时间序列回归与预测

含外生回归变量时的动态因果效应估计

- 分布滞后模型

$$Y_t = \beta_0 + \beta_1 X_t + \cdots + \beta_{r+1} X_{t-r} + u_t$$

分布滞后模型的假设

1. X是外生的,$E(u_t|X_t, X_{t-1}, X_{t-2}, \ldots) = 0$

2. (a) Y和X是平稳分布　　(b) 当j足够大时,(Y_t, X_t)与(Y_{t-j}, X_{t-j})独立

3. Y_t, X_t有大于八阶的非零有限矩

4. 不存在完全多重共线性

时间序列回归中的外生性

- 分布滞后模型：

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 X_{t-1} + \cdots + \beta_{\mathrm{r}+1} X_{t-r} + u_t$$

回顾：在横截面数据回归中，为一致的估计因果效应，我们需要误差之间彼此不相关。在时间序列回归中，为估计因果关系，我们需要相同的条件。

- 外生性（过去和现在外生性）
  - X是外生的，若$E(u_t|X_t, X_{t-1}, X_{t-2}, \ldots) = 0$
- 严格外生性（过去、现在和未来外生性）
  - X严格外生，若$E(u_t|\ldots, X_{t+1}, X_t, X_{t-1}, X_{t-2}, \ldots) = 0$
- 严格外生性意味着外生性

利用回归模型进行预测

- 预测和估计因果关系具有完全不同的目的
- 对于预测
  - R^2十分重要！
  - 在预测模型中不需要考虑系数的解释能力，遗漏变量偏差不会对估计结果造成影响！
  - 外部有效性最为重要：利用历史数据进行估计的模型必须在（较近的）未来也成立！

R^2以及$\bar{R}^2$

$$R^2 = \frac{ESS}{TSS} = \frac{\sum(\hat{y}_i - \bar{Y})^2}{\sum(y_i - \bar{Y})^2}$$

或者，

$$R^2 = 1 - \frac{SSR}{TSS} = 1 - \frac{\sum \hat{u}_i^2}{\sum(y_i - \bar{Y})^2}$$

- R^2会随着回归变量的增加而变大。（为什么？）

证明：对于二元回归：

$$Q(k=2) = \min_{b_0,b_1,b_2} \sum_{i=1}^{n} [y_i - (b_0 + b_1 x_{1i} + b_2 x_{2i})]^2 = \sum_{i=1}^{n} \hat{u}_i^2 = SSR$$

对于一元回归：

$$Q(k=1) = \min_{b_0,b_1,b_2} \sum_{i=1}^{n} [y_i - (b_0 + b_1 x_{1i} + b_2 x_{2i})]^2 \quad s.t. b_2 = 0$$

R^2以及$\bar{R}^2$

为了修正R^2会随着回归变量的增加而变大的问题，引入$\bar{R}^2$（调整R^2）

$$\bar{R}^2 = 1 - \frac{n-1}{n-k-1}\frac{SSR}{TSS} = 1 - \frac{\frac{1}{n-k-1}\sum \hat{u}_i^2}{\frac{1}{n-1}\sum(y_i - \bar{Y})^2}$$

也就是

$$\bar{R}^2 = 1 - \frac{s_{\hat{u}}^2}{s_y^2}$$

由于$\frac{n-1}{n-k-1} > 1$，所以有$\bar{R}^2 < R^2$。且$\bar{R}^2$可能为负。

- 主要有以下两种模型进行预测：
  - 自回归（AR）模型
  - 自回归分布滞后（ADL）模型

模型选择

- AR模型

$$AIC(p) = \ln\left(\frac{SSR(p)}{T}\right) + (p+1)\frac{2}{T}$$

$$BIC(p) = \ln\left(\frac{SSR(p)}{T}\right) + (p+1)\frac{\ln T}{T}$$

- ADL模型

$$BIC(K) = \ln\left(\frac{SSR(K)}{T}\right) + K\frac{\ln T}{T}$$

其中，*K*=模型中系数的个数（包括截距，*Y*的迟滞项，*X*的迟滞项）

预测误差与预测区间

为什么我们需要计算预测的不确定性？

- 需要知道预测的准确性
- 需要计算预测区间

考虑如下预测值

$$\hat{Y}_{T+1|T} = \hat{\beta}_0 + \hat{\beta}_1 Y_T + \hat{\beta}_2 X_T$$

预测偏差：

$$Y_{T+1} - \hat{Y}_{T+1|T} = u_{T+1} + \left(\beta_0 - \hat{\beta}_0\right) + \left(\beta_1 - \hat{\beta}_1\right) Y_T + (\beta_2 - \hat{\beta}_2) X_T$$

The mean squared forecast error (MSFE)

- MSFE:

$$E\left(Y_{T+1}-\hat{Y}_{T+1|T}\right)^2=E(u_{T+1})^2+E\left[\left(\beta_0-\hat{\beta}_0\right)+\left(\beta_1-\hat{\beta}_1\right)Y_T+(\beta_2-\hat{\beta}_2)X_T\right]^2$$

- $MSFE=Var(u_{T+1})$ +参数估计引起的方差

- 如果样本比较大，参数估计引起的方差比$Var(u_{T+1})$小很多。在这种情况下

$$MSFE\approx Var(u_{T+1})$$

- The root mean squared forecast error (RMSFE) 定义为：

$$RMSFE=\sqrt{E\left(Y_{T+1}-\hat{Y}_{T+1|T}\right)^2}$$

参数估计的方差

$\hat{\beta}_1$的方差为：

$$\sigma_{\hat{\beta}_1}^2 = \frac{1}{n-1}\left[\frac{1}{1-R_{X_1,X_2}^2}\right]\frac{\sigma_u^2}{S_{X_1}^2}$$

其中，R_{X_1,X_2}^2是$X_{1i} = \gamma_0 + \gamma_1 X_{2i} + v_i$模型回归得到的拟合优度。

The mean squared forecast error (MSFE)

The root mean squared forecast error (RMSFE)：

$$RMSFE = \sqrt{E\left(Y_{T+1} - \hat{Y}_{T+1|T}\right)^2}$$

- RMSFE度量预测误差的发散程度
- RMSFE的意义和随机变量u_t的标准误比较像
- RMSFE度量预测“错误”的程度

如何估计MSFE

- 由于$MSFE \approx Var(u_{T+1})$，所以我们通过估计$u_{T+1}$的标准误估计MSFE
- 伪样本外MSFE估计

$$\widehat{MSFE} = \frac{1}{T - t_1 + 1} \sum_{t=t_1-1}^{T-1} \left(Y_{t+1} - \hat{Y}_{t+1|t}\right)^2$$

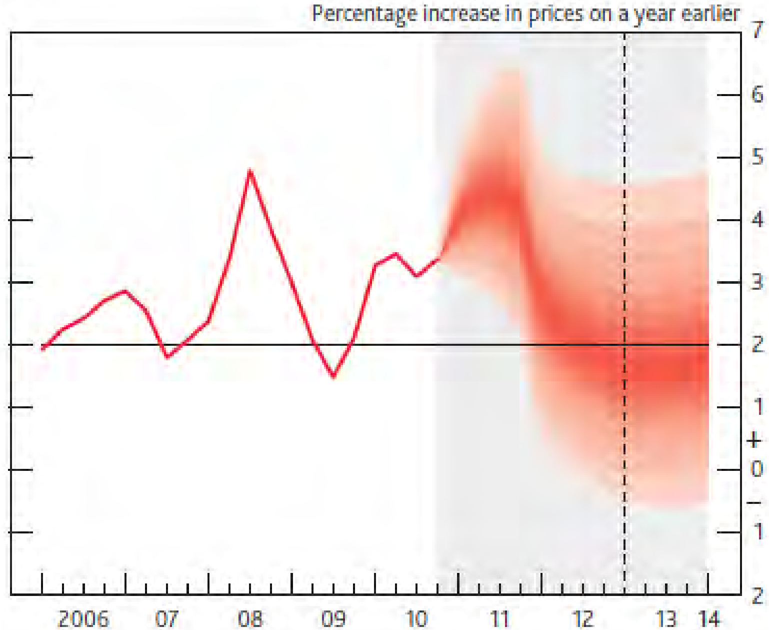
- $\hat{Y}_{T+1|T}$的95%置信区间为：

$$\hat{Y}_{T+1|T} \pm 1.96 \times R\widehat{MSFE}$$

- $\hat{Y}_{T+1|T}$的67%置信区间为：

$$\hat{Y}_{T+1|T} \pm R\widehat{MSFE}$$

Chart 3 CPI inflation projection based on market interest rate expectations and £200 billion asset purchases

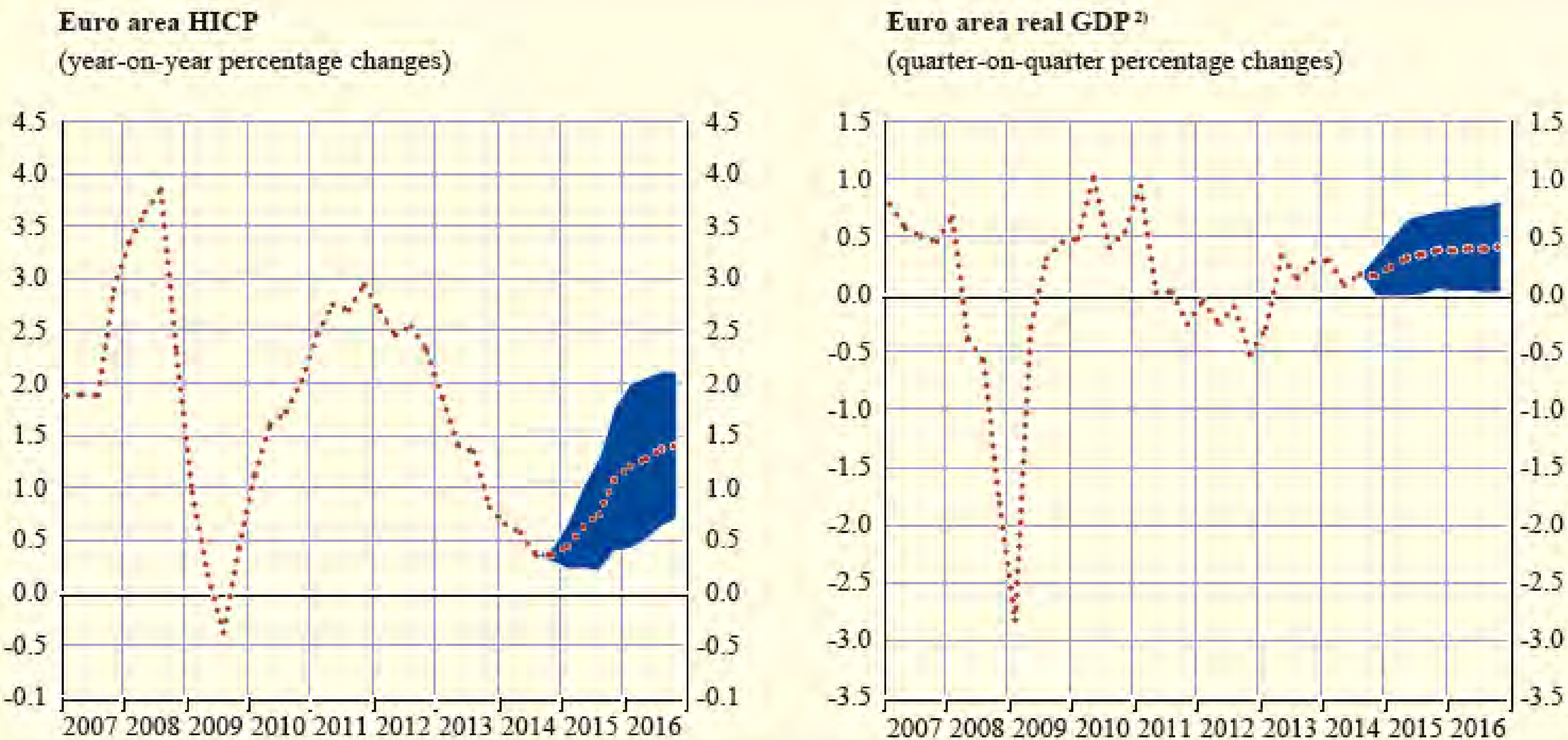


The fan chart depicts the probability of various outcomes for CPI inflation in the future. It has been conditioned on the assumption that the stock of purchased assets financed by the issuance of central bank reserves remains at £200 billion throughout the forecast period. If economic circumstances identical to today's were to prevail on 100 occasions, the MPC's best collective judgement is that inflation in any particular quarter would lie within the darkest central band on only 10 of those occasions. The fan chart is constructed so that outturns of inflation are also expected to lie within each pair of the lighter red areas on 10 occasions. In any particular quarter of the forecast period, inflation is therefore expected to lie somewhere within the fan on 90 out of 100 occasions. And on the remaining 10 out of 100 occasions inflation can fall anywhere outside the red area of the fan chart. Over the forecast period, this has been depicted by the light grey background. In any quarter of the forecast period, the probability mass in each pair of identically coloured bands sums to 10%. The distribution of that 10% between the bands below and above the central projection varies according to the skew at each quarter, with the distribution given by the ratio of the width of the bands below the central projection to the bands above it. In **Chart 3**, the ratios of the probabilities in the lower bands to those in the upper bands are approximately 4:6 at Years 2 and 3. The upward skew at Year 1 is smaller. See the box on pages 48–49 of the May 2002 *Inflation Report* for a fuller description of the fan chart and what it represents. The dashed line is drawn at the two-year point.

http://www.bankofengland.co.uk/publications/inflationreport/ir11feb.pdf

Chart 1 Macroeconomic projections[1]

(quarterly data)



1) The ranges shown around the central projections are based on the differences between actual outcomes and previous projections carried out over a number of years. The width of the ranges is twice the average absolute value of these differences. The method used for calculating the ranges, involving a correction for exceptional events, is documented in *New procedure for constructing Eurosystem and ECB staff projection ranges*, ECB, December 2009, available on the ECB's website.
2) Working day-adjusted data.

http://www.ecb.europa.eu/pub/economic-bulletin/mb/html/index.en.html

大数据中模型选择方法

时间序列数据模型的分解

经典的时间序列分解模型：

$$y_t = m_t + s_t + x_t$$

- m_t:趋势项
- s_t： 季节性
- x_t： 平稳部分

季节性

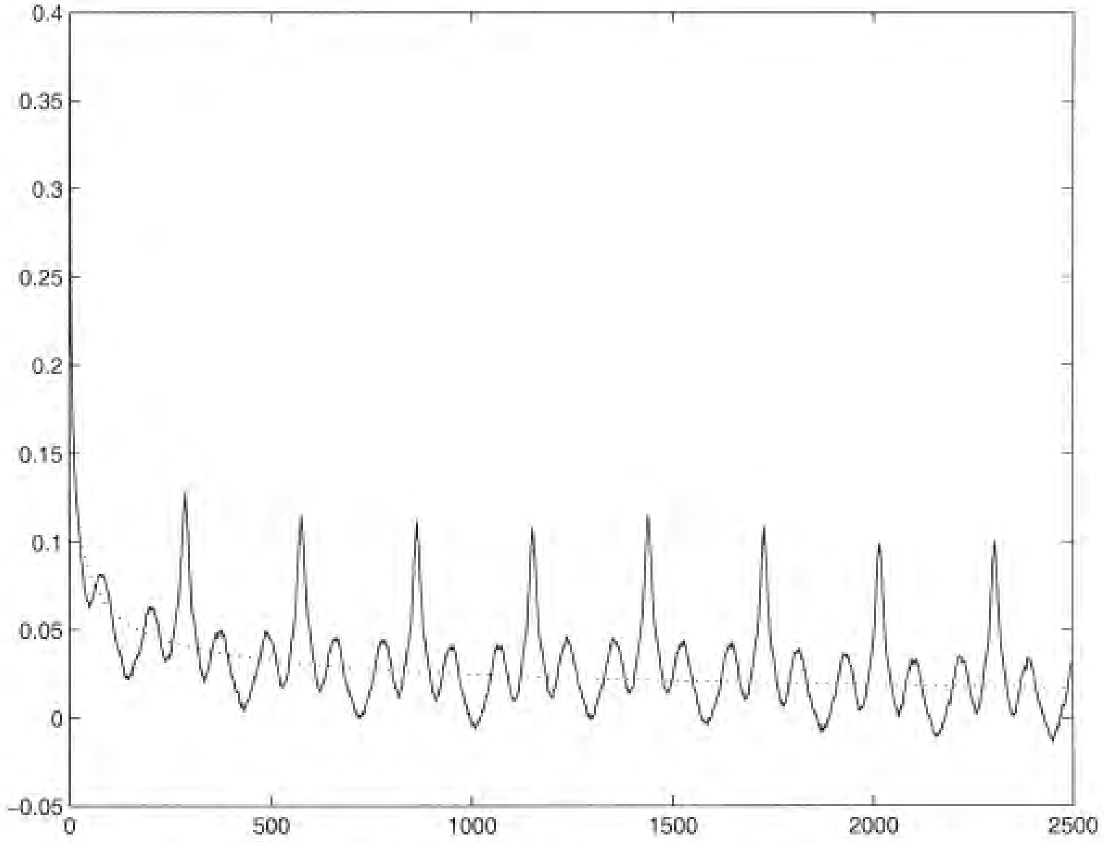


Fig. 1. Sample and benchmark model population autocorrelograms. The solid line graphs the sample autocorrelogram for the 5-min log-squared ¥–$ series. The sample period extends from December 1, 1986 through December 1, 1996, for a total of $T = 751{,}392$ observations. The dotted line refers to the theoretical autocorrelogram for the FISV model defined by Eqs. (2) and (11) with parameters $d = 0.3$, $\sigma_\varepsilon^2 = 4.10^{-4}$, $\phi = 0.6$ and $\sigma_u^2 = 0.25$.

Bollerslev and Wright, 2000

季节性调整及X-11 滤波

月度数据：X-11 filter

$$
\begin{aligned}
SM(L) &= 1-\frac{1}{24}(1+L)(1+L+\cdots+L^{11})L^{-6} \\
&\approx -0.042L^6-0.083L^5-0.083L^4-0.083L^3 \\
&\quad -0.083L^2-0.083L+0.917-0.083L^{-1} \\
&\quad -0.083L^{-2}-0.083L^{-3}-0.083L^{-4}-0.083L^{-5}
\end{aligned}
$$

季度数据：X-11 filter

$$
\begin{aligned}
SQ(L) &= 1-\frac{1}{8}(1+L)(1+L+L^2+L^3)L^{-2} \\
&= 0.125L^2-0.250L+0.750-0.250L^{-1}-0.125L^{-2} \\
HQ(L) &= -0.073L^2+0.294L+0.558+0.294L^{-1}
\end{aligned}
$$

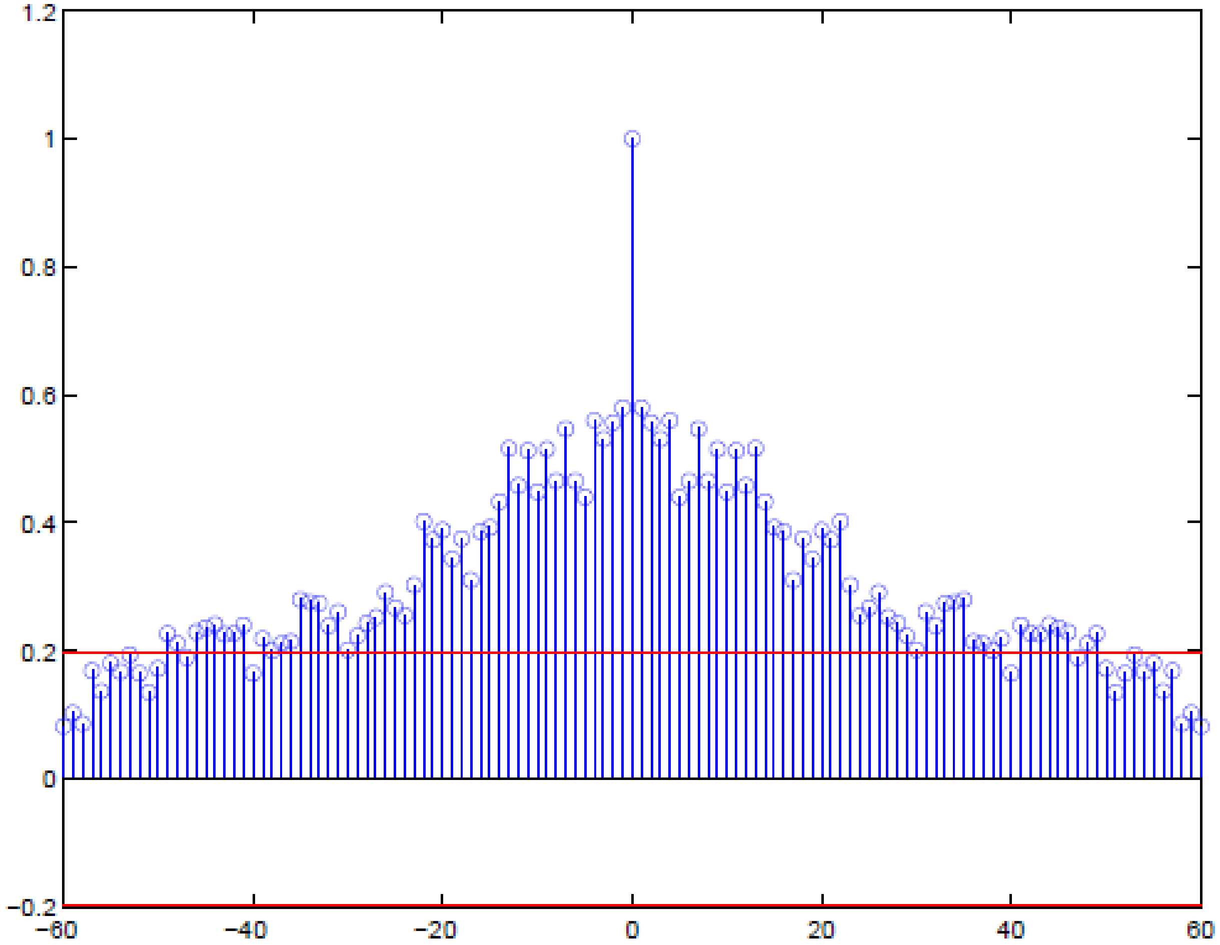
趋势

- 确定性趋势：关于时间的非随机函数（$y_t = \beta t, y_t = \beta t^2$）
- 随机趋势:随机的随着时间变化的
- 随机趋势的典型例子就是随机游走过程:

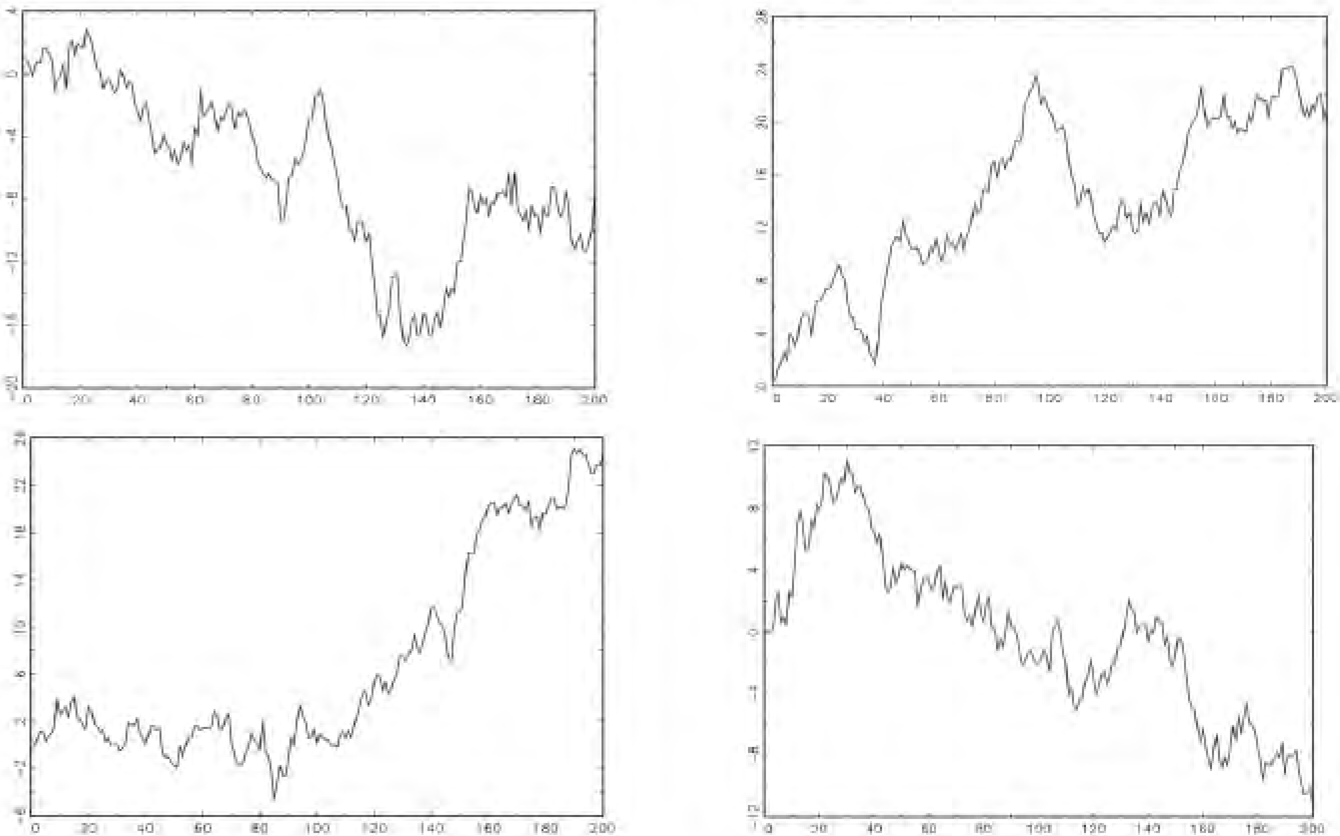
$$y_t = y_{t-1} + u_t$$

- 如果y_t是随机游走过程，那么当前的y取值就是下一期y的预测值
- 如果$y_0 = 0$，那么$Var(y_t) = t\sigma_u^2$

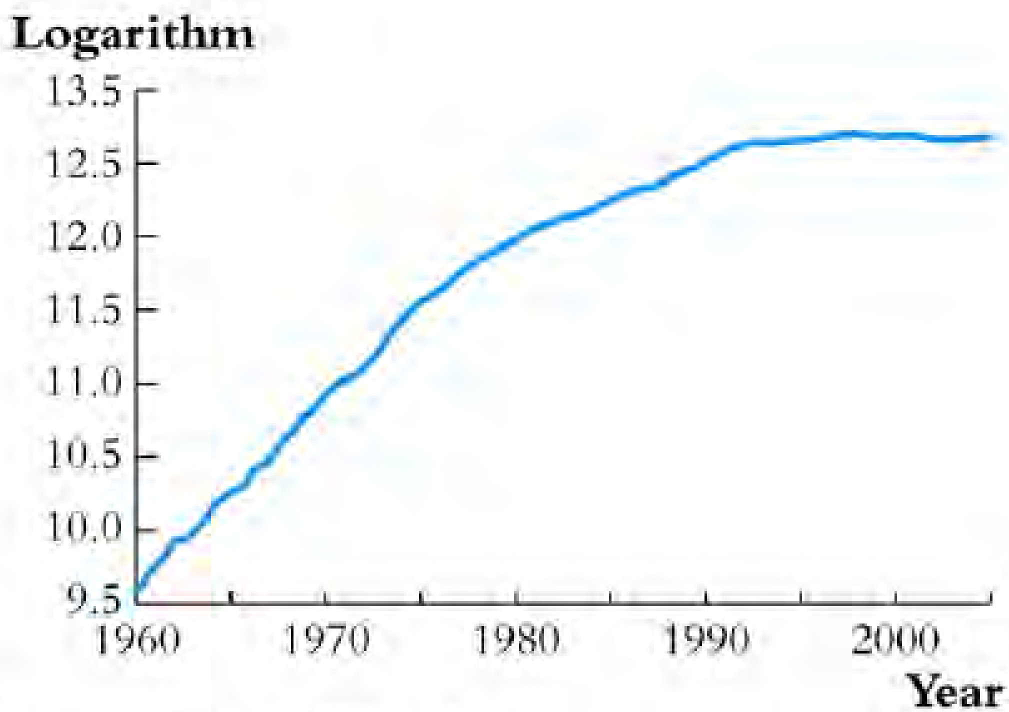
Sample ACF



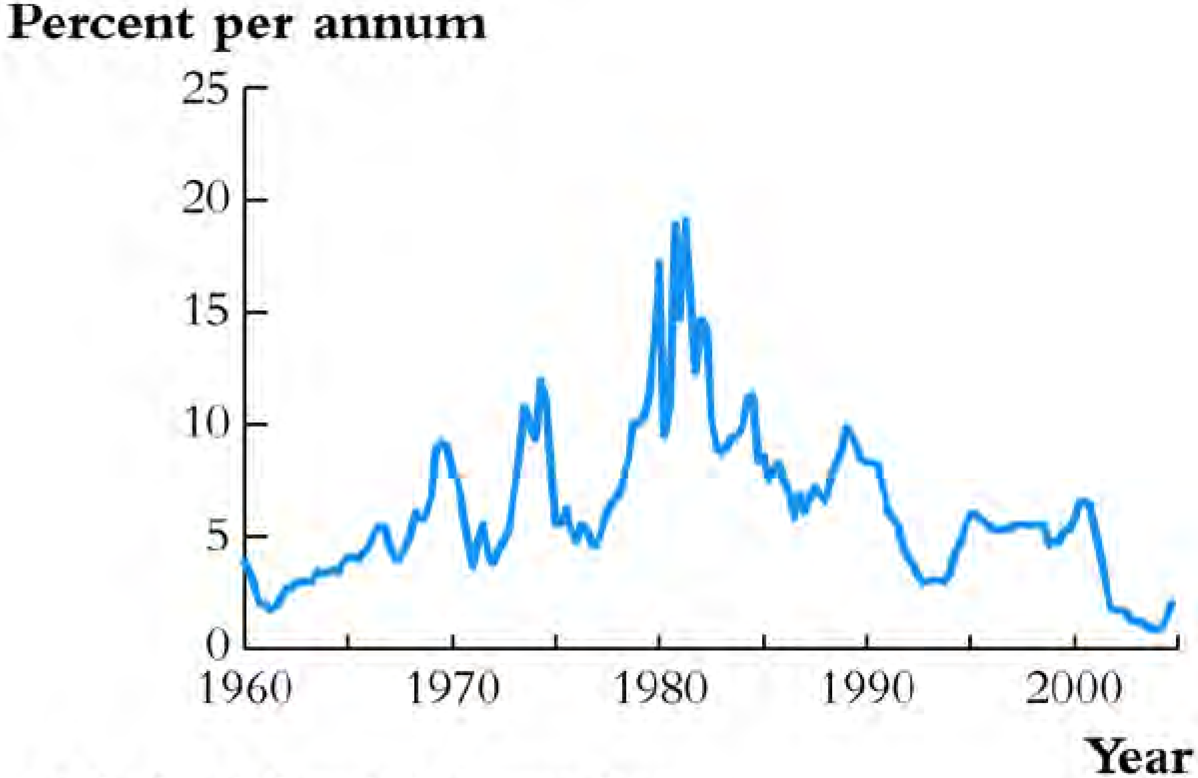
Four artificially generated random walks, T = 200:



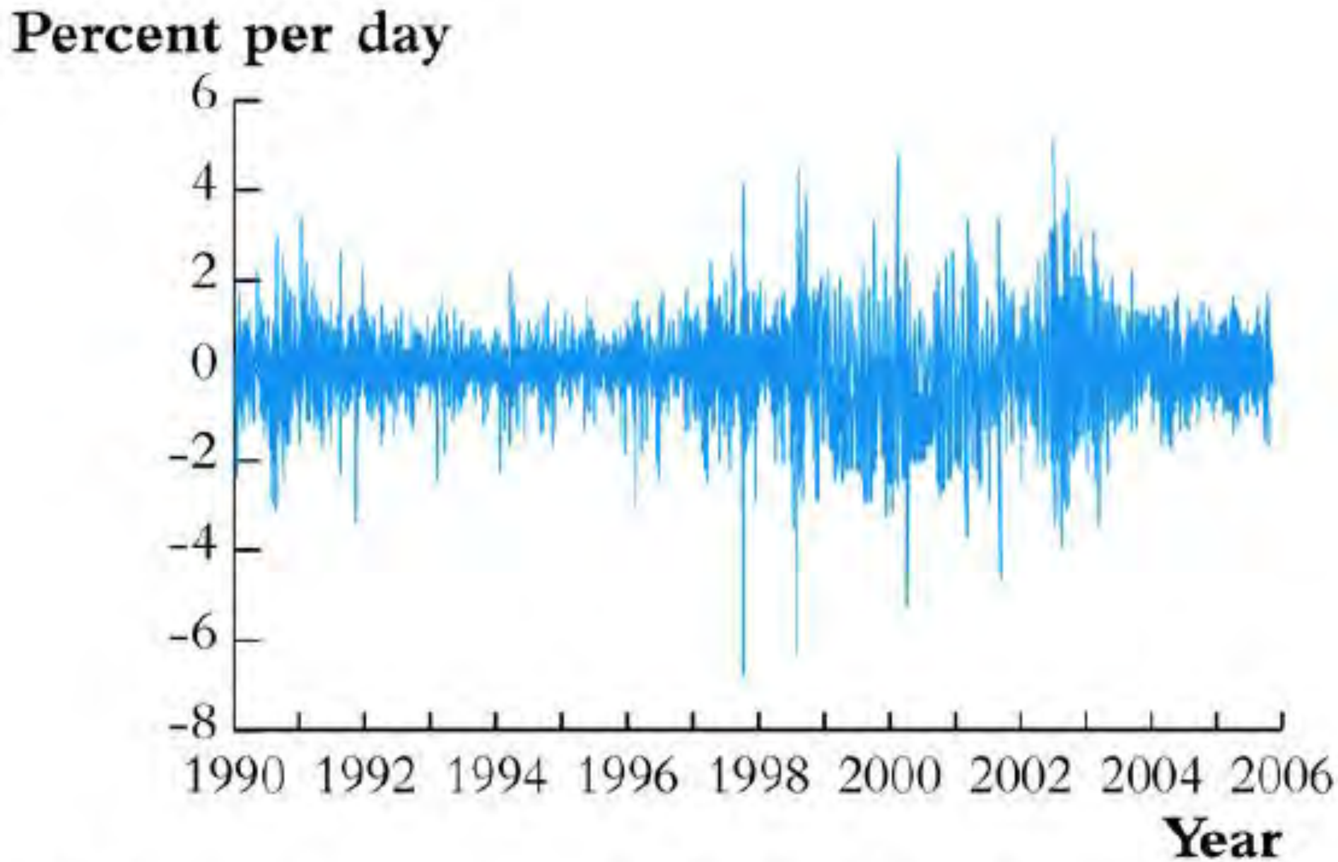
趋势



(c) Logarithm of GDP in Japan



(a) Federal Funds Interest Rate

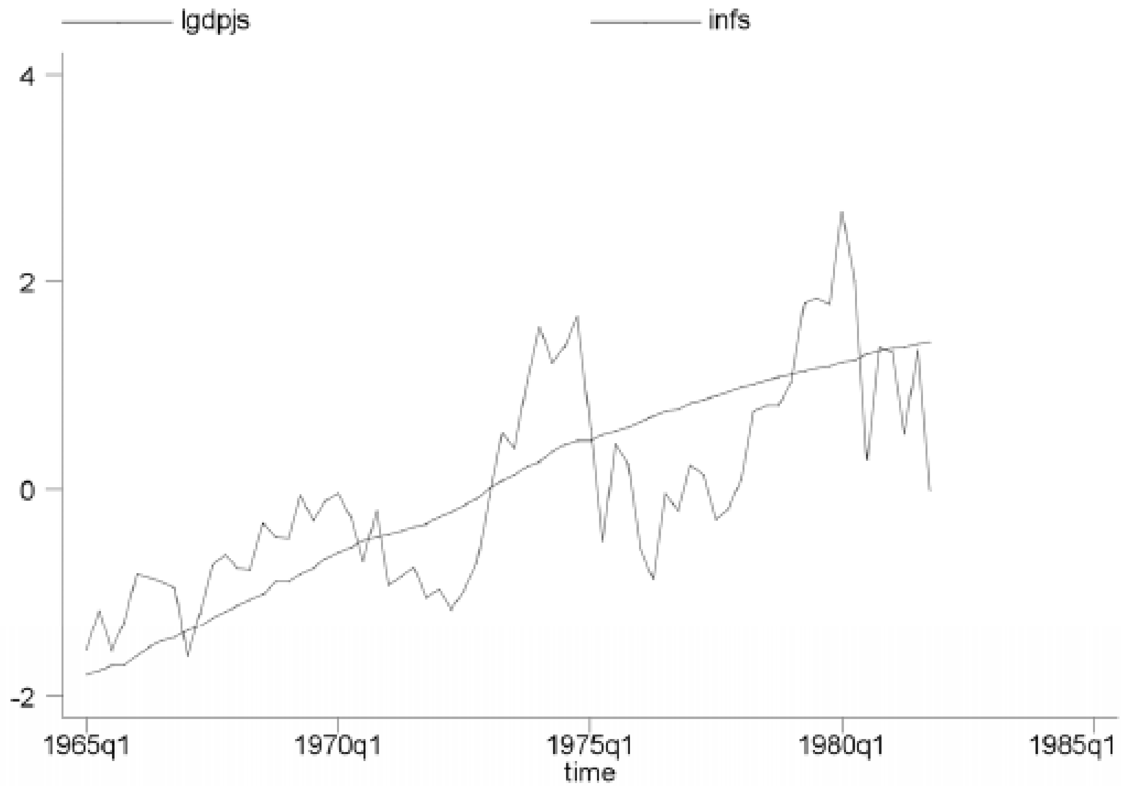


(d) Percentage Changes in Daily Values of the NYSE Composite Stock Index

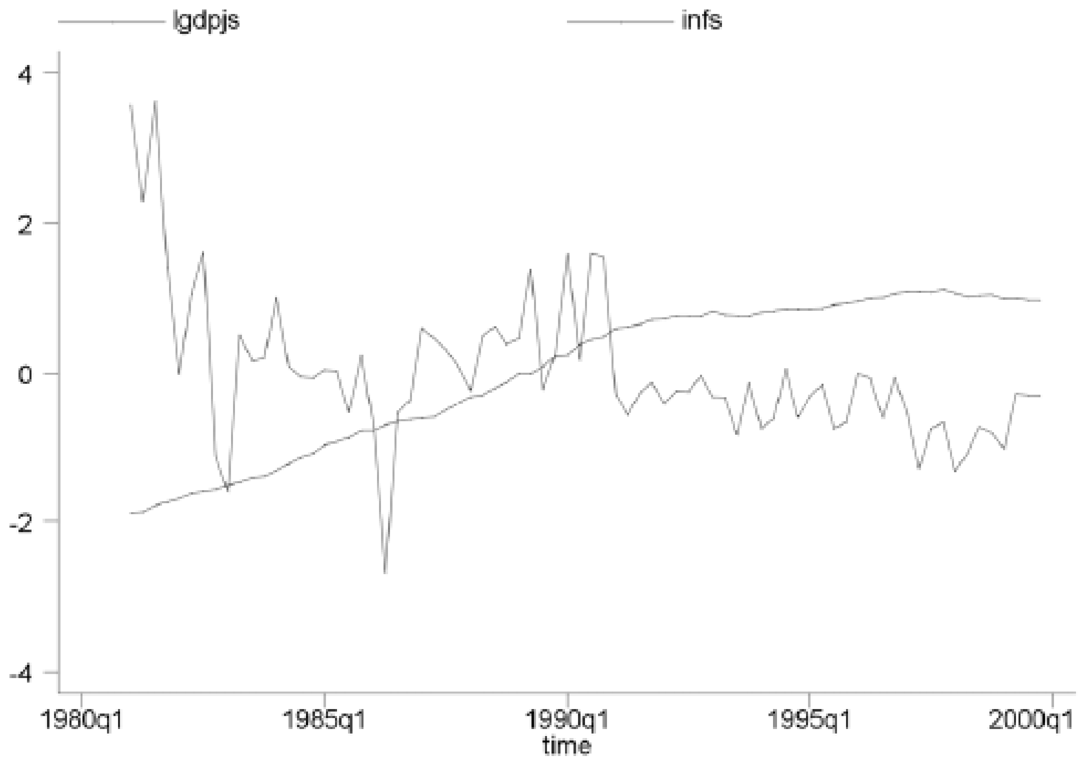
模型中没有考虑趋势项会有什么问题?

- 如果数据具有确定性趋势，用AR模型做回归，参数估计有偏（绝对值偏小）
- 如果数据具有随机趋势，AR模型回归的参数估计的t统计量即使在大样本情况下也不具有正态分布
- 如果Y和X都具随机趋势，那么即使他们不具有因果关系，也可能估计出因果关系（参数估计不一致）

日本GDP和美国Inflation,1965−1981



日本GDP和美国Inflation,1982−1999



如何发现随机趋势?

- 画图:观测是否具有长期的趋势
- 回归方法检验是否具有随机趋势

Dickey-Fuller 单位根检验：

$$y_t = \beta_0 + \beta_1 y_{t-1} + u_t$$

或者

$$\Delta y_t = \beta_0 + \delta y_{t-1} + u_t$$

H_0: $\delta = 0$ 也就是$\beta_1 = 1$。H_1: $\delta < 0$

注意： Dickey-Fuller 单位根检验是单边检验，也就是备择假设y是平稳的。

也可以在检验中加入确定性趋势：

$$\Delta y_t = \beta_0 + \mu t + \delta y_{t-1} + u_t$$

Dickey-Fuller 单位根检验

- 原假设下，根据y_{t-1}的系数估计$\hat{\delta}$构造的t统计量在大样本情况下不服从正态分布，其分布函数和确定性趋势的形式有关。大样本下的临界值的确定也和确定性趋势有关。

(a) $\Delta Y_t = \beta_0 + \delta Y_{t-1} + u_t$ (intercept only)

(b) $\Delta Y_t = \beta_0 + \mu t + \delta Y_{t-1} + u_t$ (intercept and time trend)

TABLE 14.5 Large-Sample Critical Values of the Augmented Dickey–Fuller Statistic

Deterministic Regressors	10%	5%	1%
Intercept only	−2.57	−2.86	−3.43
Intercept and time trend	−3.12	−3.41	−3.96

AR（p）过程的平稳性

AR(p):

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \cdots + \beta_p y_{t-p} + u_t$$

上式可以改写为：

$$A(L)y_t = \beta_0 + u_t$$

其中，$A(L) =$（$1 - \beta_1 L - \beta_2 L^2 - \cdots - \beta_p L^p$）

若$A(z) = 0$的p个根都大于1，那么此AR(p)过程平稳。若最大的根等于1，就说此时间序列存在单位根。

例如：

$$y_t = 0.5y_{t-1} + +u_t$$

平稳吗？

AR（2）过程的单位根检验

AR(2):

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + u_t$$

上式可以改写为：

$$y_t = \beta_0 + (\beta_1 + \beta_2) y_{t-1} - \beta_2 y_{t-1} + \beta_2 y_{t-2} + u_t$$

方程两边同时减去y_{t-1}：

$$\Delta y_t = \beta_0 + (\beta_1 + \beta_2 - 1) y_{t-1} - \beta_2 \Delta y_{t-1} + u_t$$

或者

$$\Delta y_t = \beta_0 + \delta y_{t-1} + \gamma \Delta y_{t-1} + u_t$$

其中，$\delta = \beta_1 + \beta_2 - 1$，$\gamma = -\beta_2$

Dickey-Fuller 单位根检验

一般形式的Dickey-Fuller 单位根检验：

$$\Delta y_t = \sum_{i=0}^{k} \mu_i t^i + \delta y_{t-1} + \sum_{j=1}^{p} \gamma_j \Delta y_{t-j} + u_t$$

原假设下，根据y_{t-1}的系数估计$\hat{\delta}$构造的t统计量在大样本情况下不服从正态分布，其分布函数和确定性趋势的形式以及滞后项的阶数有关。所以，大样本下的临界值的确定也和确定性趋势以及滞后项的阶数有关。

美国的通胀率是否有随机趋势?

(a) U.S. CPI Inflation Rate

DF test for a unit root in U.S. inflation – using p = 4 lags

```
. reg dinf L.inf L(1/4).dinf if tin(1962q1,2004q4);
      Source |       SS       df       MS              Number of obs =     172
-------------+------------------------------           F(  5,   166) =   10.31
       Model |  118.197526     5  23.6395052           Prob > F      =  0.0000
    Residual |  380.599255   166   2.2927666           R-squared     =  0.2370
-------------+------------------------------           Adj R-squared =  0.2140
       Total |  498.796781   171  2.91694024           Root MSE      =  1.5142

------------------------------------------------------------------------------
        dinf |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         inf |
         L1. |  -.1134149   .0422339    -2.69   0.008    -.1967998     -.03003
        dinf |
         L1. |  -.1864226   .0805141    -2.32   0.022    -.3453864   -.0274589
         L2. |   -.256388   .0814624    -3.15   0.002     -.417224   -.0955519
         L3. |    .199051   .0793508     2.51   0.013     .0423842    .3557178
         L4. |   .0099822   .0779921     0.13   0.898     -.144002    .1639665
       _cons |   .5068071    .214178     2.37   0.019     .0839431     .929671
------------------------------------------------------------------------------
```

DF t-statistic = −2.69

Don't compare this to −1.645 – use the Dickey-Fuller table!

DF *t*-statistic = −2.69 (intercept-only):

TABLE 14.5 Large-Sample Critical Values of the Augmented Dickey–Fuller Statistic

Deterministic Regressors	10%	5%	1%
Intercept only	−2.57	−2.86	−3.43
Intercept and time trend	−3.12	−3.41	−3.96

- 通胀率是否具有单位根?

$$y_t = \beta z_t + \alpha y_{t-1} + \sum_{i=1}^{k} d_i \triangle y_{t-i} + e_{tk}$$

Both the ADF and Phillips-Perron type tests are widely used in practice and implemented in most statistical softwares, such as Stata, EViews, TSP and so on. These software packages offer default values for the implementation of the tests. For the ADF, the crucial auxiliary parameter to choose is the truncation lag k. The default value in Stata is

$$k = int[4(T/100)^{2/9}]$$

Unit Root Tests in ARMA Models with Data-Dependent Methods for the Selection of the Truncation Lag

Serena Ng; Pierre Perron

Journal of the American Statistical Association, Vol. 90, No. 429 (Mar., 1995), 268-281.

Econometrica, Vol. 69, No. 6 (November, 2001), 1519–1554

LAG LENGTH SELECTION AND THE CONSTRUCTION OF UNIT ROOT TESTS WITH GOOD SIZE AND POWER

BY SERENA NG AND PIERRE PERRON[1]

Ng and Perron (1995) argued for the use of data-dependent methods to select the truncation lag in the ADF test. They considered four possibilities

1. AIC: $k = \operatorname{argmax} \min_k \ln(\hat{\sigma}^2_{ek}) + 2k/T$ where $\hat{\sigma}^2_{ek} = T^{-1} \sum_{t=k+1}^{T} \hat{e}^2_{tk}$ with $\hat{e}_{tk}$ the residuals from the Dickey-Fuller regression.
2. BIC: $k = \operatorname{argmax} \min_k \ln(\hat{\sigma}^2_{ek}) + \ln(T)k/T$
3. $t_{sig}(5)$: A general to specific testing procedure on the last autoregressive lag using 5% two-sided tests (starting at $k\max = 10$ and stopping when a rejection occurs).
4. $t_{sig}(10)$: same except that a 10% two-sided test is used.

(a) ADF $t_{\hat{\alpha}}$ test

θ	size $k=2$	$k=4$	$k=6$	$k=8$	$k=12$	power ($\alpha = 0.85$) $k=2$	$k=4$	$k=6$	$k=8$	$k=12$
0.0	0.068	0.064	0.078	0.086	0.106	0.557	0.472	0.406	0.354	0.303
0.5	0.052	0.064	0.071	0.085	0.102	0.378	0.411	0.377	0.348	0.308
0.8	0.037	0.051	0.072	0.082	0.109	0.267	0.302	0.313	0.308	0.273
-0.2	0.063	0.065	0.084	0.092	0.111	0.591	0.490	0.421	0.358	0.304
-0.5	0.116	0.076	0.073	0.086	0.105	0.868	0.626	0.512	0.434	0.350
-0.8	0.677	0.343	0.201	0.142	0.120	1.00	0.988	0.900	0.772	0.523

(b) Z_α test

θ	size $m=2$	$m=4$	$m=6$	$m=8$	$m=12$	power ($\alpha = 0.85$) $m=2$	$m=4$	$m=6$	$m=8$	$m=12$
0.0	0.044	0.049	0.049	0.049	0.055	0.772	0.776	0.784	0.794	0.805
0.5	0.010	0.021	0.022	0.021	0.014	0.301	0.524	0.534	0.500	0.378
0.8	0.003	0.020	0.024	0.022	0.015	0.247	0.513	0.526	0.474	0.330
-0.2	0.134	0.105	0.110	0.115	0.133	0.960	0.940	0.946	0.952	0.967
-0.5	0.537	0.417	0.428	0.451	0.516	1.00	1.00	1.00	1.00	1.00
-0.8	0.997	0.998	0.988	0.991	0.995	1.00	1.00	1.00	1.00	1.00

Size of Unit Root Tests;Moving-Average Case, $T = 200$, $k_{\max} = 12$

α	θ	$t_{sig}(10)$	$t_{sig}(5)$	AIC	BIC
1.0	.80	.056	.060	.059	.063
1.0	.50	.061	.064	.056	.064
1.0	.30	.061	.064	.061	.066
1.0	.00	.064	.066	.059	.057
1.0	-.30	.067	.076	.076	.102
1.0	-.50	.085	.110	.121	.168
1.0	-.80	.177	.250	.366	.557

Size and Power of Unit Root Tests;Moving-Average Case, $T = 100$, $k_{\max} = 10$

α	θ	$t_{sig}(10)$	$t_{sig}(5)$	AIC	BIC
1.0	.80	.069	0.73	.068	.071
1.0	.50	.083	.087	.082	.088
1.0	.30	.075	.077	.070	.069
1.0	.00	.063	.059	.052	.046
1.0	-.30	.097	.126	.127	.174
1.0	-.50	.116	.158	.167	.244
1.0	-.80	.304	.424	.561	.733
.95	.80	.136	.151	.146	.158
.95	.50	.164	.184	.170	.196
.95	.30	.158	.162	.152	.144
.95	.00	.153	.151	.140	.126
.95	-.30	.228	.292	.294	.393
.95	-.50	.254	.336	.377	.510
.95	-.80	.534	.704	.877	.963
.85	.80	.347	.387	.405	.451
.85	.50	.445	.510	.520	.586
.85	.30	.456	.513	.536	.505
.85	.00	.486	.540	.580	.575
.85	-.30	.555	.682	.758	.859
.85	-.50	.627	.753	.860	.936
.85	-.80	.825	.908	.996	1.000

1. With AIC and BIC, the size distortions are still very large in the negative MA case (especially BIC) and remain large even at $T = 200$.

2. The size distortions with information criteria arise because the chosen value of k is too small. Furthermore, k increases very slowly with T at a rate proportional to $\log(T)$. Recall that this rate was forbidden by the assumptions of Said-Dickey in their proof which is why Ng-Perron (1995) relaxed the lower bound they had imposed to show the asymptotic validity of the tests with k selected using an information criterion.

3. Size distortions are smaller with t_{sig}, especially $t_{sig}(10)$ with $T = 200$.

4. But in all cases, we are still far from the nominal 5% target.

Summary: detecting and addressing stochastic trends

1. The random walk model is the workhorse model for trends in economic time series data

2. To determine whether Y_t has a stochastic trend, first plot Y_t. If a trend looks plausible, compute the DF test (decide which version, intercept or intercept + trend)

3. If the DF test fails to reject, conclude that Y_t has a unit root (random walk stochastic trend)

4. If Y_t has a unit root, use ΔY_t for regression analysis and forecasting. If no unit root, use Y_t.

John Y. Campbell & Pierre Perron (1991) :
Pitfalls and Opportunities: What Macroeconomists Should Know about Unit Roots

We use the notation DV_t (*deterministic variables*) for the set of variables that appears in the deterministic trend under the maintained DGP. In most applications $DV_t = \{1\}$, a constant, or $DV_t = \{1, t\}$, allowing a first-order polynomial in t. However DV_t can be more complicated; for example, the nonlinear structural model with a deterministic change in the intercept at date T_B has $DV_t = \{1, t, 1(t > T_B)\}$. Since we are interested in the properties of the noise function, a natural strategy is first to "detrend" the series and analyze the time series behavior of the estimated residuals. We use the notation $\tilde{y}_t$ for the residuals of a projection of y_t on a set of *deterministic regressors* DR_t.

Rule 1: Suppose that the deterministic regressors DR_t used to construct $\tilde{y}_t$ in (2.7) contain at least the deterministic variables DV_t included in the maintained data generating process. Then under the null hypothesis of a unit root, the asymptotic distribution of $t_{\hat{\pi}}$ is nonnormal and varies with the set DR_t. In the case where the maintained DGP has a linear trend, the same result holds for regression Equation (2.8) when the deterministic regressors DR_t^* include at least the variables DV_t.

Critical values for the asymptotic distribution of $t_{\hat{\pi}}$ can be found in the following sources for different sets of included deterministic regressors. For $DR_t = \{\emptyset\}$, $\{1\}$ or $\{1, t\}$, see Fuller (1976); for $DR_t = \{1, t, t^2, \ldots, t^p; p = 2, \ldots, 5\}$, see Ouliaris, Park, and Phillips (1989); for $DR_t = \{1, 1(t > T_B)\}$, see Perron (1990a); for $DR_t = \{1, t, 1(t > T_B)\}$, $\{1, t, (t - T_B)1(t > T_B)\}$ and $\{1, t, 1(t > T_B), t\,1(t > T_B)\}$, see Perron

Rule 2: Under the null hypothesis of a unit root, the left-tailed critical values of the asymptotic distribution of $t_{\hat{\pi}}$ increase in absolute value with the number of included deterministic regressors.

Rule 3: Suppose that DR_t omits a variable in DV_t that is growing at a rate at least as fast as any of the elements of DR_t. Then under the null hypothesis of a unit root, the statistic $t_{\hat{\pi}}$ in (2.7) can be normalized in such a way that its asymptotic distribution is standard normal. In the case where the maintained DGP has a linear trend, a similar result describes the set of regressors DR_t^* and the distribution of $t_{\hat{\pi}}$ in the one-step regression (2.8).

The **power** of a significance test measures its ability to detect an alternative hypothesis.

The power against a specific alternative is calculated as the probability that the test will reject H_0 when that specific alternative is true.

Rule 4: (1) Assume that DR_t omits a variable in DV_t that is growing at a rate at least as fast as any of the elements of DR_t. Then the power of the statistic $t_{\hat{\pi}}$ in (2.7) goes to zero as the sample size increases. (2) Suppose that DR_t fails to include a variable in DV_t that is nontrending (e.g., a mean or a change in mean). Then $t_{\hat{\pi}}$ in (2.7) is a consistent test but the finite sample power is adversely affected and decreases as the coefficient on the omitted component increases. Similar results apply to the set of regressors DR_t^* in the one-step procedure (2.8).

Rule 5: Suppose that $t_{\hat{\pi}}$ is constructed using a set of deterministic regressors, DR_t, that includes at least all the deterministic components under the relevant DGP. The power of a test of the unit root hypothesis against stationary alternatives decreases as additional deterministic regressors are included.

Rule 6: A nonrejection of the unit root hypothesis may be due to misspecification of the deterministic components included as regressors.

Rule 7: In finite samples, any trend-stationary process can be approximated arbitrarily well by a unit root process (in the sense that the autocovariance structures will be arbitrarily close).

Rule 9: In finite samples, any unit root process can be approximated arbitrarily well by a trend-stationary process (in the sense that the autocovariance structures will be arbitrarily close).

Summary: detecting and addressing stochastic trends

1. The random walk model is the workhorse model for trends in economic time series data

2. To determine whether Y_t has a stochastic trend, first plot Y_t. If a trend looks plausible, compute the DF test (decide which version, intercept or intercept + trend)

3. If the DF test fails to reject, conclude that Y_t has a unit root (random walk stochastic trend)

4. If Y_t has a unit root, use ΔY_t for regression analysis and forecasting. If no unit root, use Y_t.

Cointegration

Suppose that two time series are each *integrated*, i.e. have unit roots, and hence moving average representations

$$(1-L)y_t = a(L)\delta_t$$

$$(1-L)w_t = b(L)\nu_t$$

In general, linear combinations of y and w also have unit roots. However, if there is some linear combination, say $y_t - \alpha w_t$, that is stationary, y_t and w_t are said to be *cointegrated*, and $[1 - \alpha]$ is their *cointegrating vector.*

Suppose y_t and w_t are cointegrated, so that $y_t - \alpha w_t$ is stationary. Now, consider running

$$y_t = \beta w_t + u_t$$

by OLS. OLS estimates of β converge to α, even if the errors u_t are correlated with the right hand variables w_t!

As an example, consider the textbook simultaneous equations problem:

$$y_t = c_t + a_t$$

$$c_t = \alpha y_t + \epsilon_t$$

a_t is a shock ($a_t = i_t + g_t$); a_t and ϵ_t are iid and independent of each other. If you estimate the c_t equation by OLS you get biased and inconsistent estimates of α. To see this, you first solve the system for its reduced form,

$$y_t = \alpha y_t + \epsilon_t + a_t = \frac{1}{1-\alpha}(\epsilon_t + a_t)$$

$$c_t = \frac{\alpha}{1-\alpha}(\epsilon_t + a_t) + \epsilon_t = \frac{1}{1-\alpha}\epsilon_t + \frac{\alpha}{1-\alpha}a_t$$

Then,

$$\hat{\alpha} \to \frac{plim(\frac{1}{T}\sum c_t y_t)}{plim(\frac{1}{T}\sum y_t^2)} = \frac{plim(\frac{1}{T}\sum(\alpha y_t + \epsilon_t)y_t)}{plim(\frac{1}{T}\sum y_t^2)} = \alpha + \frac{plim(\frac{1}{T}\sum \epsilon_t y_t)}{plim(\frac{1}{T}\sum y_t^2)}$$

$$= \alpha + \frac{\frac{\sigma_\epsilon^2}{1-\alpha}}{\frac{\sigma_\epsilon^2+\sigma_a^2}{(1-\alpha)^2}} = \alpha + (1-\alpha)\frac{\sigma_\epsilon^2}{\sigma_\epsilon^2 + \sigma_a^2}$$

As a result of this bias and inconsistency a lot of effort was put into estimating "consumption functions" consistently, i.e. by 2SLS or other techniques.

$$y_t = c_t + a_t$$

$$c_t = \alpha y_t + \epsilon_t$$

$$\hat{\alpha} \rightarrow \frac{plim(\frac{1}{T}\sum c_t y_t)}{plim(\frac{1}{T}\sum y_t^2)} = \frac{plim(\frac{1}{T}\sum(\alpha y_t + \epsilon_t)y_t)}{plim(\frac{1}{T}\sum y_t^2)} = \alpha + \frac{plim(\frac{1}{T}\sum \epsilon_t y_t)}{plim(\frac{1}{T}\sum y_t^2)}$$

$$= \alpha + \frac{\frac{\sigma_\epsilon^2}{1-\alpha}}{\frac{\sigma_\epsilon^2+\sigma_a^2}{(1-\alpha)^2}} = \alpha + (1-\alpha)\frac{\sigma_\epsilon^2}{\sigma_\epsilon^2 + \sigma_a^2}$$

Now, suppose instead that $a_t = a_{t-1} + \delta_t$. This induces a unit root in y and hence in c as well. But since ϵ is still stationary, $c - \alpha y$ is stationary, so y and c are cointegrated. Now $\sigma_a^2 \rightarrow \infty$, so $\text{plim}(\frac{1}{T}\sum y_t^2) = \infty$ and $\hat{\alpha} \rightarrow \alpha$! Thus, none of the 2SLS etc. corrections are needed!