**ECO 214 学习笔记**

**Lecture 1**

**1. 多重回归模型**



其中：Y为因变量

为自变量

为误差项

OLS估计量就是使实际真值和基于估计线的预测之间的平均平方差最小化

最小二乘假设：

1. 给定x时，误差项的条件均值为0 -> 误差项和自变量之间没有关系
2. 独立同分布
3. 大的异常值几乎不可能
4. 不存在完全多重共线性

**2. 同方差和异方差**

同方差：误差项的分布不随x的变化而变化，即为一个常数

异方差：误差项的分布随x的变化而变化，即为关于x的一个公式

**3. Hypothesis testing (假设检验)**

在大样本情况下，t统计量服从正态分布，F统计量服从分布

在有限的样本情况下，如果误差项时同方差且服从正态分布，则t统计量服从分布，F统计量服从分布 (n为样本数量，k为非零自变量)

假设检验的方法和ECO205中相同

如果P值小于5%，则拒绝原假设

**4. 面板数据**

* 面板数据集包含了对多个因素（个人、州、公司等）的观察结果，在两个或多个时间点观察每个实体 （n个不同个体在T个不同时间上的观测数据）

-> 在不同时间节点的实体必须相同

举例：420个加利福尼亚学校在1999年和2000年的数据

美国48个州7年的数据

* 符号

当描述截面数据时，用下标表示个体，例如表示第i个个体的变量Y

当描述面板数据时，我们需要其他符号来同时表达个体和时期

：代表实体 (entity), n为实体的数量 -> 

：代表时间段，T为时间段的数量 -> 

如果我们只有一个自变量，则数据为其中，

如果有k个自变量，则数据为其中，

如果为平衡面板（balanced panel），就是没有遗漏的观测量，即变量在每个个体和每一时期中都能被观测到；如果至少有一个个体在至少一个时期中的数据没有被观测到，则称为非平衡面板（unbalanced panel）

* 用处

可以用来控制下面这种特殊的遗漏变量：

1. 随着实体的变化而变化，但不随时间改变
2. 通常无法观测或衡量 -> 不可以被包含在多重回归中

如果某个遗漏变量不随时间改变，则任何因变量的改变都不是由这个遗漏变量所造成的

Fixed effect模型会自动控制这种遗漏变量

这些被称为idiosyncraises

**Lecture 2**

**1. 面板数据**

* 举例：

实体：美国48个州

时间：7年

如果为balanced panel，则总共有336个观测值

因变量：交通死亡率

自变量：酒精税

通过回归所有的数据我们发现这两者成正相关，这和我们的推断不符

有其他的因素可能会影响交通死亡率，例如：汽车的质量、道路的质量、关于喝酒和开车的文化、道路上的车辆密度等，这些因素都是idiosyncraises

遗漏这些变量会造成遗漏变量偏差

但因为这些因素在给定某个州内不会随时间变化，那么我们可以采用固定效应OLS回归来解决问题

当得到每个州T=2个时期上的数据后，就可以比较第一个时期和第二个时期的因变量取值，这种比较通过关注因变量的变化而实际上固定了随个体变化但不随时间变化的不可测因素

* 两种方向使用面板数据：

1. Pooled regression：汇集不同年份的观测数据以增加样本量
2. Differenced regression：取起点和终点之间的差值去消除从不随时间变化的变量中的遗漏变量偏差

**2. 估计方法**

* n-1 binary regressors regression model



其中

举例：D2=1 如果state = 2

回归方程中不能包含所有n个二值变量和一个公共截距，因为这样会存在多重共线性

所以一般省略第一组二值变量

方法：

1. 创造二元变量
2. 使用OLS估计模型
3. 假设检验、置信区间（和ECO205中所学相同）

当样本数量很大时，可能不可行

* fixed effects regression model

固定效应模型有n个不同的截距，其中一个截距对应一个个体，可以用一系列二值或指示变量来表示这些截距，它们包含了所有在个体间不同但不随时间变化的遗漏变量的影响



其中被称为 state fixed effect

总体回归线的斜率系数对所有个体都是相同的，但总体回归线的截距对每个个体有不同的取值，个体固定效应的差异来源于随个体变化但随时间不变的遗漏变量

* entity-demanded OLS regression

其中entity average满足

则从上面的数据的偏差为

或

方法：

首先建立entity-demanded 变量，即和

使用OLS回归和

在交通事故死亡人数中的应用：

从7年的所有数据中得到的死亡率和啤酒税的固定效应回归线的OLS估计：



这个式子中的回归系数估计值为负数，和我们预测的一样，较高的啤酒税会减少交通死亡人数

回归中包含的州固定效应让我们避免了那些随州变化但不随时间变化的遗漏变量的影响

* time fixed effects

和个体固定效应能控制不随时间变化但个体间不同的变量一样，时间固定效应能控制个体间相同但随时间变化的变量

指遗漏变量会随时间而改变但不会随实体而变化

举例：安全气囊、国家法律的变化

（估计步骤和前面相同）

所有的实体有相同的截距，但截距可能会随时间的改变而改变

或

* 个体和时间固定效应

Entity and time fixed effects regression model为



为 个体固定效应，为时间固定效应

* changes specification

如果T=2,则三种表述方法是相同的

如果T>2,则前两种相同，但和第三种不同

**Lecture 3**

**1. 面板数据的LS假说**

* 误差项的条件均值为0

给定自变量x的历史数据和entity fixed effect，误差项的均值为0

-> 不存在遗漏落后因素

-> 误差项的现值和自变量x的现值和历史值不相关

误差项的现值也和x的未来值不相关

* 为独立同分布

单个个体的变量分布相同，且与另一个个体的变量独立

这个条件当实体为随机抽样时自动满足

但比不要求相同实体不同时间的观测值为独立同分布

举例：一个州今年的税收高一般暗示明年的税收也高

When we talk about observations across the entity, we assume they are i.i.d.

But we don’t assume that observations belonging to the same entity are also i.i.d.

* 大的异常值不太可能
* 不存在完美的多重共线性

当n充分大时，固定效应估计量是一致的并服从正态分布

**2. Autocorrelation**

在不同的时间节点，和它自己相关

即对于一些

则被称为autocorrelation of 

举例：交通事故死亡率的例子中，第t年第i个州的啤酒税存在自相关-> 多数情况下州议会不太会改变啤酒税，如果今年的税收相对于其平均值而言较高，则明年往往也会较高

误差项也存在自相关的可能性

**3. Clustered standard error**

之前在ECO205中使用的homoskedsticity-only 和 heteroskedasticity-robust 标准误都不合适，因为它们没有考虑到误差项是相关的

通常来说使用这两种得到的结果会biased downward，即小于真实值，更容易拒绝原假设

所以我们要使用新的标准误，称为clustered standard error

Clustered standard error estimates the variance of when the error terms are i.i.d. across entities but are potentially autocorrelated within an entity.

在大样本情况下，使用群居标准误进行推断时，可以采用通用的大样本正太分布临界值作为t统计量临界值，用的临界值作为F统计量的临界值

计算方法：



则的估计量为

由于我们知道观测量对于不同实体是i.i.d.的

如果样本容量很大，根据中心极限定理可得

由于不可知，我们需要估计它

它的估计量就是clustered standard error，即，其中

在面板数据的回归中，不论是否有异方差或序列相关，clustered standard error都成立

总结：前三个lecture介绍了如何利用一个个体在多个时间点上的观测控制个体间不同但随时间上相同的遗漏变量。为了利用这个，就需要同一个个体在不同时间点上的数据-面板数据，有了相关数据后，就可以把ECO205中学过的多元回归模型推广到这里，即固定效应回归模型

**Lecture 4**

**1. 时间序列数据**

时间序列数据就是相同的主体不同的时间区间的数据

举例：一个国家20年的总消费和通货膨胀率数据

时间序列数据可能带来的问题：

Time lags

Correlation over time

Calculation of standard errors when the errors are serially correlated

**2. 使用回归模型来预测**

注意：预测和因果关系时不同的！！！

举例：日常温度可能是随时间相关的，所以他们对于以后可能的温度有很强的预测性，但这并不意味着明天的温度是由今天的温度来决定的，明天的温度是由其他因素所决定

对于预测而言：

1. 很重要
2. 遗漏变量偏差不是问题
3. 不用解释系数的含义
4. 外部可行性很重要

External validity实现不了的原因可能是因为regime shift和non-statsionary

**3. 滞后、一阶差分、对数和增长率**

数据集 (data set) ：时间序列变量Y的T个观测量

我们只考虑连续的、相同间隔的数据

举例：月度的，从1960年到1999的失业率

Y的前一期的数据称为一阶滞后（first lagged value）-

J期前的取值为j阶滞后值-

序列的一阶差分（first difference）为序列在t-1期间和t期间的变化，即

一阶对数差分为

时间序列在t-1期间和t期间的百分比变化近似等于

使用对数的原因：

很多经济序列，例如国内生产总值，具有近似指数的增长速度-> 序列长期而言趋向于每年以一定的百分率增长-> 序列的对数具有近似于线性的增长速度

注意：一般来说，数据都是以年度数据表示

如果计算的时候用月度数据或季度数据来进行计算的，要相应的乘上12或4

**4. 自相关**

时间序列数据的某一期取值通常与下一期的取值相关

序列与其自身滞后值之间的相关性称为自相关 (autocorrelation) 或序列相关 (serial correlation)

一阶自相关为和的相关系数，即两个相邻时期Y取值的相关系数

二阶自相关为和的相关系数

-> 序列的j阶自相关系数为和的相关系数

自协方差为和其j阶滞后的协方差

即j阶自协方差=

j阶自相关=

这些都是population correlation

* Sample autocorrelation

由于总体的值难以得到，我们通过用样本来估计总体

样本自相关：

其中，

前面一项以T-j结束是因为如果以T结束，则t-j=T-> t=j+T没有这一项

同理后一项为T-j结束

**5. 自回归**

如果想要预测时间序列的未来值，一个较好的出发点是从最近的过去值出发

举例：想要预测下个季度的通货膨胀率可以观察一下前几个季度的通货膨胀率

自回归表示的是将一个时间序列变量的条件均值以它的滞后值的线性组合所表示出来

The number of lags used as regressors is called the order of the autoregression

* 一阶自回归

一阶自回归简记为AR (1)，其中1表示这是一阶的

总体AR (1)模型为：



如果，则代表一阶滞后对于预测没有用

对于是否有用，我们可以进行significance test即t-test

测试和是在测试假说一阶滞后对于预测没有用处

如果我们不能拒绝原假设，则说明一阶滞后对于预测没有用

**6. 预测误差**

Predicted value是in the sample

Forecast是out of sample

预测误差为由于预测造成的错误

它是实际发生的和基于的预测值的差额

即 预测误差=

预测不是OLS预期值，预测误差不是OLS残差

OLS预期值计算的是样本内用于估计回归的观测的预期值

预测是针对用于估计回归的数据以外的某个日期所作的

-> 预测的因变量实际值不用于估计回归样本中

**7. 预测误差均方根**

预测误差均方根度量了预测误差的大小，即预测模型所范典型错误的大小

A measure of the spread of the forecast error distribution

RMSFE有两种误差来源：u未来值未知产生的误差和估计系数产生的误差

方法：

* 用SER

当样本容量较大时，第一种误差远大于第二种误差，则RMSFE近似于SER

* Final prediction error

其中P为number of lags

T为sample size

* Pseudo out of sample forecasting

完全来自于数据

利用到的数据来得到

得到预测误差-

对于重复上述步骤

对于得到的总共T-+1个数据，可以得到

**8. AR(P)模型**

AR(1)模型利用来预测

但这样忽略了较远的过去值种的潜在有用信息

包含这些有用心西的就是p阶自回归模型-AR(P)模型



滞后数P称为自回归阶数或滞后长度

系数没有casual interpretation

用t检测还是F检测来检测历史数据是否对预测有用取决于p的值

**Lecture 5**

**1. 自回归分布滞后模型（ADL）**

具有因变量的p阶滞后和其他预测变量的q阶滞后的自回归分布滞后模型称为ADL(p,q) 模型



ADL模型中不包含Y或x的其他滞后值

-> p和q是真实的滞后长度而其他滞后项的系数为0

**2. 平稳性**

时间序列数据的回归分析用过去来分析历史关系

如果时间序列数据的概率分布不随时间变化，即的联合分布不依赖于s，则称Y是平稳的 (stationary)，反之则为非平稳

就是未来要类似于过去，则这些历史关系可以用来预测未来

**3.AR模型和ADL模型的 LS假说**

将回归模型推广到包含多个预测变量



其中

* 误差项的条件均值为0

给定所有回归变量和它们的滞后项，误差项的均值为0

-> 误差项和自变量无关

但是误差项对于因变量Y的未来值的条件均值不为0

-> AR模型的系数是有偏的

一般来说是，biased downward

* 是平稳的

数据是从平稳分布中抽取的

今天的数据分布与过去的分布是相同的

这个假设和截面数据中的假设中的identically distributed类似

但这个假设在实践中往往不成立

* 当间隔时间较长时，随机变量是独立分布的

这个假设有时候被称为弱相关

-> 在大样本条件下，LS估计量是一致的但有偏，asymptotically normal

* 所有变量有非零四阶矩
* 不存在完全多重共线性

**4. Granger因果检验**



Granger因果关系统计量为检验原假设为所有X前面的系数为0的F统计量

原假设暗示这些自变量没有关于因变量Y的预测内容

检验这个原假设的检验为Granger因果关系检验

当我们能够拒绝原假设时，我们说X Granger causes Y或者X has predict power over Y

他并不是真实的因果关系

**5. 基于信息准则的滞后长度选择**

使用更多阶数的好处：可以包含更多在较远滞后中的有用信息

使用更多阶数的坏处：要估计不必要的系数，会在预测中引入额外的估计误差

需要平衡多加入一个滞后项的边际收益和边际成本

* F统计量方法

P的一种选择方法是从包含很多滞后项的模型开始，

然后进行有关最后一项滞后的假设检验

举例：从估计AR(8)开始，检验八阶滞后系数是不是显著的，如果不显著，就去掉

* BIC



其中SSR(p)为AR(p)估计的残差平方和

选择使BIC最小的p

前一项随着p的增加而减少

后一项随着p的增加而增加

P的BIC估计量是一致的

* AIC



AIC和BIC的差别是AIC中的2取代了BIC中的lnT

所以AIC中的第二项较小

P的AIC估计量不是一致的

通常情况下会高估p

-> AIC estimates more lags than BIC

当使用AIC和BIC去选择p时，the regression must use the same sample

**Lecture 6**

**1. 非平稳性：趋势**

前面我们都假设自变量和因变量都是平稳的，但如果现在这个假设不成立，即自变量或因变量都是非平稳的，那么常规的假设检验、置信区间和预测都是不可靠的

非平稳性的两种常见形式：

* 趋势

趋势是指变量随时间持续长期的运动

时间序列数据围绕其趋势波动

分为确定性和随机性两种趋势

* 突变 (以后会讲)

**2. 确定性趋势**

确定性趋势是时间的非随机函数

例如：确定性趋势是时间的线性函数 eg. Y=t+u

**3. 随机性趋势**

随机性趋势是随机的且随时间变化

例如：较长时间的上升之后伴随着较长时间的下降

一般来说，建立随机性的趋势模型比建立确定性 趋势更为合适

随机性趋势模型中最简单的是随机游走模型

**4. 趋势的随机游走模型**

如果时间序列的变化是独立同分布的，则称服从随机游走

即如果中的是独立同分布的

但一般来说我们会用的条件均值等于0这个条件

-> 序列明天的取值等于今天的取值再加上一个不可预测的变化

从误差项的条件均值为0，我们可以得到基于直到时间t-1的的条件均值为

-> 明天的取值的最佳预测是今天的值

不带漂移随机游走的两个基本特征：

* Your best forecast of the value of Y in the future is the value of Y today
* The variance of Y depends on t

带漂移的随机游走模型：



包含朝一个方向或另一个方向的变动趋势

上面的就代表漂移

如果大于0 -> 平均而言呈上升趋势

序列明天的值的最佳预测是今天的值加上

注意：随机游走模型是非平稳的！！！

**5. 自回归模型、单位根**

我们知道随机游走模型为

而AR(1)模型为

所以随机游走模型就是AR(1)模型中=1的特殊情况

如果AR(P)模型中有一个等于1的根，则称序列有一个单位自回归根

多项式为：

如果序列存在单位根，那么它就包含了随机性趋势

如果是平稳的，就不包含单位根

如果序列包含随机性的趋势，那么大多数情况下是平稳的

-> 回归应该使用而不是

当时，AR(1)模型就变成了

我们可以得到AR(2)模型：



左右两边同时减，得到：



或者表示为：



其中，并且

如果序列包含随机趋势当且仅当只有单位根时

把z=1带人上面的公式，我们可以得到

那么AR(2)模型就变成了

-> 如果AR（2）模型有单位根，那么它可以被写为AR(1)模型使用一阶差分

这个结论也可以被推广到AR(P)模型中

**Lecture 8**

**1. 趋势造成的问题**

* 偏向于0的自回归系数

-> 当Y服从随机游走时，基于AR (1) 模型的预测要比基于施加真值的随机游走模型预测糟糕的多

* T统计量的非正态分布

若回归变量中包含随机趋势，则常用的OLS t统计量在原假设成立时即使在大样本情况下也服从非正态分布

-> 常用的置信区间是不正确的，也不能像往常一样进行假设检验

* 伪回归

随机性趋势会使两个没有相关关系的实际序列呈现出相关性

e.g. 假设 

X和Y都是独立的

回归Y on X

我们期望随着样本容量的增加，回归系数接近于0，t统计量有正态分布

我们通过monte carlo simulation来获得相关分布

如果和都是stationary的，那么回归系数就是无偏的和一致的，t值有正态分布

-> 我们可以用传统的临界值1.96

如果和都是有单位根的，那么回归系数是有偏的和不一致的，t值的分布随样本量的增加而增加

-> 我们不可以使用传统的临界值

**2. 随机性趋势探测**

使用Dickey-Fuller test

* AR (1)模型中的Dickey-Fuller检验

Dickey-Fuller检验的出发点是自回归模型

随机游走模型是AR (1)模型中的特例

如果，则是非平稳的，包含了随机性趋势

所以对于AR (1)模型，包含趋势的假设检验为检验

中 vs 

或中 vs 

原假设为它有单位根，备择假设为它是平稳的

其中检验的t统计量就是Dickey-Fuller 统计量

如果，那么OLS估计量就存在我们想要的特点

如果，那么OLS估计量就是有偏的，t统计量也不服从正态分布

我们也是通过monte carlo simulation来获得分布的

**3. ADF统计量的临界值**

ADF统计量在单位根的原假设下即使大样本也不服从正态分布

因为分布不是标准的正态分布，所以使用ADF统计量检验单位根的时候不能用正态分布临界值

ADF检验是单边检验

如果不包含时间趋势，当ADF统计量小于-2.86时，在5%的玄珠水平下拒绝原假设

如果包含时间趋势，那么临界值变为-3.41

是否有时间趋势，可以从视觉上来检验

* 如果时间序列不存在长期的增加或下降，那么在不存在随即趋势时，时间序列是 stationary AR process around a constant

即 intercept only: 

* 如果时间序列存在长期趋势，那么在不存在随即趋势时，时间序列是stationary process around a deterministic trend

即intercept and time trend: 

**4. AR (p)模型中的Dickey-Fuller检验**

在AR (p)模型中，ADF检验是基于转换形式



其中

如果有随即趋势，那么

如果是平稳的，那么

然后步骤和之前的AR（1）模型相同

其中滞后长度可以用AIC或者BIC估计

如果有随即趋势，最简单的方法去避免问题就是使用模型 in first difference

**Lecture 9**

**1. 突变**

突变来自总体回归系数在某一特定日期上的离散变化或来自系数在较长时期内的渐变

如果总体回归函数在样本内发生了突变，则全样本的OLS回归估计是平均成立的关系

-结合了两个不同的时期

**2. 突变的检验**

* 突变时间已知的突变检验

为了简便，以带截距项，只有 和的一阶滞后 ADL (1,1) 模型为例

假设为发生突变的时间

在突变发生前等于0，突变发生后等于1

则

如果不存在突变，那么两部分样本上的总体回归函数相等->

如果至少有一个不等于0，则发生过突变

可以用Chow检验来检验

* 突变时间未知的突变检验

突变的可能时间一般是未知的或只知道在某一范围内

e.g. 在和之间发生突变

这时候我们可以修改Chow检验来让他能够处理在这两者之间所有可能日期上的突变检验

利用得到的最大F统计量取值来检验未知时间上的突变

修改后的Chow检验称为QLR检验

注意：QLR统计量的临界值必须从特定分布中得到

他的分布依赖于待检验的约束个数q

要使QLR统计量分布的大样本近似较好，子样本的和不可以过于接近样本起点或终点

通常选择剔除15%

即

类似于F统计量，QLR统计量可以用来检验所有或部分回归系数中的突变

注意：你认为你知道时你也可能不知道突变时间！

**3. 伪样本外预测**

是一种模拟预测模型在真实时间上表现的方法

思想：挑选一个接近样本末尾的时间，用直到这个时间的数据估计预测模型，然后利用这个预测模型估计做预测

它被称为伪样本外预测是因为它不是真正的样本外预测

步骤：

选择观测次数P-> 你要进行的伪样本外预测

设s=T-P

利用缩短后的数据t=1，2，…，s估计预测回归

计算这个缩短样本外的第一期s+1的预测，记做

计算预测误差->

对剩下的时间s=T-P+1到T-1重复上面的步骤

伪样本外预测的第二个用途时估计RMSFE

伪样本外预测反映了与误差项未来值有关的不确定性，还反映了由于估计系数带来的不确定性

伪样本外预测误差的样本标准差是RMSFE的估计量

**Lecture 10**

**1. 动态因果效应**

X变化时对Y的现在或将来效应的估计问题

e.g. 一个冻度日对橘子价格变化的因果效应

动态效应必定是随时间发生的，所以动态因果效应的模型中必定包含滞后项

上述模型被称为Y的分布滞后模型（distributed lag model）

其中，为immediate effect of change in X

为1-period dynamic multiplier

为2-period dynamic multiplier

X变化一个单位对h期后的Y的效应，称为h期的动态乘数

累积动态因果效应就是把这个时期之前的所有dynamic multiplier加起来

X变化一个单位对接下来h期的Y的累积效应为h期累计动态乘数

e.g. the 2-period cumulative dynamic multiplier =

**2. 两类外生性**

* 给定X的现在和所有过去的值时误差项具有条件零均值

->

-> 所有更远的距离出的因果效应，就是滞后超过r的因果效应为0

-> r阶分布之后系数构成了非零的动态因果效应

过去和现在外生性

* 给定X的所有过去值、现在值和未来值时误差项具有条件零均值

过去、现在和未来外生性 / 严格外生性

当X严格外生时，分布滞后回归系数的OLS估计量更有效的动态因果效应估计量

外生性和严格外生性的区别就是严格外生性包括了X的未来值

-> 严格外生性成立，则外生性一定成立，反之并不成立

e.g. 由于试验中的肥料是随机施用的，所以是外省的

由于今天的番茄产量不依赖于未来施用的肥料，所以也是严格外生的

**3. 包含外生变量的动态因果效应估计**

* 分布滞后模型假设

X为外生的

->

变量具有平稳分布

当间隔时间变大时，他们是独立分布的

变量有大于八阶的非零有限矩阵

不存在多重共线性

* 自相关的误差项、标准误和推断

在分布之后回归模型中，误差项可能是自相关的，即误差项和滞后值有关

误差项的自相关不会影响OLS的一致性，也不会引入偏差

但如果误差项是自相关的，那么常用的OLS标准误一般是非一致的

解决办法使用异方差和自相关一直的HAC标准误

Newey-West方差估计量是HAC估计量的一种

HAC估计量涉及了m-1个自协方差和方差的估计

在实践中，用HAC标准误需要选取截断参数m的值

**Lecture 12**

**1. VAR模型**

具有两个时间序列变量和的向量自回归有两个方程组成

其中一个方程的因变量为

另一个方程的因变量为

两个方程中的回归变量为两个变量的滞后值

在VAR的假设下，OLS估计量是一致的，且大样本下服从联合正态分布

-> 可以用常用的方法进行推断

Var每个方程的系数个数与var中的变量个数成比例

e.g. 含5个变量和4个滞后的var的5个方程中每一个都有21个系数

实践告诉我们需要确保var变量中的变量个数较少，特别是要确保变量之间相关

这有助于互相之间的预测

Var中的滞后长度可以用F检验或信息准则确定

e.g. BIC

var中基于BIC的滞后长度估计是在p的候选值集合中，选择使BIC最小的的那个p

**2. 因果分析**

Var模型可以用来分析经济时间序列中的因果关系

e.g. 通货膨胀率和失业率的var模型

我们可以用var方程进行Granger因果关系检验

检验通货膨胀方程中滞后长度为1到4的失业率的系数为零的假设检验

得到的F统计量为11.04，p值小于0.001-> 拒绝原假设

可以得到结论：给定通货膨胀率滞后时，失业率有助于预测通货膨胀的变化

即失业率Granger导致了通货膨胀的变化

**3. 多期预测**

之前的讨论都集中于未来一期的预测，但通常要求预报员预测较远的未来

* + 迭代多期预测

利用预测模型作向前一期的预测

e.g. 预测T+1期用到直到T期的数据

迭代预测模型：AR、ADL、VAR

1. 迭代多期AR预测

首先计算向前一期预测，然后利用它计算向前二期预测，以此类推

1. 迭代多期VAR预测

首先计算VAR中所有变量的向前一期预测，然后利用这些预测计算向前二期预测，反复继续这一过程直至想要的预测范围

* + 直接多期预测

假如你想利用直至时间T的数据预测

直接多元方法以ADL模型作为出发点，但对预测变量多取一起滞后

e.g. 如果利用预测变量的二阶滞后，则因变量为，回归变量为 利用这些数据可以直接计算的预测而不需要任何迭代

如果要得到向前h期的预测所有变量都要滞后h期

由于多起回归中的因变量发生在两期以后，因此多期回归中的误差项时序列相关的

-> 直接多期回归中必须采用异方差和自相关一致的HAC标准误差

**4. 选择哪种方法？**

在大多数情况下建议使用迭代方法进行多期预测

如果向前一期模型设定正确，则用向前一期回归得到的系数估计要比向前多期得到的系数估计有效

迭代预测的时间路径相比直接预测在不同时期较为稳定

直接预测在每一范围内采用的是不同的模型，因此系数估计中的抽样误差为增加直接多期预测序列时间路径的随机波动

但在某些情况下，直接预测要优于迭代预测

其中一种情况就是你有理由相信向前一期模型设定不正确

或者是多元预测模型中有很多预测变量

**Lecture 13**

**1. 趋势的其他模型和单整阶数**

一种平滑趋势的模型为趋势的一阶差分服从随机游走

即

误差项序列不相关

服从随机游走-> 平稳

一阶差分称为的二阶差分

具有随机游走趋势的序列称为一阶单整-> 具有单位自回归根

具有上面那个式子的趋势的序列称为二阶单整-> 具有单位自回归根

不具有趋势且平稳的序列称为零阶单整

检验方法就是通过检验是由具有单位自回归根

如果拒绝具有单位自回归根的假设，则拒绝为二阶单整的假设，接受为一阶单整

**2. DF-GLS检验**

原假设为服从可能带漂移的随机游走趋势

备择假设为围绕线性时间趋势平稳

DF-GLS检验分为两步：

1. 利用广义最小二乘假设估计截距和趋势
2. 利用DF检验来检验中的单位自回归根，其中DF回归中不包含截距项或时间趋势

如果DF-GLS检验统计量小于临界值，则拒绝具有单位根的原假设

**3. 协整**

两个或多个具有随机性趋势的时间序列在长期内的共同变动如此接近以至于似乎包含了相同的趋势成分

-> 他们看来具有共同趋势

两个或多个具有共同随机性趋势的序列称为是协整的

假设都是一阶单整

若对某个系数，为零阶单整，则称是协整的，称为协整系数

如果是协整的，则它们具有相同的或共同的随机性趋势

的差异计算消除了这种共同的随机性趋势

判断方法：

* + 已知的协整检验

在某些情况下，专业知识或经济理论建议了的取值

当已知的时候，可以用DF和DF-GLS单位根检验作协整检验

* + 未知的协整检验

如果协整系数未知，则在检验误差修正项中的单位跟之前必须先估计它

这个步骤使得下面的单位根检验中必须采用不同的临界值

1. 利用下面的回归的OLS估计估计协整系数
2. 利用DF的t检验检验这个回归残差是否具有单位根

这两步被称为EG-ADF检验

警示：若两个或两个以上变量协整，则误差修正项有助于这些变量及其他相关变量的预测

但协整要求变量具有相同的随机性趋势

**4. 向量误差修正模型**

如果是协整的，消除趋势的一种方法是计算

其中选择的要消除差分中的趋势

事实上，如果是协整的，则的一阶差分可以用VAR进行建模

其中加入新的回归变量做改进

这一项就被称为误差修正项

模型为成为向量误差修正模型