrularım.

Tarih

30/07/2010

Sinem PEKER

İmza

ÖZET

Doktora Tezi

Veri Dağılımının En Yakın Bulanık Gösterimine Dayalı Zaman Serisi
Etiketlendirmesi

Sinem PEKER

Dokuz Eylül Üniversitesi
Sosyal Bilimler Enstitüsü
Ekonometri Anabilim Dalı

Bu çalışmada üç yeni zaman serileri etiketlendirme yöntemi geliştirildi. Yöntemlerde, danışmansız öğrenme yöntemlerinden biri olan bulanık c-ortalamalar yöntemi zaman serilerinde uygulandı. Her bir küme, merkez değerinin büyüklüğüne göre etiketlendi. Gözlemler küme merkezlerine olan uzaklıklarına göre atandı ve atandığı kümenin etiket değerini aldı. Böylece gözlemlere ait zaman serilerinden, etiketlere ait zaman serileri çıkartılmış olundu. Sonraki adımda, daha düzgün etiket eğrileri elde edebilmek için, K -en yakın komşuluk kuralı etiketlere uygulandı. Klasik yöntemden farklı olarak, komşuluklar bulunurken uzaklıklar değil, verinin zaman bazında kendinden önce ve sonra gelen etiketleri dikkate alındı. Önerilen yöntemlerin etkinliği beyin aktiviteleri ile ilgili olan bispektral endeks veri setlerinde araştırıldı ve K -en yakın komşuluk kuralının zaman bazında çalıştırılmasının ortalama sınıflama kesinliğinde bir artışa yol açtığı kanıtlandı.

Çalışmanın ikinci kısmında, üyelik fonksiyonlarının sınıflama kesinliklerinde artırıcı rolü dikkate alınarak, veri dağılımının en yakın bulanık gösterimleri ile ilgili olan dört yeni teorem geliştirildi. İlgili teoremlerin oluşturulmasında iki yaklaşım kullanıldı. İlk yaklaşımda, veri dağılımının beş noktası parametrik üçgen ve yamuk üyelik fonksiyonlarının beş noktası ile eşleştirildi. İkinci yaklaşımda frekans tabloları kullanıldı. Frekans tablolarındaki normalleştirilmiş yüzdelerle değerleri ile sınıf aralıkların orta noktaları dikkate alınarak, amaç fonksiyonları kuruldu. Minimisasyon problemi yardımıyla, verilerin histogramı ile uyumlu parametrik üçgen ve üssel üyelik fonksiyonları elde edildi. Önerilen teoremlerin, sınıflama kesinliğinde artırıcı etkiye sahip olup olmadığını görebilmek için bispektral endeks veri setlerinde sınıflama işlemi yapıldı. Bulunan sınıflama kesinlikleri, literatürde kullanılmış olan başka bir üyelik fonksiyonu yoluyla elde edilenler ile karşılaştırıldı. Veri setlerinin analizi sonucunda, bu tezde geliştirilen üyelik fonksiyon yaklaşımlarının ortalama sınıflama kesinliğinde artırıcı bir etkiye sahip oldukları kanıtlandı.

Anahtar Kelimeler: Zaman Serisi, Kümeleme, Sınıflama, Veri Dağılımı, Üyelik Fonksiyonu, Bispektral Endeks.

ABSTRACT

Doctoral Thesis

Time Series Labeling Based on Nearest Fuzzy Representation of Data

Distribution

Sinem PEKER

**Dokuz Eylül University
Institute of Social Sciences
Department of Econometrics**

In this study, three new time series labeling methods have been generated. Fuzzy c-means clustering which is an unsupervised learning method have been used in these methods. Each cluster has been labeled regard to its center magnitude. Observations have been assigned to the clusters with respect to their distances to the cluster's centers. Hence, the time series of labels have been extracted from time series of observations. As a next step, K -nearest neighbor rule have been performed on time series of labels to obtain smoother curve of labels. As difference from classical method, instead of distance, the previous and next labels of data have been considered in the determination of neighbors. The efficiency of offered methods has been tested on bispectral index data sets which are related with brain activity and it has been proved that the application of KNN rule on time domain satisfies an increasing on average of classification accuracies.

Considering increasing effect of membership function on classification accuracies, four new theorems about nearest fuzzy representations of data distributions have been offered, as second part of this study. Two perspectives have been followed in the constructions of theorems. In the first perspective, five points of data distribution have been matched with five points of parametric triangular and trapezoidal type membership functions. In the second perspective, frequency tables have been used. The objective functions have been constructed considering the normalized percentages and midpoints of frequency tables. Parametric triangular and exponential membership functions which are consistent with histogram of data have been evaluated via minimization problem. The classification processes have been performed on bispectral index data sets to see whether the offered theorems have increasing effects on classification accuracies or not. The obtained classification accuracies are compared by other ones which are evaluated by using another approach that previously used in the literature. At the end of the analysis of data sets, it has been proved that the membership function approaches offered in this thesis have increasing effects on average of classification accuracies.

Key Words: Time Series, Clustering, Classification, Data Distribution, Membership Function, Bispectral Index.

İÇİNDEKİLER

YEMİN METNİ	ii
TUTANAK	iii
ÖZET	iv
ABSTRACT	v
İÇİNDEKİLER	vi
KISALTMALAR LİSTESİ	viii
TABLolar LİSTESİ	ix
ŞEKİLLER LİSTESİ	x
GİRİŞ	1
1. LİTERATÜR TARAMASI	3
2. YÖNTEM ve METHODOLOJİ	7
2.1 Sınıflama.....	7
2.2 Bulanık Sayı.....	10
2.3 Bulanık c-Ortalamlar Kümeleme Yöntemi.....	11
2.4 Bulanık c-Ortalamlar Etiketlendirme Yöntemi.....	13
2.5 Zaman Serisi.....	14
3. K - EN YAKIN KOMŞULUK KURALINA DAYANAN ZAMAN SERİSİ ETİKETLENDİRME YÖNTEMİ.....	16
3.1. Mod Değerine Dayanan Zaman Serisi Kümelemesi (TCM).....	16
3.2 Ağırlıklandırılmış Mod Değerine Dayanan Zaman Serisi Kümelemesi (TCWM).....	19
3.3. Etiketlerin Ağırlıklandırılmış ortalamasına dayanan Zaman Serisi Kümelemesi (TCWA).....	21
4. VERİ DAĞILIMINA DAYANAN BULANIK SAYI ELDE ETME TEKNİKLERİ.....	24
4.1 Veri Setinin Merkezi Eğilimine Dayanan Parametrik Üçgen Bulanık Sayı Türetme.....	24
4. 2 Frekans Tablosuna Dayanan Bulanık Sayı Türetme.....	29
4. 2. 1 Frekans Tablosuna Dayanan Üçgen ve Parametrik Üçgen Bulanık Sayı Türetme.....	29
4. 2. 2 Frekans Tablosuna Dayanan Parametrik Üssel Bulanık Sayı Türetme	34

5. BİSPEKTRAL ENDEKS ZAMAN SERİLERİNİN K - EN YAKIN KOMŞULUK KURALINA DAYANAN ZAMAN SERİSİ ETİKETLENDİRME YÖNTEMLERİ İLE ETİKETLENDİRİLMESİ	39
6. SINIF VERİ DAĞILIMLARINA YAKINSAYAN BULANIK ÜYELİK FONKSİYONLARININ OLUŞTURULMASI VE SINIFLAMADAKİ ETKİLERİNİN DEĞERLENDİRİLMESİ	46
7. BİSPEKTRAL ENDEKS VERİ SETLERİNİN ZAMAN SERİLERİ	70
8. DIŞ TİCARET ENDEKSLERİNİN DEĞİŞİM ORANLARININ TCM YÖNTEMİ KULLANILARAK ETİKETLENDİRİLMESİ	96
SONUÇ.....	98
KAYNAKÇA.....	100

KISALTMALAR LİSTESİ

- ARCH : Otoregresif koşullu heteroskedastisiti (Autoregressive conditional heteroskedasticity)
- GARCH : Genelleştirilmiş otoregresif koşullu heteroskedastisiti (Generalized autoregressive conditional heteroskedasticity)
- FCML : Bulanık c-ortalamalar etiketlendirme yöntemi (Fuzzy c-means labeling method)
- TCM : Mod değerine dayanan zaman serisi kümelemesi (Time series clustering based on mode)
- TCWM : Ağırlıklandırılmış mod değerine dayanan zaman serisi kümelemesi (Time series clustering based on weighted mode)
- TCWA : Etiketlerin ağırlıklandırılmış ortalamasına dayanan zaman serisi kümelemesi (Time series clustering based on weighted average of labels)
- AR : Otoregresif (Autoregressive)
- ARMA : Otoregresif hareketli ortalama (Autoregressive moving average)
- ARIMA : Otoregresif entegre edilmiş hareketli ortalama (Autoregressive integrated moving average)
- CART : Sınıflama ve regresyon ağaçları (Classification and regression trees)
- SLIQ : Veri madenciliği için hızlı arttırılabilir sınıflayıcı (A fast scalable classifier for data mining)
- SPRINT : Veri madenciliği için arttırılabilir paralel sınıflayıcı (Scalable parallel classifier for data mining)
- BIS : Bispektral (Bispectral)
- EEG : Elektroansefalografi (Electroencephalography)
- GED : Genelleştirilmiş hata dağılımı (Generalized error distribution)
- ADF : Augmented Dickey-Fuller
- KPSS : Kwiatkowski-Phillips-Schmidt-Shin
- SK : Sınıflama kesinliği

TABLULAR LİSTESİ

Tablo 4.1: Frekans Tablosu.....	29
Tablo 5.1: Bölüm 3’te Önerilen Algoritmaların Hata Miktarları.....	39
Tablo 5.2: FCML, TCM, TCWM, TCWA Algoritmalarının Sınıflama Kesinlikleri.	42
Tablo 5.3: FCML, TCM, TCWM, TCWA Yöntemlerinin Sınıflama Kesinliklerinin Normallik Testi Sonuçları	44
Tablo 5.4: TCM, TCWM, TCWA, FCML Yöntemlerinin İkili-t Testi Sonuçları	45
Tablo 6.1: Bispektral Endeks Uyku Düzeylerinin Tanımlayıcı İstatistikleri	47
Tablo 6.2. Uyku Düzeylerinin Kolmogorov-Simirnov Test Sonuçları.....	48
Tablo 6.3: Teorem 2 için Beş Uyku Düzeyinin Üyelik Bilgileri	48
Tablo 6.4: Uyku Düzeyi 1’in Frekans Tablosu.....	52
Tablo 6.5: Uyku Düzeyi 2’in Frekans Tablosu.....	52
Tablo 6.6: Uyku Düzeyi 3’ün Frekans Tablosu.....	53
Tablo 6.7: Uyku Düzeyi 4’ün Frekans Tablosu.....	53
Tablo 6.8 Uyku Düzeyi 5’in Frekans Tablosu.....	54
Tablo 6.9: Fomül (6.1), Toerem 2, Teorem (5) Üyelik Fonksiyon Yaklaşımlarının Sınıflama Kesinlikleri	59
Tablo 6.10: Sınıflama Kesinliklerinin Normallik Testi Sonuçları.....	60
Tablo 6.11: Teorem 2 , Teorem 5, (6.1)’in İkili-t Testi Sonuçları.....	61
Tablo 6.12: Formül (6.1) ve Teorem 5 Yolu ile Elde Edilen Üyelik Fonksiyon Yaklaşımlarının Sınıflama Kesinlikleri	68
Tablo 6.13: Teorem 5, Formül (6.1)’in İkili-t Testi Sonuçları	69
Tablo 7.1: 5-21 arası veri setlerin ARCH/GARCH modelleri.....	90

ŞEKİLLER LİSTESİ

Şekil 1: Zaman Serisi Kümelemesinin Görsel Olarak Gösterimi.....	1
Şekil 3.1: TCM Algoritmasının Akış Diyagramı.....	18
Şekil 3.2: TCWM Algoritmasının Akış Diyagramı.....	20
Şekil 3.3: TCWA Algoritmasının Akış Diyagramı.....	23
Şekil 4.1: Teorem 2'nin Görsel Gösterimi.....	25
Şekil 4.2: Teorem 3'ün görsel gösterimi.....	27
Şekil 4.3: Teorem 4'de yer alan amaç fonksiyonun görsel olarak gösterimi.....	31
Şekil 4.4: Teorem 5'de yer alan amaç fonksiyonun görsel olarak gösterimi.....	35
Şekil 5.1: TCM, TCWM ve TCWA Algoritmalarının ve Uzman Görüşünün Çıktıları.....	41
Şekil 6.1: Uyku Düzeyleri'nin Teorem 2 ve (6.1) Nolu Formül Kullanılarak Elde Edilen Üyelik Fonksiyonları.....	51
Şekil 6.2: Uyku Düzeyleri'nin Sonuç 4.1 ve (6.1) Nolu Formül Kullanılarak Elde Edilen Üyelik Fonksiyonları.....	56
Şekil 6.3: Uzun Kuyruk Etkisi Kaldırıldığında Uyku Düzeyleri'nin Sonuç 4.1 ve (6.1) Nolu Formül Kullanılarak Elde Edilen Üyelik Fonksiyonları	58
Şekil 6.4: Uyku Düzeyleri'nin Teorem 5 ve (6.1) Nolu Formül Kullanılarak Elde Edilen Üyelik Fonksiyonları	64
Şekil 6.5: Uyku Düzeyleri'nin Teorem 5 ve (6.1) Nolu Formül Kullanılarak Elde Edilen Üyelik Fonksiyonları	67
Şekil 7. 1: Birinci Bireyin Bispektral Endeks Zaman Serisi Grafiği.....	70
Şekil 7. 2: Birinci Veri Setin Ln Değerlerinin Birim Kök Testi E-Views Çıktısı....	71
Şekil 7. 3: Birinci Veri Setinin Ln Değerlerinin Otokorelasyon ve Kısmi Otokorelasyon Fonksiyonları.....	72
Şekil 7. 4: Birinci Veri Setin Ln Değerlerine Dayanan AR(2) Modeli	73
Şekil 7. 5: Birinci Veri Setin AR(2) Modelinin Tahmin Ve Artık Grafiği.....	74
Şekil 7. 6: Birinci Veri Setin AR(2) Modelinin Hata Terimlerinin Histogramı.....	74
Şekil 7. 7: Birinci Veri Setin ARCH Modeli.....	75
Şekil 7. 8: İkinci Bireyin Bispektral İndeks Zaman Serisi Grafiği.....	76
Şekil 7. 9: İkinci Veri Setin Birim Kök Testi.....	77

Şekil 7. 10: İkinci Veri Setin Otokorelasyon Ve Kısmi Otokorelasyon Fonksiyonları.....	78
Şekil 7. 11 İkinci Veri Setin AR(2) Modeli.....	79
Şekil 7. 12: İkinci Veri Setin AR(2) Modelinin Tahmin ve Artık Grafiği.....	79
Şekil 7. 13: İkinci Veri Setin ARCH Modeli.....	80
Şekil 7. 14: Üçüncü Bireyin Bispektral Endeks Zaman Serisi Grafiği.....	81
Şekil 7. 15: Üçüncü Veri Setinin Ln Değerlerinin Birim Kök Testi.....	82
Şekil 7. 16: Üçüncü Veri Setinin Ln Değerlerinin Otokorelasyon Ve Kısmi Otokorelasyon Fonksiyonları.....	83
Şekil 7. 17: Üçüncü Veri Setine Ait Modelin Tahmin Ve Artık Grafiği.....	83
Şekil 7. 18: Üçüncü Veri Setin ARCH Modeli.....	84
Şekil 7. 19: Dördüncü Bireye Ait Zaman Serisi.....	85
Şekil 7. 20: Dördüncü Veri Setinin Birim Kök Testi E-Views Çıktısı.....	86
Şekil 7. 21: Dördüncü Veri Setin Ln Değerlerinin Otokorelasyon Ve Kısmi Otokorelasyon Fonksiyonları	87
Şekil 7. 22: Dördüncü Veri Setin Zaman Serisi Modeli.....	88
Şekil 7. 23: Dördüncü Veri Setine Ait Modelinin Tahmin ve Artık Grafiği.....	88
Şekil 7. 24: Dördüncü Veri Setin ARCH Modeli.....	89
Şekil 7.25: 5-21 Arası Veri Setlerin Gözlem, Tahmin, Artık Grafikleri.....	93
Şekil 8.1: 1995-2004 Yılları Arası Aylık Dış Ticaret Endekslerinin Etiket Serisi ...	96

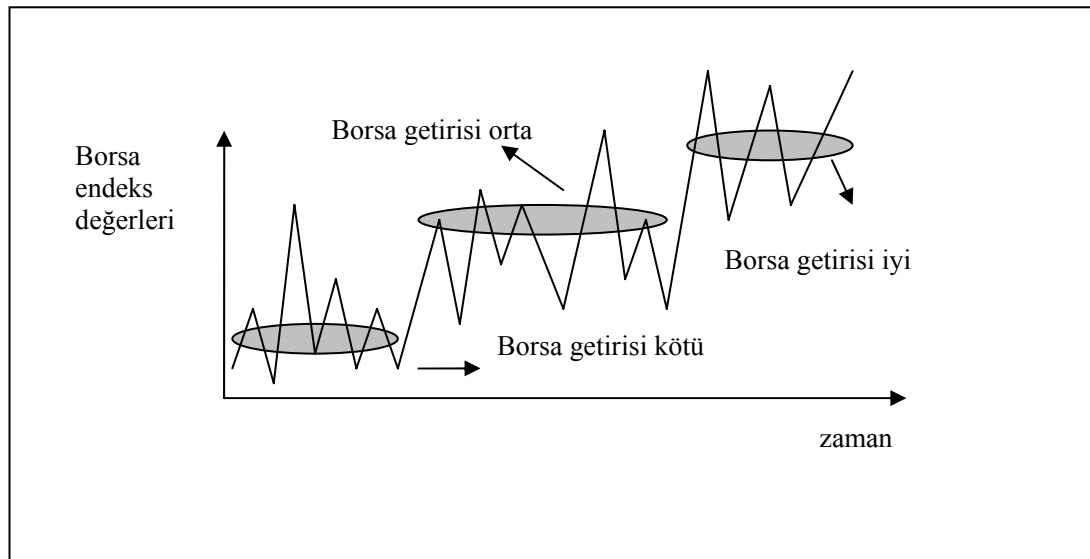
GİRİŞ

Zaman serisi, gözlemlerin kronolojik bir dizisidir. Bu dizilerde gözlemler ardışık olarak elde edilmektedir ve genellikle birbirinden bağımlı gözlemler söz konusudur. Bağımlılığın dikkate alındığı tahmin modellerinde ise, verilerin geçmiş değerlerinden yararlanır ve ileriye yönelik tahminlerde var olan kalıpların gelecekte de devam edeceği varsayılır (Bowerman ve diğerleri, 2005).

Tıp, borsa, finans gibi çeşitli alanlarda etkin olarak kullanılan zaman serisi kümelemesi, zaman serisi veri madenciliğinin önemli görevlerinden biridir. Klasik kümeleme yöntemlerinden farklı olarak bu kümeleme yönteminde sadece gözlem değerleri değil aynı zamanda gözlemlerin kronolojiside dikkate alınır (Guo ve diğerleri 2008).

Şekil 1’de zaman serisi kümeleme yönteminde kullanılan mantık görülebilir. Burada dikkat edilecek olunursa, bazı gözlem değerleri ait olduğu küme merkezi yerine diğer küme merkezine daha yakın gözüksede, zaman bazında komşuları ile aynı kümeye düşmektedir.

Şekil 1: Zaman Serisi Kümelemesinin Görsel Olarak Gösterimi



Bu tezde literatürde yapılan çalışmalar ışığı altında, 3 yeni zaman serisi kümeleme yöntemi geliştirildi (Bakınız Bölüm 3). Sunulan yöntemlerde, büyük veri setlerinin yoğunluk tahmininde ve sınıflanmasında iyi sonuç veren tekniklerden biri olan K - en yakın komşuluk yöntemi kullanıldı. Bulanık c-ortalama etiketlendirme yöntemi K -en yakın komşuluk mantığı ile birleştirildi. Bu mantığın, zaman serisi kümelemesinde ve etiketlere ait zaman serilerinin çıkartılmasında istatistiksel olarak anlamlı farklar yarattığı, Bölüm 5'te bispektral endeks (BIS) veri setleri üzerinde gösterildi. Mod değerine dayanan zaman serisinin sosyal bilimlerde bir uygulaması ise dış ticaret endeksleri üzerinde Bölüm 8'de verildi.

Tezin ikinci kanadında, üyelik fonksiyonunun sınıflamadaki etkinliği dikkate alınarak 4 yeni üyelik fonksiyon yaklaşımı geliştirildi (Bakınız Bölüm 4). Üyelik fonksiyonları elde edilmesinde, veri setinin dağılımı referans kaynağı olarak görüldü. Parametrik üçgen, yamuk ve üssel fonksiyonların kullanıldığı üyelik fonksiyonları üzerinde duruldu. Teorem 2'de, veri dağılımının merkez değerinde 1 üyelik derecesine; uygun olan sağ ve sol tarafta sabitlenmiş X_i ve X_j noktalarında α_D üyelik derecesine sahip olan parametrik üçgen bulanık sayı elde etme tekniği sunuldu. Teorem 3'de, Teorem 2'nin parametrik yamuk bulanık sayıya genelleştirilmesi yapıldı ve her iki teoremde de Nasibov ve Peker'de (2008) ortayan atılan Teorem 1'den yararlandı. Teorem 4'de frekans tablosunda yer alan orta nokta değerleri ile bunlara karşılık gelen normalleştirilmiş yüzdeler, parametrik üçgen bulanık sayının elde edilmesinde kullanıldı; böylece verilerin histogramına yakınsayan parametrik üçgen bulanık sayı elde edildi. Sonuç 4.1'de, Teorem 4'ün üçgen bulanık sayılar için özel hali verildi. Teorem 5'de, frekans tablosunun normalleştirilmiş yüzdeler ile bulanık sayının seviyeleri arasındaki fark minimize edilerek üssel parametrik bulanık sayı elde edildi. Bölüm 6'da, ortaya atılan bu teoremlerin, sınıflama kesinliğinde artırıcı bir rol alıp almadığı bispektral endeks veri setleri üzerinde istatistiksel olarak analiz edildi. Analiz sonucunda ortaya atılan üyelik fonksiyonlarının ilgili veri setleri üzerinde sınıflamayı artırıcı bir rol oynadıkları görüldü. Bölüm 7'de değişen varyans durumu göz önüne alınarak, bispektral endeks veri setlerinin ARCH/GARCH modelleri oluşturuldu.

1 LİTERATÜR TARAMASI

Literatürde araştırmacılar kümelemede ardışıklığı ve bağımlılığı dikkate alarak, zaman serisi kümelemesi ile ilgili birçok yöntem ortaya sundular. Weng ve Shen (2008) çok değişkenli zaman serisi örneklemelerinin satır-satır ve sütun-sütun kovaryans matrisinin özvektörlerine dayanan çok değişkenli sınıflama yöntemi önerdiler. Guo ve diğerleri (2008) yaptıkları çalışmada, bağımsız bileşen analizini modife edilmiş k-ortalama algoritması ile birleştirerek yeni bir zaman serisi kümeleme yöntemi ortaya koydular ve borsa verilerini grupladılar. Chandrakala ve Sekhar (2008) çok değişkenli zaman serisi kümelemede çekirdek özellik uzayında yoğunluk tabanlı kümeleme yöntemini önerdiler. İlgili algoritmanın, küme şekli ve sayısı ile ilgili bir varsayıma dayanmamasını; büyük veri setlerini ve sapanları dikkate almasını bir avantaj olarak gördüler. Zhang ve diğerleri (2004) sabit aralıklı elde edilen veriler yerine, işlemde işleme elde edilen veriler üzerine yoğunlaştılar ve koşullu otoregresif süreç (conditional autoregressive duration) verileri için model tabanlı kümeleme yöntemi sundular. Kung ve Su (2007) ilk etapta girdi-çıkı veri ürün uzayını, bulanık c-regresyon modeli kullanarak hiper düzlem şekilli kümelere ayırdılar ve uygun küme seçiminde yeni geçerlilik kriteri kullandılar. Modelin kesinliğini ayarlamak için gradyan indirgeme algoritması uyguladılar. Li ve diğerleri (2008) farklı-hacimli aralıkların, tahminlemede daha iyi sonuç verebileceğini belirttiler ve zaman serilerinde kesin olmayan yapıyı dikkate aldılar. Bulanık c-ortalama algoritmasını farklı-boyutlu aralıklar elde etmek için kullandılar. Teoh ve diğerleri (2009) zaman serilerinde linguistik aralık uzunluklarını dikkate alarak, bulanık lojistik ilişkiler için yeni bir algoritma önerdiler. Cheng ve diğerleri (2008) bulanık kümelemeye dayanan çok değişkenli bulanık zaman serisi yöntemi geliştirdiler. Weib ve Gob (2008) itere edilmiş fonksiyon sistemi üreten kesirsel kullanarak, ardışık kalıp analizi için yaklaşım sundular. Analizlerinde kategorik alfabenin vektöre dönüşümünü kullandılar. Chouakria ve diğerleri (2009) gen davranışının ve değerinin benzerlik endeksine dayanan gen anlatım profillerinin yeni bir kümelemesini sundular. Kümelemede ardışık gözlemlerin yapısını dikkate aldılar. Otranto (2008) oynaklık bilgisi için kümeleme yöntemi geliştirdi ve GARCH süreçleri arasındaki uzaklığı ölçebilmek için otoregresif metrikler kullandı.

Literatürde yapılan birçok arařtırmada, histogram, mod vs. gibi veri seti bilgisine dayanan üyelik fonksiyon elde etme teknikleri ortaya sunuldu. Örneğın, Devi ve Sarma (1985) histograma dayanan olasılık yoğunluk fonksiyonunun, rasyonel fonksiyon yaklaşımını sundular. İlgili fonksiyonda en yüksek ordinat 1 olacak şekilde normalleřtirme yaptılar. Floreaa (2008) rasgele veri setlerinde bulanık fonksiyon dönüşümünü ele aldı. Luo ve diğeri (2008) eđer-bölüm bulanık küme şekillerinin, bulanık sistem kapasitelerini etkilediklerini belirttiler. Eğitim verileri ile kernel-şekilli eđer-bölüm bulanık kümeyi kapsayan bulanık sistemler kurdular. Qiu ve Joe (2006) veri noktasının adi küme ayırma endekslerini elde etmek için geometrik bir yaklaşım sundular. Greco ve diğeri (2008) değışken kesinlikli (variable precision) rough setlerinin genelleřtirmesini ortaya attılar. Yapılan genelleřtirmeyi mutlak ve relatif rough üyelikleri kavramlarına dayandırdılar. Yazarlar parametrik rough set modelinde, frekans dağılımına dayanan veri ilişkilerinin modelleřtirmesini ele aldılar. Alcalá-Fdez ve diğeri (2009) üyelik fonksiyonları ve bulanık ilişki kurallarını veri seti kullanarak çıkaran bulanık veri madenciliğı algoritması sundular. Üyelik fonksiyonlarının ve bulanık ilişki kurallarının oluşturulmasında genetik öğrenme tekniğini kullandılar. Abbasbandy ve Amirfakhrian (2006) kesikli noktalar setine dayanan, bulanık fonksiyon evrensel yaklaşımı olarak polinom bulanık bir yaklaşım önerdiler. Nock ve diğeri (2009) spektral kümelemenin, yumuşak üyelik fonksiyonuna genişlemesi üzerinde çalıştılar. Comby ve Strauss (2007) video bölümlerinin ana hareket tahmini için bulanık çalışma çerçevesinde, olabilirlik teorisi ve kesin olmayan olasılıklar yardımı ile ana harekete ilişkin kvazi-sürekli histogramın mod deęerinin kesin bir tahminini elde ettiler. Ayhan ve diğeri (2007) histograma dayanan yeni bir üyelik fonksiyonu tasarım yöntemi geliřtirdiler ve bu yöntemi sincap kafesi (squirrel-cage) indüksiyon motorlarında bozuk rotor bar belirleme probleminde uyguladılar. Yanyan ve diğeri (2008) kanonik korelasyon analizinin sınıflamadaki performansını yükseltmek için tasvir örneklemlerinin dağılımını gösteren bulanık üyelik fonksiyonu tasarladılar. Mansoori ve diğeri (2007) bulanık sayının α eřiğine dayanan, bulanık kuralı iki parçaya ayıran bir yaklaşım sundular. Ayırma ölçeğini, kalıbın dağılımını kullanarak buldular. Choi ve Chung-Hoon (2009) histograma ve aralık tip-2 bulanık c-ortalamlar yöntemine dayanan bulanık üyelik fonksiyon algoritmaları geliřtirdiler.

Wu ve Chen (1999) sayısal deęişkenli öğrenme setinden bulanık kurallar çıkartmak ve bulanık kuralların girdi ve çıktı üyelik fonksiyonlarını oluşturabilmek için, eşitliğin α - kesitine dayanan bulanık öğrenme algoritması geliştirdiler. Chang ve Lilly (2004) bulanık sınıflama sistem sürecinde, üyelik fonksiyon ve kuralların otomatik olarak çıkartıldığı yaklaşım sundular. Sistemi, veri dağılımına ilişkin herhangi bir varsayım veya önbilgi olmaksızın verilerden elde ettiler. Chen ve diğerleri (2009) minimum yayılma, üyelik fonksiyon ve bulanık ilişki kuralları elde edebilmek için genetik bulanık madencilik algoritması geliştirdiler. Yang ve Bose (2006) danışmansız otomatik bulanık üyelik fonksiyon çıkartma teknięi önerdiler.

Yapılan dięer bazı arařtırmalarda ise üyelik fonksiyonlarının sınıflamada kesinlięi arttırmak için kullanıldığı görüldü. Örneęin, Au ve diğerleri (2006) sınıflama sonuçlarında iyileştirme yapabilmek için, veri setlerinden üyelik fonksiyonunu hesaplayan bir yöntem sundular. Chen ve Shie (2009) yeni bir bulanık bilgi kazanım ölçümü kullanarak, sınıflama probleminde yeni bir yöntem ve üyelik fonksiyonlarının oluşturulmasında, yeni bir algoritma sundular. Simpson (1992) minimum-maksimum noktalarının kullanıldığı bulanık sayılar ile danışmanlı sinir aęları sınıflayıcısı önerdi. Hu (2006) sınıflama problemlerinde bulanık karar kurallarının elde edilmesinde genetik algoritma kullanan yeni bir yöntem geliştirdi. Üyelik fonksiyon sayısını ve şeklini içeren parametreleri, ikili kromozom dizileri kullanılarak otomatik olarak belirledi. Shyi-Ming ve Chi-Hao (2005) bulanık sınıflama problemleri için herhangi uzman görüşü olmaksızın veri setlerinden, niteliklerin üyelik fonksiyonlarını ve karar kurallarını çıkaran bir yöntem sundular. Shuang ve diğerleri (2009) gaus üyelik fonksiyonuna dayanan hiyerarjik karma bulanık- sinir aęları için bir algoritma geliştirdi. Tenga ve diğerleri (2004) bölge tabanlı üssel fonksiyonu ve eğitim verisini kullanarak, birçok karakteristik bölgesini, girdi uzayına ayıran algoritma geliştirdiler. Devillez (2004) metal kesim sürecini izleyebilmek için bulanık danışmanlı sınıflama yönteminde, üssel fonksiyonlu bulanık kalıp eşleştirme algoritması sundu. McNicholas (2010) gaus fonksiyonlarının kullanıldığı sınıflama teknięi önerdi. Yang ve Wu (2006) gürültülü verilere ve sapanlara karşı dayanıklı olan ve üssel üyelik fonksiyon veren olabirlilik kümeleme

algoritması sundular. Agrawal ve diğeri (2007) rough üyelik fonksiyonuna dayanan danışmanlı sinir ağı sınıflama modeli önerdi.

2 YÖNTEM ve METODOLOJİ

2.1 Sınıflama

Karar vermede kullanılan “gizli-bilgi” veri tabanlarında yoğun olarak bulunmaktadır. Gizli- bilginin ortaya çıkartılmasında kullanılan yöntemlerden birisi de sınıflamadır (Han, 2001). Sınıflama, veri setinde yer alan bir ya da daha fazla değişken değerini kullanarak, kategorik değişken değerinin tahmin edilmesi işlemidir. Tahmin edilmek istenen kategorik değişken değerleri/elemanları, sınıf etiketleri olarak adlandırılır. Sınıflama sonucunda, veri setinde yer alan tahmin edici değişkenler ile sınıf etiketlerini tahmin etmeye yarayan modeller kurulur. Modellerin kurulmasında yapay sinir ağları, genetik algoritma, bayes yöntemi, karar ağaçları, k-en yakın komşuluk kuralı gibi çeşitli yöntemler kullanılabilir (Ye, 2003).

Model kurmak için analiz edilen satır verileri “eğitim setini” oluşturur ve bu setteki her bir satır verisi “eğitim örnekleme” olarak adlandırılır. Eğitim setinde, her örneklemin hangi sınıfa ait olduğu bilindiği için bu yöntem danışmanlı öğrenme yöntemi olarak bilinir (Witten ve Frank, 2005).

Sınıflama işlemi genel olarak iki adımlı süreçten oluşur. Birinci adımda sınıfların kümesini anlatan model, nitelikler ile tanımlanan satır verilerinin analiz edilmesi ile kurulur. İkinci adımda ise model sınıflama için kullanılır ve modelin tahmin etme kesinliği hesaplanır (Han ve diğerleri, 2001).

Sınıflamanın kesinliğinin tahmin edilmesinde bir çok yol bulunmaktadır. Burada sınıflama kesinliğinin yansız olarak tahmin edilebilmesi için eğitim örneklemlerinden bağımsız olan rasgele örneklemler ile test seti oluşturulur. Test setinin sınıf etiketleri, model ile tahmin edilir ve gözlemlenen gerçek sınıf etiketleri ile tahmin edilen sınıf etiketleri karşılaştırılır (Rencher, 2002). Sınıflama modelinin kesinliği, model tarafından doğru olarak tahmin edilen sınıf etiket oranının hesaplanmasıyla elde edilir. Eğer modelin kesinliği uygun bulunursa, o zaman model gelecek satır verilerinin veya nesnelere sınıflarının tahmin edilmesinde kullanılır.

Modelin kesinliđi, eđitim veri seti üzerinde de yapılabilir ancak bu durumda sınıflama kesinliđi gerçekte olduğundan daha yüksek çıkabilir; bu yüzden test setinin kullanılması sınıflama kesinliđinin hesabı açısından daha uygundur. (Han, 2001).

Bu bölümde en çok kullanılan sınıflama yöntemlerinden Bayes, karar ağacı ve K-en yakın komşuluk kuralı kısaca açıklanmaya çalışıldı

Bayes Sınıflama

Bayes sınıflama Bayes Teoremine dayanır ve işleyişi aşağıdaki gibidir (Han ve Kamber, 2004).

1. D eğitim setini oluşturur. Buradaki her bir satır, n boyutlu nitelik vektörü olan $X = (x_1, x_2, \dots, x_n)$ ile gösterilir.

2. C_1, C_2, \dots, C_m şeklinde m tane sınıfın olduğu varsayılır. Sınıflayıcı, X 'in hangi sınıfa ait olduğunu en yüksek ardıl (posterior) olasılığı dikkate alarak belirler. Diğer bir deđişle X , C_i sınıfına ancak ve ancak

$$P(C_i|X) > P(C_j|X), \quad 1 \leq j \leq m, j \neq i \quad (2.1)$$

ise aittir. Bu yüzden $P(C_i|X)$ 'in maksimize edilmesi gerekir. Bayes Teoremine göre

$$P(C_i|X) = \frac{P(X|C_i)P(C_i)}{P(X)} \quad (2.2)$$

dir.

3. $P(X)$ tüm sınıflar için sabittir; bu yüzden maksimizasyonda $P(X|C_i)P(C_i)$ 'nin maksimize edilmesi gerekir. $P(C_i)$ sınıf (prior) olasılıkları bilinmiyorsa, genelde bu olasılıklar eşit kabul edilirler. Ancak $P(C_i) = \frac{|C_{i,D}|}{|D|}$ ile de tahmin edilebilirler. Burada $|C_{i,D}|$, D 'de C_i sınıfına ait eleman sayısını göstermektedir.

4. $P(X|C_i)$ 'nin hesaplanmasında, nitelik deđerlerinin birbirlerinden bađımsız olduğu dikkate alınarak $P(X|C_i) = \prod_{k=1}^n P(x_k|C_i)$ hesaplanır.

5. X 'in sınıf etiketini tahmin edebilmek için, her sınıfa ait $P(X|C_i)P(C_i)$ hesaplanır. Sınıflayıcı, X 'in sınıf etiketini ancak ve ancak $P(X|C_i)P(C_i) > P(X|C_j)P(C_j)$ koşulu sağlandığında C_i olarak tahmin eder, $1 \leq j \leq m, j \neq i$.

Karar Ağacı

Karar ağacı kök düğümden yaprak düğümlerine kadar uzanan, aradaki karar düğümleri kollarla birleştirilen bir sınıflama şeklidir. Karar ağacı algoritmaları danışmanlı öğrenme yöntemleridir. Üzerinde çalışılan eğitim veri setinin, algoritmayı oluşturacak şekilde zengin ve çeşitli olmasına dikkat edilir (Larose, 2005).

Karar ağaçları tek bir düğümlerle başlar. Eğer örneklerin hepsi aynı sınıfa dahil ise o zaman düğüm yaprak haline gelir ve o sınıf ile etiketlenir. Aksi halde, algoritma örnekleri tekil sınıflara en iyi şekilde ayıran özelliği seçer. Bu nitelik düğümdeki “test” veya “karar” nitelik haline gelir. Uygun ayırma işlemi ile her düğümde eğitim veri seti küçük alt gruplara ayrılır ve bu işlem durma kriteri sağlanana kadar devam ettirilir. Eğitim setindeki tüm örnekler tek bir sınıfa aitse; maksimum ağaç derinliğine ulaşılmışsa; terminal düğümdeki örnek sayısı, aile düğümdeki minimum örnek sayısından azsa; düğüm ayrıldıktan sonra bir ya da daha fazla çocuk düğümdeki örnek sayısı, minimum örnek sayısından azsa; en iyi ayırma kriteri belli bir eşiğin altında kalmışsa algoritma durdurulur (Rokach ve Maimon, 2008).

Sınıflama ve regresyon (CART) karar ağacı tiplerinden bir tanesidir. İki çatalı veri setleri için uygundur (Larose, 2005). Algoritma çapraz-doğrulamayı, greedy yerel arama yöntemini kullanmaktadır. Çapraz sorgulama skor fonksiyonunu kullanması, onu diğer algoritmalarından ayıran önemli bir özelliğidir (Hand, 2001). ID3 ve CD5 algoritmaları bilgi entropisini kullanan algoritmalar iken CART algoritması düğüm katkısını (impurity) kullanmaktadır. C4.5 algoritması ID3 algoritmasının bir sonraki versiyonudur ve her ikisinde küçük veri setleri için uygundur (Hellerstein ve Stonebraker, 2005). SLIQ ve SPRINT algoritmaları

büyük veri setleri için olan karar ağacı algoritmalarıdır. Her ikisi de kategorik ve sürekli nitelikleri ele alabilir. Her iki algoritmada, hafızada tutulamayacak kadar büyük disk-yerleşik veri setleri üzerinde önceden ayırma tekniklerini önermektedir (Han, 2001).

K- En Yakın Komşuluk Kuralı

K- en yakın komşuluk kuralı veri madenciliğinde, kalıp tanımlarında vs. kullanılan önemli tekniklerden biridir. Etiketli belli olmayan nesneyi sınıflayabilmek için, öğrenme setinde yer alan nesnelere ile yeni nesnenin arasındaki uzaklıklar hesaplanır. Etiketli belli olmayan yeni nesnenin sınıfı, öğrenme veri setinde yer alan kendisine en benzer nesnenin/nesnelerin sınıfı ile belirlenir. Benzerliğin hesabında birçok uzaklık metriği kullanılabilir; ancak en çok tercih edilen metrik Öklit (Euclidean) uzaklığıdır (Parthasarathy ve Chatterji, 1990).

K- en yakın komşuluklar bulunduğundan sonra, yeni nesnenin etiketi K- en yakın komşulukların sınıf etiketlerini sayma işlemi ile gerçekleştirilir. Sınıf belirleme sürecinde “çoğunluk sayımı” ve “ağırlıklandırılmış toplam sayım” şeklinde iki farklı sayma işlemi kullanılabilir. Çoğunluk sayımında K-en yakın komşuluğunda en sık gözüken sınıf etiketi, yeni nesnenin sınıf etiketi olarak belirlenir. Ağırlıklandırılmış toplam sayımda ise her sayım, yakınlığı ile doğru orantılı olarak ağırlıklandırılmaktadır (He ve Fault, 2007).

2.2 Bulanık Sayı

$A \in F(E)$ bulanık sayısının LR-parametrik şekli aşağıdaki gibidir,

$$\forall \alpha \in (0,1]: A^\alpha = [L_A(\alpha), R_A(\alpha)] \subset (-\infty, \infty). \quad (2.3)$$

Burada $F(E)$ bulanık sayılar uzayını, A^α bulanık sayının α -seviye kümesini $L: [0,1] \rightarrow (-\infty, \infty)$ bulanık sayısının sol tarafının monotik azalmayan sol sürekli fonksiyonu, $R: [0,1] \rightarrow (-\infty, \infty)$ bulanık sayının sağ tarafının monotik artmayan sağ

sürekli fonksiyonunu gösterir. (Nasibov, 2002; Nasibov vd., 2005). Aşağıdaki koşulların sağlandığı varsayılmaktadır.

$$\lim_{\alpha \rightarrow 0} A^\alpha = \lim_{\alpha \rightarrow 0} [L_A(\alpha), R_A(\alpha)] = [L_A(0), R_A(0)] \quad (2.4)$$

$$L_A(0) > -\infty, R_A(0) < \infty, \quad (2.5)$$

B bulanık sayısının üyelik fonksiyonu, $L_A(1) - d_1$ sol öz sınır noktasına ve $R_A(1) + d_2$ sağ öz sınır noktasına; σ_B sol yayılmaya, β_B sağ yayılmaya sahip $\max(0, 1 - |x|^s)$ şeklinde parametrik kalıp fonksiyon şeklinde tanımlansın.

Teorem 1 (Nasibov ve Peker, 2008): $0, 1$ ve α_d seviyesinde A bulanık sayısı ile çakışan B parametrik kalıbı elde etmek için, d_1, d_2, s_L, s_R parametreleri aşağıdaki koşulları sağlamalıdır.

$$d_1 = 0, \quad d_2 = 0, \quad (2.6)$$

$$\sigma_B = \sigma_A, \quad \beta_B = \beta_A, \quad (2.7)$$

$$s_L = \frac{\text{Log}(1 - \alpha_d)}{\text{Log}\left(\frac{L_A(1) - L_A(\alpha_d)}{\sigma_A}\right)} \quad \text{ve} \quad s_R = \frac{\text{Log}(1 - \alpha_d)}{\text{Log}\left(\frac{R_A(\alpha_d) - R_A(1)}{\beta_A}\right)} \quad (2.8)$$

Teorem 1'in ispatı (Nasibov ve Peker, 2008)'de yer almaktadır.

2.3 Bulanık c-Ortalamlar Kümeleme Yöntemi

Kümeleme gruplar arasındaki benzerliği maksimize, aynı zamanda gruplar içindeki benzerliği minimize eden, verileri ayıran danışmansız bir yöntemdir (Yu ve Yin, 2005; Pedrycz, 2005)

Bulanık c-ortalamlar algoritması en çok kullanılan yöntemlerden biridir ve bu algoritma aşağıdaki gibi bir yapıya sahiptir.

c bir tamsayı, $1 < c < n$ ve $X = \{x_1, x_2, \dots, x_n\}$, R^p 'de sınıf etiketi olmayan veri setidir. $(n \times c)$ boyutunda $U = \{u_{iz}\}$ matrisi X 'in c kümeye ayrımını gösterir. Bulanık c -ortalamalar yönteminde, $U = \{u_{iz}\}$ hesaplama sonunda aşağıdaki özellikleri sağlayacak şekilde iteratif olarak hesaplanır,

$$(i) \ 0 \leq u_{iz} \leq 1, \ i = 1, \dots, n; \ z = 1, \dots, c,$$

$$(ii) \ 0 < \sum_{i=1}^n u_{iz} < n, \ z = 1, \dots, c,$$

$$(iii) \ \sum_{z=1}^c u_{iz} = 1, \ i = 1, \dots, n.$$

Bulanık c -ortalamalar kümeleme yöntemi aşağıdaki (J_m) amaç fonksiyonunun minimizasyonuna dayanmaktadır.

$$J_m(U, \alpha, X) = \sum_{z=1}^c \sum_{i=1}^n (u_{iz})^m \|x_i - \alpha_z\|^2 \quad (2.9)$$

Burada α_z , z 'inci kümenin bilimeyen merkez vektörüdür, $1 \leq z \leq c$. $\|x_i - \alpha_z\|$ ise R^p 'de farklılık ölçüsüdür ve genellikle hesaplanmasında Öklit normu kullanılır. $m > 1$, modelin ağırlıklandırma üssüdür.

Eğer $\|x_i - \alpha_z\| > 0$ her i ve z için sağlanıyorsa, o zaman U matrisi J_m 'i iteratif olarak aşağıdaki fonksiyonları güncelleştirerek minimize eder,

$$u_{iz} = \left[\sum_{k=1}^c \left(\frac{\|x_i - \alpha_z\|}{\|x_i - \alpha_k\|} \right)^{2/(m-1)} \right]^{-1}, \ i = 1, \dots, n; \ z = 1, \dots, c, \quad (2.10)$$

ve

$$\alpha_z = \frac{\sum_{i=1}^n (u_{iz})^m x_i}{\sum_{i=1}^n (u_{iz})^m}, \ z = 1, \dots, c. \quad (2.11)$$

Burada $m > 1$ 'dir. İterasyon $|U_{r+1} - U_r| \leq \varepsilon$ kısıtı sağlanana kadar devam ettirilir (r iterasyon sayısıdır).

Eğer iterasyonda verilen x_i ve bir yada daha fazla küme merkezi için $\|x_i - \alpha_z\| = 0$ durumu mevcutsa, tekillik söz konusudur ve u_{iz} hesaplanamaz. Bu durumun üstesinden gelebilmek için tekil olmayan kümelere 0 ve tekil olanlara $\sum_{z=1}^c u_{iz} = 1$ koşulunu sağlayacak şekilde aynı üyelik derecesi atanır (Pal and Bezdek, 1995).

2.4 Bulanık c-Ortalamlar Etiketlendirme Yöntemi (FCML)

x_i , i . zamana ait gözlemlenen değeri ve s_i gözlemlenen değerlerin sınıfını gösterir; s_i 'nin ait olabileceği c tane sınıfın olduğu varsayılır.

Bulanık c-ortalamlar etiketlendirme yönteminde s_i değerlerini belirleyebilmek için, ilk olarak birleştirilmiş veri seti üzerindeki x_i değerlerine bulanık c-ortalamlar algoritması uygulanır. Her bir sınıfın ortalaması (2.11) formülünü kullanarak aşağıdaki gibi hesaplanır.

$$\alpha_z = \frac{\sum_{i=1}^n (u_{iz})^m x_i}{\sum_{i=1}^n (u_{iz})^m}, \quad z = 1, \dots, c, \quad (2.12)$$

Burada u_{iz} , z . sınıfa ait x_i değerinin üyelik derecesini göstermektedir.

İlgili ortalama değerleri kullanılarak, herbir sınıfın üyelik fonksiyonu (2.10)'daki formüle göre aşağıdaki gibi oluşturulur.

$$\mu_z(x) = \left[\sum_{k=1}^c \left(\frac{\|x - \alpha_z\|}{\|x - \alpha_k\|} \right)^{2/(m-1)} \right]^{-1}, \quad z = 1, \dots, c, \quad (2.13)$$

Daha önceki bölümde belirtildiği üzere, iterasyonlarda verilen x değeri için bir ya da daha fazla $\|x - \alpha_z\|$ uzaklık “0” olarak bulunursa, tekil olmayan kümeler 0 üyelik derecesi, tekil kümeler ise $\sum_{z=1}^c \mu_z(x) = 1$ koşulunu sağlayacak şekilde aynı üyelik dereceleri atanır.

Sürecin sonunda, s_i atanmaları aşağıdaki kritere uygun olarak gerçekleştirilir.

$$s_i = \arg \max_{z=1, \dots, c} \mu_z(x_i), \quad i = 1, \dots, n, \quad (2.14)$$

Burada, sınıfların adı veya sembolik karakteri, sınıfların merkezleri dikkate alınarak verilebilir. Örneğin, en düşük merkez değerine sahip küme içinde bulunan elemanlar “1:çok düşük”, maksimum merkez değerine sahip küme içerisinde yer alan elemanlar “5:çok yüksek” olarak adlandırılabilir. Sürecin sonunda, etiket değerleri orijinal değerler ile yer değiştirerek etiketlere ait zaman serileri oluşturulur.

2.5 Zaman Serisi

Zaman serisi gözlemlerin zaman parametresine göre ardışık olarak sıralanmasıdır. Bu seriler sürekli veya kesikli olarak ölçülebilirler. Sürekli zaman serileri anlık ve sürekli olarak kayıt edilir. Örneğin osilograf, ses yükselticinin harmonik salınımlarını sürekli olarak kayıt eder. Bununla birlikte sosyal bilimlerde bir çok gözlem düzenli aralıklar ile elde edilir. Bu zaman serisi veri tipide kesikli olarak nitelendirilir (Yaffee, 2000).

Zaman serisi bileşenleri trend, döngü, mevsimsel değişme, düzensiz dalgalanmalar şeklindedir. Trend, zaman serisini zamana karşı karakterize eden, aşağı

ya da yukarı doğru olan eğilimdir. Döngü, trent çevresinde yukarı ve aşağı tekrarlayan hareketlerdir. Mevsimsel değişme, belli bir yıl içinde tamamlanan zaman serisindeki periyodik kalıplardır. Düzensiz dalgalanmalar ise zaman serisinde belli bir kalıba göre hareket etmeyen kararsız dalgalanmalardır (Boweman ve diğerleri, 2005).

Zaman serisi analizinin önemli varsayımlardan biri sabit varyans durumudur. Değişen varyans durumunun olduğu durumlarda, özellikle finansal verilerinde oynaklığın kümelenmesi için ARCH modellerine başvurulur (Brooks, 2008).

ARCH modellerinde,

i) modeldeki u_t hata teriminin, serisel olarak ilişkisiz ancak bağımlı olduğu;

ii) u_t 'nin bu bağımlılığının, u_t 'nin geçmiş değerlerinin karelerinin kullanıldığı bir model ile açıklanabileceği temel fikir olarak benimsenmektedir (Tsay, 2005).

GARCH modelinde, u_t^2 'nin otoregresif formunun yanısıra, koşullu beklenen değere karşılık gelen, σ_t^2 'nin de otoregresif yapası dikkate alınmaktadır (Lütkepohl ve Kratzig, 2004).

ARCH veya GARCH modellerinde: (Brooks, 2008).

1) Ortalama ve varyans için uygun formül belirlenir, (Örneğin AR(1)-GARCH(1,1) modeli)

$$y_t = \mu + \phi y_{t-1} + u_t, \quad u_t \sim N(0, \sigma_t^2)$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \beta \sigma_{t-1}^2.$$

2) Birinci aşamada kullanılan model dikkate alınarak, normallik varsayımı altında log-olabilirlik fonksiyonu maksimize edilir.

$$L = -\frac{T}{2} \log(2\pi) - \frac{1}{2} \sum_{i=1}^T \log(\sigma_i^2) - \frac{1}{2} \sum_{i=1}^T (y_i - \mu - \phi y_{i-1})^2 / \sigma_i^2$$

Maksimizasyon işleminde çoğunlukla bilgisayar programlarından yararlanılır ve parametreler tahmin edilir.

3 K - EN YAKIN KOMŞULUK KURALINA DAYANAN ZAMAN SERİSİ ETİKETLENDİRME YÖNTEMİ

Büyük veri setleri için K - en yakın komşuluk yöntemi, yoğunluk tahmininde ve sınıflama işleminde iyi sonuç veren yöntemlerden biridir. Bu yöntemde, bir noktanın K - en yakın komşuluğundaki noktalar dikkate alınarak yoğunluklar hesaplanmaktadır (Parthasarathy ve Chatterji 1990). Bu çalışmada, birbiri ile ilişkili olan gözlemlerin yer aldığı zaman serisinde, zaman çizelgesi üzerinde bir gözlemin K -komşuluğundaki noktaların dikkate alınmasının sınıflama işleminde kesinliği artırıcı bir rol oynayacağı düşünüldü. Bu noktadan yola çıkarak, gözlemlerin kendinden önceki ve sonraki K -en yakın komşuluğundaki değerleri dikkate alan 3 yeni danışmansız zaman serisi etiketlendirme yöntemi ortaya atıldı.

3.1 Mod Değerine Dayanan Zaman Serisi Kümelemesi (TCM)

Ortaya atılan TCM ve sonraki iki algoritmada, ilk olarak M_1, M_2, \dots, M_k zaman serisi veri setleri, $M = \{(x_1, x_2, \dots, x_n)\}$ şeklinde birleştirilir. Daha sonra bulanık c-ortalama etiketlendirme yöntemi (FCML) M üzerinde uygulanır. Uygulama sonunda, gözlemlenen zaman serisi değeri ile onun tahmin edilen sınıf etiketinden oluşan (x_i, s_i) ikililer yardımıyla, $X = \{(x_1, s_1), (x_2, s_2), \dots, (x_n, s_n)\}$ elde edilir. Burada, j . zaman serisinin etiket serisi $L_j = \{s_1, s_2, \dots, s_h\}$ ile gösterilir (h, j . zaman serisi veri kümesinde yer alan örneklem sayısını göstermektedir)

Sonraki aşamada daha düzgün bir eğri elde edebilmek için algoritmada K -en yakın komşuluk kurallarından biri olan “çoğunluk sayımı” L_j etiket seti üzerinde uygulanır. Bunun için, i . etiketin K –komşuluğunda yer alan etiketler sayılır, onlar arasından en sık tekrarlanan etiket s_i^* bulunur, s_i^* ilgili gözlemin etiket değeri olarak atanır. Bir ya da daha fazla mod gözükme durumunda ise, rasgele olarak bir tanesi seçilir. Atama işlemlerinde $(k+1)$. gözlemden başlanılır, atama işlemi yapıldıktan sonra veri seti güncelleştirilerek $(k+2)$. gözleme gidilir.

TCM Algoritması

Girdi. h : ele alınan zaman serisinin uzunluğu, K : komşuluk nokta sayısı,
 $k = \frac{K}{2}$, c : küme sayısı, $M = (x_1, x_2, \dots, x_n)$ birleştirilmiş veri seti.

Adım 1. M birleştirilmiş veri seti üzerinde bulanık c-ortalamalar etiketlendirme yöntemini uygula, ele alınan zaman serisinin, zaman etiket serisini (L) çıkart.

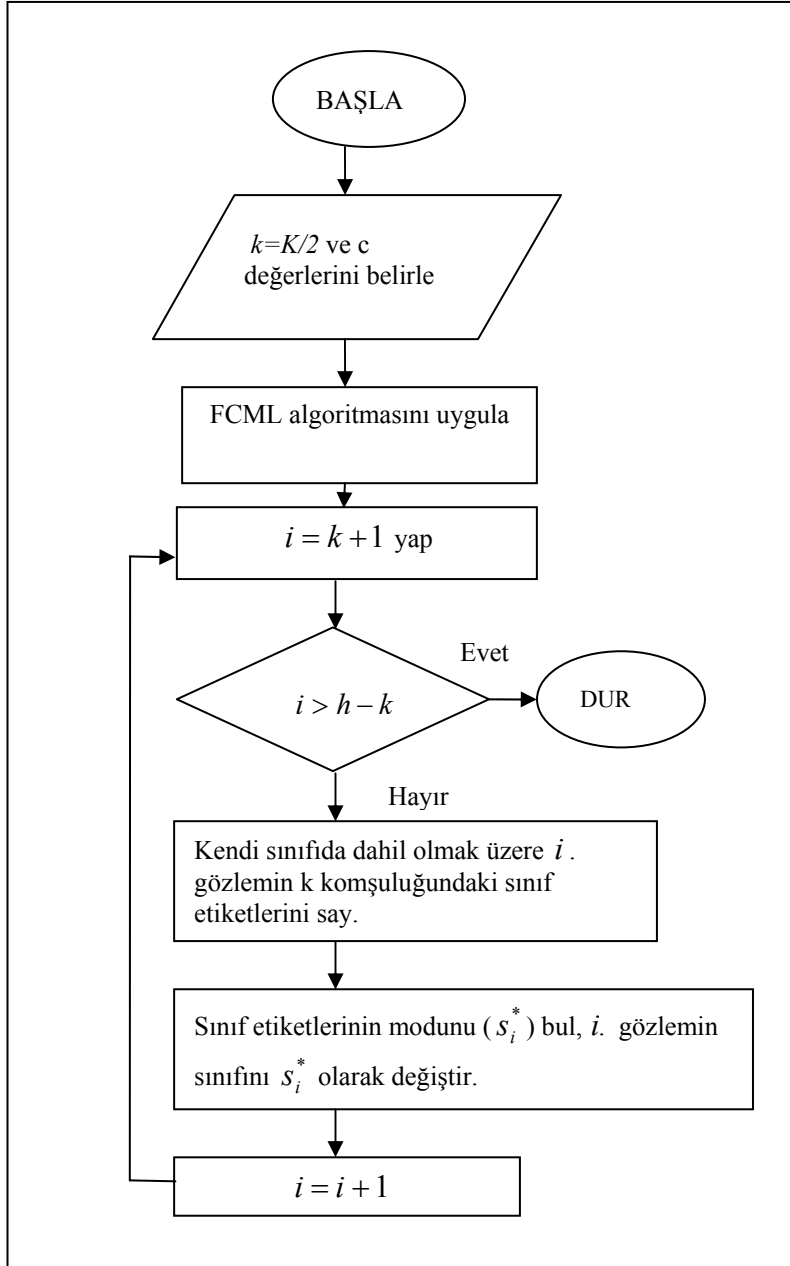
Adım 2. $i = k + 1$ 'den $i = h - k$ kadar Adım 3 ve Adım 4'ü tekrar et.

Adım 3. L 'de i . etiketin k -sağ tarafındaki ve k -sol tarafındaki etiketleri say, ve (s_i^*) çoğunluk etiketini bul.

Adım 4. s_i 'yi s_i^* ile yer değiştirerek, L 'yi tekrar yapılandır, $i = i + 1$ yap ve adım 3'e git.

Bu algoritmanın akış diyagramı Şekil 3.1'de yer almaktadır.

Şekil 3.1: TCM Algoritmasının Akış Diyagramı



3.2 Ağırlıklandırılmış Mod Değerine Dayanan Zaman Serisi Kümelemesi (TCWM)

Bu algoritmada, yine ilk olarak bulanık c-ortalamlar etiketlendirme yöntemi, birleştirilmiş veri seti M üzerinde uygulanır. Diğer yöntemden farklı olarak “yakındaki gözlemler uzaktaki gözlemlere göre daha çok sayılmalıdır” prensibi etiketlere ait zaman serilerinin düzleştirilmesi kullanılır. Bu şekilde en yakın komşuların tahmin üzerindeki etkisi arttırılır.

Sayma sürecinde, i . gözlemin sınıf etiketi (s_i) $k+1$ kere; $(i-1)$. ve $(i+1)$. komşulukların sınıf etiketleri k kere, $(i-2)$. ve $(i+2)$. komşulukların sınıf etiketleri $k-1$ kere, ..., en son olarak $(i-k)$. ve $(i+k)$. komşulukların sınıf etiketleri 1 kez sayılır.

TCWM Algoritması

Girdi. h : ele alınan zaman serisinin uzunluğu, K : komşuluk nokta sayısı, $k = \frac{K}{2}$, c : küme sayısı, $M = (x_1, x_2, \dots, x_n)$ birleştirilmiş veri seti.

Adım 1. M birleştirilmiş veri seti üzerinde bulanık c-ortalamlar etiketlendirme yöntemini (FCML) uygula; ele alınan zaman serisinin, zaman etiket serisini (L) çıkart.

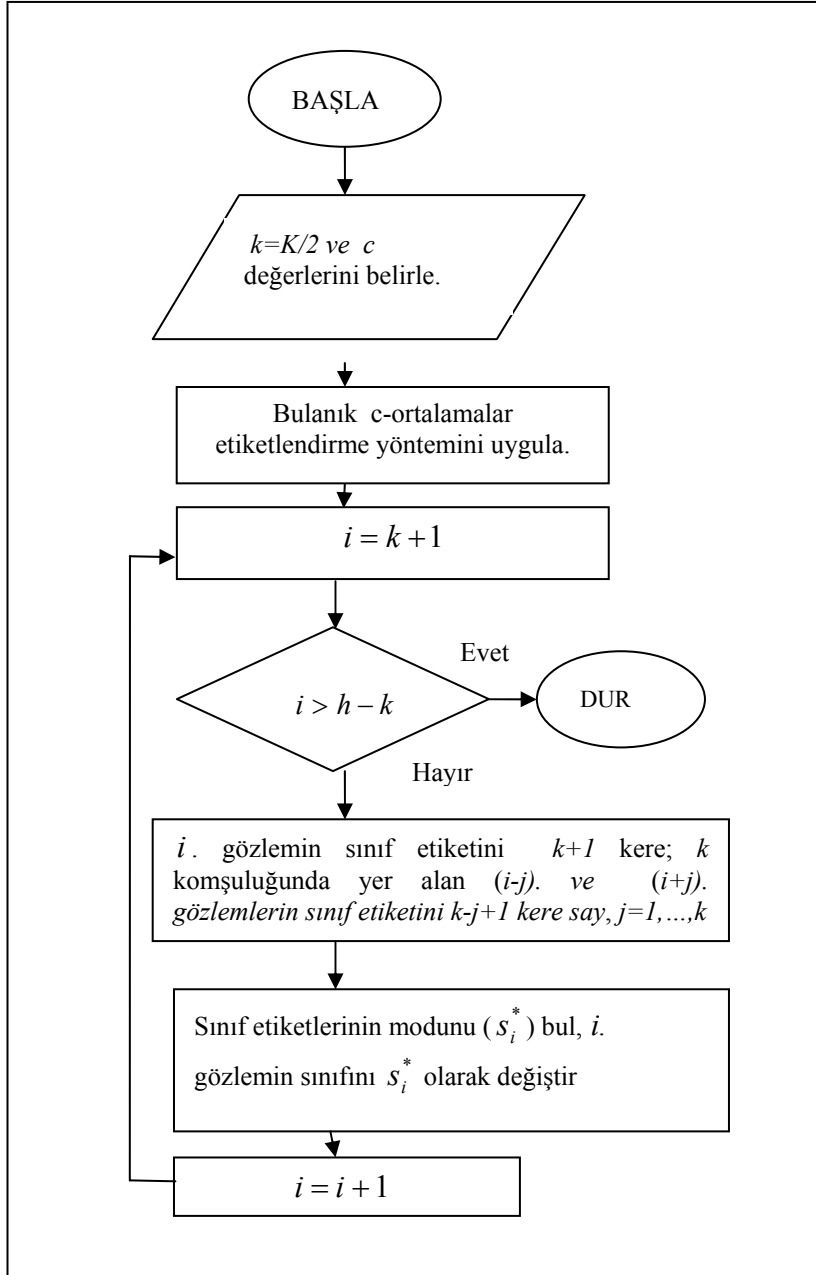
Adım 2. $i = k+1$ 'den $i = h-k$ kadar Adım 3 ve Adım 4'ü tekrar et.

Adım 3. i . gözlemin sınıfını $k+1$ kere; $(i-j)$. ve $(i+j)$. gözlemlerin sınıflarını $k-j+1$ kere say, $j = 1, \dots, k$. Sınıf etiketlerinin modunu (s_i^*) bul.

Adım 4. s_i 'yi s_i^* ile yer değiştirerek L 'yi yeniden yapılandır, $i = i+1$ yap ve Adım 3'e geç.

Bu algoritmanın akış diyagramı Şekil 3.2'deki gibidir.

Şekil 3.2: TCWM Algoritmasının Akış Diyagramı



3.3 Etiketlerin Ağırlıklandırılmış Ortalamasına Dayanan Zaman Serisi Kümelemesi (TCWA)

Bu çalışmada etiketlerin zaman serileri (L), TCM ve TCWM algoritmalarında yer alan aynı yöntem kullanılarak çıkartılır. Daha sonra i . gözlemin sınıf etiketi, K -komşuluktaki sınıf etiketlerini kapsayan bir model yardımıyla tahmin edilir; herhangi bir sayma işlemi uygulanmaz.

Bu yöntemde i . gözlemin sınıf etiketinin tahmini aşağıdaki model ile bulunur:

$$s_i^* = \sum_{j=1}^k \beta_i^{(j)} s_{L,j} + \beta_i s_i + \sum_{j=1}^k \beta_i^{(j)} s_{R,j}, \quad (3.1)$$

Burada, s_i , $s_{L,j}$, $s_{R,j}$ sırasıyla i . gözlemin sınıf etiketini, j . sol taraf komşunun sınıf etiketini, j . sağ taraf komşunun sınıf etiketini; $\beta_i^{(j)}$ j . sol ve sağ taraf komşu sınıf etiketlerinin ağırlık katsayılarını göstermektedir ($j = 1, \dots, k$). İlgili katsayıların ön değerleri aşağıdaki formüller kullanılarak bulunur.

$$\beta_i = 1, \beta_i^{(j)} = 1 - \frac{j}{k+1}, \quad j = 1, \dots, k. \quad (3.2)$$

Daha sonra, katsayı toplamında 1 değerini elde edebilmek için, β_i ve $\beta_i^{(j)}$ parametreleri, tüm parametrelerin toplam değerlerine bölünür ve katsayıların son tahminleri elde edilir.

$$\beta_i \leftarrow \frac{\beta_i}{\beta_i + \sum_{j=1}^k \beta_i^{(j)}} = \frac{1}{k+1}, \quad (3.3)$$

$$\beta_i^{(j)} \leftarrow \frac{\beta_j}{\beta_i + \sum_{j=1}^k \beta_i^{(j)}} = \left(1 - \frac{j}{k+1}\right) / (k+1), \quad j = 1, \dots, k. \quad (3.4)$$

Bu şekilde model parametreleri için aşağıdaki normallik koşulu sağlanmış olunur.

$$\beta_i + 2 \sum_{j=1}^k \beta_i^{(j)} = 1. \quad (3.5)$$

(3.3) ve (3.4)'yer alan parametreler (3.1)'deki modelde yerine konulursa, tahmin modeli aşağıdaki gibi tekrar yazılabilir.

$$s_i^* = \sum_{j=1}^k \left(\frac{1}{k+1} - \frac{j}{(k+1)^2} \right) s_{L,j} + \frac{1}{k+1} s_i + \sum_{j=1}^k \left(\frac{1}{k+1} - \frac{j}{(k+1)^2} \right) s_{R,j}. \quad (3.6)$$

Anlamli bir yorum yapılabilmesi için s_i^* değeri tamsayıya yuvarlanır.

TCWA Algoritması

Girdi. h : ele alınan zaman serisinin uzunluğu, K : komşuluk nokta sayısı, $k = \frac{K}{2}$, c : küme sayısı, $M = (x_1, x_2, \dots, x_n)$ birleştirilmiş veri seti.

Adım 1. M birleştirilmiş veri seti üzerinde bulanık c-ortalamlar etiketlendirme yöntemini uygula (FCML); ele alınan zaman serisinin etiket serisini (L) çıkart.

Adım 2. $i = k + 1$ 'den $i = h - k$ kadar Adım 3 ve Adım 4'ü tekrar et.

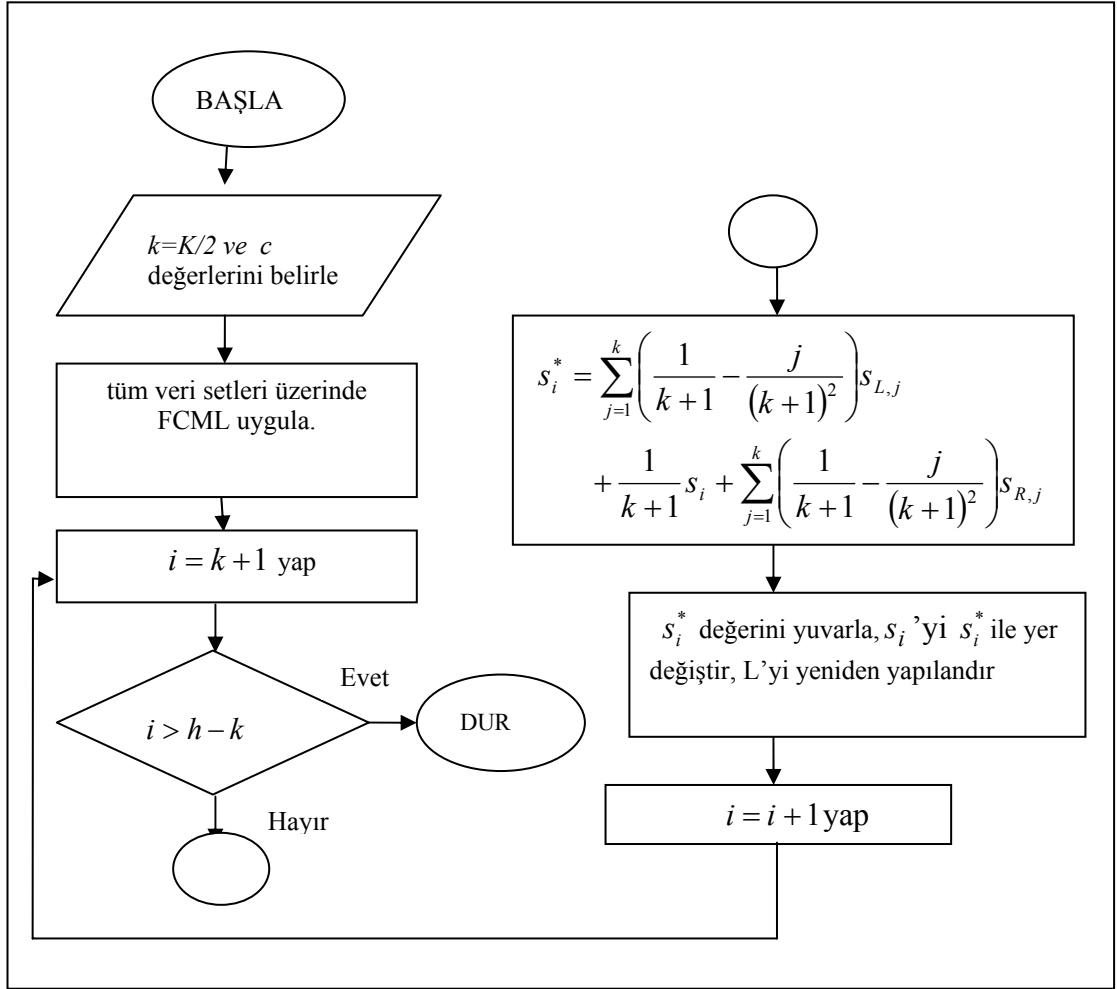
Adım 3. i . gözlemin sınıf etiketini aşağıdaki formülü kullanarak bul

$$s_i^* = \text{yu var la} \left[\sum_{j=1}^k \left(\frac{1}{k+1} - \frac{j}{(k+1)^2} \right) s_{L,j} + \frac{1}{k+1} s_i + \sum_{j=1}^k \left(\frac{1}{k+1} - \frac{j}{(k+1)^2} \right) s_{R,j} \right]$$

Adım 4. s_i 'yi s_i^* ile yer değiştirerek L 'yi yeniden yapılandır, $i = i + 1$ yap ve Adım 3'e geç.

Bu algoritmanın akış diyagramı Şekil 3.3'de yer aldığı gibidir.

Şekil 3.3. TCWA Algoritmasının Akış Diyagramı



4 VERİ DAĞILIMINA DAYANAN BULANIK SAYI ELDE ETME TEKNİKLERİ

Veri dağılımı histogram, kutu grafiği, kök-yaprak gibi grafiklerin yanı sıra; çeyreklikler, ortalama, varyans, standard sapma, mod, medyan vs. gibi tanımlayıcı istatistikler ile özetlenebilir. Buradan hareketle dağılım simetrik mi, sola mı yoksa sağa mı çarpık, yayılım miktarı büyük mü, küçük mü sorularına yanıt alınabilir.

Veri dağılımına uygun üyelik fonksiyonların oluşturulması hem bulanıklığa dayanan çalışmalar ile istatistik arasında bir geçiş; hem de sınıflamada üyelik fonksiyonlarına dayanan yöntemlerde bir iyileştirme sağlayacaktır. Bu noktadan yola çıkarak bu çalışmada, veri dağılımını en iyi yansıtan yeni üyelik fonksiyonu elde etme teknikleri ile ortaya atıldı.

4.1 Veri Setinin Merkezi Eğilimine Dayanan Parametrik Üçgen Bulanık Sayı Türetme

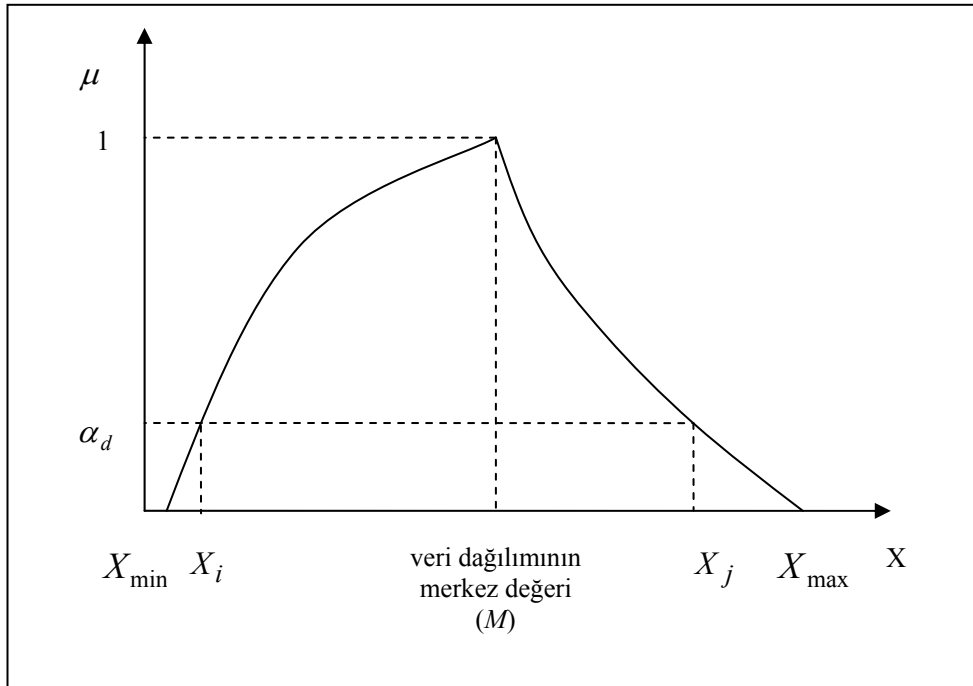
Bu bölümde, X_1, X_2, \dots, X_N veri seti dikkate alınarak Nasibov ve Peker (2008)'in ortaya attığı Teorem 1 revize edildi ve veri dağılımının 5 noktası ile çakışan parametrik üçgen bulanık sayı oluşturuldu. Hesaplamalarda aşağıdaki üçgen parametrik bulanık sayı tanımı dikkate alındı.

$$\mu(X) = \begin{cases} 1 - \left(\frac{M - X}{M - X_{\min}} \right)^{S_L} & ; X_{\min} < X \leq M \\ 1 - \left(\frac{X - M}{X_{\max} - M} \right)^{S_R} & ; M < X < X_{\max} \\ 0 & ; \text{aksihalde} \end{cases} \quad (4.1)$$

Burada X_{\min} , veri aralığının alt sınırını; X_{\max} , veri aralığının üst sınırını göstermektedir. s_L ve s_R katsayıları ise bulanık sayının sağ ve sol tarafının şeklini belirleyen katsayılardır ($s_L > 0, s_R > 0$).

Teorem 2’de bulanık sayı ile veri dağılımı arasında yapılan eşleştirmeler Şekil 4.1’de görsel olarak yer almaktadır. Şekilden de anlaşılacağı üzere, oluşturulan bulanık sayıda, bulanık sayının merkez değeri veri dağılımının merkez değerine eşit kılınmakta ve X_i, X_j noktaları bulanık sayıda α_d seviyesi ile eşleştirilmektedir ($X_i < M < X_j$).

Şekil 4.1: Teorem 2’nin Görsel Gösterimi



Teorem 2: Veri dağılımının merkez değerinde (mod, medyan, ortalama) 1 üyelik derecesine; X_{\min} ve X_{\max} noktalarında 0 üyelik derecesine; sırasıyla $L_A(\alpha_d) \neq M$ ve $R_A(\alpha_d) \neq M$ koşulunu sağlayan sol (X_i) ve sağ taraf (X_j) veri noktalarında α_D üyelik derecesine sahip olan, (4.1) üyelik fonksiyonun parametreleri aşağıdaki gibidir:

$$\begin{aligned}
 L_A(0) &= X_{\min}, \\
 R_A(0) &= X_{\max}, \\
 M &\in \{Mod, Medyan, Ortalama\}, \\
 s_L &= \frac{\text{Log}(1 - \alpha_d)}{\text{Log}\left(\frac{M - L_A(\alpha_d)}{M - L_A(0)}\right)}, \quad s_R = \frac{\text{Log}(1 - \alpha_d)}{\text{Log}\left(\frac{R_A(\alpha_d) - M}{R_A(0) - M}\right)}. \quad (4.2)
 \end{aligned}$$

İspat: (2.6)- (2.8) formüllerinde aşağıdaki eşitlikler dikkate alınırsa,

$$L_A(1) = R_A(1) = M$$

$$L_A(0) = X_{\min}$$

$$R_A(0) = X_{\max}$$

$$\sigma_A = M - X_{\min}$$

$$\beta_A = X_{\max} - M,$$

(4.2)'deki formüller kolayca elde edilebilir.

Teorem 2'de, veri dağılımının merkez noktası $\alpha = 1$ üyelik derecesine karşılık gelmektedir. Burada, veri dağılımı çarpıksa veya bir ya da daha fazla sapan gözlem bulunuyorsa merkez değer olarak ortalama yerine mod veya medyan kullanılmalıdır.

Veri dağılımının şekli iki tepeli ise iki tane mod değeri söz konusudur yani dağılım iki merkez değerine sahiptir. Bu durumda parametrik yamuk üçgen bulanık sayının kullanımı daha uygundur. Teorem 3'te ilgili konu dikkate alınarak parametrik

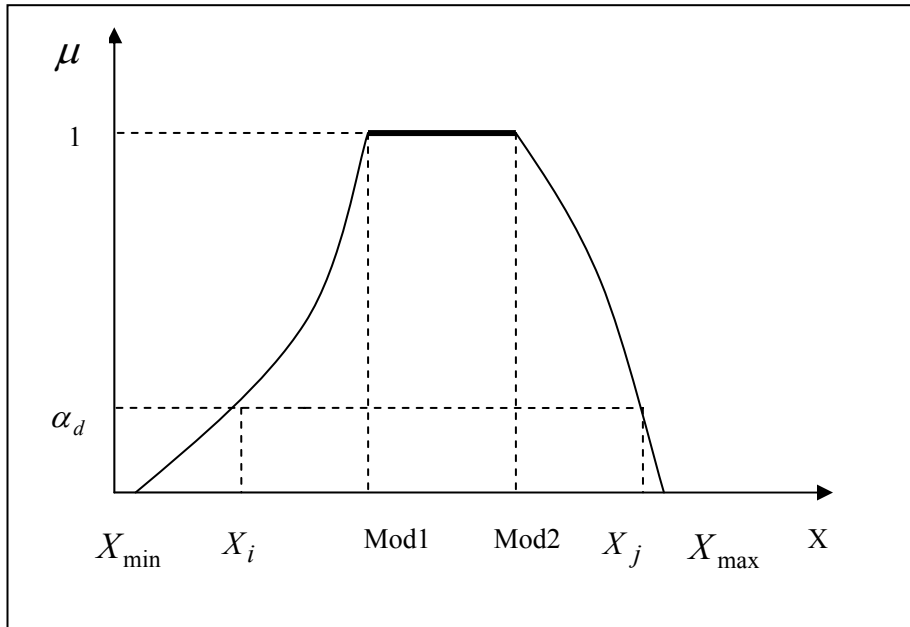
yamuk bulanık sayının parametre formülleri ortaya atıldı. Bu teorem aşağıdaki parametrik yamuk bulanık sayı tanımını için geçerlidir,

$$\mu(X) = \begin{cases} 1 - \left(\frac{M1 - X}{M1 - X_{\min}} \right)^{s_L} & ; X_{\min} < X < M1 \\ 1 & ; M1 \leq X \leq M2 \\ 1 - \left(\frac{X - M2}{X_{\max} - M2} \right)^{s_R} & ; M2 < X < X_{\max} \\ 0 & ; \text{aksihalde} \end{cases} \quad (4.3)$$

Burada X_{\min} , veri aralığının alt sınırını; X_{\max} , veri aralığının üst sınırını göstermektedir.

Teorem 3’de bulanık sayı ile veri dağılımı arasında yapılan eşleştirmeler Şekil 4.2’de görsel olarak yer almaktadır.

Şekil 4.2: Teorem 3’ün Görsel Gösterimi



Teorem 3: Veri dağılımının $[M1, M2]$ aralığında 1 üyelik derecesine, X_{\min} ve X_{\max} veri noktalarında 0 üyelik derecesine, sırasıyla $L_A(\alpha_d) \neq M1$ ve $R_A(\alpha_d) \neq M2$ koşulunu sağlayan sol (X_i) ve sağ taraf (X_j) veri noktalarında α_D üyelik derecesine sahip olan, (4.3) üyelik fonksiyonun parametreleri aşağıdaki gibidir:

$$L_A(0) = X_{\min},$$

$$R_A(0) = X_{\max},$$

$$M1 = \text{Mod1},$$

$$M2 = \text{Mod2},$$

$$s_L = \frac{\text{Log}(1 - \alpha_d)}{\text{Log}\left(\frac{M1 - L_A(\alpha_d)}{M1 - L_A(0)}\right)}, \quad s_R = \frac{\text{Log}(1 - \alpha_d)}{\text{Log}\left(\frac{R_A(\alpha_d) - M2}{R_A(0) - M2}\right)}. \quad (4.4)$$

İspat: Aşağıdaki eşitlikler (2.6)- (2.8) formüllerinde yerine konulduğunda,

$$L_A(1) = M1$$

$$R_A(1) = M2$$

$$L_A(0) = X_{\min}$$

$$R_A(0) = X_{\max}$$

$$\sigma_A = M1 - X_{\min}$$

$$\beta_A = X_{\max} - M2,$$

(4.4)'deki formüller kolayca elde edilebilir.

Not: Teorem 3'de, $[M1, M2]$ mod aralığında bulunan veri noktalarının üyelik dereceleri 1 olarak alınmaktadır; bu nedenle Teorem 3'ün, iki modu arasındaki frekansları birbirine yakın olan dağılımlar için kullanılması daha uygundur. Hem Teorem 2 hem Teorem 3'de, veri dağılımının uzun kuyrukları daha küçük yayılma elde edebilmek adına dikkate alınmayabilir.

4.2 Frekans Tablosuna Dayanan Bulanık Sayı Türetme

Önceki teoremlerde, bulanık sayı türetiminde veri dağılımının sadece beş noktası dikkate alınmaktadır. Veri dağılımının beş noktası yerine daha çok niteliğinden yararlanılmak istenilirse, frekans tablolarının kullanımı alternatif bir yol olarak düşünülebilir. Böylece dağılımın grafiği ile uyumlu bulanık sayı türetilmiş olunur. Bu noktadan yola çıkarak, bu bölümde frekans tablo bilgilerini kullanan yeni üçgen, parametrik üçgen ve üssel bulanık sayı elde edilme teknikleri ortaya atıldı ve yapılan işlemlerde X_1, X_2, \dots, X_N veri setinin frekans tablosunun, aşağıdaki satır ve sütun bilgilerini kapsadığı varsayıldı.

Tablo 4.1: Frekans Tablosu

Sınıf Aralığı	Orta Nokta	Frekans	Yüzdellik
$X_1 - X_2$	$M_1 = (X_1 + X_2)/2$	f_1	$p_1 = f_1 / N$
$X_2 - X_3$	$M_2 = (X_2 + X_3)/2$	f_2	$p_2 = f_2 / N$
...
$X_{k-1} - X_k$	$M_k = (X_{k-1} + X_k)/2$	f_k	$p_k = f_k / N$
Toplam		N	1

Bu tabloda mod sınıf aralığı, maksimum yüzdellik değerine sahip sınıf aralığını göstermektedir.

4.2.1 Frekans Tablosuna Dayanan Üçgen ve Parametrik Üçgen Bulanık Sayı Türetme

Bu bölümde, sınıf aralıklarının normalleştirilmiş yüzdellikleri (\tilde{p}_i) ve orta-nokta değerleri kullanılarak (M_i) üçgen ve parametrik üçgen bulanık sayı türetme teknikleri ortaya atıldı. İlgili teknikler

$$\mu = \begin{cases} 1 - \left(\frac{M - X}{\sigma} \right)^{s_L} & ; M - \sigma < X \leq M \\ 1 - \left(\frac{X - M}{\beta} \right)^{s_R} & ; M < X < M + \beta \\ 0 & ; \text{aksihalde} \end{cases} \quad (4.5)$$

bulanık sayı tanımı için geçerlidir. Burada M frekans tablosunda maksimum yüzdeye sahip sınıf aralığının orta noktasına göstermektedir.

İlk etapta, söz konusu tekniklerde (Teorem 4 - Teorem 5), 1 üyelik derecesi mod sınıf aralığının orta noktasına (M) atanır ve frekans tablosundaki diğer yüzdelikler aşağıdaki eşitlik yardımıyla normalleştirilirler,

$$\tilde{p}_i = \frac{p_i}{p_m}. \quad (4.6)$$

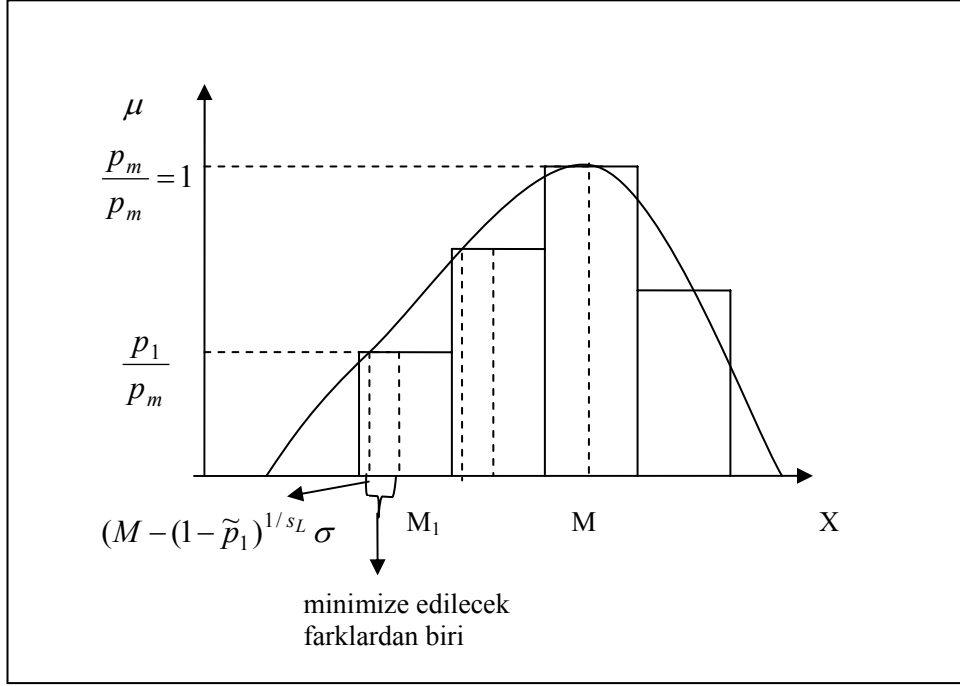
Burada, p_i ilgili sınıf aralığının yüzdesini, p_m ise frekans tablosundaki maksimum yüzde değerini göstermektedir.

Daha sonraki etapta, (4.7)'deki amaç fonksiyonu, (4.5)'in bilinmeyen s_L , σ ve s_R , β parametrelerine göre minimize edilerek veri dağılıma yakınsayan üçgen ve parametrik üçgen bulanık sayı türetilir.

$$D = \sum_{i=1}^{m-1} \left(M_i - (M - (1 - \tilde{p}_i)^{1/s_L} \sigma) \right)^2 + \sum_{i=m+1}^k \left(M_i - (M + (1 - \tilde{p}_i)^{1/s_R} \beta) \right)^2 \quad (4.7)$$

Amaç fonksiyonunun minimizasyonunda kullanılan kriterler görsel olarak Şekil 4.3'te yer almaktadır. Şekilden de anlaşılacağı üzere, bulanık sayının sol tarafı için $L(\tilde{p}_i)$ ve M_i arasındaki farkların karelerinin toplamları ile, bulanık sayının sağ tarafı için $R(\tilde{p}_i)$ ve M_i arasındaki farkların karelerinin toplamları minimize edilmektedir.

Şekil 4.3: Teorem 4’de Yer Alan Amaç Fonksiyonunun Görsel Olarak Gösterimi



Teorem 4: (4.7)’deki amaç fonksiyonunu minimize eden, frekans tablosu ile uyumlu parametrik üçgen bulanık sayı elde edebilmek için, (4.5)’de yer alan üyelik fonksiyonunun parametreleri aşağıdaki eşitlikleri sağlamalıdır.

$$M = Mod ,$$

$$\sigma = \frac{\sum_{i=1}^{m-1} (M - M_i)(1 - \tilde{p}_i)^{1/SL}}{\sum_{i=1}^{m-1} (1 - \tilde{p}_i)^{2/SL}}, \quad \beta = \frac{\sum_{i=m+1}^k (M_i - M)(1 - \tilde{p}_i)^{1/SR}}{\sum_{i=1+m}^k (1 - \tilde{p}_i)^{2/SL}},$$

$$\begin{aligned} & \sum_{i=1}^{m-1} (M_i - M)(1 - \tilde{p}_i)^{1/SL} \text{Log}(1 - \tilde{p}_i) \\ & + \frac{\sum_{i=1}^{m-1} (M - M_i)(1 - \tilde{p}_i)^{1/SL}}{\sum_{i=1}^{m-1} (1 - \tilde{p}_i)^{2/SL}} \sum_{i=1}^{m-1} (1 - \tilde{p}_i)^{2/SL} \text{Log}(1 - \tilde{p}_i) = 0 \end{aligned},$$

$$\begin{aligned} & \sum_{i=m+1}^k (M_i - M)(1 - \tilde{p}_i)^{1/s_R} \text{Log}(1 - \tilde{p}_i) \\ & - \frac{\sum_{i=m+1}^k (M_i - M)(1 - \tilde{p}_i)^{1/s_R}}{\sum_{i=1+m}^k (1 - \tilde{p}_i)^{2/s_R}} \sum_{i=m+1}^k (1 - \tilde{p}_i)^{2/s_R} \text{Log}(1 - \tilde{p}_i) = 0 \end{aligned} \quad (4.8)$$

Burada *Mod*, maksimum yüzde değerinin yer aldığı mod sınıf aralığının orta noktasını göstermektedir.

İspat: (4.7)'deki amaç fonksiyonunu minimize edebilmek için

$$\frac{\partial D}{\partial \sigma} = 0, \quad \frac{\partial D}{\partial s_L} = 0, \quad \frac{\partial D}{\partial \beta} = 0, \quad \frac{\partial D}{\partial s_R} = 0$$

eşitliklerinin sağlanması gerekmektedir.

Bu amaçla aşağıdaki eşitlik elde edilir.

$$\frac{\partial D}{\partial \sigma} = 2 \sum_{i=1}^{m-1} \left(M_i - (M - (1 - \tilde{p}_i)^{1/s_L} \sigma) \right) (1 - \tilde{p}_i)^{1/s_L} = 0$$

Buradan

$$\sum_{i=1}^{m-1} (M_i - M)(1 - \tilde{p}_i)^{1/s_L} + \sum_{i=1}^{m-1} \sigma (1 - \tilde{p}_i)^{2/s_L} = 0$$

elde edilir ve son olarak aşağıdaki σ formülüne ulaşılır.

$$\sigma = \frac{\sum_{i=1}^{m-1} (M - M_i)(1 - \tilde{p}_i)^{1/s_L}}{\sum_{i=1}^{m-1} (1 - \tilde{p}_i)^{2/s_L}}$$

s_L 'yi elde edebilmek için,

$$\frac{\partial D}{\partial s_L} = -\frac{2\sigma}{s_L^2} \sum_{i=1}^{m-1} \left(M_i - (M - (1 - \tilde{p}_i)^{1/s_L} \sigma) \right) (1 - \tilde{p}_i)^{1/s_L} \text{Log}(1 - \tilde{p}_i) = 0$$

eşitliği çözülür ve buradan

$$\frac{\partial D}{\partial s_L} = \sum_{i=1}^{m-1} \left[(M_i - M)(1 - \tilde{p}_i)^{1/s_L} \text{Log}(1 - \tilde{p}_i) + \sigma (1 - \tilde{p}_i)^{2/s_L} \text{Log}(1 - \tilde{p}_i) \right] = 0$$

bulunur. Yukarıdaki eşitlikte σ yerine koyularak aşağıdaki eşitlik elde edilir.

$$\sum_{i=1}^{m-1} (M_i - M)(1 - \tilde{p}_i)^{1/s_L} \text{Log}(1 - \tilde{p}_i) + \frac{\sum_{i=1}^{m-1} (M - M_i)(1 - \tilde{p}_i)^{1/s_L}}{\sum_{i=1}^{m-1} (1 - \tilde{p}_i)^{2/s_L}} \sum_{i=1}^{m-1} (1 - \tilde{p}_i)^{2/s_L} \text{Log}(1 - \tilde{p}_i) = 0$$

Benzer yolla,

$$\beta = \frac{\sum_{i=m+1}^k (M_i - M)(1 - \tilde{p}_i)^{1/s_R}}{\sum_{i=m+1}^k (1 - \tilde{p}_i)^{2/s_R}}$$

ve

$$\sum_{i=m+1}^k (M_i - M)(1 - \tilde{p}_i)^{1/s_R} \text{Log}(1 - \tilde{p}_i) - \frac{\sum_{i=m+1}^k (M_i - M)(1 - \tilde{p}_i)^{1/s_R}}{\sum_{i=m+1}^k (1 - \tilde{p}_i)^{2/s_R}} \sum_{i=m+1}^k (1 - \tilde{p}_i)^{2/s_R} \text{Log}(1 - \tilde{p}_i) = 0$$

eşitlikleri sırasıyla $\frac{\partial D}{\partial \beta} = 0$, $\frac{\partial D}{\partial s_R} = 0$ eşitlikleri kullanılarak bulunur.

Sonuç 4.1: (4.7)'deki amaç fonksiyonunu minimize eden, frekans tablosu ile uyumlu üçgen bulanık sayı elde edebilmek için, (4.5)'de yer alan üyelik fonksiyonunun parametreleri aşağıdaki eşitlikleri sağlamalıdır.

$$M = \text{Mod} ,$$

$$s_L = 1, s_R = 1,$$

$$\sigma = \frac{\sum_{i=1}^{m-1} (M - M_i)(1 - \tilde{p}_i)}{\sum_{i=1}^{m-1} (1 - \tilde{p}_i)^2}, \beta = \frac{\sum_{i=m+1}^k (M_i - M)(1 - \tilde{p}_i)}{\sum_{i=m+1}^k (1 - \tilde{p}_i)^2} . \quad (4.9)$$

Burada *Mod*, maksimum yüzde değerinin yer aldığı mod sınıf aralığının orta noktasını göstermektedir.

4. 2. 2 Frekans Tablosuna Dayanan Parametrik Üssel Bulanık Sayı

Türetme

Bazen veri dağılımı üssel bir yapı sergileyebilir ve araştırmacılar bulanık sayıda parametrik üçgen form yerine üssel form kullanmayı tercih edebilir. Bu durumu dikkate alarak, bu bölümde frekans tablosuna dayanan yeni bir üssel bulanık sayı türetme tekniği ortaya atıldı. İlgili teknik

$$\mu(X) = \begin{cases} e^{-\left(\frac{M-X}{\sigma}\right)^{s_L}} & ; X \leq M \\ e^{-\left(\frac{X-M}{\beta}\right)^{s_R}} & ; M < X \\ 0 & ; \text{aksihalde} \end{cases} \quad (4.10)$$

bulanık sayı tanımı için geçerlidir.

Diğer bölümden farklı olarak amaç fonksiyon, frekans tablosundaki sınıf aralıklarının orta noktalarına karşılık gelen normalleştirilmiş yüzdeler ile, o noktalara karşılık gelen bulanık sayının seviyeleri arasındaki farkın minimize edilmesine dayandırıldı ve bu amaçla,

$$\tilde{p}_i = \exp\left(-\left(\frac{M - M_i}{\sigma}\right)^{s_L}\right), \tilde{p}_i = \exp\left(-\left(\frac{M_i - M}{\beta}\right)^{s_R}\right) \quad (4.11)$$

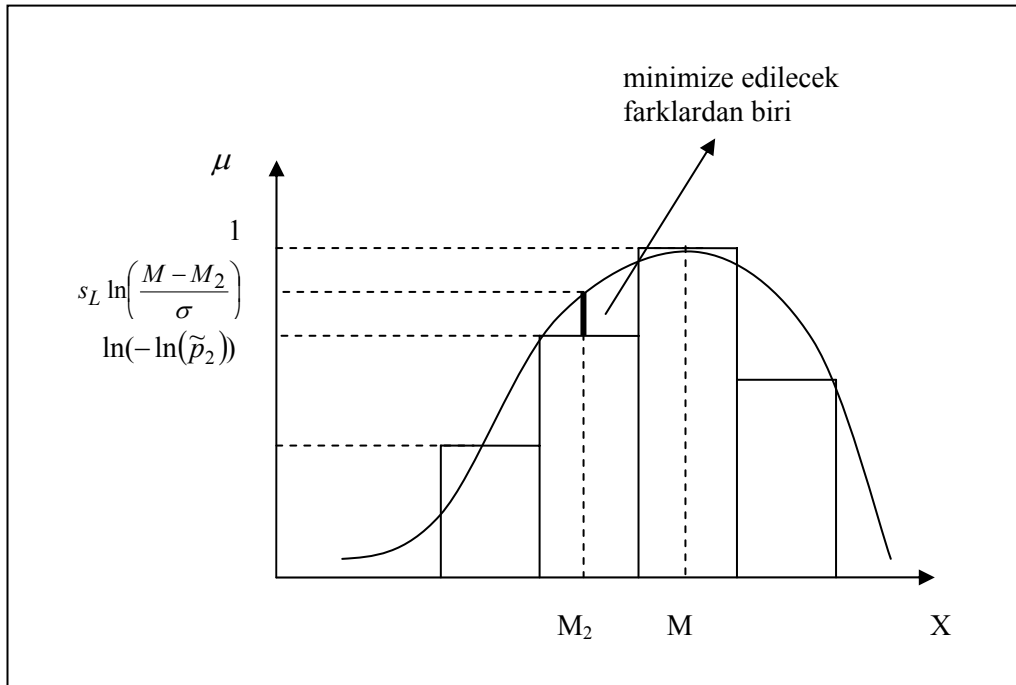
eşitlikleri (4.12)'deki amaç fonksiyonun kurulumunda dikkate alındı.

$$f(\sigma, \beta, s_L, s_R) = \sum_{i=1}^{m-1} \left(\ln(-\ln(\tilde{p}_i)) - s_L \ln\left(\frac{M - M_i}{\sigma}\right) \right)^2 + \sum_{i=m+1}^k \left(\ln(-\ln(\tilde{p}_i)) - s_R \ln\left(\frac{M_i - M}{\beta}\right) \right)^2 \quad (4.12)$$

(4.12)'de yer alan amaç fonksiyonunun minimizasyonundan önce, ortaya atılan diğer teoremlerde olduğu üzere maksimum yüzdeliğe sahip sınıf aralığının orta noktası bulanık sayıda 1 üyelik derecesine karşılık getirilir ve yüzdelikler (4.6)'daki formül kullanılarak normalleştirilir. Daha sonra (4.10)'daki üssel bulanık sayının bilinmeyen parametreleri (4.12)'deki amaç fonksiyonunun s_L , σ , s_R , β 'lere göre minimize edilmesi ile bulunur.

Amaç fonksiyonunun minimizasyonunda kullanılan kriterler görsel olarak Şekil 4.4'te gösterildiği gibidir.

Şekil 4.4: Teorem 5'de Yer Alan Amaç Fonksiyonun Görsel Olarak Gösterimi



Teorem 5: (4.12)'deki amaç fonksiyonunu minimize eden, frekans tablosu ile uyumlu parametrik üssel bulanık sayı elde edebilmek için, (4.10)'de yer alan üyelik fonksiyonunun parametreleri aşağıdaki eşitlikleri sağlamalıdır.

$$\begin{aligned}
s_L &= \frac{\sum_{i=1}^{m-1} \ln(-\ln(\tilde{p}_i))}{\sum_{i=1}^{m-1} \ln\left(\frac{M - M_i}{\sigma}\right)}, \\
\sigma &= \exp\left(\frac{\sum_{i=1}^{m-1} \ln(-\ln(\tilde{p}_i)) \sum_{i=1}^{m-1} (\ln(M - M_i))^2 - \sum_{i=1}^{m-1} (\ln(-\ln(\tilde{p}_i)) \ln(M - M_i)) \sum_{i=1}^{m-1} \ln(M - M_i)}{\sum_{i=1}^{m-1} \ln(-\ln(\tilde{p}_i)) \sum_{i=1}^{m-1} \ln(M - M_i) - (m-1) \sum_{i=1}^{m-1} (\ln(-\ln(\tilde{p}_i)) \ln(M - M_i))}\right), \\
s_R &= \frac{\sum_{i=m+1}^k \ln(-\ln(\tilde{p}_i))}{\sum_{i=m+1}^k \ln\left(\frac{M_i - M}{\beta}\right)}, \\
\beta &= \exp\left(\frac{\sum_{i=m+1}^k \ln(-\ln(\tilde{p}_i)) \sum_{i=m+1}^k (\ln(M_i - M))^2 - \sum_{i=m+1}^k (\ln(-\ln(\tilde{p}_i)) \ln(M_i - M)) \sum_{i=m+1}^k \ln(M_i - M)}{\sum_{i=m+1}^k \ln(-\ln(\tilde{p}_i)) \sum_{i=m+1}^k \ln(M_i - M) - (k-m) \sum_{i=m+1}^k (\ln(-\ln(\tilde{p}_i)) \ln(M_i - M))}\right),
\end{aligned} \tag{4.13}$$

Not: Teoremde,

$$\sum_{i=m+1}^k \ln(-\ln(\tilde{p}_i)) \sum_{i=m+1}^k \ln(M_i - M) \neq (k-m) \sum_{i=m+1}^k (\ln(-\ln(\tilde{p}_i)) \ln(M_i - M)) \tag{4.14}$$

ve

$$\sum_{i=1}^{m-1} \ln(-\ln(\tilde{p}_i)) \sum_{i=1}^{m-1} \ln(M - M_i) \neq (m-1) \sum_{i=1}^{m-1} (\ln(-\ln(\tilde{p}_i)) \ln(M - M_i)) \tag{4.15}$$

koşullarının sağlanması gerekmektedir.

İspat: (4.10)'da yer alan üyelik fonksiyonunun seviyelerinin normalleştirilmiş \tilde{p}_i seviyelerine yakınsatabilmek için (4.12)'de yer alan amaç fonksiyonlarının bilinmeyen s_L , σ , s_R , β parametrelere göre minimize edilmesi gerekmektedir. Bunun için

$$\frac{\partial f}{\partial s_L} = 0, \quad \frac{\partial f}{\partial \sigma} = 0, \quad \frac{\partial f}{\partial s_R} = 0, \quad \frac{\partial f}{\partial \beta} = 0$$

eşitlikleri sağlanmalıdır.

s_L ve σ 'yı bulabilmek için aşağıdaki eşitliklerinin çözülmesi gerekir.

$$\frac{\partial f}{\partial \sigma} = -2 \sum_{i=1}^{m-1} \left(\ln(-\ln(\tilde{p}_i)) - s_L \ln\left(\frac{M - M_i}{\sigma}\right) \right) \frac{s_L}{\sigma} = 0$$

$$\frac{\partial f}{\partial s_L} = -2 \sum_{i=1}^{m-1} \left(\ln(-\ln(\tilde{p}_i)) - s_L \ln\left(\frac{M - M_i}{\sigma}\right) \right) \ln\left(\frac{M - M_i}{\sigma}\right) = 0$$

Buradan

$$\sum_{i=1}^{m-1} \ln(-\ln(\tilde{p}_i)) - s_L \sum_{i=1}^{m-1} \ln\left(\frac{M - M_i}{\sigma}\right) = 0$$

ve

$$\sum_{i=1}^{m-1} \ln(-\ln(\tilde{p}_i)) \ln\left(\frac{M - M_i}{\sigma}\right) - s_L \sum_{i=1}^{m-1} \left(\ln\left(\frac{M - M_i}{\sigma}\right) \right)^2 = 0$$

eşitlikleri sırayla elde edilir. İşlemlerin sonunda

$$s_L = \frac{\sum_{i=1}^{m-1} \ln(-\ln(\tilde{p}_i))}{\sum_{i=1}^{m-1} \ln\left(\frac{M - M_i}{\sigma}\right)}$$

ve

$$\ln \sigma = \frac{\sum_{i=1}^{m-1} \ln(-\ln(\tilde{p}_i)) \sum_{i=1}^{m-1} (\ln(M - M_i))^2 - \sum_{i=1}^{m-1} (\ln(-\ln(\tilde{p}_i)) \ln(M - M_i)) \sum_{i=1}^{m-1} \ln(M - M_i)}{\left(\sum_{i=1}^{m-1} \ln(-\ln(\tilde{p}_i)) \sum_{i=1}^{m-1} \ln(M - M_i) - (m-1) \sum_{i=1}^{m-1} (\ln(-\ln(\tilde{p}_i)) \ln(M - M_i)) \right)}$$

bulunur. Buradan σ aşağıdaki gibi tekrar yazılabilir,

$$\sigma = \exp \left(\frac{\sum_{i=1}^{m-1} \ln(-\ln(\tilde{p}_i)) \sum_{i=1}^{m-1} (\ln(M - M_i))^2 - \sum_{i=1}^{m-1} (\ln(-\ln(\tilde{p}_i)) \ln(M - M_i)) \sum_{i=1}^{m-1} \ln(M - M_i)}{\sum_{i=1}^{m-1} \ln(-\ln(\tilde{p}_i)) \sum_{i=1}^{m-1} \ln(M - M_i) - (m-1) \sum_{i=1}^{m-1} (\ln(-\ln(\tilde{p}_i)) \ln(M - M_i))} \right).$$

Benzer yolla, aşağıdaki eşitlikler eşanlı olarak çözülürler,

$$\frac{\partial f}{\partial \beta} = -2 \sum_{i=m+1}^k \left(\ln(-\ln(\tilde{p}_i)) - s_R \ln\left(\frac{M_i - M}{\beta}\right) \right) \frac{s_R}{\beta} = 0,$$

$$\frac{\partial f}{\partial s_R} = -2 \sum_{i=1}^{m-1} \left(\ln(-\ln(\tilde{p}_i)) - s_R \ln\left(\frac{M_i - M}{\beta}\right) \right) \ln\left(\frac{M_i - M}{\beta}\right) = 0$$

ve sırasıyla,

$$s_R = \frac{\sum_{i=m+1}^k \ln(-\ln(\tilde{p}_i))}{\sum_{i=m+1}^k \ln\left(\frac{M_i - M}{\beta}\right)},$$

$$\ln \beta = \frac{\sum_{i=m+1}^k \ln(-\ln(\tilde{p}_i)) \sum_{i=m+1}^k (\ln(M_i - M))^2 - \sum_{i=m+1}^k (\ln(-\ln(\tilde{p}_i)) \ln(M_i - M)) \sum_{i=m+1}^k \ln(M_i - M)}{\sum_{i=m+1}^k \ln(-\ln(\tilde{p}_i)) \sum_{i=m+1}^k \ln(M_i - M) - (k - m) \sum_{i=m+1}^k (\ln(-\ln(\tilde{p}_i)) \ln(M_i - M))}$$

elde edilir. Buradan β

$$\beta = \exp \left(\frac{\sum_{i=m+1}^k \ln(-\ln(\tilde{p}_i)) \sum_{i=m+1}^k (\ln(M_i - M))^2 - \sum_{i=m+1}^k \ln(-\ln(\tilde{p}_i)) \ln(M_i - M) \sum_{i=m+1}^k \ln(M_i - M)}{\sum_{i=m+1}^k \ln(-\ln(\tilde{p}_i)) \sum_{i=m+1}^k \ln(M_i - M) - (k - m) \sum_{i=m+1}^k (\ln(-\ln(\tilde{p}_i)) \ln(M_i - M))} \right)$$

şeklinde tekrar yazılabilir.

5 BİSPEKTRAL ENDEKS ZAMAN SERİLERİNİN *K* - EN YAKIN KOMŞULUK KURALINA DAYANAN ZAMAN SERİSİ ETİKETLENDİRME YÖNTEMLERİ İLE ETİKETLENDİRİLMESİ

Bispektral endeks, sakinleştirici ilaç ve anestezinin uygulanmasında beyin durumunun izlenmesi için kullanılır (Kreuer,2001;Gan, 1997). İstatistiksel olarak zaman ve frekans tabanlı, yüksek dereceli spektral alt-parametreleri içeren kompleks parametrelere dayanır (Vretzakis ve diğerleri, 2005). Analizi farklı EEG tanımlarını içerir ve algoritmasında insan hatası kabul etmeyen ileri teknikli kompleks formüller kullanılır (Johansen, 2006). Bu endeks uyku seviyesi veya hipnoz seviyesi ile yüksek ilişkilidir ve 0 ile 100 arasında değer alır (Elina ve diğerleri, 2006; Özgören ve diğerleri, 2008). BIS endeksin “100” civarında olması uyanıklığı, “0” olması ölüm durumunu göstermektedir (Agarwal ve Griffiths, 2004). Endekste yetmiş-doksan aralığı hafif sedasyonu, yetmişler yoğun sedasyonu, kırk-yetmiş aralığı genel anestesiye, daha düşük değerler ağır hipotonik durumu göstermektedir (Misis ve diğerleri, 2008).

Bireylerin durumları sadece BIS endeks değerleri ile değil ayrıca klinik araştırmalarla da değerlendirilmektedir. Bununla birlikte, Özgören (2008) bispektral endeks değerlerinin non-REM uyku düzeyleri ile yüksek ilişkili olduğunu göstermiştir. Bu çalışmada, söz konusu ilişkiden yola çıkarak, bireylerin durumları bölüm 3’te önerilen *K* - En Yakın Komşuluk Kuralına Dayanan Zaman Serisi Etiketlendirme Yöntemleri ile tahmin edilmeye çalışıldı. Hesaplamalarda sadece bireylere ait bispektral endeks değerleri kullanıldı.

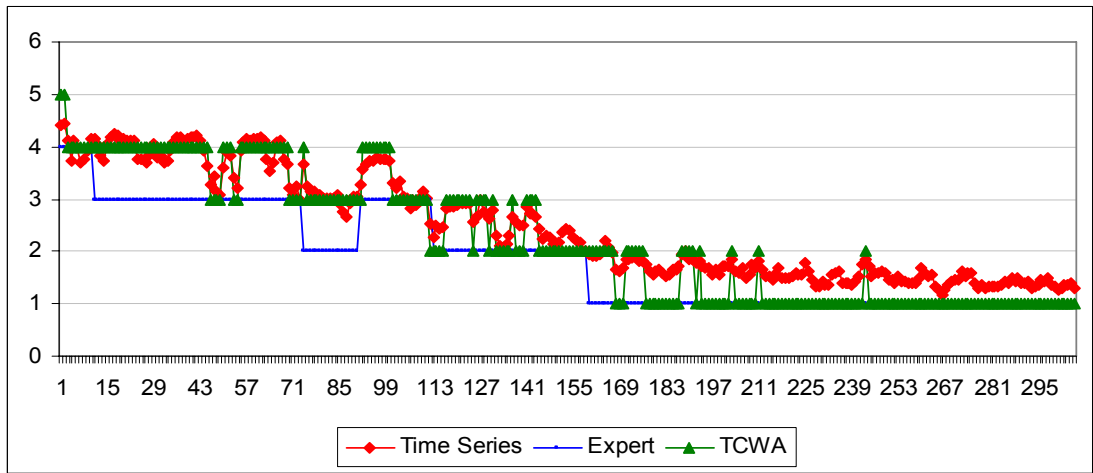
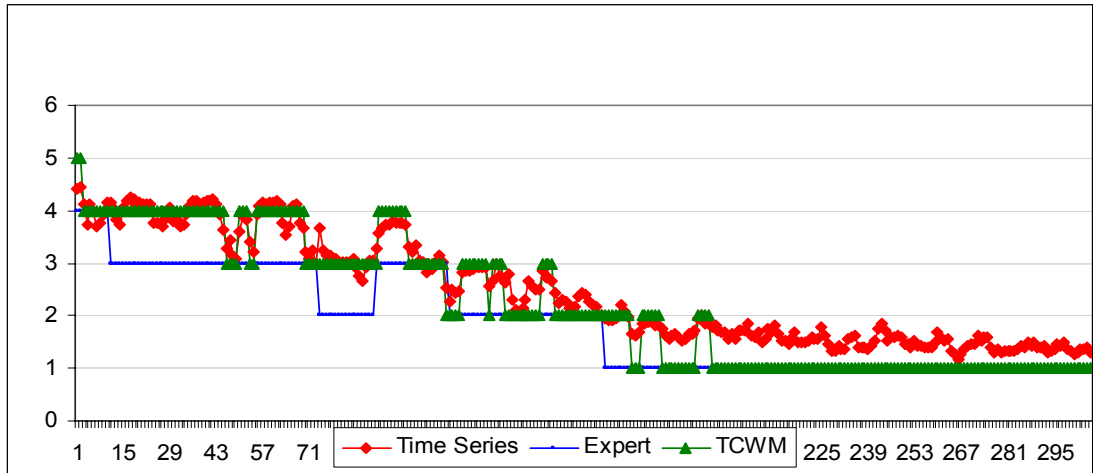
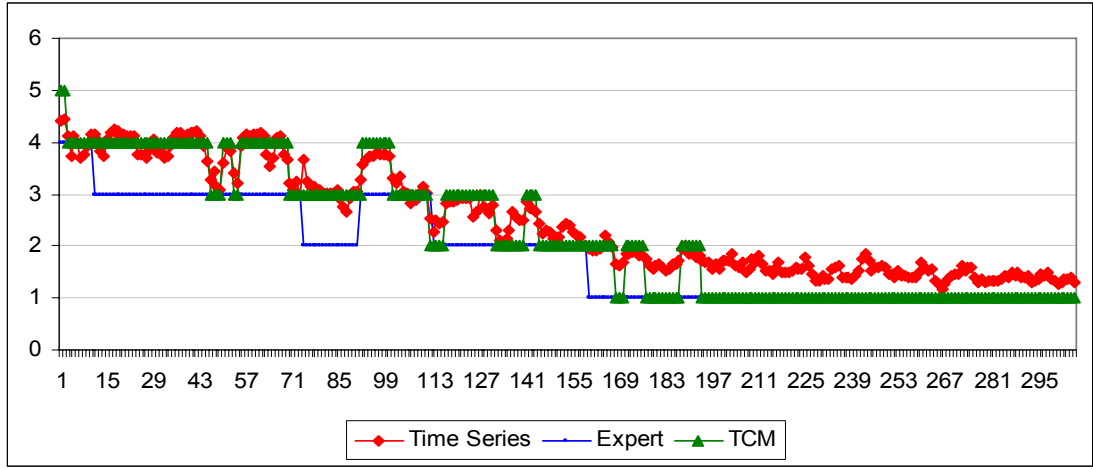
Önerilen algoritmalar, Dr. Murat Özgeren’den alınan veri setleri üzerinde çalıştırıldı ve çıkan sonuçların tutarlı olup olmadığı uzman kararları ile karşılaştırıldı. Algoritmalarından elde edilen tahminlerin hata payını bulmak için, gözlemlenen ile beklenen sınıf etiketleri arasındaki farkların kareler toplamının kara kökü hesaplandı ve Tablo 5.1’de gösterildi. Her bir algoritmaya ait ortalama hata miktarı ise Tablo 5.1’in son satırında belirtildi.

Tablo 5.1: Bölüm 3’te Önerilen Algoritmaların Hata Miktarları

Veri Seti	FCML	TCM		TCWM		TCWA	
	HKT	HKT	k	HKT	K	HKT	k
1	12.6095	11.8743	2	11.619	3	12.53	2
2	14.5945	14.4914	1	14.4568	2	14.4914	10
3	16.5227	15.0333	4	14.8661	5	16.4924	4
4	12.2066	11.8743	1	11.5326	4	12.2066	0
5	15.1987	13.9642	7	14.1421	5	15.1327	2
6	8.6603	6.0828	8	6.0828	10	8.3666	7
7	14.1774	13.8564	4	14.0357	3	14.1774	0
8	12.2882	11.619	3	11.3578	8	11.9583	8
9	11.0454	11.0454	0	10.8167	1	11.0454	0
10	11.1803	11	1	10.7703	1	11.1803	0
11	18.2209	13.3791	10	16.6733	10	18.0555	10
12	18.2757	17.7482	3	17.1464	10	18.2757	0
13	11.7047	11.2694	2	11.2694	3	11.7047	0
14	18.1659	18.1659	0	18	2	17.7764	10
15	14.3527	14.2829	5	14.3527	0	14.1421	6
16	8	6.9282	7	6.7082	10	7.4833	7
17	8.3666	5.5678	9	6.6332	5	8	9
18	18.8149	18.5203	6	18.412	4	18.8149	0
19	19.1833	18.4662	4	18.303	4	18.9737	10
20	12.8062	12.1655	4	11.9164	6	12.8062	0
21	13.3041	12.2066	3	11.9164	5	13.3041	0
Ortalama	13.7942	12.8353		12.9053		13.6627	
Stand. S.	3.4576651	3.7105351		3.6914081		3.5065441	

Bu tabloda herbir veri seti için hangi $k = K / 2$ komşuluğunda minimum hata miktarına ulaşıldığı k sütununda gösterildi; algoritmalar arasındaki en iyi sonuçlar ise sarı renkte işaretlendi. Tablo incelendiğinde TCWM algoritmasının 13 veri setinde, TCM algoritmasının 4 veri setinde, TCWA algoritmasının 2 veri setinde tek başına minimum hata ürettiği görüldü. 2 veri setinde TCWM ve TCM algoritmaları aynı hata miktarına sahiplerdir. Yöntemlerin 10. veri seti üzerinde verdiği sonuçlar bir örnek olarak Şekil 5.1’de gösterildi.

Şekil 5.1: TCM, TCWM ve TCWA Algoritmalarının ve Uzman Görüşünün Çıktıları



Algoritmalarla ait sınıflama kesinlikleri aşağıdaki oran (Nasibov ve Ulutagay, 2010) kullanılarak hesaplandı ve Tablo 5.2 oluşturuldu. Tabloda maksimum sınıflama kesinlikleri sarı renkte işaretlendi.

$$SK = \frac{\text{Seride doğru tahmin edilen nokta sayısı}}{\text{Serideki toplam nokta sayısı}}$$

Tablo 5.2: FCML, TCM, TCWM, TCWA Algoritmalarının Sınıflama Kesinlikleri

Veri Seti	FCML		TCM		TCWM		TCWA	
	Doğru Tahmin Sayısı	SK	Doğru Tahmin Sayısı	SK	Doğru Tahmin Sayısı	SK	Doğru Tahmin Sayısı	SK
1	168	0.549	177	0.578	177	0.578	170	0.556
2	162	0.529	165	0.539	166	0.542	165	0.539
3	99	0.324	116	0.379	121	0.395	100	0.327
4	199	0.65	207	0.676	209	0.683	199	0.650
5	144	0.471	147	0.480	163	0.533	146	0.477
6	231	0.755	269	0.879	269	0.879	236	0.771
7	125	0.408	132	0.431	121	0.395	126	0.412
8	173	0.565	189	0.618	177	0.578	181	0.592
9	186	0.608	187	0.611	189	0.618	187	0.611
10	184	0.601	185	0.605	190	0.621	184	0.601
11	152	0.497	160	0.523	147	0.480	137	0.448
12	116	0.379	108	0.353	126	0.412	116	0.379
13	214	0.699	215	0.703	215	0.703	214	0.699
14	66	0.216	66	0.216	66	0.216	80	0.261
15	175	0.572	195	0.637	175	0.572	181	0.592
16	251	0.82	258	0.843	261	0.853	259	0.846
17	236	0.771	275	0.899	262	0.856	242	0.791
18	30	0.098	29	0.095	21	0.069	30	0.098
19	28	0.092	34	0.111	31	0.101	36	0.118
20	148	0.578	161	0.526	170	0.556	148	0.484
21	177	0.578	193	0.631	200	0.654	177	0.578
Ort.	158.048	0.516	165.143	0.540	164.571	0.538	157.810	0.516
S.s.	59.206	0.193	66.073	0.216	65.802	0.215	59.950	0.196

Tablo 5.2 incelendiğinde, TCWM algoritmasının 13 veri setinde, TCM algoritmasının 8 veri setinde, TCWA algoritmasının 3 veri setinde, FCML algoritmasının ise sadece 1 veri setinde maksimum sınıflama kesinliğine ulaştığı görüldü. Buna göre sınıflama etiketlerinde uygulanan KNN kuralının sınıflamada bir iyileştirmeye yol açtığı yorumu yapıldı.

TCM, TCWM ve TCWA yöntemlerinin ortalama sınıflama kesinliklerinin FCML yönteminin ortalama sınıflama kesinliğinden fazla olup olmadığı istatistiksel olarak test edilmek istendi ve aşağıdaki hipotezler kuruldu. Burada bölüm 3’te önerilen algoritmaların ortalama sınıf kesinlikleri arasında bir fark olup olmadığı da dikkate alındı.

Hipotez 5.1

$$H_0 : \mu_{TCM} = \mu_{FCML}$$

$$H_1 : \mu_{TCM} > \mu_{FCML}$$

Hipotez 5.3

$$H_0 : \mu_{TCWA} = \mu_{FCML}$$

$$H_1 : \mu_{TCWA} > \mu_{FCML}$$

Hipotez 5.5

$$H_0 : \mu_{TCM} = \mu_{TCWA}$$

$$H_1 : \mu_{TCM} \neq \mu_{TCWA}$$

Hipotez 5.2

$$H_0 : \mu_{TCWM} = \mu_{FCML}$$

$$H_1 : \mu_{TCWM} > \mu_{FCML}$$

Hipotez 5.4

$$H_0 : \mu_{TCM} = \mu_{TCWM}$$

$$H_1 : \mu_{TCM} \neq \mu_{TCWM}$$

Hipotez 5.6

$$H_0 : \mu_{TCWM} = \mu_{TCWA}$$

$$H_1 : \mu_{TCWM} \neq \mu_{TCWA}$$

SPSS paket programı ile ikili-t testi uygulandı ve Tablo 5.4’de yer alan analiz sonuçlarına ulaşıldı. Bu analizlerin hemen öncesinde ikili-t testi için gerekli olan normallik koşulunun sağlanıp sağlanmadığına bakıldı (Bakınız Tablo 5.3) ve yöntemlerin sınıflama kesinliklerinin normal dağılıma sahip olduğu görüldü (Sıfır hipotezi rededilemez).

H_0 : Sınıflama kesinlikleri normal dağılıma sahiptir.

H_1 : Sınıflama kesinlikleri normal dağılıma sahip değildir.

Tablo 5.3. FCML, TCM, TCWM, TCWA Yöntemlerinin Sınıflama Kesinliklerinin Normallik Testi Sonuçları

	Kolmogorov-Smirnov ^a		Shapiro-Wilk			
	İstatistik	s.d	İstatistik	s.d	İstatistik	s.d
FCML	0.141	21	0.200*	0.944	21	0.264
TCM	0.136	21	0.200*	0.950	21	0.339
TCWM	0.157	21	0.189	0.938	21	0.203
TCWA	0.118	21	0.200*	0.961	21	0.528

a. Lilliefors anlamlılık düzeltmesi

*. Bu gerçek anlamlılığın alt sınırındır.

Tablo 5.4 incelendiğinde Hipotez 5.4 haricinde diğer hipotezlerde yer alan sıfır hipotezlerinin rededilmelerine karar verildi ve bispektral endeks veri setleri için aşağıdaki sonuçlara ulaşıldı.

Sonuç 5.1: TCM yönteminin ortalama sınıflama kesinliği, FCML yönteminin ortalama sınıflama kesinliğinden fazladır.

Sonuç 5.2: TCWM yönteminin ortalama sınıflama kesinliği, FCML yönteminin ortalama sınıflama kesinliğinden fazladır.

Sonuç 5.3: TCWA yönteminin ortalama sınıflama kesinliği, FCML yönteminin ortalama sınıflama kesinliğinden fazladır.

Sonuç 5.4: TCM ve TCWM yöntemlerinin ortalama sınıflama kesinlikleri arasında fark yoktur.

Sonuç 5.5: TCM ve TCWA yöntemlerinin ortalama sınıflama kesinlikleri arasında fark vardır.

Sonuç 5.6: TCWM ve TCWA yöntemlerinin ortalama sınıflama kesinlikleri arasında fark vardır.

Tablo 5.4: TCM, TCWM, TCWA, FCML Yöntemlerinin İkili-t Testi Sonuçları

İkililer	Ortalama	Standart Sapma	Ort.'un	95% Güv.	95% Güv.	t	s.d	P
			Standart Hatası	Ar. Alt Sınırı	Ar. Üst Sınırı			
TCM- FCML	0.03175	0.03854	0.00841	0.01420	0.04929	3.774	20	0.001
TCWM - FCML	0.02988	0.03875	0.00846	0.01224	0.04752	3.533	20	0.002
TCWA - FCML	0.00778	0.01806	0.00394	-0.00044	0.01600	1.974	20	0.062
TCM- TCWM	0.00187	0.03141	0.00685	-0.01243	0.01616	0.272	20	0.788
TCM - TCWA	0.02397	0.03960	0.00864	0.00594	0.04199	2.773	20	0.012
TCWA - TCWM	-0.02210	0.04033	0.00880	-0.04046	-0.00374	-2.511	20	0.021

6 SINIF VERİ DAĞILIMLARINA YAKINSAYAN BULANIK ÜYELİK FONKSİYONLARININ OLUŞTURULMASI VE SINIFLAMADAKİ ETKİLERİNİN DEĞERLENDİRİLMESİ

Bu bölümde dördüncü bölümde ortaya atılan formüllerin bispektral endeks serilerinin sınıflanmasında bir iyileştirme yaratıp yaratmadığı kanıtlanmak istendi. Bu nedenle bölüm dörtte önerilen veri dağılımına yakınsayan üyelik fonksiyonları kullanarak bispektral endeks değerlerinin sınıflandırılmasına ve sınıflama kesinliklerinin karşılaştırılmasına karar verildi. Kıyaslama için daha önceki çalışmalarda (Nasibov ve Ulutagay, 2010) bispektral endeks değerlerinin üyelik fonksiyonlarının elde edilmesinde kullanılan (6.1) nolu formülün kullanılması uygun görüldü.

$$\mu_k(x) = e^{-\frac{1}{2} \left(\frac{x - \alpha_k}{\sigma_k} \right)^2}, k=1, \dots, 5. \quad (6.1)$$

$$\alpha_k = \frac{\sum x_i}{n_k}, \sigma_k = \sqrt{\frac{\sum (x_i - \alpha_k)^2}{n_k}}. \quad (6.2)$$

Önceki bölümde yer alan aynı veri setleri üzerinde çalışıldı ve ilk etapta uyku düzeylerinin üyelik fonksiyonları (6.1) nolu formül ile aşağıdaki gibi hesaplandı,

$$\mu_1 = e^{-\frac{1}{2} \left(\frac{0.31121 - X}{0.055421} \right)^2}, \mu_2 = e^{-\frac{1}{2} \left(\frac{0.49281 - X}{0.122669} \right)^2}, \mu_3 = e^{-\frac{1}{2} \left(\frac{0.69081 - X}{0.128966} \right)^2}$$
$$\mu_4 = e^{-\frac{1}{2} \left(\frac{0.83307 - X}{0.088472} \right)^2}, \mu_5 = e^{-\frac{1}{2} \left(\frac{0.88617 - X}{0.084955} \right)^2}.$$

Teorem 2 kullanılarak uyku düzeylerine ait parametrik üçgen bulanık sayı elde edebilmek için 21 veri seti birleştirildi ve beş farklı uyku düzeyine ait bispektral endeks değerleri, kendi sınıfları içerisinde toplandı. Bu sınıflara ait tanımlayıcı istatistikler SPSS paket programı kullanılarak Tablo 6.1'deki gibi bulundu.

Tablo 6.1: Bispektral Endeks Uyku Düzeylerinin Tanımlayıcı İstatistikleri

	Uyku Düzeyi 1	Uyku Düzeyi 2	Uyku Düzeyi 3	Uyku Düzeyi 4	Uyku Düzeyi 5
N	833	1742	2694	737	726
Ortalama	0.31121	0.49281	0.69081	0.83307	0.88617
Standard Hata	0.001921	0.002940	0.002485	0.003261	0.003155
Medyan	0.30400	0.46700	0.70800	0.84500	0.90300
Mod	0.262	0.409	0.818	0.871	0.911
Std. Sapma	0.055454	0.122704	0.128990	0.088532	0.085014
Varyans	0.003	0.015	0.017	0.008	0.007
Çarpıklık	1.249	0.557	-0.452	-1.359	-3.550
Çarpıklığın std. hatası	0.085	0.059	0.047	0.090	0.091
Basıklık	4.444	-0.352	-0.570	2.732	14.459
Basıklığın std. hatası	0.169	0.117	0.094	0.180	0.181
Aralık	0.527	0.758	0.691	0.540	0.587
Minimum	0.167	0.205	0.285	0.434	0.390
Maksimum	0.694	0.963	0.976	0.974	0.977
Yüzdeler					
5	0.24100	0.32800	0.44600	0.65880	0.77175
15	0.26010	0.37200	0.53700	0.74300	0.86200
50	0.30400	0.46700	0.70800	0.84500	0.90300
85	0.36700	0.63700	0.82375	0.91000	0.93800
95	0.40100	0.71800	0.86300	0.95210	0.96265

Farklı sınıflara ait verilerin normal dağılıma sahip olup olmadığını belirleyebilmek için Kolmogorov-Smirnov uyum iyiliği testine başvuruldu (Bakınız Tablo 6.2). Analiz sonucunda, sıfır hipotezleri 0.05 anlamlılık düzeyinde reddildi; uyku düzey dağılımlarının normal dağılıma uymadığı sonucuna varıldı.

H_0 : Populasyon dağılımı normal.

H_1 : Populasyon dağılımı normal değil.

Tablo 6.2. Uyku Düzeylerinin Kolmogorov-Simirnov Test Sonuçları

		Uyku Düze y i 1	Uyku Düze y i 2	Uyku Düze y i 3	Uyku Düze y i 4	Uyku Düze y i 5
N		833	1742	2694	737	726
Normal Parametreler ^{a,b}	Ortalama	0.31121	0.49281	0.69081	0.83307	0.88617
	Std. Sapma	0.055454	0.122704	0.128990	0.088532	0.085014
En fazla uç farklılıklar	Mutlak	0.066	0.096	0.072	0.142	0.241
	Positif	0.066	0.096	0.047	0.057	0.144
	Negatif	-0.059	-0.041	-0.072	-0.142	-0.241
Kolmogorov-Smirnov Z		1.901	4.001	3.737	3.864	6.504
D		0.001	0.000	0.000	0.000	0.000

a. Test dağılımı normal. b. Veriden hesaplandı

Teorem 2'ye göre parametrik üçgen bulanık sayı üretebilmek için, bu aşamada karar vericinin aşağıdaki maddelere karar verdiği varsayıldı.

- i) $L(0.5) = 15. \text{ yüzd$ elik , $R(0.5) = 85. \text{ yüzd$ elik .
- ii) Merkez değer uyku düzeyinin mod değeridir.
- iii) Uzun kuyruklardan etkilenmemek için verilerin uçlardaki %5'lik kısmını dikkate alma.

Buna göre uyku düzeylerinin parametrik üçgen bulanık sayı oluşturulmasında kullanılan kritik değerleri Tablo 6.3'deki gibi özetlendi.

Tablo 6.3: Teorem 2 için Beş Uyku Düzeyinin Üyelik Bilgileri

	Uyku Düze y i 1	Uyku Düze y i 2	Uyku Düze y i 3	Uyku Düze y i 4	Uyku Düze y i 5
X_{\min}	0.24100	0.32800	0.44600	0.65880	0.77175
$L(0.5)$	0.26010	0.37200	0.53700	0.74300	0.86200
$M : Mod$	0.262 ^a	0.409	0.818	0.871	0.911
$R(0.5)$	0.36700	0.63700	0.82375	0.91000	0.93800
X_{\max}	0.40100	0.71800	0.86300	0.95210	0.96265
Sol yayılma ($\sigma_A = \sigma_B$)	0.021	0.081	0.372	0.2122	0.13925
Sağ yayılma ($\beta_A = \beta_B$)	0.139	0.309	0.045	0.0811	0.05165
s_L	0.288491	0.884645	2.470768	1.371214	0.663648
s_R	2.470993	2.280122	0.336894	0.946765	1.068594

Bu deęerler kullanılarak ařaęıdaki üyelik fonksiyonlarına ulařıldı.

$$\mu_1 = \begin{cases} 1 - \left(\frac{0.262 - X}{0.021} \right)^{0.288491} & ; \quad 0.241 < X \leq 0.262 \\ 1 - \left(\frac{X - 0.262}{0.139} \right)^{2.470993} & ; \quad 0.262 < X \leq 0.401, \\ 0 & ; \quad \text{aksihalde} \end{cases}$$

$$\mu_2 = \begin{cases} 1 - \left(\frac{0.46717 - X}{0.13912} \right)^{0.957255} & ; \quad 0.32805 < X \leq 0.46717 \\ 1 - \left(\frac{X - 0.46717}{0.2503} \right)^{0.899564} & ; \quad 0.46717 < X \leq 0.71747, \\ 0 & ; \quad \text{aksihalde} \end{cases}$$

$$\mu_3 = \begin{cases} 1 - \left(\frac{0.818 - X}{0.372} \right)^{2.470768} & , \quad 0.446 < X \leq 0.818 \\ 1 - \left(\frac{X - 0.818}{0.045} \right)^{0.336894} & , \quad 0.818 < X \leq 0.863, \\ 0 & , \quad \text{aksihalde} \end{cases}$$

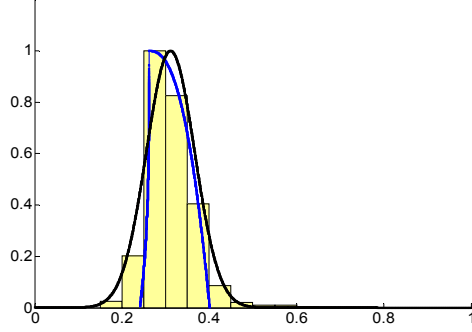
$$\mu_4 = \begin{cases} 1 - \left(\frac{0.871 - X}{0.2122} \right)^{1.371214} & ; \quad 0.6588 < X \leq 0.871 \\ 1 - \left(\frac{X - 0.871}{0.0811} \right)^{0.946765} & ; \quad 0.871 < X \leq 0.9521, \\ 0 & ; \quad \text{aksihalde} \end{cases}$$

$$\mu_5 = \begin{cases} 1 - \left(\frac{0.911 - X}{0.13925} \right)^{0.663648} & ; \quad 0.77175 < X \leq 0.911 \\ 1 - \left(\frac{X - 0.911}{0.05165} \right)^{1.068594} & ; \quad 0.911 < X \leq 0.96265. \\ 0 & ; \quad \text{aksihalde} \end{cases}$$

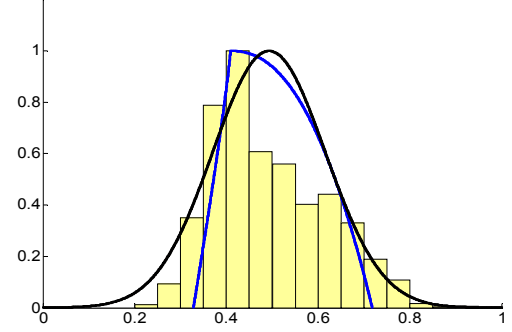
Teorem 2 yolu ile bulunan üyelik fonksiyonları ile, formül (6.1) yolu ile bulunan üyelik fonksiyonları arasındaki farkı görsel olarak görebilmek için Şekil 6.1 elde edildi. Burada histogramlar uyku düzeylerinin normalleştirilmiş yüzdeliklerine dayandırıldı.

Şekil incelendiğinde, Teorem 2 yolu ile bulunan üyelik fonksiyonunun, formül (6.1) yolu ile bulunan üyelik fonksiyonuna göre, veri dağılımına daha iyi uyum sağladığı görüldü.

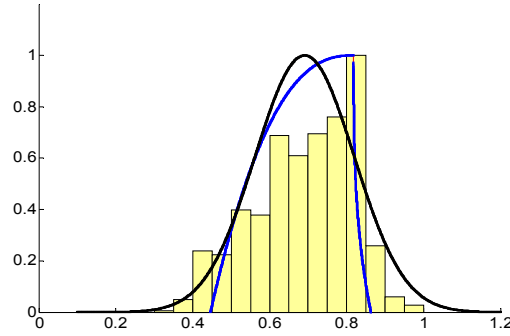
Şekil 6.1: Uyku Düzeyleri'nin Teorem 2 ve (6.1) Nolu Formül Kullanılarak Elde Edilen Üyelik Fonksiyonları



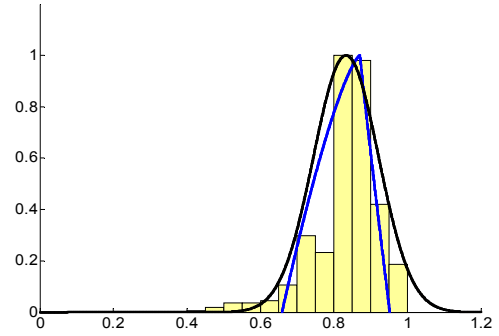
Uyku Düzeyi 1



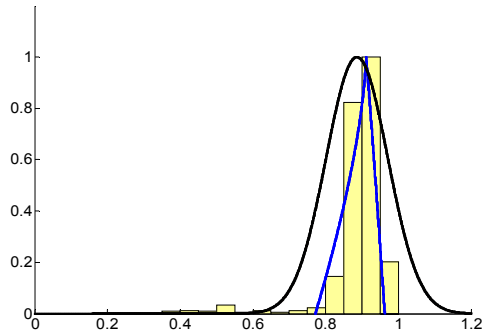
Uyku Düzeyi 2



Uyku Düzeyi 3



Uyku Düzeyi 4



Uyku Düzeyi 5

Şekillerde mavi eğri bu tezde ortaya atılan Teorem 2 yolu ile elde edilen üyelik fonksiyonunu, siyah eğri formül (6.1) yolu ile elde edilen üyelik fonksiyonunu göstermektedir.

Sonuç 4.1 kullanılarak üyelik fonksiyonu elde edebilmek için, beş uyku düzeyine ait frekans tabloları Tablo 6.4-Tablo 6.8'deki gibi oluşturuldu. Üyelik derecesinin elde edilmesinde kritik rol oynayan maksimum yüzdeliğe sahip sınıf aralığı, tablolarda sarı renkte gösterildi.

Tablo 6.4: Uyku Düzeyi 1'in Frekans Tablosu

Sınıf Aralığı	Orta Nokta	Frekans	Yüzelik	Birikimli Yüzelik
0.15-0.20	0.175	8	1	1
0.20-0.25	0.225	65	7.8	8.8
0.25-0.30	0.275	322	38.7	47.4
0.30-0.35	0.325	266	31.9	79.4
0.35-0.40	0.375	130	15.6	95
0.40-0.45	0.425	28	3.4	98.3
0.45-0.50	0.475	7	0.8	99.2
0.50-0.55	0.525	3	0.4	99.5
0.55-0.60	0.575	3	0.4	99.9
0.60-0.65	0.625	1	0.1	100
Toplam		833	100	

Tablo 6.5: Uyku Düzeyi 2'in Frekans Tablosu

Sınıf Aralığı	Orta Nokta	Frekans	Yüzelik	Birikimli Yüzelik
0.20-0.25	0.225	4	0.2	0.2
0.25-0.30	0.275	33	1.9	2.1
0.30-0.35	0.325	124	7.1	9.2
0.35-0.40	0.375	280	16.1	25.3
0.40-0.45	0.425	355	20.4	45.7
0.45-0.50	0.475	216	12.4	58.1
0.50-0.55	0.525	199	11.4	69.5
0.55-0.60	0.575	143	8.2	77.7
0.60-0.65	0.625	157	9	86.7
0.65-0.70	0.675	117	6.7	93.5
0.70-0.75	0.725	67	3.8	97.3
0.75-0.80	0.775	38	2.2	99.5
0.80-0.85	0.825	6	0.3	99.8
0.85-0.90	0.875	1	0.1	99.9
0.90-0.95	0.925	2	0.1	100
Toplam		1742	100	

Tablo 6.6: Uyku Düzeyi 3'ün Frekans Tablosu

Sınıf Aralığı	Orta Nokta	Frekans	Yüzdeler	Birikimli Yüzdeler
0.25-0.30	0.275	1	0	0
0.30-0.35	0.325	3	0.1	0.1
0.35-0.40	0.375	25	0.9	1.1
0.40-0.45	0.425	119	4.4	5.5
0.45-0.50	0.475	112	4.2	9.7
0.50-0.55	0.525	199	7.4	17
0.55-0.60	0.575	189	7	24.1
0.60-0.65	0.625	344	12.8	36.8
0.65-0.70	0.675	304	11.3	48.1
0.70-0.75	0.725	347	12.9	61
0.75-0.80	0.775	379	14.1	75.1
0.80-0.85	0.825	499	18.5	93.6
0.85-0.90	0.875	129	4.8	98.4
0.90-0.95	0.925	30	1.1	99.5
0.95-1	0.975	14	0.5	100
Toplam		2694	100	

Tablo 6.7: Uyku Düzeyi 4'ün Frekans Tablosu

Sınıf Aralığı	Orta Nokta	Frekans	Yüzdeler	Birikimli Yüzdeler
0.40-0.45	0.425	1	0.1	0.1
0.45-0.50	0.475	4	0.5	0.7
0.50-0.55	0.525	8	1.1	1.8
0.55-0.60	0.575	8	1.1	2.8
0.60-0.65	0.625	10	1.4	4.2
0.65-0.70	0.675	23	3.1	7.3
0.70-0.75	0.725	65	8.8	16.1
0.75-0.80	0.775	51	6.9	23.1
0.80-0.85	0.825	219	29.7	52.8
0.85-0.90	0.875	215	29.2	82
0.90-0.95	0.925	92	12.5	94.4
0.95-1	0.975	41	5.6	100
Toplam		737	100	

Tablo 6.8 Uyku Düzeyi 5'in Frekans Tablosu

Sınıf Aralığı	Orta Nokta	Frekans	Yüzdeler	Birikimli Yüzdeler
0.35-0.40	0.375	3	0.4	0.4
0.40-0.45	0.425	4	0.6	1
0.45-0.50	0.475	3	0.4	1.4
0.50-0.55	0.525	11	1.5	2.9
0.55-0.60	0.575	2	0.3	3.2
0.60-0.65	0.625	4	0.6	3.7
0.65-0.70	0.675	2	0.3	4
0.70-0.75	0.725	4	0.6	4.5
0.75-0.80	0.775	7	1	5.5
0.80-0.85	0.825	46	6.3	11.8
0.85-0.90	0.875	260	35.8	47.7
0.90-0.95	0.925	316	43.5	91.2
0.95-1	0.975	64	8.8	100
Toplam		726	100	

Sonuç 4.1 yolu ile uyku düzeylerinin üyelik fonksiyonları aşağıdaki gibi elde edildi.

$$\mu_1 = \begin{cases} 1 - \left(\frac{0.275 - X}{0.08656643} \right) & ; 0.18843357 < X \leq 0.275 \\ 1 - \left(\frac{X - 0.275}{0.25225700} \right) & ; 0.275 < X \leq 0.52725700, \\ 0 & ; \text{aksihalde} \end{cases}$$

$$\mu_2 = \begin{cases} 1 - \left(\frac{0.425 - X}{0.18034198} \right) & ; 0.24465802 < X \leq 0.425 \\ 1 - \left(\frac{X - 0.425}{0.39616265} \right) & ; 0.425 < X \leq 0.82116265, \\ 0 & ; \text{aksihalde} \end{cases}$$

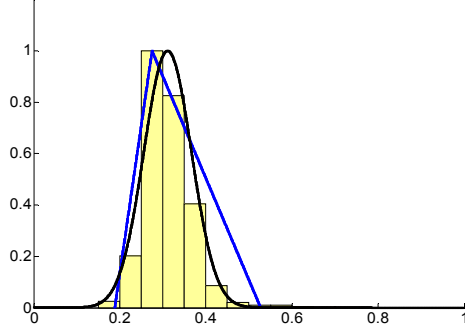
$$\mu_3 = \begin{cases} 1 - \left(\frac{0.825 - X}{0.47426030} \right) & ; \quad 0.35073970 < X \leq 0.825 \\ 1 - \left(\frac{X - 0.825}{0.11641292} \right) & ; \quad 0.825 < X \leq 0.94141292, \\ 0 & ; \quad \text{aksihalde} \end{cases}$$

$$\mu_4 = \begin{cases} 1 - \left(\frac{0.825 - X}{0.25813816} \right) & ; \quad 0.56686184 < X \leq 0.825 \\ 1 - \left(\frac{X - 0.825}{0.18153951} \right) & ; \quad 0.825 < X \leq 1.00653951, \\ 0 & ; \quad \text{aksihalde} \end{cases}$$

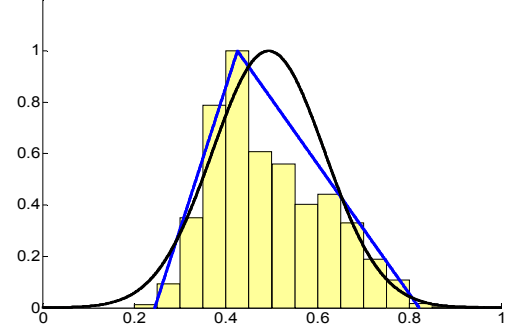
$$\mu_5 = \begin{cases} 1 - \left(\frac{0.925 - X}{0.33673093} \right) & ; \quad 0.58826907 < X \leq 0.925 \\ 1 - \left(\frac{X - 0.925}{0.06268012} \right) & ; \quad 0.925 < X \leq 0.98768012. \\ 0 & ; \quad \text{aksihalde} \end{cases}$$

Sonuç 4.1 ve formül (6.1) yolları ile elde edilen üyelik fonksiyonlarını görsel olarak karşılaştırabilmek amacıyla Şekil 6.2 oluşturuldu. Şekil incelendiğinde, Sonuç 4.1 yolu ile elde edilen üyelik fonksiyonunun özellikle uyku düzeyi 2 ve 3'te veri dağılımına daha iyi uyum sağladığı görüldü.

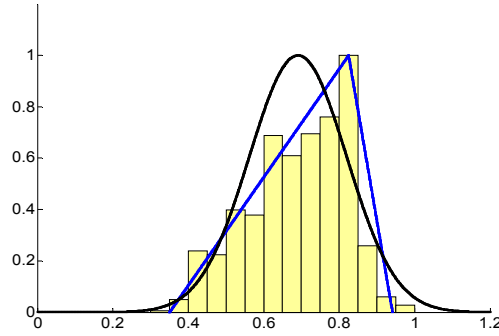
Şekil 6.2: Uyku Düzeyleri'nin Sonuç 4.1 ve (6.1) Nolu Formül Kullanılarak Elde Edilen Üyelik Fonksiyonları



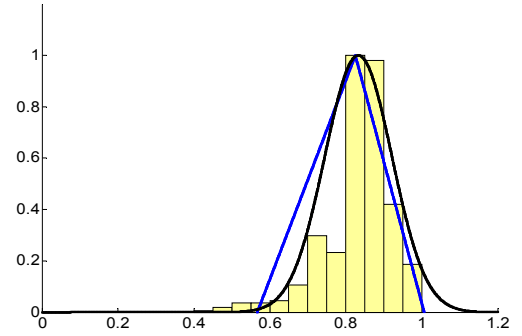
Uyku Düzeyi 1



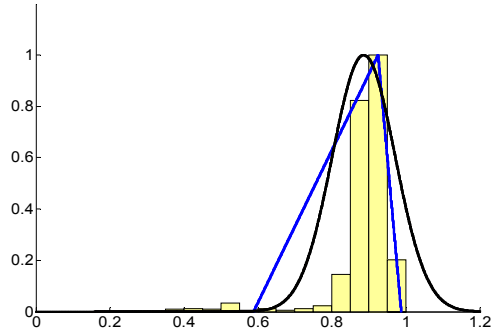
Uyku Düzeyi 2



Uyku Düzeyi 3



Uyku Düzeyi 4



Uyku Düzeyi 5

Şekillerde mavi eğri bu tezde ortaya atılan Sonuç 4.1 yolu ile elde edilen üyelik fonksiyonunu, siyah eğri formül (6.1) yolu ile elde edilen üyelik fonksiyonunu göstermektedir.

Veri dağılımda, uzun kuyruk etkisini görebilmek için, frekans tablosu %2.5 ve daha küçük yüzdeliğe sahip sınıftan itibaren kesildi. Sonuç 4.1, bu kesilmiş frekans tabloları kullanılarak tekrar uygulandı ve aşağıdaki üyelik fonksiyonları elde edildi.

$$\mu_1 = \begin{cases} 1 - \left(\frac{0.275 - X}{0.06262136} \right) & ; \quad 0.21237864 < X \leq 0.275 \\ 1 - \left(\frac{X - 0.275}{0.16839080} \right) & ; \quad 0.275 < X \leq 0.44339080, \\ 0 & ; \quad \text{aksihalde} \end{cases}$$

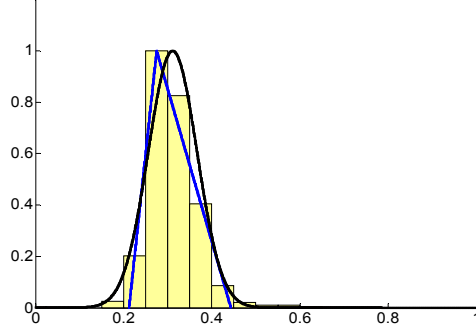
$$\mu_2 = \begin{cases} 1 - \left(\frac{0.425 - X}{0.161316409} \right) & ; \quad 0.263683591 < X \leq 0.425 \\ 1 - \left(\frac{X - 0.425}{0.317711516} \right) & ; \quad 0.425 < X \leq 0.742711516, \\ 0 & ; \quad \text{aksihalde} \end{cases}$$

$$\mu_3 = \begin{cases} 1 - \left(\frac{0.825 - X}{0.46261070} \right) & ; \quad 0.36238930 < X \leq 0.825 \\ 1 - \left(\frac{X - 0.825}{0.06751825} \right) & ; \quad 0.825 < X \leq 0.89251825, \\ 0 & ; \quad \text{aksihalde} \end{cases}$$

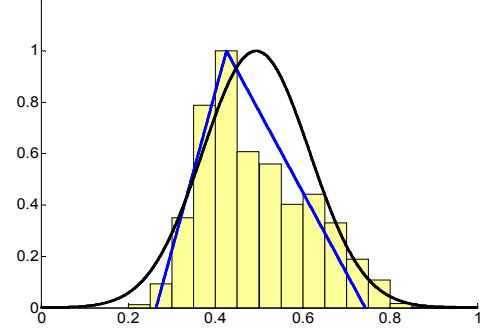
$$\mu_4 = \begin{cases} 1 - \left(\frac{0.825 - X}{0.12885033} \right) & ; \quad 0.69614967 < X \leq 0.825 \\ 1 - \left(\frac{X - 0.825}{0.18153951} \right) & ; \quad 0.825 < X \leq 1.00653951, \\ 0 & ; \quad \text{aksihalde} \end{cases}$$

$$\mu_5 = \begin{cases} 1 - \left(\frac{0.925 - X}{0.123736254} \right) & ; \quad 0.80126375 < X \leq 0.925 \\ 1 - \left(\frac{X - 0.925}{0.062680115} \right) & ; \quad 0.925 < X \leq 0.98768012. \\ 0 & ; \quad \text{aksihalde} \end{cases}$$

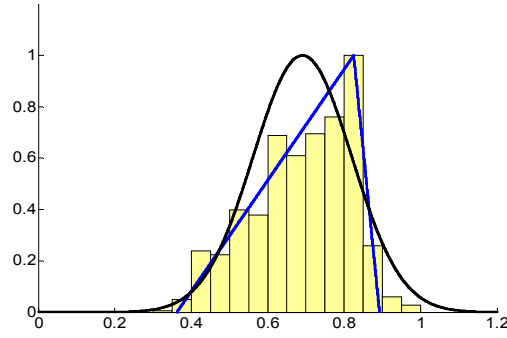
Şekil 6.3: Uzun Kuyruk Etkisi Kaldırıldığında Uyku Düzeyleri'nin Sonuç 4.1 ve (6.1) Nolu Formül Kullanılarak Elde Edilen Üyelik Fonksiyonları



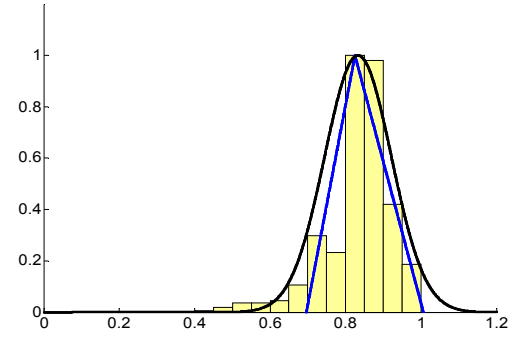
Uyku Düzeyi 1



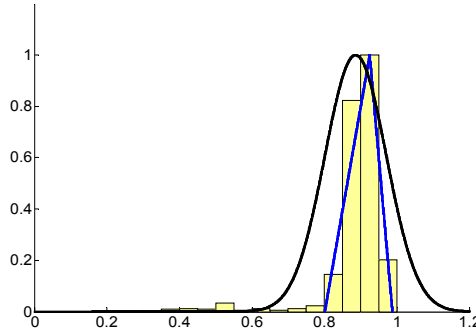
Uyku Düzeyi 2



Uyku Düzeyi 3



Uyku Düzeyi 4



Uyku Düzeyi 5

Şekillerde mavi eğri, uzun kuyruk etkisi ortadan kaldırıldığında Sonuç 4.1 yolu ile elde edilen üyelik fonksiyonunu, siyah eğri formül (6.1) yolu ile elde edilen üyelik fonksiyonunu göstermektedir.

Uzun kuyruk etkisinin yok edilmesinin, Sonuç 4.1 yolu ile elde edilen üyelik fonksiyonunda ne gibi bir değişmeye yol açtığı görsel olarak Şekil 6.3'te incelendi. İnceleme sonunda uzun kuyruk etkisinin yok edilmesinin bulanık sayıda daha küçük bir yayılmaya neden olduğu görüldü.

Üyelik fonksiyonları bulunduktan sonra, üyelik fonksiyonlarının sınıflamadaki etkinliğini görebilmek için veri setleri üzerinde sınıflama işlemi gerçekleştirildi. Atamalar maksimum üyelik seviyesine göre yapıldı ve etiketlere ait zaman serilerine ulaşıldı.

Tablo 6.9'da Şekil 6.1-Şekil 6.3'de yer alan üyelik fonksiyonları kullanılarak ulaşılan sınıflama kesinlikleri yer almaktadır. Burada sınıflama kesinlikleri (SK) veri setlerinde doğru tahmin edilen sınıf etiketlerinin oranını göstermektedir.

Tablo 6.9: Formül (6.1), Teorem 2 ve Sonuç 4.1 Üyelik Fonksiyon Yaklaşımlarının Sınıflama Kesinlikleri

Veri Seti	Formül (6.1)		Teorem 2		Sonuç 4.1		Sonuç 4.1 Uzun kuyruk kesilmiş	
	Doğru Tahmin Sayısı	SK	Doğru Tahmin Sayısı	SK	Doğru Tahmin Sayısı	SK	Doğru Tahmin Sayısı	SK
1	190	0.620915	215	0.702614	216	0.705882	217	0.70915
2	165	0.539216	176	0.575163	179	0.584967	192	0.627451
3	154	0.503268	186	0.607843	191	0.624183	190	0.620915
4	217	0.70915	220	0.718954	235	0.767974	234	0.764706
5	196	0.640523	218	0.712418	228	0.745098	221	0.722222
6	215	0.702614	210	0.686275	222	0.72549	238	0.777778
7	175	0.571895	211	0.689542	216	0.705882	225	0.735294
8	167	0.545752	162	0.529412	136	0.444444	143	0.46732
9	233	0.761438	238	0.777778	249	0.813725	243	0.794118
10	232	0.75817	246	0.803922	242	0.79085	238	0.777778
11	148	0.48366	152	0.496732	142	0.464052	134	0.437908
12	131	0.428105	142	0.464052	128	0.418301	139	0.454248
13	222	0.72549	205	0.669935	226	0.738562	217	0.70915
14	53	0.173203	61	0.199346	56	0.183007	65	0.212418
15	118	0.385621	103	0.336601	123	0.401961	111	0.362745
16	262	0.856209	239	0.781046	266	0.869281	251	0.820261
17	255	0.833333	262	0.856209	264	0.862745	275	0.898693
18	124	0.405229	188	0.614379	201	0.656863	196	0.640523
19	68	0.222222	142	0.464052	163	0.53268	153	0.5
20	131	0.428105	146	0.477124	113	0.369281	138	0.45098
21	192	0.627451	197	0.643791	194	0.633987	197	0.643791
Ort.	173.7143	0.567694	186.619	0.609866	190	0.620915	191.2857	0.625117
Ss.	56.77336	0.185534	49.70259	0.162427	56.12664	0.18342	53.79697	0.175807

Not: Teorem 2 kullanılarak yapılan sınıflamada, birinci uyku düzeyinin alt sınırından daha küçük olan değerlerin üyelik seviyeleri, ilk sınıfa 1 olarak; beşinci uyku

düzeyinin üst sınırından büyük olan değerlerin üyelik sevipleri, beşinci sınıfa 1 olarak atandı.

Bu tezde ortaya atılan Teorem 2, Sonuç 4.1 üyelik fonksiyonu yaklaşımlarının sınıflama kesinliğinde bir artışa yol açıp açmadığını görebilmek için aşağıdaki hipotezler oluşturuldu.

Hipotez 6.1

$$H_0 : \mu_{Teorem2} = \mu(6.1)$$

$$H_1 : \mu_{Teorem2} > \mu(6.1)$$

Hipotez 6.3

$$H_0 : \mu_{Sonuç4.1(kesilmiş)} = \mu(6.1)$$

$$H_1 : \mu_{Sonuç4.1(kesilmiş)} > \mu(6.1)$$

Hipotez 6.5

$$H_0 : \mu_{Teorem2} = \mu_{Sonuç4.1(kesilmiş)}$$

$$H_1 : \mu_{Teorem2} \neq \mu_{Sonuç4.1(kesilmiş)}$$

Hipotez 6.2

$$H_0 : \mu_{Sonuç4.1} = \mu(6.1)$$

$$H_1 : \mu_{Sonuç4.1} > \mu(6.1)$$

Hipotez 6.4

$$H_0 : \mu_{Teorem2} = \mu_{Sonuç4.1}$$

$$H_1 : \mu_{Teorem2} \neq \mu_{Sonuç4.1}$$

Hipotez 6.6

$$H_0 : \mu_{Sonuç4.1} = \mu_{Sonuç4.1(kesilmiş)}$$

$$H_1 : \mu_{Sonuç4.1} \neq \mu_{Sonuç4.1(kesilmiş)}$$

İkili-t testi için gerekli olan normallik varsayımının sağlanıp sağlanmadığını görebilmek için Kolmogorov-Smirnov testi doğru tahmin edilen etiket sayılarına uygulandı ve normallik koşulunun sağlandığı görüldü (Bakınız Tablo 6.10)

Tablo 6.10: Sınıflama Kesinliklerinin Normallik Testi Sonuçları

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	İstatistik	s.d	p	İstatistik	s.d	p
(6.1)	0.100	21	0.200*	0.967	21	0.665
Teorem 2	0.120	21	0.200*	0.950	21	0.340
Sonuç 4.1	0.155	21	0.200*	0.942	21	0.237
Sonuç 4.1 (Uzun kuyruklar kesilmiş)	0.160	21	0.171	0.942	21	0.236

a. Lilliefors anlamlılık düzeltmesi

*. Bu gerçek anlamlılığın alt sınırındır.

Daha sonra ikili-t testi SPSS paket programı kullanılarak yapıldı ve Tablo 6.11'deki sonuçlara ulaşıldı.

Tablo 6.11: Teorem 2 , Sonuç 4.1, Formül (6.1)'in İkili-t Testi Sonuçları

İkililer	Ort.	Standart Sapma	Ort.'un	95% Güv.	95% Güv.	t	s.d	p
			Standart Hatası	Ar. Alt Sınırı	Ar. Üst Sınırı			
Teorem 2 - (6.1)	0.04217	0.07819	0.01706	0.00658	0.07776	2.472	20	0.023
Sonuç 4.1- (6.1)	0.05322	0.09398	0.02051	0.01044	0.09600	2.595	20	0.017
Sonuç 4.1(kesilmiş) - (6.1)	0.05742	0.08771	0.01914	0.01750	0.09735	3.000	20	0.007
Teorem2 - Sonuç 4.1	-0.01105	0.05012	0.01094	-0.03386	0.01177	-1.010	20	0.324
Teorem2 - Sonuç 4.1 (kesilmiş)	-0.01525	0.03706	0.00809	-0.03212	0.00162	-1.886	20	0.074
Sonuç 4.1- Sonuç4.1(kesilmiş)	-0.00420	0.03443	0.00751	-0.01988	0.01147	-0.559	20	0.582

Bu tabloya göre, 6.1, 6.2 ve 6.3 nolu hipotezlerin sıfır hipotezleri rededilirken, 6.4, 6.5 ve 6.6 nolu hipotezlerin sıfır hipotezleri kabul edildi ve bispektral endeks veri setlerinin sınıflama işlemlerinde aşağıdaki sonuçlara ulaşıldı.

Sonuç 6.1: Teorem 2 kullanılarak elde edilen ortalama sınıflama kesinliği , formül (6.1) kullanılarak elde edilen ortalama sınıflama kesinliğinden fazladır.

Sonuç 6.2: Sonuç 4.1 kullanılarak elde edilen ortalama sınıflama kesinliği , formül (6.1) kullanılarak elde edilen ortalama sınıflama kesinliğinden fazladır.

Sonuç 6.3: Frekans tablosunda uzun kuyrukların dikkate alınmadığı durumlarda Sonuç 4.1 kullanılarak elde edilen ortalama sınıflama kesinliği, formül (6.1) kullanılarak elde edilen ortalama sınıflama kesinliğinden fazladır.

Sonuç 6.4: Teorem 2 ve Sonuç 4.1 kullanılarak elde edilen ortalama sınıflama kesinlikleri arasında fark yoktur.

Sonuç 6.5: Teorem 2 ve frekans tablosunda uzun kuyrukların dikkate alınmadığı durumlarda Sonuç 4.1 kullanılarak elde edilen ortalama sınıflama kesinlikleri arasında fark yoktur.

Sonuç 6.6: Frekans tablosunda uzun kuyrukların dikkate alındığı durumlarda Sonuç 4.1 kullanılarak elde edilen ortalama sınıflama kesinliği ile dikkate alınmadığı durumlarda Sonuç 4.1 kullanılarak elde edilen ortalama sınıflama kesinliği arasında bir fark yoktur.

Bu bölümde ikinci aşama olarak, bispektral endeks sınıflama işleminde, veri dağılımına yakınsayan üssel yaklaşım üyelik fonksiyon etkisi görülmek istendi. Teorem 5 birleştirilmiş veri setlerine uygulanarak aşağıdaki üyelik fonksiyonlarına ulaşıldı.

$$\mu_1 = \begin{cases} e^{-\left(\frac{0.275-X}{0.03366153}\right)^{1.19058453}} & ; X \leq 0.275 \\ e^{-\left(\frac{X-0.275}{0.10954908}\right)^{1.75612535}} & ; X > 0.275 \end{cases},$$

$$\mu_2 = \begin{cases} e^{-\left(\frac{0.425-X}{0.098304385}\right)^{2.12921635}} & ; X \leq 0.425 \\ e^{-\left(\frac{X-0.425}{0.140004284}\right)^{0.97277506}} & ; X > 0.425 \end{cases},$$

$$\mu_3 = \begin{cases} e^{-\left(\frac{0.825-X}{0.23903161}\right)^{1.16559615}} & ; X \leq 0.825 \\ e^{-\left(\frac{X-0.825}{0.03494202}\right)^{0.91429043}} & ; X > 0.825 \end{cases},$$

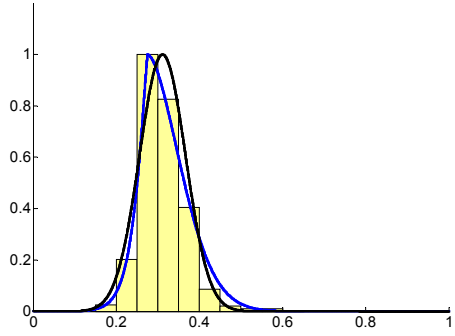
$$\mu_4 = \begin{cases} e^{-\left(\frac{0.825-X}{0.04244326}\right)^{0.67457267}} & ; X \leq 0.825 \\ e^{-\left(\frac{X-0.825}{0.12082079}\right)^{4.33694334}} & ; X > 0.825 \end{cases},$$

$$\mu_5 = \begin{cases} e^{-\left(\frac{0.925-X}{0.08407821}\right)^{1.03766617}} & ; X \leq 0.925 \\ 1 - \left(\frac{X - 0.925}{0.075}\right)^{0.55743699} & ; 0.925 < X \leq 1. \\ 0 & ; \text{aksihalde} \end{cases}$$

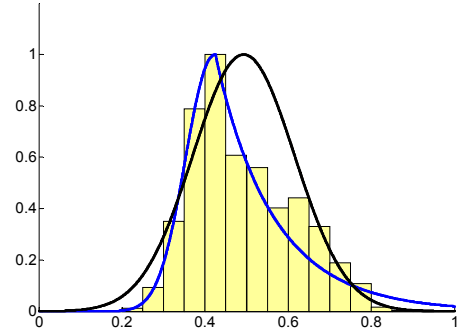
Burada, uyku düzeyi 5'e ait olan üssel üyelik fonksiyonun çıkartılması sırasında Teorem 5 için gerekli olan varsayımlardan birinin sağlanılmadığı görüldü. Varsayımın sağlanmadığı taraf için Teorem 2 uygulandı.

Teorem 5 ve (6.1) nolu formül kullanılarak elde edilen üyelik fonksiyonları Şekil 6.4'de gösterildi ve Teorem 5'in (6.1) nolu formüle göre veri dağılımına daha iyi uyum sağladığı görüldü.

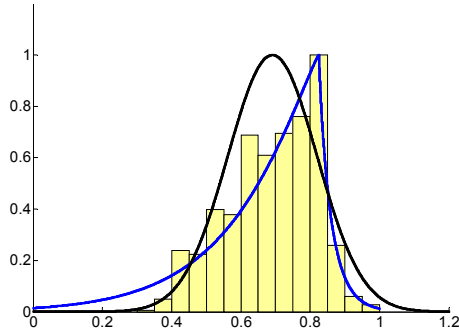
Şekil 6.4: Uyku Düzeyleri'nin Teorem 5 ve (6.1) Nolu Formül Kullanılarak Elde Edilen Üyelik Fonksiyonları



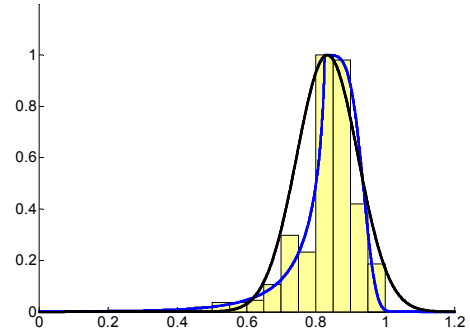
Uyku Düzeyi 1



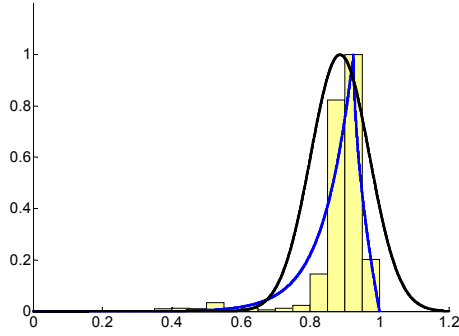
Uyku Düzeyi 2



Uyku Düzeyi 3



Uyku Düzeyi 4



Uyku Düzeyi 5

Şekillerde mavi eğri, Teorem 5 ile elde edilen üyelik fonksiyonunu gösterirken, siyah eğri formül (6.1) nolu formül ile elde edilen üyelik fonksiyonunu göstermektedir.

Veri dağılımındaki uzun kuyruk etkisini elimine etmek için frekans tablosu %2.5 ve daha az yüzdeliğe sahip sınıftan itibaren kesildi ve Teorem 5 yolu ile aşağıdaki üssel üyelik fonksiyonlarına ulaşıldı. Teorem 5'in varsayımlarının sağlanılmadığı durumlar için Teorem 2 uygulandı.

$$\mu_1 = \begin{cases} 1 - \left(\frac{0.275 - X}{0.075} \right)^{0.55512401} & ; X \leq 0.275 \\ e^{-\left(\frac{X - 0.275}{0.10279565} \right)^{2.29751301}} & ; X > 0.275 \\ 0 & ; \text{aksihalde} \end{cases}$$

$$\mu_2 = \begin{cases} e^{-\left(\frac{0.425 - X}{0.097529073} \right)^{2.15661786}} & ; X \leq 0.425 \\ e^{-\left(\frac{X - 0.425}{0.191654345} \right)^{0.61076913}} & , X > 0.475 \end{cases}$$

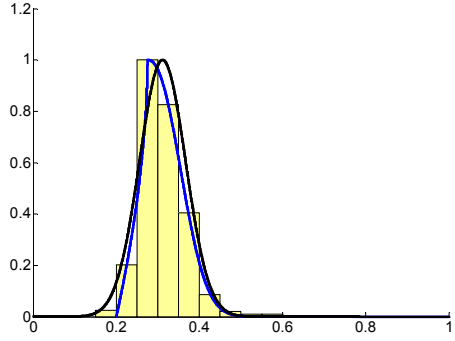
$$\mu_3 = \begin{cases} e^{-\left(\frac{0.825 - X}{0.30999222} \right)^{0.84923878}} & , X \leq 0.825 \\ 1 - \left(\frac{X - 0.825}{0.075} \right)^{0.74081565} & , X > 0.825 \\ 0 & , \text{aksi halde} \end{cases}$$

$$\mu_4 = \begin{cases} e^{-\left(\frac{0.825 - X}{0.02201361} \right)^{0.32668092}} & , X \leq 0.825 \\ e^{-\left(\frac{X - 0.825}{0.12082079} \right)^{4.33694334}} & , X > 0.825 \end{cases}$$

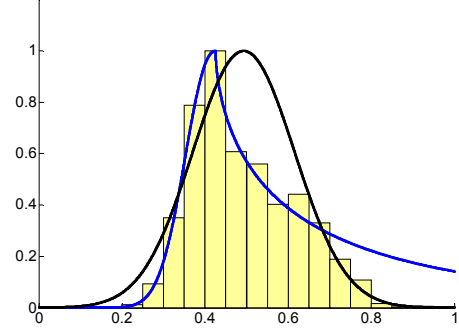
$$\mu_5 = \begin{cases} e^{-\left(\frac{0.925-X}{0.08195605}\right)^{3.31009069}} & , X \leq 0.925 \\ 1 - \left(\frac{X - 0.925}{0.075}\right)^{0.55743699} & , X > 0.925 \\ 0 & , \text{aksi halde} \end{cases}$$

Uzun kuyruk etkisinin veri dağılımında yarattığı etkiyi görsel olarak incelenmek için Şekil 6.5 oluşturuldu. Şekiller incelendiğinde üssel üyelik fonksiyon kollarının uç taraftalarda daha yüksek üyelik seviyelerine ulaştığı görüldü.

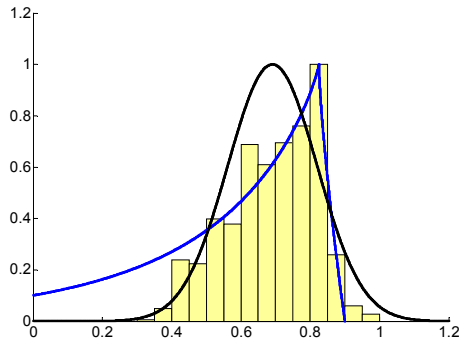
Şekil 6.5: Uzun Kuyruk Etkisi Kaldırıldığında Uyku Düzeyleri'nin Teorem 5 ve (6.1) Nolu Formül Kullanılarak Elde Edilen Üyelik Fonksiyonları



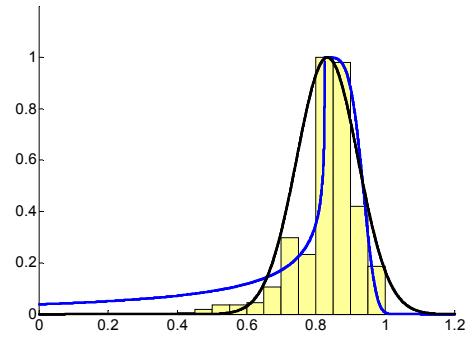
Uyku Düzeyi 1



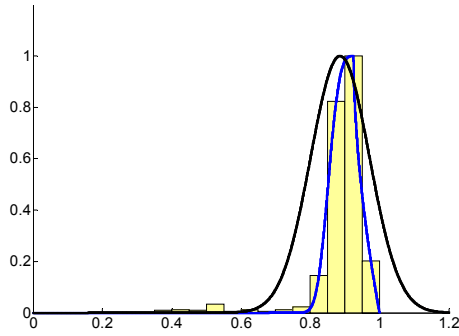
Uyku Düzeyi 2



Uyku Düzeyi 3



Uyku Düzeyi 4



Uyku Düzeyi 5

Şekil 6.5'de mavi eğri, uzun kuyruk etkisi ortadan kaldırıldığında Teorem 5 yolu ile elde edilen üyelik fonksiyonunu, siyah eğri formül (6.1) yolu ile elde edilen üyelik fonksiyonunu göstermektedir.

Teorem 5 yolu ile elde edilen üyelik fonksiyonları, bispektral endeks değerlerinin sınıflama işleminde kullanıldı. Atama işleminde maksimum üyelik seviyesi kriteri kullanıldı ve Tablo 6.12’de yer alan sonuçlara ulaşıldı.

Tablo 6.12: Formül (6.1) ve Teorem 5 Yolu İle Elde Edilen Üyelik Fonksiyon Yaklaşımlarının Sınıflama Kesinlikleri

Veri Seti	Formül (6.1)		Teorem 5		Teorem 5 Uzun kuyuk kesilmiş	
	Doğru Tahmin Sayısı	SK	Doğru Tahmin Sayısı	SK	Doğru Tahmin Sayısı	SK
1	190	0.620915	212	0.69281	211	0.689542
2	165	0.539216	199	0.650327	201	0.656863
3	154	0.503268	191	0.624183	182	0.594771
4	217	0.70915	233	0.761438	233	0.761438
5	196	0.640523	215	0.702614	214	0.699346
6	215	0.702614	237	0.77451	238	0.777778
7	175	0.571895	230	0.751634	228	0.745098
8	167	0.545752	145	0.473856	150	0.490196
9	233	0.761438	202	0.660131	237	0.77451
10	232	0.75817	233	0.761438	231	0.754902
11	148	0.48366	136	0.444444	134	0.437908
12	131	0.428105	146	0.477124	145	0.473856
13	222	0.72549	165	0.539216	211	0.689542
14	53	0.173203	68	0.222222	71	0.232026
15	118	0.385621	106	0.346405	121	0.395425
16	262	0.856209	216	0.705882	249	0.813725
17	255	0.833333	280	0.915033	279	0.911765
18	124	0.405229	191	0.624183	190	0.620915
19	68	0.222222	153	0.5	149	0.486928
20	131	0.428105	158	0.51634	144	0.470588
21	192	0.627451	200	0.653595	198	0.647059
Ort.	173.7143	0.567694	186.4762	0.609399	191.2381	0.624961
Ss.	56.77336	0.185534	49.64536	0.16224	51.01167	0.166705

Bu tezde ortaya atılan Teorem 5 üyelik fonksiyonu yaklaşımının, formül (6.1)’e nazaran sınıflama kesinliğinde bir artışa yol açıp açmadığını görebilmek için aşağıdaki hipotezler oluşturuldu. Hipotezlerde veri dağılımlarındaki uzun kuyruk etkileri de dikkate alındı.

Hipotez 6.7

$$H_0 : \mu_{Teorem5} = \mu_{(6.1)}$$

$$H_1 : \mu_{Teorem5} > \mu_{(6.1)}$$

Hipotez 6.8

$$H_0 : \mu_{Teorem5(kesilmiş)} = \mu_{(6.1)}$$

$$H_1 : \mu_{Teorem5(kesilmiş)} > \mu_{(6.1)}$$

Hipotez 6.9

$$H_0 : \mu_{Teorem5} = \mu_{Teorem5(kesilmiş)}$$

$$H_1 : \mu_{Teorem5} \neq \mu_{Teorem5(kesilmiş)}$$

SPSS paket programı kullanılarak sınıflama kesinlikleri işleme tabi tutuldu (Bakınız Tablo 6.13).

Tablo 6.13: Teorem 5, Formül (6.1)'in İkili-t Testi Sonuçları

İkililer	Ort.	Standart Sapma	Ort.'un	%95 Güv.	%95 Güv.	t	s.d	p
			Standart	Ar. Alt	Ar. Üst			
			Hatası	Sınırı	Sınırı			yanlı
Teorem5 – (6.1)	0.04171	0.11414	0.02491	-0.01025	0.09366	1.674	20	0.110
Teorem5kesilmiş – (6.1)	0.05727	0.08335	0.01819	0.01933	0.09521	3.149	20	0.005
Teorem5 – Teorem5kesilmiş	-0.01556	0.04919	0.01073	-0.03795	0.00683	-1.450	20	0.163

Buna göre Hipotez 6.8'in sıfır hipotezi red edilirken, Hipotez 6.7 ve Hipotez 6.9'un sıfır hipotezleri red edilmedi ve bispektral endeks veri setlerinin sınıflama işlemlerinde aşağıdaki sonuçlara ulaşıldı.

Sonuç 6.7: Teorem 5 ve formül (6.1) kullanılarak elde edilen ortalama sınıflama kesinlikleri birbirine eşittir.

Sonuç 6.8: Frekans tablosunda uzun kuyrukların dikkate alınmadığı durumlarda Teorem 5 kullanılarak elde edilen ortalama sınıflama kesinliği, formül (6.1) kullanılarak elde edilen ortalama sınıflama kesinliğinden fazladır.

Sonuç 6.9: Teorem 5 kullanılarak elde edilen ortalama sınıflama kesinliği ile frekans tablosunda uzun kuyrukların dikkate alınmadığı durumda Teorem 5 kullanılarak elde edilen ortalama sınıflama kesinliği arasında fark yoktur.

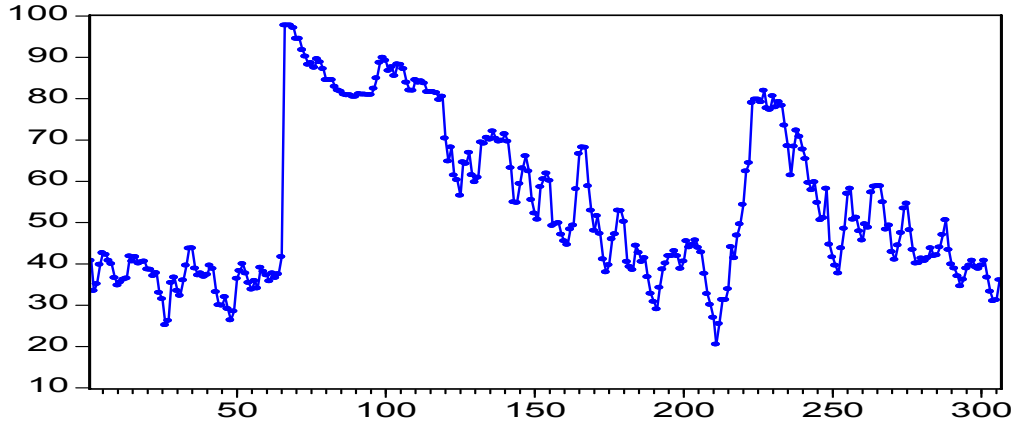
7 BİSPEKTRAL ENDEKS VERİ SETLERİNİN ZAMAN SERİLERİ

Wong ve diğerleri (2006) anestezi sırasında kaydedilmiş olan EEG verilerini GARCH modeli kullanarak modelledi. Bu çalışmaya istinaden, bu tezde 21 bireye ait bispektral endeks değerlerinin zaman serileri analizi, GARCH/ARCH modeli kullanılarak yapıldı.

Birinci Veri Setinin Zaman Serisi Analizi

Birinci bireye ait Şekil 7.1'deki bispektral değerler incelendiğinde durağan bir yapının bulunmadığı saptandı. Verileri durağan hale getirebilmek için Ln dönüşümü uygulandı.

Şekil 7.1: Birinci Bireyin Bispektral Endeks Zaman Serisi Grafiği



Ln dönüşümünden sonra verilerin durağan hale geldiği görüldü (Bakınız Şekil 7.2)

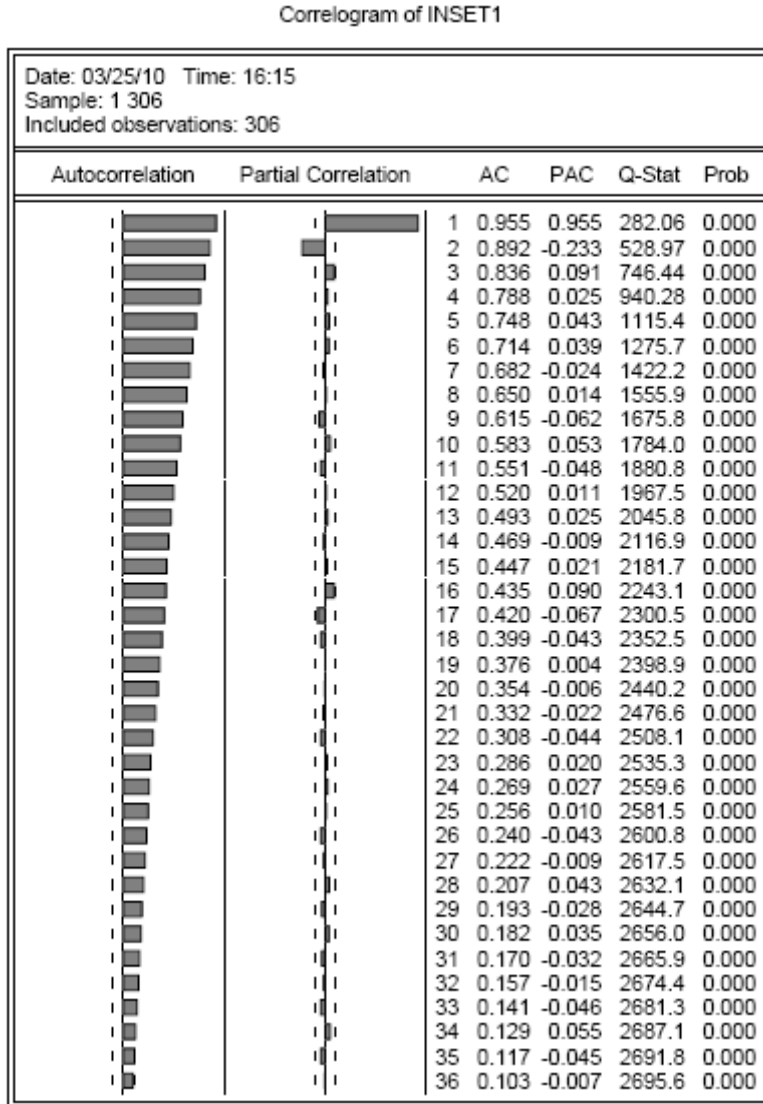
Şekil 7.2: Birinci Veri Setin Ln Değerlerinin Birim Kök Testi E-Views Çıktısı

Null Hypothesis: INSET1 has a unit root				
Exogenous: Constant				
Lag Length: 1 (Automatic based on SIC, MAXLAG=15)				
			t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic			-3.050064	0.0316
Test critical values:	1% level		-3.451703	
	5% level		-2.870836	
	10% level		-2.571794	
*MacKinnon (1996) one-sided p-values.				
Augmented Dickey-Fuller Test Equation				
Dependent Variable: D(INSET1)				
Method: Least Squares				
Date: 03/25/10 Time: 16:13				
Sample (adjusted): 3 306				
Included observations: 304 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
INSET1(-1)	-0.046734	0.015322	-3.050064	0.0025
D(INSET1(-1))	0.221126	0.056085	3.942727	0.0001
C	-0.030763	0.011479	-2.679885	0.0078
R-squared	0.067512	Mean dependent var		0.000246
Adjusted R-squared	0.061316	S.D. dependent var		0.093948
S.E. of regression	0.091023	Akaike info criterion		-1.945598
Sum squared resid	2.493821	Schwarz criterion		-1.908916
Log likelihood	298.7308	F-statistic		10.89616
Durbin-Watson stat	1.965641	Prob(F-statistic)		0.000027

Şekil 7.3’de yer alan otokorelasyon ve kısmi otokorelasyon fonksiyonları incelendiğinde aşağıdaki AR(2) modelinin kullanılabileceği düşünüldü.

$$z_t = c + \phi_1 z_{t-1} + \phi_2 z_{t-2} + u_t$$

Şekil 7.3: Birinci Veri Setinin Ln Değerlerinin Otokorelasyon ve Kısmi Otokorelasyon Fonksiyonları



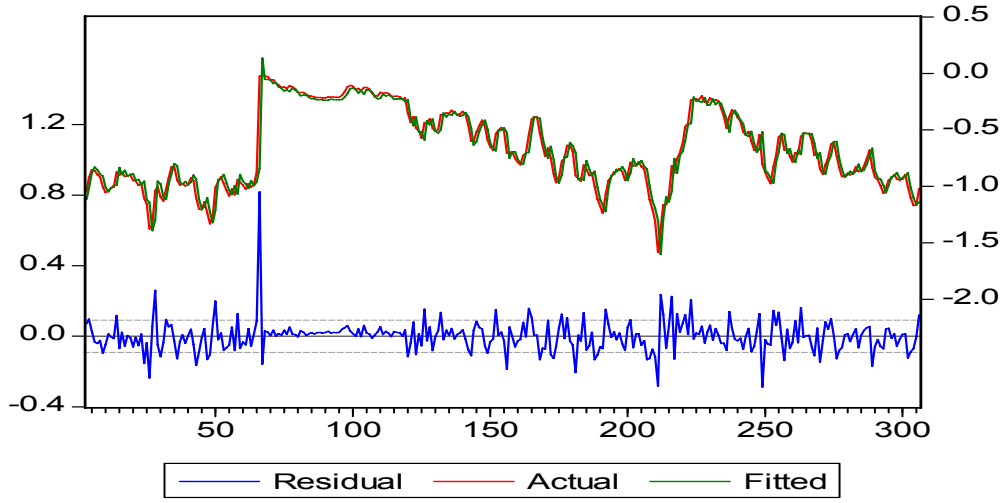
İlk etapta, serinin AR(2) modeli sonuçları e-views paket programı yardımıyla Şekil 7.4'deki gibi bulundu. Oluşturulan modelin hata terimleri arasında bir ilişki olmadığı Durbin-Watson test istatistiğine göre belirlendi. Modelin tüm katsayılarının anlamlı olduğu görüldü. Modelin %93.07'lik yüksek bir determinasyon katsayısına sahip olduğu bulundu.

Şekil 7.4: Birinci Veri Setin Ln Değerlerine Dayanan AR(2) Modeli

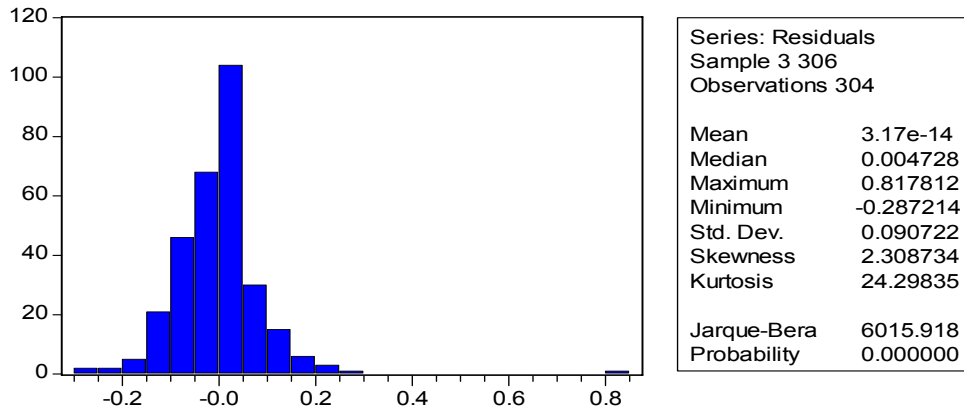
Dependent Variable: INSET1				
Method: Least Squares				
Date: 03/25/10 Time: 16:21				
Sample (adjusted): 3 306				
Included observations: 304 after adjustments				
Convergence achieved after 3 iterations				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.658261	0.111750	-5.890489	0.0000
AR(1)	1.174392	0.055933	20.99647	0.0000
AR(2)	-0.221126	0.056085	-3.942727	0.0001
R-squared	0.930738	Mean dependent var		-0.667452
Adjusted R-squared	0.930277	S.D. dependent var		0.344717
S.E. of regression	0.091023	Akaike info criterion		-1.945598
Sum squared resid	2.493821	Schwarz criterion		-1.908916
Log likelihood	298.7308	F-statistic		2022.397
Durbin-Watson stat	1.965641	Prob(F-statistic)		0.000000
Inverted AR Roots	.94	.24		

Bununla birlikte Şekil 7.5’de yer alan artık değerlerin grafiği incelendiğinde değişen varyansın olduğu tespit edildi. Şekil 7.6’da yer alan hata terimlerine ait histogram ve Jarque-Berra sonuçlarına bakıldığında ise hata terimlerinin normal dağılıma sahip olmadığı belirlendi.

Şekil 7.5: Birinci Veri Setin AR(2) Modelinin Tahmin ve Artık Grafiği



Şekil 7.6: Birinci Veri Setin AR(2) Modelinin Hata Terimlerinin Histogramı



Tüm bu kriterler dikkate alınarak ARCH modeli uygulandı ve Şekil 7.7'de yer alan sonuçlara ulaşıldı. ARCH modelinin oluşturulmasında normal dağılım varsayımı sağlanamadığı için genelleştirilmiş hata dağılımı (GED) seçeneği yardımıyla tahminler elde edildi.

Şekil 7.7: Birinci Veri Setin ARCH Modeli

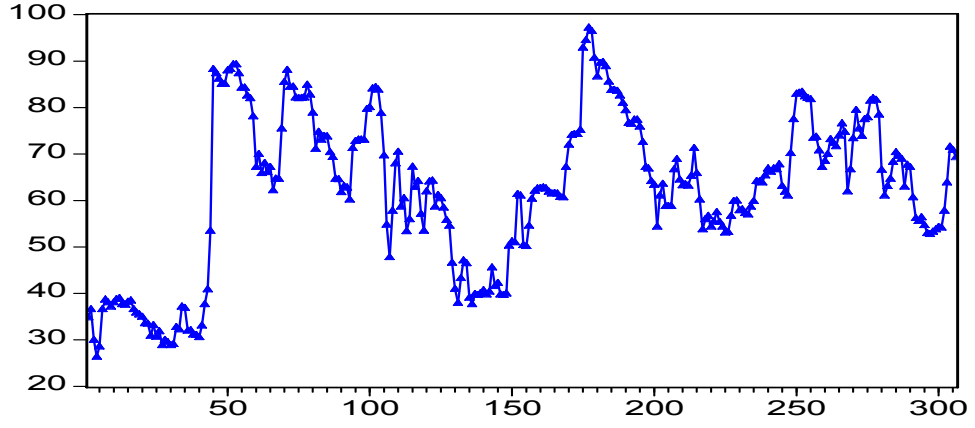
Dependent Variable: INSET1				
Method: ML - ARCH (Marquardt) - Generalized error distribution (GED)				
Date: 03/30/10 Time: 17:11				
Sample (adjusted): 3 306				
Included observations: 304 after adjustments				
Convergence achieved after 23 iterations				
Variance backcast: ON				
GARCH = C(4) + C(5)*RESID(-1)^2				
	Coefficient	Std. Error	z-Statistic	Prob.
C	-0.898272	0.328755	-2.732342	0.0063
AR(1)	1.230317	0.044309	27.76653	0.0000
AR(2)	-0.241408	0.044404	-5.436660	0.0000
Variance Equation				
C	0.005363	0.001097	4.890330	0.0000
RESID(-1)^2	0.381120	0.194212	1.962391	0.0497
GED PARAMETER	0.876305	0.075518	11.60398	0.0000
R-squared	0.929300	Mean dependent var		-0.667452
Adjusted R-squared	0.928113	S.D. dependent var		0.344717
S.E. of regression	0.092424	Akaike info criterion		-2.290378
Sum squared resid	2.545597	Schwarz criterion		-2.217015
Log likelihood	354.1374	F-statistic		783.3943
Durbin-Watson stat	2.035756	Prob(F-statistic)		0.000000
Inverted AR Roots	.99	.25		

ARCH modeli bulunurken oluşturulan AR modelinin katsayılarına bakıldığında durağanlık koşulunun sağlandığı görüldü ($\phi_1 + \phi_2 < 1$ ve $\phi_2 - \phi_1 < 1$).

İkinci Veri Setinin Zaman Serisi Analizi

İkinci bireyin bisapektral endeks değerlerinin grafiği Şekil 7.8'deki gibi oluşturuldu.

Şekil 7.8: İkinci Bireyin Bispektral Endeks Zaman Serisi Grafiği



Grafikte serinin durağan olup olmadığı tam olarak belirlenemedi ve Şekil 7.9'da yer alan birim kök testi sonuçları incelendi ve birim köke sahip olmadığı sonucuna varıldı.

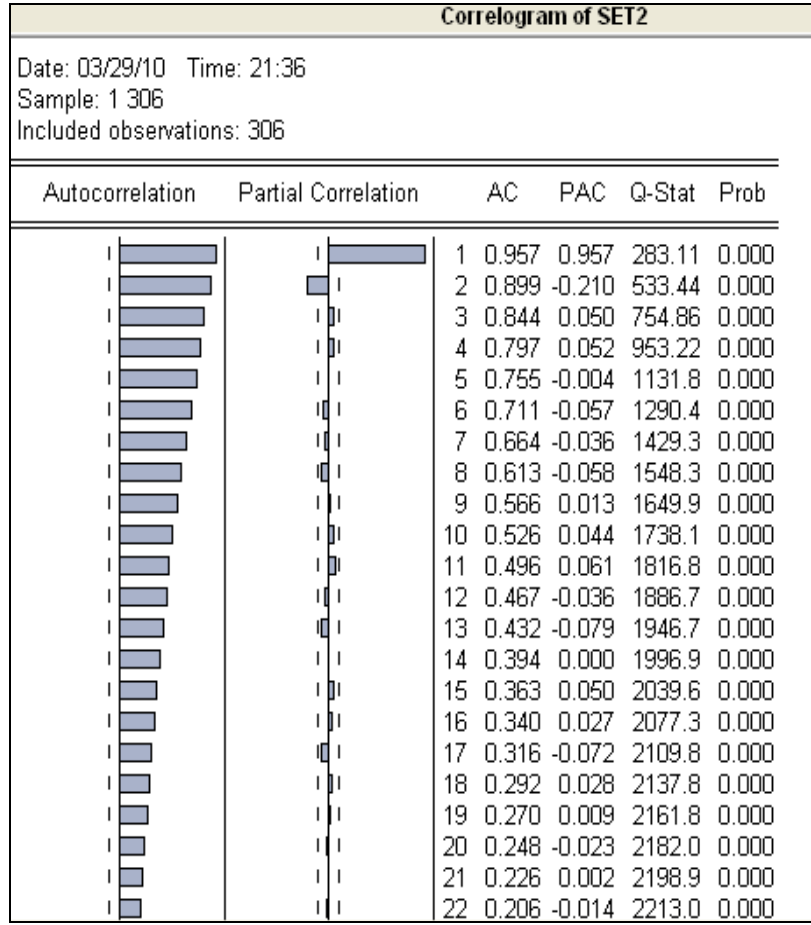
Şekil 7.9: İkinci Veri Setin Birim Kök Testi

Null Hypothesis: SET2 has a unit root				
Exogenous: Constant				
Lag Length: 1 (Automatic based on SIC, MAXLAG=15)				
			t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic			-3.250009	0.0182
Test critical values:	1% level		-3.451703	
	5% level		-2.870836	
	10% level		-2.571794	
*MacKinnon (1996) one-sided p-values.				
Augmented Dickey-Fuller Test Equation				
Dependent Variable: D(SET2)				
Method: Least Squares				
Date: 03/29/10 Time: 21:25				
Sample (adjusted): 3 306				
Included observations: 304 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
SET2(-1)	-0.049655	0.015278	-3.250009	0.0013
D(SET2(-1))	0.244901	0.055650	4.400707	0.0000
C	3.179375	0.985909	3.224817	0.0014
R-squared	0.081841	Mean dependent var		0.107566
Adjusted R-squared	0.075740	S.D. dependent var		4.559946
S.E. of regression	4.383861	Akaike info criterion		5.803556
Sum squared resid	5784.691	Schwarz criterion		5.840238
Log likelihood	-879.1406	F-statistic		13.41491
Durbin-Watson stat	1.949760	Prob(F-statistic)		0.000003

İkinci veri setin Şekil 7.10’da yer alan otokorelasyon ve kısmi otokorelasyon fonksiyonları incelendiğinde modelin, aşağıdaki AR(2) modeline uyabileceği düşünüldü.

$$z_t = c + \phi_1 z_{t-1} + \phi_2 z_{t-2} + u_t$$

Şekil 7.10: İkinci Veri Setin Otokorelasyon ve Kısmi Otokorelasyon Fonksiyonları



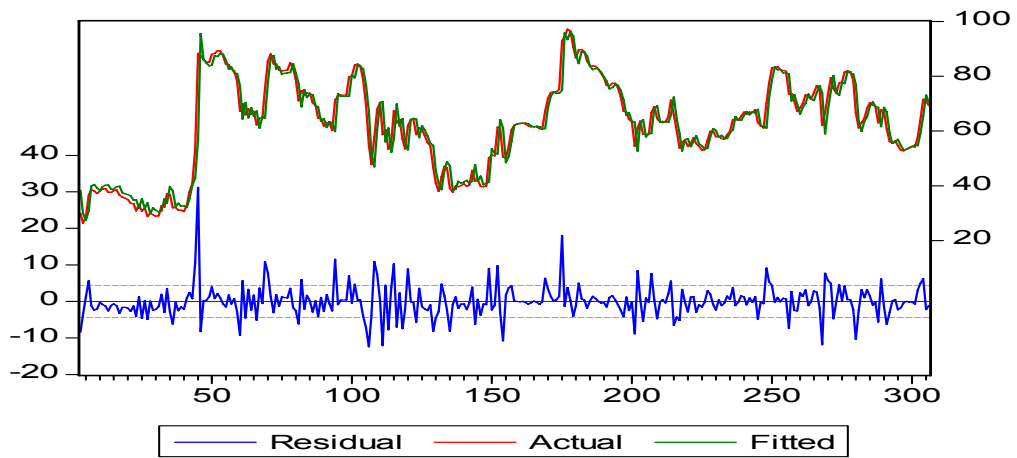
Oluşturulan AR(2) modelinin Durbin Watson test istatistiği incelendiğinde modelin hata terimlerinin birbirinden bağımsız olduğu ve yüksek bir determinasyon katsayısına ulaşıldığı gözlemlendi (Bakınız Şekil 7.11).

Şekil 7.11 İkinci Veri Setin AR(2) Modeli

Dependent Variable: SET2				
Method: Least Squares				
Date: 03/29/10 Time: 21:53				
Sample (adjusted): 3 306				
Included observations: 304 after adjustments				
Convergence achieved after 3 iterations				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	64.02924	5.090196	12.57893	0.0000
AR(1)	1.195245	0.055877	21.39057	0.0000
AR(2)	-0.244901	0.055650	-4.400707	0.0000
R-squared	0.930269	Mean dependent var		62.55461
Adjusted R-squared	0.929805	S.D. dependent var		16.54645
S.E. of regression	4.383861	Akaike info criterion		5.803556
Sum squared resid	5784.691	Schwarz criterion		5.840238
Log likelihood	-879.1406	F-statistic		2007.785
Durbin-Watson stat	1.949760	Prob(F-statistic)		0.000000
Inverted AR Roots	.93	.26		

Bununla birlikte Şekil 7.12’de yer alan tahmin, gözlem ve artık değerlerinin grafiği incelendiğinde hata terimlerinde eşit varyans durumu sağlanamadığı belirlendi.

Şekil 7.12: İkinci Veri Setin AR(2) Modelinin Tahmin ve Artık Grafiği



Artıklarda sabit varyans durumu olmadığı için ARCH modeli uygulandı. Model oluşturulurken artıkların normal dağılmadığı göz önünde bulundurularak e-views programındaki GED seçeneği ile Şekil 7.13’de yer alan sonuçlara ulaşıldı.

Şekil 7.13: İkinci Veri Setin ARCH Modeli

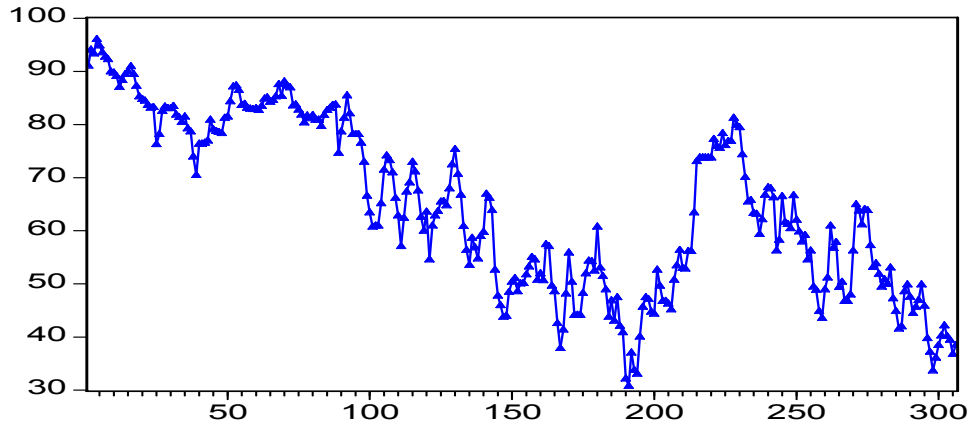
Dependent Variable: SET2				
Method: ML - ARCH (Marquardt) - Generalized error distribution (GED)				
Date: 03/29/10 Time: 22:12				
Sample (adjusted): 3 306				
Included observations: 304 after adjustments				
Convergence achieved after 41 iterations				
Variance backcast: ON				
GED parameter fixed at 1.5				
GARCH = C(4) + C(5)*RESID(-1)^2				
	Coefficient	Std. Error	z-Statistic	Prob.
C	58.01098	9.418065	6.159543	0.0000
AR(1)	1.144006	0.060953	18.76852	0.0000
AR(2)	-0.168336	0.060460	-2.784267	0.0054
Variance Equation				
C	11.70479	0.968236	12.08878	0.0000
RESID(-1)^2	0.284840	0.064153	4.439989	0.0000
R-squared	0.929168	Mean dependent var	62.55461	
Adjusted R-squared	0.928221	S.D. dependent var	16.54645	
S.E. of regression	4.433063	Akaike info criterion	5.559607	
Sum squared resid	5875.963	Schwarz criterion	5.620742	
Log likelihood	-840.0602	F-statistic	980.5709	
Durbin-Watson stat	1.837512	Prob(F-statistic)	0.000000	
Inverted AR Roots	.97	.17		

Modelin katsayıları bakıldığında $\phi_1 + \phi_2 < 1$ ve $\phi_2 - \phi_1 < 1$ durağanlık koşullarının sağlandığı görüldü. Modelin açıklayıcılık düzeyinin yüksek olduğu ($R^2 = \%92.9$) ve hata terimlerinin birbirinden bağımsız olduğu sonucuna varıldı.

Üçüncü veri setin zaman serisi analizi

Üçüncü bireyin bispektral endeks değerlerinin grafiği Şekil 7.14'deki gibi oluşturuldu. Serinin durağan bir yapıya sahip olmadığı belirlenerek ilk etapta Ln dönüşümü uygulandı.

Şekil 7.14: Üçüncü Bireyin Bispektral Endeks Zaman Grafiği



Şekil 7.15'de yer alan birim kök testi sonucuna göre, Ln değerlerinde birim kök olmadığı tespit edildi.

Şekil 7.15: Üçüncü Veri Setinin Ln Değerlerinin Birim Kök Testi

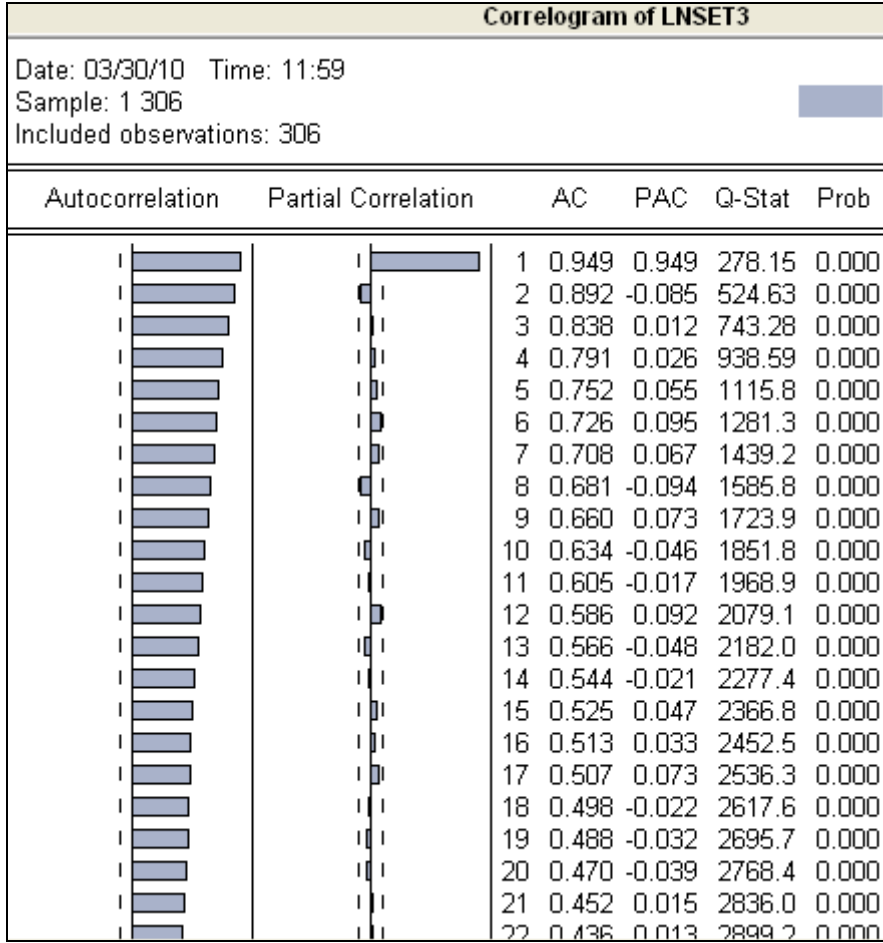
Null Hypothesis: LNSET3 has a unit root				
Exogenous: Constant, Linear Trend				
Lag Length: 1 (Automatic based on SIC, MAXLAG=15)				
			t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic			-3.563708	0.0347
Test critical values:	1% level		-3.988534	
	5% level		-3.424676	
	10% level		-3.135407	
*MacKinnon (1996) one-sided p-values.				
Augmented Dickey-Fuller Test Equation				
Dependent Variable: D(LNSET3)				
Method: Least Squares				
Date: 03/30/10 Time: 11:53				
Sample (adjusted): 3 306				
Included observations: 304 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
LNSET3(-1)	-0.072646	0.020385	-3.563708	0.0004
D(LNSET3(-1))	0.151841	0.057185	2.655266	0.0083
C	0.322365	0.091319	3.530108	0.0005
@TREND(1)	-0.000159	5.99E-05	-2.652832	0.0084
R-squared	0.053185	Mean dependent var		-0.002940
Adjusted R-squared	0.043717	S.D. dependent var		0.064278
S.E. of regression	0.062857	Akaike info criterion		-2.682834
Sum squared resid	1.185306	Schwarz criterion		-2.633926
Log likelihood	411.7908	F-statistic		5.617261
Durbin-Watson stat	2.003284	Prob(F-statistic)		0.000929

Ln değerlerinin Şekil 7.16'de yer alan otokorelasyon ve kısmi otokorelasyon fonksiyonları ve zaman serisindeki trent faktörü gözönüne alınarak,

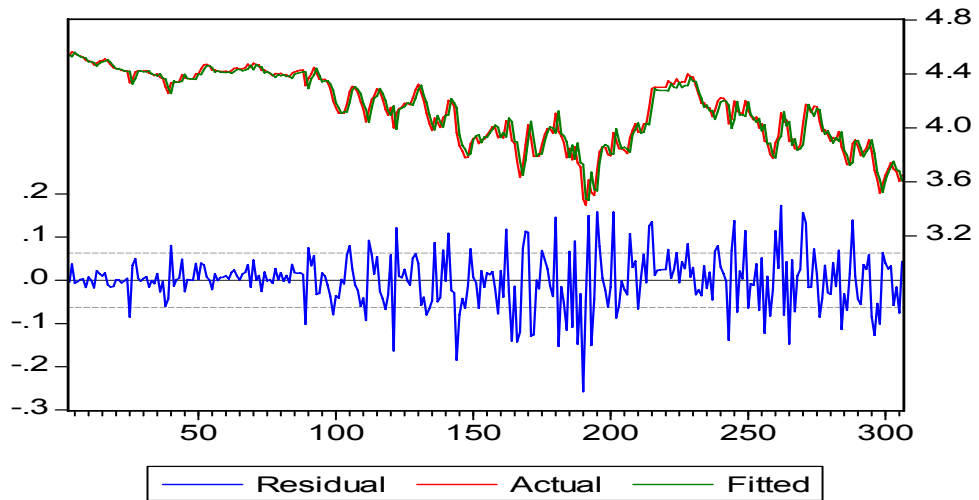
$$y_t = \delta + \beta t + \phi_1 y_{t-1} + \phi_2 y_{t-2} + u_t$$

modeline dayanan ARCH modelinin uygulanabileceği düşünüldü. İlgili modelin oluşturulmasında, hata terimlerinin normal dağılıma uymadığı gözönüne alınarak genelleştirilmiş hata dağılımları seçeneği kullanıldı (Bakınız Şekil 7.17, Şekil 7.18).

Şekil 7.16: Üçüncü Veri Setinin Ln Değerlerinin Otokorelasyon ve Kısmi Otokorelasyon Fonksiyonları



Şekil 7.17: Üçüncü Veri Setine Ait Modelin Tahmin ve Artık Grafiği



En son oluşturulan ARCH modelin hata terimlerinin birbirinden bağımsız olduğu ve modelin açıklayıcılık düzeyinin yüksek olduğu sonuçlarına varıldı.

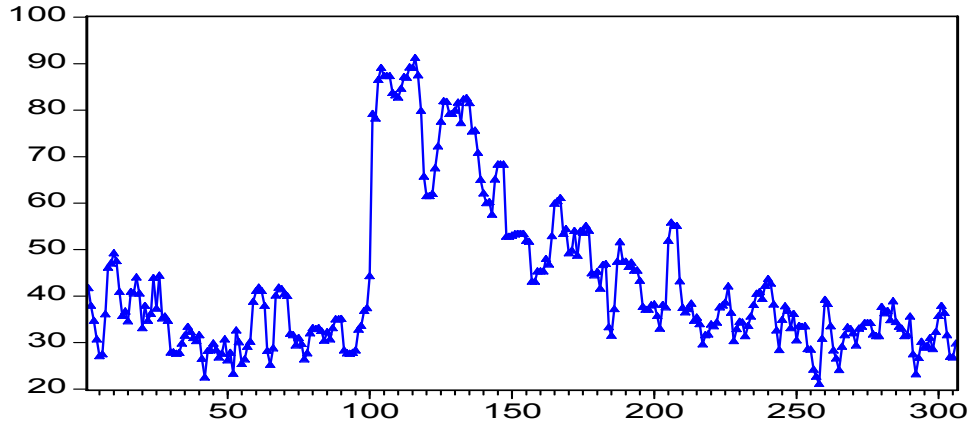
Şekil 7.18: Üçüncü Veri Setin ARCH Modeli

Dependent Variable: LNSET3				
Method: ML - ARCH (Marquardt) - Generalized error distribution (GED)				
Date: 03/30/10 Time: 17:40				
Sample (adjusted): 3 306				
Included observations: 304 after adjustments				
Convergence achieved after 38 iterations				
Variance backcast: ON				
GED parameter fixed at 1.5				
GARCH = C(5) + C(6)*RESID(-1)^2				
	Coefficient	Std. Error	z-Statistic	Prob.
C	4.491394	0.112610	39.88439	0.0000
@TREND	-0.002151	0.000544	-3.957803	0.0001
AR(1)	1.093044	0.060888	17.95182	0.0000
AR(2)	-0.170085	0.061479	-2.766565	0.0057
Variance Equation				
C	0.002757	0.000298	9.255358	0.0000
RESID(-1)^2	0.284793	0.140311	2.029729	0.0424
R-squared	0.942428	Mean dependent var	4.133251	
Adjusted R-squared	0.941462	S.D. dependent var	0.260976	
S.E. of regression	0.063142	Akaike info criterion	-2.778126	
Sum squared resid	1.188101	Schwarz criterion	-2.704764	
Log likelihood	428.2752	F-statistic	975.6325	
Durbin-Watson stat	2.027172	Prob(F-statistic)	0.000000	
Inverted AR Roots	.91	.19		

Dördüncü veri setinin zaman serisi analizi

Dördüncü bireyin bispektral endeks değerleri Şekil 7.19'daki gibi gözlemlendi ve serinin durağan bir yapıya sahip olmadığına karar verildi. Seride durağanlık yakalayabilmek için Ln dönüşümü uygulandı.

Şekil 7.19: Dördüncü Bireye Ait Zaman Serisi



ADF testine göre Ln değerleri alınmış değerlerin birim köke sahip olduğu bulundu. Ancak ADF testinin bazı durumlarda farklı sonuçlar verebileceği dikkate alınarak KPSS testi uygulandı ve bu teste göre birim kök durumu gözlemlenmedi.

Şekil 7.20: Dördüncü Veri Setinin Birim Kök Testi E-Views Çıktısı

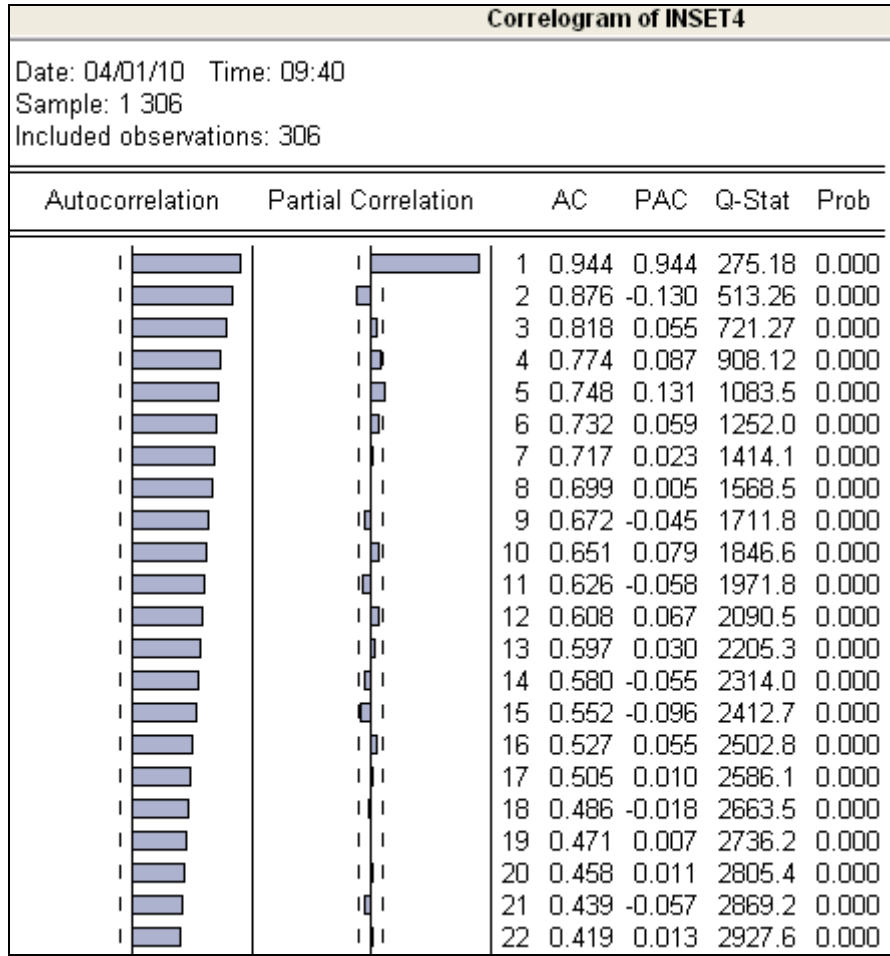
Null Hypothesis: INSET4 is stationary Exogenous: Constant Bandwidth: 14 (Newey-West using Bartlett kernel)				
				LM-Stat.
Kwiatkowski-Phillips-Schmidt-Shin test statistic				0.367321
Asymptotic critical values*:				
1% level				0.739000
5% level				0.463000
10% level				0.347000
*Kwiatkowski-Phillips-Schmidt-Shin (1992, Table 1)				
Residual variance (no correction)				0.115789
HAC corrected variance (Bartlett kernel)				1.383391
KPSS Test Equation Dependent Variable: INSET4 Method: Least Squares Date: 03/31/10 Time: 17:43 Sample: 1 306 Included observations: 306				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3.697157	0.019484	189.7511	0.0000
R-squared	0.000000	Mean dependent var		3.697157
Adjusted R-squared	0.000000	S.D. dependent var		0.340835
S.E. of regression	0.340835	Akaike info criterion		0.688426
Sum squared resid	35.43140	Schwarz criterion		0.700595
Log likelihood	-104.3292	Durbin-Watson stat		0.103049

Dördüncü veri setinin Ln değerlerinin otokorelasyon ve kısmi otokorelasyon fonksiyonları incelendiğinde

$$y_t = \delta + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \phi_5 y_{t-5} + u_t$$

modelinin kullanılabileceği düşünüldü.

Şekil 7.21: Dördüncü Veri Setin Ln Değerlerinin Otokorelasyon ve Kısmi Otokorelasyon Fonksiyonları

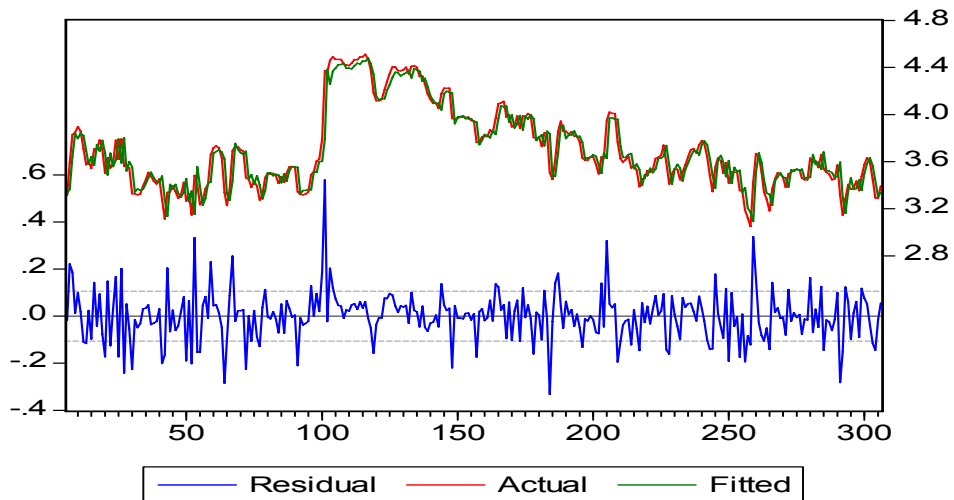


Model uygulandığında model parametrelerinin anlamlı olduğu, modelin %90.47'lik bir R^2 sahip olduğu görüldü. Durban-Watson istatistiğine bakıldığında hata terimleri arasında bir ilişki olmadığı kanısına varıldı. Bununla birlikte modelin tahmin ve artık grafiği incelendiğinde, değişen varyans durumu gözlemlendi ve ARCH yöntemine başvuruldu.

Şekil 7.22: Dördüncü Veri Setin Zaman Serisi Modeli

Dependent Variable: INSET4				
Method: Least Squares				
Date: 04/20/10 Time: 16:23				
Sample (adjusted): 6 306				
Included observations: 301 after adjustments				
Convergence achieved after 3 iterations				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3.704049	0.145277	25.49647	0.0000
AR(1)	1.069203	0.056614	18.88581	0.0000
AR(2)	-0.206553	0.062554	-3.301976	0.0011
AR(5)	0.095203	0.032014	2.973797	0.0032
R-squared	0.904721	Mean dependent var	3.699981	
Adjusted R-squared	0.903759	S.D. dependent var	0.342388	
S.E. of regression	0.106218	Akaike info criterion	-1.633440	
Sum squared resid	3.350852	Schwarz criterion	-1.584177	
Log likelihood	249.8328	F-statistic	940.0530	
Durbin-Watson stat	2.047723	Prob(F-statistic)	0.000000	
Inverted AR Roots	.96	.40+.51i	.40-.51i	-.35-.34i
	-.35+.34i			

Şekil 7.23: Dördüncü Veri Setine Ait Modelinin Tahmin ve Artık Grafiği



Dördüncü veri setin ARCH modeli oluşturulurken hata terimlerinin normal dağılmadığı göz önünde bulundurularak genelleştirilmiş hata dağılımı kullanıldı (Bakınız Şekil 7.24).

Şekil 7.24: Dördüncü Veri Setin ARCH Modeli

Dependent Variable: INSET4				
Method: ML - ARCH (Marquardt) - Generalized error distribution (GED)				
Date: 04/20/10 Time: 16:22				
Sample (adjusted): 6 306				
Included observations: 301 after adjustments				
Convergence achieved after 58 iterations				
Variance backcast: ON				
GED parameter fixed at 1.5				
GARCH = C(5) + C(6)*RESID(-1)^2				
	Coefficient	Std. Error	z-Statistic	Prob.
C	3.743079	0.209782	17.84274	0.0000
AR(1)	1.091826	0.065385	16.69853	0.0000
AR(2)	-0.189347	0.070715	-2.677622	0.0074
AR(5)	0.070593	0.029543	2.389521	0.0169
Variance Equation				
C	0.008044	0.000884	9.095120	0.0000
RESID(-1)^2	0.254789	0.081044	3.143830	0.0017
R-squared	0.904135	Mean dependent var	3.699981	
Adjusted R-squared	0.902510	S.D. dependent var	0.342388	
S.E. of regression	0.106905	Akaike info criterion	-1.737346	
Sum squared resid	3.371475	Schwarz criterion	-1.663450	
Log likelihood	267.4706	F-statistic	556.4458	
Durbin-Watson stat	2.078907	Prob(F-statistic)	0.000000	
Inverted AR Roots	.98	.38-.46i	.38+.46i	-.32+.32i
	-.32-.32i			

Kalan 17 veri seti benzer şekilde analiz edilerek ARCH modelleri Tablo 7.1'deki gibi bulundu.

Tablo 7.1: 5-21 arası veri setlerin ARCH/GARCH modelleri

Veri Seti	Veri Tipi	Birim Kök Testi	ARCH/GARCH Modeli	R^2	Durbin Watson
5	Ln	KPSS 0.196764	$y_t = 4.300953 + 0.971310y_{t-1} + \mathbf{u}_t$ $\hat{\sigma}_t^2 = 0.005914 + 0.549238\mathbf{u}_{t-1}^2$	0.940751	1.801116
6	Or.	ADF -3.319499	$y_t = 70.75216 - 0.244393t$ $+ 0.000416t^2 + 0.921307y_{t-1}$ $- 0.137699y_{t-3} + \mathbf{u}_t$ $\hat{\sigma}_t^2 = 9.473616 + 0.208946\mathbf{u}_{t-1}^2$	0.905925	1.849098
7	Ln	KPSS 0.603626	$y_t = 3.3315 + 0.012248t$ $- 0.0000349t^2 + 1.035234y_{t-1}$ $- 0.111182y_{t-2} + \mathbf{u}_t$ $\hat{\sigma}_t^2 = 0.000855 + 0.157503\mathbf{u}_{t-1}^2$ $+ 0.6333303\sigma_{t-1}^2$	0.952469	1.943458
8	Or.	KPSS 0.295640	$y_t = 86.73834 - 0.0004793t^2$ $+ 0.00000161t^3 + 0.927494y_{t-1}$ $- 0.11479y_{t-3} + \mathbf{u}_t$ $\hat{\sigma}_t^2 = 8.8181984 + 0.244301\mathbf{u}_{t-1}^2$	0.795795	1.930754
9	Or.	ADF -3.365400	$y_t = 71.50423 + 0.05958t$ $+ 0.928686y_{t-1} + \mathbf{u}_t$ $\hat{\sigma}_t^2 = 9.868144 + 0.588863\mathbf{u}_{t-1}^2$	0.857309	1.820936
10	Or.	ADF -4.297635	$y_t = 81.54482 - 0.203219t$ $+ 0.931769y_{t-1} + \mathbf{u}_t$ $\hat{\sigma}_t^2 = 6.96792 + 0.266771\mathbf{u}_{t-1}^2$	0.974959	1.904377

Tablo 7.1'in devamı

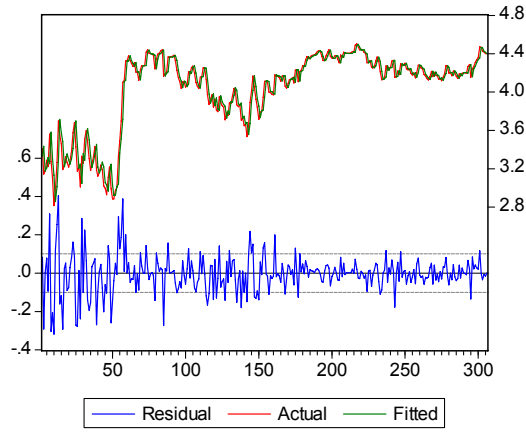
Veri Seti	Veri Tipi	Birim Kök Testi	ARCH/GARCH Modeli	R^2	Durbin Watson
11	Or.	ADF -4.121706	$y_t = 51.5446 + 1.234453y_{t-1} - 0.250644y_{t-2} + \mathbf{u}_t$ $\hat{\sigma}_t^2 = 16.15273 + 0.2206271\mathbf{u}_{t-1}^2$	0.938765	2.108657
12	Or.	ADF -4.088154	$y_t = 65.25941 + 1.167985y_{t-1} - 0.21484y_{t-2} + \mathbf{u}_t$ $\hat{\sigma}_t^2 = 23.70096 + 0.19229\mathbf{u}_{t-1}^2$	0.894303	2.010752
13	∇	ADF -12.91784	$z_t = -0.579353z_{t-4} - 0.240129z_{t-5} + 0.233813u_{t-3} - 0.528582u_{t-4} - 0.281298u_{t-5} + u_t$ $\hat{\sigma}_t^2 = 0.051849 - 0.034064\mathbf{u}_{t-1}^2 + 1.027953\sigma_{t-1}^2$	0.073702	2.047416
14	Ln	KPSS 0.589097	$y_t = 3.819415 + 1.12428y_{t-1} - 0.152496y_{t-2} + \mathbf{u}_t$ $\hat{\sigma}_t^2 = 0.000655 - 0.57564\mathbf{u}_{t-1}^2 + 0.48317\sigma_{t-1}^2$	0.902689	1.951843
15	Or.	KPSS 0.157924	$y_t = 85.42221 - 0.099484t + 1.026317y_{t-1} - 0.152496y_{t-2} + \mathbf{u}_t$ $\hat{\sigma}_t^2 = 7.765283 + 0.4577\mathbf{u}_{t-1}^2$	0.953593	1.890167
16	Or.	ADF -4.193462	$y_t = 103.9072 - 0.075407t + 1.044336y_{t-1} - 0.279595y_{t-3} + 0.169995y_{t-4} + \mathbf{u}_t$ $\hat{\sigma}_t^2 = 3.793587 + 0.870455\mathbf{u}_{t-1}^2$	0.913624	1.966509

Tablo 7.1'in devamı

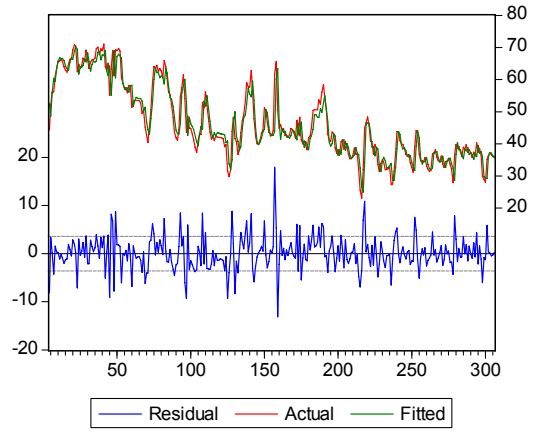
Veri Seti	Veri Tipi	Birim Kök Testi	ARCH/GARCH Modeli	R^2	Durbin Watson
17	Or.	ADF -6.595522	$y_t = 68.11425 - 0.112573t + 0.928179y_{t-1} - 0.261186y_{t-3} + 0.16539y_{t-4} + \mathbf{u}_t$ $\hat{\sigma}_t^2 = 14.42416 - 0.018669\mathbf{u}_{t-1}^2 - 0.617979\sigma_{t-1}^2$	0.913624	1.966509
18	Or.	ADF -5.005639	$y_t = 75.5117 + 1.009396y_{t-1} - 0.107319y_{t-2} + \mathbf{u}_t$ $\hat{\sigma}_t^2 = 13.65155 + 0.159929\mathbf{u}_{t-1}^2$	0.768490	2.011200
19	Or.	ADF -3.519596	$y_t = 81.0726$ $\hat{\sigma}_t^2 = 5.273568 + 1.25205\mathbf{u}_{t-1}^2$	-0.67396	0.092539
20	Or.	ADF -4.162386	$y_t = 76.18508 + 1.06898y_{t-1} - 0.1221y_{t-3} + \mathbf{u}_t$ $\hat{\sigma}_t^2 = 15.62801 + 0.171086\mathbf{u}_{t-1}^2$	0.898286	2.041062
21	Or.	ADF -3.822158	$y_t = 50.57105$ $\hat{\sigma}_t^2 = 22.04934 + 0.83584\mathbf{u}_{t-1}^2$	-0.22139	0.153689
Or.: Orjinal veri; Ln: Ln değeri alınmış veri; ∇ : Birinci fark					

Bu veri setlerin gözlem, tahmin, artık grafikleri ise Şekil 7.25'deki gibi elde edildi.

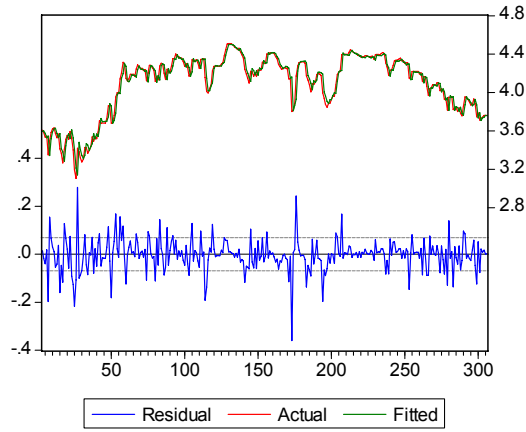
Şekil 7.25: 5-21 Arası Veri Setlerin Gözlem, Tahmin, Artık Grafikleri



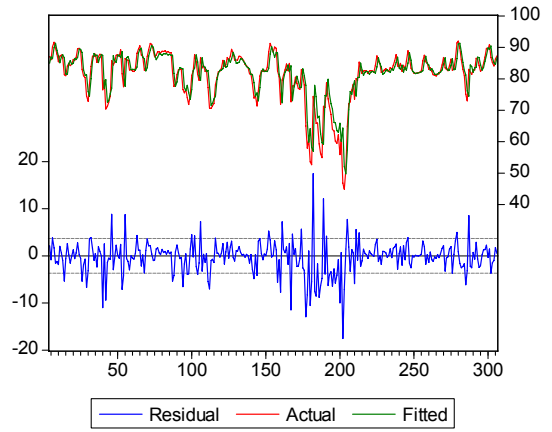
Set 5



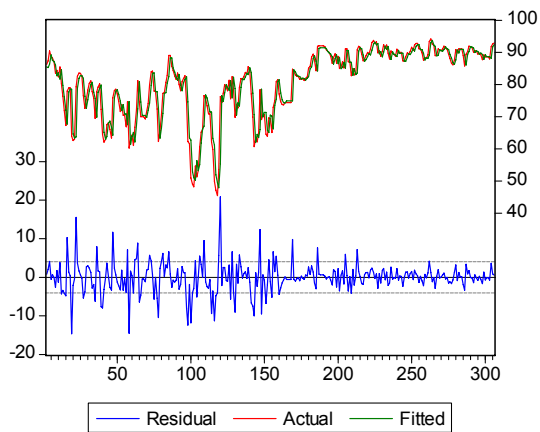
Set 6



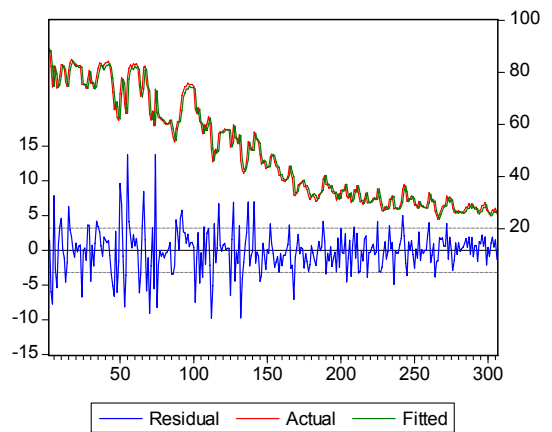
Set 7



Set 8

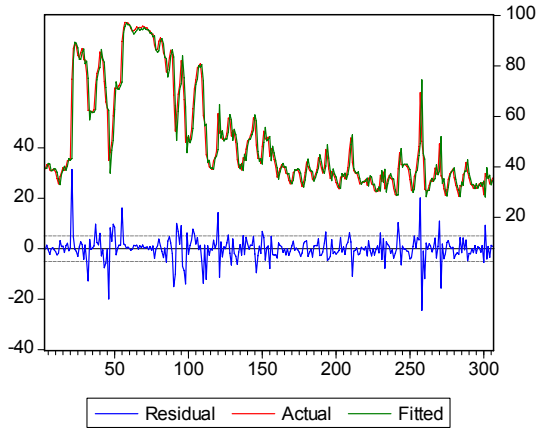


Set9

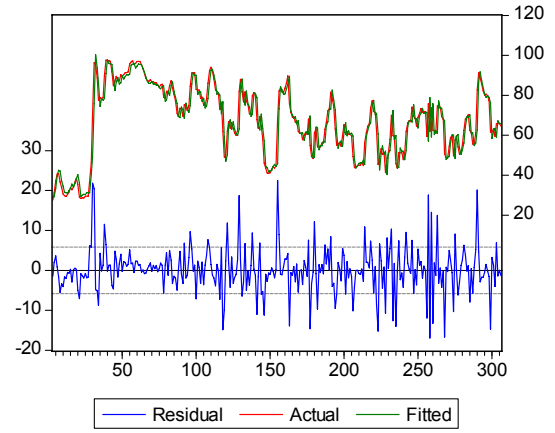


Set 10

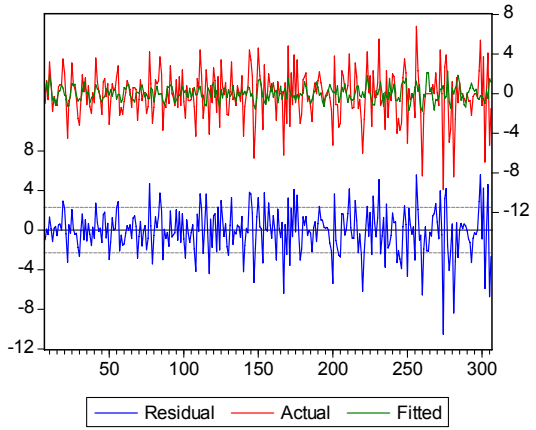
Şekil 7.25'in devamı



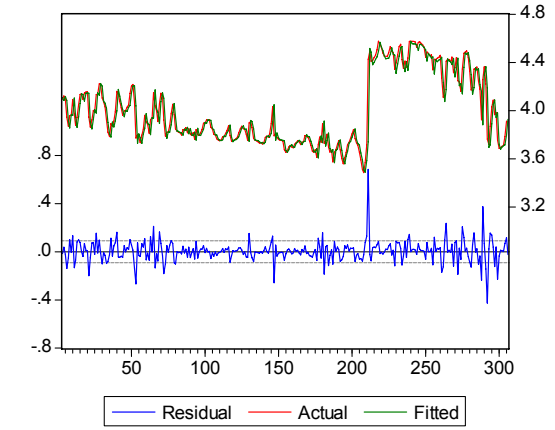
Set 11



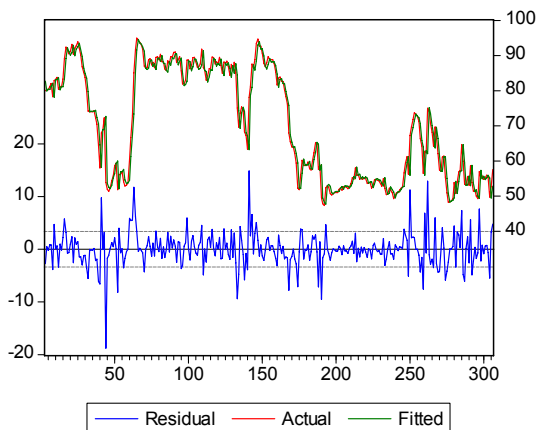
Set 12



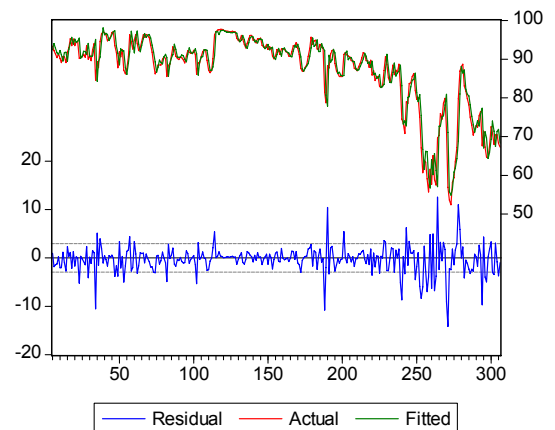
Set 13



Set 14

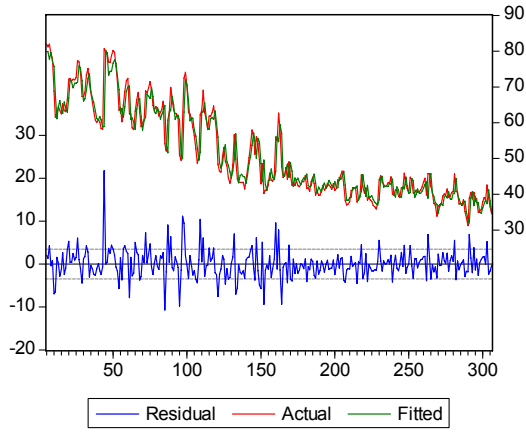


Set 15

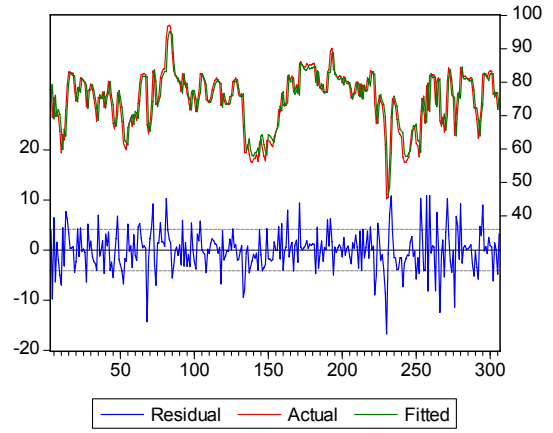


Set 16

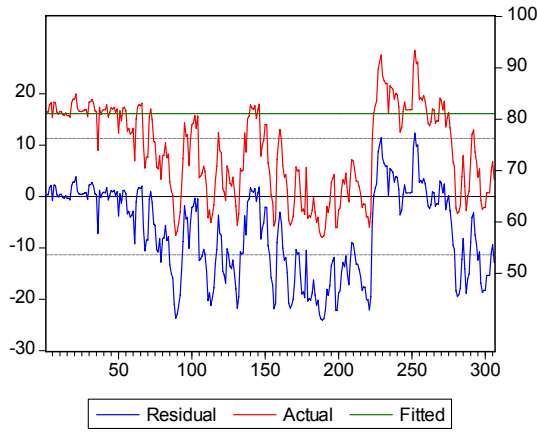
Şekil 7.25'in devamı



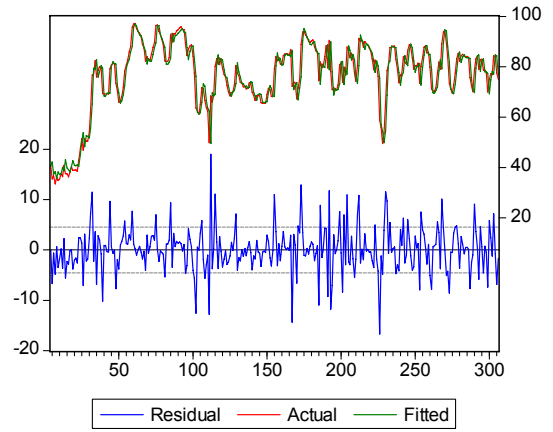
Set 17



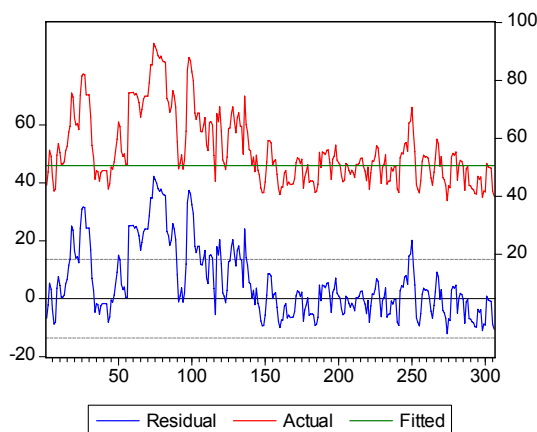
Set 18



Set 19



Set 20



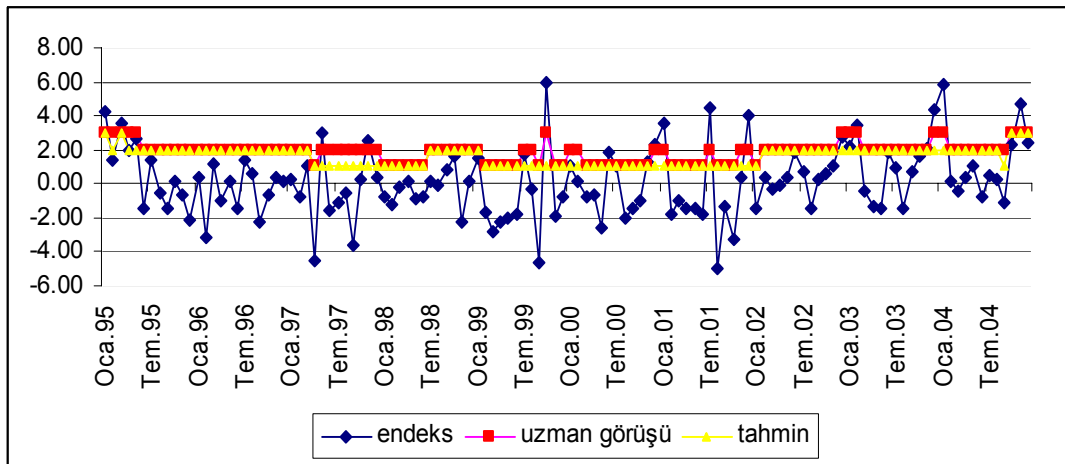
Set 21

8 DIŐ TİCARET ENDEKSLERİNİN DEĐİŐİM ORANLARININ TCM YÖNTEMİ KULLANILARAK ETİKETLENDİRİLMESİ

Bu bölümde, önerilen yöntemlerden biri olan TCM yönteminin sosyal bilimler alanında bir uygulaması yapılmak istendi. Bu amaçla dış ticaret endekslerinin önceki aya göre deđişim oranları üzerinden, ülkemizin uluslararası ticaret durum deđerlendirilmesi yapıldı. Ülkemizin yaşadığı büyük krizlerden biri olan 2001 krizininin verilerinde yer aldığı 1995-2004 yılları arasındaki endekslere ilgili zaman serisi etiketlendirme yöntemi uygulandı. Algoritmada üç küme merkezi ve $k = K / 2 = 3$ komşu sayısı ile çalışıldı ve Şekil 8.1’de yer alan grafiđe ulaşıldı. Grafikte uluslararası ticaret etiketleri aőađıdaki gibi numaralandırılarak etiketlere ait seriler elde edildi.

- 1: Uluslararası ticaret krizde
- 2: Uluslararası ticaret olađan seviyede
- 3: Uluslararası ticaret olađan üstü seviyede

Şekil 8.1: 1995-2004 Yılları Arası Aylık Dıő Ticaret Endekslerinin Etiket Serisi



Şekil 8.1 incelendiđinde Ocak 1995’de uluslararası ticaretin olađanüstü olduđu, Şubat 1995’de olađan seviyeye indiđi, Ocak 1995’de uluslararası ticaretin olađanüstü seviyeye çıktıđı, ancak Nisan 1995-Mart 1997 tarihleri arasında tekrar olađan seviyeye gerilediđi, Nisan 1997-Haziran 1998 tarihleri arasında kriz yaőandıđı,

Temmuz 1998-Ocak 1999 tarihleri arasında krizden çıkılarak uluslararası ticaretin olağan seviyesine geri döndüğü, Şubat 1999-Ocak 2002 tarihleri arasında tekrar kriz yaşandığı, Şubat 2002-Ağustos 2004 tarihleri arası krizden çıkılarak uluslararası ticaretin olağan seviyesine geri döndüğü, Eylül 2004'de uluslararası ticaretin krize girdiği buna karşın Ekim 2004'te bu krizden çıkılarak Aralık 2004'e kadar olağanüstü bir seviyenin yakalandığı TCM yöntemi kullanılarak tahmin edildi.

Uygulanan yöntem danışmansız bir yöntem olmasına karşın, tahmin gücünün belirlenebilmesi amacıyla tahminler uzman görüşüyle karşılaştırıldı ve %77.5'lik bir sınıflama kesinliğine ulaşıldığı görüldü.

SONUÇ

Bu tez iki ana bölümden oluşturuldu. İlk bölümde, K - en yakın komşuluk kuralına dayanan yeni danışmansız zaman serisi etiketlendirme algoritmaları geliştirildi. İlgili algoritmalarda, bulanık c -etiketlendirme yöntemi birleştirilmiş veri setleri üzerinde çalıştırıldı ve buradan her bir veri setinin etiketlerine ait zaman serileri çıkartıldı. Algoritmalarda, serilerdeki anlamsız oynaklığın azaltılması, daha düzgün eğri elde edebilmesi için K - en yakın komşuluk kuralı etiketlere ait zaman serilerine uygulandı.

Bu tezde geliştirilen algoritmaların etkinliğini analiz edebilmek için, ilgili algoritmalar beyin aktiviteleri ile ilgili olan bispektral endeks veri setleri üzerinde çalıştırıldı. Algoritmalara ait sınıflama kesinlikleri ikili- t testi ile karşılaştırıldı. Analiz sonucunda mod ve ağırlıklandırılmış mod değerine dayanan zaman serileri kümeleme yöntemlerinin (TCM ve TCWM) ağırlıklandırılmış ortalamaya dayanan zaman serileri kümeleme yöntemi (TCWA) ve bulanık c -ortalamlar etiketlendirme (FCML) yöntemine göre daha iyi sonuç verdiği anlaşıldı.

Sosyal bilimlerde bir örnek olarak, ülkemizin uluslararası ticaretinin zaman bazında nasıl bir seyir izlediği TCM yöntemi kullanılarak tahmin edildi. Bu yöntem danışmansız bir yöntem olmasına karşın tahminlemedeki etkinliğini görebilmek adına tahminler uzman görüşleriyle karşılaştırıldı ve tahminlerin %77.5'inin uzman görüşüyle uyumlu olduğu görüldü.

Tezin ikinci ana bölümünde, üyelik fonksiyonunun sınıflama kesinliğindeki arttırıcı rolü dikkate alındı. Veri dağılımına yakınsayan bulanık fonksiyon formüllerinin yer aldığı dört yeni teorem ortaya atıldı. Ortaya atılan ilk iki teoremden, veri dağılımının beş noktası, parametrik yamuk ve parametrik üçgen bulanık sayının çıkartılmasında kullanıldı. Geriye kalan teoremlerde, verilerin histogramıyla uyumlu olan üçgen, parametrik üçgen, parametrik üssel bulanık sayı türetme teknikleri üzerinde duruldu. Bulanık sayının çıkartılmasında kullanılan minimizasyon

problemlerinde, frekans tablolarındaki yüzdeler ve sınıf aralıklarına ait ortanoktalar referans olarak alındı.

Geliştirilen veri dağılımına yakınsayan üyelik fonksiyonlarının, sınıflamadaki etkinliklerini görebilmek için, bu üyelik fonksiyonları bispektral endeks veri setleri üzerinde uygulandı. Sınıflama maksimum üyelik derecesi kriterine göre yapıldı ve sınıflama kesinlikleri hesaplandı. Bulunan sınıflama kesinlikleri, literatürde kullanılmış olan bir başka üyelik fonksiyonu yoluyla elde edilenler ile istatistiksel olarak karşılaştırıldı ve ilgili veri setlerinde bu tezde geliştirilen üyelik fonksiyon yaklaşımlarının sınıflandırma kesinliğinde artışa yol açtıkları belirlendi.

Tez çalışması kapsamında aşağıdaki makaleler ve uluslararası bildiri yazılmıştır.

1. Time Series Labeling Algorithms Based on the K-Nearest Neighbors' Frequencies, Expert Systems with Applications.
2. Parametric Triangular Fuzzy Number Approximations Based On Data Frequencies, Knowledge and Information Systems.
3. Exponential Membership Function Evaluation Based on Frequency, Asian Journal of Mathematics and Statistics.
4. The Efficiency of Multi-Time Domain Clustering in Time Series Labeling, Ninth International Conference on Applications of Fuzzy Systems and Soft Computing, "ICAFS 2010", August 26-27, 2010, Prague, Czech Republic.

KAYNAKÇA

- [1] Agrawal, A., Kumar, N. ve Radhakrishna, M. (2007). Multispectral image classification: a supervised neural computation approach based on rough-fuzzy membership function and weak fuzzy similarity relation. *International Journal of Remote Sensing*. 28 (20):4597–4608
- [2] Abbasbandy, S. ve Amirfakhriani M. (2006). A new approach to universal approximation of fuzzy functions on a discrete set of points. *Applied Mathematical Modelling*. 30:1525-1534.
- [3] Alcalá-Fdez, J., Alcalá. R., Gacto M. J. ve Herrera, F. (2009). Learning the membership function contexts for mining fuzzy association rules by using genetic algorithms. *Fuzzy Sets and Systems*. 160:905-921.
- [4] Au, W.H., Chan, K.C.C. ve Wong, A.K.C. (2006). A Fuzzy Approach to Partitioning Continuous Attributes for Classification. *IEEE Transactions On Knowledge And Data Engineering*.18(5):715-719.
- [5] Ayhan, B., Chow M.Y. ve Trussell H.J. (2007). A Fuzzy Membership Function Design Methodology Based on Histogram and ROC Curve Analyses for Broken Rotor Bar Detection Industrial Electronics. *ISIE 2007: IEEE International Symposium on Digital Object Identifier*. 1080 – 1085.
- [6] Bowerman, B.L., O’Connell, R.T. ve Koehler, A.B. (2005). *Forecasting, Time Series, and Regression*. Amerika: Thomson Brooks Cole.
- [7] Brooks, C. (2008). *Introductory Econometrics for Finance*. Cambridge: Cambridge University Press.
- [8] Chandrakala, S. ve Sekhar, C.C. (2008). A Density based Method for Multivariate Time Series Clustering in Kernel Feature Space. *IEEE International Joint Conference on Neural Networks*. 1885-1890.
- [9] Chang, X. ve Lilly J.H. (2004) Evolutionary Design of a Fuzzy Classifier From Data. *IEEE Transactions On Systems, Man And Cybernetics, Part B: Cybernetics*. 34(4):1894-1906.
- [10] Chen, S.M. ve Shie J.D. (2009). Fuzzy classification systems based on fuzzy information gain measures. *Expert Systems with Applications*. 36: 4517–4522.
- [11] Chen, C.H., Hong, T.P. ve Tseng V.S. (2009). An improved approach to find membership functions and multiple minimum supports in fuzzy data mining. *Expert Systems with Applications*. 36:10016–10024

- [12] Cheng, C.H., Cheng, G.W. ve Wang J.W. (2008). Multi-attribute fuzzy time series method based on fuzzy clustering. *Expert Systems with Applications*. 34:1235-1242.
- [13] Chis, M. ve Grosan, C. (2006). Evolutionary Hierarchical Time Series Clustering. *Sixth International Conference on Intelligent Systems Design and Applications*. 1:451-455.
- [14] Choi, B.I. ve Chung-Hoon, F. (2009). Interval type-2 fuzzy membership function generation methods for pattern recognition. *Information Sciences*. 179: 2102–2122.
- [15] Chouakria, A.D., Diallo, A. ve Giroud, F. (2009). Adaptive clustering for time series: Application for identifying cell cycle expressed genes. *Computational Statistics and Data Analysis*. 53:1414 - 1426.
- [16] Comby, F. ve Straussi, O. (2007). Using quasi-continuous histograms for fuzzy main motion estimation in video sequence. *Fuzzy Sets and Systems*. 158:475 -495.
- [17] Devi, B.B. ve Sarma, V.V.S. (1985). Estimation of fuzzy memberships from histograms. *Information Sciences*. 35 (1):43-59.
- [18] Devillez, A. (2004). Four fuzzy supervised classification methods for discriminating classes of non-convex shape. *Fuzzy Sets and Systems*. 141:219–240.
- [19] Floreaa, M.C., Jouselme, A.L., Grenier, D. ve Bossé, E. (2008). Approximation techniques for the transformation of fuzzy sets into random sets. *Fuzzy Sets and Systems*. 159:270- 288.
- [20] Gan, T.J., Glass, P.S., Windsor, A., Payne, F., Rosow, C., Sebel, P. ve Manberg, P. (1997). Bispectral Index Monitoring Allows Faster Emergence and Improved Recovery from Propofol, Alfentanil, and Nitrous Oxide Anesthesia. *Anesthesiology*. 87(4):808-815.
- [21] Gonzalez-Abril, L., Velasco, F., Ortega, J.A. ve Cuberos, F.J. (2009). A new approach to qualitative learning in time series. *Expert Systems with Applications*. 36: 9924-9927
- [22] Greco, S., Matarazzo, B. ve Słowiński R. (2008). Parameterized rough set model using rough membership and Bayesian confirmation measures. *International Journal of Approximate Reasoning*. 49:285–300.

- [23] Guo, C., Jia, H. ve Zhang, N. (2008). Time Series Clustering Based on ICA for Stock Data Analysis. *Fourth International Conference on Wireless Communications, Networking and Mobile Computing*. 1-4
- [24] Han, J. ve Kamber, M. (2001). *Data Mining: Concepts and Techniques*. London: Morgan Kaufmann Publishers.
- [25] Hand, D., Manilla H. ve Symth P. (2001). *Principles of Data Mining*. Cambridge: MIT Press.
- [26] He, Q.P. ve Fault, W.J. (2007). Fault Detection Using the k-Nearest Neighbor Rule for Semiconductor Manufacturing Processes. *IEEE Transactions On Semiconductor Manufacturing*, 20(4):345-354.
- [27] Hellerstein, J.M. ve Stonebraker, M. (2005). *Readings in database systems*. Cambridge: MIT Press.
- [28] Hu, Y.C. (2006). Determining membership functions and minimum fuzzy support in finding fuzzy association rules for classification problems. *Knowledge-Based Systems*. 19:57–66
- [29] Kreuer, S., Biedler, A., Larsen, R., Schoth, S., Altmann, S. ve Wilhelm, W. (2001). The Narcotrend™ a new EEG monitor designed to measure the depth of anaesthesia: a comparison with bispectral index monitoring during propofol-remifentanil-anaesthesia. *Anaesthesist*. 50:921–925
- [30] Kung, C.C., Su, J.Y. (2007). Affine Takagi-Sugeno fuzzy modelling algorithm by fuzzy c-regression models clustering with a novel cluster validity criterion. *Control Theory and Applications*.1(5):1255-1265
- [31] Larose, D.T. (2005). *Discovering Knowledge in Data: An Introduction to Data Mining*. İngiltere: John and Wiley Sons.
- [32] Lütkepohl, H. ve Kratzig, M. (2004), *Applied Time Series Econometrics (Themes in modern econometrics)*, Cambridge: Cambridge University Press.
- [33] Luo, Q., Yang, W. ve Yi D. (2008). Kernel shapes of fuzzy sets in fuzzy systems for function approximation. *Fuzzy Sets and Systems*.159:270 – 288.
- [34] Li, S.T., Cheng Y.C. ve Lin S.Y. (2008). A FCM-based deterministic forecasting model for fuzzy time series. *Computers and Mathematics with Applications*, 56:3052-3063.

- [35] Mansoori, E.G., Zolghadri, M.J., Katebi, S.D. (2007). A weighting function for improving fuzzy classification systems performance. *Fuzzy Sets and Systems*. 158:583-591.
- [36] McNicholas, P.D. (2010). Model-based classification using latent Gaussian mixture models. *Journal of Statistical Planning and Inference*. 140:1175–1181.
- [37] Nasibov, E.N. (2002). Certain Integral Characteristics of Fuzzy Numbers and a Visual Interactive Method for Choosing the Strategy of Their Calculation. *Journal of Computer and Systems Sciences International*. 41(4): 584-590.
- [38] Nasibov, E.N., Baskan, O. ve Mert, A. (2005). A Learning Algorithm for Level Sets Weights in Weighted Level-based Averaging Method. *Fuzzy Optimization and Decision Making*. 4:279-291.
- [39] Nasibov, E.N., Peker, S. (2008). On the nearest parametric approximation of a fuzzy number. *Fuzzy Sets and Systems*. 159(11):1365-1375.
- [40] Nasibov EN, Ulutagay G (2010) Comparative clustering analysis of bispectral index series of brain activity. *Expert Systems with Applications* 37:2495-2504.
- [41] Nock, R., Vaillant, P., Henry, C. ve Nielsen F. (2009). Soft memberships for spectral clustering with application to permeable language distinction. *Pattern Recognition*. 42:43-53.
- [42] Otranto, E. (2008). Clustering heteroskedastic time series by model-based procedures. *Computational Statistics and Data Analysis*. 52:4685 - 4698.
- [43] Özgören, M., Kocaaslan, S., Öñiz, A. (2008). Analysis of Non-REM sleep staging with electroencephalography bispectral index. *Sleep Biol Rhythms*. 6(4):249-255.
- [44] Pal, N.R., Bezdeki J.C. (1995). On Cluster Validity for the Fuzzy c-Means Model. *IEEE Transactions on Fuzzy Systems*, 3(3):370-379.
- [45] Parthasarathy, G. ve Chatterji, B.N. (1990). A Class of New KNN Methods for Low Sample Problems. *IEEE Trans. Systems, Man and Cybernetics*. 20(3):715 -718.

- [46] Pedrycz, W. (2005). *Knowledge-Based Clustering: From Data to Information Granules*, New-York: Wiley-Interscience Publications.
- [47] Qiu, W. ve Joe, H. (2006). Separation index and partial membership for clustering. *Computational Statistics and Data Analysis*. 50:585 – 603.
- [48] Rencher, A.C.(2002). *Methods for Multivariate Analysis*. Kanada: John-Wiley Sons Publications.
- [49] Rokach, L. ve Maimon, O.Z. (2008). *Data mining with decision trees: theory and applications*. Singapore: Word Scientific Publishing.
- [50] Shuang, F., Hongxing, L., Dan, H. (2009). A new training algorithm for HHFNN based on Gaussian membership function for approximation. *Neurocomputing*. 72:1631–1638.
- [51] Shyi-Ming, C. ve Chi-Hao. C. (2005). A New Method To Construct Membership Functions And Generate Weighted Fuzzy Rules From Training Instances. *Cybernetics and Systems*. 36(4):397-414.
- [52] Simpson, P.K. (1992). Fuzzy Min-Max Neural Networks-Part 1 : Classification. *IEEE Transactions On Neural Networks*. 3(5):776-786.
- [53] Tenga, Y.W., Wanga, W.J. ve Chiu, C.H. (2004). Function approximation via particular input space partition and region-based exponential membership functions. *Fuzzy Sets and Systems*. 142:267–291.
- [54] Teoh, H.J., Chen, T.L., Cheng, C.H. ve Chu, H.H. (2009). A hybrid multi-order fuzzy time series for forecasting stock markets. *Expert Systems with Applications*. 36:7888-7897.
- [55] Tsey, R.S. (2005). *Analysis of Financial Time Series*, New-Jersey: Wiley Series.
- [56] Yang, C.C. ve Bose, N.K. (2006). Generating fuzzy membership function with self-organizing feature map. *Pattern Recognition Letters*. 27:356–365.
- [57] Yang, M.S ve Wu, K.L. (2006). Unsupervised possibilistic clustering. *Pattern Recognition*. 39:5-21

- [58] Weib, C.H. ve Gob, R. (2008). Discovering patterns in categorical time series using IFS. *Computational Statistics and Data Analysis*. 52:4369 – 4379.
- [59] Weng, X. ve Shen, J. (2008). Classification of multivariate time series using two-dimensional singular value decomposition. *Knowledge Based Systems*. 21:535 – 539.
- [60] Witten, I.H. ve Frank E. (2005). *Data Mining: Practical Machine Learning Tools and Techniques*. San Francisco: Morgan Kaufman Publishers.
- [61] Wong, K.F.K., Galka, A., Yamashita, O. ve Ozaki, T. (2006). Modelling non-stationary variance in EEG time series by state space GARCH model *Computers in Biology and Medicine*. 36(12):1327-1335.
- [62] Wu, T.P. ve Chen, S.M. (1999). A New Method for Constructing Membership Functions and Fuzzy Rules from Training Examples. *IEEE Transactions On Systems-Man And Cybernetics, Part B: Cybernetics*. 29(1): 25-40.
- [63] Yaffee, R.A. (2000). *Introduction to Time Series Analysis and Forecasting with Applications of SAS and SPSS*. New-York: Academic Press.
- [64] Yanyan, L., Xiuping, L. ve Zhixun, S. (2008). A new fuzzy approach for handling class labels in canonical correlation analysis. *Neurocomputing*. 71:1735–1740
- [65] Ye, N. (2003). *The handbook of data mining*. Routledge-Taylor and Francis. ISBN 0-8058-4081-8.
- [66] Yu, X.G. ve Yin, J. (2005). A new clustering algorithm based on KNN and DENCLUE. *Proceedings of 2005 International Conference on Machine Learning and Cybernetics*, 4: 2033 – 2038.
- [67] Zhang, X.T., Zhang, W., Xiong, X., Wang, Q.W. ve Li C.Y. (2004). A model-based clustering for time-series with irregular interval. *Proceedings of 2004 International Conference on Machine Learning and Cybernetics*. 5:2883 – 2888.