Bölüm 4

İki Değişkenli Bağlanım Modeli - Tahmin Sorunu

4.1 Sıradan En Küçük Kareler Yöntemi

Sıradan En Küçük Kareler Yöntemi

- Bağlanım çözümlemesinde amaç, örneklem bağlanım işlevi (ÖBİ) temel alınarak anakütle bağlanım işlevinin (ABİ) olabildiğince doğru biçimde tahmin edilmesidir.
- Bunun için kullanılan en yaygın yol *"sıradan en küçük kareler"* (ordinary least squares), kısaca *"SEK"* (OLS) yöntemidir.
- SEK yönteminin 1794 yılında Alman matematikçi Carl Fredrich Gauss tarafından bulunduğu kabul edilir.
- SEK yöntemini anlamak için iki değişkenli ABİ'yi anımsayalım:

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

- ABİ gözlenemediğinden ÖBİ kullanılarak tahmin edilir:

$$\begin{aligned} Y_i &= \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{u}_i \\ &= \hat{Y}_i + \hat{u}_i \end{aligned}$$

- ÖBİ'nin kendisini bulmak için ise *"kalıntılar"* (residuals), diğer bir deyişle hata terimi kullanılır:

$$\begin{aligned}\hat{u}_i &= Y_i - \hat{Y}_i \\ &= Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i\end{aligned}$$

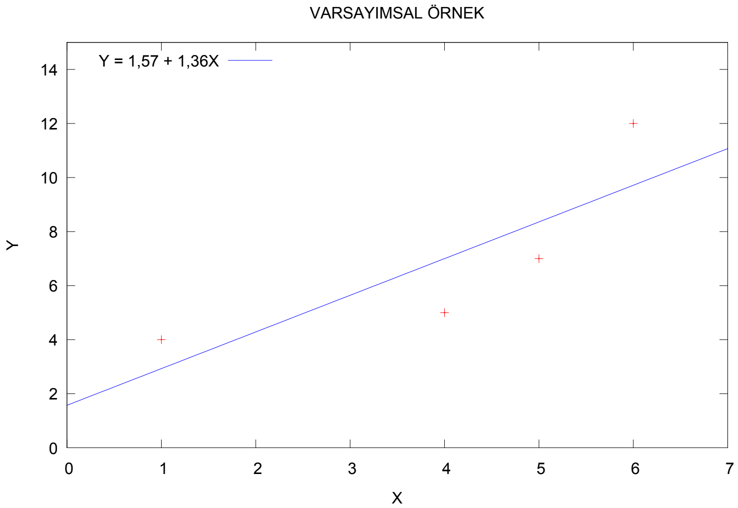
- Elimizde n tane X ve Y varken, ÖBİ'yi gözlenen Y'lere olabildiğince yakın biçimde belirlemek istiyoruz.
- Bunun için şu ölçüt benimsenebilir:

$$\min\left(\sum \hat{u}_i\right) = \min\left(\sum (Y_i - \hat{Y}_i)\right)$$

- Ancak bu durumda artı ve eksi değerli hatalar büyük ölçüde birbirlerini etkisiz hale getirecektir.
- Ayrıca burada ÖBİ'ye ne kadar yakın ya da uzak olursa olsun tüm kalıntılar eşit önem taşımaktadır.
- Öyleyse, ÖBİ'yi kalıntılar toplamı en küçük olacak şekilde seçmek iyi bir ölçüt değildir.
- Herhangi bir veri seti için farklı $\hat{\beta}_1$ ve $\hat{\beta}_2$ değerleri farklı $\hat{u}_i$ ve dolayısıyla da farklı $\sum \hat{u}_i{}^2$ toplamları verir.
- Ancak hatalar toplamı $\sum \hat{u}_i$ her zaman sıfır çıkar.
- Örnek olarak, varsayımsal bir veri seti için aşağıdaki iki ÖBİ'yi ele alalım:

$$\begin{aligned}\hat{Y}_{1i} &= 1{,}572 + 1{,}357X_i \\ \hat{Y}_{2i} &= 3{,}000 + 1{,}000X_i\end{aligned}$$

	Y_i	X_i	$\hat{Y}_{1i}$	$\hat{u}_{1i}$	$\hat{u}_{1i}^2$	$\hat{Y}_{2i}$	$\hat{u}_{2i}$	$\hat{u}_{2i}^2$
	4	1	2,929	1,071	1,147	4	0	0
	5	4	7,000	-2,000	4,000	7	-2	4
	7	5	8,357	-1,357	1,841	8	-1	1
	12	6	9,714	2,286	5,226	9	3	9
Toplam	28	16		0	12,214		0	14



- Artı ve eksi değerler alabilen kalıntıların toplamının küçük çıkma sorunundan kurtulmak için en küçük kareler ölçütü kullanılır:

En Küçük Kareler Ölçütü

$$\begin{aligned}\min\left(\sum \hat{u}_i{}^2\right) &= \min\left(\sum (Y_i - \hat{Y}_i)^2\right)\\ &= \min\left(\sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)^2\right)\end{aligned}$$

- Yukarıdaki gösterimin $\hat{\beta}_1$ ve $\hat{\beta}_2$ tahmincilerine dayanan bir matematiksel işlev olduğuna dikkat ediniz.

4.1.1 SEK Tahmincilerinin Türetilmesi

Normal Denklemler

- SEK, kalıntı kareleri toplamını *"enazlamak"* (minimize) için, ÖBİ değiştirgelerini hesaplamada basit bir *"eniyileme"* (optimization) yönteminden yararlanır.
- $\sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)^2$ teriminin $\hat{\beta}_1$ ve $\hat{\beta}_2$'ya göre kısmi türevlerini alalım:

$$\sum Y_i = n\hat{\beta}_1 + \hat{\beta}_2 \sum X_i$$

$$\sum Y_i X_i = \hat{\beta}_1 \sum X_i + \hat{\beta}_2 \sum X_i^2$$

- Burada n örneklem büyüklüğüdür.
- Yukarıdaki denklemler *"normal denklemler"* (normal equations) olarak adlandırılırlar.
- $\hat{\beta}_1$ ve $\hat{\beta}_2$ değiştirgeleri, normal denklemlerin eşanlı olarak çözülmesi ile bulunur:

$$\hat{\beta}_2 = \frac{n\sum X_iY_i - \sum X_i \sum Y_i}{n\sum X_i^2 - (\sum X_i)^2}$$
$$= \frac{\sum x_iy_i}{\sum x_i^2}$$

$$\hat{\beta}_1 = \frac{\sum X_i^2 \sum Y_i - \sum X_i \sum X_iY_i}{n\sum X_i^2 - (\sum X_i)^2}$$
$$= \bar{Y} - \hat{\beta}_2\bar{X}$$

- $\bar{X}$ ve $\bar{Y}$ terimleri X ile Y'nin örneklem ortalamalarıdır.
- Küçük harfler ise *"ortalamadan sapma"* (deviation from the mean) olarak kullanılmıştır:

$$x_i = (X_i - \bar{X})$$
$$y_i = (Y_i - \bar{Y})$$

SEK Bağlanım Doğrusunun Özellikleri

İkili bağlanım SEK tahmincileri $\hat{\beta}_1$ ve $\hat{\beta}_2$'nın şu özelliklerine dikkat edelim:

- Bunlar birer nokta tahmincisidirler.
- Gözlemlenebilen örneklem değerleri (X_i ve Y_i) cinsinden gösterilir ve dolayısıyla kolayca hesaplanabilirler.
- Örneklem verileri kullanılarak $\hat{\beta}_1$ ve $\hat{\beta}_2$ hesaplandıktan sonra, örneklem bağlanım doğrusu da kolayca çizilebilir.

SEK yöntemi ile bulunan örneklem bağlanım doğrusu aşağıda verilen özellikleri taşır:

1. Örneklem bağlanım doğrusu, X ve Y'nin örneklem ortalamalarından geçer. ($\bar{Y}_i = \hat{\beta}_1 + \hat{\beta}_2\bar{X}_i$)

2. $\hat{u}_i$ kalıntılarının ortalaması sıfırdır. ($\bar{\hat{u}}_i = 0$)

3. $\hat{u}_i$ kalıntıları tahmin edilen Y_i'lerle ilişkisizdir. ($\sum \hat{u}_i \hat{Y}_i = 0$)

4. $\hat{u}_i$ kalıntıları X_i'lerle ilişkisizdir. ($\sum \hat{u}_i X_i = 0$)

5. Tahmin edilen $\hat{Y}_i$'ların ortalaması, gözlemlenen Y_i değerlerinin ortalamasına eşittir. Bu ÖBİ'den görülebilir:

$$\begin{aligned} \hat{Y}_i &= \hat{\beta}_1 + \hat{\beta}_2 X_i \\ &= (\bar{Y} - \hat{\beta}_2 \bar{X}) + \hat{\beta}_2 X_i \\ &= \bar{Y} + \hat{\beta}_2 (X_i - \bar{X}) \end{aligned}$$

Son satırın her iki yanı örneklem üzerinden toplanıp n'ye bölünürse, $\bar{\hat{Y}} = \bar{Y}$ olarak bulunabilir.

ÖBİ'nin Sapma Biçiminde Gösterimi

- ÖBİ'nin *"sapma biçimi"* (deviation form) gösterimini bulmak için $Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{u}_i$ işlevinin her iki yanını toplayalım:

$$\begin{aligned} \sum Y_i &= n\hat{\beta}_1 + \hat{\beta}_2 \sum X_i + \sum \hat{u}_i \\ &= n\hat{\beta}_1 + \hat{\beta}_2 \sum X_i \qquad (\textstyle\sum \hat{u}_i = 0 \text{ olduğu için}) \end{aligned}$$

- Daha sonra bu denklemin her iki yanını n'ye bölelim:

$$\bar{Y} = \hat{\beta}_1 + \hat{\beta}_2 \bar{X}$$

- Yukarıdaki eşitlik, örneklem bağlanımı doğrusunun X ve Y'nin örneklem ortalamalarından geçtiğini göstermektedir.

- Son olarak yukarıdaki eşitliği ilk eşitlikten çıkaralım:

$$\begin{aligned} Y_i - \bar{Y} &= \hat{\beta}_2 (X_i - \bar{X}) + \hat{u}_i \\ y_i &= \hat{\beta}_2 x_i + \hat{u}_i \end{aligned}$$

- Sapma gösteriminde $\hat{\beta}_1$'nın bulunmadığına dikkat ediniz.

4.1.2 SEK Tahmincilerinin Arzulanan Özellikleri

Gauss - Markov Kanıtsavı

- Klasik Doğrusal Bağlanım Modeli (KDBM) varsayımları geçerli iken, en küçük kareler yöntemi ile elde edilen tahminler arzulanan bazı özellikler taşırlar.
- Gauss - Markov kanıtsavına göre $\hat{\beta}$ SEK tahmincilerine *"En iyi Doğrusal Yansız Tahminci"* (Best Linear Unbiased Estimator), kısaca *"EDYT"* (BLUE) adı verilir.

EDYT olan $\hat{\beta}$ şu üç arzulanan özelliği taşır:

1. Doğrusaldır. Diğer bir deyişle bağlanım modelindeki Y bağımlı değişkeninin doğrusal bir işlevidir.
2. Yansızdır. Beklenen değeri $E(\hat{\beta})$, anakütleye ait gerçek β değerine eşittir.
3. Tüm doğrusal ve yansız tahminciler içinde enaz varyanslı olandır. Kısaca en iyi ya da *"etkin"* (efficient) tahmincidir.

Gauss - Markov kanıtsavı hem kuramsal olarak hem de uygulamada önemlidir.

SEK Tahmincilerinin Doğrusallık Özelliği

- SEK tahmincilerinin *"doğrusallık"* (linearity) arzulanan özelliğini gösterebilmek için $\hat{\beta}_2$ formülünü şöyle yazalım:

$$\hat{\beta}_2 = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{\sum x_i(Y_i - \bar{Y})}{\sum x_i^2} = \frac{\sum x_i Y_i - \bar{Y}\sum x_i}{\sum x_i^2} = \frac{\sum x_i Y_i}{\sum x_i^2}$$

- Bu basitçe şu şekilde de gösterilebilir:

$$\frac{\sum x_i Y_i}{\sum x_i^2} = \sum k_i Y_i, \qquad k_i = \frac{x_i}{(\sum x_i^2)}$$

- x_i değerleri olasılıksal olmadığına göre k_i'ler de gerçekte Y_i'lerin önüne gelen birer *"ağırlık"* (weight) katsayısıdırlar.
- $\hat{\beta}_2$ bu durumda Y_i'lerin doğrusal bir işlevidir. Basitçe $\hat{\beta}_2$'nın Y_i'lerin bir ağırlıklı ortalaması olduğu da söylenebilir.
- $\hat{\beta}_1$'nın doğrusal olduğu da benzer biçimde kanıtlanabilir.

SEK Tahmincilerinin Yansızlık Özelliği

SEK tahmincilerinin *"yansızlık"* (unbiasedness) arzulanan özelliğini gösterebilmek için ağırlık terimi k'nin şu beş özelliği önemlidir:

1. X_i'ler olasılıksal olmadığından k_i'ler de olasılıksal değildir.
2. $\sum k_i = 0$'dır. ($\sum x_i = 0$ olduğu için)
3. $\sum k_i^2 = \sum x_i^2 / \sum (x_i^2)^2 = 1/\sum x_i^2$ olur.
4. $\sum k_i x_i = \sum x_i^2 / \sum x_i^2 = 1$'dir.
5. $\sum k_i x_i = \sum k_i X_i$ olur. ($\sum k_i x_i = \sum k_i(X_i - \bar{X}) = \sum k_i X_i - \bar{X} \sum k_i$ olduğu için)

Dikkat: Tüm bu özellikler k_i'nin tanımından türetilebilmektedir.

- $\hat{\beta}_2$'nın yansız olduğunu kanıtlamak için $Y_i = \beta_1 + \beta_2 X_i + u_i$ biçimindeki ABİ'yi $\hat{\beta}_2$ formülünde yerine koyalım:

$$\begin{aligned}\hat{\beta}_2 &= \sum k_i Y_i \\ &= \sum k_i(\beta_1 + \beta_2 X_i + u_i) \\ &= \beta_1 \sum k_i + \beta_2 \sum k_i X_i + \sum k_i u_i \\ &= \beta_2 + \sum k_i u_i\end{aligned}$$

- Yukarıdaki son adımda k_i'nin az önce sözü edilen ikinci, dördüncü ve beşinci özelliklerinden yararlanılmıştır.
- β_2 ve k_i'nin olasılıksal olmadığını ve $E(u_i) = 0$ varsayımını anımsayalım ve her iki yanın beklenen değerini alalım:

$$\begin{aligned}E(\hat{\beta}_2) &= E(\beta_2) + \sum k_i E(u_i) \\ &= \beta_2\end{aligned}$$

- $E(\hat{\beta}_2) = \beta_2$ olduğuna göre $\hat{\beta}_2$ yansız bir tahmincidir.

SEK Tahmincilerinin Enaz Varyanslılık Özelliği

- SEK tahmincilerinin *"enaz varyans"* (minimum variance) arzulanan özelliğini gösterebilmek için ise β_2'nin en küçük kareler tahmincisinden yola çıkalım:

$$\hat{\beta}_2 = \sum k_i Y_i$$

- Şimdi β_2 için başka bir doğrusal tahminci tanımlayalım:

$$\tilde{\beta}_2 = \sum w_i Y_i$$

- Buradaki ($\tilde{}$) işareti *"dalga"* (tilde) diye okunur.
- w_i'ler de birer ağırlıktır ama $w_i = k_i$ olmak zorunda değildir:
- $\tilde{\beta}_2$'nın yansız olabilmesi için gerekli koşullara bir bakalım:

$$\begin{aligned} E(\tilde{\beta}_2) &= \sum w_i E(Y_i) \\ &= \sum w_i(\beta_1 + \beta_2 X_i) \\ &= \beta_1 \sum w_i + \beta_2 \sum w_i X_i \end{aligned}$$

- Buna göre, $\tilde{\beta}_2$'nın yansız olabilmesi için şunlar gereklidir:

$$\sum w_i = 0, \qquad \sum w_i x_i = \sum w_i X_i = 1$$

- var$(\hat{\beta}_2) \leq$ var$(\tilde{\beta}_2)$ savını kanıtlamak istiyoruz. Bunun için şimdi $\tilde{\beta}_2$'nın varyansını ele alalım:

$$\begin{aligned} \text{var}(\tilde{\beta}_2) &= \text{var}(\sum w_i Y_i) \\ &= \sum w_i^2 \text{var}(Y_i) \qquad \text{[Dikkat: var}(Y_i) = \text{var}(u_i) = \sigma^2] \\ &= \sigma^2 \sum w_i^2 \qquad \text{[Dikkat: cov}(Y_i, Y_j) = 0,\ (i \neq j)] \\ &= \sigma^2 \sum \left(w_i - \frac{x_i}{\sum x_i^2} + \frac{x_i}{\sum x_i^2} \right)^2 \\ &= \sigma^2 \sum \left(w_i - \frac{x_i}{\sum x_i^2} \right)^2 + \sigma^2 \sum \left(\frac{x_i}{\sum x_i^2} \right)^2 + 2\sigma^2 \sum \left(w_i - \frac{x_i}{\sum x_i^2} \right) \left(\frac{x_i}{\sum x_i^2} \right) \\ &= \sigma^2 \sum \left(w_i - \frac{x_i}{\sum x_i^2} \right)^2 + \sigma^2 \left(\frac{1}{\sum x_i^2} \right) \end{aligned}$$

- Son satırda bulmuş olduğumuz şey şudur:

$$\text{var}(\tilde{\beta}_2) = \sigma^2 \sum \left(w_i - \tfrac{x_i}{\sum x_i^2} \right)^2 + \sigma^2 \left(\tfrac{1}{\sum x_i^2} \right)$$

- Yukarıda en sağdaki terim w_i'den bağımsızdır.
- Öyleyse var$(\tilde{\beta}_2)$'yı enazlayabilmek ilk terime bağlıdır ve ilk terimi sıfırlayan w_i değeri de şudur:

$$w_i = \tfrac{x_i}{\sum x_i^2} = k_i$$

- Bu durumda aşağıdaki eşitlik geçerlidir:

$$\text{var}(\tilde{\beta}_2) = \frac{\sigma^2}{\sum x_i^2} = \text{var}(\hat{\beta}_2)$$

- Demek ki w_i ağırlıkları k_i ağırlıklarına eşit olduğunda $\tilde{\beta}_2$'nın varyansı enazlanarak $\hat{\beta}_2$'nın varyansına eşitlenmektedir.
- Sonuç olarak, en küçük kareler tahmincisi $\hat{\beta}_2$ tüm yansız ve doğrusal tahminciler içinde enaz varyanslı tahmincidir.

4.1.3 SEK Yönteminin Ardındaki Varsayımlar

- Ekonometrik çözümlemenin amacı yalnızca β_1 ve β_2 gibi değiştirgeleri tahmin etmek değildir. Bu değerlere ilişkin çıkarsamalar yapmak da istenir.
- Örnek olarak, $\hat{Y}_i$'ların gerçek $E(Y|X_i)$ değerlerine ne kadar yakın olduklarını bilmek önemlidir.
- Anakütle bağlanım işlevini anımsayalım:

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

- Görülüyor ki Y_i hem X_i'ye hem de u_i'ye bağlıdır.
- Öyleyse Y_i, β_1 ve β_2'ye ilişkin istatistiksel çıkarım yapmak için X_i ve u_i'nin nasıl oluşturulduğunu bilmek gereklidir.
- Bu noktada Gaussçu *"Klasik Doğrusal Bağlanım Modeli"* (Gaussian Classical Linear Regression Model), kısaca *"KDBM"* (CLRM) 10 temel varsayım yapar.

Varsayım 1
Bağlanım modeli değiştirgelerde doğrusaldır:

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

- Ancak değişkenlerde doğrusallık zorunlu değildir.
- Değiştirgelerde doğrusallık varsayımı KDBM'nin başlangıç noktasıdır.

Varsayım 2
X değerleri tekrarlı örneklemelerde değişmez.

- Bu varsayım X'in olasılıksal olmadığını söyler.
- Buna göre X ve Y değerlerinin rastsal $\{X,Y\}$ çiftleri şeklinde elde edilmemiş olduğu kabul edilir.
- Diğer bir deyişle, gelir düzeyi başta örneğin 80 olarak belirlendikten sonra rastsal bir aile seçildiğini varsayıyoruz.
- Buna göre elimizdeki çözümleme açıklayıcı X değişkenine göre bir koşullu bağlanım çözümlemesidir.
- X ve Y değerlerinin birlikte örneklenebilmesi, bazı ek koşulların sağlanması ile geçerli olur. Bu duruma ise *"Neo-Klasik Doğrusal Bağlanım Modeli"* (NKDBM) denir.

Varsayım 3
u_i hata teriminin ortalaması sıfırdır:

$$E(u_i|X_i) = 0$$

- Buna göre, modelde açıkça yer almayan ve dolayısıyla u_i içine katılmış olan etmenlerin Y'yi kurallı bir şekilde etkilemediği varsayılmaktadır.
- Artı değerli u_i'ler eksi değerli u_i'leri götürmeli ve böylece bunların Y üzerindeki ortalama etkileri sıfır olmalıdır.

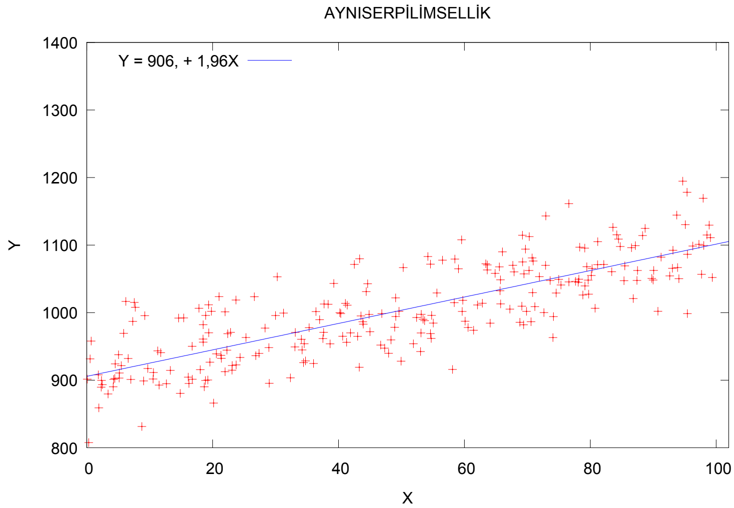
Varsayım 4
u_i hata teriminin varyansı tüm gözlemler için sabittir:

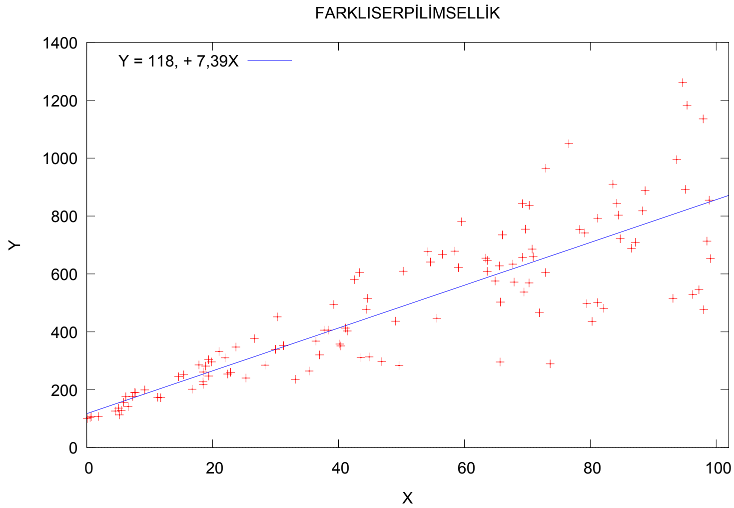
$$\text{var}(u_i|X_i) = \sigma^2$$

- *"Aynıserpilimsellik"* (homoscedasticity) varsayımına göre farklı X değerlerine karşılık gelen tüm Y'ler eşit önemdedir.
- Tersi durum ise *"farklıserpilimsellik"* (heteroscedasticity) durumudur:

$$\text{var}(u_i|X_1) \neq \text{var}(u_2|X_2) \neq \cdots \neq \text{var}(u_n|X_n).$$

- Farklıserpilimsellik durumunda çeşitli X değerlerine karşılık gelen Y değerlerinin güvenilirlikleri aynı olmaz.
- Bu yüzden kendi ortalaması etrafında farklı sıklıkta yayılan Y'leri farklı ağırlıklar vererek değerlendirmek gereklidir.

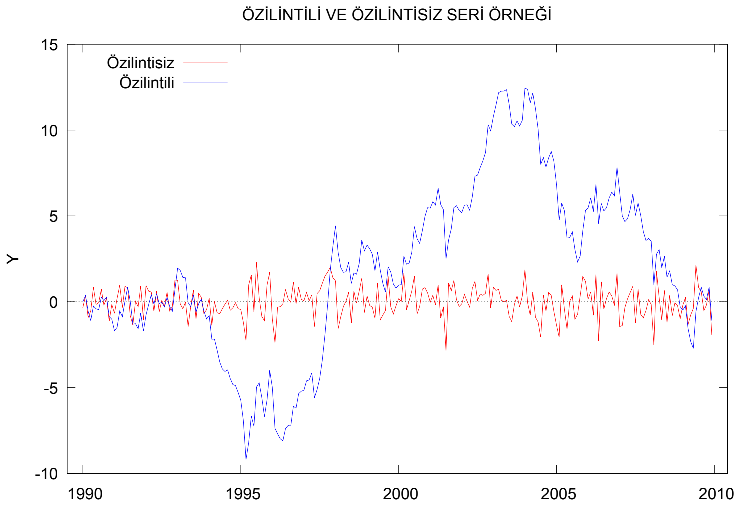


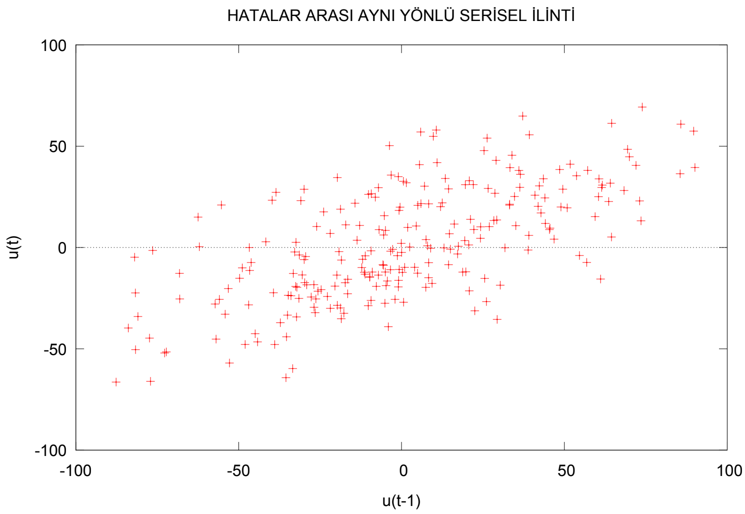


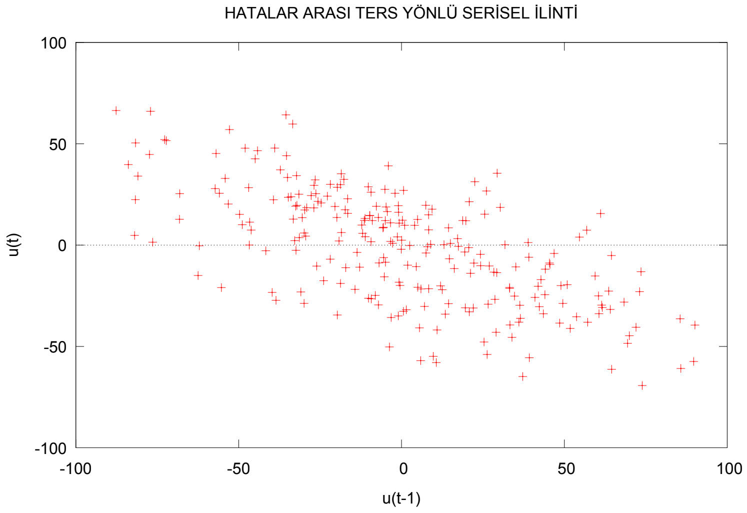
Varsayım 5
Hatalar arasında *"özilinti"* (autocorrelation) yoktur.

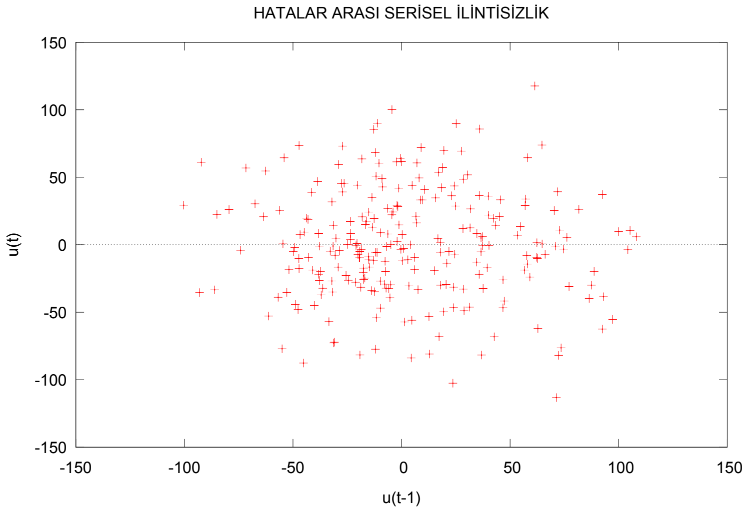
- Eğer *"bozukluklar"* (disturbances) birbirlerini kurallı biçimde izlerlerse özilinti ortaya çıkar.
- ABİ'yi $Y_t = \beta_1 + \beta_2 X_t + u_t$ olarak kabul edelim ve u_t ile u_{t-1} de aynı yönde ilişkili olsun.
- Bu durumda, Y_t yalnızca X_t'ye değil u_t'ye de bağlı olur ve bu nedenle u_t'yi bir ölçüde u_{t-1} belirler.

- Bu sorunla karşılaşmamak için hatalar arasında *"serisel ilinti"* (serial correlation) olmadığı varsayılır.









Varsayım 6
Hata terimi u_i ile X_i'nin kovaryansı sıfırdır:

$$\text{cov}(u_i, X_i) = 0$$

- Eğer X ve u ilişkiliyse, ikisinin de Y üzerindeki tekil etkilerini bulmak olanaksızlaşır.
- Ayrıca, eğer X ile u aynı yönde ilişkiliyse, X arttıkça u da artarak farklıserpilimsellik sorununa yol açar.

- Eğer 2. varsayım (X'in rastsal olmaması) ve 3. varsayım ($E(u_i|X_i) = 0$) geçerliyse, 6. varsayım da kendiliğinden gerçekleşmiş olur.

Varsayım 7
Gözlem sayısı n, tahmin edilecek anakütle katsayısından fazla olmalıdır.

- İki bilinmeyeni (β_1 ve β_2) bulmak için en az iki noktaya gereksinim vardır.
- Bu koşul çözümlemenin matematiksel olarak yapılabilmesi için gereklidir.
- Diğer yandan n serbestlik derecesi açısından önemlidir. Bu nedenle sağlıklı sonuçlar için örneklemin yeterince büyük olmasının ayrıca gerekli olduğu unutulmamalıdır.

Varsayım 8
Belli bir örneklemdeki X değerlerinin hepsi aynı olamaz:

$$\text{var}(X) \neq 0$$

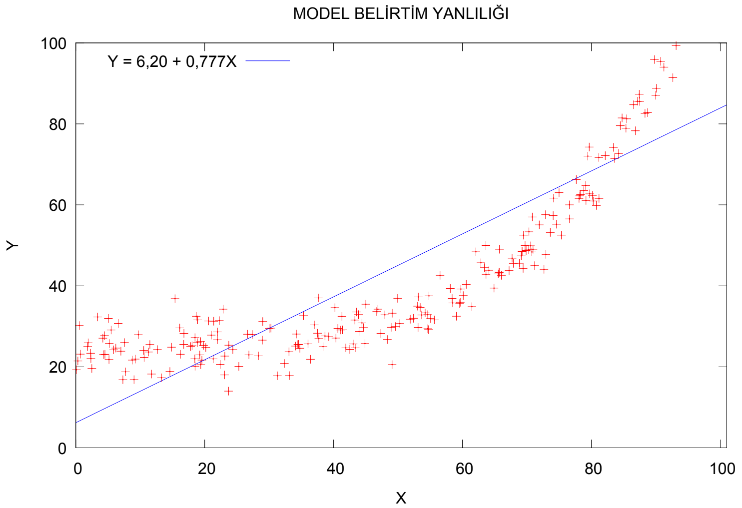
- Eğer bütün X değerleri aynı olursa:

$$\begin{aligned} & X_i = \bar{X}, & \\ & x_i = X_i - \bar{X} & \text{olduğundan} \\ & \hat{\beta}_2 = \tfrac{\sum x_i y_i}{\sum x_i^2} & \text{formülünün paydası sıfır çıkar.} \end{aligned}$$

- Kısaca değişkenler değişmelidir.

Varsayım 9
Bağlanım modeli doğru biçimde belirtilmiş olmalıdır.

- Bağlanım çözümlemesi sonuçlarının güvenilirliği, seçilen modele bağlıdır.
- Özellikle de bir iktisadi olguyu açıklayan birden fazla kuram bulunuyor ise ekonometrici çok dikkatli olmalıdır.
- Her durumda modelin işlev biçiminin ne olduğu, değişken ve değiştirgelerde doğrusal olup olmadığı konuları iyice sorgulanmalıdır.
- Bağlanım modeli yanlış olduğu zaman *"model belirtim hatası"* (model specification error) ortaya çıkar.



Varsayım 10
"Tam çoklueşdoğrusallık" (exact multicollinearity) yoktur.

- Tam çoklueşdoğrusallık durumunda bağlanım katsayıları belirsiz ve bu katsayıların ölçünlü hataları da sonsuz olur.

KDBM Varsayımlarının Gerçekçiliği
Ünlü ekonomist Milton Friedman'ın "varsayımların yersizliği" tezine göre gerçek dışılık bir üstünlüktür:

> "Önemli olabilmek için . . . bir önsav, varsayımlarında betimsel olarak gerçek dışı olmalıdır."

- Ekonometrideki KDBM'nin, fiyat kuramındaki tam rekabet modelinin karşılığı olduğu söylenebilir.
- Diğer bir deyişle öne sürmüş olduğumuz bu 10 varsayım gerçekleri tümüyle yansıtmak için değil, konuyu yavaş yavaş geliştirebilmeyi kolaylaştırmak amacıyla önemlidir.
- Bu varsayımların gerçekleşmemesi durumunda doğacak sonuçları ise ilerideki bölümlerde inceleyeceğiz.

4.2 SEK Yönteminin Güvenilirliği

4.2.1 SEK Tahmincilerinin Ölçünlü Hataları

- Sıradan en küçük kareler tahmincilerinin örneklem verilerinin birer işlevi olduğunu anımsayalım:

$$\hat{\beta}_2 = \frac{n\sum X_iY_i - \sum X_i \sum Y_i}{n\sum X_i^2 - (\sum X_i)^2}$$
$$= \frac{\sum x_iy_i}{\sum x_i^2}$$

$$\hat{\beta}_1 = \frac{\sum X_i^2 \sum Y_i - \sum X_i \sum X_iY_i}{n\sum X_i^2 - (\sum X_i)^2}$$
$$= \bar{Y} - \hat{\beta}_2\bar{X}$$

- Veriler örneklemden örnekleme değişeceği için tahminler de buna bağlı olarak değişecektir.
- Öyleyse $\hat{\beta}_1$ ve $\hat{\beta}_2$ tahmincilerinin güvenilirliği için bir ölçüte gereksinim vardır.
- İstatistikte rastsal bir değişkenin doğruluk derecesi *"ölçünlü hata"* (standard error), kısaca *"öh"* (se) ile ölçülür:

Ölçünlü Hata
Ölçünlü hata, bir tahminciye ait örneklem dağılımının kendi ortalamasından ortalama olarak ne kadar saptığını gösterir. Örneklem dağılımı varyansının artı değerli kare köküdür.

- Başta sözü edilmiş olan Gaussçu varsayımlar geçerli iken SEK tahmincilerinin ölçünlü hataları aşağıdaki gibidir:

$$\text{var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_i^2}$$

$$\text{öh}(\hat{\beta}_2) = \frac{\sigma}{\sqrt{\sum x_i^2}}$$

$$\text{var}(\hat{\beta}_1) = \frac{\sum X_i^2}{n \sum x_i^2} \sigma^2$$

$$\text{öh}(\hat{\beta}_1) = \sqrt{\frac{\sum X_i^2}{n \sum x_i^2}} \sigma$$

- Burada
  var değişirlik ya da varyansı,
  öh ölçünlü hatayı,
  σ^2 ise bağlanımın sabit varyansını
  göstermektedir.

- u_i'nin sabit varyansını veren σ^2 şöyle tahmin edilir:

$$\hat{\sigma}^2 = \frac{\sum \hat{u_i}^2}{n-2}$$

- Buradaki $\hat{\sigma}^2$, bilinmeyen σ^2'nin SEK tahmincisidir.

- $\sum \hat{u_i}^2$ terimine *"kalıntı kareleri toplamı"* (residual sum of squares), kısaca *"KKT"* (RSS) denir ve şöyle bulunur:

$$\sum \hat{u_i}^2 = \sum y_i^2 - \hat{\beta}_2^2 \sum x_i^2$$

- $n - 2$ değeri ise iki değişkenli çözümleme için geçerli serbestlik derecesidir.

Serbestlik Derecesi
"Serbestlik derecesi" (degree of freedom), örneklemdeki toplam gözlem sayısı (n) eksi bunlar üzerine konulmuş olan bağımsız ve doğrusal sınırlama sayısıdır.

- Örnek olarak, KKT'nin hesaplanabilmesi için önce $\hat{\beta}_1$ ve $\hat{\beta}_2$ değerlerinin bulunmuş olması gereklidir:

$$\sum \hat{u_i}^2 = \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)^2$$

- Dolayısıyla bu iki tahminci KKT üzerine iki sınırlama getirir.

- Bu durumda, KKT'yi ve dolayısıyla da ölçünlü hatayı doğru hesaplayabilmek için aslında elde n değil $n - 2$ sayıda bağımsız gözlem vardır.

SEK tahmincileri $\hat{\beta}_1$ ve $\hat{\beta}_2$'nın varyans formüllerini anımsayalım:

$$\text{var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_i^2} \qquad \text{var}(\hat{\beta}_1) = \frac{\sum X_i^2}{n \sum x_i^2}\sigma^2$$

$\hat{\beta}_1$ ile $\hat{\beta}_2$ tahmincilerinin varyanslarının ve dolayısıyla bunların ölçünlü hatalarının şu özellikleri önemlidir:

1. Örneklem büyüklüğü n arttıkça $\sum x_i^2$ toplamındaki terim sayısı da artar. Böylece n büyüdükçe $\hat{\beta}_1$ ve $\hat{\beta}_2$'nın doğruluk dereceleri de artar.
2. $\hat{\beta}_1$ ve $\hat{\beta}_2$, verili bir örneklemde birbirleri ile ilişkili olabilirler. Bu bağımlılık aralarındaki kovaryans ile ölçülür:

$$\text{cov}(\hat{\beta}_1, \hat{\beta}_2) = -\bar{X}\ \text{var}(\hat{\beta}_2)$$

3. Eğer $\bar{X}$ artı değerli ise kovaryans da eksi değerli olur. Bu durumda eğer β_2 katsayısı olduğundan büyük tahmin edilir ise β_1 de olduğundan küçük tahmin edilmiş olur.

4.2.2 Belirleme Katsayısı r^2

- Eldeki gözlemler çoğunlukla bağlanım doğrusu üzerinde yer almazlar.
- Artı ya da eksi işaretli $\hat{u}_i$ hataları ile karşılaşıldığına göre örneklem bağlanım doğrusunun eldeki verilerle ne ölçüde örtüştüğünü gösteren bir ölçüte gereksinim vardır:

Belirleme Katsayısı
"Belirleme katsayısı" (coefficient of determination) ya da r^2 (çoklu bağlanımda R^2), örneklem bağlanım işlevinin verilere ne kadar iyi yakıştığını gösteren özet bir ölçüttür.

Belirleme katsayısını hesaplamak için, $y_i = \hat{y}_i + \hat{u}_i$ eşitliğinin iki yanının karesi alınır ve örneklem boyunca toplanır:

$$\begin{aligned}
\sum y_i^2 &= \sum \hat{y}_i{}^2 + \sum \hat{u}_i{}^2 + 2\sum \hat{y}_i\hat{u}_i \\
&= \sum \hat{y}_i{}^2 + \sum \hat{u}_i{}^2 \\
&= \hat{\beta}_2^2 \sum x_i^2 + \sum \hat{u}_i{}^2 \\
\text{TKT} &= \text{BKT} + \text{KKT}
\end{aligned}$$

Burada
TKT *"Toplam Kareleri Toplamı"* (Total Sum of Squares),
BKT *"Bağlanım Kareleri Toplamı"* (Regression Sum of Squares),
KKT *"Kalıntı Kareleri Toplamı"* (Residual Sum of Squares)
anlamına gelmektedir.

- Yukarıdaki $\sum \hat{y}_i \hat{u}_i$ teriminin SEK bağlanım doğrusunun 3. özelliğinden dolayı sıfıra eşit olduğuna dikkat ediniz.

$$\begin{array}{ccccc} \sum y_i^2 & = & \hat{\beta}_2^2 \sum x_i^2 & + & \sum \hat{u}_i{}^2 \\ \text{TKT} & = & \text{BKT} & + & \text{KKT} \end{array}$$

Yukarıdaki eşitliğin her iki yanını TKT'ye bölelim:

$$1 = \tfrac{\text{BKT}}{\text{TKT}} + \tfrac{\text{KKT}}{\text{TKT}}$$

Buna göre r^2 aşağıdaki gibi tanımlanır:

Belirleme Katsayısı

$$r^2 = \frac{\sum \hat{y}_i{}^2}{\sum y_i^2} = \frac{\sum(\hat{Y}_i - \bar{Y})^2}{\sum(Y_i - \bar{Y})^2} = \frac{\text{BKT}}{\text{TKT}} = 1 - \frac{\text{KKT}}{\text{TKT}}$$

r^2'nin iki temel özelliğinden söz edilebilir:

1. r^2 eksi değer almayan bir büyüklüktür.
2. Sınırları $0 \leq r^2 \leq 1$'dir.

Buna göre:

- Eğer $r^2 = 1$ olursa bu kusursuz bir yakışma demektir. Bu durumda rastsal hata yoktur ve tüm gözlemler bire bir bağlanım doğrusu üzerinde yer almaktadır.
- Sıfıra eşit bir r^2 ise bağımlı değişkenle açıklayıcı değişken arasında hiçbir ilişkinin olmadığı ($\hat{\beta}_2 = 0$) anlamına gelir.

r^2 ile yakın ilişkili ama kavramsal olarak çok uzak bir büyüklük *"ilinti katsayısı"* (coefficient of correlation), kısaca r'dir:

İlinti Katsayısı

$$r = \pm\sqrt{r^2}$$

- r değeri, bağımlı ve açıklayıcı değişkenler arasındaki doğrusal bağımlılığın bir ölçüsüdür.
- -1 ve $+1$ arasında yer alır: $-1 \leq r \leq 1$.
- Bakışımlıdır: $r_{XY} = r_{YX}$.
- Sıfır noktasından ve ölçekten bağımsızdır.
- Herhangi bir neden-sonuç ilişkisi içermez.
- İki değişken arasında sıfır ilinti ($r = 0$) mutlaka bağımsızlık göstermez çünkü r yalnızca doğrusal ilişkiyi ölçer.

4.2.3 Monte Carlo Yöntemi

- KDBM varsayımları altında SEK tahmincilerinin EDYT (En iyi Doğrusal Yansız Tahminci) olmalarını sağlayan bazı arzulanan özellikler taşıdıklarını anımsayalım.
- EDYT özelliklerinin geçerliliği, bir *"benzetim"* (simulation) yöntemi olan Monte Carlo deneyleri ile doğrulanabilir.
- Bu yöntem, anakütle katsayılarını tahmin eden süreçlerin istatistiksel özelliklerini incelemede sıkça kullanılmaktadır.
- Monte Carlo aynı zamanda istatistiksel çıkarsamanın temeli sayılan *"tekrarlı örnekleme"* (repeated sampling) kavramının anlaşılması için de yararlı bir araçtır.

Bir Monte Carlo deneyi aşağıdaki gibi yapılır:

1. Anakütle katsayıları seçilir. Örnek: $\beta_1 = 20$ ve $\beta_2 = 0{,}6$.
2. Bir örneklem büyüklüğü seçilir. Örnek: $n = 25$.
3. Her gözlem için bir X değeri belirlenir.
4. Bir rastsal sayı oluşturucu kullanılarak u_i kalıntıları üretilir.
5. β_1, β_2, X_i'ler ve u_i'ler kullanılarak Y_i değerleri bulunur.
6. Bu şekilde üretilen Y_i değerleri X_i'ler ile bağlanıma sokulur ve $\hat{\beta}_1$ ve $\hat{\beta}_2$ SEK tahmincileri hesaplanır.
7. İşlem tekrarlanır (örneğin 1000 kez) ve rastsallıktan dolayı her seferde değişen tahminlerin ortalamaları ($\bar{\hat{\beta}}_1, \bar{\hat{\beta}}_2$) alınır.
8. Eğer $\bar{\hat{\beta}}_1$ ve $\bar{\hat{\beta}}_2$ değerleri β_1 ve β_2'ye aşağı yukarı eşit ise, deney SEK tahmincilerinin yansızlığını, diğer bir deyişle $E(\hat{\beta}_1) = \beta_1$ ve $E(\hat{\beta}_2) = \beta_2$ olduğunu saptamış sayılır.

4.3 Sayısal Bir Örnek

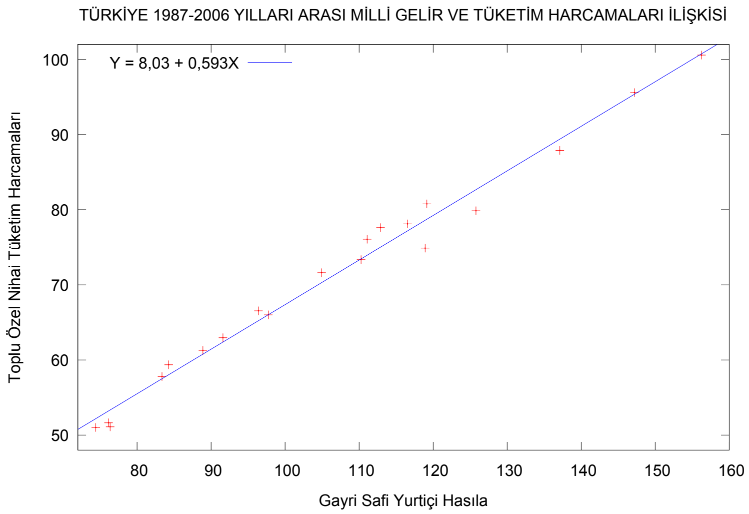
Sayısal Bir Örnek

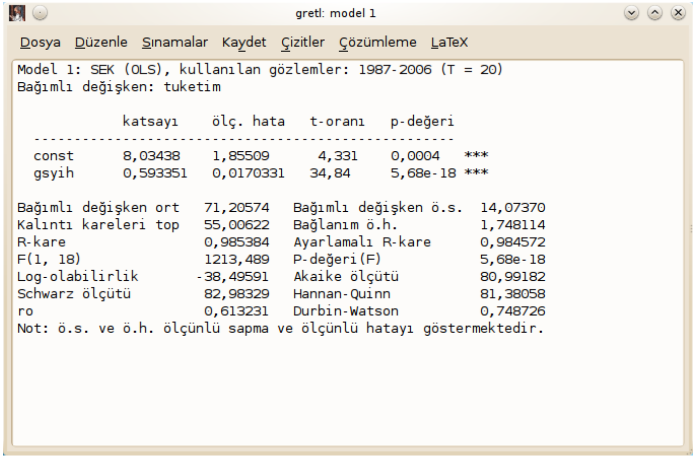
- Ele almış olduğumuz bazı kavramları sayısal bir örnek yardımı ile gözden geçirelim. Türkiye'de 1987–2006 arası toplam tüketim harcamaları ve GSYH verileri şöyledir:

Çizelge: Türkiye'de Tüketim ve GSYH (1987–2006)

Yıl	C	Y	Yıl	C	Y
1987	51.019	74.416	1997	77.620	112.892
1988	51.638	76.143	1998	78.113	116.541
1989	51.105	76.364	1999	76.077	111.083
1990	57.803	83.371	2000	80.774	119.147
1991	59.366	84.271	2001	73.356	110.267
1992	61.282	88.893	2002	74.894	118.923
1993	66.545	96.391	2003	79.862	125.778
1994	62.962	91.600	2004	87.897	137.110
1995	66.011	97.729	2005	95.594	147.200
1996	71.614	104.940	2006	100.584	156.249

- Toplu özel nihai tüketim harcamalarını (Y), gayri safi yurtiçi hasıla (X) ile ilişkilendirmek istiyor olalım.





gretl: model 1

Dosya Düzenle Sınamalar Kaydet Çizitler Çözümleme LaTeX

```
Model 1: SEK (OLS), kullanılan gözlemler: 1987-2006 (T = 20)
Bağımlı değişken: tuketim

             katsayı    ölç. hata   t-oranı   p-değeri
  ---------------------------------------------------------
  const      8,03438    1,85509      4,331    0,0004   ***
  gsyih      0,593351   0,0170331   34,84     5,68e-18 ***

Bağımlı değişken ort   71,20574   Bağımlı değişken ö.s.  14,07370
Kalıntı kareleri top   55,00622   Bağlanım ö.h.          1,748114
R-kare                 0,985384   Ayarlamalı R-kare      0,984572
F(1, 18)               1213,489   P-değeri(F)            5,68e-18
Log-olabilirlik       -38,49591   Akaike ölçütü          80,99182
Schwarz ölçütü         82,98329   Hannan-Quinn           81,38058
ro                     0,613231   Durbin-Watson          0,748726
Not: ö.s. ve ö.h. ölçünlü sapma ve ölçünlü hatayı göstermektedir.
```

- Gretl çıktısına göre marjinal tüketim eğilimi (MTE) 0,59'dur.
- Buna göre gelir 1 lira arttığında tüketimin de 59 kuruş artması beklenmektedir.
- Sabit terim, toplam gelir sıfır olduğunda toplam tüketimin yaklaşık 8 milyon lira olacağını göstermektedir.
- Sıfır gelirin gözlem aralığı dışında kalan ve gerçek hayatta olanaksız bir değer olmasından dolayı, sabit terimin böylesi bir mekanik yorumu iktisadi anlam içermemektedir.
- Gretl $\hat{\beta}_1$, $\hat{\beta}_2$ ve $\hat{u}_i$ için ölçünlü hataları sırasıyla 1,85509 ve 0,0170331 ve 1,748114 olarak hesaplamıştır.
- Yukarıdaki değerlerin karesi alınarak var$(\hat{\beta}_1) = 3{,}44136$ ve var$(\hat{\beta}_2) = 0{,}000290126$ ve $\hat{\sigma}^2 = 3{,}05590$ varyansları da kolayca bulunabilir.
- $r^2 = 0{,}985$ değeri ise bağlanım modelinin verilere gerçekçi kabul edilemeyecek kadar iyi yakıştığını göstermektedir.

Önümüzdeki Dersin Konusu ve Ödev

Ödev
Kitaptan *Bölüm 3* “Two-Variable Regression Model: The Problem of Estimation” okunacak.

Önümüzdeki Ders
Normallik Varsayımı ve Ençok Olabilirlik Yöntemi

UADMK Açık Lisans Bilgisi

A. Talha Yalta
TOBB Ekonomi ve Teknoloji Üniversitesi
Ekim 2011