ZAMAN SERİLERİ ANALİZİ DÖNEM SONU PROJESİ



HAZIRLAYANLAR 20201101061 Hatice Simay ÖZGÜL 20201101058 Ataman Önol ÜK

Bir Havayolu Şirketinin Yolcuları Üzerinden Zaman Serisi Analizi

Verinin Linki: https://www.kaggle.com/datasets/ashfakyeafi/air-passenger-data-for-time-series-analysis

<u>İÇERİK</u>

- Veri Setinin Tanımı
 - A) Korelasyon Testi
 - **B)** Otokorelasyon Testi
 - C) Normallik Testi
- Mevsimsellik ve Trend Analizi
- Durağanlaştırma (Fark Alma) İşlemi
- Uygun Model Karşılaştırması
 - 1. Doğrusal Regresyon Modelİ
 - 2. Birinci Farklar Modeli
 - 3. Üstel Regresyon Modeli
 - 4. Karesel Trend Modeli
 - 5. Lojistik Trend Modeli
 - 6. Kübik Trend Modeli
 - 7. Logaritmik Trend Modeli
- SONUÇ
- Karşılaşılabilecek Sorunlar
- Üstel Düzleştirme
- Veriyi Bileşenlere Ayırma
- ARIMA MODELLERİ

Veri Setinin Tanımı

Ele almış olduğumuz bu veri seti 1949 ile 1960 yılları arasında bir havayolu şirketine ait yolcu bilgisini vermektedir. Toplamda iki değişken (ay ve yolcu) ile 144 gözlemden oluşmaktadır.

A) Korelasyon Testi

H0: Gözlemler arasında ilişki olmadığından bu veri bir zaman serisi değildir.

H1: Gözlemler arasında ilişki olduğundan bu veri bir zaman serisidir.

Covariance Analysis: Ordinary Date: 06/10/23 Time: 23:20 Sample: 1949M01 1960M12 Included observations: 144					
Correlation Probability MONTH	MONTH 1.000000	PASSEN			
_PASSENGERS	0.923853 0.0000	1.000000			

Tabloya baktığımızda en altta yer alan "0.000" değeri bize prob değerini vermektedir. Bu değer 0.05'ten küçük olduğu için H0 hipotezi reddedilir.

Veri setindeki gözlemlerin ilişkili olduğunu ve bu sayede veri setinin bir zaman serisi olduğunu söylemek mümkündür.

B) Otokorelasyon Testi

H0: Otokorelasyondan söz edilemez.

H1: Otokorelasyon söz konusudur.

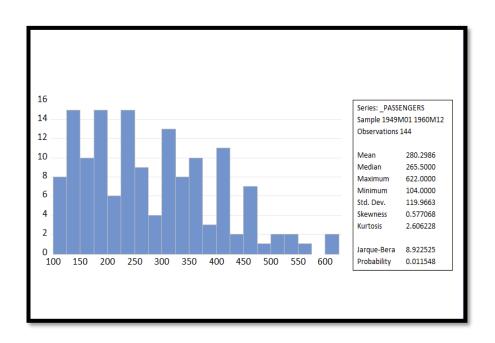
D + 00/40/00 T	00.04					
Date: 06/10/23 Tim						
Sample: 1949M01 19						
Included observation Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
-		1	0.979	0.979	140.96	0.000
		2		-0.010	276.94	0.000
		3		-0.011	408.00	0.000
		4		-0.010	534.19	0.000
		5		-0.010	655.59	0.000
	- i i	6		-0.010	772.26	0.000
1	- I i i	7		-0.011	884.28	0.000
1	ı İı i	8		-0.010	991.71	0.000
1	i	9		-0.010	1094.6	0.000
ı	. t i t i	10	0.792	-0.010	1193.1	0.000
1		11		-0.011	1287.3	0.000
ı ——	i i	12	0.751	-0.011	1377.1	0.000
ı İ	- I (I	13		-0.010	1462.8	0.000
ı ——	[14	0.710	-0.011	1544.4	0.000
	1 1	15	0.690	-0.011	1621.9	0.000
1	1 1	16	0.669	-0.011	1695.5	0.000
ı ———	[17	0.649	-0.011	1765.3	0.000
1		18	0.629	-0.010	1831.3	0.000
1		19	0.609	-0.011	1893.6	0.000
		20	0.589	-0.010	1952.3	0.000
ı —		21	0.569	-0.011	2007.6	0.000
ı —	- I (I - I	22	0.549	-0.011	2059.5	0.000
ı 	- [[]	23	0.529	-0.011	2108.1	0.000
ı 		24	0.509	-0.011	2153.6	0.000
ı 		25	0.490	-0.010	2195.9	0.000
ı 		26	0.470	-0.011	2235.3	0.000
ı 	- I (I	27	0.451	-0.011	2271.8	0.000
· 	[28	0.431	-0.011	2305.5	0.000
· 🗀 📗		29	0.412	-0.011	2336.6	0.000
· 🗀 📗		30	0.393	-0.011	2365.1	0.000
· 🗀 📗	1 1 1	31		-0.011	2391.1	0.000
ı 		32	0.355	-0.011	2414.8	0.000
· 🗀 📗		33	0.337	-0.011	2436.3	0.000
· 🗀 📗		34	0.318	-0.011	2455.6	0.000
· 🗀 📗		35		-0.011	2472.9	0.000
- 🗀	[[]	36	0.281	-0.011	2488.3	0.000

Bu hipotez testinde ele alınan konu için "Autocorrelation" adlı sütuna bakılması gerekmektedir. Güven sınırlarını aşan değerler görüldüğünden bu veri setinde otokorelasyon söz konusudur diyebilir ve H0 hipotezini reddedebiliriz.

C) Normallik Testi

H0: Veri normal dağılmaktadır.

H1: Veri normal dağılmamaktadır.



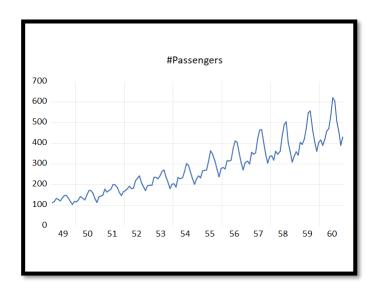
Yukarıdaki grafiğe baktığımızda veri setimizin sağa çarpık olduğunu görmekteyiz. (Skewness değeri 0.577 > 0)

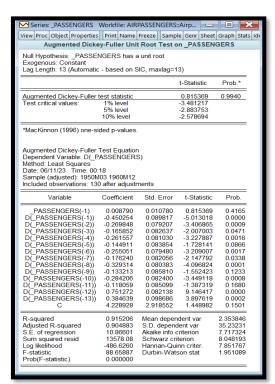
Aynı zamanda basıklık olduğunu da söyleyebiliriz. (Kurtosis değeri 2.60 < 3)

Son olarak Prob değerimiz 0.011 < 0.05 olduğu için H0 hipotezi reddedilir. Veri setinin normal dağılmadığı söylenebilir.

Mevsimsellik ve Trend Analizi

Ele alınan veri setinde mevsimsellik ve trend olup olmadığı incelenmesi amacı ile aşağıda uygulanmış olan testler görülmektedir. Öncelikle grafik incelendiğinde veride mevsimsellik olduğu, yukarı yönlü bir artış görüldüğünden ise trend olduğu yorumu yapılabilir. Ancak ne yazık ki grafik tek başına yeterli olmadığından hipotez testi ile bu yorum netleştirilmelidir.





Grafiğin yanında verilmiş olan "Augmented Dickey-Fuller" testi incelenecek şekilde bir hipotez testi yapılacak olursa,

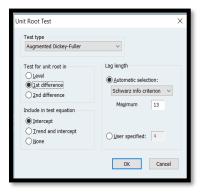
H0: Seri durağan değil, stokastik trende sahip, birim kök içermektedir.

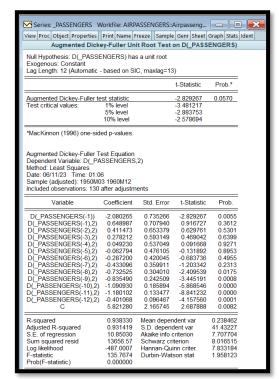
H1: Seri durağan, stokastik trende sahip değil, birim kök içermemektedir.

Tabloda verilmiş olan Prob değerinin 0.9940 geldiği görülmektedir. Bu değer 0.05'ten büyük olduğu için H0 reddedilemez. Serinin durağan olmadığı, dolayısıyla durağanlaştırma yoluna gidilmesi ve bu şekilde analize devam edilmesi gerektiği kanısına varmaktayız.

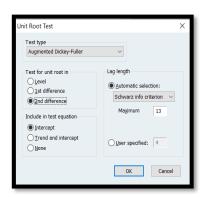
Durağanlaştırma (Fark Alma) İşlemi

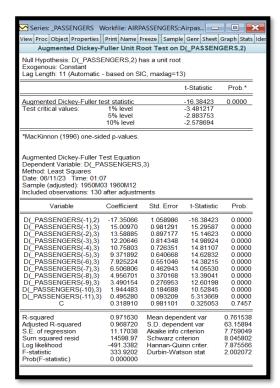
Serinin birinci farkı alınarak durağanlık incelendiğinde, yanda da görüldüğü üzere yeni Prob değeri 0.057 gelmiştir. Bu değer hala 0.05'ten büyük olduğu için ikinci defa fark alma işlemi uygulanacaktır.





İkinci fark alma işlemi sonrası Prob değeri 0.000 olduğundan 0.05 değerinden küçük gelmektedir. Artık bir önceki bölümde yazmış olduğumuz H0 hipotezini reddedebiliriz.

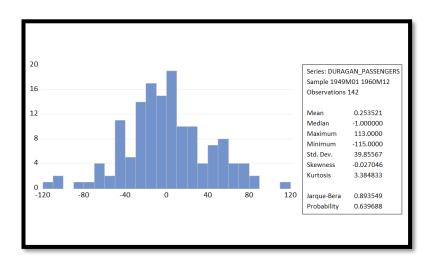




Özetle, seri artık birim kök içermemekte ve durağanlaşmış olup 2. dereceden entegre bir seridir denilebilir.

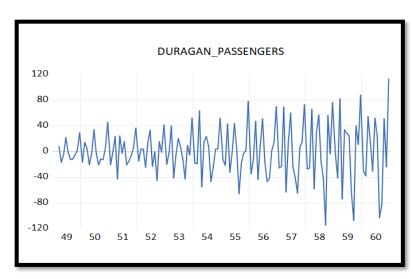
Durağanlaştırma işlemi sonrası veriye ait yeni bilgiler aşağıda verilmiştir.

Date: 06/11/23 Time: 01:33 Sample (adjusted): 1949M03 1960M12 Included observations: 142 after adjustments						
Autocorrelation	Partial Correlation	ents	AC	PAC	Q-Stat	Prob
—	·		-0.191		5.2908	0.021
			-0.196		10.901	0.004
1 🗓 1	-		-0.049		11.247	0.010
			-0.207		17.575	0.001
' !! '	' □ '	5		-0.104	19.703	0.001
1 1 1	— 1	6		-0.157	19.802	0.003
' <u> </u>		7		0.077	23.013	0.002
		-	-0.226		30.816	0.000
<u>- </u>	<u> </u>	9		-0.062	30.847	0.000
		i	-0.208		37.566	0.000
' <u>'</u>	<u> </u>	11	-0.121		39.848	0.000
'	! ' ᡛᢇ	12	0.782	0.231	136.10	0.000
<u> </u>	! ' _ ₽'		-0.096	0.114	137.56	0.000
	' ! '		-0.195		143.66	0.000
<u> </u>	ļ 'Щ'		-0.070		144.46	0.000
	! <u>! </u>	-	-0.148	0.006	148.01	0.000
' ! '	! <u>"</u> "	17	0.113	0.153	150.10	0.000
' <u>L</u> '	! <u>' !!</u> '	18	0.019	0.049	150.17	0.000
	! !!	19	0.137	0.084	153.28	0.000
	! !! !		-0.240		162.95	0.000
- -	! ! !	21		-0.015	163.07	0.000
. .	! ! !		-0.173		168.17	0.000
! -	! ! !		-0.084		169.39	0.000
	1	24		-0.010	240.79	0.000
	; ;		-0.051		241.25	0.000
I 78 :			-0.186		247.33 248.39	0.000
	; <u>;</u>		-0.077 -0.080	0.033	248.39 249.54	0.000
■ :5:		29		-0.066	249.54 249.95	0.000
■ : X:		30		-0.057	250.30	0.000
■ ; * ;	; ;	31		-0.057	250.30 252.37	0.000
I	¦ ; ™ ;		-0.182	0.137	252.37 258.50	0.000
	¦ ; , ; ; ; ;		-0.162		258.62	0.000
	; , ,		-0.025	0.124	259.96	0.000
	i 1		-0.064		262.14	0.000
		36		-0.036	314.61	0.000
	1 1 1	. 55	0.022	5.014	314.01	5.000



Jarque-Bera testinde olasılık değerine (prob=0.64) bakıldığında değerin 0.05'ten büyük olduğu görülmektedir. Dolayısıyla seri normal dağılım gösterir.

Durağanlaşmış veriye ait Çizgi Grafiği yanda verilmiştir.



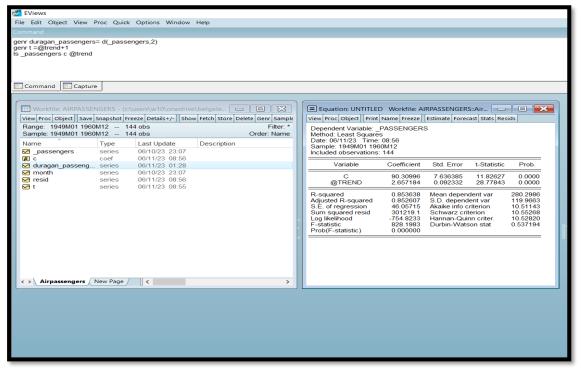
Uygun Model Karşılaştırması

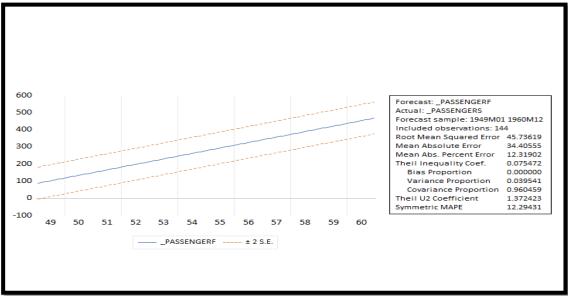
Seri durağan olduğundan trend modellerine girilebilir ve en küçük kareler yönteminden faydalanılarak hangi modelin uygun olduğu hesaplanabilir.

1. Doğrusal Regresyon Modeli

H0: Beta ve a katsayısı anlamsız, deterministik trend söz konusu değildir.

H1: Beta ve a katsayısı anlamlı, deterministik trend söz konusudur.





(genr t =@trend+1 ve ls _passengers c @trend kodları kullanılmıştır.)

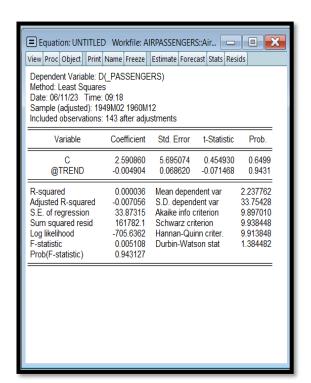
Her iki hipotezde de Prob değerinin, "0.05" değerinden küçük olduğu görülmektedir. Dolayısıyla H0 reddedilir, katsayı anlamlı ve deterministik trend vardır.

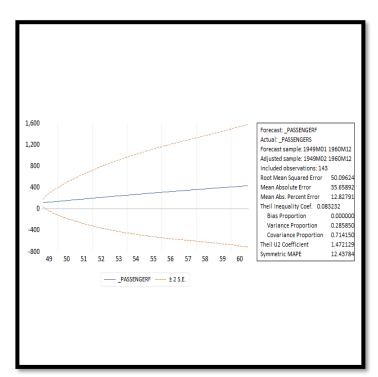
Özetle doğrusal regresyon modeli anlamlıdır.

2. Birinci Farklar Modeli

H0: Beta ve a katsayısı anlamsız, deterministik trend söz konusu değildir.

H1: Beta ve a katsayısı anlamlı, deterministik trend söz konusudur.





(ls d(_passengers) c @trend kodu kullanılmıştır.)

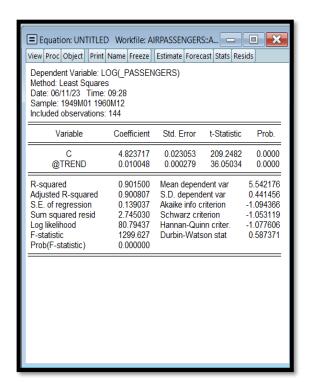
Her iki hipotezde de Prob değerinin, "0.05" değerinden büyük olduğu görülmektedir. Dolayısıyla H0 reddedilemez, katsayı anlamsız ve deterministik trend söz konusu değildir.

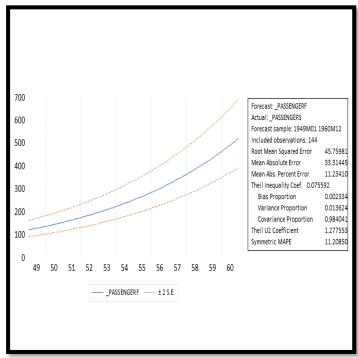
Özetle birinci farklar modeli anlamsızdır.

3. <u>Üstel Regresyon Modeli</u>

H0: Beta ve a katsayısı anlamsız, deterministik trend söz konusu değildir.

H1: Beta ve a katsayısı anlamlı, deterministik trend söz konusudur.





(ls log(_passengers) c @trend kodu kullanılmıştır.)

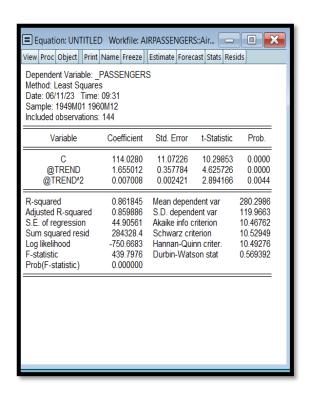
Her iki hipotezde de Prob değerinin, "0.05" değerinden küçük olduğu görülmektedir. Dolayısıyla H0 reddedilir, katsayı anlamlı ve deterministik trend vardır.

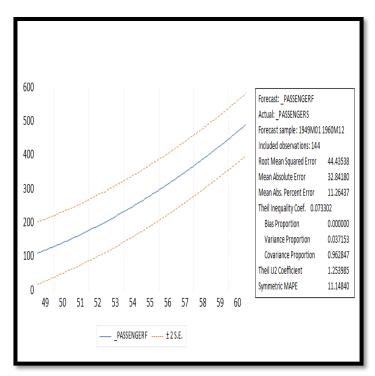
Özetle üstel regresyon modeli anlamlıdır.

4. Karesel Trend Modeli

H0: Beta ve a katsayısı anlamsız, deterministik trend söz konusu değildir.

H1: Beta ve a katsayısı anlamlı, deterministik trend söz konusudur.





(ls_passengers c @trend @trend^2 kodu kullanılmıştır.)

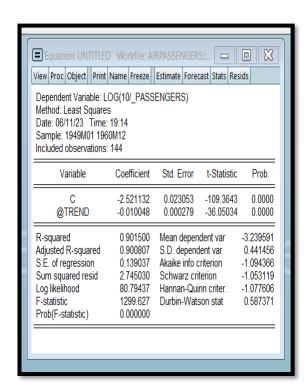
Her hipotezde Prob değerinin, "0.05" değerinden küçük olduğu görülmektedir. Dolayısıyla H0 reddedilir, katsayı anlamlı ve deterministik trend vardır.

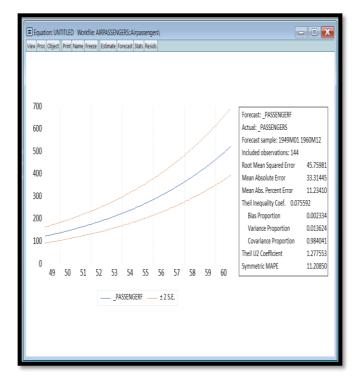
Özetle karesel trend modeli anlamlıdır.

5. Lojistik Trend Modeli

H0: Beta ve a katsayısı anlamsız, deterministik trend söz konusu değildir.

H1: Beta ve a katsayısı anlamlı, deterministik trend söz konusudur.





(ls log(10/ passengers-1) c @trend kodu kullanılmıştır.)

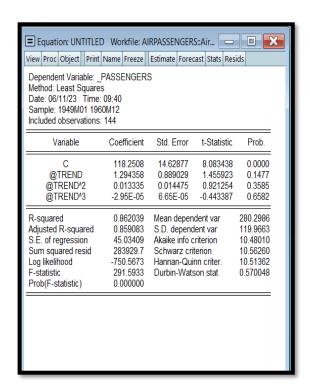
Her hipotezde Prob değerinin, "0.05" değerinden küçük olduğu görülmektedir. Dolayısıyla H0 reddedilir, katsayı anlamlı ve deterministik trend vardır.

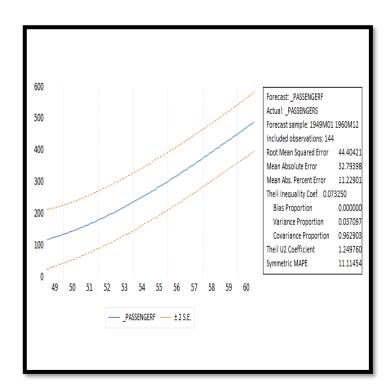
Özetle lojistik trend modeli anlamlıdır.

6. Kübik Trend Modeli

H0: Beta ve a katsayısı anlamsız, deterministik trend söz konusu değildir.

H1: Beta ve a katsayısı anlamlı, deterministik trend söz konusudur.





(ls _passengers c @trend @trend^2 @trend^3 kodu kullanılmıştır.)

Bu model ele alındığında a katsayısının Prob değerinin "0.05" değerinden küçük olması nedeniyle anlamlı olduğu ancak beta katsayılarının Prob değerlerinin "0.05" değerinden büyük olduğu görülmektedir.

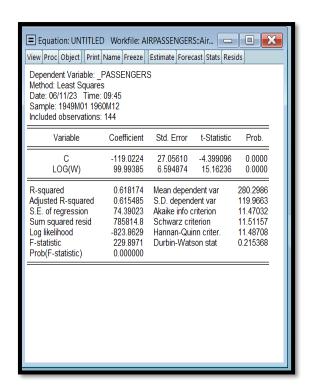
Dolayısıyla H0 hipotezi reddedilemez, deterministik trend söz konusu değildir.

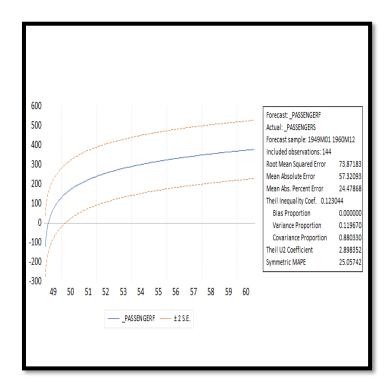
Özetle kübik trend modeli anlamsızdır.

7. Logaritmik Trend Modeli

H0: Beta ve a katsayısı anlamsız, deterministik trend söz konusu değildir.

H1: Beta ve a katsayısı anlamlı, deterministik trend söz konusudur.





(series w=@trend+1 ve ls passengers c log(w) kodları kullanılmıştır.)

Her iki hipotezde de Prob değerinin, "0.05" değerinden küçük olduğu görülmektedir. Dolayısıyla H0 reddedilir, katsayı anlamlı ve deterministik trend vardır.

Özetle logaritmik trend modeli anlamlıdır.

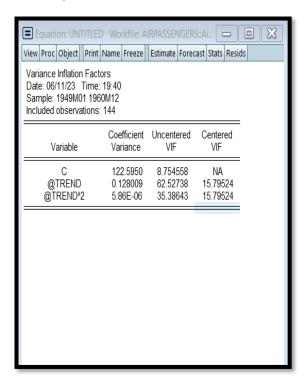
SONUÇ OLARAK;

Birinci Farklar Modeli ve Kübik Trend Modeli bu veri seti için anlamsız çıktığından dolayı karşılaştırmaya alınmayacaktır. Diğer modeller arasında "Root Mean Squared Error" değerleri karşılaştırılacak olur ise en düşük değerin "44.435" olarak Karesel Trend Modelinde olduğu görülmektedir.

En uygun model "Karesel Trend Modeli" seçilmiştir.

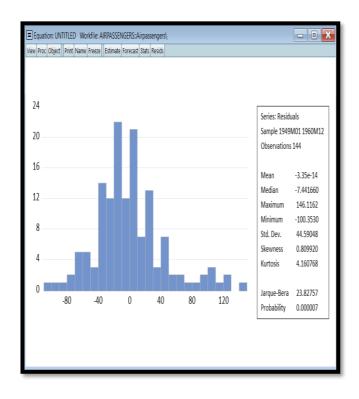
KARŞILAŞILABİLECEK SORUNLAR NELERDİR?

• ÇOKLU BAĞLANTI PROBLEMİ VAR MI?



Centered VIF değerine bakılarak bu problemin var olup olmadığına bakılır. Görüldüğü üzere değerler 1-5 aralığı için fazla büyük olduğundan çoklu bağlantı probleminin varlığından söz edilebilir.

• VERİ NORMAL DAĞILIYOR MU?



Prob değeri "0.05" değerinden küçük olduğu için H0 hipotezi reddedilmektedir. Dolayısıyla veri normal dağılmamaktadır diyebiliriz.

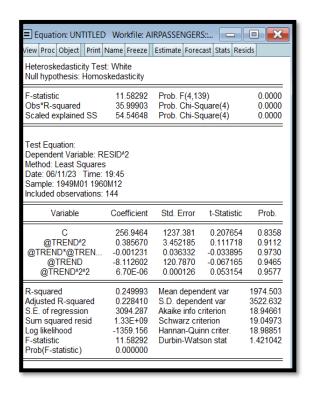
• OTOKORELASYON VAR MI?

H0 : Otokorelasyon yoktur.H1 : Otokorelasyon vardır.

Date: 06/11/23 Tim Sample: 1949M01 1						
Included observation						
Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
. —		۱ ،	0.700	0.700	70.000	0.000
		1 2	0.709	0.709	73.808 82.043	0.000
			-0.163		85.982	0.000
	i ii ii		-0.103	-	111.54	0.000
		:	-0.469		144.80	0.000
			-0.467		177.98	0.000
i .	i 🔳 :		-0.438		207.39	0.000
<u> </u>	 	8	-0.369	-0.321	228.45	0.000
		9	-0.135	0.144	231.27	0.000
· 🗀		10	0.200	-0.037	237.57	0.000
ı İ		11	0.589	0.544	292.41	0.000
ı	<u> </u>	12	0.823	0.185	400.26	0.000
	ļ <mark>I</mark>	13		-0.456	454.77	0.000
<u> </u>	! ' !!	14	0.172	0.159	459.53	0.000
'		i	-0.169	0.145	464.18	0.000
			-0.370		486.67	0.000
	<u> </u>	:	-0.418		515.54	0.000
			-0.431	-	546.50	0.000
			-0.422 -0.372		576.42 599.88	0.000
	<u>" </u>		-0.372	-	604.14	0.000
		22		-0.019	608.08	0.000
		23	-	-0.069	650.44	0.000
		24		-0.102	735.15	0.000
	101	25		-0.084	778.13	0.000
ı İ	[[[]	26		-0.004	781.58	0.000
-		27	-0.148	-0.002	785.50	0.000
<u> </u>	-		-0.313		803.24	0.000
<u> </u>			-0.356	-0.018	826.37	0.000
<u> </u>	ļ (ļ)		-0.361	0.081	850.39	0.000
			-0.354		873.67	0.000
<u> </u>	<u> </u>	-	-0.317		892.57	0.000
		i	-0.143		896.46	0.000
<u> </u>	' '	34	0.123	0.030	899.36	0.000
	 	35		-0.136	931.31	0.000
I		36	0.585	0.039	997.89	0.000

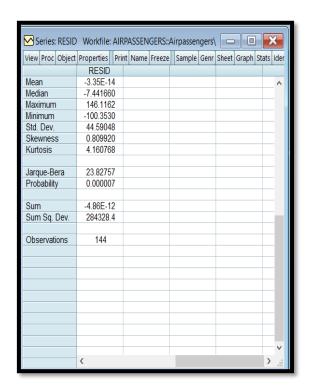
Güven sınırlarını aşan değerler olduğundan H0 hipotezi reddedilir, otokorelasyon vardır yorumunu yapabiliriz.

• DEĞİŞEN VARYANS PROBLEMİ VAR MI?



White Testi'ne bakıldığında Prob değerine (0.97) bakıldığında "0.05" değerinden büyük olduğundan değişen varyans sorunu söz konusu değildir.

• SIFIR ORTALAMA VARSAYIM KONTROLÜ

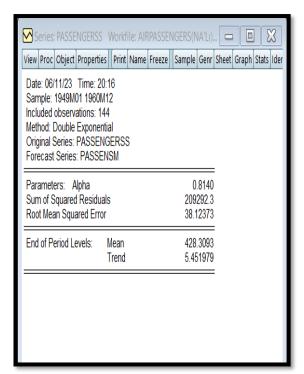


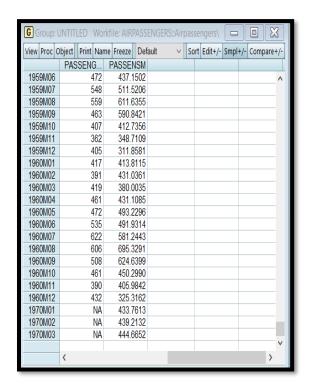
Hatalar ortalamasına bakacak olursak sonuç "-3.35E-14" geldiğinden 0'a çok yakın bir değer olduğundan bu varsayımdan söz edilebilir.

ÜSTEL DÜZLEŞTİRME

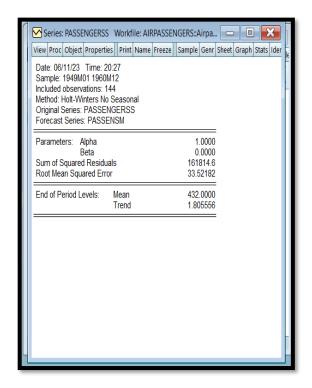
Bir önceki bölümde trend analizi yapılmıştı. Bu bölümde ise üstel düzleştime yöntemi yardımı ile modele ait zaman aralığı genişletilerek ileri vade için tahminleme yapılacaktır.

1. Çifte Üstel Düzleştirme





2. Holt-Winters Üstel Düzleştirme



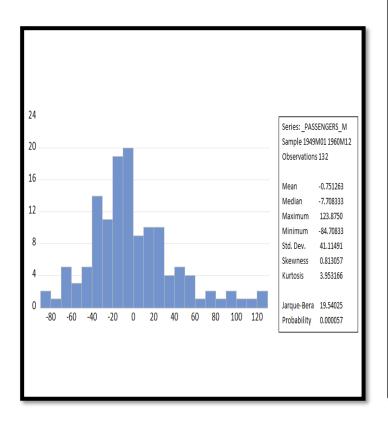
Tahminleme yapmak için Çifte Üstel ve Holt-Winters Üstel Düzleştirme testleri yapıldığında Çifte Üstel Düzleştirme değerleri daha küçük çıktığından tercih edilen yöntem olmuştur.

Boş gözlemler yerine yazılmış olan tahmin değerleri yukarıdaki tabloda gösterilmektedir.

Veriyi Bileşenlere Ayırma

Aşağıda verilmiş olan kodlar kullanılarak veri bileşenlere ayrılmış bilgiler olup yanda aşağıda ve verilmiştir.

"genr passengers2=0 passengers2=@movavc(_passengers,12) genr_passengers_m=0 _passengers_m=_passengers- _passengers2"



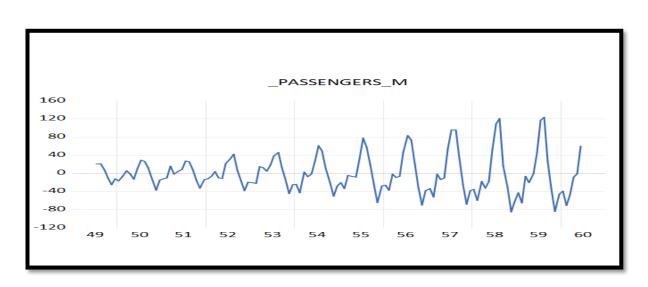
Null Hypothesis: _PASSENGERS_M has a unit root Exogenous: Constant Lag Length: 11 (Automatic - based on SIC, maxlag=12)								
			t-Statistic	Prob.*				
Augmented Dickey-Fuller test statistic -3.254976 0.0193								
Test critical values:	1% level		-3.485586	0.0.00				
Tool onlinear valueer	5% level		-2.885654					
-	10% level		-2.579708					
*MacKinnon (1996) one-sided p-values.								
-								
Augmented Dickey-Fuller Test Equation Dependent Variable: D(_PASSENGERS_M) Method: Least Squares Date: 06/11/23 Time: 15:13 Sample (adjusted): 1950M07 1960M06 Included observations: 120 after adjustments								
Variable	Coefficient	Std. Error	t-Statistic	Prob.				
_PASSENGERS_M(-1)	-1.985453	0.609975	-3.254976	0.0015				
D(_PASSENGERS_M(-1))	1.037644	0.563252	1.842237	0.0682				
D(_PASSENGERS_M(-2)) D(PASSENGERS M(-3))	0.750055 0.646896	0.514368	1.458208 1.380582	0.1477 0.1703				
D(_PASSENGERS_M(-3)) D(_PASSENGERS_M(-4))	0.646696	0.468568 0.414690	0.969112	0.1703				
D(_PASSENGERS_M(-4))	0.350464	0.368603	0.950789	0.3439				
D(_PASSENGERS_M(-6))	0.071815	0.317563	0.330703	0.8215				
D(_PASSENGERS_M(-7))	-0.027325	0.271187	-0.100762	0.9199				
D(_PASSENGERS_M(-8))	-0.322941	0.215669	-1.497388	0.1372				
D(_PASSENGERS_M(-9))	-0.363606	0.170156	-2.136906	0.0349				
D(PASSENGERS M(-10))	-0.674004	0.115218	-5.849818	0.0000				
D(_PASSENGERS_M(-11))	-0.683767	0.086083	-7.943138	0.0000				
C C	-0.141054	0.827482	-0.170462	0.8650				
R-squared	0.934204	Mean deper	ndent var	0.414236				
Adjusted R-squared	0.926825	•						
S.E. of regression	8.989424							
Sum squared resid	8646.644							
Log likelihood	-426.9187	Hannan-Quinn criter. 7.45461						
C atatiatia	400 0004	Durbin Waters stat 4.74205						

126.6031

0.000000

Durbin-Watson stat

1.743855

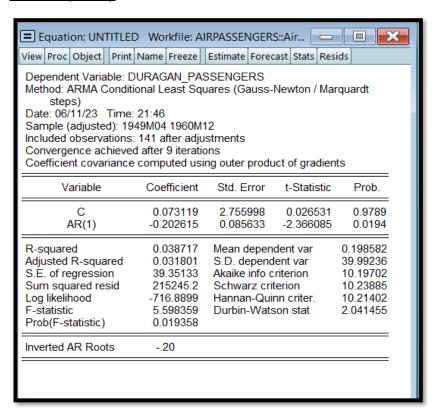


F-statistic

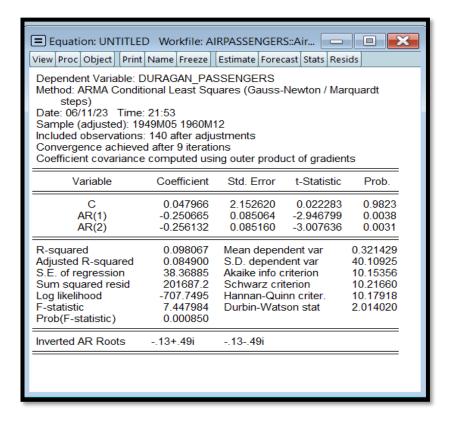
Prob(F-statistic)

ARIMA MODELLERİ

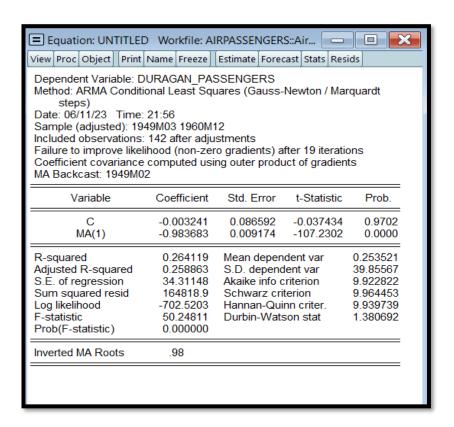
ARMA(1,0,0)



ARMA(2,1,0)



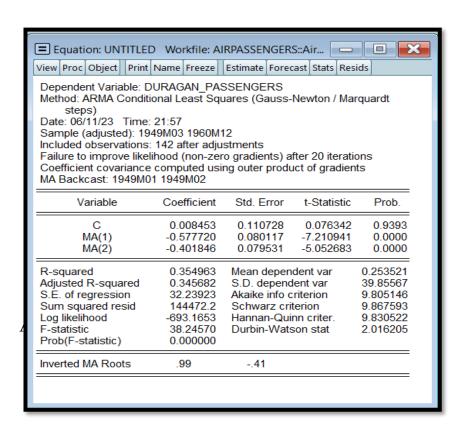
ARMA(0,0,1)



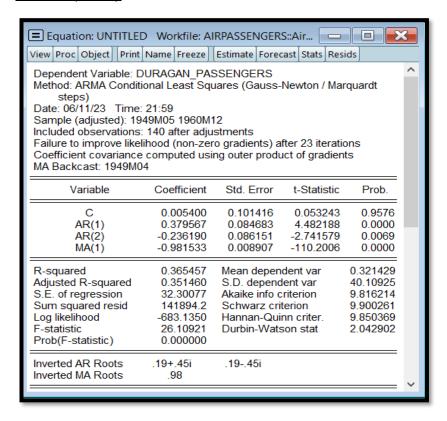
ARMA(0,1,1)

= Equation: UNTITLED) Workfile: Al	RPASSENGERS	S::Air	X		
	Name Freeze			ids		
Dependent Variable: DURAGAN_PASSENGERS Method: ARMA Conditional Least Squares (Gauss-Newton / Marquardt steps) Date: 06/11/23 Time: 21:58 Sample (adjusted): 1949M04 1960M12 Included observations: 141 after adjustments Failure to improve likelihood (non-zero gradients) after 20 iterations Coefficient covariance computed using outer product of gradients MA Backcast: 1949M03						
Variable	Coefficient	Std. Error	t-Statistic	Prob.		
C AR(1) MA(1)	0.006718 0.302361 -0.980806	0.128366 0.081533 0.009633	0.052334 3.708460 -101.8136	0.9583 0.0003 0.0000		
R-squared Adjusted R-squared S.E. of regression Sum squared resid Log likelihood F-statistic Prob(F-statistic)	0.330352 0.320647 32.96286 149943.9 -691.4029 34.03916 0.000000	Mean depend S.D. depend Akaike info c Schwarz crit Hannan-Quir Durbin-Wats	ent var riterion erion nn criter.	0.198582 39.99236 9.849687 9.912427 9.875182 1.826454		
Inverted AR Roots Inverted MA Roots	.30 .98					

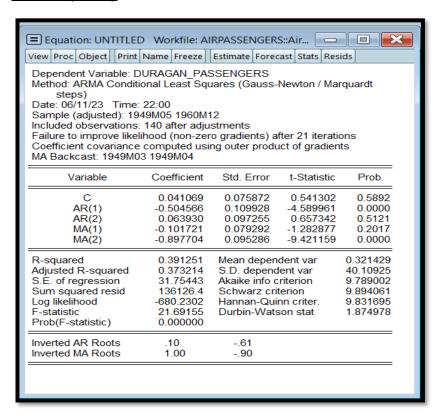
ARMA(0,1,2)



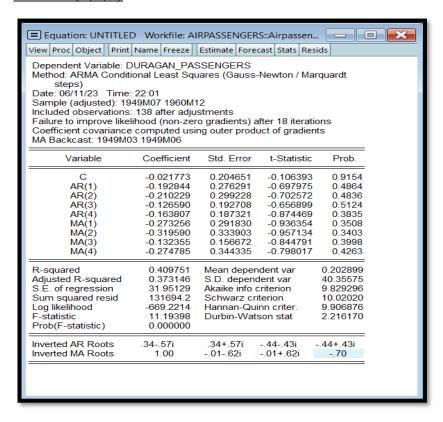
ARMA(2,1,1)



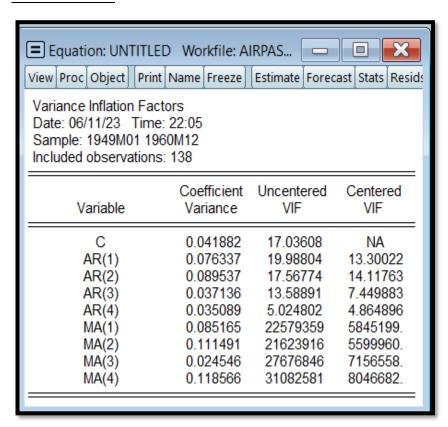
ARMA(2,1,2)



ARMA(4,2,4)



VİF DEĞERİ



SONUÇ OLARAK

ARIMA modelleri ile seçmiş olduğumuz Karesel Trend Modeli'ne bakıldığında ARMA(4,2,4) modelinde R^2 değeri "0,409" çıktığından ve VİF değerinde de bu doğrulandığından en anlamlı modelin ARMA(4,2,4) çıktığı söylenebilir.