V.C. 14.381 Ders notları[1]

# 1. Regresyonun (Bağlaşımın) Temelleri

**1.1. Regresyon ve Koşullu Beklenti Fonksiyonu.** Varsayalım $y_t$ gerçek bir tepki (response) değişkeni ve $w_t$ birlikte değişkenlerin (covariates) bir d-vektörü olsun. Burada ilgilendiğimiz, $y_t$'nin $w_t$ verildiğindeki koşullu ortalamasıdır (beklentisidir):

$$g(w_t) := E[y_t|w_t].$$

Alışılageldiği üzere, bir regresyon denklemini de şu şekilde tanımlayabiliriz:

$$y_t = g(w_t) + \varepsilon_t, \quad E[\varepsilon_t|w_t] = 0,$$

burada $y_t$ bir bağımlı değişken olarak, $w_t$ bağımsız değişkenler olarak ve $\varepsilon_t$ bozukluklar olarak düşünülmektedir.

Regresyon fonksiyonu, kayıpların karesi değerlendirilerek en iyi koşullu öndeyi (conditional prediction) probleminin çözümü olarak da tanımlanabilir: her bir $w$ için şunu elde edebiliriz

$$g(w) = \arg\min_{\tilde{g}\in\mathbb{R}} E[(y_t - \tilde{g})^2|w].$$

Böyle olunca, koşullu ortalama fonksiyonu koşulsuz öndeyi (unconditional prediction) problemini de çözmektedir:

$$g(\cdot) = \arg\min_{\tilde{g}(\cdot)\in\mathcal{G}} E[(y_t - \tilde{g}(w_t))^2],$$

burada argmin, $w$'nin ölçülebilir tüm fonksiyonlarının sınıfı olan $\mathcal{G}$ üzerinden hesaplanmaktadır. Bu düzenleme ne tahmine ne de hesaplamaya kolaylıkla çevrilebilmektedir.

Neticede, bu derste öğreneceklerimiz şunlardır

- önce, $g(w_t)$'nin $\beta \in \mathbb{R}^K$ ve orijinal bağlaştıranın (regressor) dönüşümleri olan $x_t$ için $x_t'\beta$ ile yaklaştırması,

  $$x_t = f(w_t),$$

  burada dönüşümün seçimleri olan $f$ yaklaştırım (approximation) teorisine bağlı olarak ifade edilmektedir,

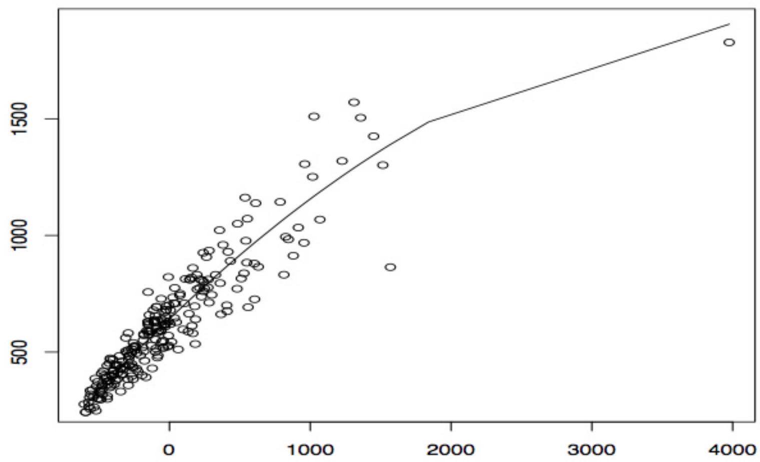
[1]Dersin orijinal notlarındaki dipnot: Bu notlar taslak halindedir. Hata bulduğunuzda lütfen bana e-mail ile haber veriniz (vchern@mit.edu).

- sonra, veriyi mümkün oluğu kadar iyi kullanarak $x_t'\beta'$yı tahmin etmek ve $x_t'\beta$ ve ilgili miktarlar hakkında küçük-örneklem ve büyük-örneklemlere yönelik çıkarsamalarda bulunmak.

**Örnek 1**: Engel (1857) çalışmasında, $y_t$ hanehalkının gıda harcaması ve $w_t$ hanehalkı geliri olarak verilmiştir. İyi bir yaklaştırma olarak şöyle bir kuvvet serisi kullanılabilir: $g(w_t) = \beta_0 + \beta_1 w_t + \beta_2 w_t^2$. Engel aslında Haar serilerini kullanmıştır (yaklaştırma birçok kukla terim kullanılarak yapılmıştır). Özellikle ilgimizi çeken, gelirin gıda tüketimi üzerindeki etkisidir

$$\frac{\partial g(w)}{\partial w}.$$

Gıda Harcaması




Gelir

ŞEKİL 1

**Örnek 2**: Varsayalım, $y_t^*$ doğumdaki ağırlık ve $w_t$ sigara içme ya da tıbbi bakımın kalitesi olsun. Açıktır ki, $E[y_t^*|w_t]$ ilgi çekici bir hedeftir. Varsayalım, $w_t$'nin çok düşük doğum ağırlıkları üzerindeki etkisiyle ilgilenelim; bunu yapmanın bir yolu olarak şunu tanımlayabiliriz

$$y_t := 1\{y_t^* < c\},$$

burada $c$ doğum ağırlığının kritik bir seviyesidir. Sonrasında şunu inceleyebiliriz

$$g(w_t) = E[y_t|w_t] = P[y_t^* < c|w_t].$$

Bu regresyon fonksiyonu uçdeğer (extreme) doğum ağırlığı gerçekleşmesi ihtimalinin birlikte değişkenlere bağlılığını ölçmektedir.[2]

Varsayalım $E[y_t|w_t]$'i elde edebildik. Bundan nasıl faydalanabiliriz?

Düşünce akımları:

**1. Betimsel.** İlginç biçimlendiren olguları (stylized facts) ortaya çıkarmak. Sigara içmek ortalama doğum ağırlığını "azaltır" (ancak uçdeğer doğum ağırlığının gerçekleşmesini azaltmaz).

**2. İşlem Etkileri.** Nedensel etkiler hakkında çıkarsamada bulunmak için ideal bir durum tam kontrol edilen rastsal deneme olarak düşünülür. Tam kontrol edilen rastsal deneme mevcut değilse, doğal deneyler de aranabilir.

**3. Yapısal Etkiler.** Ekonomik (nedensel, yapısal) bir modelin parametrelerini tahmin edelim. Ekonomi bilgisini kullanarak $E[y_t|x_t]$'nin neden ekonomik parametreleri özdeşleştirebileceğini ispat edelim. Varian'ın ilgili bölümüne bakınız.

## 1.2. **Kitlede ve Sonlu Örneklemde OLS (sıradan en küçük kareler, Ordinary Least Squares).**

### 1.2.1. *EDÖ (en iyi doğrusal öndeyici) Özelliği.*

Kitlede en küçük karelerin $\beta$'sı aşağıdaki terimin enküçültücüsüdür (argmin)

$$Q(b) = E[y_t - x_t'b]^2,$$

dolayısıyla $x_t'\beta$, $y_t$'nin kitlede kare kayıp altında en iyi doğrusal öndeyicisidir. Sonlu örneklemlerde en küçük karelerin $\hat{\beta}'$ı aşağıdaki terimin en küçültücüsüdür

$$Q_n(\beta) = E_n[y_t - x_t'\beta]^2,$$

burada $E_n$ görgül beklentidir ($\frac{1}{n}\sum_{t=1}^{n}$'in kısaltması). Yani, $x_t'\hat{\beta}$, $y_t$'nin örneklemde kare kayıp altında en iyi doğrusal öndeyicisidir.

$\beta$ için sarih bir çözüm de belirtebiliriz. Dikkat ederseniz $\beta$, $E[x_t x_t']$ tam kerteli (full rank) olduğunda, şu birinci derece koşulu çözmektedir

$$E[x_t(y_t - x_t'\beta)] = 0, \text{ yani } \beta = E[x_t x_t']^{-1}E[x_t y_t],$$

[2]Diğer bir yöntem ise $y_t^*$'ın bölütlerine (quantile) $w_t$'nin bir fonksiyonu olarak bakmaktır, bu da bölüt regresyonunun gerçekleştirdiğidir.

Örneklemde en küçük kareler, $E_n[x_t x_t']$ tam kerteli olduğunda, kitle beklemlerini (moments) görgül beklemler ile değiştirmektedir:

$$E_n[x_t(y_t - x_t'\hat{\beta})] = 0, \text{ yani } \hat{\beta} = E_n[x_t x_t']^{-1} E_n[x_t y_t].$$

1.2.2. *En İyi Doğrusal Yaklaştırma olarak OLS.* Şunu gözlemleyebiliriz

$$Q(b) = E[y_t - x_t'b]^2 = E[E[y_t|w_t] - x_t'b + \varepsilon_t]^2 = E[E[y_t|w_t] - x_t'b]^2 + E[\varepsilon_t]^2,$$

burada $\varepsilon_t = y_t - E[y_t|w_t]$'dır. Yani $\beta$ şunu çözmektedir

$$\min_b E(E[y_t|w_t] - x_t'b)^2,$$

ve $x_t'\beta$, koşullu ortalama fonksiyonu $E[y_t|w_t]$ için en iyi doğrusal yaklaştırmadır. Bu da yaklaşıklık teorisine bir bağlantı oluşturmaktadır.

1.2.3. *İşlevsel Biçimleri (Functional Form) Oluşturmak.* Yaklaşıklık teorisine verilecek bağlantı, yaklaşıklık teorisinin regresyonlarımız için iyi işlevsel (fonksiyonel) biçimler oluşturmada kullanılabilmesi açısından faydalıdır. Burada, yaklaşıklık tasarılarından ekonometrik uygulamalarda en çok yararlı olacaklar üzerine odaklanacağız.[3]

1. Kama (Spline) Yaklaşıklığı: Varsayalım $w$ tek bağlaştıran olsun. Sonrasında doğrusal kama (1'inci dereceden kama), sonlu sayıda eşit mesafeli bağlantılar (knots) olan $k_1, k_2, ..., k_r$ ile şu şekilde ifade edilebilir:

$$x_t = f(w_t) = (1, w_t, (w_t - k_1)_+, ..., (w_t - k_r)_+)'$$

burada $(u_t)_+$ terimi $u$ x $1(u > 0)$'ı ifade eder. Kübik kama şu şekilde ifade edilebilir:

$$x_t = f(w_t) = (1, (w_t, w_t^2, w_t^3), (w_t - k_1)_+^3, ..., (w_t - k_r)_+^3)'.$$

Kamaları tanımlarken $x_t$'nin boyutu olan $K'$yi kontrol edebiliriz. Kamalar kullanılarak oluşturulan fonksiyon $w \mapsto f(w)'b$'nın, herhangi $b$ için, $w$'da iki kez türevi alınabilir.

2. Güç (Power) Yaklaşıklığı: Varsayalım $w$ tek bağlaştıran (regressor) olsun ve $[0, 1]$'de dönüştürülmüş olsun; sonrasında, $r$-inci dereceden çokterimli serileri şu şekilde ifade edilebilir:

$$x_t = f(w_t) = (1, w_t, ..., w_t^r)'.$$

Chebyshev çokterimlileri çoğu kez basit çokterimliler yerine kullanılmaktadır. Varsayalım $w_t$ $[-1, 1]$ arasında değerler alacak şekilde dönüştürülmüş olsun. Chebyshev çokterimlileri

[3] W. Newey'in çalışması tahmin açısından faydalı bir uygulama sunmaktadır. Yaklaşıklık yöntemlerine faydalı bir başlangıç için K. Judd'ın kitabına bakınız.

şu şekilde oluşturulabilir

$$x_t = f(w_t) = (cos(j \cdot cos^{-1}(w_t)), j = 0, ..., r)$$

(Bunlara çokterimli denme sebebi şudur: $f(w_t) = (1, w_t, 2w^2 - 1, 4w_t^3 - 3w_t, ..., )$, ve böylece aslında $w_t$ de çokterimlilerdir.)[4]

3. Dalgacık (Wavelet) Yaklaşıklığı: Varsayalım $w$ tek bağlaştıran olsun ve $[0, 1]$'de dönüştürülmüş olsun; sonrasında, $r$-inci dereceden dalgacık serisi şu şekilde ifade edilebilir:

$$x_t = f(w_t) = (e^{j(i2\pi w_t)}, j = 0, ..., r),$$

veya sinüs ve kosinüs tabanları ayrı ayrı kullanılabilir.

Çoklu bağlaştıranların olduğu durum da benzer şekilde incelenebilir: Varsayalım temel (basic) bağlaştıranlar şunlar olsun $w_{1t}, ...w_{dt}$, sonrasında her bir temel bağlaştıran için $d$ serileri oluşturulabilir. Sonrasında $d$ serilerinin tüm etkileşimleri, bunlara geren (tensor) çarpımları denir, oluşturulabilir ve bunların hepsi bağlaştıran vektörü $x_t$ içinde toplanabilir. Eğer temel bir bağlaştıran için herbir seri $J$ terime sahipse, sonuç (final) regresyonu $K \approx J^d$ boyutundadır, bu da $d$ boyutunda üstel olarak patlamaktadır (boyutsallık belasının (curse of dimensionality) tezahürü). Geren çarpımlarının düzgün bir tanımı için bkz. Newey.

**Teorem 1.1.** *Varsayalım $w_t$'nin $\mathbb{R}^d$'de tanımlı bir küpte sınırlı bir desteği (bounded support) ve pozitif, sınırlı bir yoğunluğu olsun. Eğer $g(w)$'nin sınırlı türevlerle (sabit bir $M$ ile) s-kere sürekli türevselliği varsa, yukarıda belirtilen şekilde $K$-terimli $x = f(w)$ serileri kullanılarak, yaklaşıklık hatası şu şekilde kontrol edilebilir:*

$$\min_b [E[g(w_t) - x_t' b]^2]^{1/2} \leq \ const_M \cdot K^{-\gamma/d},$$

*burada kuvvet serileri ve dalgacıklar için $\gamma = s$ ve kamalar için $\gamma = \min(s, r)$ kullanılmıştır, $r$ ise kamanın derecesidir.*

Sonuçta, $K \to \infty$ iken, yaklaşıklık hatası sıfıra yakınsamaktadır. Teorem aynı zamanda düzgün (smooth) bir fonksiyonun yaklaşımının daha kolay olduğunu ve birçok temel bağlaştıranlı bir fonksiyonun yaklaşımının daha zor olduğunu söylemektedir (boyutsallık belasının başka bir tezahürü). Dikkat çeken başka bir husus da, $K \to \infty$ iken, yaklaşıklık hatası sıfıra yakınsamaktadır ifadesinin

[4] Daha fazla detay için K. Judd'ın kitabına bakınız; ayrıca şuraya da bakabilirsiniz http://mathworld.wolfram.com/

*düzgünlük (smoothness)* varsayımları olmadan dahi geçerli olduğudur; aslında $g(w)$'nin ölçülebilir olması ve kare (square) tümlevlenebilirliği (integrable) yeterlidir.

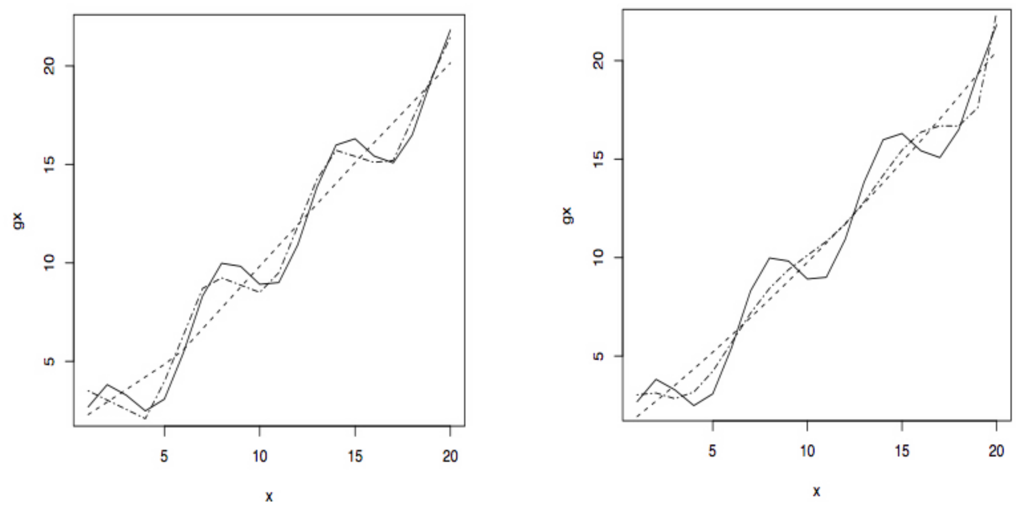
En küçük kareler kullanılarak fonksiyonların yaklaşıklığında kamaların ve Chebyshev serilerinin kullanımının olumlu özellikleri vardır, bu sadece ortalama kare yaklaşıklık hatasını enküçük yapmak için değil, aynı zamanda yaklaşılan terime (approximand) ençok mesafeyi elde etmede de geçerlidir (son özelliğe ortak-enküçüklük denmektedir (co-minimality)).

1.3. **Yaklaşıklık Örnekleri.** Örnek 1. (Yapay, synthetic) Varsayalım şu fonksiyon verilmiş olsun $g(w) = w + 2sin(w)$, ve $w_t$ tekdüze (uniform) şekilde şu tamsayılar üzerinde dağılmış olsun $\{1, ..., 20\}$. Bu durumda OLS kitlede şu yaklaşıklık problemini çözer:

$$\beta = \arg\min_b E[g(w_t) - x_t'b]^2$$

burada $x_t = f(w_t)$'dir. Şimdi $f$ için farklı işlevsel biçimleri deneyelim. Bu örnekte, $f(w)$'yi şu şekilde oluşturuyoruz (a) doğrusal kama (Şekil 2,sol) ve (b) Chebyshev serileri (Şekil 2,sağ), bu durumda $f(w)$'nin boyutları ya 3 ya da 8'dir.

Doğrusal kama ile yaklaştırma, K=3 ve 8   Chebyshev serileri ile yaklaştırma, K=3 ve 8




ŞEKİL 2

Sonrasında ise $g(w)$ fonksiyonunu grafik yardımıyla doğrusal yaklaşıklık $f(w)'\beta$ ile karşılaştıralım. Şekil 2'de $K = 3$ olan tutumlu (parsimonious) modelin küresel şekli (global shape) koşullu beklenti fonksiyonunda doğru olarak yaklaştırdığını ("büyük değişiklikler"), ancak yerel şekli doğru olarak yaklaştıramadığını ("küçük değişiklikler") görüyoruz. Daha esnek işlevsel biçimler $K = 8$ parametreleri ile kullanıldığında, yerel şekil için daha iyi bir yaklaşıklık elde ediliyor. Bu örnekte kamaların Chebyshev çokterimlilerinden çok daha iyi performans gösterdiklerini görüyoruz.

Yaklaşıklık hatasının düzgün ölçümlerine de bakabiliriz, mesela kök ortalama kare yaklaşıklık hatası gibi (KOKYH, root mean square approximation error, RMSAE):

$$[E[g(w_t) - f(w_t)'\beta]^2]^{1/2},$$

ve ençok yaklaşıklık hatası (EYH, maximum approximation error, MAE):

$$\max_{w} |g(w) - f(w)'\beta|.$$

Bu ölçüler aşağıdaki tabloda hesaplanmıştır:

| | kama $K = 3$ | kama $K = 8$ | Chebyshev $K = 3$ | Chebyshev $K = 8$ |
|---|---|---|---|---|
| KOKYH | 1.37 | 0.65 | 1.39 | 1.09 |
| EYH | 2.27 | 0.95 | 2.19 | 1.81 |

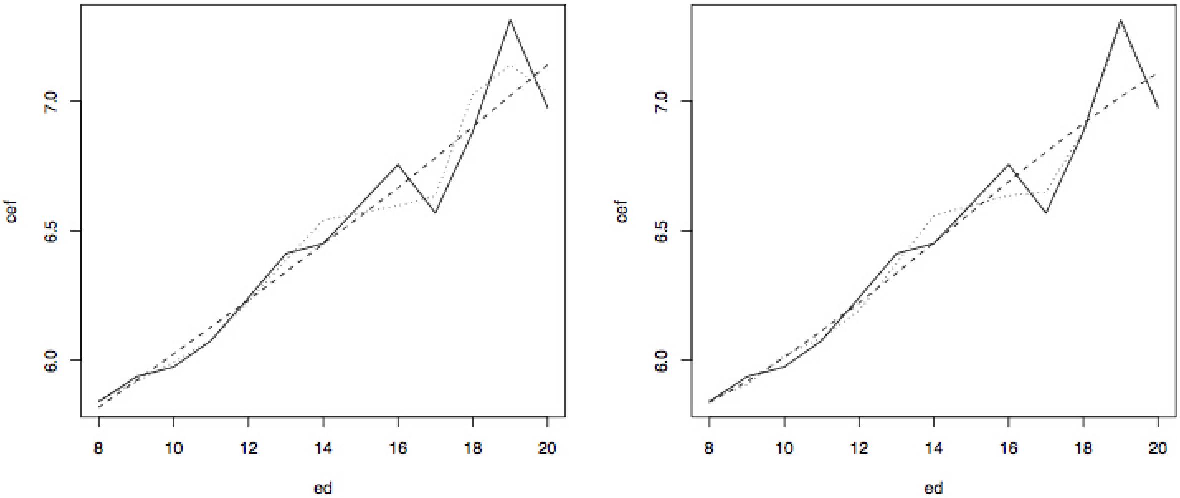
Örnek 2.(Reel) Burada $g(w)$ ücretin $(y)$ logaritmasının ortalamasıdır ve eğitim üzerine koşulludur

$$w \in \{8, 9, 10, 11, 12, 13, 14, 16, 17, 18, 19, 20\}.$$

$g(w)$ fonksiyonu kitle verisi kullanılarak hesaplanmıştır – kullanılan veriseti ABD 1990 Nüfus Sayımı'ndaki yetişkin erkeklerdir [5]. Burada bu fonksiyonun bağlaştıranlarını oluşturmak için yaygın yaklaştırma yöntemleri kullanıldığında, OLS tarafından ne kadar iyi yaklaştırıldığını bulmak istiyoruz. Problemi basitleştirmek için, $w_t$'nin tekdüze olarak dağıldığını varsayabiliriz (diğer durumda sıklık (frequency) ile ağırlıklandırabiliriz). Kitlede, OLS şu yaklaştırma problemini çözmektedir: $\min E[g(w_t) - x_t'b]^2$ burada $x_t = f(w_t)$'dır ve $f(w)$ şu şekilde oluşturulur (a) doğrusal kama (Şekil 3,sol) ve (b) Chebyshev serileri (Şekil 3,sağ), bu durumda $f(w)$'nin boyutları ya $K = 3$ ya da $K = 8$'dir.

[5] Daha fazla detay için bkz. Angrist, Chernozhukov, Fernandez-Val, 2006, Econometrica, çalışması.

Doğrusal kama ile yaklaştırma, K=3 ve 8 Chebyshev serileri ile yaklaştırma, K=3 ve 8




ŞEKİL 3

Sonrasında, grafiksel olarak $g(w)$ fonksiyonunu $f(w)'\beta$ doğrusal yaklaştırması ile karşılaştırıyoruz. Aynı zamanda KOKYH ve EYH'yi kaydediyoruz. Yaklaştırma hataları aşağıdaki tabloda verilmiştir:

| | kama $K=3$ | kama $K=8$ | Chebyshev $K=3$ | Chebyshev $K=8$ |
|---|---|---|---|---|
| KOKYH | 0.12 | 0.08 | 0.12 | 0.05 |
| EYH | 0.29 | 0.17 | 0.30 | 0.12 |

Referanslar:

1. Newey, Whitney K. Convergence rates and asymptotic normality for series estimators. J. Econometrics 79 (1997), no. 1, 147–168. (Regresyon kamalarının tanımları bu referansı takip etmektedir).

2. Judd, Kenneth L. Numerical Methods in Economics. MIT Press, Cambridge, MA, 1998. (Bölüm: Approximation Methods. Bu oldukça ileri seviye bir referanstır, ancak birçok ekonometrik ve ekonometrik olmayan uygulama için faydalıdır.)

3. Hal R. Varian. Microeconomic Analysis (Bölüm: Econometrics. Bu kitap Yapısal Ekonometri (Structural Econometrics) hakkında çok iyi bir okumadır ve içinde birçok ilginç fikir mevcuttur).

Bahsedilen materyaller şu adrese eklenmiştir: stellar.mit.edu.

## 2. Regresyon Cebiri (Regression Calculus)

**2.1. Matris (Dizey, Matrix) Hesaplamaları.** Aşağıdaki hesaplamalar sonlu örneklem çıkarsamaları için faydalıdır. $Y = (y_1, ..., y_n)'$ olsun ve $X$ de $n \times K$ boyutlu bir matris olsun: bu matrisin satırları şunlardır $x_t'$, $t = 1, ..., n$. Bu gösterim ile şu ifadeyi yazabiliriz

$$\hat{\beta} = \arg\min_{b \in \mathbb{R}^k} (Y - Xb)'(Y - Xb),$$

Eğer kerte$(X) = K$ ise, yukarıdaki ifadenin Hessian'ı olan $2X'X$ pozitif tanımlıdır; bu da katı dışbükeyliği sağlar ve çözümün tekil (unique) olduğunu ifade eder. $\hat{\beta}$'nın çözümü normal denklemler (normal equations) denen birinci derece koşulları (first-order conditions, foc) tarafından belirlenir:

$$X'(Y - X\beta) = 0. \tag{1}$$

Bu denklemler çözülünce şu ifade elde edilir

$$\hat{\beta} = (X'X)^{-1}X'Y.$$

Uyumlu (fitted) veya öndeyilen (predicted) değerler şu vektörler (yöney, vector) ile ifade edilir

$$\widehat{Y} := X\hat{\beta} = X(X'X)^{-1}X'Y = P_X Y,$$

Kalıntılar (residual) da şu şekilde tanımlanır:

$$\hat{e} := Y - X'\hat{\beta} = (I - P_X)Y = M_X Y.$$

**Geometrik yorumlama:** $L$ :=yayılma (span)$(X)$ $:= \{Xb : b \in \mathbb{R}^k\}$ $X$'in sütunları tarafından yayılan doğrusal uzay (linear space) olsun. $P_X$ matrisi izdüşüm matrisi (projection matrix) olarak adlandırılır, çünkü $Y$'nin $L$ üzerine izdüşümünü verir. $M_X$ matrisi ise $Y$'nin $L$'ye dikeysel (orthogonal) olan alt uzay (subspace) üzerine olan izdüşümünü veren izdüşüm matrisidir.

Aslında, LS (en küçük kareler tahmincisi) şu problemi çözer:

$$\min_{Y^* \in L} (Y - Y^*)'(Y - Y^*)$$

Çözüm olan $Y^* = \widehat{Y}$ ise Y'nin L üzerine olan dikeysel izdüşümüdür, yani

$$\text{(i) } \hat{Y} \in L \quad \text{(ii) } S'\widehat{e} = 0,\ \forall S \in L,\ \widehat{e} := (Y - \widehat{Y}) = M_X Y.$$

Bunu gözümüzde canlandırmak için şöyle bir basit örnek kullanalım: $n = 2$, tek boyutlu bağlaştıran olsun ve kesme (intercept) olmasın, yani şunlar geçerli olsun $Y = (y_1, y_2)' \in \mathbb{R}^2$, $X = (x_1, x_2)' \in \mathbb{R}^2$ ve $\beta \in \mathbb{R}$.

Dikkat etmemiz gereken bazı özellikler şunlardır:

1. Eğer regresyonda kesme varsa, yani $X$'in bir sütunu şu ise $\mathbf{1} = (1, .., 1)'$, bu takdirde $\bar{Y} = \bar{X}'\hat{\beta}$'dir. Regresyon doğrusu verinin ortalamalarının içinden geçmektedir. Buna eşdeğer olarak, $\mathbf{1} \in L$ olduğu için, $\mathbf{1}'\hat{e} = 0$ veya $\bar{\hat{e}} = 0$'dır.

2. İzdüşüm ("şapka" (hat)) matrisi $P_X = X(X'X)^{-1}X'$ simetriktir ($P_X' = P_X$) ve denkgüçlüdür (idempotent) ($P_X P_X = P_X$). $P_X$ bir dikey izdüşüm işlemcisidir ve $\mathbb{R}^n$'deki vektörleri $L$ ile eşleştirir (mapping). Özellikle, $P_X X = X$, $P_X \hat{e} = \mathbf{0}$, $P_X Y = X\hat{\beta}$'dır.

3. İzdüşüm matrisi $M_X = I - P_X$ de simetrik ve denkgüçlüdür. $M_X$ ise $\mathbb{R}^n$'deki vektörleri $L$'ye dikey olan doğrusal uzayla eşleştirir. Özellikle, $M_X Y = \widehat{e}$, $M_X X = 0$, ve $M_X \hat{e} = \hat{e}$'dır. Ayrıca dikkat ediniz, $M_X P_X = 0$'dır.

2.2. **Bölüntülü (Partitioned) veya Kısmi (Partial) Regresyon.** $X_1$ ve $X_2$, $X$'i şu şekilde bölüntülesinler

$$X = [X_1, X_2]$$

ve sonra bir regresyon modeli düşünelim

$$Y = X_1\beta_1 + X_2\beta_2 + \varepsilon.$$

Daha sonra ise, $P_{X_2} = X_2(X_2'X_2)^{-1}X_2'$ ve $M_{X_2} = I - P_{X_2}$ olsun. Ek olarak, $\hat{V}_1 = M_{X_2}X_1$ ve $\hat{U} = M_{X_2}Y$ olsun, yani $X_1$'in sütunlarının $X_2$ üzerine olan regresyonundan $\hat{V}_1$ kalıntı matrisi olarak elde edilsin ve $Y$'nin $X_2$ üzerine olan regresyonundan $\hat{U}$ kalıntı matrisi olarak elde edilsin.

**Teorem 2.1.** *Şu tahminciler eşdeğerlidir: 1. $Y$'nin $X_1$ ve $X_2$ üzerine regresyonundan elde edilen $(\hat{\beta}_1', \hat{\beta}_2')'$ vektörler tahmininin $\hat{\beta}_1$ bileşeni, 2. $Y$'nin $\hat{V}_1$ üzerine regresyonundan elde edilen $\tilde{\beta}_1$, 3. $\hat{U}$'in $\hat{V}_1$ üzerine regresyonundan elde edilen $\bar{\beta}_1$.*

**İspat**. Yukarıda denklem (1) de gösterilen olguyu hatırlayınız:

(2) $\hat{\gamma}$ , ancak ve ancak $Z'(Y - Z\hat{\gamma}) = 0$ ise, $Y$'nin $Z$ üzerindeki OLS'idir.

Şunu yazalım

$$Y = X_1\hat{\beta}_1 + X_2\hat{\beta}_2 + \hat{e} = \hat{V}_1\hat{\beta}_1 + \underbrace{(X_1 - \hat{V}_1)\hat{\beta}_1 + X_2\hat{\beta}_2 + \hat{e}}_{\hat{\eta}}.$$

Yukarıdaki olgu (2) kullanılarak $\hat{\beta}_1 = \bar{\beta}_1$ eşitliği ancak ve ancak $\hat{V}_1'\hat{\eta} = 0$ ise gerçekleşir. Son çıkarsama şuradan gelir: $\hat{V}_1 = M_{X_2}X_1$ $\in$yayılma$(X)$'den dolayı $\hat{V}_1'\hat{e} = 0$, $X_2'\hat{V}_1 = X_2'M_{X_2}X_1 = 0$ ve $(X_1 - \hat{V}_1)'\hat{V}_1 = (P_{X_2}X_1)'M_{X_2}X_1 = 0$.

$\hat{\beta}_1 = \bar{\beta}_1$ eşitliğini göstermek için, şöyle yazalım

$$M_{X_2}Y = M_{X_2}(X_1\hat{\beta}_1 + X_2\hat{\beta}_2 + \hat{e}),$$

bu da eşdeğer olarak şu şekilde belirtilebilir ($M_{X_2}X_2 = 0$ olduğuna dikkat ediniz)

$$\widehat{U} = \widehat{V}_1\hat{\beta}_1 + M_{X_2}\hat{e}.$$

$\widehat{V}_1'M_{X_2}\hat{e} = (M_{X_2}X_1)'(M_{X_2}\hat{e}) = X_1'M_{X_2}\hat{e} = \hat{V}_1'\hat{e} = 0$ olduğu için, şu sonuç elde edilir $\hat{\beta}_1 = \bar{\beta}_1$.

**Kısmi Regresyonun Uygulamaları:**
1. Yorum: $\hat{\beta}_1$ bir *kısmi* regresyon katsayısıdır.

2. Ortalamadan Fark Alma (De-meaning): Eğer $X_2 = \mathbf{1}$ ise, bu taktirde $M_{X_2} = I - \mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}' = I - \mathbf{1}\mathbf{1}'/n$ ve $M_{X_2}Y = Y - \mathbf{1}\bar{Y}$, $M_{X_2}X_1 = X_1 - \mathbf{1}\bar{X}_1'$'dır. Bu sebeple, $Y$'nin bir sabit $X_2 = \mathbf{1}$ ve diğer bağlaştıranların bir kümesi olan $X_1$ üzerine regresyonu şu şekilde elde edilenlerle aynı eğim katsayılarını verir: (1) $Y$'nin kendi ortalamasından sapmasının $X_1$'in kendi ortalamasından sapması üzerine regresyonu, veya (2) $Y$'nin $X_1$'in kendi ortalamasından olan sapması üzerine regresyonu.

3. Ayrı Regresyonlar: Eğer $X_1$ ve $X_2$ dikeysel ise, yani $X_1'X_2 = 0$, sonrasında $Y$'nin $X_1$ ve $X_2$ üzerine regresyonundan elde edilen $\hat{\beta}_1$, $Y$'nin $X_1$ üzerine regresyonundan elde edilen $\tilde{\beta}_1$'e eşdeğerdir. Bunu göstermek için şöyle yazalım $Y = X_1\hat{\beta}_1 + (X_2\hat{\beta}_2 + \hat{e})$ ve, dikkat ediniz, $X_1'X_2 = 0$ ve $X_1'\hat{e} = 0$ kullanılarak $X_1'(X_2\hat{\beta}_2 + \hat{e}) = 0$'dır. Olgu (2) kullanılarak, $\hat{\beta}_2 = \tilde{\beta}_2$ elde edilir.

4. Dışlanmış Değişken Yanlılığı (Omitted Variable Bias): Eğer $X_1$ ve $X_2$ dikeysel değil ise, yani $X_1'X_2 \neq 0$, sonrasında $Y$'nin $X_1$ ve $X_2$ üzerine regresyonundan elde edilen $\hat{\beta}_1$, $Y$'nin $X_1$ üzerine regresyonundan elde edilen $\tilde{\beta}_1$'e eşdeğer değildir. Ancak, şunu gösterebiliriz

$$\tilde{\beta}_1 = (X_1'X_1)^{-1}X_1'(X_1\hat{\beta}_1 + X_2\hat{\beta}_2 + \hat{e}) = \hat{\beta}_1 + (X_1'X_1)^{-1}X_1'(X_2\hat{\beta}_2).$$

Yani, "kısa" regresyondaki katsayı, "uzun" regresyondaki katsayı artı "dışlanmış" terim olan $X_2\hat{\beta}_2$'nin regresyona dahil olan $X_1$ bağlaştıranı üzerine yapılan regresyonundan elde edilen katsayıya eşittir. Buradaki ilişkinin kitle için de kolaylıkla kullanılabileceğini görebiliriz.

**Örnek:** Varsayalım $Y$ kazanç, $X_1$ eğitim ve $X_2$ gözlemlenemeyen yetenek olsun. "Uzun" katsayı $\hat{\beta}_1$ ile "kısa" katsayı $\tilde{\beta}_1$'i karşılaştıralım.

2.3. **İzdüşümler (Projections), $R^2$ ve Varyans (Değişirlik) Çözümlemesi (ANOVA).** Faydalı bir uygulama $R^2$'nin türetimidir; bu kriter, $Y$'deki değişimin ne kadarının $X$'deki değişim ile açıklandığını gösterir. Orijinden geçen regresyonda, aşağıdaki varyans çözümlemesi ayrıştırmasını gösterebiliriz (analysis of variance, ANOVA)

$$Y'Y = \widehat{Y}'\widehat{Y} + \widehat{e}'\widehat{e}.$$

Buradan $R^2 := \widehat{Y}'\widehat{Y}/Y'Y = 1 - \widehat{e}'\widehat{e}/Y'Y$ ve oluşum gereği $0 \leq R^2 \leq 1$.

Eğer regresyon kesme içeriyorsa, yukarıdaki değerlerin ortalamadan farkını almak mantıklıdır. Bu durumda formül şuna dönüşür

$$(Y - \bar{Y})'(Y - \bar{Y}) = (\hat{Y} - \bar{Y})'(\hat{Y} - \bar{Y}) + \widehat{e}'\widehat{e},$$

ve

$$R^2 := (\widehat{Y} - \bar{Y})'(\widehat{Y} - \bar{Y})/((Y - \bar{Y})'(Y - \bar{Y})) = 1 - \widehat{e}'\widehat{e}/(Y - \bar{Y})'(Y - \bar{Y}),$$

burada $0 \leq R^2 \leq 1$'dır ve kalıntılar tanım gereği, aşağıda gösterildiği gibi, sıfır ortalamasına sahiptir.

# 3. Sonlu Örneklemlerde Tahmin ve Basit Çıkarsama

**3.1. Gauss-Markov (GM) Modelinde Tahmin ve Çıkarsama.** GM modeli $(Y, X)$ verisi için ihtimal yasalarının $\{P_\theta, \theta \in \Theta\}$ bir toplamıdır ve şu özelliklere sahiptir:

| | | | |
|---|---|---|---|
| GM1 | $Y = X\beta + \varepsilon$, | $\beta \in \mathbb{R}^k$ | doğrusallık |
| GM2 | kerte$(X) = k$ | | özdeşleştirme (identification) |
| GM3 | $E_\theta[\varepsilon \mid X] = 0$ | $\forall\, \theta \in \Theta$ | dikeysellik, doğru belirginleştirme (specification), dışsallık (exogeneity) |
| GM4 | $E_\theta[\varepsilon\varepsilon' \mid X] = \sigma^2 I_{n\times n}$ | $\forall\, \theta \in \Theta$ | küresellik (sphericality) |

Burada, ihtimal modeli $P_\theta$'yı tanımlayan $\theta$ parametresi şunlardan oluşur

$$(\beta, \sigma^2, F_{\varepsilon|X}, F_X),$$

burada $\beta$ regresyon parametre vektörü, $\sigma^2$ bozukluk varyansı (variance of disturbances), $F_{\varepsilon|X}$, $\epsilon$ hatalarının $X$ verildiğinde koşullu dağılım fonksiyonu (conditional distribution function) ve $F_X$ ise $X$'in dağılım fonksiyonudur. Hem $F_{\epsilon|X}$ hem de $F_X$ belirginleştirilmemişlerdir (unspecified), yani parametrik değillerdir (non-parametric).

Bu model hakiki (actual) verinin birçok gerçekçi özelliğini yok sayar, ancak analiz için faydalı bir başlangıç noktasıdır. Ayrıca, bu bölümün ilerki kısımlarında hatalarının dağılımını normal dağılım takip edecek şekilde kısıtlayacağız.

**GM1 & GM3:** Bu varsayımlar birlikte değerlendirilmelidir, ancak GM3'ten aşağıda daha detaylı bahsedilecektir. Model, parametrelerin ve hata teriminin doğrusal bir fonksiyonu olarak yazılabilir, yani $y_i = \beta' x_i + \varepsilon_i$, burada GM3 $E[\epsilon_i|x_i] = 0$ olduğunu, ve, aşağıda da bahsedildiği gibi, aslında daha fazlasını da belirtir.

Burada ifade edilen, $Y$'nin koşullu ortalama fonksiyonunu parametrelerde doğrusal olan bir işlevsel biçim ile doğru şekilde belirginleştirdiğimizdir. Bu noktada, işlevsel biçimleri yaklaştırma teorisi ile nasıl oluşturduğumuzu hatırlayabiliriz. Bahsedilen noktada, $x_t$'yi bazı basit bağlaştıranlar olan $f(w_t)$'nin dönüşümleri (transformations) olarak oluşturmuştuk. Neticede, GM1 ve GM3 varsayımları şunu belirtecek şekilde yorumlanabilir: $E[y_t|w_t] = E[y_t|x_t] = x_t'\beta$, bu da bir tam (perfect) işlevsel biçimle ve yaklaştırma hatası nümerik olarak önemsemeye değmez olduğunda çalıştığımız bir varsayımdır. Bir çok ekonomik işlevsel biçim belirtilen varsayımlar ile gayet iyi eşleşmektedir.[6]

**GM2: Özdeşleştirme.**[7] Bu varsayım açıklayıcı değişkenlerin doğrusal olarak bağımsız olduklarını ifade etmektedir. Takip eden örnek bu koşulun arkasındaki düşüncenin altını çizmektedir. Varsayalım şu ücret denklemini tahmin etmek istiyorsunuz:

$$\log(\ddot{u}cret_i) = \beta_1 + \beta_2 e\breve{g}itim_i + \beta_3 s\ddot{u}re_i + \beta_4 tecr\ddot{u}be_i + \varepsilon_i$$

burada $e\breve{g}itim_i$ kaç yıl okul eğitimi alındığını, $s\ddot{u}re_i$ mevcut işte çalışılan süreyi ve $tecr\ddot{u}be_i$ işgücündeki tecrübeyi (yani şu anki iş dahil olmak üzere tüm işlerde çalışılan toplam süreyi) göstermektedir.

Ancak örneklemde hiç kimsenin işini değiştirmediğini düşünelim, yani tüm $i$'ler için $s\ddot{u}re_i = tecr\ddot{u}be_i$ olsun. Bu eşitliği regresyon denkleminde yerine koyduğumuzda şunu görürüz:

$$\log(\ddot{u}cret_i) = \beta_1 + \beta_2 e\breve{g}itim_i + (\beta_3 + \beta_4) tecr\ddot{u}be_i + \varepsilon_i.$$

Sonuçta $\beta_3 + \beta_4$ doğrusal kombinasyonunu tahmin edebilirken $\beta_3$ ve $\beta_4$'ü ayrı ayrı tahmin edemeyiz. Bu olay *kısmi özdeşleştirme* (*partial identification)* olarak bilinir. Ekonometride bu durumla düşünüldüğünün tersine sıkça karşılaşılır.

**GM3: Dikeysellik veya Katı Dışsallık (Strict Exogeneity).** Burada, bozukluk teriminin beklenen değeri açıklayıcı değişkenlere bağlı değildir:

$$E[\varepsilon|X] = 0.$$

Dikkat ederseniz, bu sadece $E[\varepsilon_i|x_i] = 0$ olduğunu degil, aynı zamanda tüm $j$'ler için $E[\varepsilon_i|x_j] = 0$ olduğunu ifade eder. Yani, $i$ gözlemi için bozukluk teriminin beklenen değeri sadece o gözlemin açıklayıcı değişkenlerine değil, aynı zamanda diğer tüm gözlemlerin açıklayıcı değişkenlerine de bağlı değildir. Son ifade zaman serileri için mantıksız bir koşul olabilse de, bu koşulu bir sonraki bölümde kaldıracağız.

---

[6] Örneğin, Cobb-Douglas üretim fonksiyonu gibi doğrusal olmayan bir model $y_i = AK_i^{\alpha} L_i^{1-\alpha} e^{\varepsilon_i}$ logaritmalar alınarak kolaylıkla doğrusal bir modele dönüştürülebilir:

$$\ln y_i = \ln A + \alpha \ln K_i + (1-\alpha)\ln L_i + \varepsilon_i$$

Bu aynı zamanda daha önce türettiğimiz çokterimli yaklaştırmalara güzel bir bağlantı oluşturur. Aslında, yukarıdaki denkleme $(\ln L)^2$ ve $(\ln K)^2$ gibi ek terimler katıldığında bir translog işlevsel biçimi elde edilir ve bu da ikinci dereceden bir çokterimli yaklaştırmasıdır. Açıktır ki, yaklaştırma teorisi ve ekonomik modelleme arasındaki bağlantıları keşfetmek için ilginç bir fırsat mevcuttur.

[7] Bu örnek ve aşağıdaki örneklerden bazıları Raymond Guiterras tarafından hazırlanmıştır.

Ek olarak, daha önce de vurgulandığı üzere, bu varsayım koşullu ortalama fonksiyonunda tam yaklaştırma (perfect approximation) olduğunu varsaydığı için de mantıksız olabilir.

Bu varsayımın neden dikkatle değerlendirilmesi gerektiğine yönelik başka bir sebep de vardır. Ekonometrik analizin esas amaçlarından birinin nedensel ya da yapısal etkileri ortaya çıkartmak olduğunu hatırlayınız. Eğer regresyonun nedensel bir yorumlaması olacaksa, doğru nedensel denklemin bozuklukları

$$y_i = x_i'\gamma + u_i$$

$E[u_i|x_i] = 0$ gibi dikeysellik kısıtlamalarını sağlamak zorundadır. Bu gerçekleşirse, sonrasında nedensel etki fonksiyonu $x_i'\gamma$ regresyon denklemi $x_i'\beta$ ile çakışacaktır.

Aşağıdaki standart örnek bu fikri netleştirmemizde yardımcı olacaktır. Gelir ve eğitim seviyesi ilişkisi ve gerçek model hakkındaki düşüncemiz şöyledir:

$$y_i = \gamma_1 x_i + u_i, \quad u_i = \gamma_2 A_i + \varepsilon_i$$

burada $x_i$ eğitim seviyesi iken, $u_i$ ise yetenek etkisi olan $\gamma_2 A_i$'den ve $x_i$, $A_i$ ve $\varepsilon$'dan bağımsız olan ikinci bir kalıntı $\varepsilon_i$'den oluşmaktadır. Varsayalım hem eğitimin hem de yeteneğin ortalamadan farkı alınsın, yani, $x_i$ ortalama eğitimden sapmayı ölçsün ve $A_i$ ortalama yetenekten sapma olsun. Genel olarak, eğitim ile yetenek alakalıdır, yani

$$E[u_i|x_i] = \gamma_2 E[A_i|x_i] \neq 0.$$

Bu nedenle, dikeysellik bozulmaktadır ve $\gamma_1$'i $y_i$'nin $x_i$ üzerine regresyonu ile tahmin etmek mümkün değildir. Ancak dikkat ederseniz, eğer yetenek olan $A_i$'yi gözlemleyebilsek ve $y_i$'nin eğitim $x_i$ ve yetenek $A_i$ üzerine uzun regresyonunu yapabilseydik, bu taktirde regresyon katsayıları olan $\beta_1$ ve $\beta_2$ vasıtasıyla nedensel fonksiyonun katsayıları olan $\gamma_1$ ve $\gamma_2$ hesaplanabilecekti.

**GM4**: **Küresellik.** Bu varsayım iki önemli gerekliliği içerir. Bunlardan birincisi *eşit yayılım'dır (homoscedasticity)*: $E\left[\varepsilon_i^2|X\right] = \sigma^2, \quad \forall\, i$. Bunun manası, her bir bozukluk teriminin koşullu varyansının (conditional variance) tüm gözlemler için aynı olmasıdır. Bu genellikle gerçekçilikten oldukça uzak bir varsayımdır.

İkinci koşul ise *kendiyle ilgileşim olmaması'dır (nonautocorrelation)*: $E\left[\varepsilon_i\varepsilon_j|X\right] = 0 \quad \forall\, i \neq j$. Bunun manası da herhangi iki farklı gözleme ait bozuklukların ilgileşimsiz (uncorrelated) olmasıdır. Zaman serisi verisinde bozukluklar çoğu kez kendiyle ilgileşimlidir.

Ek olarak, bu varsayım dolaylı olarak birçok ikili tepki modelini (binary response models) (ve diğer kesikli tepki (discrete response) modellerini) ortadan kaldırır. Örneğin, varsayalım $y_i \in \{0,1\}$ olsun, bu durumda

$$y_i = E[y_i|x_i] + \varepsilon_i = Pr[y_i = 1|x_i] + \varepsilon_i,$$

burada $\varepsilon_i$'nin varyansı $P[y_i = 1|x_i](1 - P[y_i = 1|x_i])$'dir ve bu da $x_i$'ye bağımlıdır; bunun istisnası ilgimizi çekmeyen $P[y_i = 1|x_i]$ durumudur, bu da $x_i$'ye bağımlı değildir.

3.2. **Gauss-Markov Modelinde OLS'nin Özellikleri.** İlgimizi çeken, $\beta$'nın çeşitli işlevselleridir, örneğin,

- $\beta_j$, $\beta$'nın ilgilenebileceğimiz bir $j$-inci bileşeni,
- $(x_1 - x_0)'\beta$, bağlaştıran değerlerindeki bir değişiklikten kaynaklanan koşullu ortalamanın bir kısmi farkı (partial difference),
- $\frac{\partial x(w)}{\partial w_k}'\beta$, koşullu ortalamanın temel (elementary) bağlaştıran $w_k$ ile bir kısmi türevi.

Bu işlevseller şu biçim ile oluşur

$$c'\beta \quad \text{geçerli olduğu} \quad c \in \mathbb{R}^K \text{ için.}$$

Bu durumda, $\hat{\beta}$'nın etkinliğini (efficiency) bu tür işlevselleri mümkün olabildiği kadar kesin (precise) tahmin etme yeteneği olarak tanımlamak mantıklıdır.

Belirtilen varsayımlar ile, şu durumlar elde edilir

$$E_\theta[\hat{\beta}|X] = E_\theta[(X'X)^{-1}X'(X\beta + \varepsilon) \mid X] = I\beta + 0 = \beta \quad \forall\theta \in \Theta.$$

Bu özellik ortalama-yansızlık (mean-unbiasedness) olarak adlandırılır. Bunun manası da, özellikle, doğrusal işlevsellerin tahminlerinin de yansız olduğudur $E_\theta[c'\hat{\beta}|X] = c'\beta$.

Şimdi, regresyon katsayısı $\beta$ için OLS'nin etkinliğini diğer tahminciler ile karşılaştıralım. Rekabet edecek aday tahminci $\tilde{\beta}$ doğrusal ve yansız olsun. Doğrusallığın manası $\tilde{\beta} = a + AY$ olmasıdır, burada $a$ ve $A$, $X$'in ölçülebilir fonksiyonudur (measurable function) ve tüm $\Theta$ içindeki $\theta$ için $E_\theta[\tilde{\beta}|X] = \beta$'dır. Dikkat ederseniz, yansızlık koşulu tüm $\beta \in \mathbb{R}^k$ için $a + AX\beta = \beta$ olmasını gerektirir. Yani,

$$AX = I \text{ ve } a = 0\text{'dır.}$$

**Teorem 3.1.** *Gauss-Markov Teoremi. GM modelinde, $X$ üzerine koşullu olarak, $\beta$'nın en küçük varyans doğrusal yansız tahmincisi (EVDYT, minimum variance linear unbiased estimator, MVLUE) $\hat{\beta}$'dır; bunun manası da herhangi bir diğer doğrusal yansız tahminci $\tilde{\beta}$'nın şu ilişkiyi sağlayacağıdır:*

$$Var_\theta[c'\tilde{\beta} \mid x] \geq Var_\theta[c'\hat{\beta} \mid X], \quad \forall c \in \mathbb{R}^K, \quad \forall \theta \in \Theta.$$

Yukarıdaki özellik şuna denktir: $c'\text{Var}_\theta[\tilde{\beta} \mid X]c - c'\text{Var}_\theta[\hat{\beta} \mid X]c \geq 0 \quad \forall c \in \mathbb{R}^K, \quad \forall \theta \in \Theta$, bu da şunu söylemekle aynıdır

$$\text{Var}_\theta[\tilde{\beta} \mid X] - \text{Var}_\theta[\hat{\beta} \mid X] \text{ , } \forall \theta \in \Theta \text{ için } \textit{pozitif tanımlıdır}.$$

**Örnek.** Varsayalım $y_t$ kazançları, $x_t$ alınan eğitim süresini temsil etsin. Eğitim süresindeki değişikliğin ortalama etkisi şudur $E[y_t \mid x_t = x^*] - E[y_t \mid x_t = x] = (x^* - x)'\beta$. GM Teoremi kullanılarak, $(x^* - x)'\beta$ için $(x^* - x)'\hat{\beta}$ EVDYT'dir.

**Örnek.** OLS'in bir rakibi WLS'dir (weighted least squares, ağırlıklı en küçük kareler) ve ağırlıkları $W = \text{diag}(w(x_1), ..., w(x_n))$'dir. WLS şunu çözer: $\min_\beta E_n[(y_t - x_t'\beta)^2 w(x_t)]$, veya, denk olarak $\min_\beta (Y - X\beta)'W(Y - X\beta)$. Çözüm $\hat{\beta}_{WLS} = (X'WX)^{-1}X'WY$'dır ve $\hat{\beta}_{WLS}$ doğrusal ve yansızdır (bunu gösteriniz). GM1-GM4 varsayımları altında WLS, OLS ile çakışmadığı müddetçe, OLS'ten daha az etkindir.

3.3. **GM Teorem'inin İspatı.** Notasyonda, $\theta$ ile endekslemeden vazgeçelim. Yansızlığın sağlaması yukarıda yapılmıştı. Herhangi bir başka yansız tahminciyi ele alalım: $\tilde{\beta} = AY$. Yansızlık özelliği ile, $AX = I$'dır. Şu hususları gözlemleyebiliriz: $\text{var}[c'\tilde{\beta} \mid X] = c'AA'c\sigma^2$ ve $\text{var}[c'\hat{\beta} \mid X] = c'(X'X)^{-1}c \cdot \sigma^2$. Aşağıdaki eşitliği göstermemiz yeterlidir

$$\text{var}[c'\tilde{\beta} \mid X] - \text{var}[c'\hat{\beta} \mid X] = \text{var}[c'\tilde{\beta} - c'\hat{\beta} \mid X],$$

çünkü sağ taraf negatif değildir. Şunu yazalım

$$\begin{aligned}
\text{var}[c'\tilde{\beta} - c'\hat{\beta} \mid X] &= \text{var}[\underbrace{c'(A - (X'X)^{-1}X')}_{M}(X\beta + \epsilon) \mid X] \\
&= \text{var}[M\varepsilon \mid X] = E[M\varepsilon\varepsilon'M' \mid X] = MM'\sigma^2 \quad \text{(A4'ten dolayı)} \\
&= c'[AA' - (X'X)^{-1}]c \cdot \sigma^2 \quad (\Longleftarrow AX = I) \\
&= c'AA'c\sigma^2 - c'(X'X)^{-1}c \cdot \sigma^2 \square
\end{aligned}$$

**Açıklama:** Bu ispat, portföy teorisi ve Hausman-türü testler gibi birçok yerde kullanılan genel bir prensibi göstermektedir. Dikkat ederseniz, farkın varyansı, kovaryansı (eşdeğişirlik, covariance) gerektirmeyen çok basit bir yapıya sahiptir:

$$\mathrm{var}[c'\tilde{\beta} - c'\hat{\beta} \mid X] = \mathrm{var}[c'\tilde{\beta} \mid X] - \mathrm{var}[c'\hat{\beta} \mid X]$$

Bunun sebebi şudur

$$\mathrm{cov}[c'\tilde{\beta}, c'\hat{\beta} \mid X] = \mathrm{var}[c'\hat{\beta} \mid X].$$

Bunun manası da etkinsiz tahmin (inefficient estimate) $c'\tilde{\beta}$ şuna eşittir: etkin bir tahmin olan $c'\hat{\beta}$ artı etkin tahmin ile ilgileşimsiz (uncorrelated) ek bir tahmin rastsalı (noise).

3.4. **OLS'in Rakipleri ve Alternatifleri. Bölüm I.** Şu örnekleri değerlendirelim.

**Örnek** [*Uzman (expert) Tahminlerine karşı OLS*] Bir uygulama olarak, varsayalım $\beta = (\beta_1, ...\beta_k)'$ olsun; burada $\beta_1$ bir ürünün talep elastikiyetini ölçer. Bahsettiğimiz tanımlar ve GM teoremi çerçevesinde, iki tahminciyi analiz edelim ve karşılaştıralım: sabit tahmin $\beta_1^* = 1$ bir endüstriyel organizasyon uzmanı tarafından sağlanmış olsun ve $\hat{\beta}_1$ sıradan enküçük kareler tahmini olarak elde edilsin.

- Hangi durumda bunlardan birini diğerine tercih edersiniz?
- GM teoremi bu karar ile alakalı mıdır?

Uzman tahmini, OLS'ten ortalama kare hata açısından daha iyi olabilir:

$$E_\theta[(\beta^* - \beta)^2] < E_\theta[(\hat{\beta} - \beta)^2]$$

bu bazı veya birçok $\theta \in \Theta$ değerleri için geçerlidir, bu da doğrusal işlevsellerin daha küçük tahmin hatası yaptığı manasına gelir. Kararın kritik yönü $\Theta$'nın değeri hakkında ne düşündüğümüzdür. Daha detaylı bilgi için Bölüm 3.9'a da bakınız.

**Örnek** [*Fire (shrinkage) Tahminlerine karşı OLS*] Fire tahmincileri öğrenme teorisinde yeniden gündeme gelmektedir, bu da regresyon çözümlemesi demenin modern bir yoludur.

Fire tahmincilerinin bir örneği, şu örneği çözendir:

$$\min_b [(Y - Xb)'(Y - Xb)/2 + \lambda(b - \beta^*)X'X(b - \beta^*)]$$

İlk terime sadakat/doğruluk (fidelity) denir, çünkü uyum iyiliğini (goodness of fit) sözkonusu veri için ödüllendirir. İkinci terime ise fire denir, çünkü $Xb$'nin *önsel* (a priori) olarak (teori ya da başka verisetleri ile yapılan tahmin sonuçları vs değerlendirildiğinde) makul olduğunu düşündüğümüz $X\beta^*$ değerlerinden sapmalarını cezalandırır. Yukarıdaki tahminci için normal denklemler şu şekilde ifade edilir:

$$X'(Y - X\tilde{\beta}) + \lambda X'X(\tilde{\beta} - \beta^*) = 0,$$

ve $\tilde{\beta}$ için çözüldüğünde şu elde edilir

$$\tilde{\beta} = (X'X(1+\lambda))^{-1}(X'Y + \lambda X'X\beta^*) = \frac{\hat{\beta}}{1+\lambda} + \frac{\lambda}{1+\lambda}\beta^*$$

Dikkat ederseniz, $\lambda = 0$'a eşitlendiğinde OLS'in $\hat{\beta}$'sı elde edilir ve $\lambda \approx \infty$ olduğunda uzman tahmini $\beta^*$ elde edilir.

$\lambda$'nın seçimi sıkça uygulamacılara (practitioners) bırakılır. Tahmin amacına yönelik olarak, $\lambda$ ortalama kare hatayı minimize edecek şekilde seçilebilir. Bu da *çapraz geçerleme* (cross-validation) denen bir yolla elde edilebilir.

3.5. **GM'da Normallik Varsayımı Altında Sonlu Örneklem Çıkarsaması.** Çıkarsama amacıyla, OLS'nin varyansını tahmin etmemiz gerekir. $\sigma^2(X'X)^{-1}$ için yansız bir tahmini $s^2(X'X)^{-1}$ olarak oluşturabiliriz; burada

$$s^2 = \widehat{e}'\widehat{e}/(n-K)\text{'dır.}$$

$E_\theta[s^2|X] = \sigma^2$'in yansızlığı şuradan gelir

$$\begin{aligned} & E_\theta[\widehat{e}'\widehat{e}|X] = E_\theta[\varepsilon' M_X \varepsilon|X] = E_\theta[tr(M_X\varepsilon\varepsilon')|X] \\ = \;& tr(M_X E[\varepsilon\varepsilon'|X]) = tr(M_X\sigma^2 I) \\ = \;& \sigma^2 tr(I-P_X) = \sigma^2(tr(I_n) - tr(P_X)) = \sigma^2(tr(I_n) - tr(I_K)) = \sigma^2(n-K), \end{aligned}$$

burada şunları kullandık: $tr(AB) = tr(BA)$, iz'in (trace) ve beklenti işlemcisinin doğrusallığı ve $tr(P_X) = tr((X'X)^{-1}X'X) = tr(I_K)$ olduğu.

Aynı zamanda önemli bir ek varsayımı da eklememiz gerekti:

$$\text{GM5.} \quad \varepsilon|X \sim N(0,\sigma^2 I) \qquad \text{Normallik}$$

Bu şekilde, $Y$'nin $X$ verildiğindeki koşullu dağılım modelimiz parametrik hale geldi ve koşullu dağılımın parametre vektörünü $\theta = (\beta,\sigma^2)$'a eksiltti (reduced).

Normallik varsayımı nasıl haklı çıkartılabilir? Bazı deneme yanılma yöntemleri tartışılabilir.

**Teorem 3.2.** *GM.1-GM.5 varsayımları altında şunlar doğrudur:*

1. $\hat\beta|X \sim N(\beta,\sigma^2(X'X)^{-1})$, $Z_j := (\hat\beta_j - \beta_j)/\sqrt{\sigma^2(X'X)^{-1}_{jj}} \sim N(0,1)$.
2. $(n-K)s^2/\sigma^2 \sim \chi^2(n-K)$.
3. $s^2$ *ve* $\hat\beta$ *bağımsızdırlar.*
4. $t_j := (\hat\beta_j - \beta_j)/se(\hat\beta_j) \sim t(n-K) \approx N(0,1)$, *şunun için* $se(\hat\beta_j) = \sqrt{s^2(X'X)^{-1}_{jj}}$; *yaklaşıklık* $\approx$ *ise* $(n-K) \geq 30$ *olduğunda doğrudur (accurate).*

3.6. **3.2 Teoreminin ispatı.** Şu olguları kullanacağız:

(a) bir normal dağılımın doğrusal fonksiyonu bir normaldir, yani eğer $Z \sim N(0,\Omega)$ ise, o zaman $AZ \sim N(0, A\Omega A')$,
(b) eğer iki normal vektör ilgileşimsiz ise, o zaman ikisi bağımsızdırlar,
(c) eğer $Z \sim N(0,I)$ ve $Q$ simetrik ve denkgüçlü (idempotent) ise, o zaman $Z'QZ \sim \chi^2$kerte$(Q)$)'dır,

(d) eğer bir standart normal değişken $N(0,1)$ ve ki-kare (chi-square) değişken $\chi^2(J)$ bağımsız iseler, o zaman $t(J) = N(0,1)/\sqrt{\chi^2(J)/J}$ bir $J$ serbestlik dereceli (degrees of freedom) Student t değişkeni (Student's t) olarak adlandırılır.

Özelllik (a)-(c)'nin ispatını ödev olarak gösteriniz.
Şimdi şu iddiaların herbirini ispatlayalım:

(1) $\hat{e} = M_X\varepsilon$ ve $\hat{\beta}-\beta = (X'X)^{-1}X'\varepsilon$ sıfır ortalama ile birleşik olarak normaldirler, çünkü bir normal vektörün doğrusal fonksiyonu normaldir.
(2) $\hat{e}$ ve $\hat{\beta}$ ilgileşimsizdirler, çünkü kovaryansları $\sigma^2(X'X)^{-1}X'M_X = 0$'a eşittir ve bu sebeple birleşik normallikten dolayı bağımsızdırlar. $s^2$ ise $\hat{e}$'in bir fonksiyonudur, yani $\hat{\beta}$'ten bağımsızdır.
(3)

$$(n-K)s^2/\sigma^2 = (\varepsilon/\sigma)'M_X(\varepsilon/\sigma) \sim \chi^2(\text{kerte}(M_X)) = \chi^2(n-K).$$

(4) Özellikler 1-3 ile şunları gösterebiliriz

$$t_j = Z_j\Big/\sqrt{(s^2/\sigma^2)} \sim N(0,1)\Big/\sqrt{\chi^2(n-K)/(n-K)} \sim t(n-K).$$

□

Özellik 4 önsav sınaması (hypothesis testing) yapmamızı ve güven aralıkları (confidence intervals) kurmamızı sağlar. Şu olaya (event) sahibiz

$$t_j \in [t_{\alpha/2}, t_{1-\alpha/2}] \text{ 'nin ihtimali } 1-\alpha\text{'dır,}$$

burada $t_\alpha$, bir $t(n-K)$ değişkeninin $\alpha$-bölütünü ifade eder. Bu nedenle, $\beta_j$'yi $1-\alpha$ ihtimali ile kapsayan bir güven bölgesi (confidence region) şu şekilde ifade edilir

$$I_{1-\alpha} = [\hat{\beta}_j \pm t_{1-\alpha/2}se(\hat{\beta}_j)].$$

Bu da olay $\beta_j \in I_{1-\alpha}$'nın olay $t_j \in [t_{\alpha/2}, t_{1-\alpha/2}]$'a eşdeğer olmasından ortaya çıkar.

Ayrıca, şunu sınamak için

$$H_o : \beta_j = 0 \text{ karşı } H_a : \beta_j > 0$$

$t_j = (\hat{\beta}_j - 0)/se(\hat{\beta}_j) \geq t_{1-\alpha}$'in bir kritik değer olarak sağlandığını kontrol ediyoruz. Geleneksel uygulama, kritik değer $t_{1-\alpha}$'ın öyle seçilmesidir ki, sıfır önsavını (null hypothesis) doğru iken hatalı şekilde reddetmenin ihtimali çok küçük bir değer olan $\alpha$'ya eşit olsun; burada $\alpha$'nın kanonik (canonical) seçimi 0.01, 0.05 veya 0.1'e eşittir. Sıfır önsavı altında, $t_j$ bir $t(n-K)$ dağılımı takip eder, $t_{1-\alpha}$ değerini de standart tablolardan bulabiliriz.

Şunu test etmek için

$$H_o : \beta_j = 0 \text{ karşı } H_a : \beta_j \neq 0$$

$|t_j| \geq t_{1-\alpha/2}$ olup olmadığına bakıyoruz. Kritik değer $t_{1-\alpha/2}$ ise hatalı ret ihtimali $\alpha$ olacak şekilde seçilir.

Sadece "reddet" veya "reddetme" sonuçlarını bildirmek yerine p-değerini (p-value) – sıfır önsavı altında $t_j$'ye eşit ya da bundan büyük bir istatistiği görme ihtimalini, bildirmek de yaygın bir uygulamadır:

$$P_j = 1 - Pr[t(n-K) \leq t]\Big|_{t=t_j}$$

tek yanlı almaşıklar (one sided alternatives) için, ve $|t_j|$

$$P_j = 1 - Pr[-t \leq t(n-K) \leq t]\Big|_{t=|t_j|}$$

ise iki yanlı almaşıklar içindir. $Pr[t(n-K) \leq t]$ ihtimali, $t(n-K)$'nin dağılım fonksiyonudur, bu da Student dağılımı ile sıralanır (tabulated).

P-değerleri, önsav sınaması için t-istatistiklerini kullanmaya eşdeğer şekilde kullanılabilir. Gerçekten de, bir önsavı eğer $P_j \leq \alpha$ ise reddedebiliriz.

**Örnek. Temin'in Roma Buğday Fiyatları.** Temin, bir mesafe indirim modeli (distance discount model) tahmin etmiştir:

$$fiyat_i = \beta_1 + \beta_2 \cdot mesafe_i + \epsilon_i, \quad i = 1, ..., 6$$

burada $fiyat_i$ Roma eyaletlerinde buğdayın fiyatı ve $mesafe_i$ ise $i$ eyaletinin Roma'ya olan uzaklığıdır. Tahmin edilen model şudur

$$fiyat_i = \underset{(0.49)}{-1.09} - \underset{(0.0003)}{0.0012} \cdot mesafe_i + \hat{\epsilon}_i, \quad R^2 = 0.79,$$

standart hatalar parantez içinde gösterilmiştir. Sınama $\beta_2 = 0$'a karşı $\beta_2 < 0$ için kullanılan t istatistiği $t_2 = -3.9$'dur. Tek yanlı sınama için P-değeri $P_2 = P[t(4) < -3.9] = 0.008$'dir. $\beta_2$ için bir %90 güven bölgesi $[-0.0018, -0.0005]$'tir; bu da $[\hat{\beta}_2 \pm t_{0.95}(4) \cdot se(\hat{\beta}_2)] = [0.0012 \pm 2.13 \cdot 0.0003]$ olarak hesaplanır.

**Teorem 3.3.** *GM1-GM5 varsayımları altında, $\hat{\beta}$ en çok olabilirlik tahmincisi (maximum likelihood estimator) ve aynı zamanda $\beta$'nın en küçük varyans yansız tahmincisidir.*

**İspat.** Bu ispat, $\hat{\beta}$'in varyansının yansız tahminciler için Cramer-Rao alt sınırını (Cramer-Rao lower bound) sağladığını doğrulayarak gösterilir. Bu durumda, $y_i$'nin $y_i = y$'deki $x_i$ üzerine koşullu yoğunluğu şu şekildedir

$$f(y|x_i, \beta, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-\frac{1}{2\sigma^2}(y - x_i'\beta)^2\}.$$

Böylece, olabilirlik fonksiyonu şudur

$$L(b, \sigma^2) = \prod_{i=1}^{n} f(y_i|x_i, b, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\{-\frac{1}{2\sigma^2}(Y - Xb)'(Y - Xb)\}.$$

Burada OLS $\hat{\beta}$'nın $\beta$ üzerinde olabilirlik fonksiyonunu ençokladığı kolayca görülebilir (bunu kontrol ediniz).

Burada, $\theta = (\beta, \sigma^2)$'ın yansız tahmincilerinin varyansının (koşullu) Cramer-Rao alt sınırı şuna eşittir (sağlamasını yapınız)

$$\left[-E\left[\frac{\partial^2 \ln L}{\partial\theta\partial\theta'}\Big|X\right]\right]^{-1} = \begin{bmatrix} \sigma^2(X'X)^{-1} & 0 \\ 0 & 2\sigma^4/n \end{bmatrix}.$$

Buradan da enküçük karelerin alt sınırı sağladığı çıkartılır.

3.7. **OLS'in Rakipleri ve Alternatifleri. Bölüm II.** Hatalar normal dağılmadığı taktirde, OLS'in diğer tahmincilere göre performansı ciddi ölçüde kötüleşebilir.

Örnek. Koenker and Bassett (1978) çalışması.

3.8. **Dışlanmış Buluşsal Yöntemler (Omitted Heuristics): $\varepsilon$'un normalliği nereden gelmektedir?** Poincare: "Herkes *Gauss*'un hata yasasına (law of error) inanmaktadır; deneyciler bunun bir matematik teoremi olduğunu düşündükleri için, matematikçiler ise deneysel bir olgu olduğunu düşündükleri için."

Gauss (1809) enküçük karelerin ençok olabilirlik tahmini olduğu bir hata dağılımı oluşturmak için geriye dönük çalışmıştır. Bu nedenle, normal dağılıma bazen Gausçu da (Gaussian) denmektedir.

Merkezi Limit Teoremi (MLT, Central Limit Theorem, CLT)"ispatı":

Ekonometride, Haavelmo, 1944'teki *Econometrica*'da yayınlanan "Probability Approach to Econometrics (Ekonometriye Olasılık Yaklaşımı)" makalesinde, bu ispatın seçkin bir destekçisi olmuştur. MLT ispatında, her bir hata $\varepsilon_i$'nin büyük sayıdaki küçük ve bağımsız basit (elementary) hatalar olan $v_{ij}$'nin toplamı olduğu düşünülür, bu sebeple de bu hatalar merkezi limit teoremi ışığında yaklaşık olarak Gausçu'dur.

Eğer basit hatalar olan $v_{ij}, j = 1, 2, ...$ sıfır ortalamalı, özdeş ve bağımsız dağılımlı (i.i.d., identically and independently distributed) ise ve $E[v_{ij}^2] < \infty$ ise, o zaman *büyük* $N$ için

$$\varepsilon_i = \sqrt{N}\left[\frac{\sum_{j=1}^{N} v_{ij}}{N}\right] \approx_d N(0, Ev_{ij}^2),$$

bu MLT'den elde edilmektedir.

Ancak, eğer basit hatalar $v_{ij}$ iid ve simetrik iseler ve $E[v_{ij}^2] = \infty$ ise, o zaman *büyük* $N$ *için* ($v_{ij}$'nin kuyruk (tail) hareketlerine getirilecek ek teknik kısıtlar ile)

$$\varepsilon_i = N^{1-\frac{1}{\alpha}}\left[\frac{\sum_{j=1}^{N} v_{ij}}{N}\right] \approx_d \textit{Kararlı (Stable)}\text{'dır}$$

burada $\alpha$ en büyük sonlu beklemdir (finite moment): $\alpha = \sup\{p : E|v_{ij}|^p < \infty\}$. Bu da Khinchine ve Levy tarafından ispat edilen MLT'den gelmektedir. Kararlı dağılımlar aynı zamanda toplam-kararlı (sum-stable) ve *Pareto-Levy* dağılımları olarak da adlandırılırlar.

Simetrik kararlı dağılımların yoğunlukları yaklaşık olarak güç fonksiyonları (power functions) gibi hareket eden kalın kuyruklara (thick tails) sahiptirler; kuyruklarda $\alpha < 2$ ile $x \mapsto \text{sabit} \cdot |x|^{-\alpha}$'dır.

Başka bir ilginç ek gözlem: Eğer $\alpha > 1$ ise, örneklem ortalaması olan $\sum_{j=1}^{N} v_{ij}/N$ bir yakınsayan istatistiktir (converging statistic), eğer $\alpha < 1$ ise örneklem ortalaması olan $\sum_{j=1}^{N} v_{ij}/N$ bir ıraksayan istatistiktir (diverging statistic), bunun da farklılaşma ve farklılaşmama (diversification and non-diversification) için ilginç uygulamaları vardır. (Son husus için bkz. R. Ibragimov'un çalışmaları).

Referanslar: Embrechts et al. *Modelling Extremal Events (Uç Olayları Modelleme)*

3.9. **Genel Doğrusal Kısıtlamalar ile Sınama ve Tahmin Etme.** Şimdi de şu şekilde verilen bir doğrusal eşitlik kısıtlamasını sınamayı düşünelim

$$H_0 : R\beta = r, \quad \text{kerte}R = p.$$

burada $R$ bir $p \times K$ matrisi ve $r$ ise $p$-vektörüdür. Varsayım olarak $R$'nin tam satır kertesinin olması (full row rank) basitçe artık kısıtlamaların (redundant restrictions) olmadığını ifade eder – yani diğer kısıtlamaların doğrusal kombinasyonu olarak yazılabilecek kısıtlamalar mevcut değildir. Almaşık ise şudur

$$H_0 : R\beta \neq r.$$

Bu düzenleme birçok önsavı sınamamıza imkan sağlar. Örneğin,

$R = [0, 1, 0, .., 0]$ $\;r = 0$ $\;$ şu kısıtlamayı oluştur $\beta_2 = 0$

$R = [1, 1, 0, ..., 0]$ $\;r = 1$ $\;$ şu kısıtlamayı oluştur $\beta_1 + \beta_2 = 1$

$R = [\mathbf{0}_{p\times(K-p)} I_{p\times p}]$ $\;r = (0, 0, ...)'$ $\;$ şu kısıtlamayı oluştur $\beta_{K-p+1} = 0, ..., \beta_K = 0.$

$H_0$'ı sınamak için, Wald istatistiğinin bir kritik değeri geçip geçmediğine bakıyoruz:

$$W := (R\hat{\beta} - r)'[\widehat{Var}(R\hat{\beta})]^{-1}(R\hat{\beta} - r) > c_\alpha,$$

burada kritik değer olan $c_\alpha$ sıfır önsavı doğru iken hatalı ret ihtimali $\alpha$'ya eşit olacak şekilde seçilmektedir. GM1-5 ile, şunu kabul edebiliriz

$$\widehat{Var}(R\hat{\beta}) = s^2 R(X'X)^{-1}R'. \tag{3}$$

Şunu çıkartabiliriz $W_0 = (R\hat{\beta} - r)'[\sigma^2 R(X'X)^{-1}R']^{-1}(R\hat{\beta} - r) = N(0, I_p)^2 \sim \chi^2(p)$ ve bu da $s^2/\sigma^2 \sim \chi^2(n - K)/(n - K)$'ı sağlayan $s^2$'den bağımsızdır. Bu sebeple, şunu çıkartabiliriz

$$W/p = (W_0/p)/(s^2/\sigma^2) \sim \frac{\chi^2(p)/p}{\chi(n - K)/(n - K)} \sim F(p, n - K),$$

burada $c_\alpha$, $F(p, n - K)$ çarpı $p$'nin $\alpha$-bölütü olarak alınabilir. Buradaki $W/p$ istatistiğine F istatistiği (F-statistic) denir.

Hipotezi test etmenin diğer bir yolu mesafe fonksiyonu (distance function, DF) testidir (yarı-olabilirlik oranı sınaması, quasi likelihood ratio test) ve bu da ölçüt (criterion) fonksiyonunun kısıtsız (unrestricted) tahmini ve kısıtlı (restricted) tahmininden elde edilen değerlerinin farkına bağlıdır:

$$Q_n(\hat{\beta}_R) - Q_n(\hat{\beta}),$$

bizim durumumuz için şunlar geçerlidir $Q_n(b) = (Y - X'b)'(Y - X'b)/n$, $\hat{\beta} = \arg\min_{b\in\mathbb{R}^K} Q_n(b)$ ve

$$\hat{\beta}_R = \arg \min_{b\in\mathbb{R}^K : Rb=r} Q_n(b)$$

Aşağıdaki eşdeğerlik, (3)'te belirtilen kurgu için de geçerli olmaktadır:

$$DF = n[Q_n(\hat{\beta}_R) - Q_n(\hat{\beta})]/s^2 = W,$$

yani, mesafe fonksiyonu testinin kullanılması Wald testi kullanılmasına denktir. Bu denklik daha genel olarak GM modeli dışında geçerli değildir.

Başka bir önemli test prensibi de LM test prensibidir; bu da Lagrange Çarpanının (Multiplier) yukarıda belirtilen kısıtlamalı optimizasyon problemi değeri üzerinden hesaplanır. Problem için Lagrange değeri şudur

$$\mathcal{L} = nQ_n(b)/2 + \lambda'(Rb - r).$$

Optimumu tanımlayan koşullar şunlardır

$$X'(Y - Xb) + R'\lambda = 0, Rb - r = 0$$

Bu denklemler çözüldüğünde şunu elde ederiz

$$\hat{\beta}_R = (X'X)^{-1}(X'Y + R'\hat{\lambda}) = \hat{\beta} + (X'X)^{-1}R'\hat{\lambda}$$

$b = \hat{\beta}_R$ değerini $Rb - r = 0$ kısıtına koyduğumuzda, $R(\hat{\beta} + (X'X)^{-1}R'\hat{\lambda}) - r = 0$ değeri elde edilir, ya da

$$\hat{\lambda} = -[R(X'X)^{-1}R']^{-1}(R\hat{\beta} - r).$$

İktisatta çarpan, bir kısıtın gölge fiyatı (shadow price) olarak adlandırılır. Sınama problemimizde, eğer kısıtın fiyatı çok yüksek ise, hipotezi reddederiz. Test istatistiği şu formu alır:

$$LM = \hat{\lambda}'[\widehat{Var}(\hat{\lambda})]^{-1}\hat{\lambda},$$

Bizim durumumuzda, $\widehat{Var}(R\hat{\beta}) = s^2R(X'X)^{-1}R'$ için, şu değeri kullanabiliriz

$$\widehat{V}ar(\hat{\lambda}) = [R(X'X)^{-1}R']^{-1}\widehat{Var}(R\hat{\beta})[R'(X'X)^{-1}R]^{-1}$$

Bu kurgu bizim özel örneğimiz için şu eşdeğerliği de vermektedir

$$LM = W.$$

Bu eşitliğin doğrusal olmayan tahminciler için geçerli olmasının gerekmediğine dikkat ediniz (aslında genel olarak bu istatistikler için yanaşık (asymptotic) eşdeğerlik geçerlidir).

Yukarıda kısıtlamalı tahmini de türetmiştik:

$$\hat{\beta}_R = \hat{\beta} - (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}(R\hat{\beta} - r).$$

Sıfır önsavı doğru olduğunda, şunu gösterebiliriz

$$E[\hat{\beta}_R|X] = \beta$$

ve

$$Var[\hat{\beta}_R|X] = FVar(\hat{\beta}|X)F', \quad F = (I - (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}R).$$

Matris anlamında, şunu göstermek zor değildir

$$Var[\hat{\beta}_R|X] \leq Var[\hat{\beta}|X]$$

ve dolayısıyla ROLS (restricted, kısıtlı) $\hat{\beta}_R$ yansızdır ve OLS'den daha etkindir. Bu eşitsizliğin sağlaması doğrudan yapılabilir; bu sonucu göstermenin başka bir yolu aşağıda verilmiştir. Bu GM teoremi ile çelişmekte midir? Hayır. Neden?

Ortaya çıkmaktadır ki, GM modelinde *kısıtlamalı* parametre $\{\beta \in \mathbb{R}^K : R\beta = r\}$ ile, ROLS en küçük varyans yansız doğrusal tahmincidir. Benzer şekilde, GM normal modelinde *kısıtlamalı* parametre $\{\beta \in \mathbb{R}^K : R\beta = r\}$ ile, ROLS aynı zamanda en küçük varyans yansız tahmincidir. Dikkat ederseniz, daha önceden kısıtlamasız parametre uzayı olan $\{\beta \in \mathbb{R}^K\}$ ile çalışmıştık.

Kısıtlanmış regresyon problemi kısıtlamasız bir regresyona dönüştürülebilir. Bu husus ilk olarak $\beta_1 + \beta_2 = 1$ kısıtlaması kullanılan bir örnekle gösterilebilir. Şunu yazınız

$$Y = X_1\beta_1 + X_2\beta_2 + \epsilon = X_1(\beta_1 + \beta_2 - 1) + X_2\beta_2 - X_1(\beta_2 - 1) + \epsilon = (X_2 - X_1)\beta_2 - X_1 + \epsilon$$

veya

$$Y - X_1 = (X_2 - X_1)\beta_2 + \epsilon,$$

Yeni modelin Gauss-Markov varsayımlarını sağladığını kolaylıkla kontrol etmek şu parametre uzayı ile mümkündür

$$\theta = (\beta_2, \sigma^2, F_{\epsilon|X}, F_X),$$

burada $\beta_2 \in \mathbb{R}$ kısıtlamasızdır. Yani, son ifadeye LS uygulanarak elde edilen $\hat{\beta}_{R2}$ etkin doğrusal tahmincidir (normallik varsayımında etkin tahminci). Benzer durum $\hat{\beta}_{R1} = 1 - \hat{\beta}_{R2}$ için de geçerlidir, çünkü bu $\hat{\beta}_{R2}$'nin doğrusal bir işlevselidir. Bu sebeple, ROLS $\hat{\beta}_R$ kısıtlamasız doğrusal en küçük kareler tahmincisi olan $\hat{\beta}_2$'den daha etkindir.

Bu fikir rahatlıkla genelleştirilebilir. Genel durumu bozmayacak şekilde, bağlaştıranların sıralamasını yeniden düzenleyebiliriz, şöyle ki

$$R = [R_1 \quad R_2],$$

burada $R_1$ bir tam kerte $p$ olan $p \times p$ matrisidir. $H_0 : R\beta = r$'yi uygulamak, $R_1\beta_1 + R_2\beta_2 = r$ veya $\beta_1 = R_1^{-1}(r - R_2\beta_2)$'ye eşdeğer olur, bu durumda da

$$Y = X_1\beta_1 + X_2\beta_2 + \epsilon \stackrel{H_0}{=} X_1(R_1^{-1}(r - R_2\beta_2)) + X_2\beta_2 + \epsilon,$$

bu da şunu gösterir

$$Y - X_1R_1^{-1}r = (X_2 - X_1R_1^{-1}R_2)\beta_2 + \epsilon\text{'dur.}$$

Bu da bize tekrar yeni bir bağımlı değişkenli ve yeni bir bağlaştıranlı model verir ve önceki GM çerçevesine uygun olur. Bu çerçevede $\hat{\beta}_{2R}$ tahmini ve aynı zamanda $\hat{\beta}_{1R} = R_1^{-1}(r - R_2\hat{\beta}_{2R})$ tahmini etkindir.

3.10. **Sonlu Örneklemde Normallik Olmadan Çıkarsama.** Sonlu-örneklemde Monte-Carlo (MC) yönteminin temel fikri aşağıdaki örnekle şu şekilde gösterilebilir.

**Örnek 1.** Varsayalım $Y = X\beta + \epsilon$, $E[\epsilon|X] = 0$, $E[\epsilon\epsilon'|X] = \sigma^2 I$, $\epsilon = (\epsilon_1, ...\epsilon_n)' = \sigma U$ olsun, burada

$$U = (U_1, ..., U_n)|X \text{ , } F_U \text{ kuralı ile iid'dir} \tag{4}$$

burada kullanılan $F_U$ bilinmektedir. Örneğin, $F_U = t(3)$ alınırsa, birçok finansal getiri verisetlerinin özelliklerini $F_U = N(0,1)$'den daha iyi sağlayacaktır.

Şimdi $H_0 : \beta_j = \beta_j^0$'a karşı $H_A : \beta_j > \beta_j^0$'yi sınadığımızı düşünelim. $H_0$ altında

$$t_j = \frac{\hat{\beta}_j - \beta_j^0}{\sqrt{s^2 (X'X)^{-1}_{jj}}} \overset{H_0 \text{ ile}}{=} \frac{((X'X)^{-1}X'\epsilon)_j}{\sqrt{\frac{\epsilon' M_X \epsilon}{n-K}(X'X)^{-1}_{jj}}} = \frac{((X'X)^{-1}X'U)_j}{\sqrt{\frac{U' M_X U}{n-K}(X'X)^{-1}_{jj}}} \tag{5}$$

Dikkat ederseniz bu istatistik bilinmeyen parametre $\sigma^2$'yi hoş bir şekilde sadeleştirmektedir.

Bu testin p-değeri Monte-Carlo ile hesaplanabilir. T-istatistiğinin $H_0$ altında üretilebilecek çekim (draw) değerlerini simule edelim:

$$\{t^*_{j,d}, d = 1, ..., B\}, \tag{6}$$

burada $d$ çekim değerlerini sıralar, $B$ toplam çekim sayısıdır ve büyük olması gerekmektedir. Her bir çekim değeri $t^*_{j,d}$'yi oluşturmak için, $U$'nun (4)'a bağlı olarak çekimlerini oluşturalım ve bunları (5)'in sağ tarafına oturtalım. Sonrasında p-değeri şu şekilde tahmin edilebilir

$$P_j = \frac{1}{B}\sum_{d=1}^{B} 1\{t^*_{j,d} \geq t_j\}, \tag{7}$$

burada $t_j$ t-istatistiğinin görgül (empirical) değeridir. $H_0 : \beta_j = \beta_j^0$'a karşı $H_A : \beta_j \neq \beta_j^0$ 'ı sınamak için kullanılacak p-değeri şu şekilde tahmin edilebilir

$$P_j = \frac{1}{B}\sum_{d=1}^{B} 1\{|t^*_{j,d}| \geq |t_j|\}. \tag{8}$$

Güven bölgeleri ve testler için t-istatistiğine bağlı kritik değerler, (6) örnekleminin uygun bölütleri (quantile) elde edilerek oluşturulabilir.

**Örnek 2.** Şimdi, önceki örneği şu şekilde genelleştirelim: $F_U$ bilinmeyen sorunlu (nuisance) parametre $\gamma$'ya bağlı olsun ve bunun da gerçek değeri $\gamma_0$'ın $\Gamma$ bölgesinde olduğu bilinsin. Bağlılığı $F_U(\gamma)$ ile ifade edelim.

Örneğin, varsayalım $F_U(\gamma)$ "serbestlik derecesi" parametresi $\gamma \in \Gamma = [3, 30]$ olan bir t-dağılımı olsun;, bu, çok ağır kuyruktan hafif kuyruğa kadar farklı kuyruk hareketi gösteren dağılımları yuvalama (nest) imkanı sağlar. Normal dağılım da $\gamma = 30$ ile yaklaşık olarak yuvalanır.

Sonrasında, her bir $\gamma \in \Gamma$ için p-değerini elde edelim ve bunu $P_j(\gamma)$ ile ifade edelim. Sonra şunu

$$\sup_{\gamma \in \Gamma} P_j(\gamma)$$

sınama amaçlı kullanalım. $\gamma_0 \in \Gamma$ olduğu için, bu değer gerçek P-değeri $P_j(\gamma_0)$ için geçerli bir üst sınırdır. Benzer şekilde, her bir $\gamma \in \Gamma$ için kritik değerler elde edilebilir ve *en az tercih edilen* kritik değer kullanılabilir. Sonuçta elde edilen güven bölgeleri eğer $\Gamma$ büyük olursa oldukça tutucu (conservative) olabilir; ne var ki, son paragrafa bakınız.

Doğal olarak oluşan soru şudur: Neden gerçek parametre $\gamma_0$'ın tahmini olan $\hat{\gamma}$'ı MC ile $\gamma = \hat{\gamma}$'a eşitleyerek kullanıp $P_j(\hat{\gamma})$'ı ve kritik değerleri elde etmeyelim? Bu yöntem parametrik *özçıkarım* (*bootstrap)* olarak bilinir. Özçıkarım, bir MC yöntemidir ve basitçe, p-değerlerini ve kritik değerleri tahmin edilen veri üretme süreci (data generating process) ile elde etmek için kullanılır. Özçıkarım yanaşık olarak geçerli çıkarsama sağlar, ancak özçıkarım her zaman geçerli sonlu örneklem çıkarsaması sağlamayabilir. Ancak, özçıkarım sonlu örneklemlerde sıklıkla yanaşık yaklaşımın sağladığından daha doğru (accurate) çıkarsama sağlar.

Yukarıdaki sonlu örneklem yaklaşımı veriye-bağlı $\Gamma$ ile de kullanılabilir. Veriye bağlı sorunlu parametre kümesini $\hat{\Gamma}$ olarak ifade edelim. Eğer $\hat{\Gamma}$ kümesi $\gamma_0$'ı $1 - \beta_n$ ihtimaliyle içerirse, ki burada $\beta_n \to 0$'dır, p-değerinin tahminini şu olacak şekilde ayarlayabiliriz

$$\sup_{\gamma \in \hat{\Gamma}} P_j(\gamma) + \beta_n.$$

Büyük örneklemlerde, şunu bekleyebiliriz

$$\sup_{\gamma \in \hat{\Gamma}} P_j(\gamma) + \beta_n \approx P_j(\gamma_0),$$

bu, $\hat{\Gamma}$ kümesi $\gamma_0$'a yakınsadığında (converge) ve $\beta_n \to 0$ ise geçerlidir. Böylece, sonlu-örneklem yöntemi büyük örneklemlerde etkin olabilir, ancak sonlu örneklemlerde de geçerliliğini korur. Yanaşık yöntem veya özçıkarım'ın sonuncu ifadeyi sağlaması illa gerekmemektedir. Bu da yöntemleri ayırmaktadır. Ancak, başkalarınca ifade edildiği gibi, sonlu-örneklem yöntemi "süslü (fancy) özçıkarım" olarak düşünülebilir.

**Örnek 3.** **(Ödev)** Temin'in (2005) makalesini düşünelim. Makale, Roma İmparatorluğu'nda eyaletin Roma'ya olan mesafesinin buğday fiyatları üzerindeki etkisini modellemektedir. Çalışmada sadece 6 tane gözlem vardır. Etkinin sıfıra eşit olduğu sıfır önsavına karşılık etkinin negatif olduğu almaşık önsavı test etmek için kullanılacak p-değerlerini hesaplayınız. Önce normal bozuklukların olduğu duruma bakınız (bu durum için simulasyon yapılmasına gerek yoktur), sonrasında ikinci durum olarak bozuklukların 8 ve 16 "serbestlik derecelerine" sahip t-dağılımı takip ettikleri durumu analiz ediniz.

### 3.11. **Ek: Kare Kayıp Altında Biraz Formal Karar Teorisi (Some Formal Decision Theory Under Squared Loss).**

Amemiya (1985) tahmincilerin etkinliğini tartışmak için şu kuralcılıkları oluşturmaktadır.

1. $\hat{\beta}$ ve $\beta^*$ bir skalar parametre $\beta$ için skalar tahminciler olsun. Eğer $E_\beta(\hat{\beta} - \beta)^2 \leq E_\beta(\beta^* - \beta)^2$, $\forall \beta \in \mathcal{B}$ ise, $\hat{\beta} \succcurlyeq$ (en az onun kadar iyi) $\beta^*$'dır.

"Daha iyi" nin tanımı karesel (quadratic) kayba bağlıdır.

2. Bazı $\beta \in \mathcal{B}$ için, eğer $\hat{\beta} \succcurlyeq \beta^*$ ve $E_\beta(\hat{\beta} - \beta)^2 < E_\beta(\beta^* - \beta)^2$ ise, $\hat{\beta}$ **daha iyidir** (daha etkindir, $\succ$) $\beta^*$'dan,

3. $\hat{\beta}$ ve $\beta^*$ vektör parametresi $\beta$ için **vektör** tahminler olsun. Eğer tüm $c \in \mathbb{R}^k$, $c'\hat{\beta} \succcurlyeq c'\beta^*$ ($c'\beta$'yı tahmin etmek için) ise $\hat{\beta} \succcurlyeq \beta^*$'dır.

4. Eğer bazı $c \in \mathbb{R}^k$ için $c'\hat{\beta} \succ c'\beta^*$ ve tüm $c \in \mathbb{R}^k$ için $c'\hat{\beta} \succcurlyeq c'\beta^*$ ise, $\hat{\beta} \succ \beta^*$'dır.

Açıkça görülmelidir ki Tanım 3 ve Tanım 5 eşdeğerdir.

5. Eğer $A_\beta \equiv E_\beta(\hat{\beta} - \beta)(\hat{\beta} - \beta)'$ ise ve $B_\beta \equiv E_\beta(\beta^* - \beta)(\beta^* - \beta)'$ ise, $\hat{\beta} \succcurlyeq \beta^*$'dır; burada $A_\beta - B_\beta$ $\forall$ $\beta \in \mathcal{B}$ için yarı-negatif tanımlıdır veya matris kapsamında $A_\beta \leq B'_\beta$'dır.

6. $\hat{\beta}$ eğer bu sınıfta daha iyi tahminci yoksa, tahminci sınıfı içinde en iyidir.

## 4. Büyük Örneklemlerde Tahmin ve Temel Çikarsama

Bu kısım için iyi bir referans Newey'nin ders notlarıdır. Aşağıda, sadece temel sorunlara değinilecektir.

### 4.1. Temel Kurgu ve Bunların İfade Ettikleri.

Büyük örneklemde geçerli çıkarsamaları önceki GM modelinin izin verdiğinden çok daha genel koşullarda yapabiliriz.

Çalışabileceğimiz bir grup yeterli (sufficient) koşul şudur:

**L1** $y_t = x_t'\beta + e_t, \quad t = 1, ..., n$

**L2** $Ee_t x_t = 0, \quad t = 1, ..., n$

**L3** $X'X/n \rightarrow_p Q$ sonlu ve tam kerte

**L4** $X'e/\sqrt{n} \rightarrow_d N(0, \Omega)$, burada $\Omega$ sonludur ve yozlaşan değildir (non-degenerate).

**Tartışma:**

1) L1 ve L2, $\beta$'nın $E[y_t|x_t]$ koşullu beklenti fonksiyonuna en iyi doğrusal yaklaştırmayı tanımlayan parametre olduğunu ifade ederler. L1-L4 ile, OLS $\hat{\beta}$, $\beta$ için tutarlı ve yanaşık (yaklaştırma manasında) olarak normal dağılan tahminci şeklinde ortaya çıkar.

2) Şu ayrıştırmayı yapabiliriz

$$e_t = a_t + \varepsilon_t,$$

burada $\varepsilon_t = y_t - E[y_t|x_t]$ ve

$$a_t = E[y_t|x_t] - x_t'\beta\text{'dır.}$$

Hata terimi bu sebeple şunların toplamıdır: $y_t$'nin koşullu ortalamadan sapması olarak tanımlanan"alışıldık" (usual) bozukluk $\varepsilon_t$ ve doğru fonksiyon $E[y_t|x_t]$ yerine doğrusal işlevsel biçim kullanılmasından oluşan hata olarak ortaya çıkan yaklaştırma (belirginleştirme) hatası $a_t$. Önceki GM çerçevesinde, yaklaştırma hatası olmadığını varsaymıştık. Şu anki çerçevede, buna "gücümüz yeter". Açıkça belirtebiliriz ki sadece gerçek koşullu ortalama fonksiyonuna bir yaklaşıklığı tahmin ediyoruz ve açıkça hesaba katabiliriz ki yaklaşıklık hatası farklı yayılımın (heteroscedasticity) (neden?) tahmincimizin varyansını etkileyen önemli bir sebebidir.

3) L3 bir önceki özdeşleştirme (identification) koşulunun sadece bir örnekselidir (analog). Aynı zamanda, bağlaştıranların $t$ çarpımının $\{x_t x_t', t = 1, ...n\}$ Büyük Sayılar Yasası'nı (BSY, Law of Large Numbers, LLN) sağlaması gerekmektedir; bu

şekilde onların üzerine bir kararlılık (stability) yüklemektedir. Bu koşul yönelimli bağlaştıranlar (trending regressors) eklenerek gevşetilebilir (bakınız, örneğin, Newey'in notları veya Amemiya'nın "Advanced Econometrics" kitabı).

4) L4, $\{x_t e_t, t = 1, ..., n\}$ dizisinin bir MLT'i sağlaması gerektiğini ifade eder. Bu koşul, daha önce hataların normal olması hakkında yaptığımız varsayımdan oldukça daha geneldir. Büyük örneklemlerde, bu koşul bizi daha önce normallik altında elde ettiğimiz tahmin ve çıkarsama sonuçlarının benzerlerine götürecektir.

**Önerme 4.1.** *L1-L4 ile,* $\hat{\beta} \to_p \beta$.

Örneklem büyüklüğü arttıkça, $\hat{\beta}$ tahmini $\beta'$ya yaklaşacaktır. Açıkçası, tutarlılık için, L4'ü daha az kısıtlayıcı olan $X'e/n \to_p 0$ gereksinimi ile değiştirebiliriz.

**Önerme 4.2.** *L1-L4 ile,* $\sqrt{n}(\hat{\beta} - \beta) \to_d N(0, V)$, $V = Q^{-1}\Omega Q^{-1}$.

Bu sonuç büyük örneklemlerde $\hat{\beta}$'nın $\beta$ ortalaması ve $V/n$ varyansı ile yaklaşık olarak normal dağıldığını önermektedir:

$$\hat{\beta} \approx_d N(\beta, V/n).$$

**Önerme 4.3.** *L1-L4 ile, varsayalım* $\hat{V} \to V$*'dir, bu durumda*

$$t_j := (\hat{\beta}_j - \beta_j)/s.e(\hat{\beta}_j) := (\hat{\beta}_j - \beta_j)/\sqrt{\hat{V}_{jj}/n} \to_d N(0, 1).$$

*ve eğer* $R$ *için tam kerte* $p$ *ise,* $R\beta = r$ *ile*

$$W = (R\hat{\beta} - r)'[R(\hat{V}/n)R']^{-1}(R\hat{\beta} - r) \to_d \chi^2(p).$$

Büyük örneklemlerde uygun şekilde oluşturulan $t$-istatistiği ve $W$-istatistiği yaklaşık olarak standart normal değişkeni ve $p$ serbestlik dereceli ki-kare değişkeni ile dağılırlar; yani $t \approx_d N(0, 1)$ ve $W \approx_d \chi^2(p)$.

Bu sonuçların temel kullanımı sonlu-örneklem durumu ile tamamen aynıdır, istisna olarak artık hepsi yaklaşık olarak ifade edilmiştir. Örneğin, sıfır önsavı ile, bir t-istatistiği $t_j \to_d N(0, 1)$'ı sağlar ve bu da şunu ifade eder

$$\lim_{n\to\infty} Pr[t_j < c] = Pr[N(0, 1) < c] = \Phi(c),$$

bu her bir $c$ için geçerlidir, çünkü $\Phi$ süreklidir (continuous). Sonlu ve büyük örneklemlerde, sadece şu geçerlidir

$$Pr[t_j < c] \approx \Phi(c),$$

burada yaklaşımın kalitesi çeşitli durumlara bağlı olarak iyi ya da zayıf olabilir.

**Dikkat ediniz:** Kısıtlamalı en küçük kareler için, ki bu kısıtlamasız OLS'in bir doğrusal dönüşümüdür, yanaşık sonuçlar ve diğer test istatistikleri (örn. LM) yukarıda sunulan sonuçlardan kolayca çıkarılabilir.

Sonuçları ispat etmek için kullanacağımız temel araçlar Sürekli Eşleşme Teoremi (SET, Continuous Mapping Theorem, CMT) ve Slutsky Teoremi'dir. Bu sonuçlarda esas olan metrik uzaylar sonlu-boyutlu Euclid uzaylarıdır (Euclidian spaces).

**Önsav 4.1** (SET)**.** *$X$ bir rastsal eleman olsun ve $x \mapsto g(x)$ her bir $x \in D_0$ için sürekli olsun, burada bir olasılığı ile $X \in D_0$'dır. Varsayalım $X_n \to_d X$ olsun, bu durumda $g(X_n) \to_d g(X)$; eğer $X_n \to_p X$ ise, o zaman $g(X_n) \to_p g(X)$'dır.*

Bu önsavın ispatı yukarıda belirtilen bir temsil teoreminin uygulamasından ve sonra da süreklilik hipotezinin kullanılmasından elde edilir. Takip eden önsav sürekli eşleşme teoreminin bir sonucudur (corollary).

**Önsav 4.2** (Slutsky Önsavı)**.** *Matris $A_n \to_p A$ ve vektör $a_n \to_p a$ olduğunu varsayalım, burada matris $A$ ve vektör $a$ sabittir. Eğer $X_n \to_d X$ ise, o zaman $A_n X_n + a_n \to_d AX + a$'dır.*

**Önerme 1'in İspatı:** Sırasıyla L4 and L3 koşulları şunu ifade eder

$$X'e/n \xrightarrow{p} 0, \; X'X/n \xrightarrow{p} Q. \tag{2}$$

Sonrasında, $Q$'nun tekilsizliği (nonsingularity) ile, bir matrisin tersinin matrisin elemanlarının her bir tekilsiz matriste sürekli fonksiyonu olmaları bilgisi ve Slutsky Önsavı ile şu elde edilir

$$\hat{\beta} = \beta + (X'X/n)^{-1} X'e/n \xrightarrow{p} \beta + Q^{-1} \cdot 0 = \beta.$$

**Önerme 2'nin İspatı:** L4 and L3 koşulları sırasıyla şunu ifade eder

$$X'e/\sqrt{n} \xrightarrow{d} N(0, \Omega), \; X'X/n \xrightarrow{p} Q. \tag{2}$$

Slutsky Önsavı'ndan şu elde edilir

$$\sqrt{n}(\hat{\beta} - \beta) = (X'X/n)^{-1} X'e/\sqrt{n} \xrightarrow{d} Q^{-1} N(0, \Omega) = N(0, Q^{-1} \Omega Q^{-1}).$$

**Önerme 3'ün İspatı:** $\hat{V} \xrightarrow{p} V$, $V_{jj} > 0$ kullanılarak ve SET ile, $\left(V_{jj}/\hat{V}_{jj}\right)^{1/2} \xrightarrow{p}$ 1'dir. Slutsky Teorem'i kullanılarak şu çıkar

$$\left(\hat{\beta}_j - \beta_j\right) / \left[V_{jj}\right]^{1/2} = (V_{jj}/\hat{V}_{jj})^{1/2}\sqrt{n}(\hat{\beta}_j - \beta_j)/\sqrt{V_{jj}} \xrightarrow{d} 1 \cdot N(0,1) = N(0,1).$$

$\Sigma = RVR'$ olsun. $\Sigma$ matrisi $R$'nin $p$ kertesine sahip olması ve $V$'nin tekilsiz olması sebebiyle tekilsizdir, yani SET ile, $\hat{\Sigma}^{-1/2} \xrightarrow{p} \Sigma^{-1/2}$'dir. Ayrıca, Slutsky Önsavı ile $Z_n = \widehat{\Sigma}^{-1/2} R\sqrt{n}\left(\hat{\beta} - \beta\right) \xrightarrow{d} Z = \Sigma^{-1/2}N(0,\Sigma) =_d N(0,I)$. Sonrasında SET ile, $W = Z_n' Z_n \rightarrow_d Z'Z =_d \chi^2(p)$'dir.

4.2. **Bağımsız Örneklemler.** Burada iki modeli değerlendireceğiz:

**IID Modeli:** Varsayalım *(a) L1 ve L2 sağlansın, (b)* $(y_t, x_t)$ *vektörleri* $t$ *üzerinde bağımsızdırlar ve özdeş dağılırlar, (c)*

$$\Omega = Var[x_t e_t] = E[e_t^2 x_t x_t']$$

*sonludur ve yozlaşmayandır (non-degenerate) (tam kerte), ve*

$$Q = E[x_t x_t']$$

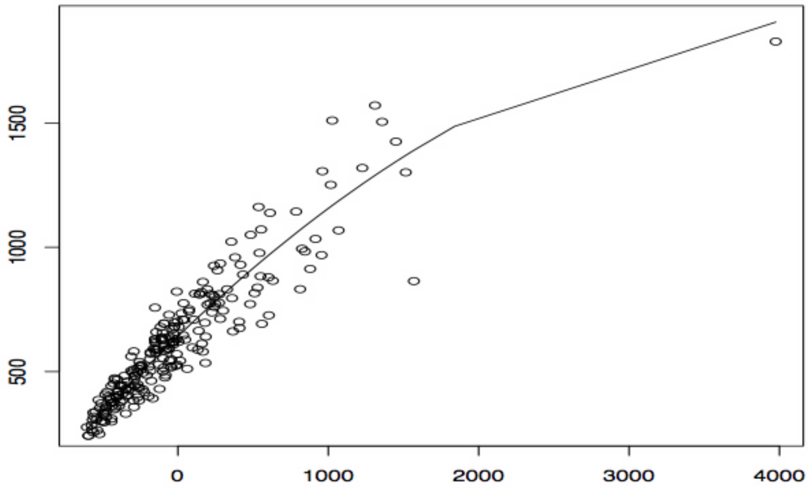
*sonludur ve tam kertedir.*

Bu koşulun $y_t$ ve $x_t$ arasındaki ilişkiyi hiçbir şekilde kısıtlamadığını vurgulamak gerekir; sadece $(y_t, x_t)$'nin birleşik dağılım fonksiyonun $t$'ye bağımlı olmaması ve verinin $t$ üzerinde bağımlılığı olmaması gerekmektedir.

Bu model, hata terimi $e_t = \varepsilon_t + a_t$ 'de farklı yayılımın iki kaynağına izin vermektedir– biri $\varepsilon_t$'nin farklı yayılımı ve diğeri de yaklaşıklı hatası $a_t = E[y|x_t] - x_t'\beta$'nın farklı yayılımıdır. Farklı yayılımdan kastımız $E[e_t^2|x_t]$'nin $x_t$'ye bağlı olmasıdır.

**Örnek:** Daha önce konuşulan Engel eğrisi örneğini hatırlayınız, ki örnekte $\varepsilon_t$ açıkça farklı yayılımlı idi. Bu sebeple, $e_t$ de farklı yayılımlıdır. İktisattaki birçok regresyon modeli farklı yayılım bozukluğuna sahiptir.

Gıda Harcaması




Gelir

ŞEKİL 4

**Örnek:** Ücret nüfus sayımı verisinde basit işlevsel biçimler için $a_t \neq 0$ olduğunu görmüştük. Bu sebeple, $a_t \neq 0$ olduğu için $e_t$ farklı yayılımlı olmalıdır.

**Teorem 4.1.** *IID modelinin koşulları L1-L4'ün yukarıda tanımlanan $\Omega$ ve $Q$ ile sağlandığını belirtmektedir. Bu sebeple, Önerme 1-3'ün sonuçları da sağlanmaktadır.*

**İspat:** Bu ispat Khinchine BSY ve çok değişkenli Lindeberg-Levy MLT'den elde edilmektedir.

$V$'nin varyansının tutarlı bir tahmincisi (farklı yayılım dirençli tahminci (heteroscedasticity robust estimator)) şu şekilde ifade edilir:

$$\hat{V} = \hat{Q}^{-1}\hat{\Omega}\hat{Q}^{-1}, \quad \hat{\Omega} = E_n[e_t^2 x_t x_t'], \quad \hat{Q} = X'X/n.$$

**Teorem 4.2.** *IID modelinin koşulları ve ek varsayımlar (örn. $x_t$'nin bağımlı dördüncü beklemleri (bounded fourth moments)) $\hat{\Omega} \to_p \Omega$ ve $\hat{Q} \to_p Q$ olduğunu ifade eder, bu şekilde $\hat{V} \to_p V$'dır.*

**İspat:** $\hat{Q} \to_p Q$'nun tutarlılığı Khinchine BSY'den çıkmaktadır ve $\hat{\Omega} \to_p \Omega$'ın tutarlılığı şu şekilde gösterilebilir. Gösterimi basitleştirmek için skalar duruma bakalım. Şunu elde etmiş durumdayız

$$\hat{e}_t^2 = e_t^2 - 2(\hat{\beta} - \beta)' x_t e_t + (\hat{\beta} - \beta)^2 x_t^2.$$

İki tarafı da $x_t^2$ ile çarpalım ve $t$ üzerinden ortalama alarak şunu elde edelim

$$E_n[\hat{e}_t^2 x_t^2] = E_n[e_t^2 x_t^2] - 2(\hat{\beta} - \beta) E_n[e_t x_t^3] + (\hat{\beta} - \beta)^2 E_n[x_t^4].$$

Sonrasında, $E_n[e_t^2 x_t^2] \to_p E[e_t^2 x_t^2]$, $E_n[x_t^4] \to_p E[x_t^4]$ ve Khinchine BSY ile $E_n[e_t x_t^3] \to_p E[e_t x_t^3]$'dir, çünkü $E[x_t^4]$ varsayımdan dolayı sonludur ve $E[e_t x_t^3]$ şundan dolayı sonludur

$$|E[e_t x_t x_t^2]|^2 \leq E[e_t^2 x_t^2] E[x_t^4]$$

bu da Cauchy-Schwartz eşitsizliğinden ve $E[e_t^2 x_t^2]$ ve $E[x_t^4]$'nin varsayılan sonluluğundan kaynaklanır. $\hat{\beta}$'nın tutarlılığından, yukarıdaki özelliklerden ve SET'den tutarlılığı elde ederiz. □.

**Eşit Yayılımlı (Homoscedastic) IID Modeli:** *IID modelinin (a)-(c) koşullarına ek olarak, varsayalım $E[e_t|x_t] = 0$ ve $E[e_t^2|x_t] = \sigma^2$ olsun, bu durumda $Var[e_t^2|x_t] = \sigma^2$'dir, bu da şu şekilde sadeleşir*

$$\Omega = \Omega_0 := \sigma^2 E[x_t x_t'].$$

Bu modelde yaklaşıklık hatası yoktur, yani $e_t = \varepsilon_t$'dir ve farklı yayılım yoktur. Bu model normalliğin dayatılmadığı bir Gauss-Markov modelidir.

**Teorem 4.3.** *Eşit yayılımlı iid modelinde, L1-L4 şunlar ile sağlanır: $\Omega_0 = \sigma^2 E[x_t x_t']$ ve $Q = E[x_t x_t']$, yani $V = V_0 := \sigma^2 Q^{-1}$.*

**İspat.** Bu, bir önceki sonuç kullanılarak doğrudan çıkartılır.

Eşit yayılımlı iid modeli için, $V$'nin varyansının tutarlı bir tahmincisi (dirençli olmayan (non-robust)) şu şekilde ifade edilir:

$$\hat{V}_0 = s^2 (X'X/n)^{-1}$$

**Teorem 4.4.** *Eş yayılımlı iid modelinin koşulları $\hat{V}_0 \to_p V_0$ olduğunu belirtir.*

**İspat.** $Q^{-1}$ matrisi $(X'X/n)^{-1}$ tarafından $X'X/n \xrightarrow{p} Q$ ile tutarlı olarak tahmin edilir, bu da BSY ve SET ile sağlanır. $s^2 = \hat{\varepsilon}'\hat{\varepsilon}/(n-K)$ olduğunu gösterebiliriz,

burada $\hat{\varepsilon} = y - X\hat{\beta}$'dir. $X'\varepsilon/n \xrightarrow{p} 0$ ve $X'X/n \xrightarrow{p} Q$ kullanılarak şu elde edilir
(4)

$$\begin{array}{cccccccc} s^2 = & [\frac{n}{n-K}] & [\frac{\varepsilon'\varepsilon}{n}+ & 2(\beta-\hat{\beta})' & (\frac{X'\varepsilon}{n})+ & (\beta-\hat{\beta})' & (\frac{X'X}{n}) & (\beta-\hat{\beta})] \underset{(SET)}{\to_p} \sigma^2, \\ & \downarrow & \downarrow p & \downarrow p & \downarrow p & \downarrow p & \downarrow p & \\ & 1 & \sigma^2 & 0 & 0 & 0 & Q & \\ & & \text{(BSY)} & \text{(OLS)} & \text{(BSY)} & \text{(OLS)} & \text{(BSY)} & \end{array}$$

burada (OLS) ile OLS tahmincisinin tutarlılığı kastedilmektedir. Yani, SET ile, $s^2(X'X/n)^{-1} \xrightarrow{p} \sigma^2 Q^{-1}$. $\square$.

**Yorum:** Farklı yayımlılık ile $V \neq V_0$'dır, ve $V_0$ ile karşılaştırıldığında $V$ daha büyük ya da küçük olabilir. Bunun sağlandığına dikkat ediniz. Uygulamada, $V$ sıklıkla $V_0$'dan büyüktür.

4.3. **Bağımlı Örneklemler.** Birçok makroekonomik ve finansal durumlarda zaman serileri serisel olarak bağımlıdır (serially dependent) (yani $t$ üzerinden bağımlıdır). Bunun bazı örneklerini düşünebilirsiniz.

Bağımlılık birçok şekilde modellenebilir. Bazı parametrik modelleri GLS ile bağlantılı olarak MIT'de verilen 14.382 dersinde görebilirsiniz. Burada temel parametrik olmayan (non-parametric) modelleri tanıtacağız.

Takip eden kısımlarda veriyi şu şekilde düşünebiliriz: $z_t = (y_t, x_t')', t = 1, ..., n$ sonsuz bir akım olan $\{z_t\} = \{z_t, t = \pm 1, \pm 2, ...\}$'nin bir alt kümesidir.

Eğer her bir $k \geq 0$ için $(z_t, ..., z_{t+k})$ dağılımı $(z_1, ..., z_{1+k})$ dağılımına eşitse, yani $t$'ye bağlı değilse, $\{z_t\}$ *durağan (stationary)* olarak tanımlanır. Bunun sonucu olarak, ortalama ve kovaryans durağanlığı da elde edilir: $E[z_t] = E[z_1]$ her bir $t$ için ve $E[z_t z_{t+k}'] = E[z_1 z_{1+k}']$ her bir $t$ için.

**(a) Karıştırma (Mixing).** Karıştırma, dönemsel bağımlılığı ifade ederken kullanılan oldukça matematiksel bir düşünce tarzıdır.

$\{h_t\}$ durağan bir süreç olsun. $k \geq 1$ için şunu tanımlayın

$$\alpha_k = \sup_{A,B} \Big| P(A \cap B) - P(A)P(B) \Big|$$

burada sup $A \in \sigma(h_0, h_{-1}, h_{-2}, ...,)$ ve $B \in \sigma(h_k, h_{k+1}, ...)$ durumları üzerinden hesaplanır. Basit olarak, bu sigma-alanlarını (sigma fields) parantezler içindeki değişkenler tarafından oluşturulmuş bilgi (information) olarak düşünebiliriz. Eğer $k \to \infty$ iken $\alpha_k \to 0$ ise, $\{h_t\}$ *kuvvetli karıştırma (strongly mixing)* veya *alfa* *karıştırma* (*alpha mixing*) olarak adlandırılır. Eğer veri iid ise, her bir $k$ için $\alpha_k = 0$'dır.

Karıştırma koşulu, birbirinden $k$ zaman birimi ayrılmış veri bloklarının bağımlılığının $k$ arttığında dağılacağını (azalacağını) belirtir. Bir çok parametrik modelin bazı düzenlilik (regularity) koşulları altında karıştırma olduğu gösterilmiştir.

**Önsav 4.3** (Bir Ergodik BSY)**.** *Eğer $\{h_t\}$ sonlu bir ortalama $E[h_t]$ ile durağan kuvvetli karıştırma ise, o zaman $E_n[h_t] \to_p E[h_t]$'dır.*

Dikkat ediniz. Bu BSY'nın daha genel bir uyarlaması Birkhoff ergodik teoremi olarak adlandırılır.[8]

**Önsav 4.4** (Gordin'in MLT'i)**.** *Eğer $\{h_t\}$ bir kuvvetli durağan karıştırma süreci ise ve $E[h_t] = 0$, $\sum_{k=1}^{\infty} \alpha_k^{\delta/(2+\delta)} < \infty$, $E\|h_t\|^{2+\delta} < \infty$ ise, o zaman*

$$\sum_{t=1}^{n} h_t/\sqrt{n} \to_d N(0, \Omega),$$

*burada*

$$\Omega = \lim_n Var(\sum_{t=1}^{n} h_t/\sqrt{n}) = \lim_n \Big(\Omega_0 + \sum_{k=1}^{n-1} \frac{n-k}{n}(\Omega_k + \Omega_k')\Big) = \Omega_0 + \sum_{k=1}^{\infty}(\Omega_k + \Omega_k') < \infty,$$

burada $\Omega_0 = Eh_1h_1'$ ve $\Omega_k = Eh_1h_{1+k}'$'dır.

Burada karıştırmanın derecesindeki (rate) kısıt ve beklem koşulu kovaryansın sonlu miktar olan $\Omega$'a toplandığını belirtir; aşağıdaki nota da bakınız. Bu gerçekleşirse, seriye zayıf-bağımlı (weakly-dependent) denir.

Kovaryans toplanabilirliğini ima eden bir yeterli koşul şu şekilde düşünülebilir: $k \to \infty$'a giderken, şuna sahip olmak yeterlidir

$$c > 1 \text{ için } \Omega_k/k^{-c} \to 0.$$

Kovaryanslar $1/k$'dan daha hızlı azalmalıdır. Bu gerçekleşmezse, serilerin uzun hafızaya (long memory) sahip oldukları söylenir. Finansta kullanılan yüksek frekanslı veri genelde uzun hafızaya sahip olarak düşünülür; bunun sebebi kovaryansın çok yavaş şekilde azalmasıdır. Uzun hafızaya yönelik yanaşık kuramı burada anlatılan yanaşık kuramından yeteri kadar farklıdır [Referans olarak bakınız: H. Koul.]

[8] Bakınız, örn. http://mathworld.wolfram.com/BirkhoffsErgodicTheorem.html

Dikkat ediniz. Gordin'in teoreminde, kovaryans toplanabilirliği Ibragimov'un durağan seriler için karıştırma eşitsizliğinden çıkartılır (burada skalarlar için ifade edilmiştir):

$$|\Omega_k| = |Eh_t h_{t+k}| \leq \alpha_k^{1-\gamma}[E[h_t]^p]^{1/p}[E[h_t]^q]^{1/q}, \quad \frac{1}{p}+\frac{1}{q} = \gamma \in (0,1).$$

$p = 2+\delta$ olarak belirlendiğinde, kovaryans toplanabilirliği $\sum_{k=-\infty}^{\infty}|\Omega_k| < \infty$ karıştırma katsayıları üzerine konulan kısıtlamadan çıkartılır.

**Teorem 4.5.** *Varsayalım $\{(y_t, x_t)\}$ durağandır, kuvvetli karıştırmadır ve L1 ve L2 geçerlidir. Varsayalım $\{\tilde{h}_t\} = \{x_t x_t'\}$ sonlu ortalamaya sahiptir, bu taktirde L3 koşulu $Q = E[x_t x_t']$ ile geçerlidir. Varsayalım $\{h_t\} = \{x_t e_t\}$ Gordin'in koşullarını sağlasın, bu taktirde L4 yukarıda belirtilen $\Omega$ formu ile geçerlidir.*

**İspat:** Sonuç önceki iki teoremden ve karıştırmanın tanımından elde edilir.

Yukarıdaki formül, $\Omega$ için şu tahminciyi önerir:

$$\hat{\Omega} = \hat{\Omega}_0 + \sum_{k=1}^{L-1}\frac{L-k}{L}(\hat{\Omega}_k + \hat{\Omega}_k'),$$

burada $\hat{\Omega}_0 = E_n[h_t h_t'] = \frac{1}{n}\sum_{t=1}^{n} h_t h_t'$ ve $\hat{\Omega}_k = E_n[h_t h_{t+k}'] = \frac{1}{n-k}\sum_{t=1}^{n-k} h_t h_{t+k}'$'dir. Belli teknik koşullar altında ve budama gecikmesi (truncation lag) üzerinde koşullar ile, mesela $L/n \to 0$ ve $L \to \infty$ gibi, tahmincinin tutarlı olduğu gösterilmiştir.

$\hat{V} = \hat{Q}^{-1}\hat{\Omega}\hat{Q}^{-1}$ tahmincisi yukarıda belirtilen formun $\hat{\Omega}$ hali ile sıklıkla HAC tahmincisi ("farklı yayılımlı ve kendisiyle ilgileşimli tutarlı tahminci," heteroscedasticity and autocorrelation consistent, HAC) olarak adlandırılmıştır. Bazı düzenlilik (regularity) koşulları ile, tahminci gerçekten de tutarlıdır. Bu koşullar ve ispat için, bakınız Newey and West çalışması.

**(b) Martingale fark dizileri (MFD, Martingale difference sequences, MDS).** Veri zamansal olarak bağımlı olabilir ama kovaryanslar $\Omega_k$ yine de sıfır olabilirler ("hepsi geçirgen" (all-pass) serileri); sonuçta yukarıdaki $\Omega$ sadeleşerek $\Omega_0$ olur. MFD bunun gerçekleşmesinin bir örneğidir. MFD akılcı beklentiler (rational expectations) ile (Hansen and Singleton'a bakınız) ve aynı zamanda etkin piyasa hipotezi (efficient market hypothesis) ile (Fama'ya bakınız) bağlantılı olarak da önemlidirler. Referans olarak detaylı bir ders kitabı olan H. White, Asymptotic Theory for Econometricians'a bakınız.

$z_t$'nin bir elemanı $h_t$ olsun ve $I_t = \sigma(z_t, z_{t-1}, ...)$ olsun. $\{h_t\}$ süreci, eğer $E[h_t|I_{t-1}] = h_{t-1}$ ise, $I_{t-1}$ süzmesine (filtration) bağlı olarak bir martingale'dir. $\{h_t\}$ süreci, eğer $E[h_t|I_{t-1}] = 0$ ise, $I_{t-1}$ süzmesine bağlı olarak bir martingale fark dizisidir (difference sequence).

Örnek. Hall'ın karesel fayda ve akılcı beklentiler ile oluşturduğu temsili tüketici modelinde şuna sahibiz

$$E[y_t|I_{t-1}] = 0,$$

burada $y_t = c_t - c_{t-1}$ 'dir ve $I_t$ ise $t$ döneminde mevcut olan bilgidir. Yani, akılcı ve en iyileme (optimization) hedefleyen bir tüketici bugünkü tüketim miktarını takip eden dönemlerdeki tüketimin ortalamasında değişiklik olmayacağı beklentisi ile oluşturur.

**Önsav 4.5** (Billingsley'in Martingale MLT'i)**.** *$\{h_t\}$ durağan ve kuvvetli karıştırmalı bir martingale fark dizisi olsun ve $\Omega = E[h_t h_t']$ sonlu olsun. Bu durumda $\sqrt{n}E_n[h_t] \to_d N(0, \Omega)$'dır.*

Bu teorem sezgisel (intuitive) olarak manalıdır, çünkü $h_t$'ler özdeş dağılımlıdır ve ilgileşimsizdirler. Aslında, $k \geq 1$ için $E[h_t h_{t-k}'] = E[E[h_t h_{t-k}'|I_{t-k}]] = E[E[h_t|I_{t-k}]h_{t-k}'] = 0$'dır, çünkü $E[h_t|I_{t-k}] = E[E[h_t|I_{t-1}]|I_{t-k}] = 0$'dır. McLeish bu teoremin güzel bir genelleştirmesini yapmıştır.

**Teorem 4.6.** *Varsayalım $\{(y_t, x_t)\}$ serisi durağan ve kuvvetli karıştırma olsun. Ayrıca, varsayalım (a) $e_t = y_t - x_t'\beta$, $I_{t-1} = \sigma((e_{t-1}, x_t)', (e_{t-2}, x_{t-1})', ...)$ süzmesine bağlı bir martingale fark dizisi olsun, yani $E[e_t|I_{t-1}] = 0$, (b) $\Omega = E[e_t^2 x_t x_t']$ sonlu olsun, ve (c) $Q = E[x_t x_t']$ sonlu ve tam kerte olsun. Bu durumda L1-L4 yukarıda tanımlanan $\Omega$ ve $Q$ ile geçerlidir. Bu sebeple, Öneriler 1-3'ün sonuçları da geçerlidir.*

İspat. Şunu biliyoruz: $E[e_t|x_t] = E[E[e_t|I_{t-1}]|x_t] = 0$, bu da L1 ve L2'yi ima eder. Şunu Ergodik BSY ile biliyoruz: $E_n[x_t x_t'] \to_p Q = E[x_t x_t']$, bu da L3'ü sağlar. $\Omega = E[e_t^2 x_t x_t']$ ile Martingale MLT'den $\sqrt{n}E_n[e_t x_t] \to_d N(0, \Omega)$ olduğunu biliyoruz, bu da L4'ü sağlar. □

**Açıklama 4.1.** *IID durumunda olduğu gibi $\Omega$ ve $Q$ için aynı tahminciyi kullanabiliriz*

$$\widehat{\Omega} = E_n[\hat{e}_t^2 x_t x_t'], \quad \hat{Q} = E_n[x_t x_t'].$$

*$\widehat{\Omega}$'ın tutarlılığı daha önce olduğu gibi $E[\|x_t\|^4] < \infty$ varsayımı ile sağlanır. İspat da daha öncesi ile aynıdır, sadece iid veri için her zamanki BSY yerine Ergodik BSY kullanılmalıdır. $\hat{Q}$'ın tutarlılığı da ergodik BSY'den elde edilir.*

Örnek. Hall'ın örneğinde,

$$E[y_t - x_t'\beta_0|x_t] = 0,$$

bu $\beta_0 = 0$ ve bilgi setine $t-1$ zamanında dahil olan ve herhangi bir değişkeni temsil eden $x_t$ için geçerlidir. Hall'ın hipotezi ile,

$$\sqrt{n}(\hat{\beta} - \beta_0) \rightarrow_d N(0, V), \quad V = Q^{-1}\Omega Q^{-1}, \quad \beta_0 = 0$$

burada $\Omega = E[y_t^2 x_t x_t']$ ve $Q = E x_t x_t'$'dir. Sonrasında, Hall'ın hipotezi Wald istatistiği kullanılarak sınanabilir:

$$W = \sqrt{n}(\hat{\beta} - \beta_0)'\widehat{V}^{-1}\sqrt{n}(\hat{\beta} - \beta_0),$$

burada, yukarıda tanımlanan $\widehat{\Omega}$ ve $\widehat{Q}$ için $\widehat{V} = \widehat{Q}^{-1}\widehat{\Omega}\widehat{Q}^{-1}$'dir.