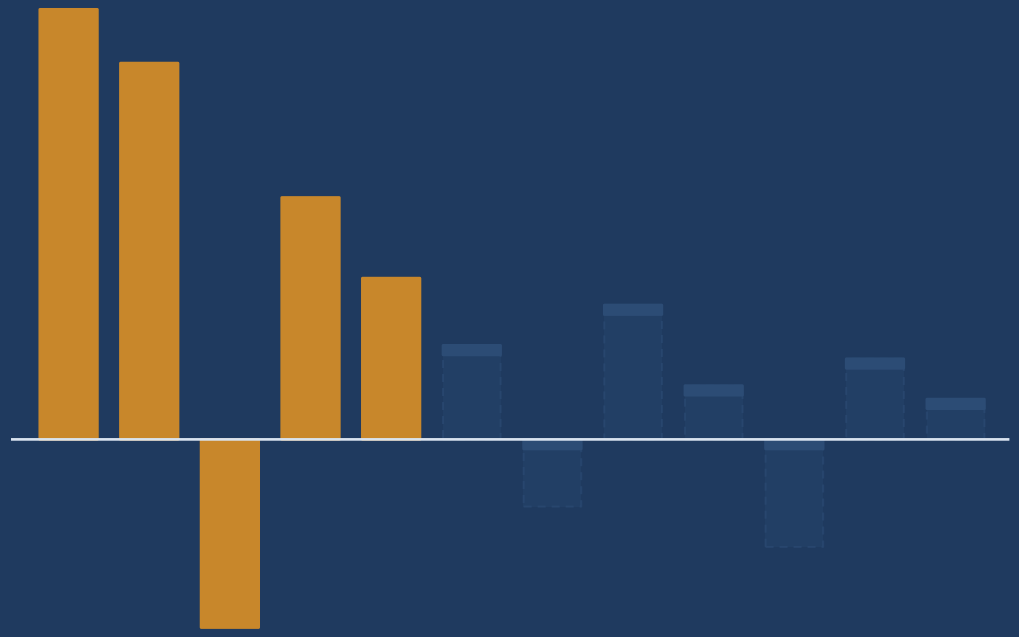
EKONOMİ VE FİNANSTA MAKİNE ÖĞRENMESİ

HAFTA 4

Katsayılara Ceza Kesmek

Regresyon ve düzenlileştirme

EKK, Ridge ve Lasso



Prof. Dr. Mehmet Songur

Dicle Üniversitesi · İİBF · İktisat Bölümü

Geçen haftadan

Bir modelin değeri, hiç görmediği veride ölçülür



Aşırı uyum

Çok esnek model gürültüyü ezberler



Yanlılık ve varyans

Esneklik ↑: yanlılık ↓, varyans ↑



Çapraz doğrulama

Hiperparametre doğrulamayla seçilir



Veri sızıntısı

Önce böl, sonra hazırla

Geçen haftanın son cümlesi: “Biraz yanlılığı kabul edip varyansı çok azaltmak toplam hatayı düşürebilir. Bunu bilinçli olarak yapan yöntemler var.” Bu hafta o yöntemler.

Bugün ne konuşacağız?

1

EKK’nın sınırı

Çok değişken, az gözlem olunca ne olur?

2

Ceza fikri

Hata + λ × ceza ve geometrisi

3

Ridge ve Lasso

İki farklı ceza, iki farklı sonuç

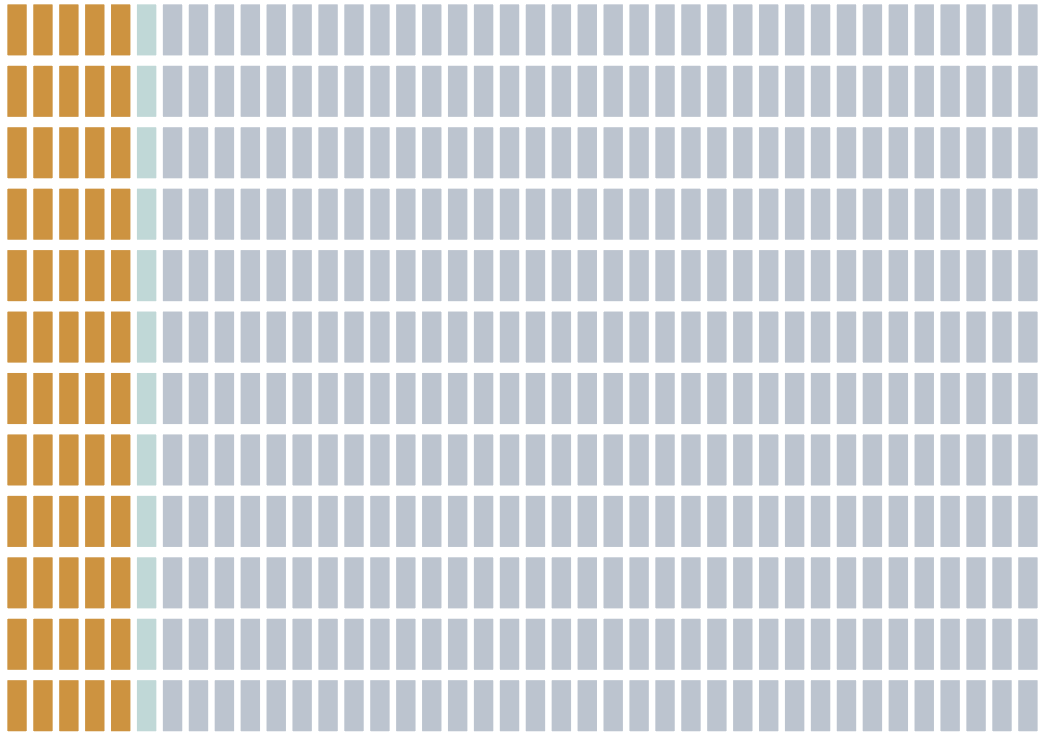
4

Seçim ve uyarılar

λ’yı seçmek, ölçeklemek, doğru yorumlamak

100 kişi, 40 değişken

Hepsini tek bir regresyona koyup EKK ile kestirirsek ne olur?



100 satır (kişi) ↓ 40 sütun (değişken) →

5 gerçek etken

eğitim, deneyim, deneyim², kıdem, büyükşehir

Yaş

deneyimle neredeyse aynı bilgi (ρ ≈ 0,97)

34 anket sorusu

ücrete hiçbir etkisi yok

Gerçekte bunu bilmiyoruz: model 40'ının arasından bulmak zorunda.

Katsayılara bir ceza kes: biraz yanlılık, çok daha az varyans.

Bu haftanın tek cümlesi

1

EKK'nın sınırı

Çok değişken, az gözlem olunca ne olur?

EKK neyi minimize eder?

Ekonometriden bildiğiniz kestirici, makine öğrenmesi gözüyle

$$\min \quad \Sigma\,(y - \hat{y})^2$$

Kalıntı kareler toplamı = Hafta 3'ün diliyle eğitim hatası

Güçlü yanı

Gauss-Markov: yansız kestiriciler arasında en küçük varyanslıdır. Değişken az, gözlem çokken çok iyidir.

Zayıf yanı

Yalnızca eğitim hatasına bakar. Değişken sayısı arttıkça "yansızların en iyisi" olmak yetmez: o en iyinin varyansı da büyür.

Değişken sayısı arttıkça üç sorun



Varyans büyür

Her ek katsayı veriden kestirilir ve biraz gürültü taşır. 100 gözlemle 40 katsayı, Hafta 3'teki derece 12 polinomuna benzer.



Katsayılar oynaklaşır

Yaş ve deneyim neredeyse aynı bilgiyi taşır: çoklu doğrusal bağlantı. Katsayılar birbirini telafi ederek örneklemden örnekleme oynar (Hafta 1).



EKK çözülemez

Değişken sayısı gözlem sayısını aşarsa (80 ülke, 150 gösterge) X'X'in tersi alınamaz; sonsuz sayıda çözüm veriye tam uyar.

Simülasyon: 40 değişkenle EKK

100 kişiyle eğit, 5.000 yeni kişide test et; 200 tekrar

0,053

Eğitim hatası

0,156

Test hatası

0,090

İndirgenemez hata

Eğitim hatası tabanın bile altında: model gürültüyü ezberlemiş. Test hatası ise tabanın neredeyse iki katı. Hafta 3'ün aşırı uyumu, bu kez çok değişkenden.

2

Ceza fikri

Katsayılar büyüdükçe bedel ödesinler

Hataya bir ceza ekle

$$\min \quad \Sigma\,(y - \hat{y})^2 \quad + \quad \lambda \times (\text{katsayıların büyüklüğü})$$

Birinci terim

Modeli veriye uymaya iter. EKK ile aynı.

İkinci terim

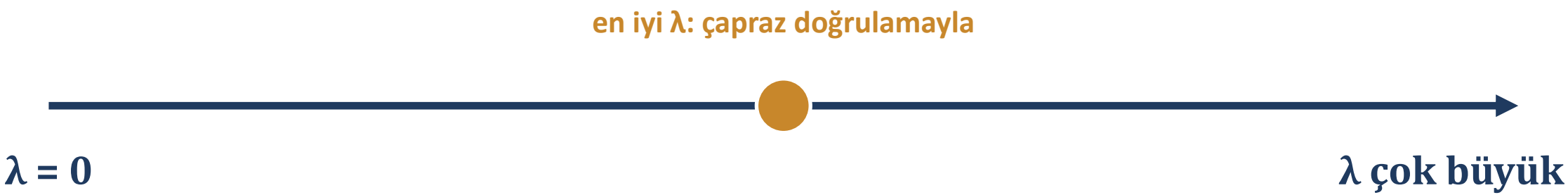
Katsayıları sıfıra doğru çeker. Büyük katsayı “pahalıdır”: ancak hatayı yeterince düşürüyorsa büyür.

λ (lambda)

Cezanın gücü. Modelin esnekliğini ayarlayan düğme: bir hiperparametre.

λ: esneklik düğmesi

Hafta 3'teki polinom derecesinin yerini alıyor, ama yön ters



Ceza yok

Sonuç tam olarak EKK. Yanlılık düşük, varyans yüksek.

Arada

Biraz yanlılık, çok daha az varyans. Hedeflediğimiz yer.

Ceza çok güçlü

Tüm katsayılar sıfıra yakın; model yalnızca ortalama ücreti öngörür. Yanlılık yüksek.

İki farklı ceza, iki farklı yöntem

“Katsayıların büyüklüğü” iki biçimde ölçülebilir

Ridge

$$\lambda \times \Sigma\, \beta^2$$

katsayıların kareleri

Tüm katsayıları sıfıra doğru küçültür. Hiçbirini tam sıfır yapmaz.

Hoerl ve Kennard (1970)

Lasso

$$\lambda \times \Sigma\, |\beta|$$

katsayıların mutlak değerleri

Katsayıları küçültür ve bazılarını tam sıfır yapar: değişken seçer.

Tibshirani (1996)

Bildiğiniz bir problem: kısıtlı optimizasyon

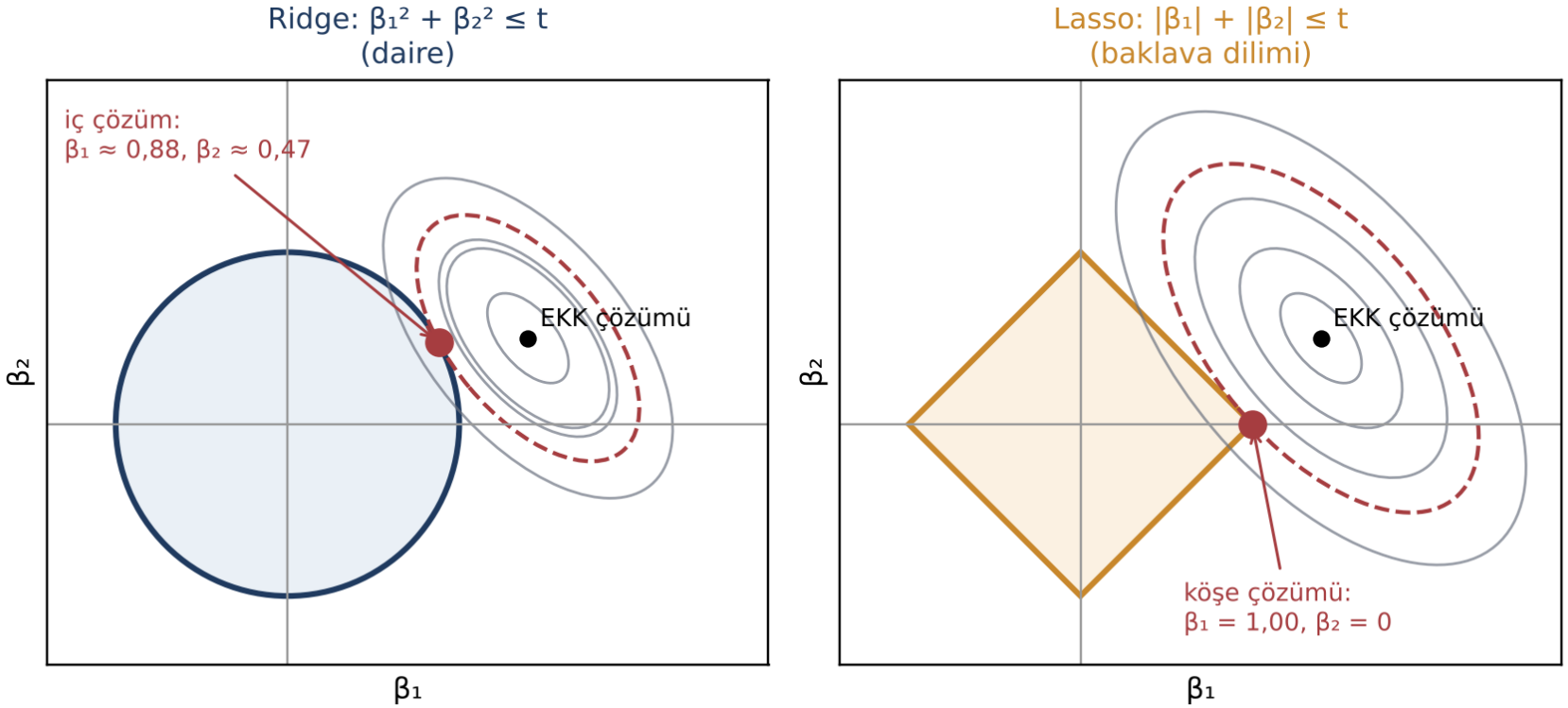
Ceza yazmanın eşdeğer yolu: katsayılara bir "bütçe" vermek

Tüketici teorisi	Ridge ve Lasso
Faydayı maksimize et	Hatayı minimize et
Bütçe kısıtı altında	Katsayı "bütçesi" altında: büyüklükleri t'yi aşamaz
Kayıtsızlık eğrileri	Eş hata eğrileri (elipsler)
Bütçe doğrusuna teğet nokta	Kısıt bölgesine teğet nokta
Köşe çözümü: bir maldan hiç alınmaz	**Lasso: bir değişkenin katsayısı tam sıfır**

Fark: tüketici faydayı artırmak ister; burada en iyi nokta elipslerin merkezi (EKK) ve kısıt bizi oradan uzakta tutar.

Daire ve baklava dilimi

İki katsayılı bir model: renkli bölge katsayı bütçesi, elipsler eş hata eğrileri



Ridge: daire pürüzsüz; teğetlik eksen dışında, iki katsayı da ≠ 0.

Lasso: sivri köşeler eksenlerde; teğetlik köşede, bir katsayı = 0.

Lasso’nun sıfır katsayıları bir köşe çözümüdür.

Tüketicinin bir maldan hiç almaması gibi, Lasso da bir değişkene hiç ağırlık vermez.

3

Ridge

Tüm katsayıları küçültmek

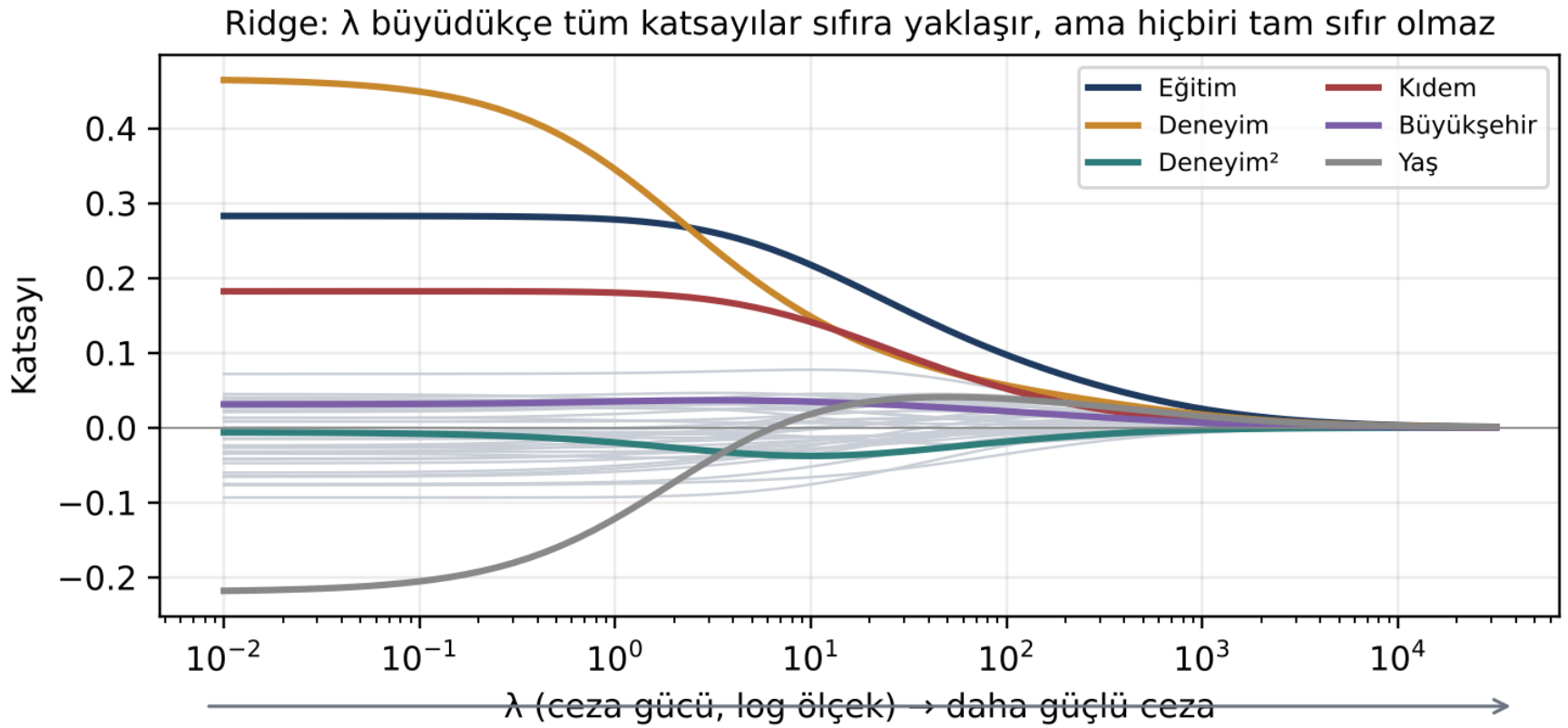
Ridge ne yapar?

$$\min \ \Sigma (y - \hat{y})^2 + \lambda \Sigma \beta^2$$

- Kare ceza büyük katsayıları çok, küçükleri az cezalandırır.
- Bu yüzden büyük ve oynak katsayıları törpüler.
- Bir katsayıyı tam sıfıra indirmeye “değmez”: sıfıra yaklaştıkça ceza zaten önemsizleşir.
- Sonuç: 40 değişkenin 40’ı da modelde kalır, ama hepsi küçülmüştür.

Ridge katsayı yolu

Soldan sağa λ artıyor. Renkli: 5 gerçek etken ve yaş. Gri: 34 alakasız anket sorusu.



Hepsi birlikte sıfıra yaklaşıyor; hiçbiri sıfıra ulaşmıyor.

Ridge’in ilk görevi: çoklu doğrusal bağlantı

Ekonometri kitaplarında “sırt regresyonu” adıyla geçer

EKK

Yaş ve deneyim neredeyse aynı. Katsayılar birbirini telafi eder: bir örneklemde deneyim +0,5 yaş −0,3, ötekinde tersi.

Ridge

Ceza aşırı uçlara izin vermez. Ağırlığı benzer değişkenler arasında paylaştırır: ikisi de orta büyüklükte, istikrarlı.

Ama dikkat: Ridge yaşa da bir katsayı veriyor. Öngörü için sorun değil; “yaşın ücrete etkisi var” demek ise yanlış olur.

Ridge ne zaman iyi çalışır?



Etki yayılmışsa

Birçok değişkenin her birinin biraz katkısı varsa: çok sayıda fiyat göstergesinin küçük katkılarla belirlediği enflasyon.



Korelasyon güçlüyse

Değişkenler birbirine çok benziyorsa ağırlığı paylaştırır.



Değişken > gözlem ise

EKK'nın çözemediği durumda bile her zaman tek bir çözüm üretir.

Zayıf yanı: "Hangi değişkenler önemli?" sorusuna cevap vermez; herkes modelde kalır. Bu soruyu soran için Lasso var.

4

Lasso

Küçültmek ve seçmek

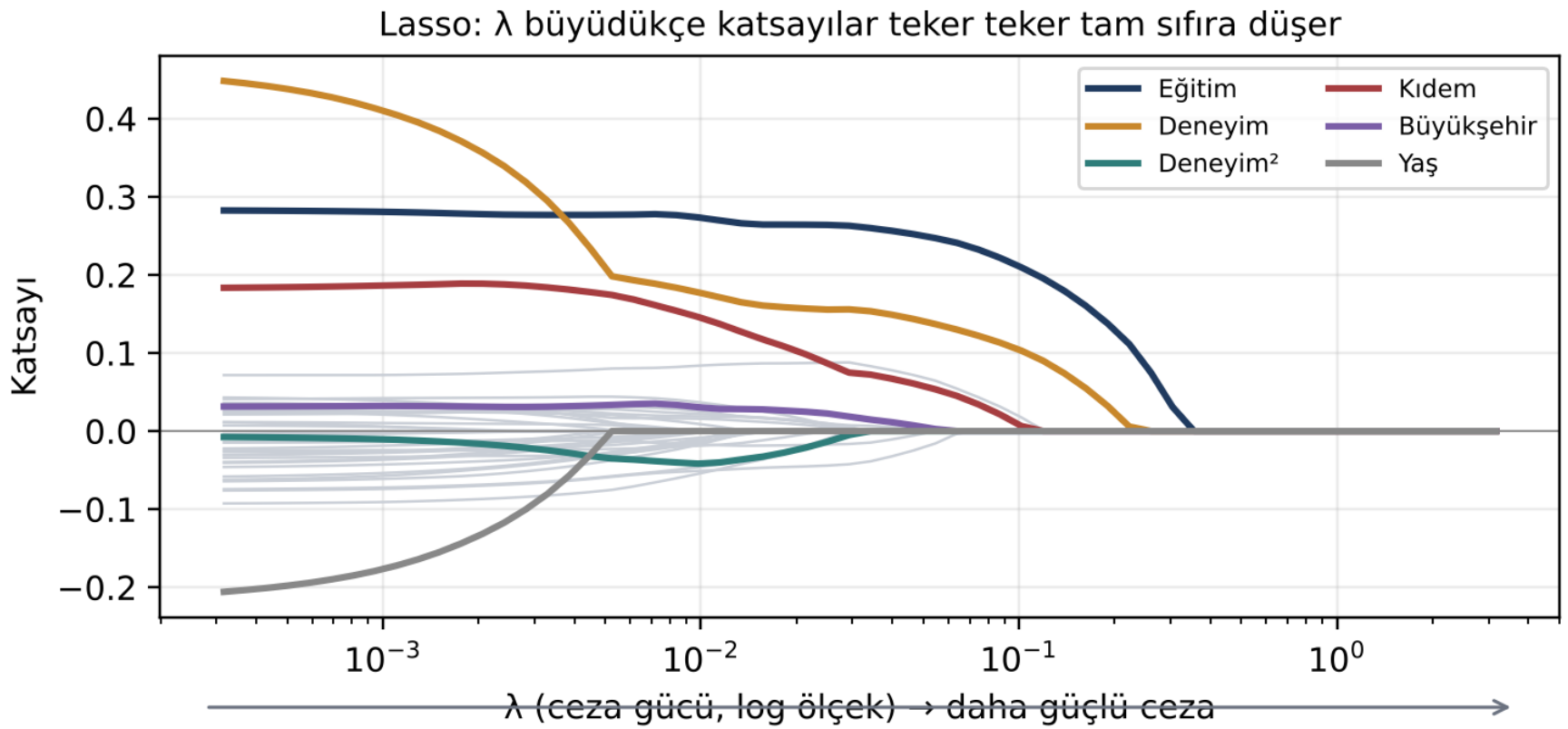
Lasso ne yapar?

$$\min \quad \Sigma\,(y - \hat{y})^2 \;+\; \lambda\,\Sigma\,|\beta|$$

- Mutlak değer cezası küçük katsayılar için de “pahalıdır”.
- Bir katsayıyı 0,01’den sıfıra indirmek bile ceza kazancı sağlar.
- Öngörüye yeterince katkı yapmayan değişkenlerin katsayısı tam sıfır olur.
- Sonuç: Lasso hem küçültür hem değişken seçer.

Lasso katsayı yolu

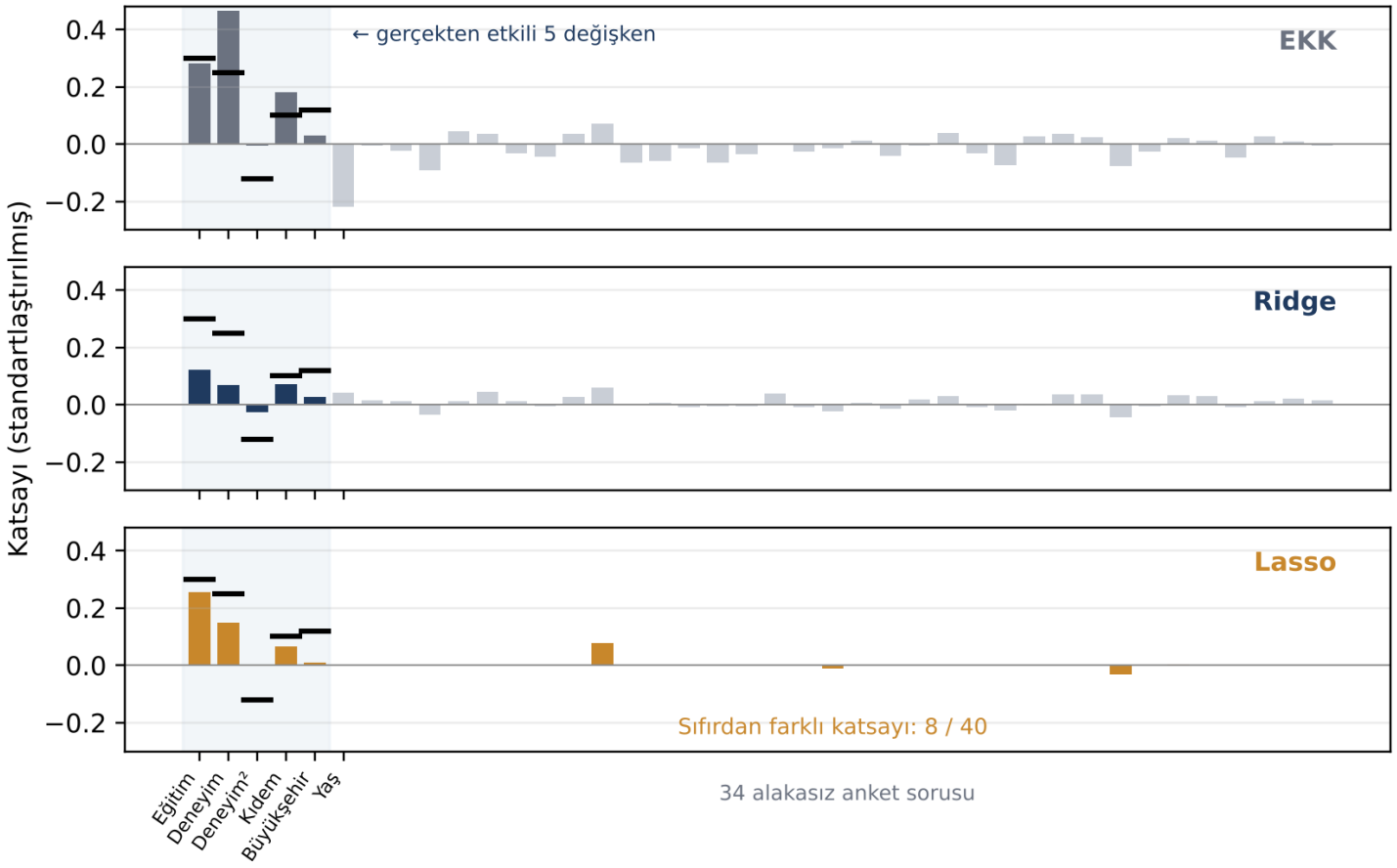
Aynı örneklem. λ arttıkça katsayılar teker teker tam sıfıra düşüyor.



Alakasız anket soruları (gri) genellikle önce elenir; en güçlü etkenler en son elenir.

Üç yöntem, aynı 100 kişi

Siyah çizgiler gerçek katsayılar (yalnızca ilk beş değişken sıfırdan farklı)



EKK

Alakasız değişkenlere de dikkate değer katsayılar veriyor.

Ridge

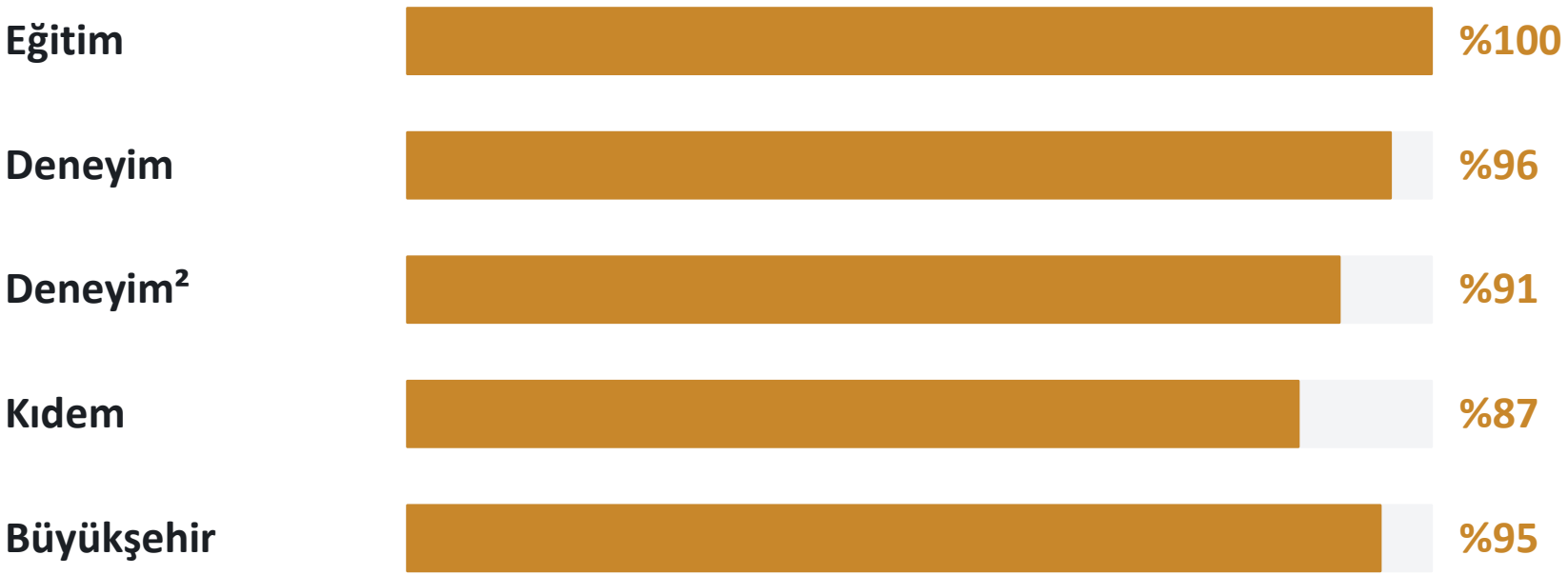
Hepsini küçültüyor, hepsini tutuyor.

Lasso

Çoğunu tam sıfır yapıyor; birkaç alakasızı da tutuyor.

Lasso ne kadar iyi seçiyor?

200 tekrarda gerçek etkenlerin modele girme oranı



11,3

alakasız değişken de her seferinde ortalama modele girdi

Lasso mükemmel bir değişken seçici değildir: λ öngörü hatasını en küçük yapmak için seçilir, “doğru” değişken kümesini bulmak için değil.

Benzer iki değişken: hangisini seçer?

Deneyim ve yaş; her korelasyon düzeyi için 150 tekrar

Deneyim–yaş korelasyonu	Yalnızca deneyim	Yalnızca yaş	İkisi birden
0,97	%59	%7	%35
0,995 (neredeyse aynı)	%50	**%23**	%27

Küçük bir fark varken

Lasso çoğunlukla gerçek değişkeni buluyor: kör bir yazı tura atmıyor.

Neredeyse aynı olunca

Seçim istikrarsızlaşıyor: örneklemlerin yaklaşık dörtte birinde deneyimi eleyip yaşı tutuyor.

Lasso’nun bir değişkeni seçmemesi, o değişkenin önemsiz olduğu anlamına gelmez.

Başka bir değişken aynı bilgiyi zaten taşıyor olabilir. Hafta 1’deki Mullainathan ve Spiess bulgusunun mekanizması bu.

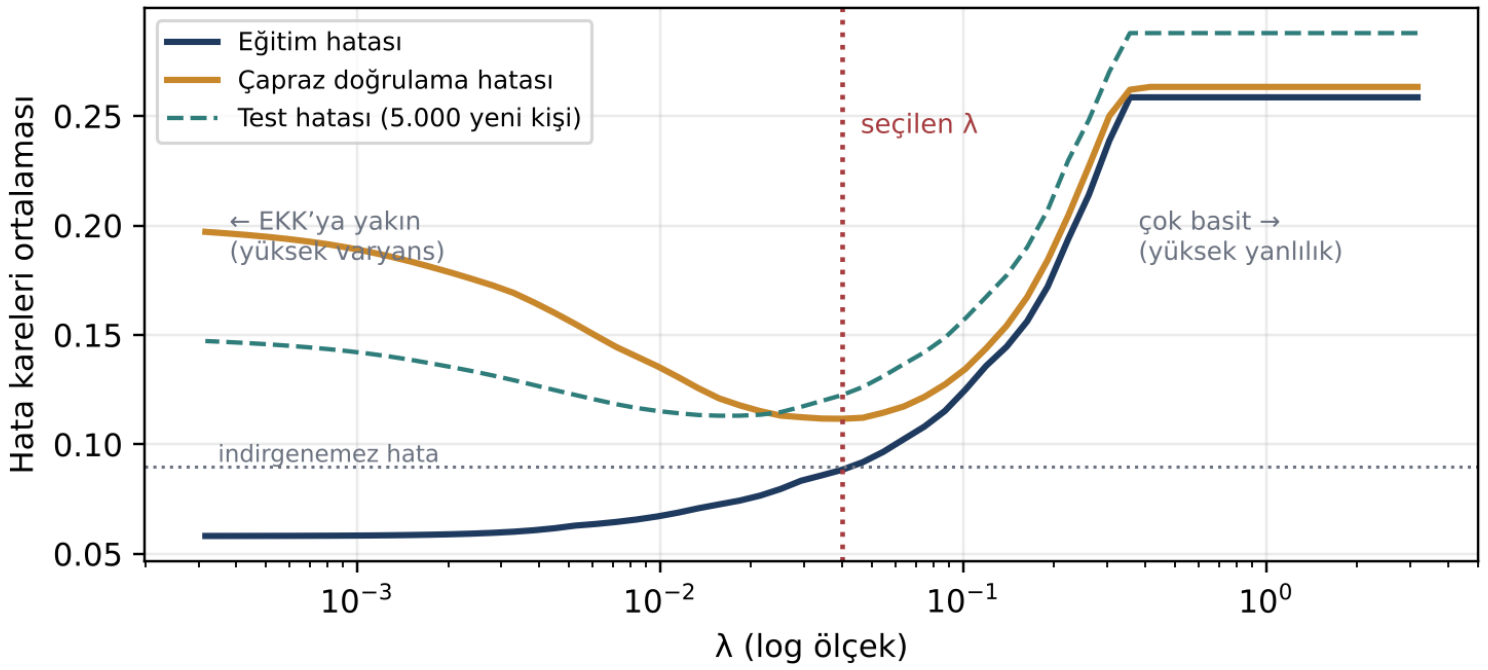
5

λ'yı seçmek ve ölçeklemek

Hafta 2 ve 3'e verilen sözler

λ çapraz doğrulamayla seçilir

Hafta 3'teki tarif aynen: λ ızgarası → her biri için 5-katlı çapraz doğrulama → en düşüğü seç



U eğrisinin aynası

Sol: ceza zayıf, EKK'ya yakın, yüksek varyans. Sağ: ceza güçlü, çok basit, yüksek yanlılık.

Hafta 3'ün kuralları

Test verisi λ seçiminde kullanılmaz. Zaman serisinde λ genişleyen pencereyle seçilir.

Önce standartlaştır

Hafta 2: “Büyük katsayıları cezalandıran yöntemlerde ölçekleme gerekir; öğrendikçe göreceğiz.” İşte o yöntemler.

EKK

- Geliri TL yerine bin TL ile ölç: katsayı 1.000 kat büyür.
- Öngörüler değişmez. Birim önemsiz.

Ridge ve Lasso

- Ceza katsayının büyüklüğüne uygulanır.
- Birim değişince ceza değişir: model gelir değişkenini sırf ölçü birimi yüzünden “önemsiz” bulabilir.

Çözüm: z = (x − ortalama) / standart sapma. Ortalama ve standart sapma yalnızca eğitim verisinden. Sabit terim cezalandırılmaz.

6

Sonuçlar ve uyarılar

Ne kazandık, neye dikkat etmeliyiz?

Üç yöntemin sonucu

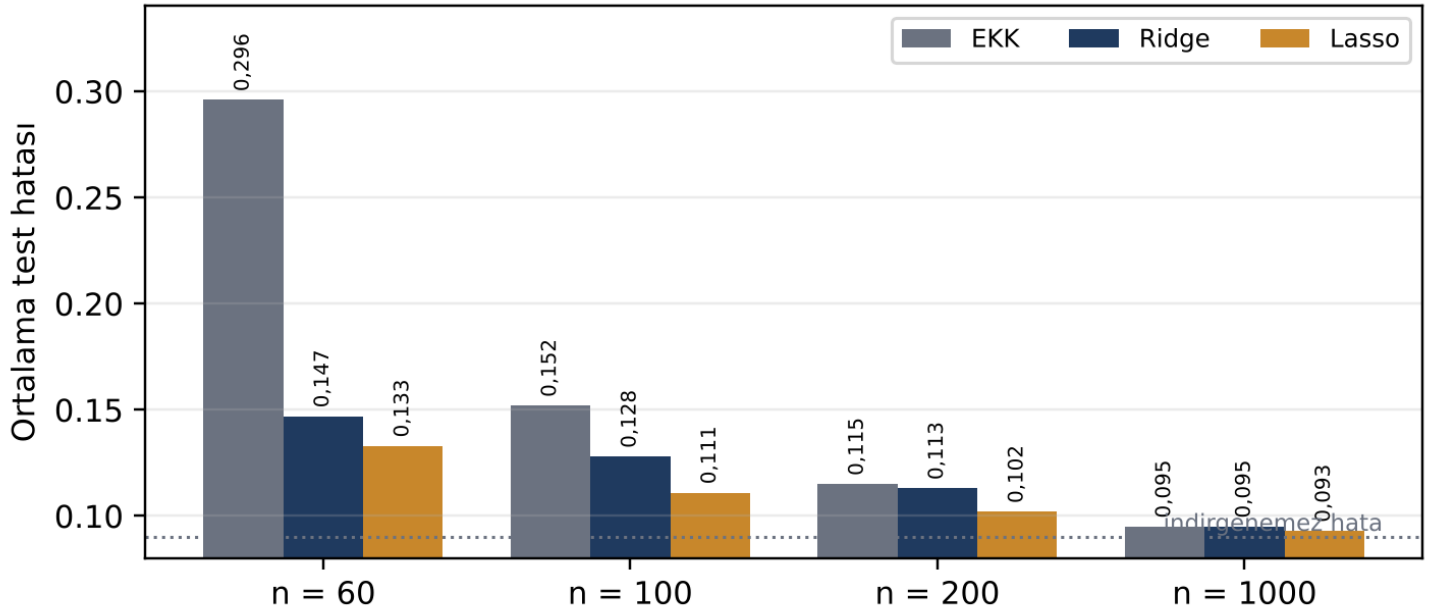
40 değişken, 100 kişi, 200 tekrar; test 5.000 yeni kişide

Model	Ortalama eğitim hatası	Ortalama test hatası
EKK	**0,053**	**0,156**
Ridge	0,072	0,130
Lasso	0,075	**0,113**
İndirgenemez hata	—	0,090

Eğitim hatasına bakan EKK'yı seçer. Ridge ve Lasso eğitimde biraz kaybediyor, testte belirgin biçimde kazanıyor.

Kazanç ne zaman büyüktür?

Aynı dünyada gözlem sayısını değiştirelim (40 değişken)



60 gözlem

EKK 0,296, Lasso 0,133. Fark çok büyük.

1.000 gözlem

EKK 0,095, Lasso 0,093. Fark neredeyse yok.

Düzenlileştirme "az veri, çok değişken" sorununun ilacıdır. Makro iktisatta tipik durum: 20–30 yıllık çeyreklik veri, yüzlerce gösterge.

Ceza kesilmiş katsayılar nedensel etki değildir



Ridge ve Lasso öngörü (ŷ) sorusu için tasarlandı. Katsayılarını nedensel bir etkinin kestirimi gibi okumak üç hata içerir:



Büyüklük yanlış

Ceza katsayıyı sıfıra doğru çekti. “Eğitimin getirisi” olarak raporlanamaz.



Seçim istikrarsız olabilir

Değişkenler çok benzerse hangisinin seçileceği örneklemden örnekleme değişir.



Seçim nedensellik değil

Seçilmek öngörüye katkı demektir; hedefi etkilemek değil. Şemsiye ve yağmur dansı (Hafta 1).

Alıştırma: EKK mı, Ridge mi, Lasso mu?

Gerekçesiyle cevap verin

1. 5.000 kişi, 6 kontrol değişkeni; eğitimin ücrete etkisini (katsayıyı) yorumlamak
2. 30 gösterge ile enflasyon öngörüsü; hepsinin biraz katkısı var, güçlü korelasyon
3. 200 gösterge ile büyüme öngörüsü; birkaçının önemli olduğundan şüpheleniyoruz
4. 80 ülke, 150 gösterge ile büyüme öngörüsü
5. Lasso yaşı seçip deneyimi eledi: “Ücreti deneyim değil yaş belirliyor”
6. 10.000 gözlem, 8 değişken; amaç öngörü

Cevaplar

Katsayı yorumu → EKK; az veri, çok değişken → ceza; bol veri → fark kalmaz

1	5.000 kişi, 6 kontrol değişkeni; eğitimin ücrete etkisini (katsayıyı) yorumlamak	**EKK**
2	30 gösterge ile enflasyon öngörüsü; hepsinin biraz katkısı var, güçlü korelasyon	**Ridge**
3	200 gösterge ile büyüme öngörüsü; birkaçının önemli olduğundan şüpheleniyoruz	**Lasso**
4	80 ülke, 150 gösterge ile büyüme öngörüsü	**Ridge ya da Lasso**
5	Lasso yaşı seçip deneyimi eledi: “Ücreti deneyim değil yaş belirliyor”	**Yanlış çıkarım**
6	10.000 gözlem, 8 değişken; amaç öngörü	**Fark etmez**

Bugünün beş mesajı

1. Çok değişken, az gözlem: EKK'nın öngörüleri yüksek varyanslı; değişken > gözlem ise çözülemez.
2. Düzenlileştirme: hata + λ × ceza. λ çapraz doğrulamayla seçilen bir hiperparametre.
3. Ridge hepsini küçültür, hiçbirini sıfırlamaz. Lasso bazılarını tam sıfır yapar: köşe çözümü.
4. Önce standartlaştır: ceza ölçü birimine göre adaletsiz uygulanmasın.
5. Ceza kesilmiş katsayılar yanlıdır: bu yöntemler öngörü içindir, nedensel yorum için değil.