جزوه اقتصاد سنجی
دکتر شاپور محمدی

mulooda

اقتصاد سنجی به مفهوم سنجش اقتصادی است که از دو واژه econo و metrics تشکیل شده این واژه همتایانی مانند Biometrics (زیست سنجی)، Technometrics (فن سنجی)، sociometrics (جامعه سنجی)، Politimetrics (سیاست سنجی) و Phsycometrics (روان سنجی) دارد

به لحاظ تاریخی Biometrics پیش از Econometrics مطرح شده اما اقتصاد سنجی از سایر metrics قدیمی‌تر است این واژه را اولین بار اقتصاددانی به نام راگنار فریش مطرح نمود.

اولین مجله در اقتصاد سنجی مجله Econometrica است که در سال‌های ۱۹۳۰ همزمان با ابداع این واژه منتشر گردید در حال حاضر مجلات متعددی مانند:

1. Econometrica
2. Journal of Econometrics
3. Econometric Theory
4. Financial Econometrics
5. Econometric journal
6. Journal of Applied Econometrics

اقتصاد سنجی به مجموعه‌ای از روش‌های آمار ریاضی و تئوری‌های اقتصادی گفته می‌شود که برای بررسی پدیده‌های اقتصادی مورد استفاده قرار می‌گیرد

1. تئوری‌های اقتصادی
2. مدلسازی ریاضی
3. مدلسازی [illegible]
4. جمع آوری داده‌ها
5. تخمین مدل
6. تفسیر نتایج و بررسی معناداری
7. آیا نتایج قابل قبول است؟ — خیر (بازگشت به ۱) / بلی
8. توسعه پیش‌بینی و کاربرد

مثال:

بازده شرکت تابعی از بازده بازار است

Y_{it}: بازده شرکت i برای پرتفولیو i ام در زمان t

$Y_{it}=\alpha+\beta r_{mt}$

$Y_{it}=\alpha+\beta r_{mt}+\varepsilon_t$

r_{mt}: بازده بازار (شاخص) در زمان t

ε_t: جزء اخلال یا Error term

داده‌های قیمت و شاخص را جمع آوری می‌کنیم و بازده را حساب می‌کنیم

$\hat{Y}_{it}=0.25+2.7 r_{mt}$

چون $\beta>1$ است سهم شرکت ریسک بالایی دارد

مدیر پرتفولیو این سهم را وارد نخواهد کرد چون مشتری ریسک‌پذیر ضعیف و عدم سلامت است

یافتمان ۱۸ ابوبکر ← آتاق ۵۴ (۴۵۳۵) داخلی

۱) اقتصاد سنجی مالی: Financial Econometrics

اقتصاد سنجی مالی شاخه‌ای از اقتصاد سنجی کاربردی است که به دلایل ذیل تمرکز خاص بر موضوعات مالی دارد:

۱) خواص داده‌های مالی از سایر داده‌ها متفاوت است و ویژگی‌هایی مانند خوشه‌ای شدن نوسانات، جهش‌ها و ... سبب می‌شود مدل‌های خاصی در اقتصاد سنجی مالی به کار گرفته شود.

۲) رویدادهایی در بازارهای مالی به وجود می‌آید که تحلیل خاصی را نیاز دارد و مدل سازی آن در حوزه اقتصاد سنجی مالی قرار می‌گیرد. به عنوان مثال مدل سازی سقوط بازار سهام، مدل سازی آثار تقویمی (calander effects) مانند اثر ژانویه، اثر آخر هفته، اثر شنبه دوم، اثر ماه رمضان، اثر مجامع شرکت‌ها و ...

۳) داده‌های پرتواتر (high Frequency Data) در بازارهای مالی ظاهر می‌شوند که تحلیل آنها روش‌های خاصی را نیاز دارد و اقتصاد سنجی مالی این روش‌ها را به دست می‌آورد.

☑ ویژگی‌های اقتصادسنجی مالی
۱- اثر تقویمی
۲- حافظه بازار و مدل‌های با حافظه طولانی
۳- پرتواتر
۴- جهش و پرش
۵- زمان

انواع داده‌ها

داده‌ها در اقتصاد سنجی به سه دسته زیر تقسیم می‌شوند:

۱- داده‌های سری زمانی: به داده‌هایی اطلاق می‌شود که رفتار پدیده معین را در زمان‌های مختلف ثبت نموده‌اند.
معمولاً سری زمانی با اندیس t نشان داده می‌شود که t دلالت بر زمان یعنی Time دارد.

	P_t
۱۳۸۳	۴۰۰۰
۱۳۸۴	۴۰۷۳
۱۳۸۵	۴۲۵۵
۱۳۸۶	۴۰۰۲
۱۳۸۷	۳۸۰۰

x_t (متغیر x) ، y_t (متغیر y) ، P_t (قیمت) ، r_t (بازده) ، $size_t$ (اندازه شرکت)

۲- داده‌های مقطعی (cross section): به داده‌هایی اطلاق می‌شود که پدیده‌های مختلفی در یک زمان معین مورد بررسی و ثبت قرار می‌دهد.
مقطع

سود تقسیمی شرکت‌های عضو بورس در سال ۱۳۸۵ یک مثال از داده‌های مقطعی است. پدیده‌های مختلف شرکت‌ها هستند و زمان که تغییری ندارد سال ۸۵ است.

i	DPS_i
ایران خودرو	۴۰۰
سایپا	۶۰۰
آذرآب	-۲۰
سیمان تهران	۱۵۰۰

۴. داده‌های تلفیقی که ترجمه Pannel Data است: در این نوع داده‌ها پدیده‌های مختلف در زمان‌های مختلف مورد بررسی قرار می‌گیرند یعنی تلفیقی از زمان و مقطع در آن‌ها دیده می‌شود.

مثال: نسبت $\frac{P}{E}$ در شرکت‌های مختلف به زمان‌های مختلف

t \ i	MSFT	GE	GM	F	IBM
2001	۲۵	۱۸	۱۹	۱۴	۲۱
2002	۲۲	۱۹	۱۸	۱۲	۲۲
2003	۲۸	۲۰	۱۷	۱۳	۲۳
2004	۳۰	۲۲	۱۶	۱۵	۲۵
2005	۳۲	۲۵	۱۶	۱۴	۲۴

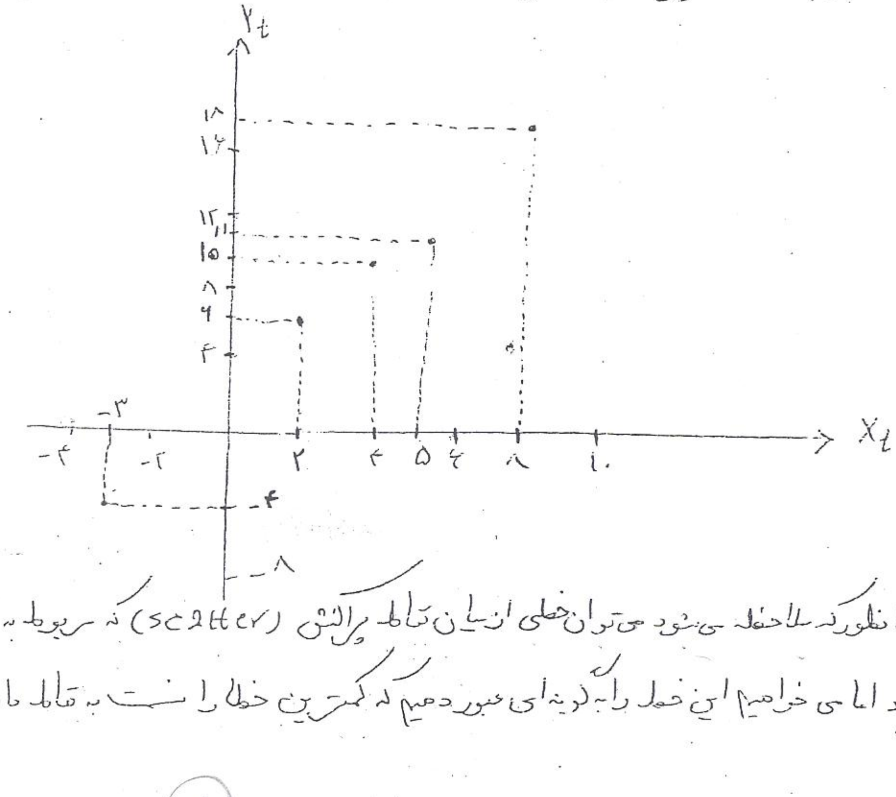
رگرسیون (Regression):

رگرسیون در لغت به معنای بازگشت و در واقع مربوط به تحقیق آقای فرانسیس گالتون است که نشان داد قد افراد تمایل به میانگین قد جامعه دارد افرادی که قد بلند هستند فرزندان کوتاه قدتر و افرادی که کوتاه‌تر هستند فرزندان قد بلندتری خواهند داشت. بنابراین تمایل قد افراد به قد متوسط افراد جامعه را بازگشت می‌نامند. برای بررسی رگرسیون از دو مفهوم ابزاری برای برازش منحنی و رگرسیون به مفهوم یک رابطه تصادفی استفاده نمود.

ذیلاً به بررسی رگرسیون با رویکرد برازش منحنی می‌پردازیم.

در رویکرد برازش منحنی سؤال پیش روی ما گذراندن خطی از میان نقاط پراکنش به نحوی که خطاها حداقل شود. فرض کنید داده‌های مربوط به بازده پرتفولیوی شرکت x و داده‌های مربوط به بازده بازار به شرح ذیل است:

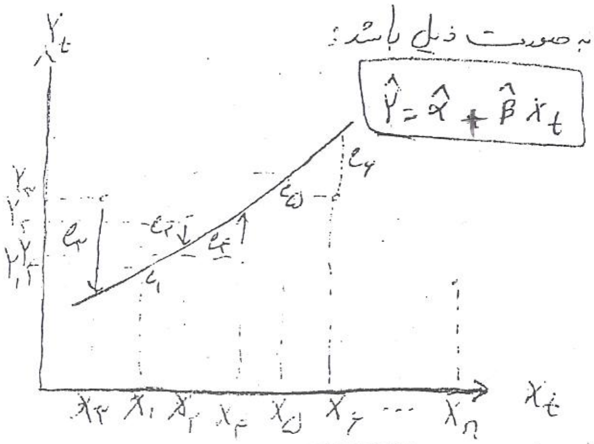
Y_t r_{it}	X_t r_{mt}
۱۰	۴
-۴	-۳
۱۸	۸
۱۱	۵
۶	۲



همانطور که ملاحظه می‌شود می‌توان خطی از میان نقاط پراکنش (scatter) که مربوط به بازده شرکت ام و بازار است عبور داد اما می‌خواهیم این خط را به گونه‌ای عبور دهیم که کمترین خطا را نسبت به نقاط داشته باشیم.

اگر نقاط پراکنش به صورت ذیل باشد:

Y_t	X_t
Y_1	X_1
Y_2	X_2
$\vdots$	$\vdots$
Y_n	X_n



$$e_t = (Y_t - \hat{Y}_t) \quad \rightarrow \quad \sum_{t=1}^{T} e_t = \sum_{t=1}^{T} (Y_t - \hat{Y}_t) \qquad \boxed{\hat{Y}_t = \hat{\alpha} + \hat{\beta} X_t}$$

$$\sum_{t=1}^{T} e_t^2 = \sum_{t=1}^{T} (Y_t - \hat{Y}_t)^2 = \sum_{t=1}^{T} (Y_t - \hat{\alpha} - \hat{\beta} X_t)^2$$

برای حداقل نمودن تابع فوق باید از تابع نسبت به پارامترها یعنی $\hat{\alpha}$ و $\hat{\beta}$ مشتق گیری نمود توجه داریم که متغیرها X_t و Y_t هستند اما مشتق گیری نسبت به $\hat{\alpha}$ و $\hat{\beta}$ انجام می‌گیرد زیرا با تغییر عرض از مبدأ و شیب می‌خواهیم مجموع مجذورات خطا را حداقل کنیم.

معادلات نرمال:

$$① \quad \frac{\partial \sum e_t^2}{\partial \hat{\alpha}} = \frac{\partial \sum (Y_t - \hat{\alpha} - \hat{\beta} X_t)^2}{\partial \hat{\alpha}} = \cancel{-2} \sum (Y_t - \hat{\alpha} - \hat{\beta} X_t) = 0$$

$$② \quad \frac{\partial \sum e_t^2}{\partial \hat{\beta}} = \frac{\partial \sum (Y_t - \hat{\alpha} - \hat{\beta} X_t)^2}{\partial \hat{\beta}} = \cancel{-2} \sum (Y_t - \hat{\alpha} - \hat{\beta} X_t)(X_t) = 0$$

$$① \quad \sum Y_t - T\hat{\alpha} - \hat{\beta} \sum X_t = 0 \quad \rightarrow \quad \hat{\alpha} = \frac{\sum Y_t - \hat{\beta} \sum X_t}{T}$$

$$② \quad \sum X_t Y_t - \hat{\alpha} \sum X_t - \hat{\beta} \sum X_t^2 = 0 \quad \xrightarrow{①} \quad \sum X_t Y_t - \left(\frac{\sum Y_t}{T} - \hat{\beta} \frac{\sum X_t}{T}\right) \sum X_t - \hat{\beta} \sum X_t^2 = 0$$

$$\sum X_t Y_t - \frac{\sum X_t \sum Y_t}{T} + \frac{\hat{\beta} (\sum X_t)^2}{T} - \hat{\beta} \sum X_t^2 = 0$$

$$\boxed{\hat{\beta} = \frac{\sum X_t Y_t - \frac{\sum X_t \sum Y_t}{T}}{\sum X_t^2 - \frac{(\sum X_t)^2}{T}}} \qquad \boxed{\hat{\alpha} = \bar{Y} - \hat{\beta} \bar{X}}$$

مثال: بازده شاخص و بازده پرتفولیوی شرکت معینی به صورت ذیل داده شده است براساس تئوری‌های مالی بازده شرکت تابعی از بازده بازار است اگر بازده شرکت را Y نامگذاری کنیم و بازده بازار را X نامگذاری کنیم ابتدا نمودار پراکنش را ترسیم نمایید و سپس خط رگرسیون این تناظر را به دست آورده و ترسیم نمایید.

بازده شرکت Y_t	بازده بازار X_t	$X_t Y_t$	X_t^2	e_t	$\hat{Y}_t$
۱	-۱	-۱	۱	۰/۲	۰/۷۹۳
۳	۰	۰	۰	۰/۲	۲/۸
۸	۳	۲۴	۹	-۰/۸	۸/۸۲۱
۱۴	۵	۷۰	۲۵	۱/۲ (۱/۱۶۵)	۱۲/۸۳۵
۱۰	۴	۴۰	۱۶	-۰/۸۲۸	۱۰/۸۲۸
۳۶	۱۱	۱۳۳	۵۱	۰	

$\bar{Y} = 7/2$ $\quad \bar{X}_t = 2/2$

$$\beta = \frac{133 - \frac{\ }{5}}{51 - \frac{(11)^2}{5}}$$

$$\hat{\beta} = \frac{133 - 79/2}{51 - 24/2} = \frac{53/8}{26/8} = 2/007$$

$$(11)^2 \div 5 = 24/2$$

$$\boxed{\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X}} = 7/2 - 2/007(2/2) = 2/8$$

$$\hat{Y}_t = \hat{\alpha} + \hat{\beta} X_t$$

درصورتی که ضرایب رگرسیون درست محاسبه شده باشد مجموع خطاها صفر خواهد بود البته اگر $\sum e_t = 0$ نمی توان لزوماً نتیجه گرفت که رگرسیون درست محاسبه شده است چون ممکن است یک اشتباه ما در جایی با اشتباه دیگر در جای خنثی شده باشد.

برای اثبات این قاعده کافی است به معادله اول نرمال توجه نمایید در معادله اول نرمال ملاحظه می شود:

$$\boxed{-2\sum(Y_t - \hat{\alpha} - \hat{\beta}X_t) = 0} \Rightarrow \sum e_t = 0$$

همان مجموع خطاهاست:

ضریب تعیین رگرسیون (Coefficient of Determination):

ضریب تعیین معیاری است که نشان می دهد چند درصد از تغییرات متغیر وابسته به وسیله متغیر مستقل توضیح (explain) داده می شود.

$$\boxed{R^2 = \frac{\sum \hat{y}_t^2}{\sum y_t^2}} \Rightarrow R^2 = \frac{\sum \hat{Y}_t^2 - \frac{(\sum \hat{Y}_t)^2}{T}}{\sum X_t^2 - \frac{(\sum X_t)^2}{T}}$$

توجه داریم که متغیرهای کوچک در اقتصادسنجی برای انحراف از میانگین داده ها استفاده می شود یعنی: $y_t = Y_t - \bar{Y}$

$$\hat{y}_t = \hat{Y}_t - \bar{Y}$$

صورت واریانس:

$$\boxed{\sum y_t^2 = \sum(Y_t - \bar{Y})^2}$$

واریانس مقادیر پیش بینی شده:

$$\boxed{\sum \hat{y}_t^2 = \sum(\hat{Y}_t - \bar{\hat{Y}})^2}$$

$$\boxed{R^2 = 1 - \frac{\sum e_t^2}{\sum y_t^2}}$$

مثال: داده های مربوط به اندازه شرکت ها به میلیارد دلار و نسبت $\frac{P}{E}$ آنها داده شده است اولاً رابطه بین اندازه و نسبت $\frac{P}{E}$ را به دست آورید ثانیاً R^2 مدل را محاسبه نمایید.

پشت صفحه

P_E Y_t	size X_t	X_tY_t	X_t^2	$\hat{Y}_t$	$\hat{Y}_t^2$	Y_t^2
۸	۳	۲۴	۹	۷,۴۳	۵۵,۲	۶۴
۱۰	۵	۵۰	۲۵	۹,۸۱	۹۶,۲۳	۱۰۰
۴	۱	۴	۱	۴,۶۵	۲۱,۶۲	۱۶
۱۲	۷	۸۴	۴۹	۱۲,۳۹	۱۵۳,۵۱	۱۴۴
۶	۲	۱۲	۴	۵,۹۴	۳۵,۲۸	۳۶
۴۰	۱۸	۱۷۴	۸۸	۴۰,۲۲	۳۶۱,۸۴	۳۶۰

$\bar{Y}_t = ۸$ $\quad \bar{X}_t = ۳,۶$

$(۱۸)^۲ = ۳۲۴ \div ۵ = ۶۴,۸$

$۴۰ \times ۱۸ = ۷۲۰ \div ۵ = ۱۴۴$

$$\hat{\beta} = \frac{۱۷۴ - \frac{(۱۸ \times ۴۰)}{۵}}{۸۸ - \frac{(۱۸)^۲}{۵}}$$

$$\boxed{\hat{B} = \frac{\sum X_tY_t - \frac{\sum X_t \sum Y_t}{T}}{\sum X_t^2 - \frac{(\sum X_t)^2}{T}}} = \frac{۱۷۴ - ۱۴۴}{۸۸ - ۶۴,۸} = \frac{۳۰}{۲۳,۲} = ۱,۲۹$$

$$\hat{Y}_t = \hat{\alpha} + \hat{\beta} X_t$$

$$\boxed{\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X}} = ۸ - ۱,۲۹(۳,۶) = ۳,۳۶ \longrightarrow \hat{Y}_t = ۳,۳۶ + ۱,۲۹ X_t$$

$$\boxed{R^2 = \frac{\sum \hat{y}_t^2}{\sum y_t^2} = \frac{\sum \hat{Y}_t^2 - \frac{(\sum \hat{Y}_t)^2}{T}}{\sum Y_t^2 - \frac{(\sum Y_t)^2}{T}}} = \frac{۳۶۱,۸۴ - \frac{۱۶۱۷,۴۴}{۵}}{۳۶۰ - \frac{۱۶۰۰}{۵}} = ۰,۹۵۸$$

$۱۶۱۷,۴۴ \longrightarrow (۴۰,۲۲)^۲ \quad ۱۶۰۰ = (۴۰)^۲$

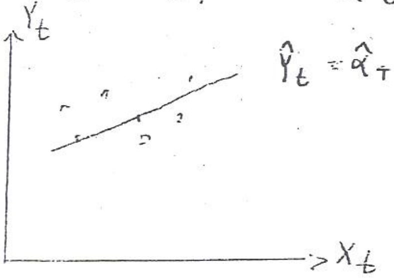
٪۹۵/۸ از تغییرات Y به وسیله X توضیح داده می‌شود.

رو یکرد تصادفی به رگرسیون و جزء خطای مدل:

$$Y_t = \alpha + \beta X_t + \varepsilon_t \longrightarrow \begin{cases} \text{Disturbance} & \text{اختلال یا اخلال} \\ \text{Error Term} & \\ \text{Stochastic Term} & \text{عبارت تصادفی} \end{cases}$$

در این حالت رگرسیون دارای دو بخش تصادفی و دترمینیستیک (معین) خواهد بود:

$$Y_t = \underbrace{\alpha + \beta X_t}_{\text{Deterministic}} + \overbrace{\varepsilon_t}^{\text{stochastic}}$$



$\hat{Y}_t = \hat{\alpha} + \hat{\beta} X_t$

دلائل وجود خطای رگرسیون:

(دلائل متعددی برای وجود خطای رگرسیون قابل ارائه است از جمله آنها می‌توان به: عوامل تصادفی. ۵- وجود یک پراکنش منفید برای متغیر اصلی

۱- خطای اندازه گیری وجود دارد. ۲- ممکن است متغیری از مدل حذف شده باشد. ۳- اگر مدل از نظر تئوری درست تصریح نشده باشد. ۴- رویدادهای غیر منتظره)

هنگامی که متغیرهای اقتصادی یا مالی اندازه گیری می‌شوند خطایی در اندازه گیری آنها وجود خواهد داشت. به عنوان مثال در محاسبه سود شرکت ممکن است خطایی وجود داشته باشد اگر چه از لحاظ حسابداری حتی به دنبال محاسبات قابل انجام است اما ممکن است برخی از اقلام از طرف حسابها حذف شده باشند و سود کمتر یا بیشتر نشان داده شود در محاسبه

GEP ممکن است بعضی از اقلام مانند کالاهایی که پر مصرف هستند لحاظ نشود، در اندازه گیری متراکم یک واحد تجاری ممکن است خطاهایی به دلیل دقت محدود ابزارهای اندازه گیری وجود داشته باشد.

۲- ممکن است متغیری از مدل حذف شده باشد در اینصورت اثر متغیر حذف شده در جزء خطا قرار خواهد گرفت به عنوان مثال اگر مدل واقعی

$$\left(\frac{P}{E}\right)_t = \alpha + \beta\, size_t + \gamma\, Inv_t + \varepsilon_t$$

باشد.

در رابطه فوق اگر یکی از متغیرهای توضیحی حذف شود اثر آن در جزء خطا ظاهر خواهد شد بنابراین متغیرهای حذف شده خودشان می توانند دلیل وجود اخلال باشند.

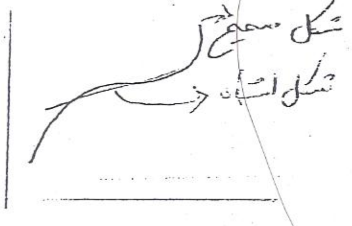
$\frac{P}{E}$ بستگی به شرایط صنعت، اقتصاد از جمله تورم و سیاستگذاری سرمایه گذاری دارد $\frac{P}{E}$ در هنگام تصاحب یک شرکت با تغییر $\frac{P}{E}$ در هنگام تعدیل EPS توسط آن شرکت متفاوت است.

۳- اگر مدل از نظر شکل تبعی (Function of Form) درست تصریح نشده یا تحدیدی به عنوان مثال به جای معادله درجه ۲، یک معادله خطی تخمین زده باشیم خطا به وجود خواهد آمد.

$$C_t = \alpha + \beta_1 Q_t + \beta_2 Q_t^2$$

$$C_t = \alpha + \beta_1 Q_t + \underbrace{\beta_2 Q_t^2 + \beta_3 Q_t^3}$$

مدل صحیح



$$C_t = \alpha + \beta_1 Q_t + \varepsilon_t$$

۴- رویدادهای غیر منتظره یا عوامل تصادفی:

مثلاً اگر سیل سبب شده باشد بازدهی بازارهای مالی کاهش پیدا کند این رویداد یک رویداد غیر منتظره بوده که به وسیله ε (اپسیلون) قابل بیان است زلزله بم، افزایش قیمت نفت، اعتصاب کارگران معدن طلا در آفریقای جنوبی، خشکسالی در برزیل، قیمت قهوه و... همه مثال‌هایی از شوک‌ها یا وجوه‌ها هستند.

ε_t : شوک، تکانه، noise

۵- پراکسی (Proxy) صفت یا متغیرهای نیابتی (جانشین) ضعیف

بعضی مواقع ممکن است متغیری که داده‌های آن وجود دارد یک Proxy ضعیف برای متغیر اصلی باشد به عنوان مثال یک مأمور مالیاتی می خواهد بر اساس سود افراد مالیات آنها را تعیین نماید با توجه به اینکه کسبه جزء دفاتر رسمی ندارند مأمور مالیاتی سود آنها را نمی تواند اندازه گیری نماید و ممکن است به جای سود از خود درآمد به عنوان یک Proxy استفاده نماید یا حتی ممکن است اجاره بهای مکان او را به عنوان یک Proxy در نظر گیرد اما می دانیم این کار موجب خطا در اندازه گیری

زیرا که ... خواهد شد و بنابراین ε را به وجود خواهد آورد.

خواص ε_t و فروض کلاسیک رگرسیون (Regression classical Assumptions)

در رگرسیون به دلیل وجود ε_t دارای خواص متغیرهای تصادفی است و متغیر Y_t که متغیر وابسته است مانند ε_t دارای توزیع آماری معین خواهد بود و برای شناسایی خواص Y_t بهتر است ابتدا خواص ε_t را بیان کنند.

فروض کلاسیک در بعضی از متون به ۵ فرض و در برخی دیگر تا ۷ فرض و در عده‌ای از متون اقتصادسنجی تا ۱۰ فرض بیان شده است ذیلاً به بررسی فروض می‌پردازیم:

۱- امید ریاضی اجزای اخلال صفر است یعنی:

$\boxed{E(\varepsilon_t)=0}$, $t=1,2,\dots,T$

از آمار خاطر داریم که امید ریاضی برابر با جمع احتمال ضرب شده در مقادیر است:

$$E(\varepsilon_t)=\sum_{i=1}^{n}(P_i\varepsilon_i)$$

معنی این فرض این است که خطاها دائماً مثبت و یا دائماً منفی نیستند.

ε_i	$P(\varepsilon_i)$
-1	$\frac{1}{3}$
0	$\frac{1}{3}$
1	$\frac{1}{3}$

ε_i	$P(\varepsilon_i)$
-2	$\frac{1}{3}$
1	$\frac{2}{3}$

$$-2\left(\frac{1}{3}\right)+1\left(\frac{2}{3}\right)=0$$

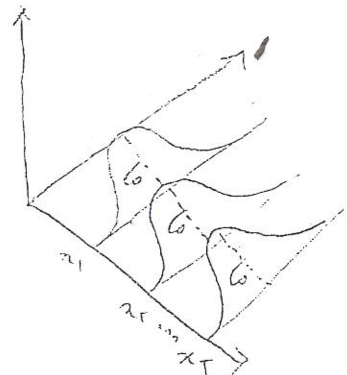
$$-1\left(\frac{1}{3}\right)+0\times\left(\frac{1}{3}\right)+1\times\left(\frac{1}{3}\right)=0$$

۲- الف) واریانس اجزای اخلال ثابت است یعنی:

$\boxed{E(\varepsilon_t^2)=\sigma^2}$

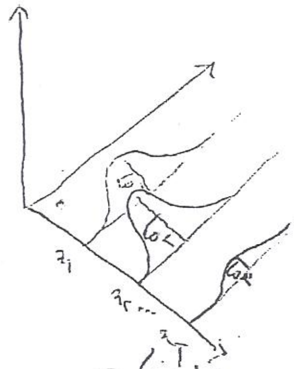
$$\sum P_i[X-E(X)]^2=E[X-E(X)]^2=E(X^2)-(E(X))^2$$

اگر امید ریاضی یک متغیر صفر باشد واریانس آن، یک عدد است.



واریانس‌ها ثابت

$\sigma_1^2=\sigma_2^2=\sigma_3^2=\sigma^2$



واریانس‌ها متغیر

$\sigma_1^2\neq\sigma_2^2\neq\sigma_3^2$

به این خاصیت، خاصیت ثابت بودن واریانس اجزای اخلال یا همسانی اجزای واریانس (Hemosecedasticity) Hemoskedasticity که در تضاد با ناهمسانی واریانس (Heteroscedasticity) است.

مفهوم این فرض آن است که دامنه تغییرپذیری اجزای اخلال طی مشاهدات مختلف به عنوان مثال سال‌های مختلف یا اگر داده‌ها مقطعی هستند به عنوان مثال بین خانوارهای مختلف تغییر نمی‌کند و واریانس اجزای اخلال برای تمامی مشاهدات یکسان و برابر σ^2 است اگر واریانس برای هر مشاهده با مشاهده دیگر متفاوت باشد دیگر نمی‌توان از نماد کلی σ^2 برای تمامی واریانس‌ها استفاده نمود و باید واریانس اجزای مختلف اخلال با σ_t^2 نمایش داد.

۲- اجزای اخلال همبستگی پیاپی ندارند یعنی اجزای اخلال به صورت متوالی مستقل از یکدیگر هستند بنابراین:

$$\boxed{E(\varepsilon_t \varepsilon_s)=0}$$

$$cov(\varepsilon_t \varepsilon_s) = E[\varepsilon_t - E(\varepsilon_t)][\varepsilon_s - E(\varepsilon_s)] = E(\varepsilon_t \varepsilon_s)$$

براساس فرض اول کلاسیک

به طور اشتباه کوواریانس صفر به مفهوم استقلال نیز است اما در مورد سایر توزیع‌ها نمی‌توان لزوماً این نتیجه را گرفت (البته به جز توزیع‌های نرمال، توزیع‌هایی وجود دارند که کوواریانس صفر، به مفهوم استقلال است البته این بحث کلی نیست) یعنی اجزای اخلال همبستگی پیاپی ندارند یعنی خطای مشاهده شده در سال ۱۳۸۱ با خطای مشاهده شده در داده‌های سال ۱۳۸۳ یا ۸۲ یا ۸۵ و... همبستگی ندارد یعنی دلیلی ندارد زلزله بم با بارانی سال ۸۵ نسبت معنی داشته باشد اجزای اخلال مستقل از هم هستند.

۳- تمامی مقادیر مستقل x_tها با یکدیگر برابر نیستند یعنی $\boxed{\Sigma x_t^2 \neq 0}$

به عبارت دیگر:

$$\Sigma x_t^2 = \sum_{t=1}^{T} (X_t - \bar{X})^2 \neq 0$$

درصورتی که مباحث قبلی را فراموش نکرده باشیم:

$$\Sigma (X_t - \bar{X})^2 = \Sigma X_t^2 - \frac{(\Sigma X_t)^2}{T}$$

همان مخرج کسر فرمول $\hat{\beta}$ است بنابراین اگر Σx_t^2 مساوی صفر باشد مخرج کسر $\hat{\beta}$ صفر خواهد شد که در این صورت $\hat{\beta}$ غیر قابل محاسبه خواهد بود.

در رگرسیون چند متغیره این فرض به شکل کاملتری تحت عنوان نبود هم خطی کامل (Perfect Multicolinearity) بیان خواهد شد همچنین این بحث معادل رتبه کامل ستونی X در بیان ماتریسی رگرسیونی است که در مباحث آتی بیشتر توضیح داده خواهد شد.

۴- Xها غیرتصادفی هستند و در صورتی که تصادفی باشند با εها همبستگی ندارند به عبارت دیگر $\boxed{E(X_t \varepsilon_t)=0}$ یعنی اجزای اخلال و متغیر توضیحی ناهمبسته هستند. معنای این فرض این است که X_t تحت کامل محقق قرار دارد یعنی محقق مقدار X را به طور دقیق می‌داند یا می‌تواند کنترل نماید یا X_t در آزمایش‌های تکراری ثابت است به بیان دیگر X_t دارای خطای اندازه‌گیری یا جزء تصادفی نیست و نمی‌توان X_t را به Z_t که یک متغیر غیرتصادفی و U_t یک متغیر تصادفی است تجزیه نمود یعنی X_t به جز اول است یعنی هیچ جز تصادفی ندارد به عنوان مثال در علم شیمی می‌خواهیم میزان گرمای ایجاد شده (Y) را در اثر اضافه کردن متغیر X به میزان مشخص در محلول بررسی نماییم اما گرمای محیط (ε) نیز بر گرمای Y که ماده مورد نظر است مؤثر است، مفهوم غیرتصادفی بودن X آن است که مقدار دقیق X به وسیله محقق قابل تعیین و قابل اعمال است یعنی در وزن کردن X خطای اندازه‌گیری وجود ندارد و در اضافه کردن X به مایع، X نه تبخیر می‌شود و نه به جز محلول به جای دیگری ریخته می‌شود و Xها همبستگی با ε ندارد محقق تصمیم می‌گیرد که ۲گرم، ۳گرم، ... به محلول اضافه نماید ε دمای محیط است و تعیین در وضعیت X ندارد.

اما در علوم انسانی و اجتماعی متغیر GEP دست محقق نیست و مصرف کردن در سطح کلان تابعی از GEP است نمی‌توان گفت تحت تأثیر یک متغیرها مانند GEP که تحت کنترل کامل است قرار دارد یعنی محقق در تعیین Xها نقشی ندارد و کنترلی بر آنها ندارد در مالی بازده بازار در دست محقق نیست که آن را به شکل دلخواه تغییر دهد تا رفتار بازده پرتفوی را بررسی نماید حال این سؤال پیش می‌آید پس چرا چنین فرضی را لحاظ می‌کنیم؟ چون داده‌ها به صورت تاریخی ثبت شده‌اند و امکان تغییر آنها وجود ندارد بنابراین در علوم اجتماعی غیرتصادفی بودن Xها و ثابت بودن آنها در آزمایش‌های تکراری به دلیل دقت اندازه‌گیری و کنترل محقق نیست بلکه به دلیل تاریخی بودن و عدم امکان تکرار تاریخ است یعنی ما نمی‌توانیم به سال ۱۳۴۸ برگردیم و GEP کشور را یک بار دیگر اندازه‌گیری کنیم عدد GEP داده شده ثابت است.

۵- توضیح اجزای اخلال نرمال است یعنی $\boxed{\varepsilon_t \sim N(0,\sigma^2)}$

۶- رابطه بین X و Y خطی است یعنی مدل بر حسب پارامترهای خطی است $Y_t=\alpha+\beta X_t+\varepsilon_t$ خطی است زیرا توان α و β هر دو یک است اما رابطه ذیل

$$Y_t=\alpha^2+\beta X_t+\varepsilon_t$$

غیرخطی است زیرا توان α به جز یک است.

در صورتی که رابطه $Y_t=\alpha+\beta X_t^{100}+\varepsilon_t$ از نظر رگرسیون خطی است چون ما می‌توانیم به جای X_t^{100}، Z_t را تعریف کرده و محاسبات را ادامه دهیم زیرا این تبدیل در حل معادلات نرمال هیچ تأثیری ندارد و به جای آینده بنویسیم:

$$\hat{\beta} = \frac{\sum X_t^{1\cdots} Y_t - \frac{\sum X_t^{1\cdots} \sum Y_t}{T}}{\sum X_t^{1\cdots} - \frac{(\sum X_t^{1\cdots})^2}{T}}$$

می‌نویسیم:

$$\hat{\beta} = \frac{\sum Z_t Y_t - \frac{\sum Z_t \sum Y_t}{T}}{\sum Z_t^2 - \frac{(\sum Z_t)^2}{T}}$$

به عنوان مثال دیگر:

$$\ln Y_t = \alpha + \beta \ln X_t + \varepsilon_t \rightarrow Y_t^* = \alpha + \beta X_t^* + \varepsilon_t$$

رابطه فوق خطی است.

استنباط آماری در رگرسیون دو متغیره:

استنباط (استنتاج): فرآیندی است ذهنی که طی آن از دانش موجود به دانش جدید با رعایت قواعد منطقی دست پیدا می‌کنیم.

استنتاج آماری: فرآیندی است که طی آن با استفاده از آماره‌های به دست آمده در نمونه نسبت به پارامترهای جامعه قضاوت می‌کنیم.

به عنوان مثال میانگین قد دانشجویان این کلاس اگر این کلاس را یک نمونه تصادفی فرض نماییم یک آماره‌ای است که می‌تواند در مورد میانگین قد دانشجویان دانشکده مدیریت که جامعه آماری است قضاوت کنیم اگر این قضاوت کردن با رعایت قواعد منطقی آماری باشد یک استنتاج آماری است در اینجا $\hat{\alpha}$ و $\hat{\beta}$ آماره‌های نمونه‌ای هستند که از نمونه به دست آمده‌اند و می‌خواهیم با استفاده از آنها در مورد پارامترهای جامعه یعنی α و β با رعایت قواعد منطقی نتیجه‌گیری نماییم و توجه داریم که $\hat{\alpha}$ با α و $\hat{\beta}$ با β متفاوت است.

جامعه آماری: گروه‌ای از عناصر که دارای حداقل یک صفت مشترک باشند جامعه آماری نامیده می‌شود.

نمونه آماری: هر زیر جامعه‌ای از یک جامعه آماری، نمونه آماری نامیده می‌شود.

انتخاب تصادفی: انتخابی است که شانس وقوع رویدادها از پیش معین باشد.

نمونه تصادفی: نمونه‌ای است که عناصر آن مستقل از یکدیگر و با رعایت شرط تصادفی بودن انتخاب شود.

آماره تابع نمونه‌ای: هر تابعی از متغیرهای نمونه، یک تابع نمونه‌ای یا آماره نامیده می‌شود.

$X_1, X_2, \dots, X_n$: عناصر نمونه　　　آماره : $T = f(X_1, X_2, \dots, X_n)$

$\bar{X} = \frac{\sum X_i}{n} = f(X_1, X_2, \dots, X_n)$　　　$S^2 = \frac{\sum (X_i-\bar{X})^2}{n-1} = g(X_1, X_2, \dots, X_n)$

تخمین زن: هر آماره‌ای که برای تخمین پارامترهای جامعه یا پارامتر جامعه استفاده شود یک تخمین زننده یا تخمین زن نامیده می‌شود (تخمین زن : Estimator).

$\bar{X} \to \mu$,　$\boxed{S^2 \to \sigma^2}$,　$\boxed{\hat{S}^2 \to \sigma^2}$

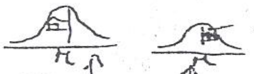
$S^2 = \frac{\sum (X_i-\bar{X})^2}{n-1}$　　　$\hat{S}^2 = \frac{\sum (X_i-\bar{X})^2}{n}$

ویژگی‌های تخمین زننده خوب:

۱- نااریب باشد　۲- کارا باشد　۳- سازگار باشد　۴- کافی باشد

۱) تعریف: تخمین زننده $T(X_1, X_2, \dots, X_n)$ برای تخمین پارامتر θ در نظر گرفته شده است این تخمین زننده در صورتی نااریب (بدون تورش، ناتورش : unbiased) نامیده می‌شود که $\boxed{E(T)=\theta}$ یا $\boxed{E(T)-\theta=0}$ باشد.

$E(T) = E(\bar{X}) = \mu$



اثبات:

$\bar{X} = \frac{\sum X_i}{n} \Rightarrow E(\bar{X}) = E\left(\frac{X_1 + X_2 + \dots + X_n}{n}\right) = \frac{1}{n}\left[E(X_1) + E(X_2) + \dots + E(X_n)\right]$

$E(\bar{X}) = E\left(\frac{\sum X_i}{n}\right) = \frac{1}{n}(\mu + \mu + \dots + \mu) = \frac{n\mu}{n} = \mu$

$E(\hat{S}^2) \neq \sigma^2$

$E(S^2) = \sigma^2$

اثبات: $E(S^2) = E\left(\frac{\sum (X_i-\bar{X})^2}{n-1}\right) = E\left[\frac{\sigma^2}{n-1} \underbrace{\sum \left(\frac{X_i-\bar{X}}{\sigma}\right)^2}_{\chi^2_{n-1}}\right] \to \sum_{i=1}^{n} \left(\frac{X_i-\bar{X}}{\sigma}\right)^2 = \sum_{i=1}^{n} (Z_i^2) = \chi^2_{n-1}$

$\Rightarrow E(S^2) = \frac{\sigma^2}{n-1} E(\chi^2_{n-1}) = \frac{\sigma^2}{n-1}(n-1) = \sigma^2 \xrightarrow{\text{پس}} E(S^2) = \sigma^2$

$E(\hat{S}^2) = E\left(\frac{\sum (X_i-\bar{X})^2}{n}\right) = \frac{\sigma^2}{n} E\left\{\sum \left(\frac{X_i-\bar{X}}{\sigma}\right)^2\right\} = \frac{\sigma^2}{n} E(\chi^2_{n-1}) = \boxed{\frac{\sigma^2}{n}(n-1)} \xrightarrow{\text{پس}} \hat{S}^2 \neq \sigma^2$

$E(\chi^2) = \chi^2$ درجه آزادی　,　$var(\chi^2) = \chi^2$ دو برابر درجه آزادی $\to 2(n-1)$

$Var(\chi^2_{n-1}) = 2(n-1)$

۲) تعریف: اگر T_1 و T_2، دو تخمین زننده برای θ باشد در صورتی T_1 کاراتر از T_2 نامیده می‌شود که واریانس T_1 کوچکتر از واریانس T_2 باشد.

$\boxed{var(T_1) < var(T_2)}$ $\to$ آنگاه تخمین زننده T_1 کاراتر نسبت به T_2 است

مثال: می دانیم تخمین زننده $\bar{x}$ و میانه هر دو تخمین زننده های بدون تورش برای μ هستند از طرفی می توان نشان داد واریانس $\bar{x}$ و Me به صورت ذیل است:

$$*\ var(\bar{x}) = \frac{\sigma^2}{n}$$

$$var(Me) = \frac{\sigma^2}{n} \cdot \frac{\pi}{2} \longrightarrow$$ $\bar{x}$ کاراتر از Me است

$$*\text{اثبات}: var(\bar{x}) = var\left(\frac{x_1 + x_2 + \dots + x_n}{n}\right) = \frac{1}{n^2}\left[var(x_1) + var(x_2) + \dots + var(x_n)\right] =$$

$$= \frac{1}{n^2}\left[\sigma^2 + \sigma^2 + \dots + \sigma^2\right] = \frac{n\sigma^2}{n^2} = \boxed{\frac{\sigma^2}{n}}$$

مثال: میزان کارایی $\hat{s}^2$ و s^2 را نسبت به یکدیگر برای تخمین σ^2 بسنجید؟

$$var(s^2) = var\left[\frac{\sum(x_i - \bar{x})^2}{n-1}\right] = var\left[\frac{\sigma^2}{n-1}\sum\left(\frac{x_i - \bar{x}}{\sigma}\right)^2\right] = \frac{\sigma^4}{(n-1)^2} var(\chi^2_{n-1}) = \boxed{\frac{\sigma^4}{(n-1)^2} 2(n-1)}$$

$$var(\hat{s}^2) = var\left[\frac{\sum(x_i - \bar{x})^2}{n}\right] = var\left[\frac{\sigma^2}{n}\sum\left(\frac{x_i - \bar{x}}{\sigma}\right)^2\right] = \frac{\sigma^4}{n^2} var(\chi^2_{n-1}) = \boxed{\frac{\sigma^4}{n^2} 2(n-1)}$$

($\hat{s}^2$ کاراتر از s^2 است چون واریانس $\hat{s}^2$ کوچکتر از واریانس s^2 می باشد. یعنی $Var(\hat{s})^2 \leq Var(s)^2$)

(۳) تعریف: اگر T یک تخمین زننده برای θ باشد در صورتی سازگار نامیده می شود که در احتمال به θ گرایش نماید:

$$\boxed{P\lim T = \theta}$$

$$\lim_{n \to \infty} P(|T - \theta| < \varepsilon) = 1 \rightarrow P\lim T = \theta$$

یعنی در بی نهایت احتمال اینکه $T = \theta$ گردد ۱ است.

با نامساوی چبی شف می توان اثبات کرد عبارت فوق قابل تبدیل به فرمول ذیل است:

$$\lim_{n \to \infty} var(T) = 0$$

مثال: $\bar{X}$ یک تخمین زننده سازگار برای μ است زیرا: $\longleftarrow$

$$\lim_{n \to \infty} Var(\bar{x}) = \lim_{n \to \infty} \frac{\sigma^2}{n} = 0$$

(۴) تعریف: بسندگی یا کافی بودن در مفهوم غیر دقیق آن به معنی استفاده کردن از تمامی داده ها در ساختن یک تخمین زن است به عنوان مثال $\bar{X}$ از تمامی مشاهدات استفاده می کند در حالی که Me و Mo از بعضی از مشاهدات استفاده می کنند تمامی مشاهدات را استفاده می کنند بنابراین $\bar{X}$ یک تخمین زننده کافی یا بسنده است در حالی که میانه و نما این گونه نیستند اما در مفهوم دقیق بسندگی یا کافی بودن به مفهوم تجزیه پذیری تابع چگالی به دو قسمت پارامتری و غیرپارامتری است یعنی یک تابع فقط تابعی از داده ها خواهد بود و تابع دیگر فقط تابعی از پارامترها خواهد بود:

اگر $\hat{\theta}$ یک تخمین زن برای پارامتر θ باشد در صورتی کافی نامیده می شود که تابعی از تمامی عناصر نمونه $x_1, x_2, \dots$ تا x_n باشد یعنی

$$\hat{\theta} = f(x_1, x_2, \dots x_n)$$

خواص آماری تخمین زننده‌های OLS: ا- خاصیت بدون تورش بودن

می‌خواهیم بررسی نماییم که آیا $\hat{\beta}$ یک تخمین زننده بدون تورش برای پارامتر β است. بنابراین می‌توان امید ریاضی $\hat{\beta}$ را محاسبه نمود (به دست آورد) و با β مقایسه نمود اگر $E(\hat{\beta})=\beta$ باشد در اینصورت $\hat{\beta}$ یک تخمین زننده بدون تورش است. به خاطر داریم که:

$$\hat{\beta}=\frac{\sum X_tY_t-\frac{\sum X_t\sum Y_t}{T}}{\sum X_t^2-\frac{(\sum X_t)^2}{T}}=\frac{\sum(X_t-\bar{X})(Y_t-\bar{Y})}{\sum(X_t-\bar{X})^2}=\frac{\sum x_t y_t}{\sum x_t^2}\ ,\quad y_t=Y_t-\bar{Y},\ x_t=X_t-\bar{X}$$

از طرفی اگر مدل $Y_t=\alpha+\beta X_t+\varepsilon_t$ به صورت انحراف از میانگین نوشته شود داریم: $\boxed{y_t=\beta x_t+\varepsilon_t}$

اثبات

$$\frac{\sum Y_t}{T}=\frac{\sum\alpha}{T}+\frac{\beta\sum X_t}{T}+\frac{\sum\varepsilon_t}{T}\rightarrow\bar{Y}=\alpha+\beta\bar{X}+\bar{\varepsilon}$$

$$Y_t-\bar{Y}=\cancel{\alpha}-\cancel{\alpha}+\beta(X_t-\bar{X})+\underbrace{\varepsilon_t-\bar{\varepsilon}}_{\varepsilon_t}\rightarrow\boxed{y_t=\beta x_t+\varepsilon_t}$$

برای اثبات بدون تورش بودن تخمین زننده $\hat{\beta}$ لازم است نشان دهیم:

$$E(\hat{\beta})=\beta$$

$$\hat{\beta}=\frac{\sum x_t y_t}{\sum x_t^2}=\frac{\sum x_t(\beta x_t+\varepsilon_t)}{\sum x_t^2}=\beta\frac{\cancel{\sum x_t^2}}{\cancel{\sum x_t^2}}+\frac{\sum x_t\varepsilon_t}{\sum x_t^2}=\beta+\frac{\sum x_t\varepsilon_t}{\sum x_t^2}$$

$$E(\hat{\beta})=\underbrace{E(\beta)}_{\beta}+E\left(\frac{\sum x_t\varepsilon_t}{\sum x_t^2}\right)=\beta+\boxed{\frac{\sum x_tE(\varepsilon_t)}{\sum x_t^2}}\rightarrow$$

امید ریاضی عدد ثابت برابر با خود آن است

کل این عبارت صفر خواهد شد.

با توجه به اینکه بر اساس فرض کلاسیک Xها غیر تصادفی هستند امید ریاضی از آنها عبور کرده و فقط بر εها اعمال می‌شود از طرف دیگر بر اساس فرض اول کلاسیک $\boxed{E(\varepsilon_t)=0}$ است بنابراین کل عبارت دوم صفر شده و $E(\hat{\beta})=\beta$ است.

تخمین زن‌های OLS $\hat{\alpha}$ و $\hat{\beta}$ بدون تورش هستند.

پس از بررسی کارایی تخمین زننده‌های OLS به محاسبه واریانس تخمین زننده‌های OLS می‌پردازیم:

$$var(\hat{\beta})=E[\hat{\beta}-E(\hat{\beta})]^2$$

بر اساس خاصیت بدون تورش بودن $\hat{\beta}$ می‌توان نوشت:

$$E(\hat{\beta})=\beta$$

$$var(\hat{\beta})=E[\hat{\beta}-\beta]^2$$

از طرفی قبلاً نشان داده شد که:

$$\hat{\beta}=\beta+\frac{\sum x_t\varepsilon_t}{\sum x_t^2}$$

با انتقال β به سمت چپ معادله خواهیم داشت:

$$\hat{\beta}-\beta=\frac{\sum x_t\varepsilon_t}{\sum x_t^2}$$

با توجه به اینکه x_t غیر تصادفی هستند می توان Σx_t^2 را به بیرون از عملگر امید ریاضی انتقال داد:

$$E(\hat{\beta}-\beta)^2 = \frac{1}{(\Sigma x_t^2)^2} E(\Sigma x_t \varepsilon_t)^2 = \frac{1}{(\Sigma x_t^2)^2} E(x_1\varepsilon_1 + x_2\varepsilon_2 + \dots + x_T\varepsilon_T)^2$$

$$= \frac{1}{(\Sigma x_t^2)^2} E[x_1^2\varepsilon_1^2 + x_2^2\varepsilon_2^2 + \dots + x_T^2\varepsilon_T^2 + 2x_1x_2\varepsilon_1\varepsilon_2 + 2x_1x_3\varepsilon_1\varepsilon_3 + \dots + 2x_{T-1}x_T\varepsilon_{T-1}\varepsilon_T]$$

$$= \frac{1}{(\Sigma x_t^2)^2}[x_1^2 \overbrace{E(\varepsilon_1^2)}^{\sigma^2} + x_2^2 \overbrace{E(\varepsilon_2^2)}^{\sigma^2} + \dots + x_T^2 \overbrace{E(\varepsilon_T^2)}^{\sigma^2} + \underline{2x_1x_2 \overbrace{E(\varepsilon_1\varepsilon_2)}^{0}} + \underline{2x_1x_3 \overbrace{E(\varepsilon_1\varepsilon_3)}^{0}} + \dots + \underline{2x_{T-1}x_T \overbrace{E(\varepsilon_{T-1}\varepsilon_T)}^{0}}]$$

$$E(\varepsilon_t^2)=\sigma^2 \qquad E(\varepsilon_t\varepsilon_s)=0 \xrightarrow{\text{اگر}} t\neq s$$

با توجه به اینکه x_tها غیر تصادفی هستند عملگر امید ریاضی از آنها عبور کرده و فقط بر εها اعمال می شود همچنین با توجه به فروض قبلی امید ریاضی عبارت‌های ضربی، یعنی عبارت‌هایی که به صورت $\varepsilon_t\varepsilon_s$ هستند و $t\neq s$ صفر است در نتیجه داریم:

$$= \frac{1}{(\Sigma x_t^2)^2}[x_1^2\sigma^2 + x_2^2\sigma^2 + \dots + x_T^2\sigma^2] = \frac{1}{(\Sigma x_t^2)^2}\sigma^2(\Sigma x_t^2) = \frac{\sigma^2}{\Sigma x_t^2}$$

$$Var(\hat{\beta}) = \frac{\sigma^2}{\Sigma x_t^2} \qquad E(\hat{\beta})=\beta$$

حال که امید ریاضی و واریانس $\hat{\beta}$ معین شده است می توان به شرط داشتن نوع توزیع، توزیع را هم از نظر نوع و هم از نظر پارامترها برای $\hat{\beta}$ مشخص کرد.

برای مشخص کردن نوع توزیع کافی است توجه نماییم که براساس فرض های کلاسیک توزیع εها نرمال بوده، می توان نشان داد $\hat{\beta}$ تابع خطی از εها است:

$$\hat{\beta} = \beta + \frac{\Sigma x_t\varepsilon_t}{\Sigma x_t^2} = \boxed{\beta + \Sigma w_t\varepsilon_t} \quad , \quad \boxed{w_t = \frac{x_t}{\Sigma x_t^2}}$$

بنابراین $\hat{\beta}$ یک تابع خطی از متغیرهای تصادفی ε_t (جزء اخلال) است. با توجه به اینکه هر تابع خطی از متغیرهایی با توزیع نرمال یک توزیع نرمال خواهد بود بنابراین $\hat{\beta}$ نیز دارای توزیع نرمال است:

$$\boxed{\hat{\beta} \sim N\left(\beta, \frac{\sigma^2}{\Sigma x_t^2}\right)}$$

حال که توزیع $\hat{\beta}$ در اختیار است می توان آزمون فرضیه و تخمین فاصله ای را در خصوص β بررسی نمود.

قبل از استنباط آماری در خصوص ضرایب رگرسیون ذکر چند نکته ضروری است.

1) $E(\hat{\alpha})=\alpha$

2) $Var(\hat{\alpha}) = \sigma^2\left[\frac{1}{T} + \frac{\bar{X}^2}{\Sigma x_t^2}\right]$

3) توزیع $\hat{\alpha}$ نرمال است.

4) کوواریانس $\hat{\alpha}$ و $\hat{\beta}$ منفی است.

قضیه گاوس - مارکف:

بدون اینکه این قضیه را اثبات کنیم فقط صورت آن را ذکر می کنیم:

قضیه: تخمین زننده های حداقل مربعات در میان تخمین زننده های خطی بدون تورش، دارای کمترین واریانس هستند به عبارت دیگر تخمین زننده های حداقل مربعات blue هستند.

(Best Linear Unbiased Estimator) BLUE

اگر بخواهیم استنباط آماری در مورد ضرایب انجام دهیم اگرچه واریانس آنها معلوم است اما σ^2 را در اختیار نداریم.

به عنوان مثال در فرمول $\boxed{var(\hat\beta)=\frac{\sigma^2}{\sum x_t^2}}$ که σ^2 نامعلوم است زیرا σ^2 واریانس جزء اخلال است که خود نامشخص است (غیر قابل مشاهده است). برای تخمین σ^2 می توان از تخمین زننده بدون تورش آن یعنی $\hat\sigma^2$ استفاده نمود:

$$\boxed{\hat\sigma^2=\frac{\sum e_t^2}{T-2}} \quad , \quad e_t=Y_t-\hat Y_t \qquad \hat Y_t=\hat\alpha+\hat\beta X_t$$

(درجه آزادی چون دو پارامتر $\hat\alpha$ و $\hat\beta$ تخمین زده شد)

همچنین می توان اثبات کرد تخمین زننده های حداقل مربعات سازگار هستند و شرط کافی بودن را دارند.

ممکن است این سؤال پیش آید تخمین زننده های حداقل مربعات به عنوان مثال $\hat\beta$ نسبت به کدام تخمین زننده بدون تورش و حداقل واریانس ها را دارند؟

در پاسخ می توان گفت برای $\hat\beta$ می توان تخمین زننده های دیگری ساخت: $\beta^*=\frac{\sum X_i}{Y+5}$ $\quad \bar\beta=\frac{\bar X}{\bar Y}$

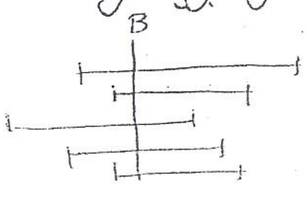
یا هر تخمین زننده دیگری تعریف نمود اما این تخمین زننده ها یا بدون تورش نیستند یا واریانس بیشتری دارند.

استنباط آماری در رگرسیون:

فرض کنید تخمین $\hat\beta$ و واریانس آن در اختیار است از محقق خواسته شده است یک فاصله اعتماد ۹۵٪ برای آن تشکیل دهد با توجه به اینکه می دانیم $\hat\beta\sim N(\beta,\frac{\sigma^2}{\sum x_t^2})$ می توان فاصله اعتماد را برای β به صورت ذیل به دست آورد:

$$P(\hat\beta - Z_{1-\frac{\alpha}{2}}\, SE(\hat\beta) < \beta < \hat\beta + Z_{1-\frac{\alpha}{2}}\, SE(\hat\beta)) = 95\%$$

البته بهتر است به جای اینکه گفته شود β در این فاصله قرار گیرد گفته شود که ۹۵٪ از فاصله هایی که به این صورت ساخته می شود در بر گیرنده β خواهد بود زیرا β یک پارامتر ثابت است و تغییر نمی کند تا بتواند در بازه های ... قرار گیرد یا قرار نگیرد در حقیقت بازه ها از نمونه های مختلف قابل محاسبه اند و مرکز بازه ها تغییر خواهد کرد و بنابراین اگر β مقدار معینی باشد فاصله های اعتماد مختلف که طول آنها یکسان است اما مرکز آن ثابت است



در بر گیرنده β خواهد بود.

آزمون فرضیه β (Hypothesis Test):

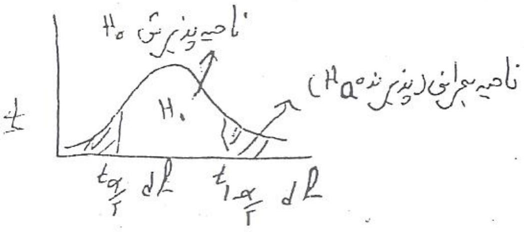
فرضیه: جمله خبری که قابل رد یا پذیرش است، می باشد.

فرضیه آماری: هر فرضیه ای راجع به چگونگی توزیع یا راجع به پارامترهای توزیع یک فرضیه آماری نامیده می شود.

1. $H_0: \theta = \theta_0$
   $H_a: \theta \neq \theta_0$

2. $t = \dfrac{\hat{\theta} - \theta_0}{SE(\hat{\theta})} \sim t_{p,\, df}$

3. α



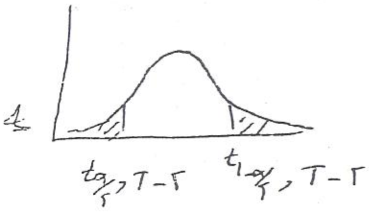
5. محاسبه آماره آزمون.

6. نتیجه گیری

1. $H_0: \beta = \beta_0$
   $H_a: \beta \neq \beta_0$

2. $t = \dfrac{\hat{\beta} - \beta_0}{SE(\hat{\beta})} \sim t_{p,\, T-2}$

3. $\alpha =$ ٪۵ یا ٪۱



5: محاسبه آماره آزمون

6. نتیجه گیری

مثال: داده های مربوط به بازده پرتفولیوی تشکیل شده در یک شرکت سرمایه گذاری و بازده بازار در جدول ارائه شده است مدیر عامل شرکت سرمایه گذاری ادعا می کند پرتفولیوی او کم ریسک است ادعا او را در سطح خطای نوع اول ٪۵ آزمون نمایید.

Y_t	X_t	X_tY_t	X_t^2	$\hat{Y}_t$	e_t	e_t^2
۱	-۱	-۱	۱	۱,۷۱	-۰,۷۱	۰,۵۰۴۱
۴	۰	۰	۰	۲,۸	۱,۲	۱,۴۴
۵	۳	۱۵	۹	۶,۷	-۱,۷	~~۱,۱۴۴۹~~ ۱,۱۴۴۹
۹	۵	۴۵	۲۵	۸,۲۵	۰,۷۵	۰,۵۶۲۵
۷	۴	۲۸	۱۶	۷,۱۶	-۰,۱۶	۰,۰۲۵۶
۲۶	۱۱	۸۷	۵۱		۰,۱ ≃ ۰	۳,۶۷۷۱

$=52 \qquad \bar{X} = 2.2$

$$\hat{\beta} = \frac{\sum X_tY_t - \dfrac{\sum X_t \sum Y_t}{T}}{\sum X_t^2 - \dfrac{(\sum X_t)^2}{T}} = \frac{29.286}{26.8} = 1.09$$

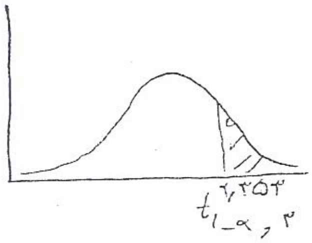
$$\frac{87 - \dfrac{(11 \times 26)}{5}}{51 - \dfrac{(11)^2}{5}} = \frac{87 - 57.2}{51 - 24.2} = \frac{29.8}{26.8} = 1.11$$

$(5.2)-(1.11)(2.2) \Rightarrow \alpha = \ldots$ $\hat{Y}_t = 2.758 + 1.11X_t$

$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X} = 2.802 + 1.02X_t$

$\hat{\sigma}^2 = \frac{\sum e_t^2}{T-2} = \frac{3.671}{3} = 1.2257$ $var(\hat{\beta}) = \frac{\hat{\sigma}^2}{\sum x_t^2} = \frac{1.2257}{26.8} = 0.045$ $SE(\hat{\beta}) = \sqrt{var(\hat{\beta})} = 0.2138$

1. $\begin{cases} H_0: \beta \leq 1 \\ H_a: \beta > 1 \end{cases}$ 2. $t = \frac{\hat{\beta}-1}{SE(\hat{\beta})} \sim t_{1-\alpha}$ 3. $\alpha = 0.05$

4. 

5. $t = \frac{1.09-1}{0.2138} = 0.42$

6.

با توجه به اینکه آماره محاسبه شده در ناحیه بحرانی قرار نگرفته در سطح خطای نوع اول ۵٪، فرضیه H_0 را مبنی بر کم کشش بودن پرتفوی نمی‌توان رد کرد.

در برخی از مراجع آزمون فرضیه برای بررسی صفر بودن یا نبودن β انجام می‌پذیرد در این صورت آزمون فرضیه، آزمون معنی داری رگرسیون نیز نامیده می‌شود زیرا اگر H_0 را نتوان رد کرد و β بتواند صفر تلقی شود رابطه بین X و Y وجود نخواهد داشت در نتیجه رگرسیون معنی دار نخواهد بود.

در صورتی که بخواهیم آزمون معنی داری انجام دهیم آماره t به سادگی با تقسیم ضریب بر SE آن محاسبه خواهد شد $t = \frac{\hat{\beta}}{SE(\hat{\beta})}$

آزمون فرضیه در مورد α نیز قابل انجام است که در اینجا نیازی به تکرار گفته‌های فوق نیست.

چرا برای محاسبه آماره از توزیع t استفاده می‌شود؟

چون σ^2 مشخص نیست و با استفاده از $\hat{\sigma}^2 = \frac{\sum e_t^2}{n-2}$، $\hat{\sigma}^2$ تخمین زده می‌شود ولی وقتی n به سمت بی‌نهایت میل کند توزیع t بر نرمال منطبق است.

$var(t) = \frac{v}{v-2}$ و $\begin{cases} \lim_{v \to \infty} \frac{v}{v-2} = 1 \\ E(t) = 0 \end{cases} \Rightarrow t \sim N(0,1)$

پیش بینی در رگرسیون دو متغیره:

prediction یا Forcasting به منظور برآورد متغیر وابسته به ازای مقادیر متغیر مستقل است اگر مقادیر درون نمونه ای باشند پیش بینی درون نمونه ای (In sample Forcasting) و اگر مقادیر، مقادیری بیرون از نمونه مورد بررسی باشند پیش بینی برون نمونه ای (out of sample) نامیده می شود.

$Y_f = \alpha + \beta X_f + \varepsilon_f$

برآورد: $\hat{Y}_f = \hat{\alpha} + \hat{\beta} X_f$ $\longrightarrow$ خطای پیش بینی: $e_f = Y_f - \hat{Y}_f$

$e_f(\alpha + \beta X_f + \varepsilon_f - \hat{\alpha} - \hat{\beta} X_f)$

$$E(e_f) = 0 \Rightarrow E(e_f) = E[\alpha + \beta X_f + \varepsilon_f - \hat{\alpha} - \hat{\beta} X_f]$$

$$= E(\alpha) + E(\beta X_f) + E(\varepsilon_f) - E(\hat{\alpha}) - X_f E(\hat{\beta})$$

بر اساس خاصیت بدون تورش بودن $\hat{\alpha}$ و $\hat{\beta}$ همچنین فرض اول کلاسیک $E(\varepsilon_f) = 0$ و $E(\hat{\beta}) = \beta$ و $E(\hat{\alpha}) = \alpha$ است. از طرفی می دانیم امید ریاضی عدد ثابت برابر با خود آن است.

$$E(e_f) = \alpha + \beta X_f + 0 - \alpha - \beta X_f = 0 \qquad E(e_f) = \underbrace{\alpha - \alpha}_{0} + \underbrace{(\beta - \beta)}_{0} X_f$$

$$Var(e_f) = E[e_f - E(e_f)]^2$$

اما چون $E(e_f) = 0$ است می توان نوشت:

$$\begin{cases} Y_f = \alpha + \beta X_f + \varepsilon_f \\ \hat{Y}_f = \hat{\alpha} + \hat{\beta} X_f \end{cases} \qquad e_f = Y_f - \hat{Y}_f$$

$$E(e_f^2) = E[\varepsilon_f - (\hat{\alpha} - \alpha) - X_f(\hat{\beta} - \beta)]^2$$

$$= E[\varepsilon_f^2 + (\hat{\alpha} - \alpha)^2 + X_f^2(\hat{\beta} - \beta)^2 - 2\varepsilon_f(\hat{\alpha} - \alpha) - 2\varepsilon_f X_f(\hat{\beta} - \beta) + 2X_f(\hat{\alpha} - \alpha)(\hat{\beta} - \beta)]$$

$$= \underbrace{E(\varepsilon_f^2)}_{\sigma^2} + \underbrace{E(\hat{\alpha} - \alpha)^2}_{Var(\hat{\alpha})} + X_f^2 \underbrace{E(\hat{\beta} - \beta)^2}_{Var(\hat{\beta})} - \underbrace{2E[\varepsilon_f(\hat{\alpha} - \alpha)]}_{0} - \underbrace{2X_f E[\varepsilon_f(\hat{\beta} - \beta)]}_{0} + 2X_f \underbrace{E(\hat{\alpha} - \alpha)(\hat{\beta} - \beta)}_{Cov(\hat{\alpha}, \hat{\beta})}$$

$$2E[\varepsilon_f(\hat{\alpha} - \alpha)] = 0 \qquad 2X_f E[\varepsilon_f(\hat{\beta} - \beta)] = 0$$

از طرفی داریم: $Var(\hat{\alpha}) = \sigma^2\left[\frac{1}{T} + \frac{\bar{X}^2}{\sum x_t^2}\right]$

$$E(e_f^2) = \sigma^2 + \sigma^2\left[\frac{1}{T} + \frac{\bar{X}^2}{\sum x_t^2}\right] + X_f^2 \frac{\sigma^2}{\sum x_t^2} - 2X_f \frac{\bar{X}\sigma^2}{\sum x_t^2} = \underbrace{\sigma^2}_{\text{فاکتورگیری از } (\sigma^2)}\left[1 + \frac{1}{T} + \frac{(X_f - \bar{X})^2}{\sum x_t^2}\right]$$

کمترین واریانس خطای پیش بینی:

$$\text{if}: X_f = \bar{X} \rightarrow var(e_f) = \sigma^2\left[1 + \frac{1}{T}\right]$$

$$(\underline{L}, \bar{L}) = e_f \pm SE(e_f)\, t_{1-\frac{\alpha}{2}}$$

$$Var(\hat{\beta}) = \frac{\sigma^2}{\sum x_t^2}$$

$$Cov(\hat{\alpha}, \hat{\beta}) = \frac{\bar{X}\sigma^2}{\sum x_t^2}$$

رگرسیون چندگانه (چند متغیره) Multiple Regression :

در دنیای واقعی به نظر بعید می‌رسد که یک پدیده فقط و فقط با یک متغیر توضیحی در ارتباط باشد و معمولاً چند متغیر توضیحی در تبیین رفتار پدیده‌ها مورد استفاده قرار می‌گیرند به عنوان مثال بعید است که قیمت سهام فقط به سود نقدی آن وابسته باشد. اخبار سیاسی، شک‌های نفتی، تغییرات نرخ ارز، تغییرات نرخ بهره، حجم پول، قوانین مالیاتی، حق بهره‌برداری از معادن، بهای مواد اولیه... همه متغیرهایی هستند که می‌توانند بر قیمت سهام موثر باشند. در نتیجه ساده انگارانه خواهد بود اگر فرض کنیم قیمت سهام با یک رگرسیون دو متغیره، قابل تبیین است.

در حالت کلی رگرسیون چندگانه را می‌توان به صورت

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \dots + \beta_k X_{kt} + \varepsilon_t \quad \text{یا} \quad Y_t = \beta_1 + \sum_{i=2}^{k} \beta_i X_{it} + \varepsilon_t$$

نمایش داد.

اگر چه بعداً رگرسیون فوق را بیشتر مورد بحث قرار می‌دهیم و آن را با استفاده از تحلیل ماتریس بررسی خواهیم نمود اما جهت سادگی، در این مرحله فقط رگرسیون با دو متغیر توضیحی را بررسی می‌کنیم.

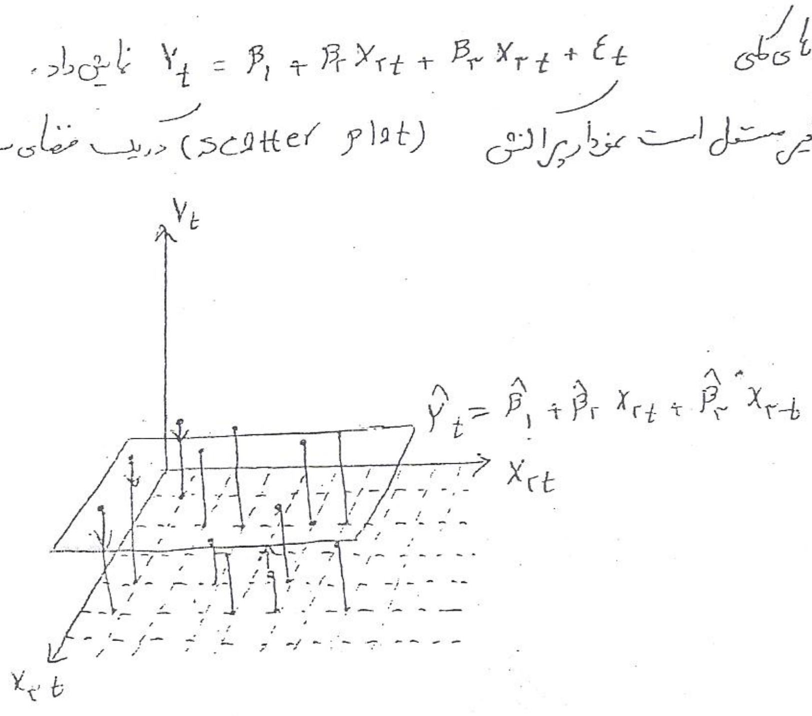
فرض کنید سود نقدی شرکت و اندازه آن تعیین کننده بازده سهام شرکت باشد بنابراین:

$$Y_{it} = \beta_1 + \beta_2 D_{it} + \beta_3 size_{it} + \varepsilon_{it}$$

البته توجه داریم در اینجا اندیس it به مفهوم داده‌های تلفیقی نیست بلکه مشخص هر کدام از i شرکت است نه اینکه همه آنها همزمان مورد تحلیل قرار گیرد.

رگرسیون فوق را می‌توان با نمادهای کلی $Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \varepsilon_t$ نمایش داد.

حال که متغیر وابسته تابعی از دو متغیر مستقل است نمودار پراکنش (scatter plot) در یک فضای سه بعدی ترسیم خواهد شد.



اکنون هدف از برازش رگرسیون عبور دادن صفحه‌ای از میان این نقاط است که این صفحه دارای دو شیب خواهد بود یک شیب نسبت X_{2t} و شیب دیگر نسبت به X_{3t} ، همچنین یک عرض از مبدأ خواهد داشت که مقدار آن همان ارتفاع آن از سطح افق است که هر دو مقدار X_{2t} و X_{3t} صفر فرض شود.

$$e_t^2 = (Y_t - \hat{Y}_t)^2 = (Y_t - \hat{\beta}_1 - \hat{\beta}_2 X_{2t} - \hat{\beta}_3 X_{3t})^2$$

$$\sum_{t=1}^{T} e_t^2 = \sum_{t=1}^{T} (Y_t - \hat{\beta}_1 - \hat{\beta}_2 X_{2t} - \hat{\beta}_3 X_{3t})^2$$

$$\begin{cases} \dfrac{\partial \sum e_t^2}{\partial \hat{\beta}_1} = -2\sum (Y_t - \hat{\beta}_1 - \hat{\beta}_2 X_{2t} - \hat{\beta}_3 X_{3t}) = 0 \\[2ex] \dfrac{\partial \sum e_t^2}{\partial \hat{\beta}_2} = -2\sum (Y_t - \hat{\beta}_1 - \hat{\beta}_2 X_{2t} - \hat{\beta}_3 X_{3t})(X_{2t}) = 0 \\[2ex] \dfrac{\partial \sum e_t^2}{\partial \hat{\beta}_3} = -2\sum (Y_t - \hat{\beta}_1 - \hat{\beta}_2 X_{2t} - \hat{\beta}_3 X_{3t})(X_{3t}) = 0 \end{cases}$$

معادله اول نرمال :

$$\sum Y_t - T\hat{\beta}_1 - \hat{\beta}_2 \sum X_{2t} - \hat{\beta}_3 \sum X_{3t} = 0 \longrightarrow \boxed{\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X}_{2t} - \hat{\beta}_3 \bar{X}_{3t}}$$

با توجه به اینکه معمولاً حل معادلات فوق طولانی است سعی می‌کنیم رگرسیون فوق ابتدا به صورت انحراف از میانگین بیان نموده سپس آن را حل نماییم.

$$1)\quad Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t \longrightarrow \frac{\sum Y_t}{T} = \frac{T\beta_1}{T} + \beta_2 \frac{\sum X_{2t}}{T} + \beta_3 \frac{\sum X_{3t}}{T} + \frac{\sum u_t}{T}$$

$$2)\quad \bar{Y} = \beta_1 + \beta_2 \bar{X}_2 + \beta_3 \bar{X}_3 + \bar{u}$$

$$1-2)\quad Y_t - \bar{Y} = (\beta_1 - \beta_1) + \beta_2 (X_{2t} - \bar{X}_2) + \beta_3 (X_{3t} - \bar{X}_3) + (u_t - \bar{u})$$

$$\boxed{y_t = \beta_2 x_{2t} + \beta_3 x_{3t} + \varepsilon_t}$$

$$\sum e_t^2 = \sum (y_t - \hat{y}_t)^2 = \sum (y_t - \hat{\beta}_2 x_{2t} - \hat{\beta}_3 x_{3t})^2 = 0$$

$$\begin{cases} \dfrac{\partial \sum e_t^2}{\partial \hat{\beta}_2} = -2\sum (y_t - \hat{\beta}_2 x_{2t} - \hat{\beta}_3 x_{3t})(x_{2t}) = 0 \\[2ex] \dfrac{\partial \sum e_t^2}{\partial \hat{\beta}_3} = -2\sum (y_t - \hat{\beta}_2 x_{2t} - \hat{\beta}_3 x_{3t})(x_{3t}) = 0 \end{cases} \longrightarrow$$

$$\begin{cases} \sum x_{2t} y_t - \hat{\beta}_2 \sum x_{2t}^2 - \hat{\beta}_3 \sum x_{3t} x_{2t} = 0 \\ \sum x_{3t} y_t - \hat{\beta}_3 \sum x_{3t}^2 - \hat{\beta}_2 \sum x_{2t} x_{3t} = 0 \end{cases}$$

روش حل معادلات به وسیله ماتریس:

$$\begin{bmatrix} \sum x_{2t}^2 & \sum x_{3t} x_{2t} \\ \sum x_{2t} x_{3t} & \sum x_{3t}^2 \end{bmatrix} \begin{bmatrix} \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix} = \begin{bmatrix} \sum x_{2t} y_t \\ \sum x_{3t} y_t \end{bmatrix}$$

$$AX = B \Rightarrow X = A^{-1} B$$

روش حل معادلات به وسیله جایگزینی:

$$\hat{\beta}_3 = \frac{\sum x_{2t} y_t - \hat{\beta}_2 \sum x_{2t}^2}{\sum x_{2t} x_{3t}} \Rightarrow \sum x_{3t} y_t - \hat{\beta}_2 \sum x_{2t} x_{3t} - \left(\frac{\sum x_{2t} y_t - \hat{\beta}_2 \sum x_{2t}^2}{\sum x_{2t} x_{3t}}\right) \sum x_{3t}^2 = 0$$

$$\sum x_{3t} y_t - \frac{\sum x_{2t} y_t \sum x_{3t}^2}{\sum x_{2t} x_{3t}} - \hat{\beta}_2 \sum x_{2t} x_{3t} + \hat{\beta}_2 \frac{\sum x_{2t}^2 \sum x_{3t}^2}{\sum x_{2t} x_{3t}} = 0 \rightarrow$$

$$\hat{\beta}_2 \left(\frac{\sum x_{2t} x_{3t} \sum x_{2t} x_{3t} - \sum x_{2t}^2 \sum x_{3t}^2}{\sum x_{2t} x_{3t}} \right) = \frac{\sum x_{3t} y_t \sum x_{2t} x_{3t} - \sum x_{2t} y_t \sum x_{3t}^2}{\sum x_{2t} x_{3t}} \rightarrow$$

$$\boxed{\hat{\beta}_2 = \frac{\sum x_{2t} y_t \sum x_{3t}^2 - \sum x_{3t} y_t \sum x_{2t} x_{3t}}{\sum x_{2t}^2 \sum x_{3t}^2 - (\sum x_{2t} x_{3t})^2}}$$

$$\boxed{\hat{\beta}_3 = \frac{\sum x_{3t} y_t \sum x_{2t}^2 - \sum x_{2t} y_t \sum x_{2t} x_{3t}}{\sum x_{3t}^2 \sum x_{2t}^2 - (\sum x_{2t} x_{3t})^2}}$$

$$\boxed{\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X}_2 - \hat{\beta}_3 \bar{X}_3}$$

مثال: رابطه بین بازده شرکت، اندازه شرکت و P/E هدف تحقیق است. اگر اندازه شرکت (size) ، X_{2t} و P/E شرکت X_{3t} و Y_t نیز بازده شرکت باشد، ضرایب رگرسیون بازده، اندازه شرکت و P/E را بدست آورید.

$(Y_t-\bar{Y})(X_{۲t}-\bar{X}_{۲t})(X_{۳t}-\bar{X}_{۳t})$

Y_t	$X_{۲t}$	$X_{۳t}$	y_t	$x_{۲t}$	$x_{۳t}$	$x_{۲t}y_t$	$x_{۳t}y_t$	$x_{۲t}x_{۳t}$	$x_{۲t}^۲$	$x_{۳t}^۲$
۷	۳	۷	۲	-۱	-۱	-۲	-۲	۱	۱	۱
۱۰	۲	۳	۵	-۲	-۵	-۱۰	-۲۵	۱۰	۴	۲۵
۳	۵	۱۰	-۲	۱	۲	-۲	-۴	۲	۱	۴
۲	۷	۱۲	-۳	۳	۴	-۶	-۱۲	۱۲	۹	۱۶
۳	۳	۸	-۲	-۱	۰	۰	۰	۰	۱	۰
۲۵	۲۰	۴۰	۰	۰	۰	-۲۱	-۴۳	۲۵	۱۶	۴۶

$$\hat{\beta}_۲=\frac{\sum x_{۲t}y_t\sum x_{۳t}^۲-\sum x_{۲t}x_{۳t}\sum x_{۳t}y_t}{\sum x_{۲t}^۲\sum x_{۳t}^۲-(\sum x_{۲t}x_{۳t})^۲}=\frac{(-۲۱)(۴۶)-۲۵(-۴۳)}{(۱۶)(۴۶)-۶۲۵}=\frac{۱۰۹}{۱۱۱}=۰/۹۸$$

$(۲۵)^۲$

$$\hat{\beta}_۳=\frac{\sum x_{۳t}y_t\sum x_{۲t}^۲-\sum x_{۲t}x_{۳t}\sum x_{۲t}y_t}{\sum x_{۲t}^۲\sum x_{۳t}^۲-(\sum x_{۲t}x_{۳t})^۲}=\frac{(-۴۳)(۱۶)-(۲۵)(-۲۱)}{۱۱۱}=-۱/۴۶$$

$$\hat{\beta}_۱=\bar{Y}-\hat{\beta}_۲\bar{X}_۲-\hat{\beta}_۳\bar{X}_۳$$

$$\hat{\beta}_۱=۵-۰/۹۸(۴)+۱/۴۶(۸)=۱۲/۷۶$$

$$y_t=\hat{\beta}_۱+\hat{\beta}_۲X_{۲t}-\hat{\beta}_۳X_{۳t}$$

$$Y_t=۱۲/۷۶+۰/۹۸X_{۲t}-۱/۴۶X_{۳t}$$

$$\hat{Y}_t=\hat{\beta}_۱-X_{۲t}$$

$\bar{Y}_t=۵$

$\bar{X}_{۲t}=۴$

$\bar{X}_{۳t}=۸$

جلسه آخر

* آنالیز واریانس در رگرسیون (Analysis of Variance):

در رگرسیون واریانس متغیر وابسته را می‌توان به دو جزء تغییرات توضیح داده شده و توضیح داده نشده یا پسماند تجزیه نمود.

همچنین می‌توان با استفاده از نسبت تغییرات توضیح داده شده به تغییرات توضیح داده نشده یک آماره F ساخت و در خصوص معنی داری کلی رگرسیون آزمون فرضیه نمود.

جدول آنالیز واریانس برای رگرسیون به صورت ذیل قابل ارائه است:

منبع تغییرات source of variation	sum of squares S.S	df	MS	
تغییرات توضیح داده شده Explain	$\boxed{ESS}$ $\sum \hat{y}^2_t$	$K-1$	$\frac{\sum \hat{y}^2_t}{K-1}$	$F=\frac{\sum \hat{y}^2_t/(K-1)}{\sum e^2_t/(T-K)}$
تغییرات توضیح داده نشده unexplain Residuals	$\boxed{RSS}$ $\sum e^2_t$	$T-K$	$\frac{\sum e^2_t}{T-K}$	
کل تغییرات Total variation	$\boxed{TSS}$ $\sum y^2_t$	$T-1$ *	$\frac{\sum y^2_t}{T-1}$	

K: تعداد پارامترها T: تعداد مشاهدات

* براساس قضیه جمع پذیری χ^2 مجموع چند χ^2 برابر با χ^2 است و درجه آزادی آن برابر با جمع درجه آزادی‌ها است

کای اسکوئر

$$T-1=(T-K)+(K-1)$$

در رگرسیون کلی چند متغیره آزمون معنی داری رگرسیون با استفاده از آماره F انجام می‌شود اما در رگرسیون دو متغیره $F=t^2_{\beta}$ است.

$$R^2=\frac{\sum \hat{y}^2_t}{\sum y^2_t} \quad , \quad R^2=1-\frac{\sum e^2_t}{\sum y^2_t}$$

$$F=\frac{\frac{\sum \hat{y}^2_t}{\sum y^2_t}\cdot\frac{1}{K-1}}{\frac{\sum y^2_t-\sum \hat{y}^2_t}{\sum y^2_t}\cdot\frac{1}{T-K}}=\frac{R^2/(K-1)}{(1-R^2)/(T-K)}$$

اگر متغیر وارد شده یک متغیر غیر مربوط باشد، R^2 کاهش خواهد [illegible]

اگر مربوط باشد، $\bar{R}^2$ را به مقداری که کاهش ناشی از [illegible] R و عبارت [illegible] جبران کند.

$$\bar{R}^2=1-\frac{T-1}{T-K}(1-R^2)$$

رگرسیون k متغیره:

در حالت کلی متغیر وابسته تابعی از k متغیر توضیحی است که به صورت ذیل بیان می‌گردد:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \dots + \beta_K X_{Kt} + \varepsilon_t \qquad X_{1t} = [1, 1, \dots, 1]$$

$$e_t = Y_t - \hat{Y}_t = Y_t - \hat{\beta}_1 - \hat{\beta}_2 X_{2t} - \hat{\beta}_3 X_{3t} - \dots - \hat{\beta}_K X_{Kt}$$

$$\sum e_t^2 = \sum (Y_t - \hat{\beta}_1 - \hat{\beta}_2 X_{2t} - \hat{\beta}_3 X_{3t} - \dots - \hat{\beta}_K X_{Kt})^2$$

یک دستگاه k متغیره و k معادله

$$(1)\quad \frac{\partial \sum e_t^2}{\partial \hat{\beta}_1} = -2 \sum (Y_t - \hat{\beta}_1 - \hat{\beta}_2 X_{2t} - \hat{\beta}_3 X_{3t} - \dots - \hat{\beta}_K X_{Kt}) = 0$$

$$\frac{\partial \sum e_t^2}{\partial \hat{\beta}_2} = -2 \sum (Y_t - \hat{\beta}_1 - \hat{\beta}_2 X_{2t} - \hat{\beta}_3 X_{3t} - \dots - \hat{\beta}_K X_{Kt})(X_{2t}) = 0$$

$$\vdots$$

$$\frac{\partial \sum e_t^2}{\partial \hat{\beta}_K} = -2 \sum (Y_t - \hat{\beta}_1 - \hat{\beta}_2 X_{2t} - \hat{\beta}_3 X_{3t} - \dots - \hat{\beta}_K X_{Kt})(X_{Kt}) = 0$$

سوال امتحان

بیان (فرم) ماتریسی رگرسیون k متغیره:

$$t=1 \qquad Y_1 = \beta_1 + \beta_2 X_{21} + \beta_3 X_{31} + \dots + \beta_K X_{K1} + \varepsilon_1$$

$$t=2 \qquad Y_2 = \beta_1 + \beta_2 X_{22} + \beta_3 X_{32} + \dots + \beta_K X_{K2} + \varepsilon_2$$

$$\vdots$$

$$t=T \qquad Y_T = \beta_1 + \beta_2 X_{2T} + \beta_3 X_{3T} + \dots + \beta_K X_{KT} + \varepsilon_T$$

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_T \end{bmatrix} = \begin{bmatrix} 1 & X_{21} & X_{31} & \dots & X_{K1} \\ 1 & X_{22} & X_{32} & \dots & X_{K2} \\ \vdots & \vdots & & & \\ 1 & X_{2T} & X_{3T} & \dots & X_{KT} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_T \end{bmatrix}$$

$$Y = X\beta + \varepsilon \qquad e = Y - \hat{Y}$$

$Y_{T\times1}$ و $X_{T\times K}$ و $\beta_{K\times1}$

$$e = Y - \hat{Y} \qquad e = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_T \end{bmatrix} \;\text{و}\; Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_T \end{bmatrix} \;\text{و}\; \hat{Y} = \begin{bmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ \vdots \\ \hat{Y}_T \end{bmatrix}$$

$$e'e = [e_1 \; e_2 \; \dots \; e_T] \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_T \end{bmatrix} = \sum e_t^2$$

$$e'e = (Y-\hat{Y})'(Y-\hat{Y}) \xrightarrow{\hat{Y} = X\hat{\beta}} (Y - X\hat{\beta})'(Y - X\hat{\beta}) = (Y' - \hat{\beta}'X')(Y - X\hat{\beta})$$

$$e'e = Y'Y - Y'X\hat{\beta} - \hat{\beta}'X'Y + \hat{\beta}'X'X\hat{\beta} \xrightarrow[1\times 1]{\underline{Y'_{1\times T}\, X_{T\times K}\, \hat{\beta}_{K\times 1}}} = Y'Y - 2Y'X\hat{\beta} + \hat{\beta}'X'X\hat{\beta}$$

$$\frac{\partial e'e}{\partial \hat{\beta}} = 0 \quad \overset{(1)}{-2X'Y} + \overset{(2)}{2X'X\hat{\beta}} = 0 \rightarrow X'X\hat{\beta} = X'Y$$

$$\underbrace{(X'X)^{-1}X'X}_{I}\hat{\beta} = (X'X)^{-1}X'Y \rightarrow \boxed{\hat{\beta} = (X'X)^{-1}X'Y}$$

:(۱)

$$[a_1 \; a_2 \; \dots \; a_k] \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{bmatrix} = \sum_{i=1}^{k} a_i x_i = A'X \rightarrow \frac{\partial (A'X)}{\partial X} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_k \end{bmatrix} = A \rightarrow$$

$$\frac{\partial\, 2Y'X\hat{\beta}}{\partial \hat{\beta}} = 2X'Y$$

(۲)

فرم مجذوری (Quadratic form) :

$$Z'AZ$$

$$[z_1 \; z_2 \; \dots \; z_k] \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1k} \\ a_{21} & a_{22} & \dots & a_{2k} \\ \vdots & \vdots & & \\ a_{k1} & a_{k2} & \dots & a_{kk} \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_k \end{bmatrix} = \sum_{j=1}^{k} \sum_{i=1}^{k} a_{ij} z_i z_j$$

اگر متقارن باشد

$$= a_{11} z_1^2 + \underbrace{a_{12} z_1 z_2 + a_{21} z_2 z_1}_{2a_{12} z_1 z_2} + \dots$$

$$\frac{\partial Z'AZ}{\partial Z} = 2AZ \rightarrow$$

$$\frac{\partial \overset{Z'}{\hat{\beta}'} \overset{A}{(X'X)} \overset{Z}{\hat{\beta}}}{\partial \hat{\beta}} = 2\overset{A}{X'X}\overset{Z}{\hat{\beta}}$$

مثال: داده‌های زیر موجود می‌باشد مطلوب است محاسبه $\hat{\beta}$؟

Y_t	X_2
۵	۲
۹	۴
۱۰	۵
۱۲	۶
۱۶	۷

$$X = \begin{bmatrix} 1 & 2 \\ 1 & 4 \\ 1 & 5 \\ 1 & 6 \\ 1 & 7 \end{bmatrix} \qquad Y = \begin{bmatrix} 5 \\ 9 \\ 10 \\ 12 \\ 16 \end{bmatrix} \qquad \hat{\beta} = (X'X)^{-1}X'Y$$

$$\hat{\beta} = \left[\begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 2 & 4 & 5 & 6 & 7 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 1 & 4 \\ 1 & 5 \\ 1 & 6 \\ 1 & 7 \end{bmatrix} \right]^{-1} \cdot \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 2 & 4 & 5 & 6 & 7 \end{bmatrix} \begin{bmatrix} 5 \\ 9 \\ 10 \\ 12 \\ 16 \end{bmatrix} = \begin{bmatrix} 5 & 24 \\ 24 & 130 \end{bmatrix}^{-1} \begin{bmatrix} 52 \\ 280 \end{bmatrix}$$

$$\hat{\beta} = \frac{1}{74} \begin{bmatrix} 130 & -24 \\ -24 & 5 \end{bmatrix} \begin{bmatrix} 52 \\ 280 \end{bmatrix} = \begin{bmatrix} \hat{\alpha} \\ \hat{\beta}_{\text{مقدم}} \end{bmatrix} = \hat{\beta}$$

بررسی خواص تخمین زننده‌های حداقل مربعات به صورت ماتریسی:

1) $E(\varepsilon) = 0$

$$E \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_T \end{bmatrix} = \begin{bmatrix} E(\varepsilon_1) \\ E(\varepsilon_2) \\ \vdots \\ E(\varepsilon_T) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

2) $E(\varepsilon\varepsilon') = \sigma^2_\varepsilon I$

$$E(\varepsilon\varepsilon') = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_T \end{bmatrix} [\varepsilon_1 \ \varepsilon_2 \ \dots \ \varepsilon_T] = E \begin{bmatrix} \varepsilon_1^2 & \varepsilon_1\varepsilon_2 & \dots & \varepsilon_1\varepsilon_T \\ \varepsilon_2\varepsilon_1 & \varepsilon_2^2 & \dots & \varepsilon_2\varepsilon_T \\ \vdots & & & \\ \varepsilon_T\varepsilon_1 & \varepsilon_T\varepsilon_2 & \dots & \varepsilon_T^2 \end{bmatrix} = E \begin{bmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & & \\ \vdots & & & \\ 0 & 0 & \dots & \sigma^2 \end{bmatrix}$$

$$E(\varepsilon\varepsilon') = \sigma^2 I$$

3) X is Nonstochastic.

4) $rank(X) = K$

$$rank = \min(rank_{cl}(X), rank_{Row}(X))$$

✓ تعداد سطرهای مستقل را رتبه سطری آن ماتریس می‌گویند.

✓ تعداد ستون‌های مستقل را رتبه ستونی آن ماتریس گویند.

۵) $Y = X\beta + \varepsilon$ رگرسیون خطی است

۶) $\varepsilon \sim N(0, \sigma^2 I)$

خواص آماری تخمین زننده‌های حداقل مربعات به صورت ماتریسی:

۱- تخمین زننده‌های حداقل مربعات بدون تورش هستند.

$E(\hat{\beta}) = \beta$

$$\hat{\beta} = (X'X)^{-1}X'Y = (X'X)^{-1}X'(X\beta + \varepsilon) = \underbrace{(X'X)^{-1}X'X}_{I}\beta + (X'X)^{-1}X'\varepsilon$$

$$\hat{\beta} = \beta + (X'X)^{-1}X'\varepsilon$$

$$E(\hat{\beta}) = E(\beta) + E\left[(X'X)^{-1}X'\varepsilon\right]$$

چون β یک عدد ثابت است و امید ریاضی عدد ثابت برابر با خود آن است بنابراین $E(\beta) = \beta$ و با توجه به اینکه Xها غیرتصادفی هستند امید ریاضی از آنها عبور کرده و فقط بر ε اعمال می‌شود:

$$E(\hat{\beta}) = \beta + (X'X)^{-1}X'E(\varepsilon)$$

با توجه به فرض اول کلاسیک $E(\varepsilon) = 0$

$$E(\hat{\beta}) = \beta + (X'X)^{-1}X'0 = \beta$$

$var(\hat{\beta}) = ?$

چون اثبات شده $E(\hat{\beta}) = \beta$ است از طریقی دیگر داریم:

$$var(\hat{\beta}) = E[\hat{\beta} - E(\hat{\beta})]^2 = E(\hat{\beta} - \beta)^2$$

$$\hat{\beta} = \beta + (X'X)^{-1}X'\varepsilon \rightarrow \hat{\beta} - \beta = (X'X)^{-1}X'\varepsilon$$

$$var(\hat{\beta}) = E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'] = E\left[(X'X)^{-1}X'\varepsilon\varepsilon'X(X'X)^{-1}\right]$$

با توجه به اینکه Xها غیرتصادفی هستند امید از آنها عبور کرده و بر ε اعمال می‌شود:

$$var(\hat{\beta}) = (X'X)^{-1}X'E(\varepsilon\varepsilon')X(X'X)^{-1}$$

براساس فرض کلاسیک:

$$var(\hat{\beta}) = (X'X)^{-1}X'\sigma^2 I X(X'X)^{-1}$$

I عضو خنثی ضرب:

$$var(\hat{\beta}) = \sigma^2(X'X)^{-1}X'IX(X'X)^{-1} = \sigma^2(X'X)^{-1}X'X(X'X)^{-1}$$

$$var(\hat{\beta}) = \sigma^2 I(X'X)^{-1} = \sigma^2(X'X)^{-1}$$

متغیرهای مجازی (Dummy Variable):

در نظریه رگرسیون برای کمی کردن صفات کیفی از رگرسیون با متغیرهای مجازی استفاده می‌شود. متغیرهای مجازی متغیرهای هستند که مقادیر ۰ و ۱ را اختیار می‌کنند. اگر مشاهده مورد نظر صفت مورد بررسی را داشته باشد مقدار یک و اگر آن صفت را نداشته باشد مقدار صفر را می‌گیرد.

یکی از کاربردهای متغیر مجازی در فاینانس بررسی اثرات تقویمی است. در حقیقت در فاینانس آثاری به نام اثرات هفته، اثرات نیمه دوم ماه، اثر ژانویه، اثر ماه رمضان در ایران و ... وجود دارد.

اگر بخواهیم این آثار را کمی کنیم باید یک متغیر مجازی برای آنها تعریف کنیم.

فرض کنید می‌خواهیم اثر ژانویه را در بازار سهام آمریکا بررسی کنیم. داده‌های ماهانه سال‌های ۲۰۰۰ تا ۲۰۰۸ را استخراج نموده و به صورت ماهانه در نرم‌افزار وارد می‌کنیم.

	Dumj	r_t	$intr_t$	e_t
۲۰۰۰ ۱	۱	۰/۱	⋮	⋮
۲۰۰۰ ۲	۰	۰/۰۵		
⋮	⋮	⋮		
۲۰۰۰ ۱۲	۰			
۲۰۰۱ ۱	۱			
۲۰۰۲ ۲	۰			
⋮	⋮			
۲۰۰۲ ۱۲	۰			
⋮				

$$r_t = \beta_1 + \beta_2\, intr_t + \beta_3\, e_t + \beta_4\, Dumj + \varepsilon_t$$

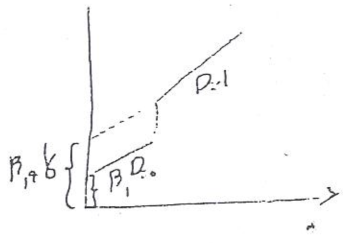
Ls Y c intr e Dumj

تأثیرات متغیر مجازی بر شیب و عرض از مبدأ:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \varepsilon_t$$

$$Y_t = \beta_1 + \beta_2 X_{2t} + \gamma D_t + \varepsilon_t$$

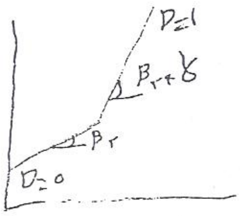
$$D_t \begin{cases} 1 \rightarrow Y_t = (\beta_1 + \gamma) + \beta_2 X_{2t} + \varepsilon_t \\ 0 \rightarrow Y_t = \beta_1 + \beta_2 X_{2t} + \varepsilon_t \end{cases}$$



تأثیر متغیر مجازی بر عرض از مبدأ:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \gamma D_t X_{2t} + \varepsilon_t$$

$$D_t \begin{cases} 1 \rightarrow Y_t = \beta_1 + (\beta_2 + \gamma) X_{2t} + \varepsilon_t \\ 0 \rightarrow Y_t = \beta_1 + \beta_2 X_{2t} + \varepsilon_t \end{cases}$$

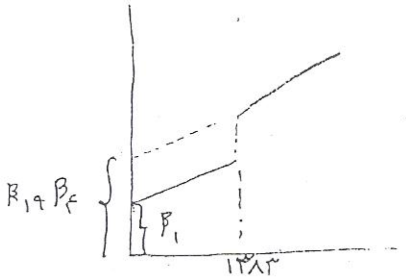


تأثیر متغیر مجازی بر شیب:

مثال: تأثیر تغییر مدیر کل بورس (دکتر صالح آبادی) بر بورس اوراق بهادار تهران

$$r_{it} = \beta_1 + \beta_2\, intr_t + \beta_3\, oil_t + \beta_4\, Dum_t + \varepsilon_t$$

$$Dum = \begin{array}{c c} \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} & \begin{array}{c} \vdots \\ ۸۲ \\ ۸۳ \\ \vdots \end{array} \end{array}$$



مثال: تأثیر شرایط رکود و رونق اقتصادی بر بازده بازار

$$r_{it} = \beta_1 + \beta_2\, r_{mt} + \beta_3\, r_{mt} D_t + \varepsilon_t$$

$$D_t = \begin{cases} 1 & \text{: رونق} \\ 0 & \text{: رکود} \end{cases}$$

شیب : $\beta_2 + \beta_3$

شیب : β_2

خود رگرسیون:

$$y_t = \alpha + \beta\, y_{t-1} + \varepsilon_t$$

مدل گام تصادفی نامیده می‌شود (اگر $\beta=1$ باشد) Random walk

نقض فروض کلاسیک :

۱- فرض اول :

$E(\varepsilon_t)=0 \qquad t=1,\dots,T$

در عمل ممکن است امید ریاضی اجزای اخلال صفر نباشد یعنی $E(\varepsilon_t)\neq 0$ در این صورت می توان گفت فرض اول کلاسیک نقض شده است در صورت نقض این فرض، فقط امید ریاضی $\boxed{\hat{\alpha}}$ برابر با α نخواهد بود در این ضریب تورش دار خواهد شد اما سایر ضرایب یعنی $\boxed{\hat{\beta}}$ بدون تورش خواهد بود.

اثبات :

$$\hat{\beta}=\frac{\sum x_t y_t}{\sum x_t^2}=\frac{\sum x_t(\beta x_t+\varepsilon_t)}{\sum x_t^2}=\beta+\frac{\sum x_t\varepsilon_t}{\sum x_t^2}$$

$$E(\hat{\beta})=E(\beta)+\frac{\sum x_t\,E(\varepsilon_t)}{\sum x_t^2}$$

با توجه به اینکه فرض اول کلاسیک نقض شده است دیگر امید ریاضی ε_t، مساوی صفر نیست بلکه مساوی یک عدد ثابت مانند c است در اینصورت :

$$E(\hat{\beta})=\beta+\frac{\sum x_t c}{\sum x_t^2}=\beta+\frac{c\cancel{\sum x_t}^{\,0}}{\sum x_t^2}=\beta$$

$$\hat{\alpha}=\bar{Y}-\hat{\beta}\bar{X}=\alpha+\beta\bar{X}+\bar{\varepsilon}-\hat{\beta}\bar{X}$$

$$E(\hat{\alpha})=\alpha+\beta\bar{X}+\frac{1}{T}\sum E(\varepsilon_t)-E(\hat{\beta})\bar{X}=\alpha+\cancel{\beta\bar{X}}+\frac{1}{T}\sum c-\cancel{\beta\bar{X}}=\alpha+c$$

کاربرد مالی در شاخص ارزیابی

$$S_p=\frac{r_p-r_f}{\sigma_p}\ ,\qquad T_p=\frac{r_p-r_f}{\beta_p}\ ,\qquad J_p=\alpha_p$$

$$r_{p_t}-r_{f_t}=\alpha+\beta(r_{mt}-r_{f_t})+\varepsilon_t$$

$$R_i-R_f=\alpha+\beta(R_m-R_f)+\varepsilon_t$$

$\alpha>0$ براساس معیار جنسن عملکرد مناسب بوده است.

$\alpha<0$ براساس معیار جنسن عملکرد نامناسب بوده است.

نقض فرض اول بر شیب تأثیری ندارد ولی بر عرض از مبدا تأثیر دارد و در نهایت بر شاخص جنسن در ارزیابی عملکرد اثر دارد.

۲- فرض عدم همبستگی پیاپی :

$E(\varepsilon_t\varepsilon_s)=0 \quad if \quad t\neq s$

$E(\varepsilon_t\varepsilon_s)=\sigma^2 \quad if \quad t=s$

اگر فرض فوق نقض شود یعنی $E(\varepsilon_t \varepsilon_s) \neq 0$ در این صورت گفته می‌شود همبستگی پیاپی بین اجزای اخلال به وجود آمده است و فرض دوم کلاسیک نقض شده است.

برای سادگی فرض کنید همبستگی وجود داشته باشد و نوع همبستگی از نوع مرتبه اول مارکف است.

$$\varepsilon_t = \rho \varepsilon_{t-1} + u_t$$

که u_t یک جزء اخلال خوش رفتار است یعنی $u_t \sim iid(0, \sigma^2)$

توضیح مستقل و یکنواخت

$$\varepsilon_t = \rho \varepsilon_{t-1} + u_t \quad , \quad \varepsilon_{t-1} = \rho \varepsilon_{t-2} + u_{t-1}$$

$$\varepsilon_t = \rho[\rho \varepsilon_{t-2} + u_{t-1}] + u_t = \rho^2 \varepsilon_{t-2} + \rho u_{t-1} + u_t \quad , \quad \varepsilon_{t-2} = \rho \varepsilon_{t-3} + u_{t-2}$$

$$\varepsilon_t = \rho^2(\rho \varepsilon_{t-3} + u_{t-2}) + \rho u_{t-1} + u_t = \rho^3 \varepsilon_{t-3} + \rho^2 u_{t-2} + \rho u_{t-1} + u_t$$

$\vdots$

$$\varepsilon_t = u_t + \rho u_{t-1} + \rho^2 u_{t-2} + \rho^3 u_{t-3} + \cdots$$

$$E(\varepsilon_t) = E(u_t) + \rho E(u_{t-1}) + \rho^2 E(u_{t-2}) + \rho^3 E(u_{t-3}) + \cdots$$

$$E(\varepsilon_t) = 0 + 0 + \cdots + 0 = 0$$

$var(\varepsilon_t) = ?$

$$\varepsilon_t = u_t + \rho u_{t-1} + \rho^2 u_{t-2} + \rho^3 u_{t-3} + \cdots$$

$$var(\varepsilon_t) = E(\varepsilon_t^2) = E[u_t^2 + \rho^2 u_{t-1}^2 + \rho^4 u_{t-2}^2 + \rho^6 u_{t-3}^2 + \cdots + 2\rho u_t u_{t-1} + 2\rho^2 u_t u_{t-2} + \cdots]$$

$$\Rightarrow E(u_t^2) + \rho^2 E(u_{t-1}^2) + \rho^4 E(u_{t-2}^2) + \cdots + 2\rho E(u_t u_{t-1}) + 2\rho^2 E(u_t u_{t-2}) + \cdots]$$

$$\Rightarrow E(\varepsilon_t^2) = \sigma_u^2 + \rho^2 \sigma_u^2 + \rho^4 \sigma_u^2 + \rho^6 \sigma_u^2 + \cdots = \sigma_u^2(1 + \rho^2 + \rho^4 + \rho^6 + \cdots) \quad \text{if } |\rho| < 1$$

$$\Rightarrow E(\varepsilon_t^2) = \sigma_u^2\left(\frac{1}{1-\rho^2}\right) = \frac{\sigma_u^2}{1-\rho^2}$$

$$var(\varepsilon_t) = \frac{\sigma_u^2}{1-\rho^2}$$

$$\begin{cases} Y_t = \alpha + \beta X_t + \varepsilon_t \\ \varepsilon_t = \rho \varepsilon_{t-1} + u_t \end{cases} \rightarrow \hat{\rho} = \frac{\sum \varepsilon_t \varepsilon_{t-1}}{\sum \varepsilon_{t-1}^2}$$

$$\operatorname{cov}(\varepsilon_t,\varepsilon_{t-1}) = E[\varepsilon_t - E(\varepsilon_t)][\varepsilon_{t-1} - E(\varepsilon_{t-1})] = E(\varepsilon_t\varepsilon_{t-1})$$

$$= E[(u_t + \rho u_{t-1} + \rho^2 u_{t-2} + \rho^3 u_{t-3} + \dots) \times (u_{t-1} + \rho u_{t-2} + \rho^2 u_{t-3} + \rho^3 u_{t-4} + \dots)]$$

$$= \rho E(u_{t-1}^2) + \rho^3 E(u_{t-2}^2) + \rho^5 E(u_{t-3}^2) + \dots + E(u_t u_{t-1}) + \rho^2 E(u_t u_{t-2}) + \dots$$

$$= \rho\sigma_u^2 + \rho^3\sigma_u^2 + \rho^5\sigma_u^2 + \rho^7\sigma_u^2 + \dots$$

$$= \rho\sigma_u^2(1 + \rho^2 + \rho^4 + \rho^6 + \dots) = \frac{\rho\sigma_u^2}{1-\rho^2}$$

✓ $$E(\varepsilon_t,\varepsilon_{t-s}) = \rho^{|s|}\frac{\sigma_u^2}{1-\rho^2}$$

ماتریس واریانس کوواریانس:

$$\operatorname{cov}(\varepsilon) = E(\varepsilon\varepsilon') = \frac{\sigma_u^2}{1-\rho^2}\begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{T-1} \\ \rho & 1 & & & \\ \vdots & \vdots & & & \\ \rho^{T-1} & \rho^{T-2} & & \dots & 1 \end{bmatrix} \neq \sigma^2 I$$

$$E(\varepsilon\varepsilon') = E\left\{\begin{bmatrix}\varepsilon_1\\ \varepsilon_2\\ \vdots\\ \varepsilon_T\end{bmatrix}[\varepsilon_1\ \varepsilon_2\ \dots\ \varepsilon_T]\right\} = E\begin{bmatrix} \varepsilon_1^2 & \varepsilon_1\varepsilon_2 & \dots & \varepsilon_1\varepsilon_T \\ \varepsilon_2\varepsilon_1 & \varepsilon_2^2 & \dots & \varepsilon_2\varepsilon_T \\ \vdots & & & \\ \varepsilon_T\varepsilon_1 & \varepsilon_T\varepsilon_2 & \dots & \varepsilon_T^2 \end{bmatrix}$$

$$= \begin{bmatrix} E(\varepsilon_1^2) & E(\varepsilon_1\varepsilon_2) & \dots & E(\varepsilon_T\varepsilon_1) \\ E(\varepsilon_2\varepsilon_1) & E(\varepsilon_2^2) & & \\ \vdots & & & \\ E(\varepsilon_T\varepsilon_1) & E(\varepsilon_2\varepsilon_T) & \dots & E(\varepsilon_T^2) \end{bmatrix}$$

درصورتی که فرض مستقل بیان شده نقض شود دیگر تخمین زننده های حداقل مربعات دارای کمترین واریانس نخواهند بود و باید از تخمین زننده های دیگری به نام حداقل مربعات تعمیم یافته GLS استفاده کرد.

Generalized Least Squares

بهترین پیامد نقض فرض عدم همبستگی پیاپی ناکارایی در تخمین زننده‌ها است یعنی اگر فرض عدم همبستگی نقض شود دیگر تخمین زننده‌های حداقل مربعات دارای کمترین واریانس نخواهند بود. بزرگ شدن واریانس به معنی کاهش کارایی است پس با نقض فرض عدم همبستگی پیاپی کارایی تخمین زننده‌ها کاهش خواهد یافت و تخمین زننده کارا GLS خواهد بود.

راه‌های تشخیص همبستگی پیاپی:

اگر فرض همبستگی پیاپی مرتبه اول را در نظر بگیریم $\varepsilon_t = \rho\varepsilon_{t-1} + u_t$ خواهد بود در نتیجه اگر $\rho = 0$ باشد همبستگی پیاپی نداریم اگر $\rho \neq 0$ باشد و از نظر آماری معنی‌دار باشد همبستگی پیاپی وجود دارد.

با توجه به اینکه همیشه ε ها غیر قابل مشاهده هستند نمی‌توان برای تخمین ρ از ε ها استفاده نمود لذا می‌توان از تخمین ε ها یعنی e_t ها استفاده نمود که e_t ها پسماندهای رگرسیون اصلی هستند.

$$e_t = \rho e_{t-1} + u_t$$

$$Y_t = \alpha + \beta X_t + \varepsilon_t$$

$$\varepsilon_t = \rho\varepsilon_{t-1} + u_t$$

$$e_t = Y_t - \hat{\alpha} - \hat{\beta} X_t$$

$$e_t = \rho e_{t-1} + u_t \qquad \hat{\rho} = \frac{\sum e_t e_{t-1}}{\sum e_{t-1}^2}$$

سوال امتحان ص ۹۵ pdf

آزمون دوربین- واتسون (Durbin - Watson):

$$DW = \frac{\sum_{t=2}^{T}(e_t - e_{t-1})^2}{\sum_{t=2}^{T} e_{t-1}^2}$$

این آماره به وسیله دوربین و واتسون مطرح شده است.

$$DW = \frac{\sum e_t^2 - 2\sum e_t e_{t-1} + \sum e_{t-1}^2}{\sum e_{t-1}^2} \simeq 2\frac{\sum e_{t-1}^2}{\sum e_{t-1}^2} - 2\frac{\sum e_t e_{t-1}}{\sum e_{t-1}^2} = 2 - 2\hat{\rho}$$

e_t	e_{t-1}
.112	–
.113	.112
.115	.113
.114	.115

$$DW \simeq 2(1-\hat{\rho})$$

وقتی همبستگی پیاپی وجود ندارد $\rho = 0$ است آماره $DW = 2$ است.

اگر همبستگی پیاپی مثبت وجود داشته باشد آماره $\hat{\rho}$ بزرگتر از صفر بوده و DW کوچکتر از ۲ خواهد شد.

در حالت حدی وقتی $\hat{\rho} = 1$ باشد $DW = 0$ خواهد بود پس همبستگی سریالی مثبت کامل، با دوربین واتسون صفر در تناظر

است .

۲) اگر همبستگی سریالی منفی داشته باشیم آماره DW بزرگتر از ۲ خواهد بود .

۳) در حالت حدی وقتی ۱- → ρ باشد DW ، ۴ خواهد بود .

بنابراین :

$0 \le DW \le 4$

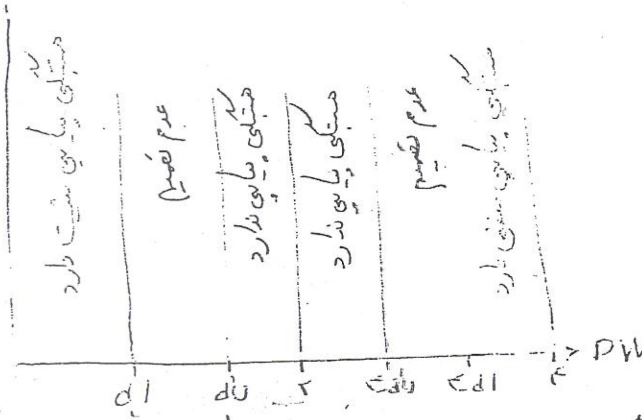
$-1 \le \rho \le 1$

جدول آماره دوربین واتسون :

تعداد پارامترها چون کلی عرض از مبدأ است

$K' = K - 1$

T	$K'=1$		$K'=2$		...
	dl	dU	dl	dU	

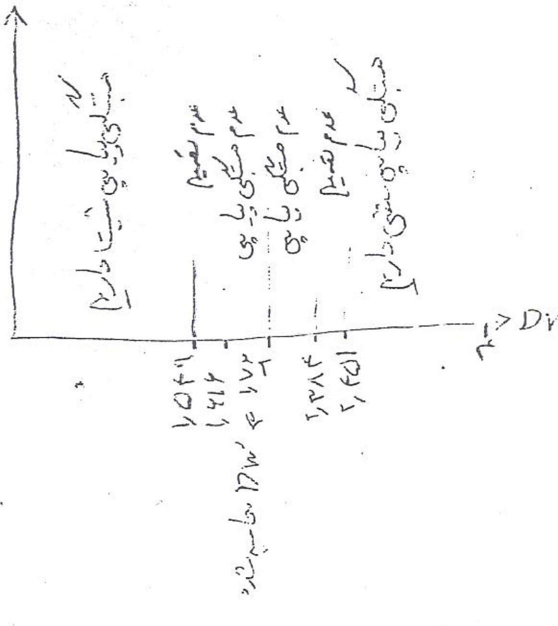


مثال: رگرسیون بین بازده بازار و بازده پرتفوی یعنی همان مدل CAPM برازش شده و آماره DW = 1/72 بدست آمده است اگر تعداد مشاهدات ۶۰ مشاهده باشد DW را برای بررسی همبستگی پیاپی آزمون نمایید .

حل: با مراجعه به جدول DW ، با توجه به اینکه تعداد مشاهدات ۶۰ و تعداد پارامترها ۲ است باید در ردیف T=60 ، $K'=K-1=2-1=1$

$dL = 1/549$

$dU = 1/616$

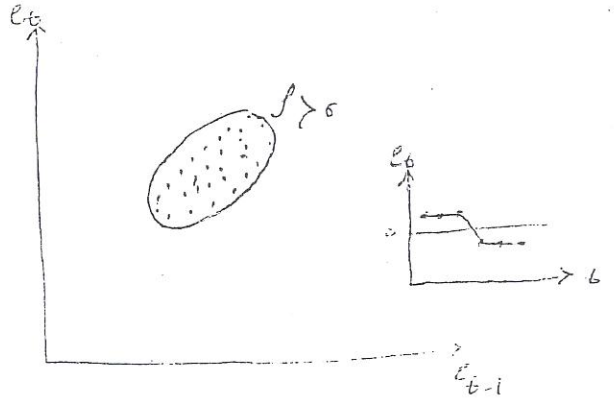


با توجه به اینکه آماره دوربین واتسون در ناحیه میانی قرار گرفته فرضیه H0 را مبنی بر نبود همبستگی پیاپی نمی توان رد کرد یعنی همبستگی پیاپی نداریم البته سطح خطا ۵٪ در نظر گرفته شده است .

آزمون همبستگی پیاپی به صورت نموداری :

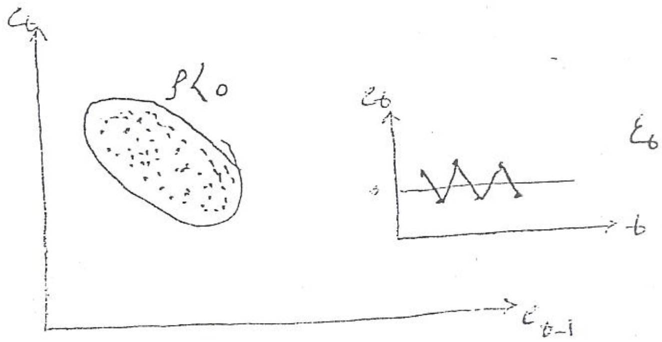
در نمودار ترسیمی ، نمودار e_t را در مقابل e_{t-1} ترسیم می کنیم :

الف) همبستگی پیاپی مثبت :



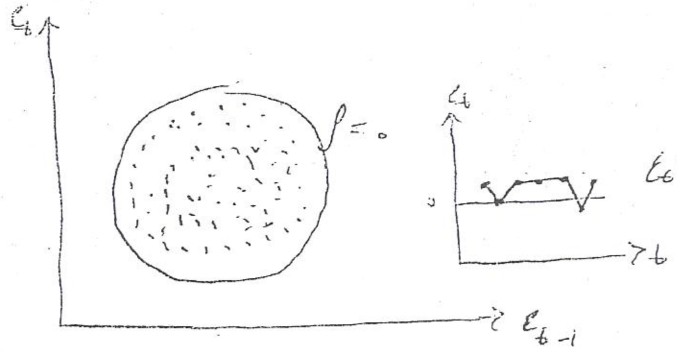
توالی مثبت : $e_t = 0.7 e_{t-1}$

ب) همبستگی پیاپی منفی :



توالی منفی : $e_t = -0.7 e_{t-1}$

ج) عدم همبستگی :



بدون الگوی مشخص : $e_t = c(e_{t-1})$

در همبستگی پیاپی مثبت e_t های مثبت با e_t های مثبت دنبال می شود و e_t های منفی با e_t های منفی دنبال می شود در حالی که در همبستگی پیاپی منفی e_t های مثبت با e_t های منفی و e_t های منفی با e_t های مثبت دنبال می شود.

آزمون ضریب لاگرانژ (LM) یا گادفری :

در این آزمون فرض محدود کننده $e_t = \rho e_{t-1} + u_t$ را کنار گذاشته و فرض کلی تری پیش می کشیم :

$$e_t = \rho_1 e_{t-1} + \rho_2 e_{t-2} + \dots + \rho_s e_{t-s} + u_t$$

که یک فرآیند مرتبه s مارکف است.

برای آزمون همبستگی فوق نمی توان از آزمون DW استفاده نمود بلکه باید از آزمونی به نام ضریب لاگرانژ LM استفاده نمود.

ضریب لاگرانژ

1- LS Y c X_1 X_2 ... X_k → e

2- LS e c e(-1) e(-2) ... e(-s) → R^2

3- $(T-s)R^2 \sim \chi^2_s$ χ^2_s $\boxed{\chi^2_5 \text{ مثال}}$

4. χ^2_s محاسبه شده را با χ^2 جدول مقایسه نموده و در صورتیکه χ^2_s محاسبه شده در سطح خطای مورد نظر از مقدار جدول بیشتر باشد گفته می شود همبستگی پیاپی وجود دارد.

۱- رگرسیون اصلی را محاسبه نموده، پسماندهای را به دست می آورید.

۲- پسماندها را روی وقفه های مختلف پسماندها رگرس نموده (رگرسیون کمکی) و R^2 رگرسیون کمکی را استخراج می کنیم.

۳- تعداد مشاهدات قابل استفاده T-s مشاهده است زیرا از s وقفه e استفاده کرده ایم.

مثال: رگرسیون اصلی بازده اسناد خزانه و تورم بصورت ذیل برآورد شده است

$\hat{r}_t = 0.02 + 1.5\,\mathrm{Inf}_t$ $T=50$

که در عبارت فوق $\hat{r}_t$ بازده اسناد خزانه و Inf_t نرخ تورم آمریکا است برای بررسی همبستگی پیاپی پسماندهای رگرسیون فوق را استخراج نموده رگرسیون کمکی به صورت ذیل برازش شده است $R^2 = 0.73$

$\hat{e}_t = 0.001 + 0.15 e_{t-1} - 0.22 e_{t-2}$ $s=2$

همبستگی پیاپی را در سطح ۵٪ (خطای نوع اول) آزمون کنید.

حل: می دانیم که $TR^2 \sim \chi^2$ و می توان برای آزمون همبستگی پیاپی R^2 مدل فوق را در تعداد مشاهدات استفاده شده در رگرسیون ضرب نموده و با جدول χ^2 مقایسه نمود توجه داریم که در این آزمون:

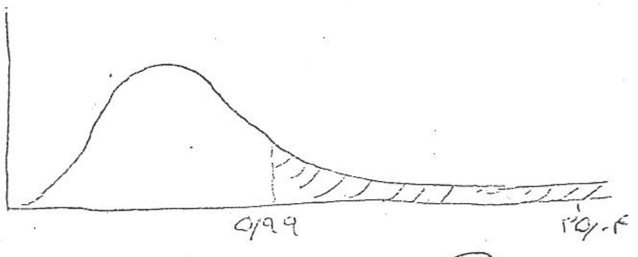
(1)
$$\begin{cases} H_0: \rho_1 = \rho_2 = 0 \\ H_1: \rho_i \neq 0 \quad i=1,2 \end{cases}$$

ρ_1 و ρ_2 = ضرایب e_{t-1} و e_{t-2}

حال می توان χ^2 را محاسبه نموده و از جدول χ^2 مقدار بحرانی را به دست آورده با آن مقایسه کرد.

$(T-2)R^2 \sim \chi^2_2 \rightarrow (50-2)\cdot 0.73 = 35.04$, $\chi^2_{5\%,2} = 5.99$

با توجه به اینکه آماره محاسبه شده در ناحیه بحرانی قرار گرفته فرضیه H_0 را مبنی بر نبود همبستگی پیاپی رد می کنیم یعنی همبستگی پیاپی وجود دارد.



مثال: با استفاده از Eeviews.

یک رگرسیون فرضی:

۱) دستور پسماندهای رگرسیون

```
Ls   tb3   c   ppi
series   e=resid
```

دستور رگرسیون پسماندها در وقفه

```
ls   e   c   e(-1)   e(-2)   ppi
```

محاسبه χ^2

```
scalar   chisq1 = R2*T = TR2
function Refrence
scalar   chisqcr=@qchisq(.90  2)
```

دریافت χ^2 جدول:

۲) روش استفاده از تست LM در نرم افزار

view → Residuals Test → serial correlation LM Test

۳) روش نمودار:

quic → Graph → e e(-1) → scatter diagram

۴) با استفاده از آزمون DW

راه‌های رفع همبستگی پیاپی:

فرض کنید براساس یکی از آزمون‌های همبستگی پیاپی (بررسی‌های دوربین واتسون یا LM) به وجود همبستگی پیاپی در رگرسیون اصلی پی برده‌ایم حال می‌خواهیم همبستگی پیاپی را رفع نماییم.

روش کوکران اورکات

رگرسیون اصلی

الگوی همبستگی پیاپی

$1.\ Y_t = \alpha + \beta X_t + \varepsilon_t$

$\varepsilon_t = \rho \varepsilon_{t-1} + u_t$

$u_t \sim N(0, \sigma^2)$

معادله اصلی را یک زمان به عقب t-1 بردیم و در ρ ضرب می‌کنیم

$2.\ Y_{t-1} = \alpha + \beta X_{t-1} + \varepsilon_{t-1}$

$3.\ \rho Y_{t-1} = \rho\alpha + \rho\beta X_{t-1} + \rho\varepsilon_{t-1}$ حاصل

تا جایی ادامه می‌دهیم که

$\rho^n \approx \rho^{n-1}$

حاصل $1-3)\ Y_t - \rho Y_{t-1} = \alpha(1-\rho) + \beta(X_t - \rho X_{t-1}) + (\varepsilon_t - \rho\varepsilon_{t-1})$

$\varepsilon_t - \rho\varepsilon_{t-1} = u_t$

$Y_t - \rho Y_{t-1} = \alpha(1-\rho) + \beta(X_t - \rho X_{t-1}) + u_t$

$Y^* = \alpha^* + \beta X_t^* + u_t$ — برازش شد

حالا مدل دیگر همبستگی پیاپی نیست حالا قاعده و تبدیل پرسین و نسبتی استفاده می‌کنیم

۱- رگرسیون اصلی را تخمین زده و $\hat{e}_t$ ها را به دست می‌آوریم. $Y_1^* = \sqrt{1-\rho^2}\, Y_1$ و $X_1^* = \sqrt{1-\rho^2}\, X_1$ تبدیل پریس وینستون

که با استفاده از فرمول $\frac{\sum e_t e_{t-1}}{\sum e_{t-1}^2}$ مقدار $\hat{\rho}$ را تخمین می‌زنیم، اگر مرحله اضافه بود تبدیل فوق به بالایی قبلی مرحله ۳ دادها در محل مرحله نسبت برگرداندن $\hat{\rho}$ با بدیع‌ها

۳- متغیرها را براساس عبارت‌های جبری ذیل تبدیل نموده و Y^* و X_t^* را به دست می‌آوریم، داده‌ها تبدیل شده و هم اکنون هم مشاهده در تحقیق تنها بعضی‌ها استفاده می‌کنیم از طرف جدول

$t=1 \qquad Y_1^* = \sqrt{1-\hat{\rho}^2}\, Y_1 \quad , \quad X_1^* = \sqrt{1-\hat{\rho}^2}\, X_1$

$t=2,\dots,T \qquad Y_t^* = Y_t - \hat{\rho} Y_{t-1} \quad , \quad X_t^* = X_t - \hat{\rho} X_{t-1}$

۴- رگرسیون Y^* روی X^* را برازش نموده و مجدداً با دهای جدید $\hat{e}_t^*$ را به دست می‌آوریم حال مرحله ۲ تا ۴ را آنقدر تکرار می‌کنیم تا $\hat{\rho}^{n} \simeq \hat{\rho}^{n-1}$ (روش تکراری کوکران اورکات).

روش دوم رفع مشکل بیاپی:

$LS \quad Y \quad C \quad X \quad AR(1)$ اتورگرسیو مرتبه ۱

روش سوم رفع مشکل بیاپی: GLS

Generalized Least squares

$$E(\varepsilon\varepsilon') = \underbrace{\frac{\sigma^2_u}{1-\rho^2}\begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{T-1} \\ \rho & 1 & & & \\ \vdots & & \ddots & & \\ \rho^{T-1} & & & & 1 \end{bmatrix}}_{\Omega} = \sigma^2_e \,\Omega$$

$$var(\hat{\beta}_{OLS}) = \sigma^2 (X'X)^{-1} X'\Omega X (X'X)^{-1}$$

$$Y = X\beta + \varepsilon$$

$$PY = PX\beta + P\varepsilon$$

$$Y^* = X^*\beta + \varepsilon^*$$

$$\hat{\beta}_{GLS} = (X^{*\prime}X^*)^{-1} X^{*\prime} Y^* = (X'P'PX)^{-1} X'P'PY \qquad \underline{P'P = \Omega^{-1}} \qquad (X'\Omega^{-1}X)^{-1} X'\Omega^{-1} Y$$

$$var(\hat{\beta}_{GLS}) = E\left[(X'\Omega^{-1}X)^{-1} X'\Omega^{-1} \varepsilon\varepsilon' \Omega^{-1} X (X'\Omega^{-1}X)^{-1}\right]$$

$$= (X'\Omega^{-1}X)^{-1} X'\Omega^{-1} E(\varepsilon\varepsilon') \Omega^{-1} X (X'\Omega^{-1}X)^{-1}$$

$$= \sigma^2_e \underbrace{(X'\Omega^{-1}X)^{-1} X'\Omega^{-1}}_{I} \overbrace{\Omega\,\Omega^{-1}}^{I} X (X'\Omega^{-1}X)^{-1}$$

$var(\hat{\beta}_{GLS}) = \sigma_u^2 (X'\Omega^{-1}X)^{-1}$

بررسی کارایی: براساس قضیه آیتکن اگر اجزای اخلال کروی نباشند $var(\hat{\beta}_{GLS}) < var(\hat{\beta}_{OLS})$

یعنی همبستگی پیاپی یا ناهمسانی داشته باشد تخمین زننده GLS در میان تمامی تخمین زننده‌ها کاراترین

$\boxed{\hat{\beta}_{FGLS} = (X'\hat{\Omega}^{-1}X)^{-1}X'\hat{\Omega}^{-1}Y}$

ناهمسانی واریانس (Hetroscedasticity):

فرض سوم از فروض کلاسیک مربوط به واریانس اجزای اخلال بود فرض می‌کردیم واریانس اجزای اخلال طی مشاهدات مختلف ثابت است و به صورت نمادین $E(\varepsilon_t^2)=\sigma^2$ نمایش داده می‌شود یا به عبارت دیگر

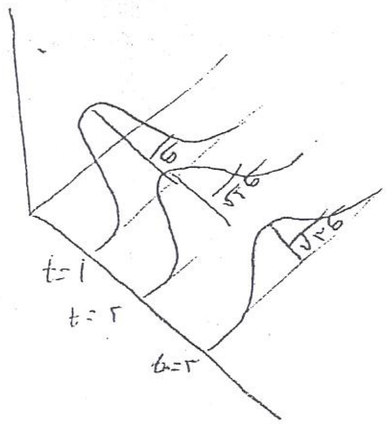
$$E(\varepsilon\varepsilon') = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & & \\ \vdots & & & \\ 0 & & \cdots & \sigma^2 \end{bmatrix}$$

اگر فرض فوق نقض شود دیگر واریانس‌ها با یکدیگر برابر نخواهند بود بلکه واریانس هر مشاهده ممکن است با دیگر واریانس‌ها متفاوت باشد:

که Z_t یک متغیر توضیحی یا حتی متغیری بیرون از مدل است: $E(\varepsilon_t^2) = Z_t\sigma^2$

t متغیر روند زمانی است و مقادیر ۱، ۲، ...، T را اختیار می‌کند یعنی: $E(\varepsilon_t^2) = t\sigma^2$

$E(\varepsilon_1^2) = \sigma^2$ ، $E(\varepsilon_2^2) = 2\sigma^2$ ، $E(\varepsilon_3^2) = 3\sigma^2$ ... $E(\varepsilon_T^2) = T\sigma^2$

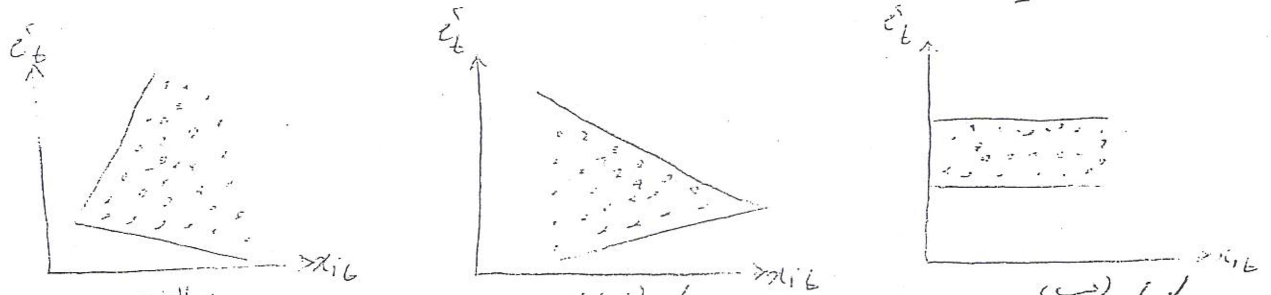


روش‌های مختلفی برای انجام ناهمسانی واریانس وجود دارد از جمله آنها می‌توان به روش‌های ترسیمی گلدفلد کوانت (Gold Feld Quant) وایت (White)، بارتلت، پارک و ... اشاره کرد که در اینجا آزمون ترسیمی و آزمون وایت مورد بحث قرار می‌گیرد.

آزمون ترسیمی:

براساس پراکندگی‌های مجذور اجزای اخلال می‌توان نسبت به همسانی یا ناهمسانی واریانس تصمیم گرفت بدین صورت که اگر متغیر مشکوک به علت ناهمسانی نسبت به x_{ij} باشد آن را روی محور افقی و $\hat{e}_i^2$ را روی محور عمودی قرار داده، نمودار پراکنش را رسم می‌کنیم اگر در نمودار پراکنش دامنه پراکندگی تغییر نماید می‌توان نتیجه گرفت ناهمسانی واریانس وجود دارد. در صورتی که رابطه‌ای بین متغیرها که روی محور افقی قرار گرفته و $\hat{e}_i^2$ ها وجود نداشته باشد ناهمسانی واریانس وجود ندارد.

نمودارهای ذیل مثال‌هایی از ناهمسانی واریانس هستند:



(الف) پراکندگی زیاد شده — (ب) پراکندگی کم شده — (پ) پراکندگی ثابت

در نمودارهای الف و ب دامنه پراکندگی با تغییر متغیر x تغییر کرده است در حالی که در نمودار پ دامنه تغییرات مجذور اجزای اخلال ثابت است و در حقیقت $\hat{e}_i^2$ به x_{ij} بستگی ندارد.

برای انجام روش ترسیمی در صورتی که ندانیم کدام یک از متغیرهای توضیحی موجب ناهمسانی شده است می‌توان بر روی محور افقی $\hat{Y}$ را به جای x_{ij} قرار داد. این روش به دلیل ترسیمی بودن و امکان ورود خطای حسی، چندان قابل استفاده نیست بنابراین از روش‌های دیگری که عمدتاً کمی هستند استفاده می‌کنیم.

آزمون ناهمسانی واریانس وایت: (white)

آزمون وایت با رگرسیون کردن مجذور اجزای اخلال بر روی متغیرهای توضیحی به توان ۲ و حاصلضرب متغیرهای توضیحی به دنبال بررسی وجود ناهمسانی واریانس است. اگر R^2 حاصل از رگرسیون کردن مجذور اجزای اخلال بر روی مجذور متغیرهای توضیحی و حاصلضرب آنها در تعداد مشاهدات ضرب شود توزیع حاصل یک χ^2 خواهد بود و می‌توان آن را با جدول χ^2 مقایسه نمود. در صورتی که TR^2 در ناحیه بحرانی قرار گیرد نتیجه خواهیم گرفت که رابطه بین اجزای اخلال و متغیرهای توضیحی معنی‌دار است و ناهمسانی واریانس وجود دارد/ به بیان دیگر هنگامی که اجزای اخلال با متغیرهای توضیحی رابطه داشته باشند که در قالب R^2 این رابطه مشخص می‌شود با توجه به اینکه متغیرهای توضیحی یعنی x ها ثابت نیستند واریانس اجزای اخلال ثابت نبوده و ناهمسانی واریانس

[illegible] برابر نبودن واریانس اجزای اخلال یعنی ناهمسانی واریانس.

اگر بخواهیم به صورت گام به گام این آزمون را بیان کنیم طی مراحل ذیل آزمون وایت انجام خواهد شد.

۱- رگرسیون اصلی را برازش نموده و پسماندهای آن را به دست می‌آوریم؛

$$LS\ Y\ C\ X_2\ X_3\ \ldots\ X_n \rightarrow e$$

۲- رگرسیون کمکی را با استفاده از مجذور اجزای اخلال برازش می‌کنیم؛

$$e^2_t = \omega_1 + \omega_2 X_{2t} \cdots \omega_k X_{kt} + \theta_2 X^2_{2t} + \theta_3 X^2_{3t} + \theta_k X^2_{kt} + \theta_2 X_{2t} X_{3t} + \cdots \theta_s X_{(k-1)t} X_{kt}$$

بدون عبارت‌های ضربی without cross terms

$$LS\ e_t^2\ C\ X_2\ X_3\ \ldots\ X_n\ X_2^2\ X_3^2\ \ldots\ X_n^2$$

$$LS\ e_t^2\ C\ X_2\ X_3\ \ldots\ X_n\ X_2^2\ X_3^2\ \ldots\ X_n^2\ X_2X_3\ X_2X_4\ \ldots\ X_{n-1}X_n$$

با عبارت‌های ضربی with cross term

۳- R^2 حاصل از رگرسیون کمکی را در تعداد مشاهدات ضرب کرده و χ^2 حاصل از این دو عبارت را با χ^2 جدول مقایسه می‌کنیم.

۴- χ^2 محاسبه شده را با χ^2 جدول با درجه آزادی برابر با تعداد پارامترها (k) مقایسه نموده در صورتی که χ^2 محاسبه شده در ناحیه بحرانی قرار گیرد رگرسیون اصلی دچار مشکل ناهمسانی واریانس است و باید رفع شود.

$$\begin{cases} H_0 : \sigma_i^2 = \sigma_j^2 \\ H_a : \sigma_i^2 \neq \sigma_j^2 \quad i \neq j \end{cases}$$

$\sigma_t^2 = \sigma^2 X_{2t}$ $\qquad$ $\sigma_t^2 = \frac{\sigma^2}{X_t^2}$

راه‌های رفع ناهمسانی واریانس: ضرب یا تقسیم بر جزء عبارتی که باعث ناهمسانی شده

اگر به وسیله آزمون‌های ناهمسانی تشخیص داده شده که ناهمسانی وجود دارد می‌توان به وسیله تبدیل مناسب که شامل تقسیم کردن یا ضرب کردن متغیرها بر عامل ناهمسان کننده واریانس است مشکل را رفع کرد. با همان مثال فرض کنید $\sigma_t^2 = \sigma^2 X_{2t}$

یعنی عامل ناهمسان کننده واریانس متغیر توضیحی دوم است به عبارت دیگر اگر مدل اصلی را تخمین بزنیم هنگامی که امید ریاضی اجزای اخلال به توان ۲ محاسبه شود (یعنی واریانس اجزای اخلال) عبارت $\sigma^2 X_{2t}$ به دست خواهد آمد یعنی $E(\varepsilon_t^2) = \sigma^2 X_{2t}$

بدیهی است اگر به جای ε_t در مدل $\varepsilon_t^* = \left(\frac{\varepsilon_t}{\sqrt{X_t}}\right)$ وجود داشته باشد در این صورت هنگام گرفتن امید ریاضی، عامل ناهمسان کننده حذف خواهد شد زیرا:

$$E(\varepsilon_t^{*2}) = E\left(\frac{\varepsilon_t^2}{X_{2t}}\right) = \frac{1}{X_{2t}} E(\varepsilon_t^2) = \frac{1}{\cancel{X_{2t}}} \sigma^2 \cancel{X_{2t}} = \sigma^2$$

$$Y_t = \alpha + \beta X_t + \varepsilon_t \rightarrow \frac{Y_t}{\sqrt{X_t}} = \frac{\alpha}{\sqrt{X_t}} + \beta\frac{X_t}{\sqrt{X_t}} + \frac{\varepsilon_t}{\sqrt{X_t}} \rightarrow Y_t^* = \alpha^* + \beta X_t^* + \varepsilon_t^*$$

به عنوان مثال اگر الگوی ناهمسانی به صورت $\sigma_t^2 = \frac{\sigma^2}{X_{3t}}$ باشد داریم:

$$Y_t \sqrt{X_{3t}} = \alpha \sqrt{X_{3t}} + \beta_1 X_{2t} \sqrt{X_{3t}} + \beta_2 X_{3t} \sqrt{X_{3t}} + \varepsilon_t \sqrt{X_{3t}}$$

در داده‌های مالی ناهمسانی واریانس اغلب به دلیل متفاوت بودن اندازه شرکت‌ها به وجود می‌آید برای خنثی کردن اثر سایز و مماشت از ناهمسان شدن واریانس از تقسیم متغیرهای وارد شده در رگرسیون به سایز یا اندازه شرکت استفاده می‌شود می‌دانیم اندازه شرکت برابر با قیمت روز سهام x تعداد سهام شرکت است که به آن Market capitalization یا Market cap یا به اختصار Mkt cap گفته می‌شود البته در مالی متغیرهای دیگر نیز می‌توانند موجب ناهمسانی شوند ممکن است دو شرکت از لحاظ اندازه یکسان باشند اما گردش مالی یک شرکت از شرکت دیگر متفاوت باشد این تفاوت می‌تواند موجب ناهمسانی شود همچنین متغیرهایی مانند فروش نیز می‌توانند موجب ایجاد ناهمسانی شود بنابراین برای رفع ناهمسانی معمولاً متغیرهای وابسته و مستقل را بر متغیرهایی که در فوق گفته شد تقسیم می‌کنند البته پیش فرض محقق این است که ناهمسانی از نوع ضربی است در صورتی که ناهمسانی از نوع تقسیمی باشد باید به متغیرهای سایز، متغیرهای رگرسیون را ضرب کرد.

روش دیگر رفع ناهمسانی واریانس استفاده از روش تخمین حداقل مربعات تعمیم یافته است.

در روش حداقل مربعات تعمیم یافته (GLS) ماتریس Ω به گونه‌ای تعریف خواهد شد که نقش تقسیم یا ضرب را ایفا نماید به عنوان مثال در موردی که عامل ناهمسان کننده واریانس X_{3t} بود و فرم الگوی ناهمسانی واریانس به صورت $\frac{\sigma^2}{X_{3t}} = \sigma_t^2$ بود ماتریس Ω به صورت

$$\Omega = \begin{bmatrix} \frac{1}{\sqrt{X_{31}}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sqrt{X_{32}}} & & \\ \vdots & & & \\ 0 & 0 & \cdots & \frac{1}{\sqrt{X_{3T}}} \end{bmatrix}$$

تشکیل خواهد شد و تخمین زننده GLS به صورت:

$$\hat{\beta}_{GLS} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Y$$

تخمین زده خواهد شد.

مثال:

$$LS \quad e_t^2 \quad c \quad X_2 \quad X_3 \quad X_2^2 \quad X_3^2 \quad X_2X_3$$

الگوی ناهمسانی:

$$\hat{\sigma}_t^2 = \hat{\alpha} + \hat{\delta}_1 X_{2t} + \hat{\delta}_2 X_{3t} + \hat{\delta}_3 X_{2t}^2 + \hat{\delta}_4 X_{3t}^2 + \hat{\delta}_5 X_{2t} X_{3t}$$

مشکلات اقتصادسنجی نشان می‌دهد که روش‌های مبتنی بر روش موزون که بر ضرب یا تقسیم عامل ناهمسانی کنند در متغیرهای دیگر مدل است، منجر به رفع ناهمسانی واریانس نخواهد شد. البته در مواردی نیز ناهمسانی رفع خواهد شد اما از نوع ناهمسانی رفع نخواهد شد. به همین دلیل اقتصادسنجی به نام آلبرت وایت که آزمون وایت نیز به وسیله او ارائه شده است روشی را برای رفع ناهمسانی واریانس پیشنهاد نمود در این روش به جای آنکه ناهمسانی را رفع کنیم از انحراف معیارهایی که سازگار هستند استفاده خواهیم کرد این انحراف‌های معیار به وسیله روشی که وایت معرفی نموده است محاسبه می‌شود که به آنها:

White Heteroscedasticitic consistent covariances

گفته می‌شود محاسبات این ماتریس واریانس کوواریانس فراتر از سطح دوره حاضر است اما در نرم‌افزار Eviewes به سادگی می‌توان گزینه

White Heteroscedasticitic consistent covariances

را انتخاب نموده و تخمین زننده‌های سازگار را به دست آورد

Residual Test → white Heteroscedasticitic

Estimate → option → Heteroscedasticitic → white ...

اگر Prob < 0.05 باشد ناهمسانی در سطح 5% وجود دارد و یا Obs* Residual خیلی بزرگ است.

همخطی (Multicolinearity):

در فروض کلاسیک دیدیم داشتیم که همخطی کامل وجود ندارد: No perfect Multicolinearity

$\sum x_t^2 \neq 0$ یا $X'X$ دارای رتبه کامل ستونی است اما در عمل اگر چه ممکن است همخطی کامل وجود نداشته باشد همخطی شدید نیز می‌تواند کارایی تخمین زننده‌ها را کاهش دهد به بیان ساده همخطی کامل به همبستگی شدید بین متغیرهای توضیحی اطلاق می‌شود.

اگر همبستگی بین متغیرهای توضیحی به گونه‌ای باشد که معنی‌داری ضرایب را تحت تأثیر قرار دهد و سبب شود ضرایب معنی‌دار معنی‌داری خود را از دست دهند گفته می‌شود همخطی شدید است. اثر همخطی کاهش کارایی است یعنی همخطی شدید سبب افزایش SE ضرایب شده و tها کوچک می‌شود، کوچک شدن tها سبب غیرمعنی‌دار شدن ضرایب شده و به نادرست ضرایب معنی‌دار را بی‌معنی تشخیص داده می‌شود.

یان کمنتا یکی از اقتصادسنجان در کتاب مبانی اقتصادسنجی اشاره می‌کند که بین متغیرهای اقتصادی همخطی وجود دارد و هنگامی که صحبت از همخطی می‌شود بحث وجود یا نبود آن نیست، بلکه شدت آن مهم است همخطی شدید مضر است و برای رفع آن باید اقدام نمود اما اگر همخطی در حدی باشد که تخمین‌ها را تحت تأثیر قرار دهد می‌توان از بررسی و حذف آن

شرایط همخطی ... همچنین پیامد همخطی، بی معنی شدن ضرایب است یا کاهش کارایی است البته اثرات منفی دیگری نیز دارد مانند بی ثبات شدن پارامترها در اثر همخطی بالا، مورد انتظار است. در صورتی که همخطی کامل باشد تخمین مدل امکان پذیر نخواهد بود و در هنگام تخمین ضرایب نرم افزاری، پیغامی مبنی بر نزدیکی بودن ماتریس ایکس پریم ایکس نمایش می دهد (near singular matrix).

ماتریس تکین به ماتریسی گفته می شود که دترمینان آن صفر است. ماتریسی که دترمینان صفر داشته باشد معکوس پذیر نخواهد بود. اگر ماتریس $X'X$ معکوس پذیر نباشد $\hat{\beta}$ قابل محاسبه نخواهد بود به خاطر داریم که معادله محاسبه $\hat{\beta}$ به صورت $\hat{\beta} = (X'X)^{-1}X'Y$ داده می شود، اگر همخطی کامل باشد تشخیص آن مشکل نیست زیرا اساساً تخمین امکان پذیر نخواهد بود و می توان با بررسی بیشتر، می توان متغیری را که موجب همخطی کامل شده است را حذف نمایید.
ممکن است این سؤال پیش آید که در دنیای واقعی مگر همخطی کامل ممکن است؟ در پاسخ این سؤال مثبت است ممکن است دانشجویی متغیری را بر روی دارایی ها، سرمایه و بدهی رگرس نماید:

$Y \sim A \quad cap \quad deb$

اما اول حسابداری بیان می کند:

دارایی = سرمایه + بدهی

بنابراین asset ترکیب خطی debt را capital است.

ترکیب خطی به رابطه خطی بین ستون های ماتریس یا سطرهای ماتریس گفته می شود اگر ماتریس X به صورت

$$X = \begin{bmatrix} 1 & X_{11} & X_{21} & \cdots & X_{k1} \\ 1 & X_{12} & X_{22} & & \\ \vdots & \vdots & \vdots & & \\ 1 & X_{1T} & X_{2T} & \cdots & X_{kT} \end{bmatrix}$$

باشد ترکیب خطی بین ستون ها به صورت $c_i - c_j + \alpha c_s$ نمایش داده می شود.

البته هر ترکیب دیگری نیز بین ستون ها ممکن است ستون ام (c_i) می تواند به هر شکل دیگری به ستون های دیگری مرتبط باشد به این رابطه خطی، همخطی کامل می گویند. اما در دنیای واقعی آنچه شایع است همخطی کامل نیست بلکه همخطی بالا است که به سادگی قابل تشخیص نیست تخمین مدل ممکن است اما کارایی آن به دلیل همخطی بالا، پایین است در نتیجه لازم است روش هایی برای تشخیص همخطی بالا در نظر گرفته شود.

راه های تشخیص همخطی:

روش های متعددی برای تشخیص همخطی وجود دارد اما توضیح همه این روش ها خارج از حوصله کلاسی است ذیلاً به برخی از این روش ها اشاره می کنیم:

۱- بررسی ثبات تخمین ها و مقایسه معنی داری تک تک ضرایب و معنی داری کلی رگرسیون:

در این روش مدل از نظر معنی داری کلی و معنی داری تک تک ضرایب بررسی می شود در صورتی که تک تک ضرایب معنی دار

نباشند اما مدل به صورت کلی معنی دار باشد یعنی آماره های t کوچک باشد اما آماره F معنی دار باشد و یا به عبارت دیگر R^2 بالا، اما تعداد ضرایب معنی دار اندک باشد امکان همخطی بالا وجود دارد همچنین بی ثباتی پارامترها دلیل دیگری بر همخطی شدید است به عنوان مثال اگر رگرسیونی با ۱۰۰ مشاهده مورد تخمین قرار گرفته باشد و ضریب تعیین X_2، ۰/۷۴ به دست آمده باشد با اضافه کردن ۲ مشاهده ممکن است ضرایب تفاوت زیادی از خودشان دهند و ضریب مذکور از ۰/۷۴ به ۰/۹ به عنوان مثال تغییر نماید این بدین معنی است که پارامترها نسبت به تغییر مشاهدات بی ثبات هستند و دلیل این بی ثباتی می تواند همخطی بالا باشد.

۲- عامل تورم واریانس:

در این روش می خواهیم بررسی نماییم که هر کدام از متغیرهای مستقل به چه میزان با سایر متغیرهای مستقل وابسته هستند برای این کار کافیست متغیرهای مستقل را به نوبت بر روی سایر متغیرهای مستقل رگرسیون نموده و R^2 را به دست آوریم پس از به دست آوردن R^2 می توان با محاسبه عامل تورم واریانس که به صورت $VIF_j=\frac{1}{1-R^2_j}$ محاسبه می شود نسبت به شدید بودن همخطی نتیجه گیری نمود که در فرمول فوق عامل تورم واریانس (Variance Inflating Factor) یا VIF_j برای متغیر j ام و R^2_j ضریب تعیین رگرسیون متغیر j ام بر روی سایر متغیرهای توضیحی است به صورت گام به گام می توان عامل تورم واریانس را به صورت ذیل محاسبه نمود:

۱- متغیر توضیحی j ام را بر روی سایر متغیرها رگرسیون می کنیم یعنی:

$$1.\ LS\ X_j\ c\ X_2\ X_3 \ldots \cancel{X_j} \ldots X_k \longrightarrow R^2_j$$

$$2.\ VIF_j=\frac{1}{1-R^2_j}$$

۳- در صورتی که VIF_j که j می تواند هر کدام از متغیرهای توضیحی باشد بزرگتر از ۱۰ باشد همخطی شدید است اما اگر هیچ کدام از متغیرهای توضیحی دارای VIF_j بزرگتر از ۱۰ باشد مدل دچار همخطی شدید نیست. بدیهی است هنگامی VIF_j بزرگتر از ۱۰ خواهد بود که R^2_j بزرگتر از ۰/۹ باشد.

۴- عدد مشخصه (condition number):

در این روش ابتدا قطر اصلی ماتریس $X'X$، λI را کم کرده و دترمینان ماتریس حاصل را محاسبه می کنیم دترمینان حاصل، یک معادله درجه n برحسب λ خواهد بود (ابعاد ماتریس $X'X$ را n فرض کردیم) از حل این معادله مقادیر ویژه به دست خواهد آمد. برای محاسبه عدد مشخصه کافی است بزرگترین مقدار ویژه (λ_{max}) به کوچکترین مقدار ویژه (λ_{min}) تقسیم شده و جذر آن

محاسبه شود به صورت مرحله‌ای می‌توان محاسبه عدد مشخصه را به صورت ذیل نمایش داد:

$1.\ |X'X - \lambda I| = 0$

$2.\ a_0\lambda^k + a_1\lambda^{k-1} + \dots + a_k = 0$

۳- از حل معادله فوق $\lambda_k, \dots, \lambda_3, \lambda_2, \lambda_1$ ← λ_{min} , λ_{max} بدست می‌آید.

$4.\ CN = \sqrt{\frac{\lambda_{max}}{\lambda_{min}}}$

5. If $CN > 20$ → هم‌خطی شدید است

If $CN < 20$ → هم‌خطی شدید نیست

مثال: داده‌های مربوط به بازده سهام، بازده ارز و نرخ بهره به صورت ذیل داده شده است. اگر داده‌های نقلی نرخ ارز به عنوان متغیر مستقل ۱، نرخ بهره به عنوان متغیر مستقل ۲ و نرخ سهام به عنوان متغیر وابسته باشد بررسی نمایید آیا هم‌خطی شدید است یا خیر؟

Y_t	X_{2t}	X_{3t}
۶	۱	۳
-۱	-۱	-۲
۱۵	۳	۹
۲۱	۵	۱۵
۱۷	۴	۱۲

$$X = \begin{bmatrix} 1 & 1 & 3 \\ 1 & -1 & -2 \\ 1 & 3 & 9 \\ 1 & 5 & 15 \\ 1 & 4 & 12 \end{bmatrix} \qquad X'X = \begin{bmatrix} 5 & 12 & 37 \\ 12 & 52 & 155 \\ 37 & 155 & 463 \end{bmatrix}$$

$\lambda_1 = 0/031923$, $\lambda_2 = 2/117214$, $\lambda_3 = 517/8509$

$CN = \sqrt{\frac{517.85}{0/031923}} = 127/212 > 20$ هم‌خطی شدید است

دستور ماتریس ترانسپوز در Excel: shift ctrl enter
$= mmult(transpose\ \ \ \ )$

راه‌های رفع هم‌خطی: به مانند سایر فروض کلاسیک راه‌های مختلف برای رفع هم‌خطی وجود دارد، از جمله این روش‌ها می‌توان به حذف متغیر توضیحی، استفاده از مشاهدات اضافی، استفاده از متغیرها به صورت انحراف از میانگین، استفاده از رگرسیون ریج و استفاده از مؤلفه‌های اصلی اشاره نمود. ذیلاً به بررسی و توضیح هر یک از این روش‌ها می‌پردازیم:

۱- حذف متغیرهای توضیحی:

یکی از راه‌های رفع هم‌خطی حذف متغیرهایی است که موجب هم‌خطی شده است البته ممکن است همه متغیرها سبب هم‌خطی باشند ولی ممکن است تعدادی از آنها بیشتر با دیگر متغیرها همبستگی داشته باشند بنابراین در صورتی که بخواهیم از استراتژی حذف متغیر ایجاد کننده متغیر هم‌خطی استفاده کنیم باید ابتدا متغیری را حذف نماییم که عامل تورم واریانس

بالاتری دارد به عنوان مثال رابطه بین حجم پول m_1، m_2 و m_3، هم‌خطی m_2 و m_3 با بقیه نسبت به m_1 شدیدتر است در نتیجه ما می‌توانیم از متغیری که مشکل بالاتری دارد که همبستگی بالاتر با VIF بالاتر نشان داده می‌شود شروع کرده و حذف متغیر را انجام دهیم. اگر چه حذف متغیر موجب افزایش کارایی تخمین زننده خواهد شد اما تورش تصریح ایجاد خواهد کرد (specification biased) به بیان دیگر می‌توان نشان داد اگر مدل صحیح به صورت (True model) TM. $Y = X_1\beta_1 + X_2\beta_2 + \varepsilon$ است، محقق با حذف متغیر آن را تخمین زده باشد.

$$FM: Y = X_1\beta_1 + \varepsilon \Rightarrow \hat{\beta}_1 = (X_1'X_1)^{-1}X_1'Y$$

با جایگذاری Y در معادله فوق می‌توان نشان داد که تورش تصریح (تورش حذف متغیر مربوط) ایجاد شده است.

$$\hat{\beta}_1 = (X_1'X_1)^{-1}X_1'(X_1\beta_1 + X_2\beta_2 + \varepsilon)$$

$$\hat{\beta}_1 = (X_1'X_1)^{-1}X_1'X_1\beta_1 + (X_1'X_1)^{-1}X_1'X_2\beta_2 + (X_1'X_1)^{-1}X_1'\varepsilon$$

$$E(\hat{\beta}_1) = E(\beta_1) + (X_1'X_1)^{-1}X_1'X_2E(\beta_2) + (X_1'X_1)^{-1}X_1'\overset{0}{E(\varepsilon)}$$

$$E(\hat{\beta}_1) = \beta_1 + (X_1'X_1)^{-1}X_1'X_2\beta_2 \neq \beta_1$$

همانطور که در معادله فوق نشان داده شده است با حذف متغیر X_2 از مدل صحیح (TM) ضرایب باقیمانده در مدل که با بردار β_1 نشان داده شده است تورش دار تخمین خواهد خورد یعنی $E(\hat{\beta}_1) \neq \beta_1$ مگر اینکه بردارهای X_1 و X_2 متعامد باشند یعنی ضرب داخلی آنها صفر باشد. صفر بودن ضرب داخلی دو بردار به مفهوم استقلال خطی آن دو بردار خواهد بود در حالی که در خصوص هم‌خطی می‌دانیم که آنها مستقل از هم نیستند زیرا اگر مستقل می‌بودند دلیلی بر حذف متغیر X_2 وجود نداشت باشد بنابراین در حالت کلی اگر چه حذف متغیر توضیحی موجب رفع هم‌خطی خواهد شد اما تورش حذف متغیر را ایجاد خواهد کرد.

۲- استفاده از مشاهدات بیشتر:

در این روش سعی می‌شود برای رفع هم‌خطی از مشاهدات بیشتری استفاده شود ممکن است این سؤال پیش آید اگر محقق مشاهدات بیشتری می‌داشت دلیلی بر استفاده نکردن آن در تخمین مرحله اول وجود نداشت، بنابراین اگر تعداد مشاهدات کم باشد محقق نمی‌تواند تعداد مشاهدات را افزایش دهد. اما این استدلال همیشه صحیح نیست اگر چه در مورد متغیرهای کلان، متغیر تولید کننده داده نیست و از داده‌های موجود استفاده می‌کند اما در خصوص داده‌های

پرسشنامه‌ای را آزمایش کند محقق می‌تواند در صورت بروز مشکل هم‌خطی داده‌های بیشتری تهیه و مشکل را رفع نماید.

۳- استفاده از مؤلفه‌های اصلی (Principale component):

مؤلفه‌های اصلی بردارهای متناظر با مقادیر ویژه هستند که با PC_1 ، PC_2 ، ... و PC_k نمایش داده می‌شود. هر کدام PC های فوق متناظر با λ_1 ، λ_2 ، ... ، λ_k هستند که λها مقادیر ویژه هستند. یکی از ویژگی‌های مؤلفه‌های اصلی این است که متعامد هستند یعنی ضرب داخلی آنها صفر است، بنابراین در صورتی که به جای متغیرها از مؤلفه‌های اصلی استفاده شود هم‌خطی به طور کامل از بین خواهد رفت زیرا مؤلفه‌های اصلی هم‌خطی با یکدیگر ندارند و ضرب داخلی آنها صفر است در نتیجه کوواریانس بین آنها صفر بوده و با استفاده کردن آنها در رگرسیون، هم‌خطی به طور کامل از بین خواهد رفت. هر مؤلفه اصلی که با بزرگترین مقدار ویژه در تناظر باشد بالاترین قدرت توضیح دهندگی را خواهد داشت و مؤلفه‌های اصلی بعدی، قدرت توضیح دهندگی کمتری خواهند داشت. معمولاً در عمل از تمامی مؤلفه‌های اصلی استفاده نمی‌شود بلکه تعدادی از مؤلفه‌های اصلی که بخش عمده‌ای از واریانس را توضیح می‌دهند معمولاً مورد استفاده قرار می‌گیرند.

اشکال اصلی روش مؤلفه‌های اصلی تفسیر آنهاست زیرا مؤلفه‌های اصلی ترکیبی از متغیرهای اولیه هستند که تفسیر روشنی ندارند به عنوان مثال نرخ بیکاری، نرخ تورم، GDP و نرخ ارز در مؤلفه‌های اصلی با هم ترکیب می‌شوند. طبیعی است که ترکیب متغیرهای فوق تفسیر اقتصادی ندارند یعنی نمی‌توان گفت PC_1 که از متغیرهای فوق به دست آمده است نشان دهنده کدام یک از ویژگی‌های اقتصاد است و پایین بودن یا بالا بودن آن نشان دهنده موضوع یا ویژگی معینی نیست.

برای انجام این روش در نرم افزار E-Views 4، 5، 6 می‌توان سری‌ها را به عنوان group باز کرده و در قسمت View پنجره باز شده مؤلفه‌های اصلی را پیدا کرد.

برای انجام تحلیل مؤلفه‌های اصلی در نرم افزار E-Views ابتدا سری‌ها را انتخاب کرده سپس به عنوان یک Group باز می‌کنیم بعد قسمت View مؤلفه‌های اصلی را انتخاب نموده:

View	proc	Edit	window	Help
→ Principal component Analysis				

با انتخاب مؤلفه‌های اصلی صفحه‌ای به شکل ذیل ظاهر خواهد شد:

correlation ⊙ covariance○

series of principale comp

PC_1 PC_2 ...
یا
$Hamed_1$ $Hamed_2$ ...

cancel | OK

سپس از بدست آوردن متغیرهای اصلی رگرسیون را به صورت ذیل برازش می کنیم:

LS Y C PC_1 PC_2 ... PC_K

۲- استفاده از رگرسیون ریج (Ridge Regression):

وقتی که شکل همخطی بروز می کند ماتریس $(X'X)$ نزدیک به تکین (ماتریسی که دترمینان آن صفر است) می شود یعنی دترمینان ماتریس $X'X$ کوچک شده و معکوس آن با دقت کمی محاسبه می شود.

$$A^{-1} = \frac{adj}{|A|}$$

چون مخرج کسر کوچک و خود کسر بزرگ می شود.

برای رفع این شکل می توان ماتریس $X'X$ با یک ماتریس قطری rI جمع نمود که r یک اسکالر مثبت و I یک ماتریس یکه $K \times K$ است.

$$|X'X+rI|$$

که دیگر دترمینان ماتریس فوق نزدیک به صفر نخواهد بود:

$$\boxed{\hat{\beta}_r = (X'X+rI)^{-1}X'Y}$$

اگر چه شکل همخطی با تبدیل فوق حل می شود اما شکل دیگری به وجود خواهد آمد. تورش دار بودن تخمین زننده فوق بدیهی است و به سادگی می توان نشان داد:

$$\hat{\beta}_r = (X'X+rI)^{-1}X'(X\beta+\varepsilon) = (X'X+rI)^{-1}X'X\beta + (X'X+rI)^{-1}X'\varepsilon$$

$$E(\hat{\beta}_r) = (X'X+rI)^{-1}X'X\beta + (X'X+rI)^{-1}X'\overset{0}{E(\varepsilon)} = (X'X+rI)^{-1}X'X\beta + 0 \neq \beta$$

آزمون نرمال بودن:

آزمون های مختلفی برای نرمال بودن اجزای اخلال وجود دارد از جمله این آزمون ها، آزمون χ^2، آزمون کولموگروف- اسمیرنف (KS)، اندرسون- دارلینگ (AD)، لیلیفورز، جارک- برا (GB) و ... اشاره نمود در اینجا جهت رعایت اختصار فقط به آزمون جارک- برا اشاره می شود. این آزمون مبتنی بر این فرض است که اگر توزیع داده ها نرمال باشد باید ضریب چولگی صفر و ضریب کشیدگی ۳ باشد بنابراین آماره GB را می توان بدین صورت نمایش

$$GB = \frac{T}{6}\left[S^2 + \frac{(K-3)^2}{4}\right] \sim \chi^2_2$$

که در عبارت فوق S ضریب چولگی، K ضریب کشیدگی و T تعداد مشاهدات استفاده شده در تخمین رگرسیون است. بدیهی است هرچه انحراف از قرینگی یعنی چولگی بیشتر باشد آماره GB بزرگتر خواهد بود و هرچه ضریب کشیدگی بزرگتر از ۳ باشد یا کوچکتر از آن باشد GB بزرگتر خواهد بود. بنابراین GB بزرگ به مفهوم نرمال نبودن است. اما اینکه به چه مقداری بزرگ گفته می‌شود در داده‌های تصادفی نیاز به استفاده از جدول مرجع دارد. پس باید آماره GB با توزیع χ^2 با درجه آزادی ۲ و سطح احتمال معین مقایسه شود. اگر آماره GB در ناحیه بحرانی قرار گیرد فرضیه H_0 که بیان می‌کند توزیع داده‌ها نرمال است رد خواهد شد و در صورتی که در ناحیه بحرانی قرار نگیرد فرضیه H_0 را نمی‌توان رد کرد. برای انجام این آزمون در نرم افزار EViews کافی است پس از تخمین رگرسیون به View خروجی رفته و آزمون پسماندها را فعال نمود (Residual Tests). در بخش آزمون پسماندها یک آزمون به نام هیستوگرام - نرمالیتی تست وجود دارد با کلیک کردن بر روی این آزمون خروجی ظاهر شده آمار توصیفی مربوط به پسماندها را ارائه نموده و آماره GB نیز در ذیل خروجی داده خواهد شد. اگر Prob این آماره زیر ۵٪ باشد نرمال بودن در سطح ۵٪ رد می‌شود (GB بالا و Prob پایین نشان دهنده نرمال نبودن است).

آزمون نرمال بودن

اینجا چک برا محاسبه که توزیع نرمال نیست پس چطور باید معنی‌داری ضرایب را آزمون کرد

چون دیگر از آماره F و t نمی‌شه استفاده کرد، در توزیع‌ها تک‌مد اگر $t > 2.7$ براساس

تفسیر چبی‌شف ضریب معنی‌دار. اگر دومد براساس نامساوی چبیشف $t > 4.27$

Dependent Variable: PPI
Method: Least Squares
Date: 01/12/09 Time: 13:18
Sample: 1959:01 1989:12
Included observations: 372

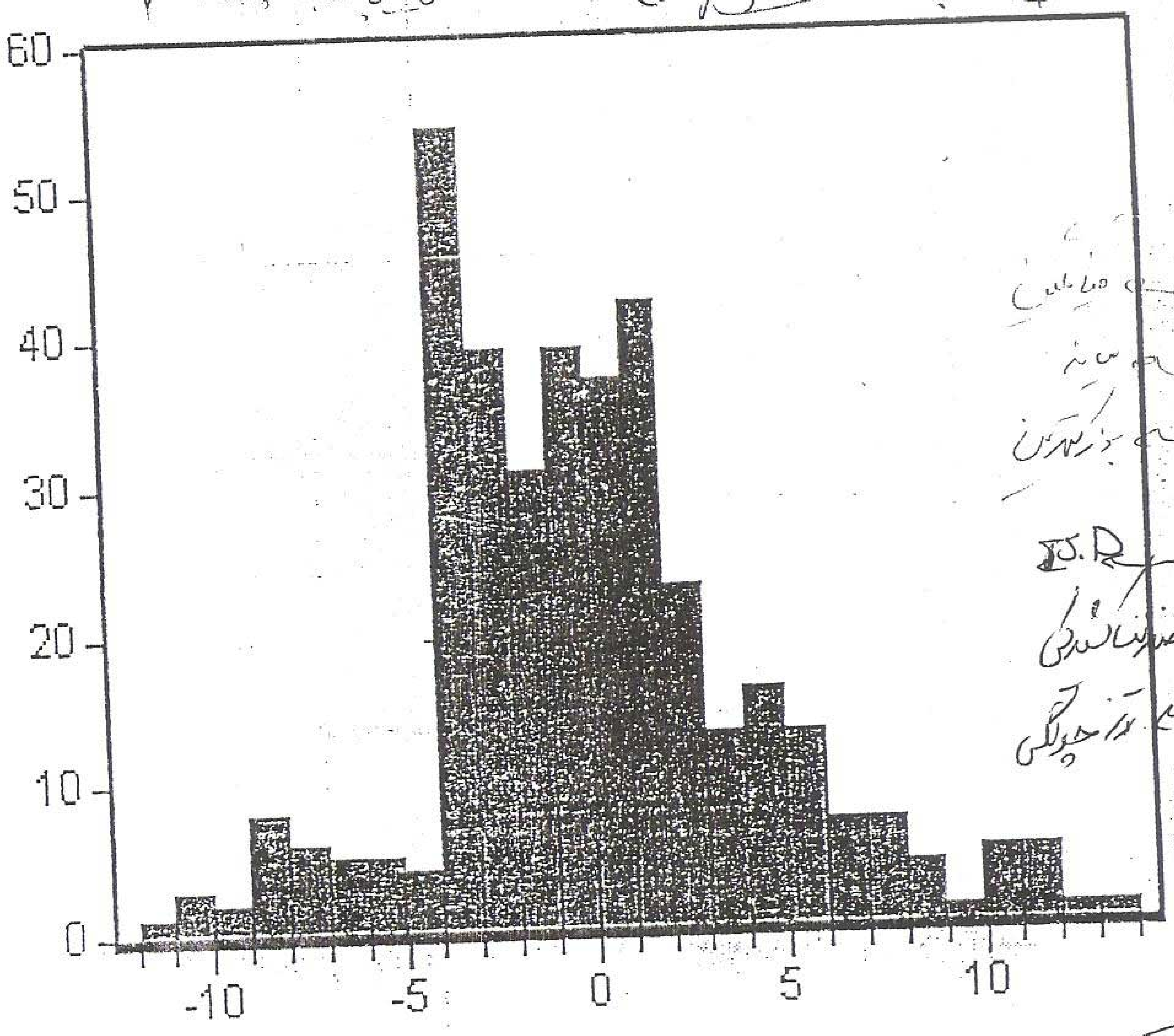
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	50.74954	1.399668	36.25826	0.0000
M1	-0.339703	0.012678	-26.79565	0.0000
M2	0.050723	0.009438	5.374295	0.0000
M3	0.043712	0.007441	5.874870	0.0000

صرف‌نظر از نرمال بودن

معنی‌دار

R-squared	0.979169	Mean dependent var	63.07016
Adjusted R-squared	0.978999	S.D. dependent var	29.64320
S.E. of regression	4.295820	Akaike info criterion	5.763856
Sum squared resid	6791.098	Schwarz criterion	5.805995
Log likelihood	-1068.077	F-statistic	5765.912
Durbin-Watson stat	0.028228	Prob(F-statistic)	0.000000

خودهمبستگی مثبت $2(1-\hat{\rho})$ نزدیک یک



Series: Residuals
Sample 1959:01 1989:12
Observations 372

Mean	2.55E-13
Median	-0.343086
Maximum	13.43333
Minimum	-11.04946
Std. Dev.	4.278416
Skewness	0.362195
Kurtosis	3.669003
Jarque-Bera	15.07074
Probability	0.000534

میانگین، میانه، بزرگترین، S.D، ضریب چولگی، ضریب کشیدگی

< 0.05

نرمال نیست

$H_0 =$ نرمال است

$H_1 =$ نرمال نیست

همانطور که مشخص است GB بالا و PROB پایین نشان دهنده نرمال نبودن است.

$$\frac{N}{6}\left(S^2 + \frac{(K-3)^2}{4}\right)$$

$$SD \rightarrow \sqrt{\frac{Y-\bar{Y}}{(T-1)}}$$

SE خطای استاندارد

سری‌های زمانی (Time series Analysis):

سری زمانی به داده‌هایی گفته می‌شود که طی زمان ثبت شده و یک دنباله زمانی را به وجود آورند:

$$\{y_t\}_{t=1}^{T} \quad \text{یا} \quad \{y_t\}_{t=-\infty}^{\infty} \qquad \{P_t\}_{t=60,\ldots,85}^{t=87,\ldots,88}$$

مدل‌های خودرگرسیو یا اتورگرسیو (Auto Regressive Model):

در این مدل‌ها متغیر وابسته سری زمانی $\{y_t\}$ و متغیرهای مستقل $y_{t-1}, y_{t-2}, \ldots, y_{t-p}$ هستند.

به عنوان مثال اگر فرض کنیم حقوق کارمندان دولت در هر سال ۱۳٪ افزایش می‌یابد می‌توان مدل اتورگرسیو برای حقوق کارمندان نوشت:

$$y_t = 1{,}13\, y_{t-1} + \varepsilon_t$$

مدل اتورگرسیو مرتبه اول (AR(1)):

$$y_t = \rho y_{t-1} + \varepsilon_t \qquad \varepsilon_t \sim N(0, \sigma^2_\varepsilon)$$

ممکن است متغیر به دو دوره قبلی خود وابسته باشد به عنوان مثال اگر یک اعتبار کلی در ۳ مرحله پرداخت شود میزان اعتبار در مرحله سوم به دو مرحله قبلی خود وابستگی پیدا می‌کند این که در مراحل قبلی چقدر پرداخت شده است با قیمانده را که در مرحله سوم باید پرداخت شود را مشخص کرده است.

$$\underset{\downarrow\ \text{مرحله سوم}}{y_t} = \rho\, \underset{\downarrow\ \text{وام مرحله اول}}{y_{t-1}} + \rho_2\, \underset{\downarrow\ \text{وام مرحله دوم}}{y_{t-2}} + \varepsilon_t \qquad AR(2)$$

البته مدل فوق نیز می‌تواند عرض از مبدأ داشته باشد و در حالت کلی یک متغیر می‌تواند به P مرحله قبلی خود وابستگی داشته باشد:

$$y_t = \alpha + \rho_1 y_{t-1} + \rho_2 y_{t-2} + \cdots + \rho_p y_{t-p} + \varepsilon_t = \alpha + \sum_{i=1}^{p} \rho_i y_{t-i} + \varepsilon_t$$

مدل‌های دیگری وجود دارد که به مدل‌های میانگین متحرک (Moving Average) مشهورند این مدل‌ها از وقفه‌های جزء اخلال برای پیش بینی متغیر وابسته استفاده می‌کنند به عنوان مثال:

$$y_t = \alpha + \theta_1 \varepsilon_{t-1} + \varepsilon_t$$

که یک فرآیند میانگین متحرک مرتبه اول MA(1) است.

$$y_t = \alpha + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \varepsilon_t \qquad MA(2)$$

$$y_t = \alpha + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t \qquad MA(q)$$

تخمین این مدل ها در نرم افزار Eviews بسیار ساده است حتی ساده تر از نیمرو درست کردن.

برای تخمین مدل اتورگرسیو در Eviews کافی است:

LS Y C AR(1) AR(2) ... AR(P)

و برای تخمین میانگین متحرک کافی است از دستور:

LS Y C MA(1) MA(2) ... MA(q)

البته می توان ۲ مدل فوق را با یکدیگر ترکیب نموده و به مدل ARMA(P, q) دست پیدا کرد.

$$y_t = \alpha + \sum_{i=1}^{p} \beta_i y_{t-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t$$

یک مدل ARMA(P, q) است.

به عنوان مثال اگر معتقد باشیم که شک قبلی بازار سهام بر قیمت اثرگذار است الا قیمت دیروز نیز بر قیمت امروز مؤثر است می توان از مدل ARMA(1,1) استفاده نمود.

$$y_t = \alpha + \beta_1 y_{t-1} + \theta_1 \varepsilon_{t-1} + \varepsilon_t$$

برای تخمین این مدل بسیار ساده:

LS y C AR(1) MA(1)

استفاده می شود.

مدل های سری زمانی در پیش بینی متغیرها بسیار قوی هستند بعضاً توان پیش بینی این متغیرها از مدل های سیستم رگرسیون با تعداد چندین معادله رگرسیونی مانند مدل DRI، S.T، مدل Warton و... بالاتر است. با توجه به قدرت بالای پیش بینی این متغیرها، استفاده از سری های زمانی به انتخاب سری های زمانی مشهور شده است.

LS ip C AR(1)

معیارهای انتخاب مدل در سری زمانی:

معیار آکائیک و شوارتز:

$$AIC = \ln \hat{\sigma}^2 + \frac{2K}{T}$$

$$SC = \ln \hat{\sigma}^2 + \frac{K \ln T}{T}$$

که در فرمول های فوق K، تعداد پارامترها و T تعداد مشاهدات است. هرچه آکائیک یا شوارتز پایین تر

...شد مدل مناسب تر است. بنابراین برای انتخاب مدل، باید مدلی را انتخاب کرد که شوارتز آن کوچک باشد

البته مدل آکائیک نیز می تواند برای انتخاب مدل استفاده شود اما معیار شوارتز مناسب تر است.

خودهمبستگی و خودهمبستگی جزئی:

اگر همبستگی متغیر را با خود آن در نظر بگیریم تابع خودهمبستگی یا همبستگی حاصل خواهد شد به عنوان مثال همبستگی Y_t و Y_{t-1}

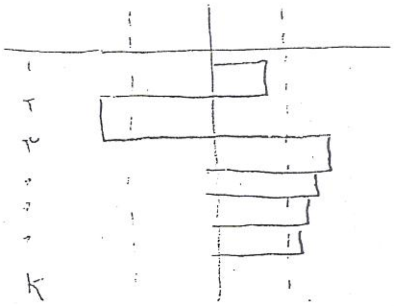
خودهمبستگی مرتبه 1 نامیده می شود:

Auto correlation function (ACF):

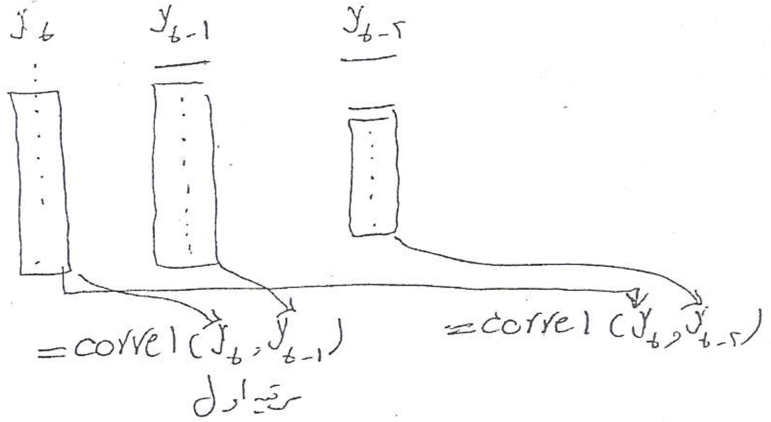
$$\rho_1 = \frac{\sum (y_t - \bar{y})(y_{t-1} - \bar{y})/T}{\sqrt{\frac{\sum (y_t - \bar{y})^2}{T}}\sqrt{\frac{\sum (y_{t-1} - \bar{y})^2}{T}}}$$

Auto correlation (K):

$$\rho_k = \frac{\sum (y_t - \bar{y})(y_{t-k} - \bar{y})/\cancel{T}}{\sqrt{\frac{\sum (y_t - \bar{y})^2}{\cancel{T}}}\sqrt{\frac{\sum (y_{t-k} - \bar{y})^2}{\cancel{T}}}} = \frac{\sum (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum (y_t - \bar{y})^2} = \frac{cov(y_t, y_{t-k})}{var(y_t)}$$



خودهمبستگی در Excel:



خودهمبستگی در EViews:

ابتدای سری داده را باز کرده و در View آن کلیک کرده سپس گزینه correlogram ← level ← lake

وقتی Prob = 0 است ضریب همبستگی معنی دار است.

JPM

MA / AR	0	1	2
0	6.041872	5.203589	5.391034
1	4.189703	4.203475	4.207158
2	4.720495	4.202263	4.720544

بهترین مدل ، مدلی است که SC کوچکتری دارد یعنی ARMA(1,0)
حال داده ها را از ۲۳۰-۲۵۲ پیش بینی می کنیم

شوارتز

Dependent Variable: P
Method: Least Squares
Date: 12/31/08 Time: 05:22
Sample(adjusted): 2 252
Included observations: 251 after adjusting endpoints
Convergence achieved after 3 iterations

Variable	Coefficient	Std. Error (SE)	t-Statistic (t آماره)	Prob.
C	39.76444	1.430644	27.79478	0.0000
AR(1)	0.914433	0.024783	36.89730	0.0000

R-squared	0.845381	Mean dependent var	39.33697
Adjusted R-squared	0.844760	S.D. dependent var	4.900274
S.E. of regression	1.930732	Akaike info criterion	4.161612
Sum squared resid	928.2038	Schwarz criterion	4.189703
Log likelihood	-520.2823	F-statistic	1361.411
Durbin-Watson stat	2.146523	Prob(F-statistic)	0.000000
Inverted AR Roots	.91		

حال پیش بینی را با مقدار واقعی پیش بینی می کنیم

P	PF	درصد خطا
44.99000	42.47788	0.055837
46.97000	42.25578	0.100367
46.14000	42.05205	0.088599
46.09000	41.86517	0.091665
46.19000	41.69374	0.097343
44.36000	41.53649	0.06365
42.48000	41.39224	0.025606
43.76000	41.25993	0.057131
44.50000	41.13855	0.075538
39.77000	41.02721	-0.03161
38.54000	40.92508	-0.06189
38.98000	40.83140	-0.0475
40.33000	40.74546	-0.0103
38.13000	40.66663	-0.06653
40.26000	40.59432	-0.0083
39.77000	40.52799	-0.01906
40.23000	40.46714	-0.00589
39.19000	40.41133	-0.03116
38.64000	40.36013	-0.04452
40.24000	40.31316	-0.00182
39.84000	40.27008	-0.0108
40.77000	40.23057	0.013231
41.05000	40.19432	0.020845

ایران خودرو

AR \ MA	0	1	2
0	13.45964	12.46077	12.52282
1	10.83802	10.83638	10.86908
2	11.30888	10.87047	11.32748

بهترین مدل ،ARMA(1,1)است چون SCکوچکتری دارد.
حال مدل را از ۱۴۰-۱۶۵ پیش بینی می کنیم.

مدل ARMA برای بررسی همبستگی

H0 : همبستگی ندارند ≥ ۰.۰۵
H1 : همبستگی دارند < ۰.۰۵

Dependent Variable: P
Method: Least Squares
Date: 12/31/08 Time: 05:35
Sample(adjusted): 2 165
Included observations: 164 after adjusting endpoints
Convergence achieved after 6 iterations
Backcast: 1

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1323.661	123.8430	10.68821	0.0000
AR(1)	0.970771	0.017351	55.94813	0.0000
MA(1)	-0.189926	0.080026	-2.373310	0.0188

H0 رد
همبستگی دارند

R-squared	0.931318	Mean dependent var	1403.476
Adjusted R-squared	0.930465	S.D. dependent var	199.2981
S.E. of regression	52.55380	Akaike info criterion	10.77968
Sum squared resid	444666.2	Schwarz criterion	10.83638
Log likelihood	-880.9334	F-statistic	1091.576
Durbin-Watson stat	1.997354	Prob(F-statistic)	0.000000

Inverted AR Roots	.97
Inverted MA Roots	.19

P	PF	درصد خطا
1330.160	1330.000	0.00012
1330.285	1290.000	0.030283
1330.406	1290.000	0.030371
1330.524	1290.000	0.030457
1330.638	1290.000	0.03054
1330.749	1300.000	0.023107
1330.857	1300.000	0.023186
1330.961	1300.000	0.023262
1331.062	1300.000	0.023336
1331.160	1300.000	0.023408
1331.255	1300.000	0.023478
1331.348	1300.000	0.023546
1331.438	1300.000	0.023612
1331.524	1300.000	0.023675
1331.609	1300.000	0.023737
1331.691	1300.000	0.023798
1331.770	1300.000	0.023855
1331.847	1270.000	0.046437
1331.922	1300.000	0.023967
1331.995	1310.000	0.016513
1332.065	1310.000	0.016565
1332.133	1310.000	0.016615
1332.200	1310.000	0.016664
1332.264	1310.000	0.016711
1332.326	1310.000	0.016757
1332.387	1310.000	0.016802

ایران خودرو

AR \ MA	0	1	2
0	13.45964	12.46077	12.52282
1	10.83802	10.83638	10.86908
2	11.30888	10.87047	11.32748

بهترین مدل، ARMA(1,1)است چون SCکوچکتری دارد.
حال مدل را از ۱۴۰-۱۶۵ پیش بینی می کنیم.

Dependent Variable: P
Method: Least Squares
Date: 12/31/08 Time: 05:35
Sample(adjusted): 2 165
Included observations: 164 after adjusting endpoints
Convergence achieved after 6 iterations
Backcast: 1

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1323.661	123.8430	10.68821	0.0000
AR(1)	0.970771	0.017351	55.94813	0.0000
MA(1)	-0.189926	0.080026	-2.373310	0.0188

R-squared	0.931318	Mean dependent var	1403.476
Adjusted R-squared	0.930465	S.D. dependent var	199.2981
S.E. of regression	52.55380	Akaike info criterion	10.77968
Sum squared resid	444666.2	Schwarz criterion	10.83638
Log likelihood	-880.9334	F-statistic	1091.576
Durbin-Watson stat	1.997354	Prob(F-statistic)	0.000000

Inverted AR Roots	.97
Inverted MA Roots	.19

P	PF	درصد خطا
1330.160	1330.000	0.00012
1330.285	1290.000	0.030283
1330.406	1290.000	0.030371
1330.524	1290.000	0.030457
1330.638	1290.000	0.03054
1330.749	1300.000	0.023107
1330.857	1300.000	0.023186
1330.961	1300.000	0.023262
1331.062	1300.000	0.023336
1331.160	1300.000	0.023408
1331.255	1300.000	0.023478
1331.348	1300.000	0.023546
1331.438	1300.000	0.023612
1331.524	1300.000	0.023675
1331.609	1300.000	0.023737
1331.691	1300.000	0.023798
1331.770	1300.000	0.023855
1331.847	1270.000	0.046437
1331.922	1300.000	0.023967
1331.995	1310.000	0.016513
1332.065	1310.000	0.016565
1332.133	1310.000	0.016615
1332.200	1310.000	0.016664
1332.264	1310.000	0.016711
1332.326	1310.000	0.016757
1332.387	1310.000	0.016802

AR(1) AR(2) M(1) M(2)

$Rge_t \sim Rge_{t-1}$

Rge_{t-2} ε_{t-1} ε_{t-2}

Prob 0.0220

AR(1)

AR(1)
AR(2)
AR(3)

خود همبستگی جزئی (Partial Auto correlation function):

PACF

$$y_t - \mu = c + \rho_1^* (y_{t-1} - \mu) + \varepsilon_t$$

ضریب همبستگی جزئی مرتبه یک

$$y_t - \mu = c + \rho_1^* (y_{t-1} - \mu) + \rho_2^* (y_{t-2} - \mu) + \varepsilon_t$$

ضریب همبستگی جزئی مرتبه دوم

$\vdots$

$$y_t - \mu = c + \rho_1^* (y_{t-1} - \mu) + \dots + \rho_k^* (y_{t-k} - \mu) + \varepsilon_t$$

ضریب همبستگی جزئی مرتبه kام

ضریب همبستگی با لحاظ کردن سایر وقفه‌ها محاسبه شده است یعنی وقتی مرتبه دوم را می‌گیریم اثرات مرتبه اول از آن جدا شده است.

اگر یک سری نامانا باشد (ریشه واحد دارد یعنی با lagهای خودش همبستگی شدید دارد) $\rho = 1 \Rightarrow$ نامانا است

مانایی متغیر (stationarity)

$|\rho| < 1$

* اگر سری زمانی $\{y_t\}$ مانا باشد لازم است گشتاورهای مراتب مختلف آن مستقل از زمان (ثابت) باشند و هنگامی که صحبت از گشتاوری مراتب مختلف می‌شود منظور این است امید ریاضی یا میانگین، گشتاور مرتبه اول، واریانس گشتاور مراتب دوم و کوواریانس‌ها هم همین گشتاور مرتبه دوم، صورت ضریب چولگی گشتاور مرتبه سوم، صورت ضریب کشیدگی گشتاور مرتبه چهارم و ... است.

گشتاورهای اولیه:

$$m_k = \frac{\sum_{i=1}^{n} x_i^k}{n}$$ میانگین $m_1 = \mu$

$$\mu_k = \frac{\sum (x_i - \mu)^k}{n}$$

$$\mu_2 = \frac{\sum (x_i - \mu)^2}{n}$$

$$\mu_3 = \frac{\sum (x_i - \mu)^3}{n}$$

$$\mu_4 = \frac{\sum (x_i - \mu)^4}{n}$$

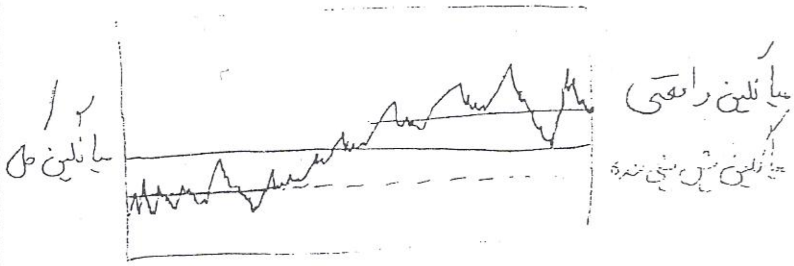
$\vdots$

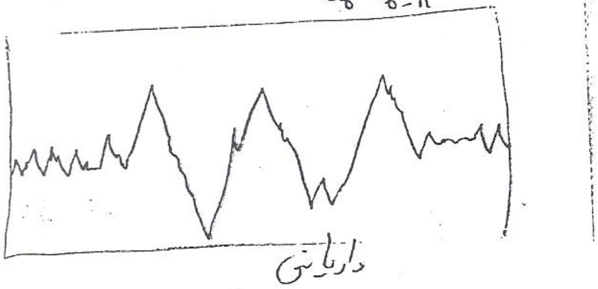
مانایی: ~~قطعی~~ مؤکد قوی / ضعیف

... یک متغیر تصادفی ثابت و مستقل از زمان باشد آن متغیر تصادفی مانای قوی (strictly stationary) یا مانای مؤکد (strong stationary) نامیده می‌شود. شکل دیگری از مانایی که تحت عنوان مانایی ضعیف نامیده می‌شود مستلزم این است که گشتاور مرتبه اول و دوم مستقل از زمان باشند یعنی اگر سری زمانی $\{y_t\}$ یک سری مانای ضعیف باشد (weakely stationary covariance stationary) باشد. لازم است

$$1.\ E(y_t) = \mu$$
$$2.\ cov(y_t, y_{t-k}) = \gamma_k$$

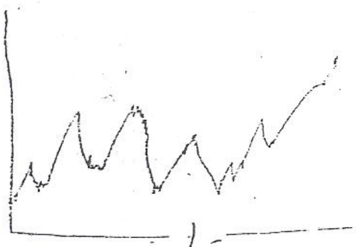
مستقل از زمان باشند.





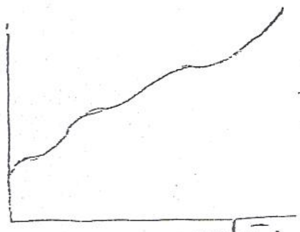
مانایی در آزمون آن :

$E(x_t) = \mu$ $\quad cov(x_t, x_{t-s}) = \gamma_s$ $\quad var(x_t) = \sigma^2$



روند تصادفی

Stochastic trend



روند قطعی

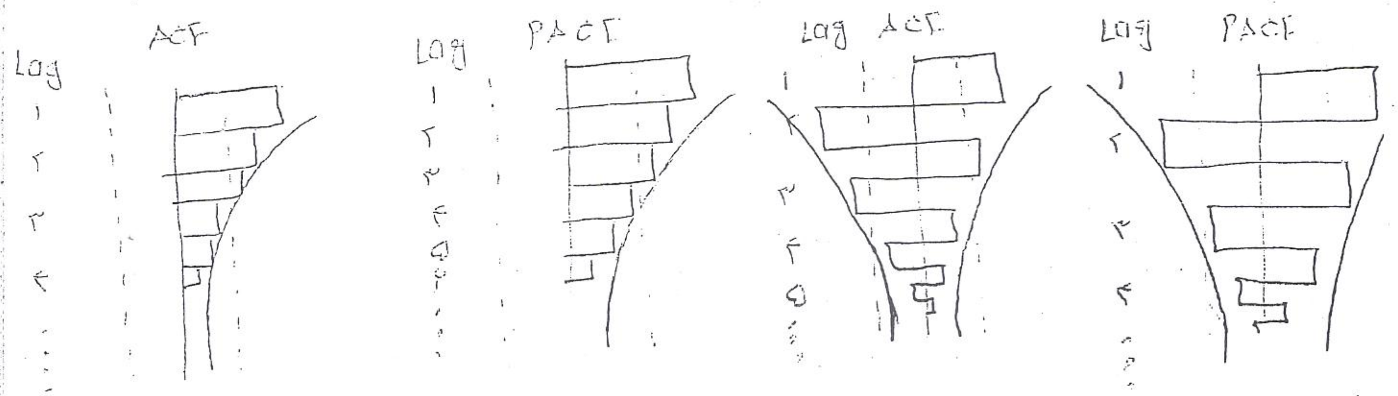
Deterministic trend

برای انجام آزمون مانایی :

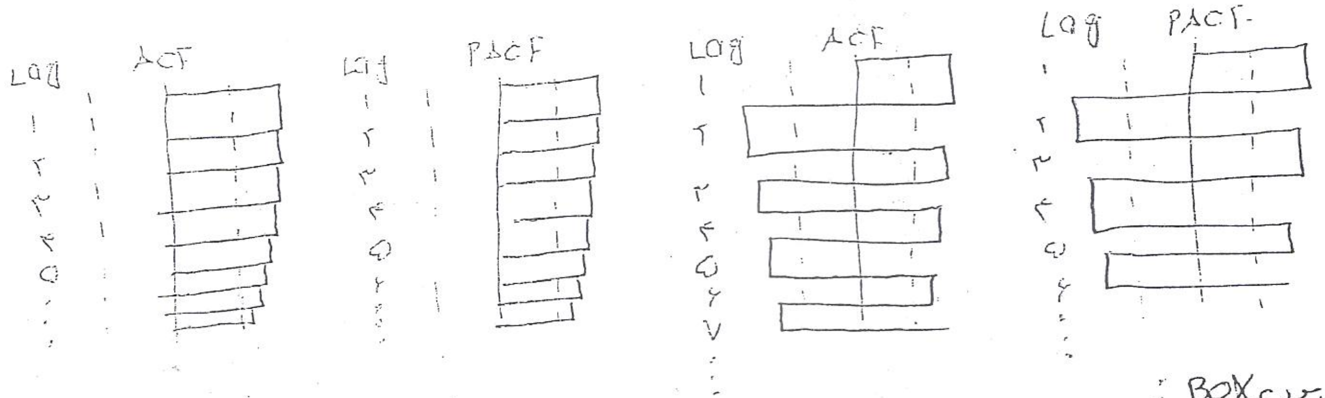
۱- می‌توان از توابع خودهمبستگی (ACF) و خودهمبستگی جزئی استفاده کرد.

۲- می‌توان از آزمون‌های ریشه واحد استفاده کرد (Unit Root Tests).

برای استفاده از ACF و PACF جهت انجام آزمون مانایی، نمودار خودهمبستگی جزئی و نمودار خودهمبستگی برای وقفه‌های مختلف ترسیم شده و رفتار آنها بررسی می‌شود اگر خودهمبستگی جزئی و خودهمبستگی هر دو به صورت نمایی کاهش یابند فرآیند مانا است یعنی :



اگر فرآیند نامانا باشد نمودار خودهمبستگی جزئی و نمودار خود همبستگی روند کاهشی سریعی نخواهند داشت یا اصلاً کاهش پیدا نخواهند کرد.



آزمون BOX :

می‌توان مجذور ضرایب همبستگی برای وقفه‌های مختلف را جمع زده $Q = T\sum_{i=1}^{k}\rho_i^2 \sim \chi^2_k$ و در جدول وقفه مربوطه نوشت البته ضریب T که تعداد مشاهدات است بر آن ضرب خواهد شد (χ^2 در اینجا تعداد مرتبه وقفه است).

درصورتی که Q در ناحیه بحرانی قرار بگیرد فرآیند نامانا خواهد بود، با توجه به اینکه دقت آزمون فوق پایین است معمولاً در اقتصاد سنجی از آزمون‌های ریشه واحد (Unit Root Test) استفاده می‌شود.

آزمون ریشه واحد (Unit Root Test):

می‌توان نشان داد فرآیندهایی که نامانا هستند ریشه واحد دارند یعنی اگر یک فرآیند خود رگرسیو مرتبه اول را در نظر بگیریم می‌توان آن را به صورت یک معادله تفاضلی مرتبه اول نوشت

عملگر وقفه (Lag oprator):

$$y_{t-1} = L y_t$$
$$y_{t-2} = L^2 y_t$$
$$\vdots$$
$$y_{t-k} = L^k y_t$$

$$y_t = \rho y_{t-1} + \varepsilon_t \rightarrow y_t - \rho y_{t-1} = \varepsilon_t \rightarrow (1-\rho L)\, y_t = \varepsilon_t$$

$$1 - \rho L = 0 \rightarrow L = \frac{1}{\rho}$$ ریشه معادله تفاضلی فوق

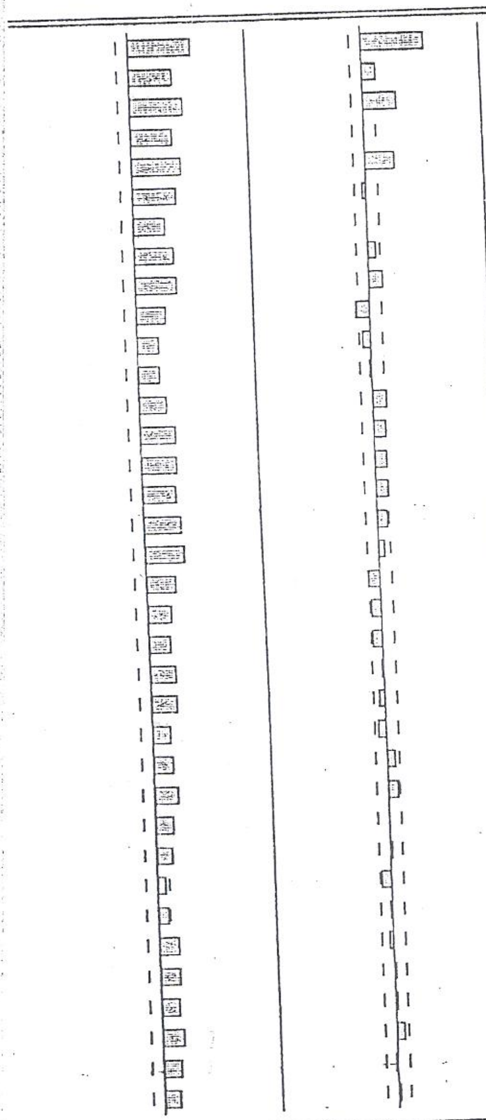
Date: 01/14/09 Time: 11:22
Sample: 1959M01 1989M12
Included observations: 372

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.991	0.991	368.59	0.000
		2	0.983	-0.003	731.82	0.000
		3	0.974	-0.010	1089.6	0.000
		4	0.966	-0.004	1442.1	0.000
		5	0.957	-0.007	1789.2	0.000
		6	0.948	-0.012	2130.8	0.000
		7	0.939	-0.002	2467.0	0.000
		8	0.930	-0.014	2797.8	0.000
		9	0.921	-0.023	3122.8	0.000
		10	0.911	-0.018	3442.1	0.000
		11	0.902	-0.014	3755.5	0.000
		12	0.892	-0.015	4063.0	0.000
		13	0.882	-0.013	4364.4	0.000
		14	0.872	-0.009	4659.9	0.000
		15	0.862	-0.012	4949.3	0.000
		16	0.851	-0.011	5232.6	0.000
		17	0.841	-0.012	5509.9	0.000
		18	0.831	-0.012	5781.0	0.000
		19	0.820	-0.005	6046.0	0.000
		20	0.810	0.001	6305.1	0.000
		21	0.799	-0.006	6558.3	0.000
		22	0.789	0.001	6805.7	0.000
		23	0.779	-0.005	7047.4	0.000
		24	0.768	-0.011	7283.3	0.000
		25	0.758	0.003	7513.6	0.000
		26	0.747	-0.017	7738.2	0.000
		27	0.737	-0.018	7957.0	0.000
		28	0.726	0.004	8170.3	0.000
		29	0.716	-0.008	8378.0	0.000
		30	0.705	-0.006	8580.2	0.000
		31	0.694	-0.011	8776.9	0.000
		32	0.683	-0.019	8968.1	0.000
		33	0.673	-0.007	9153.8	0.000
		34	0.662	0.006	9334.1	0.000
		35	0.651	-0.007	9509.2	0.000
		36	0.641	-0.011	9679.1	0.000

همانطور که مشاهده می شود ACFو PACF به طور نمایی کاهش نیافتند لذا سری M1 مانا نیست.
حال باتبدیل این سزی به یک سری مانا روند تغییرات ACFو PACF را مشاهده می کنیم .

خود همبستگی

Date: 01/14/09 Time: 11:31
Sample: 1959M01 1989M12
Included observations: 371



Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.552	0.552	114.11	0.000
		2	0.386	0.116	169.96	0.000
		3	0.459	0.302	249.28	0.000
		4	0.368	0.008	300.27	0.000
		5	0.439	0.269	373.03	0.000
		6	0.378	-0.041	427.23	0.000
		7	0.277	0.005	456.50	0.000
		8	0.341	0.074	500.78	0.000
		9	0.372	0.125	553.77	0.000
		10	0.255	-0.122	578.71	0.000
		11	0.194	-0.064	593.15	0.000
		12	0.197	-0.016	608.05	0.000
		13	0.250	0.119	632.17	0.000
		14	0.318	0.108	671.39	0.000
		15	0.307	0.112	708.07	0.000
		16	0.301	0.100	743.32	0.000
		17	0.332	0.087	786.49	0.000
		18	0.349	0.056	834.15	0.000
		19	0.265	-0.099	861.81	0.000
		20	0.200	-0.091	877.57	0.000
		21	0.196	-0.086	892.76	0.000
		22	0.224	-0.014	912.66	0.000
		23	0.229	-0.046	933.44	0.000
		24	0.158	-0.058	943.38	0.000
		25	0.165	0.076	954.32	0.000
		26	0.201	0.096	970.55	0.000
		27	0.158	0.006	980.58	0.000
		28	0.132	0.021	987.64	0.000
		29	0.069	-0.083	989.55	0.000
		30	0.084	-0.014	992.44	0.000
		31	0.166	-0.026	1003.7	0.000
		32	0.175	0.016	1016.2	0.000
		33	0.162	0.014	1026.9	0.000
		34	0.181	0.070	1040.4	0.000
		35	-0.161	-0.010	1051.1	0.000
		36	0.145	0.009	1059.8	0.000

همانطور که مشاهده می شود روند کاهشی ACFوPACFنمایی است یعنی سری مانا است.

اگر $|\rho|<1$ باشد یعنی معادله فوق را حد محدود بود ($|\rho|<1$).

امکان وجود ریشه واحد

حال ببینیم واحد بودن ریشه چه اثری بر واریانس ها کوواریانس ها خواهد داشت و به چه صورت ناپایایی را شدت تأثیر خواهد داد. به خاطر داریم که برای فرآیند های خود رگرسیون مرتبه اول داشتیم:

$$var(y_t) = \frac{\sigma^2_\varepsilon}{1-\rho^2} , \quad |\rho|<1 \qquad cov(y_t , y_{t-s}) = \frac{\rho^{|s|}\sigma^2_\varepsilon}{1-\rho^2} , \quad |\rho|<1$$

بدیهی است که در صورتی که فرآیند، ریشه واحد داشته باشد یعنی $|\rho|=1$ باشد شرط فوق برآورده نخواهد شد بنابراین فرمول های واریانس و کوواریانس عبارات فوق نیست بلکه عباراتی که تابعی از زمان هستند برای واریانس و کوواریانس به دست خواهد آمد در نتیجه داشتن ریشه واحد برای یک فرآیند به مفهوم ناپایایی است و آزمون ریشه واحد به دنبال آزمون ناپایایی سری ها از این طریق است.

دیکی و فولر DF:

$$y_t = \alpha + \beta y_{t-1} + \varepsilon_t$$

برای بررسی ریشه واحد می توان رگرسیون فوق را برآورد نموده فرضیه $\beta=1$ (ناپایایی) را در مقابل $|\beta|<1$ (پایایی) آزمون نمود اما باید توجه نمود برای آزمون فوق نمی توان از جدول t استفاده نمود بلکه در حالتی که ریشه واحد وجود داشته باشد آماره τ برابر است با:

$$\tau = \frac{\hat\beta}{SE(\hat\beta)}$$

از جدول τ تبعیت خواهد کرد. مقادیر بحرانی این جدول به وسیله دیکی و فولر محاسبه شده است در صورتی که آماره محاسبه شده یعنی $\frac{\hat\beta}{SE(\hat\beta)}$ در رگرسیون فوق در سطوح معنی داری مورد نظر، منفی تر از مقدار جدول باشد فرآیند پایا است و در غیر اینصورت نیست.

به عنوان مثال: اگر رابطه بین شاخص سهام در بورس تهران برای یک شرکت خودروسازی با یک وقفه آن به صورت ذیل تخمین زده شده باشد می توان آماره فوق را استاندارد نموده و با جدول τ مقایسه نمود البته در عمل به جای رگرسیون فوق از رگرسیون $\Delta y_t = \alpha + (\beta-1)y_{t-1} + \varepsilon_t$ استفاده می شود.

$$y_t = 1200 + 0.98\, y_{t-1}$$
$$SE \quad 800 \qquad (0.09)$$

چون توزیع τ به صورت تجربی شبیه سازی می شود که توزیع چوله به چپ است باید مقدار به دست آمده از رگرسیون $\Delta y_t = \alpha + (\beta-1)y_{t-1} + \varepsilon_t$ درست چپ قرار گیرد تا سری ناپایا نباشد. دلیل منفی تر

بودن ، آماره محاسبه شده برای مانایی همین است آماره ADF هرچه منفی تر باشد فرآیند با احتمال بیشتری مانا است .

ADF شکل کامل‌تری از DF است و علامت اختصاری Augmented Dicky Fuller که علت نامگذاری افزوده در این شکل نوشته می‌شود :

$$\Delta y_t = \alpha + \overbrace{(1-\beta)}^{\gamma} y_{t-1} + \overbrace{\sum_{j=1}^{p} \theta_j \Delta y_{t-j}}^{\text{افزوده}} + \varepsilon_t$$

علت آوردن جملات افزوده در رگرسیون نوع کاهش همبستگی پیاپی است حال در رگرسیون نوع اگر $\frac{\hat{\gamma}}{SE(\hat{\gamma})}$ کوچکتر از مقادیر بحرانی ADF باشد سری مانا است :

$t = \frac{\hat{\gamma}}{SE(\hat{\gamma})} <$ ADF مقادیر بحرانی $\Rightarrow$ سری مانا است

برای انجام این آزمون در نرم افزار Eviews می‌توان به صورت زیر عمل کرد :

بدین صورت که بر سری مورد نظر (مثلاً y_t) دو بار کلیک کرده و آن را باز می‌کنیم در قسمت view سری unit Root test را انتخاب کرده و از میان آزمون‌های مختلف ADF را انتخاب می‌کنیم ،

گزینه بدون عرض از مبدأ و روند، عرض از مبدأ و عرض از مبدأ و روند وجود دارد گزینه عرض از مبدأ و روند را انتخاب نموده و OK می‌کنیم اگر آماره به دست آمده یعنی ADF محاسبه شده منفی تر از مقادیر بحرانی باشد سری مانا است.

① Trend entres : عرض از مبدأ و روند

اگر سری مانا نبود به صورت زیر عمل می‌کنیم :

series dy = y - y(-1)

حال سری dy را باز کرده و عملیات فوق را بر روی آن انجام می‌دهیم این سری مانا است .

$Dy_t = y_t - y_{t-1}$

حال باید ARMA را پیش بینی کرد .

اگر روی Dy ، ARMA را پیش بینی کنیم ARIMA نامیده می‌شود .

y_t	Dy_t
۱۲	—
۱۳	۱
۱۵	۲
۱۸	۳
۲۰	۲
۲۲	۲

آزمون مانايي M1

ADF Test Statistic	-0.319391	1% Critical Value*	-3.9869
		5% Critical Value	-3.4237
		10% Critical Value	-3.1345

*MacKinnon critical values for rejection of hypothesis of a unit root.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(M1)
Method: Least Squares
Date: 01/12/09 Time: 13:39
Sample(adjusted): 1959:06 1989:12
Included observations: 367 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
M1(-1)	-0.000472	0.001477	-0.319391	0.7496
D(M1(-1))	0.417843	0.052912	7.896925	0.0000
D(M1(-2))	-0.080678	0.055778	-1.446412	0.1489
D(M1(-3))	0.277781	0.055953	4.964520	0.0000
D(M1(-4))	-0.043162	0.053308	-0.809677	0.4187
C	-0.157461	0.212931	-0.739492	0.4601
@TREND(1959:01)	0.005804	0.002784	2.084536	0.0378

R-squared	0.409087	Mean dependent var	1.780654
Adjusted R-squared	0.399238	S.D. dependent var	2.556742
S.E. of regression	1.981700	Akaike info criterion	4.224677
Sum squared resid	1413.769	Schwarz criterion	4.299166
Log likelihood	-768.2281	F-statistic	41.53777
Durbin-Watson stat	1.968894	Prob(F-statistic)	0.000000

چون ADFکمتر از مقدار بحراني نيست لذا سري مانا نيست براي مانا شدن-اقدامات زير را انجام مي دهيم:

SERIES DM1=M1-M1(-1)

حال داريم: