ZAMAN SERİLERİNDE REGRESYON ANALİZİ

1. GİRİŞ

- Trent, serinin genelinde yukarıya ya da aşağıya doğru olan hareketlere denmektedir. Bu hareket bazen düz bir doğru şeklinde olmaktadır. Bu tür harekete sahip trendlere *doğrusal trend* adı verilmektedir. Bazen ise bu hareket düz bir doğru şeklinde olmayıp matematiksel eğriler şeklinde olmaktadır. Böyle hareketlere sahip trendlere ise *eğrisel trend* adı verilmektedir.

- Basit regresyon analizi, doğrusal ilişkileri ya da doğrusal trendi modelleme amacıyla en sık uygulanan yöntemdir. Öte yandan eğrisel trendleri modelleyebilmek için verilere dönüşüm uygulamak ya da çoklu regresyon modelinin kurulması gerekmektedir. Ancak regresyon analizinin seriye uygulanabilmesi için serinin trendinin zaman içinde değişiklik göstermemesi gerektiğine dikkat edilmelidir.

- Zaman içinde aynı özellikte olan ve yapısal değişiklikler göstermeyen trendlere *determinisitik trend* adı verilmektedir. Eğer bir seri deterministik trende sahip ise, serinin grafiği daima trend doğrusuna ya da eğrisine dönme eğilimi gösterir.
- Zaman serileri regresyon analizinde bağımsız değişken olarak «zaman» ele alınır (t=1,2,...,T).

- Bu analizin istatistiksel olarak geçerli olabilmesi için birtakım varsayımların sağlanması ve istatistiksel testler sonucu istenilen özelliklerin elde edilebilmesi gerekmektedir. Varsayımlar sağlandıktan sonra hesaplanan öngörü değerleri güvenilir olacak ve gerçeği yansıtacaktır.

2. REGRESYON ANALİZİNDE İSTENİLEN ÖZELLİKLER

A. Normallik Testi

B. Değişen Varyans Sorunu

Bir serinin durağan olmama sorunu sadece fark işlemiyle çözülemeyebilir. Çünkü bazı seriler ortalamada durağan, varyansta durağan olmayabilirler. Bu durumda değişen varyans sorunu ortaya çıkar. Bu sorunun çözümü için seriye *Box-Cox Dönüşüm Yöntemi* uygulanır.

- Bu dönüşüm:

formülüyle ifade edilir. $z_t^{(\lambda)} = \dfrac{z_t^{\lambda} - 1}{\lambda}$

Burada λ dönüşüm parametresi olmaktadır. Eğer λ=0 ise seriye logaritma dönüşümü, λ=-1 ise seriye $1/z_t$ dönüşümü, λ=-0,5 ise seriye $1/\sqrt{z_t}$ dönüşümü, λ= 0,5 ise seriye $\sqrt{z_t}$ karekök dönüşümü uygulanır. Ancak λ=1 ise seriye dönüşüm uygulanmaz.

- Bu dönüşümlerden birinin seriye uygulanması gerekiyorsa analizin başında hatta fark işleminden de önce ilgili dönüşümün yapılması şarttır. Ayrıca logaritma ve karekök dönüşümlerinin sadece pozitif değerli serilere uygulanabileceğine dikkat edilmelidir. Eğer negatif değerli veriler varsa ve bu seriye logaritma ya da karekök dönüşümünün mutlaka yapılması gerekliyse, bu durumda serideki tüm veriler pozitif olabilecek şekilde keyfi bir sayıyla toplanmalıdır.

- Bir serinin tüm verilerine sabit bir değer eklendiğinde serinin yapısında bir değişiklik olmayacağı, dolayısıyla yapılacak analiz için bir sakınca doğurmayacağı unutulmamalıdır. Bu dönüşümler sadece değişen varyans sorununda değil, serinin normal dağılımı sahip olmadığı durumlarda seriyi normalleştirme amaçlı da kullanılabilmektedir. Hangi dönüşümün yapılmasına karar verme aşamasında HKO değerine bakılır. Hangi dönüşümde modelin HKO değeri en küçük ise seriye o dönüşüm uygulanır.

C. Otokorelasyon Testi

D. Regresyon Katsayılarının Önemlilik Testi

Eğer seriye uygulanan regresyon modellerinin hepsinde tüm terimlerin katsayıları önemli, hata serileri normal dağılıma sahip ve hiçbirinde otokorelasyon sorunu söz konusu değil ise, başka bir deyişle uygulanan tüm regresyon modelleri seriye uygun ise en uygun regresyon modelinin tespiti için regresyon modellerinin $\sqrt{HKO}$ değerleri hesaplanır.

- $\sqrt{HKO}$ değeri en küçük olan regresyon modeli seriye en iyi uyum sağlayan model olur. Zaman serileri analizinde bu tür karşılaştırmalarda modelin seriyi açıklama miktarını veren R^2 istatistiği güvenilir değildir. Bu nedenle modellerin R^2 değerlerinin karşılaştırılması yanlış yorumlara neden olmaktadır. Aynı durum düzeltilmiş R^2 istatistiği için de geçerlidir.

3. MEVSİMSEL OLMAYAN SERİLERDE REGRESYON ANALİZİ

A. Basit Doğrusal Regresyon Modeli

$$z_t = a + bt + \varepsilon_t$$

B. Birinci Farklar Regresyon Modeli

$$\Delta z_t = a + bt + \varepsilon_t$$

C. Üstel Regresyon Modeli

$$z_t = ab^t + \varepsilon_t$$

D. Karesel Regresyon Modeli

$$z_t = a + b_1 t + b_2 t^2 + \varepsilon_t$$

E. Lojistik Regresyon Modeli

Lojistik regresyon modelinde seriye:

$$z_t^* = \ln\left(\frac{L}{z_t} - 1\right)$$

dönüşümü yapılmaktadır. Burada L serideki en büyük gözlem değerinden daha büyük keyfi bir sabit değerdir. Serinin dönüşümü yapıldıktan sonra elde edilen yeni seriye; $z_t^* = a + bt + \varepsilon_t$

Basit doğrusal regresyon analizi uygulanır.

- Buradan z_t^* serisinin tahmini elde edilir. Orijinal serinin tahmini ise;

$$\hat{z}_t = \frac{L}{1+\exp(z_t^*)}$$

eşitliği ile hesaplanır.

F. Kübik Regresyon Modeli

$$z_t = a + b_1 t + b_2 t^2 + b_3 t^3 + \varepsilon_t$$

G. Diğer Regresyon Modelleri

-Logaritmik Regresyon Modeli

$$z_t = a + b \ln(t) + \varepsilon_t$$

-Artan Regresyon Modeli

$$\ln(z_t) = \ln(a) + \ln(b)t + \varepsilon_t$$

-Güç Regresyon Modeli

$$z_t = at^b$$

-S Regresyon Modeli

$$\ln(z_t) = a + b\,(1/t) + \varepsilon_t$$

-Ters Regresyon Modeli

$$z_t = a + b(1/t) + \varepsilon_t$$

4. MEVSİMSEL SERİLERDE REGRESYON ANALİZİ

1.Toplamsal Model İçin

- Eğer bir seride hem trend hem de mevsimsel dalgalanma var ise bu seriye uygulanacak regresyon modeli;

$$z_t = a + \sum_{i=1}^{m} b_i t^i + \sum_{j=1}^{[\![s/2]\!]} \left[c_j sin\left(\frac{2\pi jt}{s}\right) + d_j cos\left(\frac{2\pi jt}{s}\right)\right] + \varepsilon_t$$

biçiminde olmaktadır.

m: serinin trendinin yapısına göre polinom derecesidir.

- m=1 doğrusal trende sahip seri için
- m=2 karesel regresyona modeline uygun trende sahip seri için
- m=3 kübik regresyon modeline uygun trende sahip seri için
- $\sum_{i=1}^{m} b_i t^i$:serinin trend bileşenini açıklar.
- s:periyot olmak üzere $[\![s/2]\!]$ periyodun yarısının tamsayı kısmını gösterir. Örneğin periyot 9 ise $[\![s/2]\!]$, 4 olmaktadır.

- Burada j indisli toplam, yani köşeli parantez içindeki sinüs ve kosinüs fonksiyonları serinin mevsimsel bileşenini açıklamaktadır. Her regresyon modelinde olduğu gibi bu regresyon denkleminde de tüm katsayıların istatistiksel olarak önemli olması gerekmektedir.
- Sinüs ve kosinüs çiftine harmonik adı verilmektedir. Böylece, j=1 için birinci harmonik, j=2 için ikinci harmonik şeklinde j= $[\![s/2]\!]$ oluncaya kadar regresyon denklemine harmonik eklenir. Ancak her ekleme sonucunda c_j ve d_j regresyon katsayılarının önemlilik kontrolü yapılır.

- Bu regresyon katsayılarından biri önemsiz olduğunda harmonik ekleme işlemine son verilir ve önemsiz olan regresyon katsayısına ait terim modelden atılarak serinin tahmininde katsayılarının hepsi önemli olan regresyon modeli kullanılır.
- Uygulama verilerinde genellikle sadece birinci harmonikler seriyi açıklamada yeterli olmaktadır. Dolayısıyla, genellikle ikinci harmonikler regresyon modeline eklenmemektedir. Ayrıca, j= $[\![s/2]\!]$ için periyot çift sayı iken sinüs fonksiyonu, yani sin(πt) hep 0 değerini alacağından ilgili harmonikte sinüs serisinin oluşturulmamasına ve regresyon modeline eklenmemesine dikkat edilmelidir.

- İlgilenilen mevsimsel seri için elde edilen regresyon denkleminin hata terimi diğer yöntemlerde olduğu gibi mutlaka beyazgürültülü olmalıdır. Aksi takdirde elde edilen regresyon denklemi seriye uygun değildir ve hesaplanan tahmin değerleri güvenilir değildir. Bu durumda baştan seriye uygun yeni bir regresyon denklemi kurulmalı ya da regresyon analizinden başka bir yöntem seçilmelidir.

2.Çarpımsal Model İçin

- Bu modelde serideki dalgalanma büyüklükleri düzenli bir şekilde artmaktadır. Bu model regresyon analizine uygulandığında trende ve mevsimsel dalgalanmaya sahip bir serinin analizi için uygun olan bir regresyon denklemi:

$$z_t = a + \sum_{i=1}^{m} b_i t^i + \sum_{i=1}^{m} \sum_{j=1}^{[\![s/2]\!]} b_i t \left[c_j sin\left(\frac{2\pi jt}{s}\right) + d_j cos\left(\frac{2\pi jt}{s}\right)\right] + \varepsilon_t$$

biçiminde olmaktadır.

- Burada i indisli toplam serinin trendini, i ve j indisli iki toplam ise mevsimselliğin çarpımsal biçimde olduğunu göstermektedir. Bu denklem aşağıdaki biçimde de yazılabilir:

$$z_t = a + \sum_{i=1}^{m} b_i t^t + \sum_{i=1}^{m} \sum_{j=1}^{[|s/2|]} \left[c_j^* t \sin\left(\frac{2\pi jt}{s} \right) + \right] \left[d_j^* t \cos\left(\frac{2\pi jt}{s} \right) \right] + \varepsilon_t$$

Burada regresyon katsayıları c_j^* $=b_i c_j$ ve d_j^* $=b_i d_j$ olmaktadır.

- Bu denklemde bağımsız değişkenin t serisi ile sinüs fonksiyonu serisini çarpımı, t serisi ile kosinüs fonksiyonu serisinin çarpımı t serisi olduğuna dikkat edilmelidir.
- Mevsimsel serilere uygulanan regresyon analizinde katsayılar yine en küçük kareler yöntemi kullanılır.