

T.C.
İNÖNÜ ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ



MEKÂNSAL BÜYÜK VERİ ANALİZİ

YÜKSEK LİSANS TEZİ

DANIŞMAN

Prof. Dr. Fatma ZEREN

HAZIRLAYAN

Meral ÖNDER

MALATYA-2022

T.C
İNÖNÜ ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
EKONOMETRİ ANA BİLİM DALI

YÜKSEK LİSANS TEZİ

MEKÂNSAL BÜYÜK VERİ ANALİZİ

Meral ÖNDER

DANIŞMAN
Prof. Dr. Fatma ZEREN

MALATYA, 2022

ONUR SÖZÜ

Bu çalışmadaki bütün bilgi ve belgeleri akademik kurallar çerçevesinde elde ettiğimi, bilimsel ahlak ve geleneklere uygun olarak sunduğumu, yararlandığım bütün yapıtların hem metin içinde hem de kaynakçada yöntemine uygun biçimde gösterilenlerden oluştuğunu belirtirim, Prof. Dr. Fatma ZEREN danışmanlığında tarafımdan yazıldığını ve İnönü Üniversitesi Sosyal Bilimler Enstitüsü Tez yazım kurallarına göre yazıldığını beyan eder ve bunu onurumla doğrularım.

Tarih

Meral ÖNDER

İmza

TEŞEKKÜR

Lisans ve lisansüstü eğitim hayatım boyunca gerek kütüphanesinden gerekse de ilminden faydalandığım ve Yüksek lisans tezi olarak sunduğum bu çalışmanın her aşamasında sonsuz destek ve yardımlarını esirgemeyen değerli hocam ve tez danışmanım sayın Prof. Dr. Fatma Zeren'e sonsuz teşekkürlerimi sunuyorum.

Yüksek lisans tezimin hazırlanmasında gerekli düzeltmeler ve önerilerde bulunan değerli hocalarım Doç. Dr. Kadir KARAGÖZ ve Doç. Dr. Sinem Güler KANGALLI UYAR, bana her daim yardımlarını esirgemeyen hocamın doktora öğrencisi Ayşegül Han'a teşekkürü bir borç bilirim.

Bu günlere gelmemde büyük emekleri olan canım annem Medine Aykanat ve babam Necati Aykanat'a, tüm bu sürecin en başından yanımda bulunan ve bana her türlü imkânı sunan canım eşim Ali Önder'e, tezimi çalışma fırsatına izin veren kızım Elif Önder'e teşekkür ederim.

Meral ÖNDER
MALATYA, 2022

ÖZET

Mekânsal olarak düzenlenmiş verilerin birbirine olan bağımlılığı, bağımsızlık varsayımının ihlal edilmesine neden olmaktadır. Bu nedenle mekânsal istatistiksel analiz, diğer istatistiksel analizlerden ayrılmaktadır. Mekânsal bağılılığa neden olan komşuluk ilişkileri mekânsal ağırlık matrisleri ile tanımlanır. Mekânsal bağımlılığın yapısına göre kurulan regresyon modelleri de farklılaşmaktadır. Bu modellerden en sık kullanılan mekânsal regresyon modelleri, mekânsal hata modelleri (SEM) ve mekânsal gecikme modelleri (SLM) şeklindedir. Mekânsal modellerde içsellik problemi nedeniyle en küçük kareler (EKK) yöntemi kullanılamaz, bunun yerine genellikle maksimum olabilirlik yöntemi uygulanır ve mekânsal modellere ait logaritmik olabilirlik fonksiyonu maksimize edilir. Maksimizasyon işlemi yapılırken logaritmik determinantın hesaplanması örneklem büyüklüğüne bağlıdır ve mekânsal bağımlılık parametresinin her bir değeri için ayrı ayrı determinant hesaplanır. Büyük boyutlu verilerde ise yüksek boyutlu matrislerin determinantının hesaplanması oldukça zordur. Çünkü bu matrislerin içerisinde yüksek oranda sıfır olması öz değerlerin bulunmasında hatalı sonuçlara neden olmaktadır. Bu problemi ortadan kaldırmak için mekânsal büyük veri modellerinde geçerli yöntemler mevcuttur. Bunlar, Üstel Mekânsal Matris Tanımı (MESS), Tek Taraflı Yaklaşıklık Yöntemi ve Kompozit Olabilirlik Yaklaşımıdır. MESS, hem mekânsal hata modelleri hem de mekânsal gecikme modelleri için hesaplamada kolaylık sağlamaktadır. Bu yaklaşıma alternatif olarak kovaryans matrislerinin tersine çevrilmesindeki hesaplama zorluklarını ortadan kaldırmak için tek taraflı yaklaşıklık yöntemi kullanılmaktadır. Bu yöntemin dışında mekânsal hata terimlerindeki ilişkileri modellemek için kompozit olabilirlik yaklaşımı, içsellik problemini gidermek için Genelleştirilmiş İki Aşamalı En Küçük Kareler yöntemleri uygulanmaktadır.

Yapılan bu tez çalışmasında öncelikle mekânsal büyük veri modellerinin tahmin yöntemlerinde kullanılan bu yaklaşımlar açıklanmıştır. Daha sonra bu yaklaşımlardan MESS yöntemi kullanılarak ABD'nin 1990-2000 yılı on yıllık ilçe düzeyindeki veriler kullanılarak cinayet oranlarını etkileyen faktörlerin etkisi araştırılmıştır. Bu çalışmada kullanılan veri setinin büyük boyutlu olması nedeni ile MESS modeli ile tahmin gerçekleştirilmiştir. MESS tahmin sonuçlarına göre cinayet oranları üzerinde etkili olan değişkenler tespit edilmiştir.

Anahtar Kelimeler: Ağırlık Matrisi, Büyük Veri, Mekânsal Bağımlılık, Mekânsal Büyük Veri, Suç.



ABSTRACT

The interdependence of spatially organized data leads to a violation of the independence assumption. For this reason, spatial statistical analysis differs from other statistical analyses. Neighborhood relations that cause spatial commitment are defined by spatial weight matrices. Regression models established according to the structure of spatial dependence also differ. The most frequently used spatial regression models among these models are spatial error models (SEM) and spatial delay models (SLM). The least squares (LCS) method cannot be used in spatial models due to the interiority problem, instead, the maximum likelihood method is generally applied and the logarithmic likelihood function of the spatial models is maximized. The calculation of the logarithmic determinant while performing the maximization process depends on the sample size and the determinant is calculated separately for each value of the spatial dependence parameter. In large dimensional data, it is very difficult to calculate the determinant of high dimensional matrices. Because the high percentage of zeros in these matrices causes erroneous results in finding the eigenvalues. There are valid methods in spatial big data models to eliminate this problem. These are Exponential Spatial Matrix Definition (MESS), One Sided Approximation Method and Composite Likelihood Approach. MESS provides ease of computation for both spatial error models and spatial delay models. As an alternative to this approach, one-sided approximation method is used to eliminate computational difficulties in inverting covariance matrices. Apart from this method, composite likelihood approach is applied to model the relationships in spatial error terms, and Generalized Two-Stage Least Squares methods are applied to solve the internality problem.

In this thesis, first of all, these approaches used in the estimation methods of spatial big data models are explained. Then, using the MESS method, one of these approaches, the effects of the factors affecting the homicide rates were investigated by using the ten-year district level data of the USA for the years 1990-2000. Due to the large size of the data set used in this study, estimation was made with the MESS model. According to the MESS estimation results, the variables affecting the murder rates were determined.

Keywords: Weight Matrix, Big Data, Spatial Dependency, Spatial Big Data, Crime.

İÇİNDEKİLER

ONUR SÖZÜ.....	i
TEŞEKKÜR	ii
ÖZET	iii
ABSTRACT.....	v
TABLolar LİSTESİ	viii
ŞEKİLLER LİSTESİ	ix
KISALTMALAR	x
1. GİRİŞ	1
1.1. Mekânsal Ağırlık Matrisi	3
1.1.1. Sınırdışlığa Bağlı Ağırlıklar (Contiguity Weight)	3
1.1.2. Uzaklığa Bağlı Ağırlıklar	7
1.2. Mekânsal Etkiler	8
1.2.1. Mekânsal Bağımlılık.....	8
1.2.2.1. Global Mekânsal Otokorelasyon Testi.....	10
1.2.2.2. Lokal Mekânsal Otokorelasyon Testi	13
2. BÜYÜK VERİ	16
2.1. Büyük Veri Türleri	17
2.2. Büyük Veri Bileşenleri.....	18
2.2.1. Hacim (Volume).....	18
2.2.2. Hız (Velocity)	19
2.2.3. Çeşitlilik (Variety)	19
2.2.4. Değer (Value)	19
2.2.5. Doğrulama (Verification).....	20
3. MEKÂNSAL BÜYÜK VERİ ANALİZİ.....	21
3.1. Doğrusal Regresyon Modelleri.....	21

3.1.1. İçsellik.....	23
3.2. Mekânsal Regresyon Modelleri	25
3.2.1. Mekânsal Hata Modeli (SEM).....	26
3.2.2. Mekânsal Gecikmeli Model (SAR).....	28
3.3. Mekânsal Regresyon Modelleri İçin Tahmin Yöntemleri.....	29
3.3.1. Maksimum Olabilirlik (ML) Yöntemi.....	29
3.3.2. Genelleştirilmiş İki Aşamalı En Küçük Kareler Yöntemi.....	32
3.4. Mekânsal Büyük Veri Tahmin Edicileri.....	36
3.4.1. Mekânsal Üstel Matris Belirtimi (MESS)	37
3.4.1.1. SAR İçin Mekânsal Üstel Matris Belirtimi.....	37
3.4.1.1.1. SAR Mekânsal Üstel Matris Belirtimi için Maksimum olabilirlik yaklaşımı	41
3.4.1.2. SEM İçin Mekânsal Üstel Matris Belirtimi	44
3.4.1.2.1. Bayesian Yöntemi	45
3.4.2. Tek Taraflı Yaklaşımlar Yöntemi.....	48
3.4.2.1. Mekânsal Gecikme Modellerinde İzotropi	50
3.4.2.2. Mekânsal Gecikme Modellerinde İzotropinin Test Edilmesi	53
3.4.3. Kompozit Olabilirlik Yaklaşımı.....	54
3.4.3.1. SEM Tahmini İçin İki Değişkenli Marjinal Olabilirlik Yaklaşımı ..	55
4. AMPİRİK ANALİZ	58
5. SONUÇ	70
KAYNAKÇA.....	72

TABLULAR LİSTESİ

Tablo 1: Yapılandırılmış Veri Örneği.....	17
Tablo 2: Veri Birimleri Ölçüsü.....	18
Tablo 3: Değişken Tanımları	60
Tablo 4: Global Otokorelasyon Test Sonuçları	63
Tablo 5: Mekânsal Ekonometrik Modellerin Tahmin Sonuçları	63
Tablo 6: MESS Yöntemi Katsayı Tahmin Sonuçları.....	67
Tablo 7: MESS Ve SAR Yöntemi Katsayıları.....	68



ŞEKİLLER LİSTESİ

Şekil 1: Hücreler Arası Ortak Sınır Şekilleri.....	4
Şekil 2: Düzensiz Bir Kafes Veri Örneği	5
Şekil 3: Mekânsal Otokorelasyon Örnekleri	10
Şekil 4: Düzenli Bir Kafes Izgara Üzerine Düzenlenmiş Üç BölgeHata! Yer işareti tanımlanmamış.	
Şekil 5: Düzenli Kare Kafeslere Sahip İki Ayrı Ağırlık Matrisi Örneği.....	50
Şekil 6: ABD'de Yaşanan Cinayet Oranlarının İlçelere Göre Mekânsal Dağılımı	62
Şekil 7: Mekânsal Ekonometrik Modellerin Keman Grafiği	66



KISALTMALAR

DESTE:	En İyi Sapmasız Tahmin Edici
EGLS:	Tahmini Genelleştirilmiş En Küçük Kareler
EKK:	En Küçük Kareler Yöntemi
ESDA:	Keşifsel Mekânsal Veri Analizi
GEKK:	Genelleştirilmiş En Küçük Kareler
GISA:	Global Indicators of Spatial Association
GMM:	Genelleştirilmiş Moment Yöntemi
GSM:	Genelleştirilmiş Mekânsal Model
GS2SLS:	Genelleştirilmiş Mekânsal İki Aşamalı En Küçük Kareler
IV:	Araç Değişken
LISA:	Local Indicators of Spatial Association
LM:	Lagrange Çarpan
MESS:	Mekânsal Üstel Matris Belirtimi
ML:	Maksimum Olabilirlik Yöntemi
MLE:	Maksimum Olabilirlik Tahmin Edicileri
NIG:	Normal Ters Gama
SAC:	Mekânsal Kombo Modeli
SAR:	Mekânsal Gecikme Modeli
SEM:	Mekânsal Hata Modeli
SLM:	Mekânsal Gecikme Modeli
2SLS:	İki Aşamalı En Küçük Kareler

1. GİRİŞ

Mekânsal ekonometrik modellerin tahmini ve test edilmesi ilk olarak Moran (1948), Geary (1954), Whittle (1954) tarafından önerilmiştir ve mekânsal bağımlılığın matematiksel tanımlarını ortaya koymuşlardır. Çalışmalarında uzaydaki durağan süreçlerin, zaman serilerindeki durağan süreçlere benzemediğini ifade etmişlerdir ve bir zaman serisindeki değişim yalnızca geçmişteki değerlerden etkilenirken, mekânsal alanlardaki değişimin ise tüm yönlerden (tüm komşu/yakın konum veya tüm benzerlerden) etkilendiğini vurgulamıştır.

Büyük ölçekli geoistatistiksel verilerin modellenmesinde çok sayıda çalışmalar mevcuttur. Geoistatistiksel veri, mekânsal veri kümelerini dikkate alan bir istatistik dalıdır. Bu çalışmaların çoğunda mekânsal etkiler dikkate alınmıştır. Mekânsal etki, mekânsal bağımlılığı ve mekânsal heterojenliği ifade etmektedir. Mekânsal bağımlılığın nedeni olan komşuluk ilişkileri (ülke, bölge, şehir, vb.) ağırlık matrisleri ile ölçülmüş ve regresyon yapısında ise bu mekânsal bağımlılık veya mekânsal etkileşim dahil edilerek mekânsal regresyon modelleri elde edilmiştir.

Mekânsal otokorelasyon kavramı Washington Üniversitesi'ndeki iki coğrafyacı Garrison ve Ullman'ın çalışmalarına dayanarak doğmuştur. Bu üniversite, mekânsal ekonometri alanında çalışma yapan birçok akademisyen ve öğrenci yetiştirmiştir ve bu öğrencilerden biri de Waldo Tobler'dir. Tobler mekânsal istatistik ve mekânsal ekonometrinin önde gelen temsilcilerinden biridir. Washington Üniversitesi bu yüzden mekânsal kavramları geliştiren bir okul kabul edilir.

1968'e kadar mekânsal otokorelasyon, "mekânsal bağlılık", "mekânsal ilişki", "mekânsal etkileşim" gibi terimler ile ifade edilmiştir. 1968'den sonra, Cliff ve Ord (1968) "mekânsal otokorelasyon" terimini kullanmışlar ve böylece mekânsal otokorelasyon kavramını literatüre kazandırmışlardır. Cliff ve Ord daha sonra 1981'de mekânsal süreçlere ve bu süreçlere ait modeller hakkında detaylı bir çalışma yapmıştır.

Mekânsal ekonometri kavramını ilk kez Paelinck kullanmış ve çalışmalarında mekânsal ekonometriyi çok boyutlu ekonometrik problemlerin etkisi altında tahmin ve test problemleri olarak ele almıştır. Daha sonra 1976 yılında Klaassen ile konuya ilgili ilk kitaplardan "Spatial Econometrics" isimli kitabı yazmışlardır (Getis, 2008: 304). Yazdıkları bu kitapta mekânsal ekonometrinin tanımından bahsetmemişlerdir, bunun

yerine mekânsal ekonometrinin tanımını oluşturacak beş önemli kural açıklamışlardır. Bu kurallar;

- Mekânsal ilişkinin önemi,
- Mekânsal ilişkiler arasındaki asimetri,
- Diğer bölgelerdeki faktörlerin etkisi,
- Mekânsal etkileşim olmadan önce ve olduktan sonra arasındaki farklılaşma,
- Mekânsal modellerdeki bölgelerin modellemeleri şeklindedir.

Mekânsal ekonometri konusunda önemli bir çalışma Anselin (1988) tarafından gerçekleştirilmiştir. Anselin bu çalışmasında, “Spatial Econometrics: Methods and Models” kitabını yazmıştır. Bu kitap mekânsal ekonometrinin başeserlerinden biri olmuştur. Bu kitapta mekânsal ekonometrik modeller ve bu modellerin tahmin süreçleri hakkında oldukça ayrıntılı ve yararlı bilgiler vermiştir. Anselin (1988) mekânsal etkileri, mekânsal bağıllık ve mekânsal heterojenlik olarak ikiye ayırmıştır.

Ord (1975) regresyon modellerinin bağımlı değişken ve hata terimlerinde mekânsal bağımlılık olması durumunda mekânsal ekonometrik model kullanılmasını önermiştir. Mekânsal ekonometrik modellerin tahmini için en çok olabilirlik tahmincisini (MLE) geliştirmiştir. Aynı zamanda mekânsal ekonometrik modeller için Anselin (1980), araç değişkenler (Instrumental Variables, IV) yöntemini ve Kelejian ve Prucha (1998, 1999), genelleştirilmiş momentler yöntemini (GMM) geliştirmişlerdir. Cliff ve Ord (1972), mekânsal ekonometrik modellere ait EKK artıkları için Moran I testini geliştirmişlerdir.

Anselin (1995), local Moran I ve local Geary’ C istatistiğini geliştirerek “hot spots” kavramını ortaya çıkarmıştır. Bu istatistiklere mekânsal ilişkinin yerel göstergesi LISA istatistiği (Local Indicators of Spatial Association) adını vermiştir. Son yıllarda mekânsal ekonometri ile ilgili olarak çalışmalar giderek artmaktadır.

Mekânsal ekonometrik modellerde yazılım alanındaki gelişmeler 1990 yılında giderek artan bir yoğunluk kazanmıştır. Anselin (1992) tarafından SpaceStat yazılımı yayınlamıştır. LeSage (1999) ve Pace (1998) MATLAB için mekânsal toolbox geliştirmişlerdir. Mekânsal ekonometrik analizler için R programında, Roger Bivand (2002) ve ekibi “spdep” paketini oluşturarak yayınlamışlardır. Luc Anselin, GeoDa isimli mekânsal programı geliştirmiştir. GIS yazılımlarından ESRI ArcGIS, 2006 yılında yayınladığı 9.2 versiyonuna mekânsal toolbox eklemiştir.

1.1. Mekânsal Ağırlık Matrisi

Mekânsal ekonometrik çalışmalarında mekânsal otokorelasyonu (ardışık bağımlılığı, komşuluk ilişkisi) ifade etmek için genellikle mekânsal ağırlık matrisi oluşturulur. Bu ağırlıklar etkileşimin bir ölçüsünü verir. Bu ağırlıklara bağlı olarak mekânsal ekonometrik model oluşturulur ve daha sonra modelin tahmin aşaması gerçekleştirilir.

Cliff ve Ord (1972)'a göre, mekânsal ağırlık matrisinin içerisindeki bilgi rassal olarak elde edilmemiştir. Tamamen modelin içindeki birimlere ait bilgiler ile belirlenir, seçilen ağırlıklar oldukça önemlidir. Çünkü oluşturulan modellerin tahmini bu oluşturulan ağırlıkların seçimine karşı oldukça hassastır.

Zeren (2011)'in çalışmasında belirttiği üzere birbirinden farklı alanların komşuluk ilişkilerine bağlı olarak ağırlıklar oluşturulur. Her alanın kendine ait komşuluk ağırlığı vardır. S_n ve S_m iki farklı alan olsun. Bu iki alana ilişkin ağırlık w_{nm} , n ve m alanlarına ait ağırlıklar listesi bileşenidir. Eğer bu S_n ve S_m alanları arasında komşuluk yok ise komşuluk matrisindeki ağırlık $w_{nm} = 0$ olarak tanımlanır.

Mekânsal ekonometri yöntemlerinde ağırlık matrislerinin önemli bir yeri vardır (Arbia, 2014: 28). Tahminlerin asimptotik sapmasız olabilmesi için aşağıdaki gibi tanımlanan mekânsal ağırlık matrisi W ile verilen bağımlılığın aralığı belirlenir.

$$W = \begin{bmatrix} w_{11} & \cdots & w_{n1} \\ \cdots & \cdots & \cdots \\ \cdots & w_{ij} & \cdots \\ \cdots & \cdots & \cdots \\ w_{1n} & \cdots & w_{nn} \end{bmatrix}_{N \times N}$$

N tane alanın birbirlerine göre mekânsal olarak nasıl konumlandığını gösteren $N \times N$ boyutlu bir matris elde edilir. Birimler arası komşuluk ilişkilerini belirlemede kullanılan farklı komşuluk kriterleri mevcuttur. Bu ağırlık matrisinin elemanları sınırdaşlığa bağlı ve uzaklığa bağlı ağırlıklara göre oluşturulur.

1.1.1. Sınırdaşlığa Bağlı Ağırlıklar (Contiguity Weight)

Zeren (2011)'in çalışmasında sınırdaşlığa bağlı komşuluk kriteri, sınırları ayırt edilebilen bir haritanın varlığını varsayar ve komşulukların iki konumun birbirine olan yakınlığına bağlı olduğunu ifade eder. Yani iki konum ortak bir sınırı paylaşıyorlarsa komşu olarak ifade edilebilir. Genellikle iki değerli ağırlık matrisi kullanılır (Arbia, 2014:

37). Bu durumda iki değerli ağırlık matrisi negatif olmayan ve simetrik ağırlık matrisinin elemanlarından oluşmaktadır ve aşağıdaki gibi tanımlanır;

$$w_{ij} = \begin{cases} 1 & j \in N(i) \\ 0 & \text{diğer} \end{cases} \quad (1)$$

Bu alanlar mekânsal olarak birbirine yakın ve ortak bir sınır paylaşıyor ise komşu olarak tanımlanır ve W matrisindeki w_{ij} 1 değerini alır. Eğer alanlar mekânsal olarak birbirine yakın değil yani ortak bir sınırları mevcut değil ise komşu alanlar değildir ve $w_{ij}=0$ değerini almaktadır. Ayrıca, her alan kendisinin komşusu olamayacağı için w_{ii} değeri 0 olarak alınır. Bu yakınlık ölçüsünün seçimi ile simetrik bir komşuluk matrisi $w_{ij} = w_{ji}$ elde edilebilir.

Bazı durumlarda ağırlık matrisinde satır standartlaştırması yapılabilir (Arbia, 2014: 28-19). Bu durumda her satır toplamı $\sum_j w_{ij}^* = 1$ olur. Böylelikle ağırlık matrisinin standartlaştırılan elemanları aşağıdaki gibi yazılabilir;

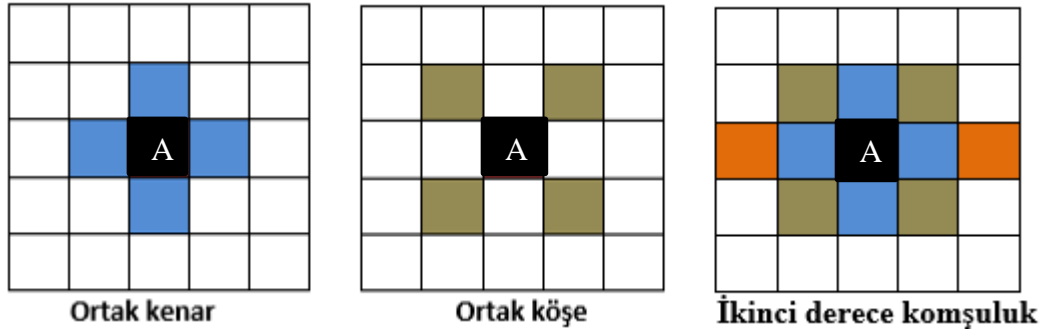
$$w_{ij}^* = \frac{w_{ij}}{\sum_j w_{ij}} \quad (2)$$

Mekânsal ağırlık matrisi satır standartlaştırması; ağırlık matrisinin her bir satırındaki elemanlar o satırın toplamına bölünür. Bu işlem mekânsal ağırlıkların 0 ile bir arasında olmasını ve modeller arasında parametrelerin kıyaslanmasını sağlar (Arbia, 2014: 28-19). Ağırlık matrisleri simetrik olmasına rağmen satır standartlaştırmasından sonra bu özelliğini kaybetmektedir.

$$\sum_j w_{ij} \neq \sum_i w_{ji}$$

Sınırdışlığın belirlenmesi için çok farklı yaklaşımlar mevcuttur. Arbia(2014), düzenli bir şebeke içerisinde A hücresi (şekildeki siyah renkli hücre) ile onu çevreleyen hücreler arasındaki ortak sınırı aşağıdaki şekillerle ifade etmiştir. Aşağıdaki şekilde her bir renk bir hücreyi ifade etmektedir.

Şekil 1: Hücreler Arası Ortak Sınır Şekilleri



Sınır komşuluğunu bir satranç oyununa benzeterek farklı komşuluk tanımları geliştirilmiştir. Bunlar kale (rook), fil (bishop) ve vezir (quenn) komşuluğudur. Bunların yanı sıra doğrusal komşuluk adı verilen bir komşuluğun varlığı da ortaya konulmuştur (LeSage, 1999. 11-13).

Doğrusal Komşuluk: $w_{ij} = 1$ ele alınan alanın sağında veya solunda ortak bir kenarı paylaşmaktadır. Ortak bir kenar paylaşmıyorsa $w_{ij} = 0$ 'dır.

Kale (rook) Komşuluk: $w_{ij} = 1$ ise, bölgeler (i ve j) ortak bir kenarı paylaşmaktadırlar.

Fil (bishop) Komşuluk: $w_{ij} = 1$ ise, i ile j ortak bir köşeyi paylaşan komşulardır.

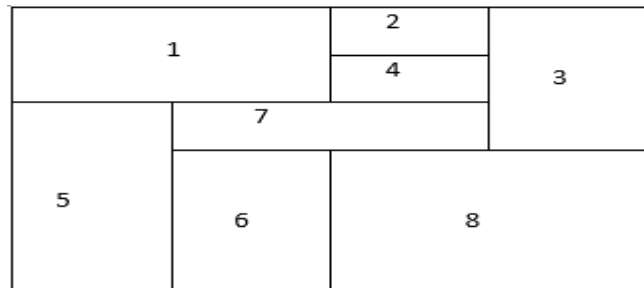
Çift Doğrusal Komşuluk: İlgilenilen alanda c bir bölge olsun. İki komşu olan i ve j, ilgili c bölgesinin hemen sağında veya solunda yer alır.

Çift Kale (rook)Komşuluğu: İlgilenilen alanda c bir bölge olsun. İki komşu olan i ve j, ilgili bölgenin sağı, solu, kuzey ve güneyi için kullanılır.

LeSage (2009) çift kale ve doğrusal komşuluğu, “ikinci dereceden komşuluk” olarak adlandırırken, diğer tanımlar “birinci dereceden komşuluk” olarak nitelendirilmiştir. Yukarıdaki son şekil ikinci dereceden komşuluğu göstermektedir.

Düzensiz bir kafes veri örneği ele alınacak olursa aşağıdaki şekil kullanılarak mekânsal ağırlık matrisi belirlenmektedir.

Şekil 2: Düzensiz Bir Kafes Veri Örneği



Şekil 2, 8 adet bölgeye ait kurgusal bir veridir. Bölge sayısı 8 olduğundan 8×8 boyutlu bir mekânsal ağırlık matrisi, 64 tane ağırlık matrisi elemanından meydana gelmektedir. Bu elemanlar 0 veya 1 değerini almaktadırlar.

Şekil 2'ye göre çeşitli komşuluk tanımları için ağırlık matrisinin çeşitli elemanları şöyle yazılabilir. Kale (rook) komşuluğuna göre; 3. satır için $w_{38} = 1$, $w_{34} = 1$, $w_{37} = 1$, $w_{32} = 1$ 'dir ve diğer satır elemanları sıfırdır.

Şekil 2 için birinci mertebeden kale (rook) komşuluğu için ağırlık matrisi aşağıdaki gibidir:

$$W = \begin{bmatrix} 0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 \end{bmatrix}$$

Yukarıdaki mekânsal ağırlık matrisi W, simetriktir ve diyagonaldır. Şebeke alanlar kendi kendine komşu olamayacağından asal köşegen elemanları sıfır kabul edilir. Bu matrisin uygun bir dönüşümü satır toplamı 1 olan bir matrise dönüştürmektedir. Buna satır standartlaştırılması yapılmış ağırlık matrisi denir. W matrisinin her satırındaki bilgiyi toplamak için standartlaştırılır. Yukarıdaki W matrisinin satır standartlaştırılması yapılmış hali (C) aşağıdaki biçimde gösterilebilir:

$$C = \begin{bmatrix} 0 & \frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{4} & 0 & \frac{1}{4} & 0 \\ \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{4} & 0 & \frac{1}{4} & 0 & 0 & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & 0 & 0 & \frac{1}{4} & 0 \\ \frac{1}{3} & 0 & 0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{6} & 0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & 0 & \frac{1}{6} \\ 0 & 0 & \frac{1}{3} & 0 & 0 & \frac{1}{3} & \frac{1}{3} & 0 \end{bmatrix}_{8 \times 8}$$

1.1.2. Uzaklığa Bağlı Ağırlıklar

Mekânsal yakınlığı ölçmek için başka alternatif çözümlerde bulunmaktadır. S_n ve S_m alanları arasında sınır olmadan da yakın çevredeki alanları kapsayacak bir komşuluk matrisi oluşturulabilir. Alanlar arasındaki uzaklık azaldıkça aralarındaki mekânsal ardışık bağımlılık ve mekânsal otokorelasyon artmaktadır. Uzaklık, öklid uzaklığı ile ifade edilebilir ve genellikle d_{ij} olarak gösterilir (Zeren, 2011: 15). Ağırlık matrisinin elemanları uzaklık fonksiyonu cinsinden aşağıdaki gibi ifade edilebilir:

$$w_{ij} = g(d_{ij}) \quad (3)$$

Kritik değer komşuluğu; $0 \leq d_{ij} < d^*$ ise, i ve j konumları birbirine komşudur. Burada d^* , kritik değerdir. O halde ağırlık matrisi W 'nin elemanlarını aşağıdaki gibi yazmak mümkündür:

$$w_{ij} = \begin{cases} 1 & 0 \leq d_{ij} < d^* \\ 0 & \text{aksi halde} \end{cases} \quad (4)$$

Zeren (2010), bu fonksiyon yapısının genellikle $w_{ij} = 1/d$ veya $w_{ij} = 1/d^\alpha$ olarak ele alındığını ifade etmiştir. i ve j bölgelerinin merkezleri arasındaki uzaklık d_{ij}

olmak üzere en sık kullanılan komşuluk yapısı en yakın k komşuluk ve kritik değer komşuluğudur.

En yakın k komşuluğu, S_n konumunun k tane komşu konumu olduğu varsayıldığında, en kısa uzaklıkta olan konuma göre komşu seçilir ve ağırlık matrisinin elemanları buna göre belirlenir. Bu özelliğe göre S_j konumu S_i konumunun en yakın komşusu ise $d_{ij} = \text{Min}(d_{ik})$ 'dir. Burada k 'nın değeri dışsal olarak belirlenir.

1.2. Mekânsal Etkiler

Mekânsal veri analizinde, mekânsal etkileşim ön plana çıkmaktadır. Mekânsal etkileşim hem mekânsal bağımlılığı hem de mekânsal heterojenliği kapsamaktadır (Anselin, 2001: 566). Mekânsal verideki bağımlılık yapısı regresyon analizini de içeren çeşitli istatistiksel yöntemlerde geçerli olan bağımsızlık varsayımının ihlal edilmesine neden olmaktadır.

1.2.1. Mekânsal Bağımlılık

Haritada konum olarak birbirine yakın konumda olan gözlemler birbirlerine benzer olma eğilimine girerler. Ünlü coğrafyacı akademisyen Waldo Tobler'e göre coğrafyanın temel yasası şöyledir; "Her şey başka her şeyle ilişkilidir. Fakat yakın şeyler, uzak şeylere göre daha ilişkilidir." (Miller, 2004: 284-285). Coğrafya ile olan ilişkisi ile bilinen Tobler tarafından mekânsal otokorelasyon dile getirilmiştir.

Miller (2004) tarafından da belirtildiği gibi, birbirine yakın olan mekânlara ilişkin verilerin uzakta bulunan verilere göre daha fazla benzerlik göstermesi bağımlılık yapısını ortaya koymaktadır. Bu verilere sıradan istatistiksel analizlerin uygulanması çeşitli problemlere neden olmaktadır.

Mekânsal veriler; bir uzaydaki R çalışma alanındaki konumları ve diğer bilgileri ile birlikte gözlenirler (Zeren, 2011: 5). Uzay-zamanın bir veya birden fazla noktasındaki kesit birimlerini oluşturur. Burada, değişkenler coğrafik konumlara göre sıralanmış ise rassal değişkenler dizisidir ve $x_{s1}, x_{s2}, \dots, x_{sn}$ şeklinde ifade edilir. S, mekânsal konuma ilişkin bir indekstir. Zamana göre sıralanmışsa $x_{t1}, x_{t2}, \dots, x_{tn}$ şeklinde gösterilir ve bu değişkenler dizisine rassal süreç adı verilir. Benzer olarak konumlara göre sıralanmış rassal değişkenler dizisine mekânsal rassal süreç veya mekânsal rassal alan adı verilir.

Çalışma alanındaki örneklem verilerinin mekânsal bağımlılığı; i konumundaki bir gözlemin j konumundaki diğer gözlemlere bağımlılığını ifade etmektedir ve aşağıdaki gibi gösterilebilir:

$$y_{si} = f(y_{sj}) , \quad i=1,2,\dots,n \text{ ve } j \neq i \quad (5)$$

Uzaydaki herhangi bir R çalışma alanındaki rastgele bir noktada gözlenen verilerinin diğer konumlarda gözlenen değerlere bağımlı olmasının genellikle iki nedeni vardır. Birinci neden, mekânsal birimlere ait verilerin ölçme hatası içermesidir. İkinci neden ise, mekânsal otokorelasyonun (ardışık bağımlılık) ortaya çıkmasıdır (Zeren, 2011: 7). Bu nedenlere bağlı olarak mekânsal bağımlılık sınıflandırılmıştır.

Matematiksel olarak mekânsal otokorelasyon aşağıdaki biçimde gösterilir;

$$Cov(Y_i, Y_j) = E(Y_i Y_j) - E(Y_i).E(Y_j) \neq 0 \quad (6)$$

Burada $i \neq j$ 'dir ve aynı zamanda Y_i ve Y_j , $s_i \in D, s_j \in D$ noktalarındaki verileri göstermektedir.

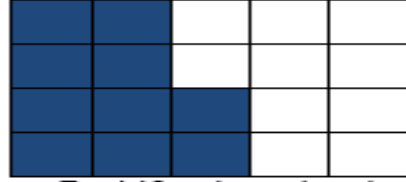
Mekânsal otokorelasyon, bir değişkene ait değerlerin komşu gözlemler ile ilişkisini kovaryans veya korelasyon gibi istatistiksel bir ölçü ile tanımlanmasıdır.

Ayrıca mekânsal otokorelasyon, gözlemlerin benzerliği ve konumlar arası benzerlikleri ortaya koymaktadır. Bu otokorelasyonun göz ardı edilmesi durumunda verilerin gerçek varyansı hatalı tahmin edilir. Birbirine yakın gözlemlerin değerleri benzerlik gösterir ise bir bütün olarak pozitif mekânsal otokorelasyon örüntüsü gösterirler. Tuzcu (2016), pozitif mekânsal otokorelasyon durumunda örneklem ortalamasının daha az hassas olacağını ve genellikle α tipi (birinci tip) hata meydana geleceğini belirtmiştir.

Aşağıda Şekil 3'teki bir bölge ile mekânsal otokorelasyon görsel olarak açıklanmıştır (Yamagata ve Seya, 2019: 3). Bu bölge birbirine komşu olan 20 değişkenlere ait gözlem değerleri göstermektedir. Bu değişkenlere ait gözlem değerlerinin birbirine benzer eğilim gösterip göstermediği araştırılmıştır. Mekânsal otokorelasyon, -1 ile 1 aralığında değer almaktadır. Komşu gözlemlerin bir yerde kümelenip benzer eğilimler göstermesi 0 ile 1 arasında değer alır, buna pozitif mekânsal otokorelasyon denir. Pehlivanoglu ve Yeler (2019) genellikle analizlerinde pozitif mekânsal otokorelasyonun dikkate alındığını ve bunun sebebinin ise yakın bölgelerin

benzer özellik göstermesi beklentisi olduğunu ifade etmişlerdir. Buna karşın farklı eğilim göstermesi negatif mekânsal otokorelasyon olarak tanımlanır.

Şekil 3: Mekânsal Otokorelasyon Örnekleri



a - Pozitif mekansal otokorelasyon



b - Rastgele mekansal dağılım



c - Negatif mekansal otokorelasyon

Günümüzde, mekânsal otokorelasyonu ölçmek için çok sayıda yöntem mevcuttur. Bu yöntemlerin bir kısmı global bir kısmı ise lokal düzeyde ölçüm yapmaktadır.

1.2.2.1. Global Mekânsal Otokorelasyon Testi

Global mekânsal otokorelasyon, ele alınan bölge için genel bir sonuç sağlamaktadır. Bu test için genellikle Global Moran I ve Global Geary C istatistikleri kullanılır. Global mekânsal otokorelasyon testi Global Indicators of Spatial Association (GISA) olarak adlandırılır. GISA'nın en güçlü göstergesi Moran I testidir. Moran I bölgenin genelini inceler ve bu bölgedeki bütün gözlemler birbirlerine bağlıdır. Bu global otokorelasyon testinde alternatif hipotez altında mekânsal korelasyonun yapısı belirgin değildir. Bu nedenle yalnızca mekânsal bağımlılığın varlığını araştırır, ancak bağımlılığın sınıfı hakkında bilgi vermez.

Global Moran I istatistiği, bir rassal değişken için ortalamadan sapmalar cinsinden aşağıdaki gibi yazılabilir (Yamagata ve Seiya, 2019: 40);

$$I = \frac{N \sum_{i=1}^N \sum_{j=1}^N w_{ij} (Y_i - \bar{Y})(Y_j - \bar{Y})}{S_0 \sum_{i=1}^N (Y_i - \bar{Y})^2} \quad (7)$$

Yukarıdaki denklemde Y_j ve Y_i gözlem değerleri ve \bar{Y} , gözlenen değerlerin ortalamasıdır. N gözlem sayısı ve $S_0 = \sum_{i=1}^N \sum_{j=1}^N w_{ij}$ 'dir. w_{ij} ağırlık matrisi elemanıdır.

Regresyon modellerinin kalıntılarındaki mekânsal otokorelasyon nedeniyle aşağıdaki regresyon modelinin EKK tahmincileri tutarsız ve sapmalı olacaktır.

$$y = x\hat{\beta} + e \quad (8)$$

Bir regresyon modelindeki mekânsal otokorelasyon EKK kalıntıları ile elde edilmektedir (Cliff ve Ord, 1972; 267). Hataların normal olarak dağıldığı varsayımına bağlı olarak yukarıdaki regresyon modeli için Global Moran I istatistiği aşağıdaki gibidir:

$$I = (N/S_0)(e'We/e'e) \quad (9)$$

Burada N , gözlem sayısıdır; $S_0 = \sum_i \sum_j w_{ij}$ sıfırdan farklı ağırlıklar toplamının standart faktörü ve $e = (y - x\hat{\beta})$, $(N \times 1)$ boyutundaki EKK modelinin kalıntılarıdır. $\hat{\beta} = (X'X)^{-1}X'Y$ 'dir ve W , mekânsal ağırlık matrisidir.

Ağırlık matrisine satır standartlaştırması uygulandığında Global Moran'ın hesaplanması daha da basit bir hale gelmektedir. Bunun nedeni yukarıdaki denklemde N/S_0 teriminin ortadan kalkmasıdır. Global Moran'ın alt ve üst değer limitleri $|I| \leq \sqrt{\text{Var}(Wy)/\text{Var}(y)}$ ile belirlenmektedir. Global Moran'ın pozitif değeri pozitif otokorelasyonun olduğunu ve negatif değeri ise negatif otokorelasyonun varlığını ifade etmektedir. (Yamagata ve Seya, 2019: 40).

Mekânsal otokorelasyonun ölçümünde kullanılan diğer bir istatistik Global Geary C istatistiğidir ve aşağıdaki gibi tanımlanmaktadır (Yamagata ve Seya, 2019: 40):

$$C = \frac{N-1}{2S_0} \frac{\sum_{i=1}^N \sum_{j=1}^N w_{ij} (Y_i - Y_j)^2}{\sum_{i=1}^N (Y_i - \bar{Y})^2} \quad (10)$$

Yukarıdaki denklemde Y_j ve Y_i gözlem değerleri ve \bar{Y} , gözlenen değerlerin ortalamasıdır. N , gözlem sayısı ve $S_0 = \sum_i \sum_j W_{ij}$ sıfırdan farklı ağırlıklar toplamının standart faktörüdür. Global Geary'nin değer aralığı 0 ile 2 arasındadır. Sıfıra yakın değerler çıktığında pozitif mekânsal otokorelasyonun varlığından, 2'ye yakın değerlerde ise negatif mekânsal otokorelasyonun varlığından söz edilebilir.

Global Moran I ve Global Geary C normallik ve rassallık varsayımı altında, beklenen değerleri ve varyansları kullanılarak standartlaştırılabilmektedir (Yamagata ve Seya, 2019: 41-43).

$$Z = \frac{I-E(I)}{\sqrt{Var(I)}} \quad Z = \frac{C-E(C)}{\sqrt{Var(C)}} \quad (11)$$

Normallik varsayımı altında Z 'nin asimptotik normal dağıldığı, mekânsal otokorelasyon olmadığı şeklinde hipotez kurulacak olursa Global Moran'ın beklen değer ve varyansı aşağıdaki gibidir (Yamagata ve Seya, 2019: 41);

$$E(I) = \frac{-1}{N-1}$$

$$Var(I) = \frac{1}{(N-1)(N+1)S_0^2} (N^2S_1 - NS_2 + 3S_0^2 - [E(I)]^2) \quad (12)$$

Burada yer alan $S_1 = \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N (w_{ij} + w_{ji})^2$ ve $S_2 = \sum_{i=1}^N (w_i + w_i^*)^2$ 'dir. S_2 'deki terimler ise; $w_i = \sum_{j=1}^N w_{ij}$ ve $w_i^* = \sum_{j=1}^N w_{ji}$ 'dir.

Global Geary'nin beklenen değeri ve varyansı aşağıdaki gibidir;

$$E(C) = 1$$

$$Var(C) = \frac{(2S_1 + S_2)(N-1) - 4S_0^2}{2(N+1)S_0^2} \quad (13)$$

Denklem 12 ve 13'e göre N 'nin artışına bağlı olarak N sıfıra ne kadar yaklaşmış olsa da Global Moran'ın beklen değeri tam olarak sıfır olmamaktadır (Yamagata ve Seya, 2019: 42).

Cliff ve Ord (1972) bu istatistiği iki farklı hipotez altında türetmiştir. Bunlar rassallık ve artıkların normalliğidir. Rassallık hipotezi altında örnekleme dağılımı, sınırlar

içerisinde gözlemlenen verilerin tüm olası permütasyonlarını (genellikle 99 ve 999 kez kullanılır) da hesaba katarak ve her birinde Moran I istatistiği hesaplayarak elde edilir. Ayrıca asimptotik dağılımın, herhangi bir hipoteze bağlı olmayan beklenen bir değerle normal dağıldığını kanıtlamışlardır:

$$E(I) = \frac{N \operatorname{tr}(M_x W)}{S_0(N-k)} \quad (14)$$

Burada $S_0 = \sum_i \sum_j w_{ij}$, $M_x = I - P_x$ ve $P_x = X(X^T X)^{-1} X^T$ 'dir. Bu işlemin varyansı, seçilen hipoteze bağlıdır. Eğer I , $E(I)$ 'dan yeterince büyükse burada pozitif otokorelasyonun varlığından söz edilebilir. Özellikle artıkların normalliği kabul edilirse varyansı şu şekilde yazılabilir:

$$\operatorname{Var}(I) = \left(\frac{N}{S_0}\right)^2 \frac{\operatorname{tr}(M_x W M_x W^T) + \operatorname{tr}(M_x W)^2 + [\operatorname{tr}(M_x W)]^2}{(N-k)(N-k+2)} - E(I)^2 \quad (15)$$

Cliff ve Ord (1972), rassallık ve artıkların normalliği varsayımı altında Moran I istatistiğinin ortalama ve varyansını hesaplayarak limit dağılımının normal dağılıma uyduğunu ispatlamışlardır. Moran I istatistiğinin sıfır hipotezi, EKK artıkları arasında mekânsal bağımlılığın olmadığını ifade eder. Moran I istatistiği normal dağılım tablosunda belirlenen kritik değerden büyük olduğunda sıfır hipotezi reddedilir ve alternatif hipotez kabul edilir. Ancak alternatif hipotez mekânsal bağımlılığın varlığını ifade etmesinin yanı sıra bu bağımlılığın kaynağı hakkında bilgi vermemektedir (Yamagata ve Seya, 2019: 42). Bu nedenle Moran I istatistiğine alternatif mekânsal bağımlılık testleri mevcuttur ve Lagrange Çarpanı (LM) testi buna örnektir.

1.2.2.2. Lokal Mekânsal Otokorelasyon Testi

Bölgeler arasındaki mekânsal etkileşimler keşifsel mekânsal veri analizi (Exploratory Spatial Data Analysis; ESDA) tekniğiyle analiz edilebilmektedir. ESDA, istatistiksel verilerin mekânsal özelliklerini, mekânsal kalıplarını, mekânsal dağılımlarını görselleştirmek ve konumlarına dayanan hipotezleri formüle etmek için kullanılır. ESDA'nın araçları; çekirdek yoğunluğu, K fonksiyonları, variogram ve LISA

istatistikleridir (Yamagata ve Seya, 2019: 41). Çekirdek yoğunluğu ve K fonksiyonları mekânsal nokta örüntüleri için kullanılır.

Lokal Indicators of Spatial Association (LISA) testi her bölge için ayrı ayrı mekânsal otokorelasyon değeri hesaplamakta ve poligon verileri için kullanılmaktadır. Dört kategoride sonuçlar vermektedir. Bunlar; yüksek-yüksek(sıcak nokta), yüksek-düşük, düşük-düşük(soğuk nokta), düşük-yüksek şeklindedir. LISA bir değışkene ait gözlem etrafındaki benzer değerlerin mekânsal kümelenmelerini belirlemektedir. Yüksek gözlem değerlerinin lokal kümelenmesi (sıcak nokta) veya düşük gözlem değerlerinin lokal kümelenmesinin (soğuk nokta) varlığını göstermektedir. Ayrıca kümelenme olmayan bölgeleri belirlemektedir.

Lokal mekânsal otokorelasyon için lokal Moran I ve lokal Geary C ölçümleri yapılır. Lokal Moran I istatistiğı aşağıdaki gibi yazılabilir (Yamagata ve Seya, 2019: 44):

$$I_i = \frac{Y_i - \bar{Y}}{m_2} \sum w_{ij} (Y_j - \bar{Y}) \quad (16)$$

Denklem 16'da yer alan $m_2 = N^{-1} \sum (Y_i - \bar{Y})^2$ şeklinde tanımlanan bir sabittir.

Lokal Moran'ın beklen değeri ve varyansı, rassallık varsayımı altında Denklem 17'deki gibi ifade edilebilir;

$$E(I_i) = \frac{-w_i}{N-1}$$

$$\text{Var}(I_i) = \frac{w_{i(2)}}{N-1} + \frac{2w_{i(kh)}(2b_2 - N)}{(N-1)(N-2)} - \frac{w_i^2}{(N-1)^2} \quad (17)$$

Buradaki bilinmeyen terimler şu şekildedir;

$$w_i = \sum_j w_{ij}, \quad w_{i(2)} = \sum_{j \neq i} w_{ij}, \quad b_2 = \frac{m_4}{m_2^2}, \quad m_4 = \frac{\sum (Y_i - \bar{Y})^4}{N},$$

$$2w_{i(kh)} = \sum_{k \neq i} \sum_{h \neq i} w_{ik} w_{ih}$$

Lokal Geary istatistiğı, Denklem 18'deki gibi ifade edilebilir:

$$C_i = \frac{1}{m_2} \sum_j w_{ij} (Y_i - Y_j) \quad (18)$$

Denklem 18'de yer alan m_2 terimi bir sabittir ve $m_2 = N^{-1} \sum (Y_i - \bar{Y})^2$ şeklinde tanımlanır.

Lokal Geary'nin rassallık varsayımı altında beklen değeri aşağıdaki gibidir:

$$E(C_i) = \frac{2Nw_i}{N-1}$$

Anselin (1995), Lokal Geary istatistiğinin varyansını ($\text{Var}(C_i)$) permutasyon testi kullanılarak geliřtirmenin mümkün olduğunu ifade etmiştir.



2. BÜYÜK VERİ

Günümüzde büyük veriye olan ilgi her geçen gün hızla artmaktadır. Büyük veri ilk olarak astronomi ve genetik alanında kullanılmaya başlanmış günümüz teknolojileri ile de her geçen gün farklı alanlarda da kullanılmaktadır. Astronomi ve genetiğin yanı sıra özellikle bankacılık sektörü, firmaların müşteri odaklı çalışmaları, sağlık, eğitim, güvenlik ve sosyal bilimler alanında da kullanıldığı görülmektedir. Büyük veriler artık hayatımızın her alanında karşımıza çıkmaktadır (Çağlayan Akay, 2020: 3-4). Kısaca kullanım alanları aşağıdaki gibidir:

- Sağlıkta; sağlık alışkanlıklarıyla ilişkili olarak yaşam kalite eğilimleri, MHRS ve e-nabız gibi uygulamalarda uzaktan hasta takip etme, hasta profillemeye, hastalık tahmini, vb.
- Pazarlamada; çeşitli satış yöntemi, mekânsal tabanlı pazarlama, müşterilerin duygu analizi, müşteri odaklı eğilim analizi, davranış analizi, vb.
- Üretimde; sensor tabanlı operasyonlar, firma odaklı tedarik zinciri, firmalarda envanter yönetimi, lojistik, vb.
- Kamuda; temel ihtiyaç tespiti, trafik sorunları çözümü, gürültü, hava ve su kirliliği engelleme, suç oranlarının çeşitliliği, vb.
- Bankacılık ve Sigortacılıkta; dolandırıcılığın, istismarın veya suiistimallerin tespiti, müşteri tahminleme, risk analizleri, vb.
- Telekom sektöründe ise; coğrafi özellikten kaynaklanan kitlelere reklam, acil müdahale, kentsel planlama, ücretlendirme, dolandırıcılık önleme, saldırı tespiti, vb. konularda faydalanılmaktadır.
- Tarım; Akıllı tarım ile toprağın mineral, nem ve benzeri ihtiyaçları belirlenerek toprağa bu teknolojiler ile bakım yapıp o yılki hasat verimliliğini arttırmak.

Günümüzde geleneksel veri işleme araçları ile analizi gerçekleştirilemeyen ve karışık bir yapıda bulunan büyük miktarlardaki veriler için kullanılan terime Büyük veri denir. (Çağlayan Akay, 2020: 3). Bu veriler çok büyük hacimlidir ve giderek büyüyen veri bütünlerinin depolanması, analiz edilmesi oldukça zordur. Büyük veri aynı zamanda bir veridir. Literatürde büyük veri ile ilgili farklı tanımlarda mevcuttur. Büyük veriyi yaygın veri türlerinde ayıran nokta verilerin depolanamaz olması, verimli bir şekilde işlenemez olması ve çok büyük boyutlarda olmasıdır.

Teknolojinin gelişmesiyle sıkça kullanılan büyük veri geleneksel veri tabanı sistemlerinin büyük hacimleri nedeni ile işlem kapasitesini aşan verilerdir. Çok büyük olan bu veriler aynı zamanda çok hızlı birikmektedir ve mevcut veri tabanlarına sığmamaktadır.

Günümüzde büyük verinin geleneksel veri türlerine göre kendine has özellikleri ve avantajları vardır. Bu avantajları aşağıdaki gibi ifade etmek mümkündür:

- Farklı veri türlerinin birleşimi olmaları nedeni ile ek bilgiler elde edebilirler
- Büyük veriler güncel olmaları nedeni ile gerçek zamanlı verileri sağlar.
- Daha önce gözlenemeyen insan davranışları hakkında bilgilerini elde edebilirler.
- Büyük örnek büyüklüğü yapılan analizlerinin daha doğru sonuçlar vermesini sağlar.

Büyük verinin bu avantajı olduğu gibi zorlukları da vardır. Bu zorluklar aşağıdaki gibidir:

- Büyük verinin tanımından da anlaşılacağı gibi saklanması ve depolanması zordur.
- Farklı veri türleri birleşimi nedeni ile bu verileri analiz ederken bazı parametrelerin tahmininde hatalara neden olabilmektedir.
- Büyük veri kullanıldığında uygun yöntemlere başvurulmaz ise hatalı tahminlere ve sahte korelasyona neden olabilmektedir.
- Veri boyutu büyük olduğu için analizleri de maliyetli olabilmektedirler.

2.1. Büyük Veri Türleri

Büyük verinin üç farklı tipi mevcuttur. Bunlar yapılandırılmış veri, yapılandırılmamış veri ve yarı yapılandırılmış büyük verilerdir.

Yapılandırılmış veri; sabit formatlarda yazı, tablo veya diğer şekillerde saklanan, erişilebilen veya işlenebilen her türlü veri yapılandırılmış veri olarak adlandırılır. Bu verilerin herhangi bir yapısı yoktur ve içerdiği bilginin nereden nasıl bulunduğu bellidir. Bunlar sayısal verilerdir. Aşağıda verilen Tablo 1'deki yapısal veri örneğinde, her bir satır birime ait verileri içeren kayıtlardan yer almaktadır, bu kayıtlar ise kendilerini oluşturan alanların gösterildiği özniteliklerden oluşur (Sağiroğlu vd. , 2017: 32).

Tablo 1: Yapılandırılmış Veri Örneği

	Adres	Şehir	P.K	Doğum	Cinsiyet	Ders	Okul
Nuriye	Melikgazi	Kayseri	38220	18.05.2005	Kadın	Fizik	Lise
Kenan	Erkilet	Kayseri	38100	20.11.2007	Erkek	Felsefe	Lise
Ela	Kocasınan	Kayseri	38100	18.07.2008	Kadın	Türkçe	Ortaokul

...
-----	-----	-----	-----	-----	-----	-----	-----

Yapılandırılmamış veri, biçimi bilinmeyen ve sabit olmayan formda yazılmış veriler olarak ifade edilir. Bu tür veriler genellikle sosyal medyadan, resimlerden videolardan veya farklı alanlardan elde edilmektedir. Yapılandırılmamış veriler büyük ve karmaşık olduğundan kolay analiz edilemezler ve yüksek maliyet ve zaman gerektiği için yapılandırılmış veriye dönüştürülemezler.

Yarı yapılandırılmış veri, basit bir şekilde ortaya çıkarılmamış, saklanması analiz edilmesi zor olan verilerdir. Yarı yapılandırılmış veriler, verilerin organizasyon özelliklerini içeren fakat geleneksel veri tabanı biçiminde olmayan bilgileri içerir.

2.2. Büyük Veri Bileşenleri

Büyük verinin oluşumunda çeşitli özellikleri vardır. Bunlar büyük veriyi anlamamızda ve zorluklara karşı önlemler almamızda bize yardımcı olmaktadır (Çağlayan Akay, 2020: 6-8). Büyük veri beş temel bileşene dayanmaktadır. Bunlar hacim (Volume), hız (Velocity), çeşitlilik (Variety), değer (Value), doğrulama (Verification)'dir. Verinin ve bilginin büyük bir önem kazandığı bu dönemde verilerin kullanımını artırmak, veriye ulaşımı kolaylaştırarak bu konuda çalışmalar yapmak da bir o kadar önemlidir. Verinin çeşitliliği kadar işlenebilirliği, kalitesi ve verinin anlamlı hale getirilmesi bunu yaparken de veri güvenliğinin sağlanması, hataların ve eksiklerin tespit edilerek önlem alınması da bir o kadar önemlidir. Literatürde büyük verinin farklı özellikleri mevcuttur. Bunlara değişkenlik, geçerlilik, görselleştirme vb. örnek verilebilir.

2.2.1. Hacim (Volume)

Büyük verinin en ayırt edici ve büyük veriyi tanımlayan özelliğinden biri hacimdir. Bu nedenle büyük verinin birinci özelliği olarak bilinir. Hacim kim tarafında nasıl ve ne şekilde üretildiğine bakılmaksızın büyük verinin miktarını ifade etmektedir. Bu verilerin hacmini ölçmek için genellikle petabayt veya exabayt gibi büyüklükler kullanılmaktadır. Videolar, resimler ve yapılan ses kayıtları gibi geniş hacimli veriler büyük verinin hacmine örnektir. Verilerin bayt cinsinden ifade edilmesi için sayıların kullanılması gerekmektedir. Bu nedenle bilgisayar veri birimleri ortaya çıkmıştır. Bu birimler 1024'er artıp veya azalmaktadır. Aşağıdaki tabloda bilgisayar veri birimleri gösterilmektedir:

Tablo 2: Veri Birimleri Ölçüsü

1 bayt	8 bit = 1 bayt
--------	----------------

1KB (kilobayt)	1024 bayt =1KB
1MB (megabayt)	1024 kilobayt =1 MB (megabayt)
1GB (gigabayt)	1024 megabayt = 1 GB (gigabayt)
1TB (terabayt)	1024 gigabayt = 1TB (terabayt)
1PB (petabayt)	1024 terabayt = 1 PB (petabayt)
1EB (exabayt)	1024 petabayt = 1EB (exabayt)
ZB (zetabayt)	1024 exabayt = 1ZB (zetabayt)

Bu birimlere örnek verilecek olunur ise; ortalama çözünürlüklü bir resim yaklaşık 750 KB büyüklüğünde, bir mp3 müzik dosyası yaklaşık 6 MB gibi bir büyüklükte, kısa bir video yaklaşık olarak 100 MB büyüklüğünde, film arşivi klasörünüz toplam 120 GB gibi bir büyüklükte ve taşınabilir disk 2 TB saklama kapasitesinde olabilir.

2.2.2. Hız (Velocity)

İşlenecek olan verinin sıklığını ifade eder ve dolayısıyla verilerin yüksek hızda üretilip veya analiz edilmesi anlamına gelmektedir. Büyük veriler yüksek hızda ve hacimde üretildiği zaman bazı sistemlerde ve veri tabanlı yazılımlarda saklanması veya depolanması zorlaşabilir ya da maliyetli olabilmektedir.

2.2.3. Çeşitlilik (Variety)

Günümüzde verilerin hacmi artarken aynı zamanda bu verilerin çeşitliliği de artmaktadır. Büyük veriler yapılandırılmış veya yapılandırılmamış şekilde çokça kaynaktan veri elde edebilmektedir. Büyük verinin çeşitlilik özelliği verilerdeki heterojenliği belirtmektedir. Yapılandırılmış, yapılandırılmamış ve yarı yapılandırılmış veriler bunlara örnek verilmektedir.

2.2.4. Değer (Value)

Büyük verilerden elde edilen değerleri ve önemini ifade etmektedir. Verilerin büyüklüğü bu verilerin değerli olduğu anlamına gelmemektedir. Değeri olmayan verileri de kullanmak anlamsız olacaktır. Bu nedenle analiz sonuçlarında verilerin değeri olduğundan emin olmak önemlidir. Verinin istatistiksel olarak hesaplanabilirliği, korelasyonlu olup olmadığı ve desen tahmininin yapılabilirliği olması bu verinin bir değeri olduğunu göstermektedir (Aktan. 2018: 6).

2.2.5. Doğrulama (Verification)

Büyük verinin bu özelliği verinin güvenilirliğini ve diğer veri türleri ile ilişkilerinin güvenilirliğini belirtir. Bir veri seti ne kadar doğru ise o veri seti hatasız ve güvenilirlerdir. Doğru olmayan veriler, arařtırmacılara yüksek oranda sorun çıkarmaktadır.



3. MEKÂNSAL BÜYÜK VERİ ANALİZİ

Mekânsal veri analizine olan ilgi her geçen gün hızla artmaktadır. Mekânsal veri analizi bir dağılımın mekânsal örüntüsünü, mekanın içerisindeki her bir varlığın düzenini ve bunlar arasındaki ilişkileri göstermek için kullanılmaktadır (Aydın vd., 0218: 29) Günümüzde değişen teknoloji ile her geçen gün farklı alanlara yayılmaktadır. Epidemiyoloji, ekoloji, madencilik gibi alanların yanı sıra istatistikçiler, salgın bilimciler, yerbilimciler vb. alanlarında da kullanılmaktadır.

Mekânsal analiz, coğrafi bölgeler üzerindeki mekânsal örüntüleri göstermek için kullanılan bir yöntemdir. Mekânsal analiz yapılırken mekânsal veriyi oluşturan nesnelere öznel ve konum bilgileri kullanılır. Mekânsal verinin içindeki birimlerin birbirine bağımlılığı, gözlemlerin birbirinden bağımsız olduğu varsayımının ihlal edilmesine neden olur ve dolayısıyla gözlemler mekânsal otokorelasyon sergiler. Bu sebeple, modelde açıklayıcı değişkenler kullanılsa bile mekânsal ilişki yakalanamamakta ve artıklarda bu otokorelasyon kalmaktadır. Artıklarda kalan bu ilişkiden dolayı doğrusal model yaklaşımlarında bağımsızlık varsayımı bozulmaktadır. Bu nedenle komşuluk ilişkilerinin (ülke, bölge, şehir, vb.) oluşturulduğu mekânsal ağırlık matrisleri ve mekânsal etkileşimlerin dikkate alındığı mekânsal regresyon modelleri kullanılır. Mekânsal verinin içindeki birimlerin birbirine olan bağımlılığın yapısına bağlı olarak tahmin yöntemi değişim göstermektedir.

3.1. Doğrusal Regresyon Modelleri

Klasik doğrusal regresyon modeli şu şekildedir:

$$Y = X\beta + \varepsilon \quad (19)$$

Burada Y ; $N \times 1$ boyutunda bağımlı değişken vektörü, β ; $K \times 1$ boyutunda sabit terim ve eğim parametresi vektörü, X ; açıklayıcı değişken matrisi, ε ; $N \times 1$ boyutlu hata terimi vektörüdür. Hata teriminin sıfır ortalama ve σ^2 varyans ile normal dağıldığı varsayılmaktadır.

$$\varepsilon | X \sim i. i. d. N(0, \sigma_\varepsilon^2 I_n)$$

Eşitlik 19’da tanımlanan doğrusal regresyonu matris formunda aşağıdaki gibi göstermek mümkündür;

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_N \end{pmatrix} = \begin{pmatrix} 1 & X_{21} & \cdots & X_{K1} \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{2N} & \cdots & X_{KN} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{pmatrix}$$

Klasik doğrusal regresyon modeli için genellikle dört varsayım yapılmaktadır (Yamagata ve Seya, 2019: 11). Bu varsayımlar aşağıdaki gibidir;

1. X açıklayıcı değişkenleri dışsaldır.
2. X açıklayıcı değişkenleri verili olduğunda, Y’nin koşullu beklenen değeri $X\beta$ ’dir ve hata terimi ε ’nin beklenen değeri 0’dır.

$$E(\varepsilon|X) = 0$$

3. X açıklayıcı değişkenlerine koşullu olarak verildiğinde, hata terimi varyansı şu koşulları sağlamalıdır:

$$Var[\varepsilon|X] = \sigma_\varepsilon^2 I_N = \begin{pmatrix} \sigma_\varepsilon^2 & 0 & \cdots & 0 \\ 0 & \sigma_\varepsilon^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & \sigma_\varepsilon^2 \end{pmatrix}$$

4. X’ in rankı K’ dir. O halde $X'X$ matrisinin tersi mevcuttur. Bu nedenle β ’nın EKK tahmincisi ve varyansı aşağıdaki gibi hesaplanır:

$$\hat{\beta}_{EKK} = (X'X)^{-1}X'y \quad (20)$$

$$Var(\hat{\beta}_{EKK}) = \sigma_\varepsilon^2(X'X)^{-1}$$

σ_ε^2 genelde bilinmediğinden aşağıdaki tahmincisi ile değiştirilir:

$$\hat{\sigma}_\varepsilon^2 = \hat{\varepsilon}'_{EKK}\hat{\varepsilon}_{EKK}/(N - K) \quad (21)$$

Bu dört varsayımın geçerli olduğu durumda EKK tahmin edicisi DESTE (En İyi Sapmasız Tahmin Edici) özelliği taşır ve bu özellik Gauss-Markov teoremi olarak bilinmektedir.

Klasik doğrusal regresyon modelinde hataların normal dağıldığı varsayılır. Bu varsayımdan hareketle EKK kriterine alternatif olarak maksimum olabilirlik fonksiyonu da kullanılmaktadır (Arbia, 2014: 13-14). Hataların olabilirlik fonksiyonu aşağıdaki gibi ifade edilir (Gujarati ve Porter, 2018: 817):

$$f_{\varepsilon_i}(\varepsilon_i) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\varepsilon_i^2\right] \quad (22)$$

Denklem 22'teki f bir yoğunluk fonksiyonudur. Bu fonksiyon eşitlik 19'daki modelin tahmini için kullanılmaktadır ve gözlenen örneğin olasılığı şu şekildedir:

$$\begin{aligned} L(\beta, \sigma_\varepsilon^2) &= \prod_{i=1}^n f_i(\varepsilon_i) = \prod_{i=1}^n \frac{1}{\sigma_\varepsilon\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma_\varepsilon^2}\varepsilon_i^2\right] \\ &= (\sigma_\varepsilon^2)^{-\frac{n}{2}} (2\pi)^{-\frac{n}{2}} \exp\left[-\frac{\varepsilon^T \varepsilon}{2\sigma_\varepsilon^2}\right] \\ &= \text{cost}(\sigma_\varepsilon^2)^{-\frac{n}{2}} \exp\left[-\frac{\varepsilon^T \varepsilon}{2\sigma_\varepsilon^2}\right] \end{aligned} \quad (23)$$

Denklem 19'daki modelden hatalar $\varepsilon = y - X\beta$ şeklinde elde edilebilir. Dolayısıyla Denklem 23'teki olasılık fonksiyonu şu şekilde tekrar yazılabilir:

$$L(\beta, \sigma_\varepsilon^2) = \text{cost}(\sigma_\varepsilon^2)^{-\frac{n}{2}} \exp\left[-\frac{(y-X\beta)^T(y-X\beta)}{2\sigma_\varepsilon^2}\right]$$

Olasılık fonksiyonunun logaritmik olasılığı şu şekildedir:

$$l(\beta, \sigma_\varepsilon^2) = \ln[L(\beta, \sigma_\varepsilon^2)] = \text{cost} - \frac{n}{2} \ln(\sigma_\varepsilon^2) - \frac{(y-X\beta)^T(y-X\beta)}{2\sigma_\varepsilon^2}$$

3.1.1. İçsellik

Doğrusal regresyon modellerindeki 1. ve 2. varsayımlar (X ' ler ile ε arasında korelasyon olması halinde) gerçekleşmediğinde EKK tahmincisi tutarsız ve yanlıdır (Yamagata ve Seya, 2019: 12). Regresyon katsayılarının tutarlı tahminler verebilmesi için bir araç değişken kullanılır. Bu yöntem araç değişken yöntemi (Instrumental Variable, IV) denir. Araç değişkenler koşullu olarak hata terimi ile korelasyonsuz, içsel açıklayıcı

değişkenler ile korelasyonlu olmalıdır. Araç değişken yöntemleri olarak genellikle iki Aşamalı EKK (Two-Stage Least Squares, 2SLS) ve Genelleştirilmiş Momentler Metodu (GMM) kullanılmaktadır.

İçsel değişkenli doğrusal regresyon modeli kurulur ise bu model aşağıdaki gibi ifade edilir (Yamagata ve Seya, 2019: 13);

$$y = X\beta + \dot{X}\dot{\beta} + \varepsilon \quad (24)$$

Burada X dışsal değişken olup $N \times K$ tipinde açıklayıcı değişken matrisi, \dot{X} 'i içsel değişkenlerden oluşan $N \times L$ tipinde açıklayıcı değişken matrisi, β dışsal değişkenlere karşılık gelen $K \times 1$ tipinde katsayı vektörü ve son olarak da $\dot{\beta}$ içsel değişkenlere karşılık gelen $L \times 1$ tipinde bir vektördür. Araç değişkenler bağımlı değişken üzerinde doğrudan etkiyi veya ihmal edilen değişkenlerin dolaylı etkisini dışlamaktadır (Yamagata ve Seya, 2019: 13).

Denklem 24'teki modelde \dot{X} içsel bir değişken olması nedeniyle hata terimi ile korelasyonludur ve $Cov(\varepsilon_i, \dot{X}_{1,i}) \neq 0 \quad i = 1, \dots, L$ biçiminde ifade edilmektedir (Yamagata ve Seya, 2019: 13). Bu nedenle \dot{X} ile ilişkili ve hata terimi ile ilişkisiz bir araç değişken (Z) dâhil edilmelidir. Z araç değişkeninin sayısı olan P , içsel değişken sayısı L 'den büyük, yani modele eklenen araç değişken sayısı modeldeki içsel değişkenden fazla ($P > L$) ise denklem aşırı belirlenmiştir. Eklenen araç değişken sayısı modeldeki içsel değişken sayısına eşit ise ($P = L$) denklem tam belirlenmiştir.

Açıklayıcı değişken matrisi $R \equiv [X; \dot{X}]$ tekrardan düzenlecek olursa iki aşamalı bir tahmin gerçekleştirilerek $R, S \equiv [X; Z]$ şeklini alır ve hata terimi ile ilişkisizdirler (Yamagata ve Seya, 2019: 13). Uygun bir araç değişken kullanılırsa iki aşamalı EKK tutarlı sonuçlar vermektedir. β 'nın 2SLS tahmincisi ve varyansı şu şekildedir:

$$\begin{aligned} \hat{\beta}_{2SLS} &= (\hat{R}'\hat{R})^{-1}\hat{R}'y \\ Var[\hat{\beta}_{2SLS}] &= \sigma_\varepsilon^2(\hat{R}'\hat{R})^{-1} \end{aligned} \quad (25)$$

Burada $\hat{R} = (S'S)^{-1}S'R$ 'dir ve $\hat{\beta}_{2SLS} = [R'S(S'S)^{-1}S'R]^{-1}R'(S'S)^{-1}S'R$ şeklinde düzenlenebilir.

3.2. Mekânsal Regresyon Modelleri

Mekânsal modeller mekânsal etkileri/mekânsal bağımlılığı modellemektedir. Bu bağımlılık ya da etki modele dâhil edilmediğinde gerekli bir değişkeni dışlama hatası yapılır. Bu dışlama hatası, sapmalı ve tutarsız tahminlere neden olacağından tahmin sonuçlarında ve hipotez testlerinde yanıltıcı sonuçlara neden olacaktır. Mekânsal regresyon modellerinden en yaygın olanları mekânsal hata modelleri (Spatial Error Model, SEM) ve mekânsal gecikme modelleri (Spatial Lag Model, SLM)'dir. Yakın ve birbiri ile ilişkili birimler arasındaki mekânsal ilişkinin yapısı ve bu nedenle yapılacak olan tahmin yöntemi de değişkenlik göstermektedir. Örneğin, şehirlerdeki suç oranlarında şehirlerarası etkileşim olması durumunda mekânsal gecikme modeli kullanılır. Şehirlerdeki suç oranlarını açıklayan modelde mekânsal korelasyonun veya etkileşimin nedeni bazen ölçme hataları olabilir. Böyle bir durumda mekânsal hata modeli kullanılır. Her iki etkileşimin olması durumunda ise genel mekânsal model veya mekânsal Kombo modeli (Spatial Combo Model, SAC) ile tahmin yapılır (Anselin ve Rey, 2014; 257).

Mekânsal modeller, klasik regresyon modellerinden farklı olarak, açıklayıcı değişkenlerin bağımlı değişken üzerindeki marjinal etkileri doğrudan veya dolaylıdır. Belirli bir birimdeki belirli bir açıklayıcı değişken değişir ise sadece bu birimdeki bağımlı değişkenin kendisi değil, diğer birimlerdeki bağımlı değişkenler de değişmektedir. Birimlerdeki bağımlı değişken değişmesi direkt etki olarak adlandırılırken, diğer birimlerdeki bağımlı değişkenlerin değişmesi dolaylı etki olarak adlandırılır. Başka bir deyişle, bağımlı değişkenlerin mekândaki değişimleri sadece mekânın kendisinde olmayıp diğer mekânlardan da etkilenmektedir (LeSage ve Pace, 2009: 21).

Mekânsal olmayan modeller genellikle EKK yöntemi ile tahmin edilirler. Mekânsal ekonometrik analizlerde tüm mekânsal etkileşimler tahmin edilmelidir. Mekânsal etkileşim varsa otokorelasyon vardır (Anselin ve Rey, 2014; 257).

Mekânsal otokorelasyon, daha önce açıklandığı üzere komşu olan coğrafi birimler arasında yakınlıktan doğan etkileşime denir. Mekânsal otokorelasyonun ihmal edilmesi EKK tahminlerinde yanlışlık ve etkinlik kaybı sorununa yol açmaktadır. Mekânsal gecikmeli bir bağımlı değişkenin ihmal edilmesi ise dışlanmış değişken hatasına yol açar. Bu durumda EKK tahminleri sapmalı ve tutarsız olur.

Kelejian ve Prucha (1998), Mekânsal hata modeli ve mekânsal gecikme modellerini kapsayan daha geniş bir model tanımlamışlardır. Bu model SARAR(1,1) olarak adlandırılmaktadır ancak literatürde bu model, Anselin (1988) tarafından Genel Mekânsal Model (GSM), Anselin (2010) ve LeSage tarafından mekânsal kombo (SAC- Spatial Combo) modeli olarak adlandırılmaktadırlar.

Mekânsal gecikme modeli ve mekânsal hata modelinin birleşiminden meydana gelen en genel yapıdaki mekânsal kombo regresyon modeli (SAC-spatial Combo) aşağıdaki gibidir (Arbia, 2014: 74-76):

$$y = \rho W_1 y + x\beta + \varepsilon \quad (26)$$

$$\varepsilon = \lambda W_2 \varepsilon + u$$

$$u \sim N(0, \sigma^2 I_n)$$

Denklem 26'da yer alan y ; $n \times 1$ boyutlu bir bağımsız değişken vektörüdür ve x ; $n \times k$ boyutlu açıklayıcı değişkenler vektörüdür. W_1 ve W_2 ; $n \times n$ boyutlu mekânsal ağırlık matrisleridir. Mekânsal hata ve mekânsal gecikme modeli, kombo modelin parametrelerine getirilen kısıtlar aracılığı ile elde edilir.

3.2.1. Mekânsal Hata Modeli (SEM)

Mekânsal hata modeli, farklı konumlara ait hatalar arasındaki kovaryans sıfırdan farklı olduğunda meydana gelen bağımlılığı ifade etmek için kullanılır. Modellerdeki ölçme hataları sonucunda ortaya çıkan bağımlılığa problemlili bağımlılık adı verilir. Bu bağımlılık regresyonların hata terimlerinde görülmektedir.

Denklem 26'daki $W_1 = 0$ olduğunda, hata terimi mekânsal otokorelasyonlu bir regresyon modeli elde edilir (Arbia, 2014: 55). Bu modellere mekânsal hata modeli (SEM) adı verilir, aşağıdaki gibi gösterilir:

$$y = x\beta + \varepsilon \quad (27)$$

$$\varepsilon = \lambda W_2 \varepsilon + u \quad |\lambda| < 1$$

$$u \sim N(0, \sigma_u^2 I_n)$$

Burada λ , ilgili konumun hata terimi ile komşu konumların hata terimleri arasındaki mekânsal bağımlılığın derecesini ölçer. $|\lambda| < 1$ kısıtının etkili olabilmesi için $u|X \approx i. i. d. N(0, \sigma_u^2, I_n)$ koşulunun sağlanması ve ağırlık matrisi W matrisine satır

standartlaştırması uygulandığında bu kısıt geçerli olacaktır. ε matrisi aşağıdaki gibi yeniden düzenlenebilir:

$$(1 - \lambda W_2)\varepsilon = u \quad (28)$$

$$\varepsilon = (1 - \lambda W_2)^{-1}u$$

Denklem 27'deki hata teriminin beklenen değeri ve varyansı aşağıdaki gibidir:

$$E(\varepsilon) = 0$$

$$E(\varepsilon\varepsilon^T) = \sigma_\varepsilon^2(I - \lambda W)^{-1}(I - \lambda W^T)^{-1} = \sigma_\varepsilon^2\Omega \quad (29)$$

$$\Omega = (I - \lambda W)^{-1}(I - \lambda W^T)^{-1}$$

Burada mekânsal hata modelleri varyans-kovaryans matrisi $\sigma_\varepsilon^2\Omega$ gibi küresel olmayan kalıntı varyansı türetilir (Tatlı, 2016; 35).

SEM modellerinde bağımlı değişkenin mekânsal gecikmeli değeri W_y ile ε arasındaki ilişki, içsellik problemine neden olmaktadır.

$$y = x\beta + \varepsilon$$

Denklem 28'deki u hata terimi yukarıdaki model aracılığıyla aşağıdaki gibi elde edilir (Arbia, 2014: 55):

$$(I - \lambda W)y = (I - \lambda W)x\beta + (I - \lambda W)\varepsilon$$

$$y = \lambda W y + x\beta - Wx\rho\beta + u$$

$$y = \lambda W y + x\beta - Wx\gamma + u \quad (30)$$

Burada $\gamma = \rho\beta$ 'dir ve bu nedenle denklem 30 doğrudan tahmin edilebilir. Ancak bu tahmin sırasında iki sorun meydana çıkmaktadır. Bunlardan birincisi; $\gamma = \rho\beta$ kısıtlaması nedeniyle aşırı parametrikleştirilmiştir. İkinci neden ise; $W y$ terimi hata terimi ile ilişkilidir ve bu ilişkide içsellik problemine neden olmaktadır (Arbia, 2014: 56). Bu içsellik denklem 31'deki gibi gösterilebilir.

$$(I - \lambda W)y = x\beta - Wx\gamma + u$$

$$y = (I - \lambda W)^{-1}(x\beta - Wx\gamma) + (I - \lambda W)^{-1}u$$

Wy ile hata terimi arasındaki ilişki sıfır değildir. $E[(Wy)\varepsilon'] \neq 0$ (Arbia, 2014: 56)

$$E[(Wy)\varepsilon^T] = E[W(I - \lambda W)^{-1}(x\beta - Wx\gamma) + W(I - \lambda W)^{-1}u]\varepsilon^T$$

$$= W(I - \lambda W)^{-1}(x\beta - Wx\gamma)E(\varepsilon^T) + W(I - \lambda W)^{-1}E[\varepsilon\varepsilon^T]$$

$$= \sigma_\varepsilon^2 W(I - \lambda W)^{-1}I \neq 0 \quad (31)$$

Denklem 31'de mekânsal gecikmeli değişken Wy , hata terimi ε ilişkilidir. Bu içsellik nedeni ile EKK optimal özelliğini kaybetmektedir (Arbia, 2014: 56). İçsellik problemini gidermek için araç değişken yöntemi uygulanmaktadır. Ancak λ parametresi bilinmediğinden Maksimum Olabilirlik (ML) veya Genelleştirilmiş EKK (GEKK) yöntemleri ile içsellik problemleri giderilebilir.

3.2.2. Mekânsal Gecikmeli Model (SAR)

Mekânsal gecikme modeli, denklem 26'nın kısıtlarla daraltılmış halidir ve $i \neq j$ iken i alanındaki bağımlı değişken Y 'nin gözlemlerinin, komşu alanlardaki gözlemlere bağlı olmasına sebebiyet vermektedir. Mekânsal gecikme modeli, 1. dereceden mekânsal otoregresif model (SAR model) olarak da adlandırılır ve model yapısı aşağıdaki gibi tanımlanabilir (Arbia, 2014: 66):

$$y = \rho Wy + x\beta + \varepsilon \quad |\rho| < 1 \quad (32)$$

Burada β , $(k \times 1)$ boyutlu eğim parametre vektörüdür ve ρ , mekânsal otokorelasyon parametresidir. Ağırlık matrisi W , i konumu üzerindeki j 'nin olası etkileşim derecesini ölçmektedir. Denklem 26'da $\rho \neq 0$, $\lambda = 0$ olduğunda ve $u|X \sim i.i.d.N(0, \sigma_u^2 I_n)$ sağlandığında Mekânsal Gecikmeli Model (SAR) elde edilir. Bu durumda SEM'de olduğu gibi bağımlı değişken y 'nin hata terimi ile ilişkili olması sebebi ile içsellik sorunu ortaya çıkmaktadır.

$$(I - \rho W)y = x\beta + \varepsilon$$

$$y = (I - \rho W)^{-1}x\beta + (I - \rho W)^{-1}\varepsilon \quad (33)$$

$E((Wy)\varepsilon^T) \neq 0$ aşağıdaki gibi bulunur;

$$\begin{aligned} E[(Wy)\varepsilon^T] &= E[W(I - \rho W)^{-1}x\beta + (I - \rho W)^{-1}\varepsilon]\varepsilon^T \\ &= W(I - \rho W)^{-1}x\beta E(u^T) + W(I - \rho W)^{-1}E[\varepsilon\varepsilon^T] \\ &= \sigma_\varepsilon^2 W(I - \rho W)^{-1}I \neq 0 \end{aligned} \quad (34)$$

Bu modelin bağımlı değişkenin beklen değeri ve varyansı şu şekildedir;

$$\begin{aligned} E(y) &= E[(I - \rho W)^{-1}x\beta + (I - \rho W)^{-1}\varepsilon] \\ &= (I - \rho W)^{-1}x\beta \\ y - E(y) &= (I - \rho W)^{-1}\varepsilon - (I - \rho W)^{-1}x\beta \\ &= (I - \rho W)^{-1}(x\beta + u) - (I - \rho W)^{-1}x\beta \\ &= (I - \rho W)^{-1}u \end{aligned} \quad (35)$$

Rassal terim σ 'nın varyansına dayalı olarak y 'nin varyansı şu şekilde yazılabilir:

$$\begin{aligned} E(yy^T) &= \sigma_\varepsilon^2 (I - \rho W)^{-1} (I - \rho W)^{-T} \\ &= \sigma_\varepsilon^2 \Omega \end{aligned} \quad (36)$$

İçsellik sorunu nedeni ile EKK uygulanmaz. Bunun yerine ML ve 2SLS yöntemleri uygulanmaktadır.

3.3. Mekânsal Regresyon Modelleri İçin Tahmin Yöntemleri

Mekânsal modeller de içsellik problemi nedeni ile en küçük kareler yöntemi hatalı sonuçlara neden olmaktadır. Bunun yerine maksimum olabilirlik (ML) yöntemi veya genelleştirilmiş iki aşamalı en küçük kareler (G2EKK) yöntemi kullanılmaktadır (Arbia, 2014: 76).

3.3.1. Maksimum Olabilirlik (ML) Yöntemi

Ord (1975), bir mekânsal gecikmenin ya da mekânsal hata teriminin yapısına mekânsal otokorelasyonu katan modellerin maksimum olabilirlik ile tahminine ilk kapsamlı yaklaşımı gelmiştir. Maksimum olabilirlik tahmin edicileri normallik hipotezini koruyarak, tutarlı ve etkin sonuçlar vermektedir. ML; gözlenen bağımlı değişken y 'nin birleşik olasılık değerine göre ana kütle parametrelerini tahmin eder.

Standart regresyon modeli ele alındığında;

$$y = Z\beta + u$$

Bu denklemden elde edilen hata terimi aşağıdaki gibidir:

$$u = y - Z\beta$$

SEM için varyans-kovaryans matrisi Denklem 36'daki gibi ifade edilir ve bu denklemden artıklar normal dağıldığı varsayımı ile olabilirlik fonksiyonunu elde edilmektedir.

$$L(\lambda, \sigma_\varepsilon^2, \beta) = \text{const} \sigma_\varepsilon^2 |\Omega|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2\sigma_\varepsilon^2} u^T \Omega^{-1} u \right\}$$

Olabilirlik fonksiyonu denklemine hata terimi eşitliği ve $\Omega = (I - \lambda W)^{-1} (I - \lambda W^T)^{-1}$ yazılırsa oluşan yeni denklem:

$$L(\lambda, \sigma_\varepsilon^2, \beta) = \text{const} (\sigma_\varepsilon^2)^{-\frac{n}{2}} |(I - \lambda W)^{-1} (I - \lambda W^T)^{-1}|^{-\frac{1}{2}} \times \exp \left\{ -\frac{1}{2} (y - x\beta)^T [(I - \lambda W)^{-1} (I - \lambda W^T)^{-1}]^{-1} (y - x\beta) \right\} \quad (37)$$

Olabilirlik fonksiyonunun logaritmik olasılığı aşağıdaki gibi ifade edilir:

$$l(\lambda, \sigma_\varepsilon^2, \beta) = \text{const} - \frac{n}{2} \ln(\sigma_\varepsilon^2) - \frac{1}{2} \ln |(I - \lambda W)^{-1} (I - \lambda W^T)^{-1}| - \frac{1}{2} (y - x\beta)^T [(I - \lambda W)^{-1} (I - \lambda W^T)^{-1}]^{-1} (y - x\beta) \quad (38)$$

Parametre tahmini için Denklem 38'deki parametreler doğrusal değildir bu yüzden maksimize edilememektedirler (Zeren, 2010: 26). Çünkü σ_ε^2 ve β tahminleri analitik açıdan λ parametresinin bir fonksiyonudur. Cochrane-Orcutt İteratif Tahmini yöntemiyle bu parametrelerin tahminleri bulunabilmektedir.

SAR için y 'nin beklen değeri ve varyansı kullanılarak olabilirlik fonksiyonu aşağıdaki gibi elde edilmektedir;

$$L(\sigma^2, \rho, \beta: y) = (2\pi)^{-\frac{n}{2}} |\sigma_\varepsilon^2 \Omega|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2\sigma_\varepsilon^2} [y - (I - \rho W)^{-1} x \beta]^T \Omega^{-1} [y - (I - \rho W)^{-1} x \beta] \right\} \quad (39)$$

Denklem 39'un logaritması alınarak log-olabilirlik fonksiyonu aşağıdaki gibidir:

$$l(\sigma^2, \rho, \beta: y) = -\frac{n}{2} \ln 2\pi - \frac{1}{2} \ln |\sigma_\varepsilon^2 \Omega| - \frac{1}{2\sigma_\varepsilon^2} [y - (I - \rho W)^{-1} x \beta]^T \Omega^{-1} [y - (I - \rho W)^{-1} x \beta] \quad (40)$$

Denklem 36'daki Ω Kullanılarak $\sigma_\varepsilon^2 \Omega$ ' nin determinanı aşağıdaki gibi yazılabilir (Arbia, 2014: 67):

$$\begin{aligned} |\sigma_\varepsilon^2 \Omega| &= |\sigma_\varepsilon^2 (I - \rho W)^{-1} (I - \rho W)^{-T}| \\ &= \sigma_\varepsilon^{2n} |(I - \rho W)^{-1} (I - \rho W)^{-T}| \end{aligned} \quad (41)$$

Denklem 41'deki bu determinant kısmı yeniden düzenlenebilir;

$$\begin{aligned} |(I - \rho W)^{-1} (I - \rho W)^{-T}| &= |(I - \rho W)^{-1}| |(I - \rho W)^{-T}| \\ &= |(I - \rho W)|^{-2} \end{aligned}$$

Bu son ifadeye bağlı olarak, denklem 41'in determinanı şöyledir:

$$|\sigma_\varepsilon^2 \Omega| = \sigma_\varepsilon^{2n} |(I - \rho W)|^{-2} \quad (42)$$

Düzenlenen denklem 42, denklem 40'ta yerine yazılacak olursa;

$$\begin{aligned}
l(\sigma^2, \rho, \beta: y) &= -\frac{n}{2} \ln 2\pi - \frac{1}{2} \ln(\sigma_\varepsilon^{2n} |I - \rho W|^{-2}) \\
&\quad - \frac{1}{2\sigma_\varepsilon^2} [y - (I - \rho W)^{-1} x\beta]^T [(I - \rho W)^{-1} (I - \rho W)^{-T}]^{-1} \\
&\quad \times [y - (I - \rho W)^{-1} x\beta] \\
&= -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \sigma_\varepsilon^2 + \ln |I - \rho W| \\
&\quad - \frac{1}{2\sigma_\varepsilon^2} [y - (I - \rho W)^{-1} x\beta]^T (I - \rho W)^T (I - \rho W) [y - (I - \rho W)^{-1} x\beta]
\end{aligned}$$

Yukarıdaki denklemde $(I - \rho W)[y - (I - \rho W)^{-1} x\beta] = (I - \rho W)y - x\beta$ 'dir.

$$l(\sigma^2, \rho, \beta: y) = -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \sigma_\varepsilon^2 + \ln |I - \rho W| - \frac{1}{2\sigma_\varepsilon^2} [(I - \rho W)y - x\beta]^T [(I - \rho W)y - x\beta] \quad (43)$$

Denklem 43'deki log-olabilirlik fonksiyonu, olabilirlik fonksiyonundan farklıdır. Bu farklılığın sebebi fonksiyondaki $\ln |I - \rho W|$ teriminin varlığıdır.

Parametrelerin tahmini için log-olabilirlik fonksiyonunun parametreleri doğrusal değildir ve bu yüzden maksimize edilememektedirler (Zeren, 2010: 26). Çünkü σ_ε^2 ve β tahminleri analitik açıdan ρ parametresinin bir fonksiyonudur. Anselin (1988), parametrelerin bazılarını orijinal olabilirlik fonksiyonunda yerine koyarak problemdeki bilinmeyen sayısını azaltmayı önermiştir.

3.3.2. Genelleştirilmiş İki Aşamalı En Küçük Kareler Yöntemi

İki aşamalı en küçük kareler yöntemi (2SLS) regresyon modellerindeki içsellik problemini gidermek için kullanılan en uygun yöntemlerden biridir. SAR modellerindeki parametre tahminleri için genellikle mekânsal iki aşamalı en küçük kareler yöntemi önerilmektedir. Bunun sebebi, 2SLS tahminçileri tutarlı sonuçlar vermektedir ve asimptotik normalliğe sahiptirler. Burada hata terimi normal dağılımlı olmadığından bu yöntemi kullanmak tutarsız sonuçlara neden olmaktadır. Bu nedenle Genelleştirilmiş Momentler Metodu (GMM) kullanılması önerilmektedir. GMM tahminçileri tutarlı ve model tahmini yönünden oldukça kolay bulunmaktadır.

Mekânsal kombo modelinin tahmini için Genelleştirilmiş Mekânsal İki Aşamalı En Küçük Kareler (GS2SLS) yöntemi kullanılmaktadır. SAC modelinin başka bir gösterimi aşağıdaki gibidir.

$$\begin{aligned} y &= X\beta + \rho Wy + \varepsilon & |\rho| < 1 \\ \varepsilon &= \lambda W\varepsilon + u & |\lambda| < 1 \end{aligned}$$

Burada $|\rho| < 1$, $(I - \rho W)$ tekil olmayan bir matristir. GS2SLS; SAR için kullanılan 2SLS yönteminin bir uzantısıdır. Yukarıdaki modelde gecikmeli Wy teriminden dolayı içsellik problemi ortaya çıkmaktadır. Bu problemten dolayı 2SLS yöntemini kullanarak mekânsal gecikme modeli için parametre tahmin yöntemi uygulanmalıdır. Bu durumda SAR modeli aşağıdaki denkleme dönüştürülür:

$$y = (I - \rho W)^{-1}x\beta + (I - \rho W)^{-1}\varepsilon \quad (44)$$

$E(y) = (I - \rho W)^{-1}x\beta$ 'dir. O halde Wy' nin beklenen değeri aşağıdaki gibidir;

$$\begin{aligned} E(Wy) &= W(I - \rho W)^{-1}x\beta \\ &= W\left[\sum_{p=0}^{\infty} \rho^p W^p\right]x\beta \end{aligned}$$

Bu durumda $W^0 = I$ olduğu söylenebilir. Dolayısıyla içsel değişkenler için araç değişkenler tanımlanır (Yamagata ve Seya, 2019: 144). Bu araç değişkenler (IV) $Z_{[Nxq]}$, $q \geq K + 1$ ile $(X, WX, W^2X, \dots, W^pX)$ 'de doğrusal bağımsız sütun vektörleri kullanılarak tanımlanabilmektedir. Ancak, söz konusu bağımsız sütun vektörleri bağımlı değişken y'yi doğrudan etkiliyorsa bu yöntemin kullanılması uygun değildir.

Denklem 24'te yer alan $\dot{X}_{[NxL]}$ endojen açıklayıcı değişkenler olduğunda regresyon denklemini yeniden yazılacak olursa;

$$y = \rho Wy + X\beta + \dot{X}\dot{\beta} + \varepsilon \quad (45)$$

Burada, Drukker vd. (2013) araç değişkeni $Z_{[Nxq]}$, $q \geq K + L + 1$ ile X değişkeni yerine X_f yazmayı önermiştir. Bu durumda açıklayıcı değişken $Z = X_f, WX_f, W^2X_f, \dots, W^pX_f$ olarak değiştirilebilir.

SAR modelleri için Anselin ve Lazano-Gracia (2008;17), enlem ve boylamı araç değişken olarak kullanmışlar ve bunları X_e olarak tanımlamışlardır. Uygun bir araç değişken kullanılarak S2SLS yönteminin parametre tahminin daha basit olduğunu savunmuşlardır. SAR aşağıdaki gibi yeniden oluşturulmaktadır:

$$y = R\xi + \varepsilon \quad (46)$$

Bu denklemde $R = (Wy; X; \dot{X})$ ve $\xi = (\rho; \beta; \dot{\beta})$ 'yı içermektedirler. S2SLS'nin ξ teriminin tahmincisi ve varyansı aşağıdaki gibidir;

$$\hat{\xi} = (\hat{R}'\hat{R})^{-1}\hat{R}'y$$

$$Var(\hat{\xi}) = \hat{\sigma}_\varepsilon^2(\hat{R}'\hat{R})^{-1} = \hat{\sigma}_\varepsilon^2[R'Z(Z'Z)^{-1}Z'R]^{-1}$$

Burada R tahmincisi, $\hat{R} = PR'$ dir ve $P = Z(Z'Z)^{-1}Z'$.

SEM modelinde S2SLS tahmincisi tutarlı olmadığından dolayı ML yöntemi ya da normallik varsayımı gerektirmeyen GMM yöntemi kullanılması önerilmektedir.

SEM için regresyon denklemi aşağıdaki gibidir (Yamagata ve Seya, 2019: 146):

$$y = X\beta + \varepsilon \quad (47)$$

$$\varepsilon = \lambda W\varepsilon + u \quad Var(u) = \sigma_u^2 I$$

Burada $\varepsilon - \lambda W\varepsilon = u$ olduğundan her iki tarafı tekrardan ağırlık matrisi ile çarpılır ise yeni denklem aşağıdaki gibi elde edilir.

$$W\varepsilon - \lambda W^2\varepsilon = Wu \quad (48)$$

Burada $W^2 = W.W$ dir. Yukarıdaki her iki hata terimi denklemleri olan 47 ve 48'deki $\varepsilon, W\varepsilon, W^2\varepsilon, u, Wu$ terimleri $\varepsilon_i, \dot{\varepsilon}_i, \ddot{\varepsilon}_i, u_i, \dot{u}_i$ olarak yeniden tanımlanır. Bu denklemler tanımlanan yeni terimler ile yeniden yazılır ise sırasıyla,

$$\varepsilon_i - \lambda \dot{\varepsilon}_i = u_i$$

$$\dot{\varepsilon}_i - \lambda \ddot{\varepsilon}_i = \dot{u}_i \quad (49)$$

Eşitlik 49'daki denklemlerin toplam kareleri alınarak N ye bölünür ise;

$$\begin{aligned}\frac{1}{N}\sum_i \varepsilon_i^2 + \lambda^2 \frac{1}{N}\sum_i \dot{\varepsilon}_i^2 - 2\lambda \frac{1}{N}\sum_i \varepsilon_i \dot{\varepsilon}_i &= \frac{1}{N}\sum_i u_i^2 \\ \frac{1}{N}\sum_i \dot{\varepsilon}_i^2 + \lambda^2 \frac{1}{N}\sum_i \ddot{\varepsilon}_i^2 - 2\lambda \frac{1}{N}\sum_i \dot{\varepsilon}_i \ddot{\varepsilon}_i &= \frac{1}{N}\sum_i \dot{u}_i^2\end{aligned}\quad (50)$$

Eşitlik 50'deki denklemlerin sağ ve sol tarafı alt alta çarpılıp aşağıdaki denklem elde edilmektedir.

$$\frac{1}{N}\sum_i \varepsilon_i \dot{\varepsilon}_i + \lambda^2 \frac{1}{N}\sum_i \dot{\varepsilon}_i \ddot{\varepsilon}_i - \lambda \left(\frac{1}{N}\sum_i \varepsilon_i \ddot{\varepsilon}_i + \frac{1}{N}\sum_i \dot{\varepsilon}_i^2 \right) = \frac{1}{N}\sum_i u_i \dot{u}_i \quad (51)$$

Denklem 50 ve 51 aracılığıyla üç moment koşulu geçerlidir. Bu üç moment koşulu aşağıdaki gibidir;

1. Koşul $E\left(\frac{u'u}{N}\right) = \sigma_u^2$
2. Koşul $E\left(\frac{u'W'Wu}{N}\right) = \sigma_u^2 N^{-1} tr(W'W)$
3. Koşul $E\left(\frac{u'W'u}{N}\right) = 0$

Üçüncü koşul denklemdaki ağırlık matrisinin köşegen elemanlarının sıfır olmasına bağlıdır. Bu üç koşul ε hata terimi cinsinde yeniden yazılırsa da sonuçlarda herhangi bir değişme olmamaktadır.

1. Koşul $E\left(\frac{\varepsilon'(I-\lambda W)'(I-\lambda W)\varepsilon}{N}\right) = \sigma_u^2$
2. Koşul $E\left(\frac{\varepsilon'(I-\lambda W)'W'W(I-\lambda W)\varepsilon}{N}\right) = \sigma_u^2 N^{-1} tr(W'W)$
3. Koşul $E\left(\frac{\varepsilon'(I-\lambda W)'W'(I-\lambda W)\varepsilon}{N}\right) = 0$

Moment koşullarının matris formu $\Gamma[\lambda, \lambda^2, \sigma_u^2]' - \gamma = 0$ bu eşitlik ile birlikte aşağıdaki gibi ifade edilir;

$$\Gamma = \begin{bmatrix} -\frac{2}{N}E(\varepsilon'\varepsilon) & \frac{1}{N}E(\dot{\varepsilon}'\dot{\varepsilon}) & -1 \\ -\frac{2}{N}E(\dot{\varepsilon}'\dot{\varepsilon}) & \frac{1}{N}E(\ddot{\varepsilon}'\ddot{\varepsilon}) & -\frac{1}{N}tr(W'W) \\ \frac{1}{N}E(\varepsilon'\ddot{\varepsilon} + \dot{\varepsilon}'\varepsilon) & \frac{1}{N}E(\dot{\varepsilon}'\ddot{\varepsilon}) & 0 \end{bmatrix}, \gamma = \begin{bmatrix} \frac{1}{N}E(\varepsilon'\varepsilon) \\ \frac{1}{N}E(\dot{\varepsilon}'\dot{\varepsilon}) \\ \frac{1}{N}E(\varepsilon'\dot{\varepsilon}) \end{bmatrix}$$

Moment koşulunda ki \hat{u} tahmincisi EKK yöntemi kullanılarak tahmin edilir. Aynı zamanda $\hat{u} \equiv W\hat{u}$ ve $\hat{u} \equiv W.W\hat{u}$ olduğu da varsayılırsa örnek olaylar aşağıdaki gibi yazılabilmektedir:

$$G[\lambda, \lambda^2, \sigma_u^2]' - g = \mu(\lambda, \sigma_u^2) \quad (52)$$

Burada $\mu(\lambda, \sigma_u^2)$, 3×1 boyutlu hata terimi vektörü olarak kabul edilir ve bu $\mu(\lambda, \sigma_u^2)' \mu(\lambda, \sigma_u^2)$ terim minimize edilir. Bu minimizasyon sonucunda GMM tahmincileri $\hat{\lambda}, \hat{\sigma}_u^2$ bulunur ve β tahmincilerine dayanarak tahmini genelleştirilmiş en küçük kareler (EGLS) yöntemi ile tahmin edilir (Yamagata ve Seya, 2019: 148).

3.4. Mekânsal Büyük Veri Tahmin Edicileri

Mekânsal regresyon modellerinde içsellik problemlerinden dolayı EKK yöntemi tahmincileri sapmalı ve tutarsız tahminlere neden olmaktadır. Bunun yerine maksimum olabilirlik yöntemi uygulanır. Benzer şekilde mekânsal büyük veriler için de maksimum olabilirlik yöntemi kullanılmaktadır. Daha önceden de belirtildiği üzere logaritmik olabilirlik fonksiyonunda parametrelerin yeterince doğrusal olmaması nümerik maksimizasyon varsayımının uygulanması gerekmektedir. Bu varsayıma dayanarak parametre tahmini yapılsa bile gözlem sayısının artmasıyla tahmin sürecinde problemler ortaya çıkmaktadır. Logaritmik olabilirlik fonksiyonunun maksimizasyonu, logaritmik determinantın bulunması ve determinantın boyutu örneklem büyüklüğüne bağlıdır. Ayrıca her bir nümerik çözümde mekânsal bağımlılık parametresinin her bir değeri için ayrı ayrı bu determinant hesaplanır. Büyük veriler için bu determinantın hesaplanması oldukça zordur. Bu problemi ortadan kaldırmak amacıyla büyük veriye dayalı mekânsal modellerde geçerli çeşitli yöntemler bulunmaktadır (Arbia, 2014: 168). Bu yöntemler şu şekildedir;

- Üstel Mekânsal Matris Tanımı (MESS),
- Tek Taraflı Yaklaşıklık Yöntemi,
- Kompozit Olabilirlik Yaklaşımı,

3.4.1. Mekânsal Üstel Matris Belirtimi (MESS)

Ülke, bölge, şehir veya mekânsal noktalardan elde edilen veriler genellikle mekânsal bağımlılık sergilerler. Bu bağımlılığı analize dâhil ederek kontrol edebilen özel tahmin yöntemleri gerekmektedir. Mekânsal modellerinin maksimum olabilirlik (ML) yöntemine göre tahmini için kullanılan logaritmik olabilirlik fonksiyon parametrelerinin doğrusal olmaması nedeniyle nümerik maksimizasyon yaklaşımı kullanılmaktadır. Bu yaklaşım ile parametre tahmini yapılsa bile gözlem sayısının artmasıyla tahmin problemleri ortaya çıkmaktadır (Arbia, 2014: 172). Log-olabilirlik fonksiyonunun maksimizasyonunda logaritmik determinantın bulunması ve determinantın boyutu örneklem büyüklüğüne bağlıdır. Bu nedenle her bir sayısal çözümde mekânsal bağımlılık parametresinin her bir değeri için ayrı ayrı determinant hesaplanır. Büyük veriler için bu determinantların hesaplanması oldukça zor olmaktadır (Kangallı Uyar, 2020: 72). Bu problemi ortadan kaldırmak için büyük verili mekânsal modellerde geçerli olan yaklaşımlardan Üstel Mekânsal Matris Belirtimi (MESS) yaklaşımı uygulanabilir.

MESS büyük örnek verilerini inceler ve maksimum olabilirlik tahmini problemine kapalı bir formda çözüm üretmektedir (Arbia, 2014: 167). MESS ile geleneksel tahmin yöntemlerine benzer tahminler elde etmektedir ve değişen varyans altında tutarlı sonuçlar sunmaktadır.

Mekânsal üstel matris belirtimini (MESS) ilk olarak LeSage ve Pace (2007) komşu konumlar arasındaki etkileşimi tanımlamak amacıyla büyük verili SAR modeli olarak öne sürmüşlerdir.

Olabilirlik maksimizasyonu ile ilgili hesaplama problemleri 1970'lerde düşük hızlı ve küçük bellekli bilgisayarların olduğu dönemlerde sıkça karşılaşılan bir konudur. Artan bilgi ve işlem gücü, mevcut bilgisayarlar açısından büyük veri tabanlarına hâkimiyeti de artmaktadır. Mekânsal ekonometrik analiz tekniklerinde mekânsal olarak birbirine bağımlı milyonlarca gözlem bulunmaktadır. Örneğin 1000×1000 boyutlu matrislerin öz değerlerini hesaplama işleminde problem ortaya çıkmaktadır. Bu problemlerden biri mekânsal modelin varyans-kovaryans matrisi ile ilgilidir. Bu problemlere çözüm olarak MESS yaklaşımı geliştirilmiştir.

3.4.1.1. SAR İçin Mekânsal Üstel Matris Belirtimi

MESS modelinin log-olasılığı SAR durumunda hesaplama problemlerinin ana kaynağı olan logaritmik determinantın hesaplanmasını gerektirmemektedir (Arbia, 2014:

170). Büyük boyutlu mekânsal veriler için SAR yöntemine uygulanan MESS spesifikasyonu aşağıdaki gibidir (Arbia, 2014: 169);

$$y = \rho W y + x\beta + \varepsilon \quad |\lambda| < 1$$

Burada y vektörü, bağımlı değişken üzerinde bir bölge veya nokta ile ilişkili n gözlem içermektedir. X matrisi her bölge için bağımsız değişken üzerindeki gözlemlere karşılık gelen $n \times k$ boyutlu matristir ve hata terimi $\varepsilon|X \approx i. i. d. N(0, \sigma_\varepsilon^2, I_n)$ 'dur.

SAR için varyans-kovaryans matrisi aşağıdaki gibi hesaplanır:

$$\begin{aligned} E(y) &= E[(I - \rho W)^{-1}x\beta + (I - \rho W)^{-1}\varepsilon] = (I - \rho W)^{-1}x\beta \\ y - E(y) &= (I - \rho W)^{-1}\varepsilon - (I - \rho W)^{-1}x\beta \\ &= (I - \rho W)^{-1}(x\beta + u) - (I - \rho W)^{-1}x\beta \\ &= (I - \rho W)^{-1}u \end{aligned} \quad (53)$$

Beklenen değerine göre y 'nin varyansı Denklem 36'daki gibi yazılır. y 'nin ortalama ve varyans-kovaryans matrisinden yararlanılarak denklem 43'deki gibi log-olabilirlik fonksiyonu yazılmaktadır. Bu fonksiyon içindeki varyans-kovaryans matrisindeki $(I - \rho W)^{-1}$ ifadesi çok büyük boyutlu veri setlerinde hesaplamayı zorlaştırmaktadır. Bu problemi ortadan kaldırmak için $MESSS = (I - \rho W)$ şeklinde düzenlenebilmektedir.

$$Sy = x\beta + \varepsilon \quad (54)$$

$$(I - \rho W)y = x\beta + \varepsilon$$

$$y = (I - \rho W)^{-1}x\beta + \varepsilon$$

$(I - \rho W)^{-1}$ ifadesi aşağıdaki gibi yazılabilir (Horn ve Johnson, 1993, s. 301);

$$(I - \rho W)^{-1} = I_n + \rho W + \rho^2 W^2 + \dots$$

Denklem 54'teki hata teriminin varyans-kovaryans matrisinde yardımcı olması için toplam gecikme işlemcisi aşağıdaki gibidir (LeSage ve Pace, 2007;192):

$$\Sigma \rho = \sigma^2 [(I - \rho W)(I - \rho W')]^{-1} \quad (55)$$

ρ , mekânsal bağımlılığın büyüklüğünü yansıtan skaler bir parametredir. Chiu ve diğerleri (1996) y'nin $S = (I - \rho W)$ dönüşümüne dayalı SAR dönüşümü yerine, yeni baştan bir dizi sergileyen y dönüşümünün matris üstel belirtimi avantajları olduğunu belirtmiş ve kovaryans matrisinin modellenmesi için üstel matrisin (S) kullanımını önermiştir.

$$\begin{aligned} S &= e^{\alpha W} \\ Sy &= x\beta + \varepsilon \end{aligned} \quad (56)$$

Burada S , gerçek pozitif tanımlı bir matristir ve bu matrisin farklı tanımları varyans-kovaryans matrisi üzerinde de bir etkiye sahiptir.

Chiu (1996;198) Denklem 56'daki S matrisinin sağladığı avantajlardan şöyle bahsetmiştir;

- Herhangi bir reel pozitif tanımlı S matrisi için her zaman reel pozitif tanımlı bir αW matris elde edilebilir.
- Ağırlık matrisi gibi herhangi bir simetrik matris için, S pozitif tanımlı bir matristir.
- S matrisinin tersi $S^{-1} = e^{-\alpha W}$ ve determinanı $|S| = e^{tr(\alpha W)}$ şeklinde hesaplanmaktadır. Ağırlık matrisinin tanımı gereği diagonal $w_{ii} = 0$ olmasından dolayı S matrisinin determinanı bire eşittir ($|S| = e^{tr(\alpha W)} = e^0 = 1$).

LeSage ve Pace (2007;193), aşağıdaki gibi $S(\alpha)$ tanımlamıştır;

$$S(\alpha) = e^{\alpha W} = I + \alpha W + \frac{\alpha^2}{2} W^2 + \dots + \frac{\alpha^t}{t!} W^t + \dots \quad (57)$$

Denklem 57, ağırlık matrisi W 'nin üstel matris sonsuz seri açılımında yerine koyulduğunda elde edilen terimi ve $-\infty < \alpha < +\infty$ aralığında tanımlanan gerçek bir skaler parametreyi içermektedir.

$S(\alpha)$ tekil değildir. Bu da Üstel matris tanımına dayalı hata kovaryans matrisinin aşağıdaki gibi ifade edilebileceğini göstermektedir (LeSage ve Pace,2007;193).

$$\begin{aligned}\Sigma \alpha &= \sigma^2(S(\alpha)S(\alpha)')^{-1} = \sigma^2(S(\alpha)')^{-1}S(\alpha)^{-1} \\ &= \sigma^2(e^{-\alpha W'}e^{-\alpha W})\end{aligned}\quad (58)$$

Strang (1976,270), geometrik ve üstel mekânsal özelliklerini karşılaştırmak amacıyla ρ ve α 'nın büyüklüğünü ilişkilendirmiştir. Bir matrisin normunun matrisin büyütme gücünün bir ölçüsü olduğunu belirtmiştir. $S = (I - \rho W) = e^{-\alpha W}$ eşitliğinde ki matrisin maksimum satır toplamları için $I - \rho W$ normu ve $e^{\alpha W}$ normunu eşitleyerek yakın bir ilişki oluşturmuştur. Bu ilişki aşağıdaki gibi tanımlanmıştır;

$$1 - \rho = e^{\alpha} \quad (59)$$

Horn ve Johnson (1993)'a göre maksimum satır toplamı normu $||| \cdot |||_{\infty}$ şeklindedir. SAR ve MESS yaklaşımında bu norm farklıdır. Satır bazında normalleştirilmiş ağırlık matrisi W içeren SAR modeli için $||| I - \rho W |||_{\infty} = 1 - \rho$ 'dir. MESS yaklaşımında ise $||| e^{\alpha W} |||_{\infty} = e^{\alpha}$ 'dir. Her iki yaklaşımdaki normlar eşitlenecek olursa $\alpha = 1 - e^{\alpha}$ ve $\rho = \ln(1 - \rho)$ bulunur. Bu sonuç yaklaşık ilişkilere yol açar ve modeli yorumlamayı kolaylaştırır. $|\rho| < 1$ verildiğinde, ρ 'nun pozitif değeri için, α aralığının $-\infty < \alpha \leq 0$ olduğu; ρ 'nun negatif değeri için α aralığının $0 < \alpha \leq 0.693147$ olduğu belirtilmiştir. Özellikle $\rho = 0$ olduğunda, $\alpha = 0$, $\lambda \rightarrow 1$ olduğunda $\alpha \rightarrow -\infty$ olur. ρ 'nun değeri 1' çok yakın değerler aldığıında $\alpha = -5$ değerini alır (LeSage ve Pace,2007; 194).

SAR modelleri α parametre tahminleri ve bunların ρ cinsinden yazılmasına izin verir. Bunu uygulama amacı $S(\alpha) = e^{\alpha W} = I + \alpha W + \frac{\alpha^2}{2} W^2 + \dots + \frac{\alpha^t}{t!} W^t + \dots$ denkleminde gösterilen sonsuz genişlemeyi q kesme terimi ile keserek $e^{\alpha W}$ 'ye yakın bir değer bulmaktır. Bu ortaya çıkan kesme hatasının büyüklüğü hakkında bir bilgi bulunmamaktadır. Fakat satır standardize edilmiş ağırlık matrislerinin özellikleri bu kesme hatasını analiz etmeye yardımcı olur ve bu problemi ortadan kaldırmaktadır (Arbia, 2014: 172)

3.4.1.1.1. SAR Mekânsal Üstel Matris Belirtimi için Maksimum Olabilirlik Yaklaşımı

MESS yaklaşımı için bir logaritmik olasılık hesaplanır. Bu logaritmik olasılık, SAR durumunda hesaplama probleminin esas sebebi olan logaritmik determinantın hesaplanmasını gerektirmemektedir.

Tahmin ve hipotez testi amaçları için denklem 54'te bulunan $I - \rho W$ yerine S yazılır ise y 'nin varyansı aşağıdaki gibidir (Arbia, 2014: 171);

$$E(yy^T) = \sigma_\varepsilon^2 \Omega = \sigma_\varepsilon^2 (I - \rho W)^{-1} (I - \rho W)^{-T} = \sigma_\varepsilon^2 S^{-1} S^{-T} \quad (60)$$

Dolayısıyla y 'nin varyansı dikkate alındığında logaritmik olasılığı aşağıdaki gibi tanımlanır:

$$\begin{aligned} l(\sigma^2, \rho, \beta; y) &= -\frac{n}{2} \ln 2\pi - \frac{1}{2} \ln |\sigma_\varepsilon^2 S^{-1} S^{-T}| - \frac{1}{2\sigma_\varepsilon^2} [y - S^{-1}x\beta]^T S S^T [y - S^{-1}x\beta] \\ &= -\frac{n}{2} \ln 2\pi - \frac{1}{2} \ln |\sigma_\varepsilon^2 S^{-1} S^{-T}| - \frac{1}{2\sigma_\varepsilon^2} [Sy - Z\beta]^T [Sy - Z\beta] \end{aligned} \quad (61)$$

Denklemin determinant kısmı $|S^{-1}S^{-T}| = |S^{-1}||S^{-T}|$ şeklinde düzenlenir ve yukarıdaki $|S| = e^{tr(\alpha W)} = e^0 = 1$ denkleminin avantajlardan yararlanılarak $|S^{-1}S^{-T}| = 1$ olarak basitleştirilerek logaritmik olasılık yeniden aşağıdaki gibi ifade edilir;

$$l(\sigma^2, \rho, \beta; y) = -\frac{n}{2} \ln 2\pi - \frac{1}{2} \sigma_\varepsilon^2 - \frac{1}{2\sigma_\varepsilon^2} [Sy - Z\beta]^T [Sy - Z\beta] \quad (62)$$

Burada S matrisinin determinanı 1'e eşittir. MESS modeli için maksimum olabilirlik yaklaşımı α 'ya göre dönüştürülmüş modelin hataların kareler toplamına karşılık gelen $[Sy - Z\beta]^T [Sy - Z\beta]$ terimini minimize etmeye eşdeğerdir.

Başka bir yol ile çözülecek olursa $S(\alpha) = e^{\alpha W} = I + \alpha W + \frac{\alpha^2}{2} W^2 + \dots + \frac{\alpha^t}{t!} W^t + \dots$ denklemindeki sonsuz genişleme yaklaşık bir sonuca götüren q teriminde kesilmelidir;

$$S = e^{\alpha W} = \sum_{i=0}^{\infty} \frac{\alpha^i W^i}{i!} \approx \sum_{i=0}^q \frac{\alpha^i W^i}{i!} \quad (63)$$

LeSage ve Pace (2007) yapmış oldukları çalışmada sonsuz genişlemeyi durduran q kesme teriminin mekânsal korelasyon (pozitif veya negatif) miktarına bağlı olduğunu, örneklem büyüklüğüne bağlı olmadığını göstermişlerdir. Simülasyon çalışması ile mekânsal korelasyonun mutlak değerinin çok yüksek olmadığına (0.95'ten daha düşük olduğunda $q = 16$ noktasında yapılan bir kesmenin çok iyi bir yaklaşımı ürettiğini ifade etmişlerdir. Denklem 63'te ρ 'nın yüksek değerleri için genişleme modelinde daha çok kesme terimine ihtiyaç bulunmaktadır. Ağırlık matrisi elemanları seyrek ise hesaplama oldukça basitleşmektedir. Eğer ağırlık matrisi elemanları oldukça yoğun olduğunda (örneğin sosyal network çalışmalarında) ise $\sum_{i=0}^{\infty} \frac{\alpha^i W^i}{i!}$ dönüşüm terimi daha yoğun olacaktır, bu nedenle hesaplama amacıyla daha fazla hafıza ve daha fazla zaman gerektirecektir.

Mekânsal üstel matris belirtimi MESS'in diğer bir avantajı parametre tahminleri için kapalı formlu çözümler üretmenin mümkün olabilme olasılığıdır (LeSage ve Pace, 2007: 195). Kapalı bir form çözümleri elde etmek için matris şeklindeki $S(\alpha)$ 'ya ve bağımlı değişken y 'nin bütün değerleri ve çeşitli derecelerdeki gecikme terimlerini içeren aşağıdaki gibi bir matrise ihtiyaç vardır.

$$Y = [y, Wy, W^2y, \dots, W^{q-1}y] \quad (64)$$

Burada q değeri S genişleme modelindeki kesme terimidir. y ; $n \times 1$ boyutlu vektördür ve Y ; $n \times q$ boyutlu bir matristir. Modelin parametre tahminlerini çözmek için $(q \times q)$ boyutlu diyagonal bir matris aşağıdaki gibi tanımlanmıştır. Bu matris, genişleme serisinin (denklem 63) katsayılarının bir kısmını içermektedir;

$$G = \begin{vmatrix} \frac{1}{0!} & 0 & \dots & 0 \\ 0 & \frac{1}{1!} & \dots & 0 \\ \dots & \dots & \dots & 0 \\ 0 & 0 & 0 & \frac{1}{(q-1)!} \end{vmatrix} \quad (65)$$

Ayrıca mekânsal korelasyonu kontrol eden α parametresinin gücünü içeren q elemanlı sütun vektörü $v(\alpha)^T$ şöyledir;

$$v(\alpha)^T = [1, \alpha, \alpha^2, \dots, \alpha^{q-1}] \quad (66)$$

Denklem 64'deki, Denklem 65 ve 66'daki ifadelerden yararlanılarak eşitlik 67'deki ifade aşağıdaki gibi yazılabilir:

$$Sy \approx YG(\alpha) \quad (67)$$

Denklem 67'in her iki tarafı idempotent projeksiyon matris $H = I - X(X^T X)^{-1} X^T$ ile çarpılır. Böylece dönüştürülmüş elemanların bir fonksiyonu olan aşağıdaki gibi kalıntı (artıkların) karelerinin toplamı elde edilir (Arbia, 2014: 173):

$$u^T u = v(\alpha)^T G(Y^T H^T H Y) G v(\alpha)$$

veya daha basit bir şekilde ifade edilecek olursa;

$$u^T u \approx v(\alpha)^T Q v(\alpha) \quad (68)$$

Bu denklemde $Q = G(Y^T H^T H Y) G$ ' dir.

Denklem 66'daki polinomial kısıta rağmen Denklem 67 α 'ya göre tek bir minimuma sahiptir (LeSage and Pace,2007:196).

$$\min(u^T u) \approx \min(v(\alpha)^T Q v(\alpha))$$

Yukarıdaki minimum değeri $\hat{\alpha}$ 'nın minimum değerine bağlı olarak aşağıdaki parametre tahmincileri elde edilmektedir.

$$\hat{\sigma}_\varepsilon^2 = \hat{\varepsilon}(\alpha)' \hat{\varepsilon}(\alpha) / n$$

$$\hat{\beta} = (X' X)^{-1} X' e^{-\hat{\alpha} W} y$$

İkinci merteye koşulları ile parametrelere özgü hipotez testleri ve çıkarılma için standart hatayı elde etme imkânı vermektedir (Arbia, 2014: 173).

3.4.1.2. SEM İçin Mekânsal Üstel Matris Belirtimi

Mekânsal gecikme modellerdekine benzer bir biçimde mekânsal hata modelleri içinde MESS yaklaşımı oldukça basittir. Bir mekânsal hata modeli aşağıdaki gibi tanımlanır:

$$y = x\beta + \varepsilon$$

$$\varepsilon = \lambda W_2 \varepsilon + u \quad |\lambda| < 1$$

$u \approx N(0, \Omega)$ 'dır. Ω burada regresyon artıkları için genel bir varyans-kovaryans matrisi denklemdir ve değeri aşağıdaki gibidir;

$$\begin{aligned} E(\varepsilon\varepsilon^T) &= \sigma_\varepsilon^2 (I - \lambda W)^{-1} (I - \lambda W^T)^{-1} = \sigma_\varepsilon^2 \Omega \\ \Omega &= (I - \lambda W)^{-1} (I - \lambda W)^{-T} \end{aligned} \quad (69)$$

Denklem 69'daki Ω 'yı MESS şeklinde $\Omega = e^{\alpha W}$ varsayıldığında aşağıdaki gibi bir açılıma sahip olur (Arbia, 2014: 174).

$$\Omega = e^{\alpha W} = \sum_{i=0}^{\infty} \frac{\alpha^i W^i}{i!} \quad (70)$$

ve

$$\Omega^{-1} = e^{-\alpha W} = \sum_{i=0}^{\infty} \frac{-\alpha^i W^i}{i!}$$

$$\Omega^{-1} = I - \alpha W + \frac{\alpha^2}{2} W^2 - \frac{\alpha^3}{6} W^3 + \dots$$

Denklem 70'deki matrisin tersi o halde;

$$\Omega^{-1} = (I - \lambda W)(I - \lambda W)^T$$

Bu denklem düzenlenecek olursa;

$$\Omega^{-1} = (I - \lambda W)(I - \lambda W)^T = (I - \lambda W)(I - \lambda W^T) \quad (71)$$

$$= I - \lambda W - \lambda W^T + 2 \lambda W W^T$$

$$= I - 2 \lambda W + \lambda^2 W^2 \quad (72)$$

Denklem 72, denklem 70'in özel bir şeklidir (Arbia, 2014: 175). Burada $-2\lambda = -\alpha$ ve $\lambda^2 = -\frac{\alpha^2}{2}$ olduğunda genişleme modeli $q=2$ 'de kesilir.

MESS yaklaşımının esnekliğini arttırmak için genel bir ağırlık matrisi kullanılabilir. Bunun için üstel matris dönüşümüne ilave bir parametre eklenir. Böylece yüksek dereceli komşularda korelasyon decay (bozunma) hızı kontrol edilebilir, denklem 56 kullanılmak üzere aşağıdaki ağırlık matrisi önerilmiştir (Arbia, 2014: 175).

$$W = \sum_{i=1}^m \frac{\phi^i W_i}{\sum_{i=1}^m \phi^i} \quad (73)$$

Burada W ; i . en yakın komşu elemanları için bir ağırlık matrisi, ϕ ; bir decay (bozunma) terimidir ve $0 \leq \phi \leq 1$ aralığındadır. Paydadaki $\sum_{i=1}^m \phi^i$ terimi bir normalleştirme faktörüdür. Tahmin edilecek m ve ϕ parametrelerinden dolayı bu esneklik yüksek hesaplama gerektirmektedir. LeSage ve Pace (2009,250) parametreler için uygun yöntem belirleyerek tahmin süreci ve hipotez kurulumu için Bayesian bir yaklaşım önermektedirler.

3.4.1.2.1. Bayesian Yöntemi

MESS modeline Bayesian bir yaklaşım; $S(\alpha)y = X\beta + \varepsilon$ modelindeki parametrelerin sonsal dağılımının analizini gerektirir. Model tahmini veren sonsal (posterior) dağılım, örneklem dağılımı ile önsel (a priori) dağılımın birleşiminden oluşmaktadır. Tahmin problemini çözmek için önsel bilgilerin kullanılması bazı araştırmacılar tarafından bu yöntemin dezavantajı olarak görülmektedir. Büyük veri örnekleri mekânsal ekonometrik modellerin oluşumu öncesi sonsal dağılımı az da olsa belirlemektedir. Bayesian yöntemi, örnek veriler ve bu verilerin bilgileri ile ağırlık matrislerinin ortalamasını bulmaya çalışmaktadır. Fakat bu ağırlıklar önsel bilgilere göre mevcut verilerin bilgilerinden etkilenmektedirler.

Mekânsal gecikmeli model $y = \rho Wy + X\beta + \varepsilon$ kullanılarak önceki ve mevcut verilerin birleşiminin olasılığı yazıla bilinir. $A = (I - \rho W)$ İfadesi kullanılarak bu olasılık şöyledir (LeSage ve Pace,2009: 128).

$$p(D|\beta, \sigma, \rho) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp[-(1/2\sigma^2)(Ay - X\beta)'(Ay - X\beta)] \quad (74)$$

Yukarıdaki modelin parametreleri için mevcut dağılımın olasılığı ile önsel dağılımlar belirtilmelidir. β ve σ^2 parametreleri için normal ters gama (NIG) tanımlanır. Bu tanım β parametresinin önsel dağılımını, σ^2 parametresi için ters gama dağılımına bağımlı hale getirmektedir. ρ parametresi modelde oldukça önemlidir ve bunun için aralık belirtilebilir ($1/\lambda_{min}, 1/\lambda_{max}$). Burada λ_{min} ve λ_{max} ağırlık matrisinin özdeğerleri dir (LeSage ve Pace,2009: 128). Mekânsal gecikmeli modellerin ağırlık matrisleri bu aralıkta ρ parametresinin oluşumuna sebebiyet vermektedir.

$$y = \rho W y + X \beta + \varepsilon$$

$$\varepsilon \sim N(0, \sigma^2 I_n)$$

$$\pi(\beta, \sigma^2) \sim NIG(c, T, a, b)$$

$$= \pi(\beta | \sigma^2) \pi(\sigma^2)$$

$$= N(c, \sigma^2 T) IG(a, b)$$

$$= \frac{b^a}{(2\pi)^{k/2} |T|^{1/2} \Gamma(a)} (\sigma^2)^{-(a+(k/2)+1)} \times \exp[-\{(\beta - c)' T^{-1} (\beta - c) + 2b\} / (2\sigma^2)]$$

Aşağıda ters gama dağılımı parametre haline getirilmiştir. Burada $\Gamma(\cdot)$ Standart gama fonksiyonunu, $\Gamma(a) = \int_0^\infty t^{a-1} e^{-t}$ temsil etmektedir. Önsel bilgileri belirtmek için kullanılan parametreler NIG' den gelen parametrelerdir (LeSage ve Pace,2009: 128-129).

$$\pi(\sigma^2) = \frac{b^a}{\Gamma(a)} (\sigma^2)^{-(a+1)} \exp(-b/\sigma^2)$$

$$\sigma^2 > 0, a, b > 0$$

$$\pi(\rho) \sim U(\lambda_{min}^{-1}, \lambda_{max}^{-1})$$

Gecikme modelinin olabilirlik fonksiyonu ile önsel bilgiyi birleştirmek için Bayesian teoremi kullanılarak çekirdek sonsal dağılım elde edilir ve bu dağılım aşağıdaki gibi gösterilir;

$$p(\beta, \sigma^2, \alpha | D) \propto \sigma^{-(n+1)} \exp[-(1/2\sigma^2)(Sy - X\beta)'(Sy - X\beta)] \pi(\alpha) \quad (75)$$

Denklem 75'teki modelde gama dağılımının özellikleri kullanılarak aşağıdaki terimleri bulmak için σ^2 parametresi bütünleştirilebilir (LeSage ve Pace, 2009: 250).

$$\begin{aligned}
p(\beta, \alpha|D) &\propto ([y'S(\alpha)'MS(\alpha)y] + [\beta - \beta(\alpha)]'X'X[\beta - \beta(\alpha)])^{-n/2}\pi(\alpha) \\
\beta(\alpha) &= (X'X)^{-1}X'S(\alpha)y \\
S(\alpha) &= e^{\alpha W}
\end{aligned} \tag{76}$$

Denklem 76'da bulunan terimlerin mekânsal bağımlılık parametresi olan α 'ya bağımlılığını belirtmek için $S(\alpha)$ ve $\beta(\alpha)$ değerleri yazılır (LeSage ve Pace, 2009: 250). Bu ortak dağılımın mekânsal bağımlılık parametresi α için sonsal dağılımı bulmak üzere β 'ya göre bütünleştirilebilen bir t dağılımına uymakta ve α koşuluna bağlıdır.

$$p(\alpha|D) \propto ([y'S(\alpha)'MS(\alpha)y])^{-(n-k)/2}\pi(\alpha) \tag{77}$$

Denklem 77'de bulunan $y'S(\alpha)'MS(\alpha)y = Z(\alpha)$ ifadesi $2q - 2$ derecelik polinomdur. α parametresinin sonsal beklentisi aşağıdaki gibidir:

$$E(\alpha|D) = \alpha^* = \frac{\int_{-\infty}^{+\infty} \alpha.p(\alpha|D)d\alpha}{\int_{-\infty}^{+\infty} p(\alpha|D)d\alpha} \tag{78}$$

Denklem 78'de yer alan uyum limitlerine dikkat etmek gerekir çünkü uyum limitinin üst sınırının sıfır ile sınırlandırılması pozitif mekânsal bağımlılığa da neden olacaktır (LeSage ve Pace, 2009: 251). Sonsal beklenti denkleminin paydasındaki integral $Z(\alpha)$ polinomu ile yeniden yazılabilir:

$$p(\alpha|D) \propto (\sum_{i=1}^{2q-1} c_i \alpha^{i-1})^{-(n-k)/2}\pi(\alpha) \tag{79}$$

Denklem 78'nin varyansı aşağıdaki gibi ifade edilir;

$$var(\alpha|D) = \frac{\int [\alpha - \alpha^*]^2 . p(\alpha|D) da}{\int p(\alpha|D) da}$$

Yukarıdaki bu denklemler SAR ile kıyaslanabilir ise,

$$\begin{aligned}
p(\rho|D) &\propto |I_n - \rho W| [(n-k)^{-1} u(\rho)' u(\rho)]^{-(n-k)/2} \pi(\rho) \\
u(\rho) &= (I_n - \rho W)y - X\beta(\rho) \\
\beta(\rho) &= \beta(\alpha) = (X'X)^{-1} X'(I_n - \rho W)y
\end{aligned} \tag{80}$$

Denklem 80'deki SAR' da yer alan ρ 'nun tek deęişkenli sonsal beklentisi ařaęıdaki gibidir (LeSage ve Pace, 2009: 251):

$$E(\rho|D) = \rho^* = \frac{\int_{-\infty}^{+\infty} \rho \cdot p(\rho|D) d\rho}{\int_{-\infty}^{+\infty} p(\rho|D) d\rho} \tag{81}$$

Burada ρ parametresi için $p(\rho|D)$ denkleminde $n \times n$ boyutlu $|I_n - \rho W|$ determinantının ve W aęırlık matrisinin öz deęerlerine ihtiya duyulduęuna dikkat edilmelidir. Bu determinant büyük örneklem verilerinde problemlere neden olmaktadır. MESS modellerinde öz deęerlere ihtiya duyulmamaktadır. Uyum limitlerinin alt ve üst sınırları yeterlidir.

Bayesian MESS modelindeki β 'nin sonsal daęılımı $\beta(\alpha^*)$ için çok deęişkenli t-yoęunluęu kullanılarak sonsal ortalama ařaęıdaki gibi ifade edilir (LeSage ve Pace, 2009: 252);

$$E(\beta|D) = (X'X)^{-1} X'S(\alpha^*)y \tag{82}$$

Bu denklemdeki α^* ; α parametresinin sonsal daęılım ortalamasıdır ve varyans-kovaryans matrisi:

$$var - cov(\beta) = \frac{1}{n-k-2} (\int Z(\alpha) p(\alpha|D) d\alpha) (X'X)^{-1}$$

Burada $p(\alpha|D)$; α parametresinin sonsal beklentisi ve $Z(\alpha)$; skaler bir polinom ifadesidir.

3.4.2. Tek Taraflı Yaklařıklar Yöntemi

Besag (1974) mekânsal modellerin büyük veri tahminleri için tek taraflı yaklařıklar yöntemini önermiştir. Bu yaklařım hesaplama için gerekli alanı, süreyi oldukça

azaltmaktadır ve mekânsal modellerde kullanılan yöntemlere bir alternatiftir. Bu yöntem için i . konumdaki y değişkenine ait mekânsal gecikmeli değer aşağıdaki gibi tanımlanmaktadır

$$L(y_i) = \sum_{j=1}^n w_{ij}^* y_j \quad (83)$$

Bu modeldeki w_{ij}^* ; standartlaştırılmış ağırlık matrisidir. Mekânsal gecikme i . konumuna komşu olan tüm konumlarda gözlemlenen y değerinin ortalaması olarak bilinir. Bu ifade ile tüm komşuların (konumları ne olursa olsun) gecikmeli değişkene aynı miktarda katkı sağladığı için yönün önemli olmadığı vurgulamaktadır (Arbia, 2014: 179). Literatürde bu durum izotropi olarak bilinmektedir. İzotropi hipotezi, birçok mekânsal istatistik dalında tartışılmaktadır. Fizik biliminden türetilen bu kavrama göre bağımlılık yapısı için yöne bağlı bir sapma ve tercih edilmiş bir yön söz konusu değildir. Meteoroloji, jeoloji gibi bilimlerde veya medikal görüntüleme uygulamalarında ise yön büyük önem taşımaktadır. Buda anizotropi olarak bilinmektedir. Örneğin, göğüs kanseri riskinin öngörülmesinde anizotropi yönteminin oldukça iyi olduğu ispatlanmıştır. İzotropi, anizotropinin aksine modelin özelliklerinin yönden tamamen bağımsız olma özelliğidir.

Anizotropinin özellikle yaygın bir tezahürü, mekânsal ilişkilerdeki asimetri olduğu vurgulanmıştır (Arbia, 2014:179). Paelinck ve Klaassen (1979) asimetriyi mekânsal ekonometrinin beş temel özelliğinden biri olarak tanımlamışlardır. Asimetrinin iyi bir örneği merkez-çevre modelidir. Arbia (2014), Paelinck ve Nijkamp (1975) çalışmasını örnek göstererek tersi yönden ziyade merkezden çevreye doğru daha güçlü bir bağımlılık olduğunu belirtmiştir. Kıyı bölgelerinden gözlemlenen pek çok ekonomik değişkende bağımlılık gözlemlenirken, kıyıdan iç bölgeye doğru ise bağımlılığın az olduğu ya da olmadığı belirtilmiştir.

Arbia (2014:180), asimetri ve anizotropi kavramları arasındaki ilişkiyi şöyle bir örnek ile açıklamıştır.

Şekil 4:Düzenli Bir Kafes İçin Ayarlanmış Üç Bölge

	l
m	i

Burada bölünmüş alan ele alınarak i , l ve m mekânsal bölgeleri temsil etmektedir. Şekil 4'teki gibi sadece üç mekânsal bölgeye (i , l ve m olarak adlandırılır) bölünmüş bir çalışma alanı ele alınsın. Sistemin topolojisinin basit bir kale-rook komşuluk yapısı tarafından temsilinin olabileceği varsayılmıştır (Arbia, 2014: 180). Sırasıyla yönler dikey (V) ve yatay (H) yönler boyunca iki farklı bağımlılık yapısı tanımlansın. Kale komşuluk yapısına göre denklem sistemi aşağıdaki gibi yazılabilir;

$$y_i = \vec{\rho}_H y_m + \vec{\rho}_V y_l + \varepsilon_i$$

$$y_l = \vec{\rho}_V y_i + \varepsilon_l$$

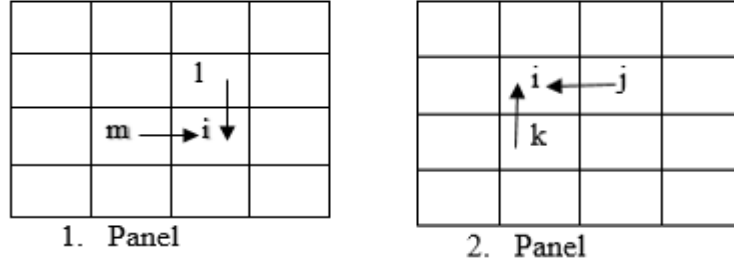
$$y_m = \vec{\rho}_H y_i + \varepsilon_m$$

Bu denklemlerde yatay (ρ_H) ve dikey (ρ_V) yön boyunca mekânsal bağımlılık parametreleri ve bu bağımlılığın yönünü gösteren terim ise üst simge olarak belirtilmektedir. Dolayısıyla i . konumunun yatay olarak soldan sağa m konumuna, dikey olarak yukarıdan aşağıya ise l konumuna bağlıdır. l konumu dikey olarak (aşağıdan yukarıya) i konumuna bağlıdır ve m konumu yatay olarak (sağdan sola) i konumuna bağlıdır. Bu bilgiler ile asimetri $\vec{\rho}_V \neq \vec{\rho}_V$ ve $\vec{\rho}_H \neq \vec{\rho}_H$ olarak gösterilir. Oysa anizotropi simetrisinin varlığında $\vec{\rho}_V = \vec{\rho}_V = \rho_V$ ve $\vec{\rho}_H = \vec{\rho}_H = \rho_H$ olmaktadır. Dikkat edilecek olursa izotropi simetri anlamına gelmektedir. Ancak simetri her zaman izotropi anlamına gelmemektedir (Arbia, 2014:181)

3.4.2.1. Mekânsal Gecikme Modellerinde İzotropi

İzotropi yönden bağımsız olma durumudur. Mekânsal gecikme modelleri için İzotropinin tanımını Arbia (2014: 181)'ya göre şöyle açıklanmıştır. n gözlemlili bir y değişkeni ve satır bazlı standartlaştırılmış W ağırlık matrisi bilgisi olsun. İki farklı birbiri ile örtüşmeyen ağırlık matrisleri W_1 ve W_2 ve bu ağırlık matrisleri ile ilişkin $W_1 + W_2 = W$ tanımlansın. Ayrıca bu iki ağırlık matrisinin iki farklı tercih edilen mekânsal bağımlılık yönünü içerdiği varsayılınsın. Şekil 5'te bir R çalışma alanı 16 mekânsal bölgeye bölünmüş düzenli kare kafeslere sahiptir. Kale komşu yapısına göre mekânsal bağımlılığı ifade etmek için iki farklı ağırlık matrisleri bu bölge üzerinde görüntülenir.

Şekil 5: Düzenli Kare Kafeslerde İki Ayrı Ağırlık Matrisi Görüntüsü



Kaynak: Arbia, 2014: 181

Yukarıdaki şekilde i hücresi W_1 ağırlık matrisindeki $1, m$ hücrelerine ve W_2 ağırlık matrisindeki j, k hücrelerine bağlıdır. 1. panelde W_1 ağırlık matrisi kuzey-batıdan güney-doğuya bağımlılığı ve 2. panel W_2 ağırlık matrisi güney-doğudan kuzey-batıya bağımlılığı ifade etmektedir. Bu paneller ele alınırsa ve model izotropik ise, bu prosedür çıkarımsal olarak tam olabilirlik yaklaşımına eşdeğerdir (Arbia, 2014: 186). Bu yöntemi tanıtmak için aşağıdaki tek taraflı mekânsal gecikme modeli;

$$y = Z\beta + \rho_1 W_1 y + \rho_2 W_2 y + \varepsilon$$

Burada y bağımlı değişkenin gözlemlerinin bir vektörüdür, $Z = [X, WX]$ stokastik olmayan dışsal bir matris regresörü; β tahmin edilecek bilinmeyen parametrelerin bir vektörü ve u stokastik bozulma vektörüdür $u|Z \approx i. i. d. N(0, \sigma^2 I)$. Ayrıca W_1 belirlenen bir yöndeki bağımlılığı içerir ve satır standardizasyonu yapılmamış bir ağırlık matrisidir. İlişkinin tek taraflı olduğu varsayılır ve öyleki $w_{ij} \in W_1$ ise j tercih edilen yönde i 'nin bir komşusu olma durumunda $w_{ij} = 1$ aksi halde $w_{ij} = 0$ 'dır. Bu durumda Şekil 5'te gösterilen üç konumlu şemaya ve tek bir regresör durumunda, tek taraflı modelin olasılığı şu şekilde ifade edilebilir:

$$L(\sigma^2, \rho, \beta) = \prod_{i=1}^n f_{y_i|y_l, y_m}(y_i|y_l, y_m) \quad (84)$$

Burada $l, m \in PN(i)$ 'dir. Denklem 84'ün sağ tarafındaki her bir yoğunluk fonksiyonu komşularına koşullu tanımlanmıştır.

Çoklu normal koşullu dağılımlara özgü standart sonuçlar kullanılarak ve farklı konumlardaki (Z ve y) değişkenlerin kovaryansların sıfır olduğu varsayımı ($Cov(Z_i Y_l) = Cov(Z_i Y_m) = 0$) ile Denklem 84'ün sağ tarafındaki bilgilere dayalı koşullu yoğunluk

fonksiyonu aşağıda tanımlanan koşullu ortalama ve varyans ile $N(\mu_c, \sigma_c^2)$ olacaktır (Arbia, 2014: 186).

$$\mu_s = \mu + \frac{\rho}{(1+\rho')} \left[\sum_{j=1}^n w_{ij} (y_j - \mu) \right] + \frac{\beta}{(1-\rho'^2)} (Z_i - \mu_Z) \quad (85)$$

$$\sigma_s^2 = \sigma^2 - \frac{2\rho^2}{(1-\rho'^2)} - \frac{\beta^2 \sigma_Z^2}{(1-\rho'^2)}$$

Burada, ortalama ve varyanstaki ρ' ; koşullu yoğunluk fonksiyonundaki y_l ve y_m arasındaki korelasyonu göstermektedir. Ayrıca bu korelasyon y 'nin ikinci mertebeden komşuları arasındaki korelasyonu göstermektedir. Denklem 85'teki dağılım ile ilişkili olabilirlik fonksiyonu;

$$\begin{aligned} L(\sigma^2, \rho, \rho', \beta) &= \prod_{i=1}^n f_{y_i|y_l, y_m}(y_i|y_l, y_m) \\ &= \text{const} (\sigma_s^2)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2\sigma_s^2} \left[\sum_{i=1}^n (y_i - \mu_s)^2 \right] \right\} \end{aligned} \quad (86)$$

Denklem 86'nın olasılık fonksiyonunun logaritmik olasılığı:

$$l(\sigma^2, \rho, \rho', \beta) = \text{const} - \frac{n}{2} \ln(\sigma_s^2) - \frac{1}{2\sigma_s^2} \left[\sum_{i=1}^n (y_i - \mu_s)^2 \right] \quad (87)$$

85'inci denklem, 87'deki denklemde yerine yazılır ise:

$$\begin{aligned} l(\sigma^2, \rho, \rho', \beta) &= \text{const} - \frac{n}{2} \ln \left[\sigma^2 - \frac{2\rho^2}{(1-\rho'^2)} - \frac{\beta^2 \sigma_Z^2}{(1-\rho'^2)} \right] \\ &\quad - \frac{1-\rho'^2}{2 \left[\sigma^2(1-\rho'^2) - 2\rho^2(1-\rho') - \beta^2 \sigma_Z^2 \right]} \\ &\times \left[\sum_{i=1}^n \left(y_i - \mu + \frac{\rho}{(1+\rho')} \left[\sum_{j=1}^n w_{ij} (y_j - \mu) \right] + \frac{\beta}{(1-\rho'^2)} (Z_i - \mu_Z) \right)^2 \right] \end{aligned}$$

gibi elde edilir.

Bu koşullu ortalama ve varyansa sahip olan logaritmik olabilirlik denklemi herhangi bir matris formu içermemektedir. Bu nedenle büyük veri boyutlarında doğrudan kolay bir şekilde hesaplanabilmektedir (Abia, 2014:187).

3.4.2.2. Mekânsal Gecikme Modellerinde İzotropinin Test Edilmesi

Büyük veri örneklerinde maksimum olabilirlik yöntemi kovaryans matrislerinin tersine çevrilmesiyle hesaplanmasında çok zorlanılmaktadır. Bu yüzden Besag (1974;212) tarafından literatüre kazandırılan tek taraflı yaklaşımlar kullanılarak hesaplamadaki bu zorluk büyük oranda giderilmektedir.

Mekânsal gecikme modellerinde izotropi bölümünde bahsettiğimiz iki farklı ağırlık matrisi $W_1 + W_2 = W$ olduğu ve bu ağırlık matrislerinin mekânsal bağımlılığın iki farklı yönünü içerdiği belirtilmiştir. İki parametrelili ve iki farklı belirli yönde bağımlılığı kapsayan anizotropik mekânsal gecikmeli model şöyledir;

İki parametrelili ve iki farklı belirli yönde bağımlılığı kapsayan anizotropik mekânsal geçilmeli model şöyledir;

$$\begin{aligned} y &= Z\beta + \rho_1 W_1 y + \rho_2 W_2 y + \varepsilon \\ \varepsilon|Z &\approx i. i. d. N(0, \sigma^2 I) \end{aligned} \quad (88)$$

Burada $Z = [X, WX]$ stokastik olmayan bağımsız değişkenler matrisidir ve bu matris gecikmeli terimler içermektedirler. W_1 belirlenen bir yöndeki bağımlılığı içerir ve satır standardizasyonu yapılmamış bir ağırlık matrisidir. W matrisi izotropi varsayımı altında oluşturulmuş tam(full) ağırlık matrisidir. $W_2 = W - W_1$ 'dir ve belirlenen bağımlılığın ikinci yönünü göstermektedir. $\rho_1 = \rho_2 = \rho$ ise denklem 88, aşağıdaki gibi tek değişkenli izotropik gecikme modeline dönüşmektedir (Arbia,2014:181).

$$\begin{aligned} y &= Z\beta + \rho W y + \varepsilon \\ \varepsilon|Z &\approx i. i. d. N(0, \sigma^2 I) \end{aligned}$$

Anizotropik ve izotropik modelinin bozucu terimleri sırasıyla aşağıdaki gibi ifade edilir;

$$\begin{aligned} \varepsilon^{ANI} &= y - Z\beta - \rho_1 W_1 y - \rho_2 W_2 y \\ \varepsilon^{ISO} &= y - Z\beta - \rho W y \end{aligned} \quad (89)$$

Yukarıdaki bu modeller için logaritmik olabilirlik fonksiyonları sırasıyla şu şekildedir (Arbia,2014:182):

$$\begin{aligned}
 l_{ANI}(\sigma^2, \rho_1, \rho_2, \beta) &= const - \frac{1}{2} \ln |I - \rho_1 W_1 - \rho_2 W_2| \frac{1}{2\sigma^2} \varepsilon^{(ANI)T} \varepsilon^{(ANI)} \\
 l_{ISO}(\sigma^2, \rho, \beta) &= const - \frac{1}{2} \ln |I - \rho W| \frac{1}{2\sigma^2} \varepsilon^{(ISO)T} \varepsilon^{(ISO)}
 \end{aligned} \tag{90}$$

İzotropi testi için olabilirlik oran test istatistiği şöyledir (Arbia vd., 2013, Arbia,2014)

$$A - test = -2[l_{ANI}(\hat{\sigma}^2, \hat{\rho}_1, \hat{\rho}_2, \hat{\beta}) - l_{ISO}(\hat{\sigma}^2, \hat{\rho}, \hat{\beta})] \tag{91}$$

Bu testte $l_{ISO}(\hat{\sigma}^2, \hat{\rho}, \hat{\beta})$ kısıtlı modele (izotropi modelin) ait log-olabilirliğin maksimum değerini ve $l_{ANI}(\hat{\sigma}^2, \hat{\rho}_1, \hat{\rho}_2, \hat{\beta})$ kısıtsız modelin (anizotropi modelin) logaritmik olabilirliğin maksimum değerini göstermektedir. Her iki ifadede bilinmeyen parametreler yerine bu parametrelerin maksimum olabilirlik tahminleri (şapka-hat) kullanılmıştır. İzotropi sıfır hipotezi altında denklem 91'deki test istatistiği χ_1^2 dağılımına sahiptir (Arbia, 2014: 182-183);

$$A - test \xrightarrow{d} \chi_1^2 \tag{92}$$

İzotropinin ihmal edilmesi parametre tahminlerinde yanlılığa ve etkinlik kaybına neden olduğu Monte Carlo simülasyonları ile Arbia vd. (2013) tarafından ortaya koyulmuştur. Ayrıca ağırlık matrisi ile ilgili yanlış belirlemenin parametre tahminlerinin standart hatalarının büyümesine ve dolayısı ile etkinlik kaybına yol açmaktadır (Arbia, 2014: 183)

3.4.3. Kompozit Olabilirlik Yaklaşımı

Kompozit olabilirlik yaklaşımı, her bir alanda tespit edilen olayların birbirinden bağımsız olup olmadığına bakılmaksızın birleştirilmiş, koşullu veya marjinal yoğunluk fonksiyonunun bir parçası şeklinde oluşturulmuş bir fonksiyondur. Psuedo-olabilirlik yaklaşımının alt bir sınıfı olarak kabul edilmektedir. Tamamen belirlenen olabilirliğin bazı özelliklerini kullanmaktadır ancak tamamen belirlenmiş bir olabilirlik değildir.

Literatürde oldukça yaygın olan bu yaklaşım ikili (pairwise) olabilirlik diye de adlandırılmaktadır. Zaman serilerinin yanı sıra panel veri, mekânsal kesit verilerde ve mekânsal kategorik veriler içinde kullanılmaktadır. Bu yaklaşım hesaplama maliyetini azaltmaktadır ve yanlış model kurulmasının bir miktarda olsa önüne geçmektedir (Arbia, 2014: 189).

3.4.3.1. SEM Tahmini İçin İki Değişkenli Marjinal Olabilirlik Yaklaşımı

Mekânsal hata modelleri için hata terimi otoregresif bir formdadır ve aşağıdaki gibidir;

$$y_i = \beta Z_i + \varepsilon_i$$

$$\varepsilon = \lambda W_\varepsilon + u \quad |\lambda| < 1$$

Z açıklayıcı değişkeni iki değişkenlidir ve $Z \sim N(X\beta, \Sigma(\sigma^2, \lambda))$.

Burada otoregresif form yerine her bir i ve l gibi hata çiftleri için birleşik iki değişkenli gauss dağılımı varsayımı aşağıdaki gibidir (Arbia, 2014: 190)

$$\begin{Bmatrix} \varepsilon_i \\ \varepsilon_l \end{Bmatrix} \approx MVN(0, \sigma^2 \Omega) \quad \forall i, l \quad l \in N(i) \quad \Omega = I \quad (93)$$

Denklem 93'ün tüm n adet hata bileşik davranışı için genel bir tanımdır. Basitlik amacıyla 2×2 formatındaki bir ilişki üzerinden açıklama yapılacaktır. Denklem 93'teki $\Omega = \begin{pmatrix} 1 & \Psi \\ \Psi & 1 \end{pmatrix}$ korelasyon matrisidir, $N(i)$; i konumundaki komşular kümesidir. Komşuluk kriterinin bu metod için önemli olmadığı belirtilmiştir. Ψ terimi ise, mekânsal hata korelasyonunu kontrol eden bir parametredir ve sıradan mekânsal hata modelindeki ρ terimi yerine tercih edilmiştir. Böylece bu iki terim arasında farklılık olduğu anlamına gelmektedir.

Konum i komşuluğunda rastgele başka bir konum seti q olsun (Arbia, 2014:191). Hata terimi ε_j ve ε_k olduğu, $k \in N(j)$ varsayılmaktadır ve $l(\varepsilon_i$ ve $\varepsilon_l)$ birbirlerine yakınlıkları nedeniyle mekânsal olarak bağımlı oldukları kabul edilir. O halde bu ortak komşuluğu $N(i, l) = \{N(i) \cup N(l)\}$ şeklinde düzenli bir kafes için tanımlanmaktadır. Düzensiz bir kafes yapısı için bu bağımsızlık varsayımının kısıtlı olduğu varsayılır ve esnetilir.

İki deęişkenli kodlama yapısındaki herhangi bir q hata çiftine ait birleşik yoğunluk fonksiyonu aşağıdaki gibi yazılabilir (Arbia, 2014: 191);

$$f_{\varepsilon_i \varepsilon_l}(\varepsilon_i \varepsilon_l) = \frac{1}{2\pi\sigma^2\sqrt{1-\psi^2}} \exp\left\{-\frac{1}{2\sigma^2(1-\psi^2)}[\varepsilon_i^2 - 2\psi\varepsilon_i\varepsilon_l + \varepsilon_l^2]\right\} \quad (94)$$

Burada $l \in N(i), i = 1, \dots, q$ 'dur. Kompozit olabilirlik, q 'nun iki deęişkenli yoğunluk fonksiyonlarının çarpımından elde edilir (Arbia,2014: 192)

$$\begin{aligned} L(\beta, \sigma^2, \psi) &= \prod_{i=1}^q f_{\varepsilon_i \varepsilon_l}(\varepsilon_i \varepsilon_l) \\ &= \prod_{i=1}^q \frac{1}{2\pi\sigma^2\sqrt{1-\psi^2}} \exp\left\{-\frac{1}{2\sigma^2(1-\psi^2)}[\varepsilon_i^2 - 2\psi\varepsilon_i\varepsilon_l + \varepsilon_l^2]\right\} \\ &= (2\pi)^{-q}(\sigma^2)^{-q}(1-\psi^2)^{-\frac{q}{2}} \exp\left\{-\frac{1}{2\sigma^2(1-\psi^2)}\sum_{i=1}^q[\varepsilon_i^2 - 2\psi\varepsilon_i\varepsilon_l + \varepsilon_l^2]\right\} \end{aligned} \quad (95)$$

Denklem 95'e uygun olarak log-olasılığı aşağıdaki gibi ifade edilir;

$$l((\beta, \sigma^2, \psi) = \text{const} - q \ln(\sigma^2) - \frac{q}{2} \ln(1 - \psi^2) - \frac{1}{2\sigma^2(1-\psi^2)} \sum_{i=1}^q [\varepsilon_i^2 - 2\psi\varepsilon_i\varepsilon_l + \varepsilon_l^2]$$

Gerekli düzenlemeler yapılarak bazı semboller aşağıdaki gibi tanımlanır:

$$\begin{aligned} a_0 &= \sum_{i=1}^q Z_i^2 + \sum_{l=1}^q Z_l^2 = \sum_{j=1}^{2q} Z_j^2 \\ a_1 &= \sum_{i=1}^q y_i^2 + \sum_{l=1}^q y_l^2 = \sum_{j=1}^{2q} y_j^2 \\ a_2 &= \sum_{i=1}^q Z_i y_i + \sum_{l=1}^q Z_l y_l = \sum_{j=1}^{2q} Z_j y_j \\ a_3 &= \sum_{i=1}^q Z_i y_l + \sum_{l=1}^q Z_l y_i \\ a_4 &= \sum_{i=1}^q Z_i Z_l \\ a_5 &= \sum_{i=1}^q y_i y_l \end{aligned} \quad (96)$$

Arbia (2014) denklem 96'daki sembolleri ve log-olasılığını kullanarak $\hat{\beta}_{BML}$, $\hat{\sigma}_{BML}^2$ ve $\hat{\Psi}_{BML}$ marjinal maksimum olabilirlik tahminçileri aşağıdaki gibi elde edilir;

$$\hat{\beta}_{BML} = \frac{a_2 - \hat{\Psi}_{BML} a_3}{a_0 - 2\hat{\Psi}_{BML} a_4}$$

$$\begin{aligned}
\hat{\sigma}_{BML}^2 &= \frac{\sum_{i=1}^q (\varepsilon_i^2 - 2\psi \varepsilon_i \varepsilon_l + \varepsilon_l^2)}{2q(1-\hat{\psi}_{BML}^2)} \\
&= \frac{a_1 + \hat{\beta}_{BML}^2 a_0 - 2\hat{\beta}_{BML} a_2 - 2\hat{\psi}_{BML} a_5 - 2\hat{\psi}_{BML} \hat{\beta}_{BML}^2 a_4 + 2\hat{\psi}_{BML} \hat{\beta}_{BML} a_3}{2q(1-\hat{\psi}_{BML}^2)} \\
\hat{\psi}_{BML} &= \frac{\sum_{i=1}^q \varepsilon_i \varepsilon_l}{q\hat{\sigma}_{BML}^2} = \frac{a_5 - \hat{\beta}_{BML} a_3 + \hat{\beta}_{BML}^2 a_4}{q\hat{\sigma}_{BML}^2}
\end{aligned} \tag{97}$$

Denklem 97’de dikkat edilmesi gereken nokta $\psi = 0$ olduğunda ortaya çıkmaktadır. O halde $\hat{\beta}_{BML}$ ve $\hat{\sigma}_{BML}^2$ yine kendilerine eşitlenmektedir ve bu hali ile maksimum olabilirlik tahmin edicilerine eşittir.

$$\begin{aligned}
\hat{\beta}_{BML} &= \frac{a_2 - \hat{\psi}_{BML} a_3}{a_0 - 2\hat{\psi}_{BML} a_4} = \frac{a_2}{a_0} = \frac{\sum_{j=1}^{2q} x_j y_j}{\sum_{j=1}^{2q} x_j^2} = \hat{\beta}_{ML} \\
\hat{\sigma}_{BML}^2 &= \frac{1}{2q} \sum_{i=1}^q [\varepsilon_i^2 + \varepsilon_l^2] = \frac{1}{2q} \sum_{j=1}^{2q} \varepsilon_j^2 = \hat{\sigma}_{ML}^2
\end{aligned}$$

$\hat{\beta}_{BML}$ tahmincisinde a_2 terimi Z ve y değişkeni arasındaki kovaryansı, $\hat{\psi}_{BML} a_3$ terimi z değişkenine komşu y değişkeni üzerindeki mekânsal yayılmayı, a_0 terimi bağımsız değişkenleri, payda kısmındaki $-2\hat{\psi}_{BML} a_4$ kısmı varyansı temsil etmektedir. a_4 terimi ise z değişkenindeki otokovaryansı temsil etmektedir.

4. AMPİRİK ANALİZ

Geçmiş insanlık kadar eski olan suç olgusu ekonomi, politika, hukuk, dini inanç ve gelenek gibi birçok sosyal kurumla ilişkilidir ve bu nedenle çok boyutlu bir yapıya sahiptir. Bu özelliğinden dolayı, suçla ve suçlularla etkin bir şekilde mücadele etmek ve suç işleme oranını en aza indirmek için değişik alanlardan uzmanların suça zemin hazırlayan sosyal etkenleri araştırmak, suçlu davranışının psikolojik ve tıbbi yönlerini belirlemek, suçu engellemek ve mağdurları korumak için gerekli hukuki altyapıyı oluşturmak üzere işbirliği yapmaları gerekmektedir.

Suçun çok yönlü bir olgu olması, suçla ilgili çok geniş bir teorik ve ampirik literatürün doğmasına neden olmuştur. Konu ile ilgili ampirik literatürde yer alan çok sayıda çalışma, suç verilerindeki mekânsal bağımlılığı hesaba katmayan, suç analizinde uzamsal olmayan regresyon modellerini kullanmıştır. Bununla birlikte son zamanlarda, araştırmaların ilgisi kriminologlar, sosyologlar, coğrafyacılara ve kolluk kuvvetleri için büyük önem taşıyan suçların coğrafi dağılımı ve belirleyici faktörlerine doğru genişlemiştir (Adeyemi vd., 2021).

Suç teorisine mekânsal veya ekolojik yaklaşımların geçmişi on dokuzuncu yüzyılın başlarına kadar uzanmaktadır. Bununla birlikte, suçun mekânsal kalıpları üzerine yürütülen modern çalışmalar Shaw (1929) ve Shaw ve Mackay'ın (1942) çalışmaları gibi daha spesifik köklere sahiptir. Bu teorik çerçeveyi kullanan araştırmaların çıkış noktası, insan davranışının kendi mekânı içinde ele alınması gerektiği anlayışıdır. Buna göre, suçun yeri (en azından) analiz edilen suçun boyutlarından biri olmalıdır. Esasen bu yaklaşım, suçun bir coğrafyası olduğunu kabul etmektedir (Andresen, 2006a).

Literatürde, suç davranışının coğrafi dağılımını analiz etmek için “sosyal düzensizlik teorisi” ve “rutin faaliyet teorisi” sıklıkla kullanılmıştır. Bu teoriler, sırasıyla, suç oranlarının mekânsal modelinin temel unsurları olarak sosyal yoksunluk ve rutin faaliyetler konularına odaklanmaktadır. Sosyal düzensizliğin etkisini yakalamak üzere sosyoekonomik yoksunluk, ailevi parçalanma, nüfus hareketi, etnik çeşitlilik gibi temel göstergelere yer verilmektedir. Komşuluk düzeyinde tanımlanan sosyal düzensizlik teorisinden farklı olarak, rutin faaliyet teorisi bağlamında yer verilen değişkenler daha ziyade suçun mağdurlarına ilişkin verilere dayanmaktadır. Bu iki suç teorisinin mekanla ilişkili doğası dikkate alındığında, suçla ilgili ampirik araştırmalarda her iki yaklaşımın bütünleştirilmeye çalışıldığını görmek şaşırtıcı olmamaktadır (Andresen, 2006b).

Andresen ve Malleson (2013) tarafından işaret edildiği gibi analizde kullanılan verilerin düzeyi elde edilen bulguların niteliği açısından önem arz etmektedir. Suçların genel olarak ele alınması veya türlerine göre ayrıştırılması (cinayet, hırsızlık, gasp, yaralama vb.), mekânsal düzey (ülke, kent, mahalle, ilçe vb.), zaman boyutu (yıl, ay, hafta, gün vb.) ilişkinin yönü, büyüklüğü ve hatta anlamlılığı üzerinde etkili olabilmektedir. Bununla birlikte, istatistiksel analizlerin genelinde olduğu gibi, bulguların güvenilirliği açısından analizde kullanılan verinin büyüklüğünün de önemli olduğu söylenebilir.

Mekânsal büyük veri ile ilgili uygulandığı çalışmaları, son yıllarda yapılan araştırmalarda çok değişkenli istatistiksel analizde ve suç oranlarındaki değişimi açıklamak için oldukça önemlidir (Baller, vd. , 2001: 562). Bu yüzden suç oranlarındaki meydana gelen artış ile suçu etkileyen faktörlerin belirlenmesi önemli hale gelmiştir. Çalışmalarda sosyo-ekonomik ve demografik değişkenlerin bölgesel olarak cinayet oranları üzerindeki etkisi modellenmeye çalışılmıştır. Bu çalışmalarda birkaç tutarsızlık ortaya çıkabilmektedir. Land vd. (1990), literatürdeki bu tutarsızlıkları çoklu bağlantı sorunu ile aşılabileceğini belirtmiştir. Land ve arkadaşlarının çalışmasından itibaren özellikle kriminologlar tarafından cinayet ve suç oranlarının mekânsal dağılımına ilgi gittikçe yaygınlaşmıştır. İstatistiksel olarak modellerde mekânsal süreçler dikkate alınmadığında hatalı sonuçlara neden olmaktadır ve bağımsız değişkenler etkin değildir ve yanlış sonuçlar vermektedir. Mekânsal büyük veri ile ilgili bu çalışmada cinayet oranların araştırılması ile ilgili veriler kullanılacaktır. Mekânsal büyük veri model tahmincilerinden yalnızca MESS yaklaşımı ile suçun ortaya çıkışında etkili olan faktörler araştırılacaktır. Çünkü MESS'e ait uygulama kodlarının bulunması nedeniyle uygulama çalışması bu yöntem içindir.

İşlenen suçların çeşidine göre suç oranlarını etkileyen faktörler değişmektedir. Literatürde suç oranlarını etkileyen faktörler hakkında birçok bilgi mevcuttur.

Özgen ve Bindak (2017) çalışmasında, 178 ülkedeki cinayet oranlarını araştırmış ve araştırma sonucuna göre bazı önerilerde bulunmuşlardır. Bu araştırmayı gerçekleştirmek için verileri dünya bankası ve The World Factbook raporundan temin edilmiştir. Analiz için çoklu lineer regresyon modelinden yararlanılmıştır. Elde edilen bulgulara göre yaş ortalaması, işsizlik ve gayri safi milli hasılanın yüksek olduğu ülkelerde cinayet eylemlerinin az olduğu bulgularına rastlamışlardır.

Costantini, Meco ve Paradiso (2017) 1978-2013 döneminde Amerika Birleşik Devletleri eyalet düzeyinde bireyleri suça yönelten faktörlerden birinin işsizlik oranı olduğu ve suç oranını arttırıcı bir etkiye sahip olduğu vurgulanmıştır.

Altindag (2012), Avrupa’da yer alan 33 ülkeye ait 1995 ve 2003 dönemindeki panel veri seti kullanılmıştır. Bu çalışmada 100.000 kişi için meydana gelmiş suç oranı araştırılmış ve bu araştırmada EKK ve 2SLS yöntemleri kullanılmıştır. İşsizlik ve suç oranı arasındaki içsellik probleminden dolayı araç değişken yöntemi kullanılmıştır. Elde edilen sonuçlara göre işsizlik oranı ile işlenen suç arasında pozitif ilişki olduğu bulgusuna erişilmiştir.

Enamorado vd. (2015), Meksikada uyuşturucu savaşında gelir eşitsizliğinin suç oranları üzerindeki etkisini araştırmıştır. Analizde, 2.000'den fazla belediye ve 20 yıllık döneme ait bir veriler kullanılarak analiz yapılmıştır. Gelir eşitsizliğinin suç üzerindeki etkisini ortaya çıkarmak amacıyla Gini katsayısı araç değişken olarak kullanılmıştır. Analiz sonuçlarına göre gelir eşitsizliği ile suç oranları arasında pozitif ilişki olduğu belirtilmiştir.

Cömertler ve Kar (2007) ise, Türkiye’nin 81 ilini kapsayan 2000 yılına ait veriler kullanarak işsizlik ve suç arasındaki ilişkiyi açıklamaya çalışmışlardır. Analiz sonucunda işsizlik ile suç oranı arasında pozitif bir ilişkinin bulgusuna rastlanmıştır.

Bell vd. (2018), Amerika Birleşik Devletleri'nde 1980 ve 2010 dönemleri arasında yürürlüğe giren eğitimdeki değişikliklerde okuldan ayrılma yaşının suç oranlarını nasıl etkilediği incelemiştir. Eğitim seviyesi yükseldikçe suç oranlarında düşüş meydana geldiği görülmüştür.

Bianchi vd. (2012), 1990-2003 döneminde polis idari kayıtlarına dayanarak İtalyan eyaletlerinde göç ve suç arasındaki ilişki incelemiş, Göç ve suç oranları arasında içsellik problemi ile karşılaşmış ve araç değişken yöntemine başvurulmuştur. Analiz sonucuna göre göçmen nüfus ile suç arasında pozitif ilişki olduğu belirtilmiştir.

Mekânsal büyük veri modeli uygulama çalışması için gözlem sayısının büyüklüğü önem arz etmektedir. Bu nedenle <https://geodacenter.github.io/data-and-lab/> adresindeki veriler kullanılmıştır. Bu veriler Amerika Birleşik Devletlerinin 3085 sayıda ilçeye ait cinayet oranları verisidir. Büyük veri olarak gözlem sayısı fazla olan bir anakütle tercih edilmiştir. Bu verilerden yararlanılarak komşu yerleşimlerdeki cinayet oranlarının belirli bir ilçedeki cinayet oranlarına nasıl etki ettiği ve cinayet oranına etki eden diğer faktörler

araştırılmıştır. Komşu yerleşimlerdeki cinayet oranının benzerlik, bilgi yayılması ve taklid gibi nedenlere dayalı etkisi mekânsal gecikmeli model ile araştırılabilir. Ayrıca MESS modelinin yalnızca SAR versiyonuna ait kod bulunmaktadır. Bu nedenle de SAR modeli tercih edilmiştir. Uygulama amacıyla kullanılan veriler 1990-2000 yıllarına ait ortalama verilerdir ve ilçe düzeyindedirler. Analizde kullanılan değişkenler Tablo 3'te gösterilmiştir.

Tablo 3: Değişken Tanımları

HR90	100.000 Kişi başına düşen cinayet oranı (murder rate percapita)
RD	Kaynak yoksunluğu (lack of resources)
PO	1990 yılı son on yıllık ilçe nüfusu (county population in the last ten years)
PS	Nüfus yapısı (population structure)
UE	İşsizlik oranı (unemployment rate)
DV	Boşanma oranı (divorce rate)
FH	Hane reisi kadın olanların yüzdesi (percentage of female heads of household)
MA	Medyan yaş ortalaması (median mean age)

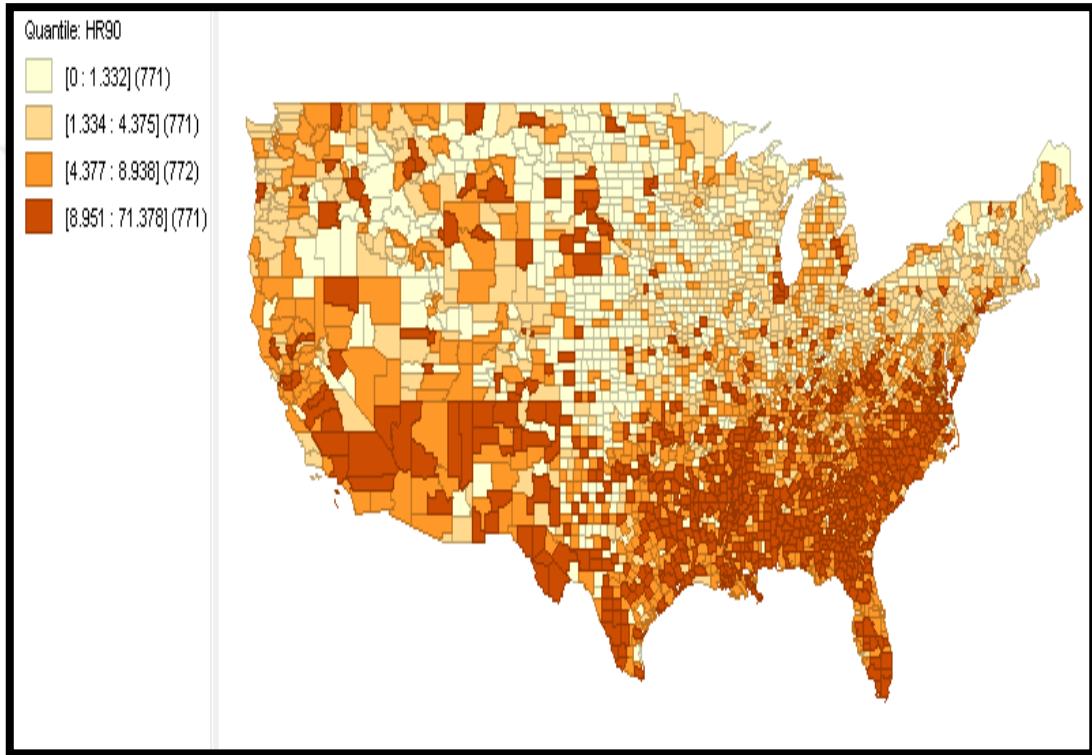
Buradaki değişkenler cinayet oranlarına etki eden en önemli sosyo-ekonomik ve demografik değişkenlerdir. Bu değişkenlerden kaynak yoksunluğu, işsizlik oranı, boşanma oranı, ilçe nüfusu, hane reisinin kadın olanların oranı suç oranları pozitif yönde etkilemesi beklenmektedir. Baller vd.(2001) analizinde,1960-1990 dönemindeki ABD'nin on yıllık dönemler halinde yılları dört zaman diliminde veri kaynaklarını ve yöntemlerini tanımladıktan sonra ESDA tekniği ile analizini gerçekleştirmiştir. Analizlerinde bağımsız değişkenler olarak kaynak yoksunluğu, nüfus yapısı, işsizlik oranı, boşanma oranı ve nüfus sayısını içeren bu değişkenler ile gerçekleştirilmiştir. Analiz sonucuna göre cinayet oranları mekânsal olarak rastgele dağılmamış ve pozitif mekânsal otokorelasyon sergilemektedir.

Yukarıda bahsedildiği gibi mekânsal bağımlılık cinayet oranını etkileyen önemli bir etmendir. Gözlem sayısının büyüklüğü nedeniyle mekânsal büyük verili SAR modeli tahmin edilecektir. Bu amaçla GeoDa programı ve R programı kullanılmıştır. GeoDa

programı verileri haritalandırarak görselleştirilmesini sağlamaktadır. Şekil 5'te GeoDa programında cinayet oranlarının mekânsal dağılımı görselleştirilmiştir.

Mekânsal veri analizinde harita önemli bir role sahiptir. Çünkü harita mekânsal dağılımı görselleştirme imkânı sağlamaktadır. Haritalandırma işlemi sırasında veri kümesini dört eşit parçaya bölen kartil haritası kullanılmıştır. Cinayet oranları için kartil haritası Şekil 5'te gösterilmiştir.

Şekil 6: ABD'de Yaşanan Cinayet Oranlarının İlçelere Göre Mekânsal Dağılımı



Şekil 5'te ABD ilçelerindeki cinayet oranları haritasında en koyu renklere sahip yerler cinayet oranlarının en fazla olduğu yerler olarak gösterilmiştir. Renkler açıldıkça cinayet oranları da azalmaktadır. En açık renkler ise cinayet oranlarının en az olduğu yerler olarak gösterilmiştir. Haritada da görüleceği üzere cinayet oranlarının en fazla yaşandığı yerler Güney bölgesindeki ilçeler olarak görülmektedir. Aynı renkte bölgelerin bir arada kümelenmesi nedeniyle görsel olarak mekânsal bağımlılığın varlığından söz edilebilmektedir.

Global otokorelasyon testleri için komşuluğun etkisi vezir (queen) ağırlık yöntemine göre ölçülmüştür. Ağırlık matrisine satır standartlaştırma işlemi uygulanmıştır ve 3085 adet ilçe olması nedeniyle ağırlık matrisi ($W_{3085 \times 3085}$)'dir.

Öncelikle mekânsal ilişkinin varlığını test eden global mekânsal otokorelasyon istatistikleri hesaplanmıştır. Global mekânsal otokorelasyon testlerinden Moran I ve Geary C testleri için elde edilen bulgular Tablo 4’te gösterilmiştir.

Tablo 4: Global Otokorelasyon Test Sonuçları

		Standart sapma	Varyans	P – değeri
Moran I	0.3833	35.703	0.000115	2.2e-16
istatistiği				
Geary c	0.6014	29.631	0.000180	2.2e-16
istatistiği				

Global Moran I ve Global Geary C istatistikleri için normallik varsayımı altında hesaplanan p değeri %1 düzeyinde istatistiksel olarak anlamlıdır. O halde mekânsal otokorelasyonun varlığından söz etmek mümkündür.

Mekânsal ilişkinin haricinde cinayet oranlarını etkileyen etmenlere ait ekonometrik model için Tablo 3’de verilen değişkenlere göre oluşturulmuştur ve şöyledir.

$$HR90 = \beta_0 + \beta_1 RD + \beta_2 PO + \beta_3 PS + \beta_4 UE + \beta_5 DV + \beta_6 MA + \beta_7 FH + \varepsilon_i \quad (98)$$

Oluşturulan bu ekonometrik modelde bağımlı değişken HR90’dır ve kalan diğer değişkenler ise bağımsız değişkenlerdir. Komşuluk kriterlerinden yine vezir (queen) kriterine göre mekânsal regresyon modelleri tahmin edilmiştir. Mekânsal hata modeli (SEM) ve mekânsal gecikme modeline (SAR) ve mekânsal kombo (SAC) modeli için elde edilen tahmin sonuçları Tablo 5’tedir.

Tablo 5: Mekânsal Ekonometrik Modellerin Tahmin Sonuçları

	SAR	SEM	SAC (SEM+SAR)
Sabit	-2.385* (0.05251)	-1.2082 (0.3565)	-2.1520** (0.07458)
RD	1.9144* (2.2e-16)	2.4116* (2.2e-16)	1.9269* (2.2e-16)
PO	2.3196e-06* (1.005e-11)	2.1221e-06* (3.463e-09)	2.3106e-06* (6.655e-12)
PS	3.3851e-02 (0.81422)	8.7771e-02 (0.5889)	5.0286e-02 (0.72071)

UE	-2.0290e-01* (6.791e-07)	-2.1156e-01* (1.601e-06)	-2.1187e-01* (1.265e-07)
DV	3.8772e-01* (1.944e-12)	4.1218e-01* (6.2145e-11)	3.8735e-01* (5.684e-13)
MA	8.0301e-03 (0.77077)	7.867e-03 (0.7953)	4.2775e-03 (0.87424)
FH	3.8109e-01* (2.2e-16)	4.2374e-01* (2.2e-16)	3.7443e-01* (2.2e-16)
Rho	2.9363e-01 (2.2e-16)		3.0077e-01 (2.2e-16)
Lambda		0.26378	-0.0492

Kaynak: Tablodaki *, **, *** ile ifade edilen değerler sırasıyla %1, %5, %10 düzeyde anlamlığı göstermektedir.

Tablo 5'te verilen hücre içindeki değerler katsayıların tahmin değerleri, parantez içi değerler ise t istatistikleridir.

SAR'ın tahmin sonuçlarına göre, mekânsal otoregresif katsayısı istatistiksel olarak anlamlı ve pozitifdir ($\rho = 0.293$). Bu sonuç, birbirlerine komşu olan ilçelerdeki cinayet oranları arasında mekânsal bağımlılığın olduğunu göstermektedir. Komşu ilçelerde cinayet oranı değişkenine ait bu katsayının pozitif olması tüm ilçelerin suç oranı artışının, ilçelerde suç oranlarını arttıracaklarını göstermektedir. Benzer olarak SAC'ın tahmin sonuçlarına göre Rho değeri de anlamlı ve pozitifdir.

İlgili konumun hata terimi ile komşu konumların hata terimleri arasındaki mekânsal bağımlılığın derecesini ölçen Lambda değeri SEM'in tahmin sonuçlarına göre 0.263 ve SAC'ın sonucuna göre ise -0.049'dır.

SAR modelindeki diğer değişkenlere ait katsayı tahminleri cinayet oranlarının belirleyicilerini yansıtmaktadır. Kaynak yoksunluğu (RD), nüfus miktarındaki artış (PO), işsizlik oranı (UE), boşanma oranı (DV), hane reisi kadın olanların yüzdesi (FH) olan bu değişkenler cinayet oranını etkilemektedir.

İyi koşullar içermeyen yerlerde yaşan kişilerin hukuk dışı eylemlere yöneldiklerini gözlemlemiştirler (Cömertler ve Kar, 2007: 45). Analizde kaynak yoksunluğunun (RD) işareti pozitifdir. Böylece kaynak yoksunluğu arttıkça suç oranının arttığı bulgusuna varılmıştır. Bu sonuç teorik beklentiyi karşılamıştır. SEM ve SAC modellerinde de kaynak yoksunluğunun işareti pozitifdir.

İlçelerdeki nüfus miktarındaki artış (PO) cinayet oranlarında da SAR, SEM ve SAC modelinde pozitif etkiye neden olmaktadır. Nüfus miktarı arttıkça cinayet oranının da arttığı gözlemlenmektedir. Bir ülkede nüfus artış hızı ne kadar yüksek olursa kalkınma hızı o derece azalır. Nüfus artış hızının sorun haline gelmesi, o ülke insanların temel ihtiyaçlarının karşılanamaması ile ortaya çıkar. Ehrlich (1976) dünyada yaşanan nüfus patlaması ve bunun sonucunda felaketlerinde artmakta olduğunu vurgulamıştır. Nüfus artışı, fakirlik ile yakından ilişkilidir (Çamurcu, 2005: 93). Örneğin, nüfus arttıkça tüketim artar, toprak daha yoğun kullanılmaya başlar. Toprak yoğun kullanılması sonucu nadas dönemleri kısalmaya başlar ve toprağın verimliliğinde düşmeye sebebiyet verir. Bunun sonucunda çevresel bozulmalar oluşur ve ürün verimliliğinde düşüş meydana gelir. Bu durum da kişi başına düşen gelirin azalmasına ve fakirliğin artmasına sebebiyet vermektedir. Fakirliğin artışı da kişilerin suça eğilimini arttırmaktadır.

İşsizlik oranı (UE) ile cinayet oranı arasında negatif bir ilişki görülmektedir çünkü işsizlik oranının işareti negatif çıkmıştır. İşsizlik oranındaki %1'lik artış, cinayet oranlarında %20 azalışa neden olmaktadır. SAR'a alternatif olan SEM ve SAC modellerinde ise bu değerler %21 azalışa neden olmaktadır. Emek piyasasının suç üzerinde önemli etkisi olduğu düşünülmektedir ve emek piyasasında, fırsatların oldukça yetersiz olması durumunda, kişiler suçlara yönelebilmektedirler (Ehrlich, 1973: 521). Teorik olarak işsizlik oranındaki artışın kişileri suça yöneltmesi beklenmektedir. Buna göre elde edilen sonuç beklentilerle uyuşmamaktadır.

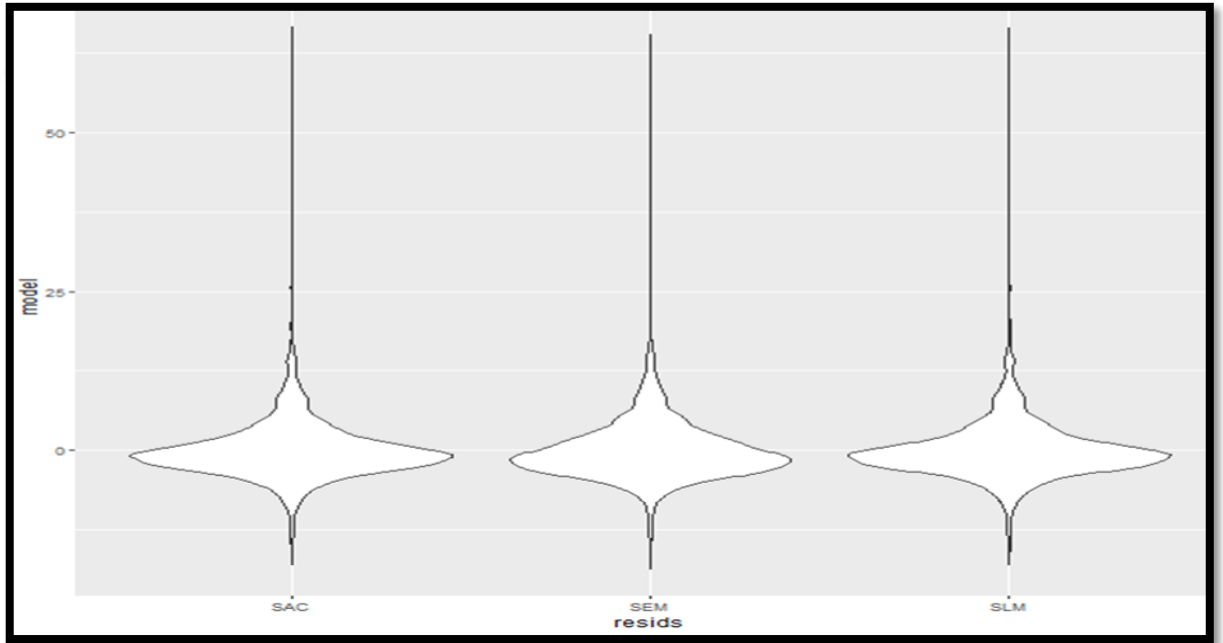
ABD ilçelerinde hane reisi kadın olanların oranı (FH) ile cinayet oranları arasında pozitif ilişki olduğu bulgusuna erişilmiştir. Hane reisi kadın olanların oranındaki %1'lik bir artış cinayet oranlarını %38 arttırmaktadır. Bu modele alternatif olan SEM ve SAC modellerinde sırasıyla cinayet oranlarını %42 ve %37 arttırmaktadır.

1990-2000 yılları arasında gerçekleşen boşanma oranlarının (DV) üç modelde de işareti pozitifdir ve boşanma oranlarındaki artış cinayet oranlarını da arttırmaktadır. Bu sonuç beklentimiz doğrultusunda sonuçlanmıştır. Çetin (2014: 53) çalışmasında, kadının boşanmak istemesi, boşanma sonucu bireyselleşme ile kadın ve erkek arasındaki çatışmanın şiddetlenmesi neticesinde cinayetlerin arttığını vurgulamıştır.

Tablo 5'te tahmin sonuçları verilen bu üç mekânsal modele ait kalıntıların normal dağılıma sahip olup olmadığı, ortalamasının sifıra yakın olup olmadığı keman grafiği üzerinden incelenmiştir. Keman grafiği verilerin dağılımı ve olasılık yoğunluğunu

görselleştirmek için yaygın bir şekilde kullanılmaktadır. Keman grafiği, kutu grafiği (box plot) ve kernel yoğunluk tahmini (Kernel Density Estimation -KDE)'nin kombinasyonundan doğmuştur. Kutu grafiği, görsel basitliği ve verilerdeki değerlerin nasıl dağıldığına ilişkin önemli ayrıntıları gizleme eğiliminde olduğundan, değişken değerlerinin dağılımının gösterilmesinde sınırları bulunmaktadır. KDE, verilerden tahmin edilen bir histogramın doğrusal anlamda toparlanmış halidir. KDE, verilerden tahmin edilen bir histogramın doğrusal anlamda toparlanmış halidir. Bu keman grafiği Şekil 6'da gösterilmektedir.

Şekil 7: Mekânsal Ekonometrik Modellerin Keman Grafiği



Şekil 6'daki keman grafiğinde modellere ait kalıntıların sıfır ortalama etrafında yayıldıkları görülmektedir, böylece üç mekânsal modelin sıfır ortalamaya sahip normal dağılımlı olduğu gösterilmektedir. Bu grafiğe göre modellerin varsayımlardan biri olan normallik koşulunu sağladığı görülmektedir.

Tablo 5'teki tahmin sonuçlarına göre bir ilçedeki cinayet oranı (HR90) sadece o ilçedeki cinayet oranı ile ilişkili değil aynı zamanda diğer ilçelerle de etkileşim içinde olduğu sonucuna varılmıştır. Ampirik analiz amacıyla kullanılan örnekleme gözlem sayısının fazla olması nedeniyle MESS yaklaşımı ile tahmin edilmesi gerekmektedir. MESS modeli özellikle büyük boyutlu mekânsal verilerin analizinde kullanılmaktadır. Büyük boyutlu verilerin ağırlık matrisleri de büyük olacağından mekânsal modellerin

tahmini zorlaşmaktadır. Eşitlik 98'deki mekânsal ekonometrik modelin MESS yöntemi ile elde edilen tahmin sonuçları Tabloda 6'da gösterilmiştir;

Tablo 6: MESS Yöntemi Katsayı Tahmin Sonuçları

	MESS
Sabit	-1.8118e+00 * (0.1318)
RD	2.1051e+00 * (2.2e-16)
PO	2.3029e-06 * (2.371e-11)
PS	7.9768e-02 (0.5779)
UE	-2.2673e-01* (9.983e-09)
DV	4.0387e-01* (2.902e-13)
MA	1.5301e-03 (0.9557)
FH	3.8942e-01* (2.2e-16)
Alpha	-0.25786* (2.22e-16)
Implied rho	0.2272962
LR test değeri	134.27 (2.22e-16)

Tablo 6'daki MESS için katsayı tahminlerinden kaynak yoksunluğu (RD), nüfus miktarı (PO), işsizlik oranı (UE) ve boşanma oranı(DV) ve hane reisi kadın olanların oranı (FH) istatistiksel olarak anlamlı çıkmıştır. Ancak medyan yaş ortalaması (MA) ve nüfus yapısı (PS) değişkenlerinin katsayıları MESS yönteminde anlamsız çıkmıştır.

MESS modeli sonucunda değişen ρ değeri için α değerleri hesaplanmıştır. Bu değer $\alpha = \ln(1 - \rho)$ ilişkisinden elde edilmektedir. Alpha($\alpha = -0.257$) katsayısı %1 düzeyinde istatistiksel olarak anlamlıdır.

MESS yöntemine ait katsayı tahmini sonucunda ima edilen ρ (implied rho = alphaya göre elde edilen rho) değeri 0.227'dir.

Olabilirlik oran testi (LR) değeri 134.27, % 1 anlamlılık düzeyine göre istatistiksel olarak anlamlıdır. Bilindiği üzere bu test, mekânsal model ile sıradan doğrusal regresyon modelinin hangisinin veri için uygun olup olmadığını test etmektedir. İstatistiksel olarak anlamlı olan bu sonuca göre mekânsal model daha uygundur.

Aşağıdaki Tablo 7’de SAR ve MESS yöntemleri ile elde edilen katsayı tahminleri verilmiştir.

Tablo 7: MESS Ve SAR Yöntemi Katsayıları

	SAR	MESS
Sabit	-2.385* (0.05251)	-1.8118e+00 * (0.1318)
RD	1.9144* (2.2e-16)	2.1051e+00 * (2.2e-16)
PO	2.3196e-06* (1.005e-11)	2.3029e-06 * (2.371e-11)
PS	3.3851e-02 (0.81422)	7.9768e-02 (0.5779)
UE	-2.0290e-01* (6.791e-07)	-2.2673e-01* (9.983e-09)
DV	3.8772e-01* (1.944e-12)	4.0387e-01* (2.902e-13)
MA	8.0301e-03 (0.77077)	1.5301e-03 (0.9557)
FH	3.8109e-01* (2.2e-16)	3.8942e-01* (2.2e-16)

Tablo 7’ye bakılacak olursa SAR ve MESS modeli tahmin sonuçları incelendiğinde, bu iki modelin katsayı tahminlerinin ve bu tahminlere karşılık gelen z istatistiklerinin birbirine genel olarak yakın değerlere sahip olduğu görülmektedir. LeSage ve Pace (2009; 246-247) MESS modelini diğer mekânsal modeller ile karşılaştırmak amacıyla Monte Carlo simülasyon çalışması yapmıştır. Yapılan 100 denemede değişen ρ değerleri için SAR ve değişen α değerleri için MESS modeli tahmin edilerek katsayı tahminlerinin ortalaması elde edilmiştir. SAR ve MESS modelinin katsayı tahminlerinin neredeyse aynı sonuçlar ürettiği gösterilmiştir. Genelde olduğu gibi bu analizde de MESS modeli SAR’a benzer sonuçlar üretmiştir. Buna göre MESS

yöntemi büyük veri setleri için mekânsal modellerin daha basit bir şekilde tahmin edilmesini sağlamaktadır.



5. SONUÇ

Mekânsal olarak düzenlenmiş verileri dikkate alan istatistiksel yöntem mekânsal veri analizidir. Mekânsal verinin ait olduğu birimlerin birbirine bağımlılığından dolayı istatistiksel ilişkilerdeki gözlemlerin bağımsızlık varsayımının ihlal edilmesi söz konusudur, bu nedenle mekânsal istatistiksel analiz, diğer analiz biçimlerinden ayrılmaktadır. Mekânsal bağımlılık, i konumundaki bir gözlemin j konumundaki diğer gözlemlerden etkilenmesi, etkilenme nedeni ise gözlemlerin birbirine olan yakınlığı olarak ifade edilir. Mekânsal bağımlılığa konu olan yakınlık veya komşuluk ilişkileri (ülke, bölge, şehir, vb.) mekânsal ağırlık matrisleri ile tanımlanır. Regresyon analizi bağlamında bu bağımlılık veya mekânsal etkileşimlerin dikkate alındığı pek çok mekânsal regresyon modeli bulunmaktadır. Yakın ve birbiri ile ilişkili birimleri ölçen mekânsal bağımlılığın yapısına göre bu regresyon modelleri değişiklik göstermektedir ve dolayısı ile bu modellerin tahmin yöntemleri de değişkenlik içermektedir. Bu modellerden en sık kullanılan modeller, mekânsal hata modelleri (SEM) ve mekânsal gecikme modelleri (SLM) şeklindedir. Mekânsal modellerde içsellik problemlerinden dolayı En Küçük Kareler (EKK) yöntemi uygulanmaz bunun yerine maksimum olabilirlik yöntemi uygulanır. Ancak maksimum olabilirlik yönteminde bulunan logaritmik olabilirlik fonksiyonunda parametrelerin yeterince doğrusal olmaması nümerik maksimizasyon varsayımının uygulanmasını gerektirmektedir. Bu varsayım uygulanıp parametre tahmini yapılsa bile gözlem sayısının artmasıyla tahmin sürecinde problemler ortaya çıkmaktadır. Logaritmik olabilirlik fonksiyonunun maksimizasyonu, logaritmik determinantın bulunması ve determinantın boyutu örneklem büyüklüğüne bağlı olup her bir nümerik çözümde mekânsal bağımlılık parametresinin her bir değeri için ayrı ayrı determinant hesaplanır. Büyük verilerde yüksek boyutlu matrislerin determinantının hesaplanması oldukça zordur. Bu problemi ortadan kaldırmak için mekânsal büyük veri modellerinde geçerli Üstel Mekânsal Matris Tanımı (MESS) kullanılmaktadır. Bu yöntemlere alternatif olarak da kovaryans matrislerinin tersine çevrilmesindeki hesaplama zorluklarını ortadan kaldırmak için Tek Taraflı Yaklaşıklar Yöntemi, mekânsal hata terimlerindeki ikişerli ilişkileri modelleyen Kompozit Olabilirlik Yaklaşımı, içsellik problemini gidermek için Genelleştirilmiş İki Aşamalı En Küçük Kareler yöntemleri uygulanmaktadır. Mekânsal büyük verilerin hesaplanmasında mekânsal hata modelleri ve mekânsal gecikme modellerinde uygulanan üstel mekânsal matris tanımı hesaplamada kolaylık sağlamaktadır.

Bu çalışmada ABD ilçelerindeki cinayet oranlarını belirleyen faktörler mekânsal modeller aracılığı ile araştırılmıştır. Öncelikle bağımlı değişken cinayet oranı için mekânsal otokorelasyonun varlığı araştırılmıştır ve Geary C ve Moran I istatistiklerine göre bu değişken de mekânsal bağımlılık olduğu sonucuna varılmıştır. Mekânsal bağımlılık etkileri dikkate alınarak çeşitli mekânsal regresyon modelleri tahmin edilmiştir. Kullanılan modeller SAR, SEM, SAC VE MESS modelleridir. Bu tahminler sonucunda cinayet oranları üzerinde en etkili olan dört değişken sırasıyla kaynak yoksunluğu, boşanma oranı, hane reisi kadın olanlar ve işsizlik oranları olarak tespit edilmiştir.

Bu tez çalışmasında büyük boyutlu verilerin varlığında kullanılan mekânsal regresyon tahmincileri MESS, Tek Taraflı Yaklaşıklar Yöntemi ve Kompozit olabilirlik yaklaşımları tanıtılmıştır. Mekânsal büyük veri analizi ile ilgili yazılan bir tez olması nedeniyle büyük veri literatürü için tanıtıcı bir katkı sunulmaya çalışılmıştır.

KAYNAKÇA

- Adeyemi, R. A., Mayaki, J., Zewotir, T. T. ve Ramroop, S. (2021) "Demography and Crime: A Spatial analysis of geographical patterns and risk factors of Crimes in Nigeria", *Spatial Statistics*, 41.
- Aktan, E., "Büyük Veri: Uygulama Alanları, Analitiği ve Güvenlik Boyutu", *Bilgi Yönetimi Dergisi*, 1/2018(1)
- Altindag. D.T., "Crime and unemployment: Evidence from Europe", *International Review of Law Economics*, 32/2012, 145-157
- Andresen, M. A. (2006a) "A spatial analysis of crime in Vancouver, British Columbia: a synthesis of social disorganization and routine activity theory", *The Canadian Geographer*, 50 (4), 487–502.
- Andresen, M. A. (2006b) "Crime measures and the spatial analysis of criminal activity", *British Journal of Criminology*, 46, 258-285.
- Andresen, M. A. ve Malleon, N. (2013) "Spatial heterogeneity in crime analysis", M. Leitner (ed.), *Crime Modeling and Mapping Using Geospatial Technologies* içinde, Springer.
- Anselin L., (1980), Estimation methods for spatial autoregressive structures, Program in Urban and Regional Studies, Cornell University. USA.
- Anselin, L., "Model Validation in Spatial Econometrics: A Review and Evaluation of Alternative Approaches", *International Regional Science Review*, 11/1988(3), 279-316.
- Anselin. L., "Spatial Econometrics", A Companion to Theoretical Econometrics, Ed.by Badi. H. Baltagi, Blackwell, 310330/2001, 310-330.
- Anselin, L. Ve Lozano Gracia, N., "Errors in Variables and Spatial Effects in Hedonic House Price Models of Ambient Air Quality". *Empirical Economics*, 34/2008(2), 1-34.
- Anselin, L., Messner, S.F., Baller, R.D., vd. ..., "Structural covariates of U.S. county homicide rates: Incorporating spatial effects", *Criminology*, 39/2001(3), 561-588.
- Anselin.L., Rey. S.J., "Modern Spatial Econometrics in Practice: A Guide to GeoDa, GeoDaSpace and PySAL", GeoDa Press LLC, 2014, 394
- Arbia, G. (2014), A Primer for Spatial Econometrics: With Applications in R, Palgrave Texts in Econometrics, (1), Palgra ve Macmillan, İngiltere.

- Aydın, O., Aslantaş-Botan, P., Özgür, E.M., “Mekânsal Veri Analizi Teknikleriyle Türkiye’de Toplam Doğurganlık Hızının Dağılımı ve Modellenmesi”, *Coğrafya Dergisi*, 37/ 2018, 27-45
- Bell.B., Costa, R. ve Machin, S., “Why Does Education Reduce Crime?”, *Cepr Discussion Papers No 13162*, 2018.
- Besag, J., “Spatial Interaction and the Statistical Analysis of Lattice Systems”, *Journal of the Royal Statistical Society. Series B (Methodological)*, 36/1974(2), 192–236.
- Bianchi, M., Buonanno, P., Pinotti, P., “Do Immigrants Cause Crime?”, *Journal of the European Economic Association*, 10/2012(86), 1318-1347.
- Costantini. M, Meco, I., ve Paradiso, A., (2017), “Do inequality, unemployment and deterrence affect crime over the long run?”, *Regional Studies*, 52/2017(1), 1-14
- Cömertler, N., Kar, M., “Türkiye’de Suç Oranının Sosyo-ekonomik Belirleyicileri: Yatay Kesit Analizi”, *Ankara Üniversitesi SBF Dergisi*, 62/2007(02), 37-57.
- Chiu.Y.M., Leonard. T, ve Tsui. K., “The Matrix-Logarithmic Covariance Model”, *Journal of the American Statistical Association*, 91/1996 (433), 198-210.
- Cliff, A.D. and Ord, J.K. (1968). The problem of spatial autocorrelation, In Scott, A.J., editors, *Studies in Regional Science*, pages 25-55, Pion, London.
- Cliff, A., ve K. Ord, “Testing for Spatial Autocorrelation among Regression Residuals”, *Geographical Analysis*, 4/1972(3), 267-284.
- Çağlayan Akay, E (2020), “Ekonometride Büyük Veri ve Makine Öğrenmesi: Temel kavramlar, D&R yayınları, İstanbul.
- Çamurcu, H., “Dünya Nüfus Artışı ve Getirdiği Sorunlar”, *Balıkesir Üniversitesi, Sosyal Bilimler Enstitüsü Dergisi*, 8/2005(13), 87-105.
- Enamorado. T., López Calva, L.F., Rodríguez Castelán, C., Winkler, H., “Income Inequality and Violent Crime: Evidence from Mexico’s Drug War”, *Journal of Development Economics*, 120/2016, 128-143
- Ehrlich, I. “Participation in Illegitimate Activities: A Theoretical and Empirical Investigation”, *The Journal of Political Economy*, 81/1973(3), 521-565.
- Ehrlich.I., Paul, (1976), *Nüfus Bombası*, (Çevirmen: Nurullah Tolon), Ayyıldız matbası, Ankara.
- Geary, R. C. (1954). “The Contiguity Ratio and Statistical Mapping.” *The Incorporated Statistician* 5/1954(3), 115-146.

- GeoDa Data and Lab. "Homicides + Socio-Economics (1960-90)". Eriřim 26 Nisan 2022.
<https://geodacenter.github.io/data-and-lab//data-and-lab/ncovr/>.
- Getis, A., " A History of the concept of spatial autocorrelation: a geographer's perspective", *Geographical Analysis*, 40/2008(3), 297–309.
- Gujarati, D. N., ve Porter, D. C. (2018). *Temel ekonometri* (Çev. Ü. řenesen ve G. G. řenesen), (5. Basımdan çeviri), Literatür Yayıncılık, İstanbul.
- Horn, R., Johnson, C., (1993). *Matrix Analysis*, Cambridge University Press, New York
- Kangallı-Uyar, S.G., "Büyük Veri Setlerinde Üstel Mekânsal Matris Tanımı Yaklaşımı: İstanbul ofis piyasası Analizi", *Sosyoekonomi*, 28/2020(43), 65-87.
- Kelejian, H., Prucha, I.R., "A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances", *Journal of Real Estate and Finance Economics*, 17/1998, 99–121.
- Kelejian, H., ve Robinson, D.P., "A Suggested Method of Estimation for Spatial Interdependent Models with Autocorrelated Errors, and an Application to a County Expenditure Model". *Papers in Regional Science* 72/1993(3), 297-312.
- Land, K. C., McCall, P.L., ve Cohen, L.E., "Structural covariates of homicide rates: Are there any invariances across time and social space?", *American Journal of Sociology*, 95/1990, 922-963.
- LeSage J.P. (1999). *Spatial Econometrics*, The Web Book of Regional Science, Regional Research Institute, West Virginia University, Morgantown.
- LeSage, J.P., Pace, R.K., "A matrix exponential spatial specification", *Journal of Econometrics* 140/2007(1), 190–214.
- LeSage, J. ve Pace, R. K. (2009), *Introduction to Spatial Econometrics*, Chapman and Hall/CRC, New York.
- Miller, H. J. "Tobler's First Law and Spatial Analysis", *Annals of the Association of American Geographers*, 94/2004(2), 284-289.
- Moran, P. "The Interpretation of Statistical Maps." *Journal of the Royal Statistical Society, series B (Methodological)*, 10/1948(2), 243–51.
- Ord, J.K., "Estimation methods for models of spatial interaction", *Journal of the American Statistical Association*, 70/1975(349), 120–126

- Özgen. N., Bindak, R., “Mekân-Cinayet Bağıntısına Küresel Bir Bakış: Bir Davranışsal Coğrafya Çalışması”, İnsan ve Toplum Bilimleri Araştırmaları Dergisi, 6/2017(2), 1070-1087
- Pehlivanoglu. G.M., Yeler. O., (2019),“ Mekânsal Otokorelasyon İndislerinin Önemi”, Fen Bilimleri ve Matematik Alanında Araştırma ve Değerlendirmeler, , (Editör, Prof. Dr. Rıdvan KARAPINAR), Gece Gece Kitaplığı, Ankara, 91-100.
- Paelinck, J.H.P., ve Klaassen, L.H., (1979),Spatial Econometrics, Aldershot, UK: Gower.
- Paelinck, J.H.P. ve Nijkamp, P., (1975), Operational Theory and Methods in Regional Analysis, Farnborough: Saxon Hause.
- Sağıroğlu, Ş., Kaya Bensghir, T., Doğdu, E., Kılıç, K., vd. (2017) Büyük Veri Ve Açık Veri Analitiği : Yöntemler Ve Uygulamalar, Grafiker yayıncılık, ANKARA
- Strang, G., (1976), Linear Algebra and Its Applications, Academic Press, New York
- Tatlı. S., (2016), Mekânsal Ekonometrik Modeller ve Türkiye’de İç Göçün Belirleyicilerinin Analizi, (Yüksek Lisans Tezi), İstanbul Üniversitesi/ Sosyal Bilimler Enstitüsü/ Ekonometri Anabilim Dalı, İstanbul
- Tobler, W.R, “A Computer Movie Simulating Urban Growth in the Detroit Region” Economic Geography, 46/1970, 234-240.
- Tosun, Ö. "Suç ve Kadın". Journal of Istanbul University Law Faculty, 33 /2011(3-4), 27-44
- Tuzcu, S.E., “Mekânsal Ekonometri ve Sosyal Bilimlerde Kullanım Alanları”, Ankara Üniversitesi SBF Dergisi, 71/2016(2), 401-436
- Yamagata, Y., Seya.H., (2019). Spatial Analysis Using Big Data: Methods and Urban Applications, Academic press.
- Zeren, F., “Mekânsal etkileşim analizi”, İstanbul Üniversitesi İktisat Fakültesi Ekonometri ve İstatistik Dergisi, 12/2010, 18–39.
- Zeren, F., (2011), Mekânsal Ekonometri ve Mekânsal Panel Ekonometri Yaklaşımları: AB Üye Ülkeleri İçin Gelir Yakınsama Hipotezi Üzerine Bir Uygulama, (Doktora Tezi), İstanbul Üniversitesi/ Sosyal Bilimler Enstitüsü/ Ekonometri Anabilim Dalı, İstanbul
- Whittle, P., “On Stationary Processes in the Plane”, Biometrika, 41/1954, 434-449

