Predictive Regression

Financial Econometrics (M524) Lecture Notes

Ke-Li Xu

Indiana University

September 20, 2022

Contents

- Predictive regression model
- Different definitions of stock returns
- OLS is biased
- Three methods of correcting the bias
  - An analytic formula (Stambaugh's estimator)
  - Jackknife method
  - Bootstrap method
- Inference: the bootstrap method
- A simulation study
- Application 1: log return vs. log dividend-price-ratio (US stock market 1945-2015)
- Application 2: Manager sentiment

Predictive regression model

- Many empirical studies in economics and finance investigate the following *predictive regression*:

$$y_t = \mu_y + \beta x_{t-1} + u_t,$$

where y_t reflects a change in an asset's price (simple return, log return, excess return), x_{t-1} is a lagged variable related to asset prices.

E.g. a fundamental issue in finance is whether future stock returns are predictable using *publicly available* information.

y_t	x_{t-1}
return on a portfolio of common stocks	dividend yield, or book-to-price ratio
return on a bond portfolio	interest rate, yield spread, or forward rate
the change in the spot rate of exchange	the spread between the forward and spot exchange rates

Testing the hypothesis $\beta = 0$

- Until the early 1980s, the standard model assumed *constant expected returns* for stocks, i.e. $\beta = 0$.
- Then, empirical evidence was uncovered showing that returns were predictable by financial ratios, such as the price-dividend or price-earnings ratio.
- Later other variables were also shown to have predictive ability.
- Expanded interest to returns on *other* asset classes, such as government bonds, currencies, real estate, and commodities, and to many countries.

Estimation of β

- A large partial equilibrium literature takes *time variation in expected returns* ($E(y_t|\mathcal{I}_{t-1}) = \mu_y + \beta x_{t-1}$) as given and asks how it affects *optimal asset allocation* decisions.
- Return predictability is of considerable interest to practitioners who can develop market-timing portfolio strategies that exploit predictability to enhance profits, if indeed such predictability is present.

Econometrics of predictability is not easy

- Despite the theoretical developments, return predictability is a subtle feature of the data.
- A parallel literature developed in the 1990s questioning the strength of the statistical evidence.
- This literature points out problems such as *biased regression coefficients*, *in-sample instability of estimates* indicating periods with and without predictability, and *poor out-of-sample performance*.

Different definitions of stock returns

Simple returns

- The return, defined as a rate, is the profit of holding an asset over a time period (quoted as a proportion of the original asset value).
- Let the asset price at time t be P_t.
- The *simple gross return* is defined as P_t / P_{t-1}.
- The *simple net return* (or simple return) is defined as a percentage change of the price:

$$R_t = P_t / P_{t-1} - 1.$$

Log returns

- The *continuously compounded return* (or log return) is defined as

$$r_t = \log(P_t / P_{t-1}) = \log(1 + R_t).$$

- In other words, r_t is the return such that if it is received continuously over the whole period, the value would increase from P_{t-1} to P_t.
- Suppose that the return is received in m sub-periods (between $t-1$ and t) (each period receives r_t / m), then $P_{t-1}(1 + r_t / m)^m = P_t$.
- That is,

$$P_t / P_{t-1} = (1 + r_t / m)^m.$$

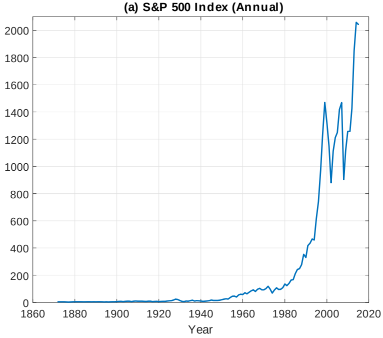
- Letting $m \to \infty$ (the return is received continuously), we have $P_t / P_{t-1} = \exp(r_t)$, which gives the definition of the continuously compounded return r_t.
- For the simple return R_t, compounding takes place only at the end of the period, i.e. $m = 1$, (unlike r_t).

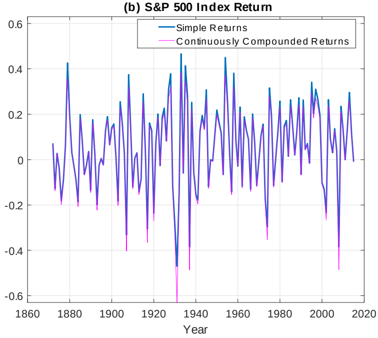
- For example, the S&P 500 index this month (Sep 1, 2016) is 2128, and the index last month is 2170. Then

$$\begin{aligned} R_t &= -1.935\%, \\ r_t &= -1.954\%, \end{aligned}$$

  which are pretty close.
- In general, R_t and r_t are close if R_t is near zero.
- This is because $\log(1+x) \approx x$ when x is close to zero. (Almost always $R_t > r_t$).
- Use the approximation $\log(1+x) \approx x - \frac{x^2}{2} + \frac{x^3}{3} - ...$ for $|x| < 1$.

- This figure shows R_t and r_t for annual S&P 500 index data from 1872-2015.





- When $|R_t|$ is large, there is some non-negiligible difference between R_t and r_t. Otherwise they are very close.

Multi-period returns

- For the return of holding for k periods, from t to $t+k-1$, the continuously compounded multi-period return (denoted as $r_t[k]$) solves

$$P_{t+k-1}/P_{t-1} = \exp(r_t[k]),$$

by definition.

- So $r_t[k] = \log(P_{t+k-1}/P_{t-1})$.
- Note that

$$r_t[k] = r_{t+k-1} + r_{t+k-2} + ... + r_t,$$

which is the sum of returns for single periods.

- This is a convenient property. Multi-period returns using R_t do not enjoy this property.

Returns including dividends

- So far we assume the asset does not pay dividends.
- If it pays dividends, the *simple return* is defined as

$$R_t = (P_t + D_t)/P_{t-1} - 1.$$

- The *continuously compounded return* (including dividends) is defined as

$$r_t = \log(P_t + D_t) - \log P_{t-1}.$$

- The dividend-price ratio is defined as D_t/P_t.
- A useful predictor for return is the logarithm of dividend-price ratio:

$$dp_t = \log(D_t/P_t).$$

OLS is biased

- Consider the following model $(t = 1, 2, ..., T)$

$$
\begin{aligned}
y_t &= \mu_y + \beta x_{t-1} + u_t \qquad (1)\\
x_t &= \mu_x + \rho x_{t-1} + v_t \qquad (2)
\end{aligned}
$$

where

$$\begin{pmatrix} u_t \\ v_t \end{pmatrix} \overset{iid}{\sim} N(0, \Sigma), \text{ with } \Sigma = \begin{pmatrix} \sigma_u^2 & \sigma_{uv} \\ \sigma_{uv} & \sigma_v^2 \end{pmatrix}_{2\times 2}.$$

- Denote the correlation between two errors: $\delta = \sigma_u^{-1}\sigma_v^{-1}\sigma_{uv}$.
- The main interest is in β.
- $\mu_y + \beta x_{t-1}$ is interpreted as expected return.
- We will also test $H_0 : \beta = 0$.
  - $\beta \neq 0$ means *predictability* of y_t (using x_{t-1}), or time varying expected return.
  - $\beta = 0$ means non-predictability, or time-invariant expected return.

Why is OLS biased?

- OLS of Eqn (1), $\widehat{\beta}$, is *biased* since u_t is correlated with x_t, a future value of the regressor x_{t-1}.
- Recall that in the standard OLS theory, unbiasdness of $\widehat{\beta}$ requires u_t to be independent of $x_1, ..., x_T$, the entire path of $\{x_t\}$, not only the concurrent value of x_t. This requirement is called *strict exogeneity*, which is violated in eqn (1).
- Normality of the errors is useful to examine the bias property.

- To make the statement more precise, assuming $\mu_x = 0$, we have

$$x_t = v_t + \rho v_{t-1} + \rho^2 v_{t-2} + ...$$

- Thus

$$Cov(u_t, x_t) = Cov(u_t, v_t) = \sigma_{uv}.$$

- So if $\sigma_{uv} \neq 0$, we have $Cov(u_t, x_t) \neq 0$, thus $\widehat{\beta}$ is biased.

- Three methods of correcting the bias
  - An analytic formula (Stambaugh's)
  - Jackknife method
  - Bootstrap method

Stambaugh Estimator

- Now we introduce a bias-reduced estimator.
- Stambaugh estimator:

$$\widehat{\beta}_{bc} = \widehat{\beta} + \widehat{\phi}\frac{1+3\widehat{\rho}}{T}, \tag{3}$$

  where $\widehat{\phi} = \sum_{t=2}^{T} \widehat{u}_t \widehat{v}_t \,/\, \sum_{t=2}^{T} \widehat{v}_t^2$.
- The right-side of (3) is called *Stambaugh's (1999) bias-corrected estimator*.
- Stambaugh estimator needs to run both regressions (1)-(2).

What does the bias look like?

- We will show that
$$E(\widehat{\beta}-\beta)=\phi E(\widehat{\rho}-\rho), \tag{4}$$
where $\phi=\sigma_v^{-2}\sigma_{uv}$. (different from δ)
- This formula gives some sense of the bias of $\widehat{\beta}$.
- First of all, if $\phi=0$, then $\widehat{\beta}$ is *unbiased*.
- In practice, often $\phi\neq 0$.
- In most applications we will see $\rho>0$, and we will show $E(\widehat{\rho}-\rho)<0$. (In other words, $E\widehat{\rho}<\rho$, or $\widehat{\rho}$ is biased downwards.)
- Then if $\phi<0$, then $E(\widehat{\beta}-\beta)>0$. (In other words, $E\widehat{\beta}>\beta$, or $\widehat{\beta}$ is biased upwards.)

Justification of $\widehat{\beta}_{bc}$

- We now justify the Stambaugh estimator, as defined in (3).
- From (4), the bias of $\widehat{\beta}$ is determined by the bias of $\widehat{\rho}$. Thus we want a less biased estimator $\widehat{\rho}$.
- An unbiased (bias-corrected) estimator should look like $\widehat{\beta} - E(\widehat{\beta} - \beta)$, which is (by (4))
$$\widehat{\beta} - \phi E(\widehat{\rho} - \rho).$$
- We will use
$$E(\widehat{\rho} - \rho) = -\frac{1 + 3\rho}{T} + O(T^{-2}). \tag{5}$$
- This formula is the well-known Kendall's expression (in statistics literature, which won't derive) for the bias of the OLS estimator of the AR(1) model.
- The Stambaugh estimator $\widehat{\beta}_{bc}$ is then obtained from (4) and (5).

Why (4) holds?

- Linear projection
$$u_t = \phi v_t + e_t, \tag{6}$$
where $\phi = \sigma_{uv}/\sigma_v^2$.
- Let $\widetilde{x}_{t-1} = x_{t-1} - (n-1)^{-1}\sum_{t=2}^{T} x_{t-1}$. (demeaning)
- OLS

$$\begin{aligned}
\widehat{\beta} - \beta &= \Big(\sum_{t=2}^{T} \widetilde{x}_{t-1}^2\Big)^{-1} \sum_{t=2}^{T} \widetilde{x}_{t-1} u_t \\
&\overset{(6)}{=} \underbrace{\Big(\sum_{t=2}^{T} \widetilde{x}_{t-1}^2\Big)^{-1} \sum_{t=2}^{T} \widetilde{x}_{t-1} e_t}_{=S} + \phi \underbrace{\Big(\sum_{t=2}^{T} \widetilde{x}_{t-1}^2\Big)^{-1} \sum_{t=2}^{T} \widetilde{x}_{t-1} v_t}_{=\widehat{\rho}-\rho} \\
&= S + \phi(\widehat{\rho} - \rho).
\end{aligned}$$

- Since $ES = 0$, so (4) holds.

Why $ES = 0$?

$$
\begin{aligned}
E[S|x_1, ..., x_T] &= E\left[(\textstyle\sum_{t=2}^T \widetilde{x}_{t-1}^2)^{-1} \textstyle\sum_{t=2}^T \widetilde{x}_{t-1} e_t | x_1, ..., x_T \right] \\
&= (\sum_{t=2}^T \widetilde{x}_{t-1}^2)^{-1} \sum_{t=2}^T \widetilde{x}_{t-1} E\left[e_t | x_1, ..., x_T \right] \\
&= 0,
\end{aligned}
$$

since $e_t \perp \{x_1, ..., x_T\}$.

- e_t is independent of $v_1, ..., v_T$
  - $e_t \perp v_t$, because of linear projection and normality.
  - $e_t \perp v_s$, for $s \neq t$, because $\{u_t, v_t\}$ are independent.
- e_t is independent of $x_1, ..., x_T$, because x_t is generated from v_t. $\square$

Application (US stock market 1945-2015): Estimation

- Annual data: $T = 71$
- The effective sample size is 70.
- Log dividend-price ratio dp_t: the most common predictive variable in the literature.
- The indices we use weigh individual stocks by their market capitalization. End-of-year dividends are obtained by summing dividends within the year.
- From 1926, the data are based on the Standard and Poosr's (S&P) 500 index provided by the Center for Research in Security Prices (CRSP). Before 1957, this was actually the S&P 90.
- E.g. for 2015, the price at the end of the year is 2043.94 (12/28/2015). Matched with Yahoo Finance.

Standard calculations:

$$
\begin{aligned}
R_t &= \frac{P_t + D_t}{P_{t-1}} - 1 \\
DP_t &= \frac{D_t}{P_t} \\
r_t &= \log(1 + R_t) \\
dp_t &= \log(DP_t)
\end{aligned}
$$

- The predictive regression

$$ret_t = \mu_y + \beta dp_{t-1} + u_t$$

- The estimator of β

$$\begin{aligned} \widehat{\beta} &= 9.2\% \\ \widehat{\beta}_{bc} &= 4.5\% \text{ (Stambaugh)} \\ \widehat{\beta}_{jk2} &= -0.4\% \text{ (Jackknife with } m = 2) \\ \widehat{\beta}_{boot-bc} &= 3.1\%. \end{aligned}$$

- Other parameters:

$$\begin{aligned} \widehat{\rho} &= 0.92 \\ \widehat{\delta} &= -0.91 \\ \widehat{\phi} &= -0.88. \end{aligned}$$

- $\widehat{\beta}_{bc}$ (Stambaugh's) and $\widehat{\beta}_{jk2}$ (Jackknife, introduced later) are computed by `pred_bias_correction.m`
- $\widehat{\beta}_{boot-bc}$ (bootstrap-based bias correction, introduced later) is computed by `boot_naive_2.m` (together with bootstrap tests)

Jackknife estimator

- Stambaugh's estimator relies on three key assumptions:
  - 1. x_t is AR(1).
    2. Two errors are jointly normal.
    3. Errors are independent over time (ruling out volatility clustering/GARCH effects).
- We now introduce a different bias correction estimator of β, which achieves the same goal, but with less assumptions.
- The jackknife (sample-splitting) estimator: a *robust* approach to bias correction.

- T is the sample size
- Decompose the sample into m consecutive (non-overlapping) subsamples, each with k observations, so that $T = m \times k$.
- Let $\widehat{\beta}$ be the OLS for the whole sample, and $\widehat{\beta}^{(j)}$ be the OLS for the j-th sample ($j = 1, ..., m$).
- The jackknife estimator

$$\widehat{\beta}_{jack} = \frac{m}{m-1}\widehat{\beta} - \frac{\sum_{j=1}^{m} \widehat{\beta}^{(j)}}{m^2 - m}.$$

- We will show that

$$E(\widehat{\beta}_{jack} - \beta) = O(T^{-2}). \tag{7}$$

Why (7) holds?

- We know that

$$\begin{aligned} E(\widehat{\beta}-\beta) &= \phi E(\widehat{\rho}-\rho) = \frac{-\phi(1+3\rho)}{T} + O(T^{-2}) \\ &= \frac{c}{T} + O(T^{-2}), \end{aligned} \tag{8}$$

where $c = -\phi(1+3\rho)$.

- For j-th subsample,

$$\begin{aligned} E(\widehat{\beta}^{(j)}-\beta) &= \frac{c}{k} + O(T^{-2}) = \frac{c}{T/m} + O(T^{-2}) \\ m^{-2}\sum_{j=1}^{m} E(\widehat{\beta}^{(j)}-\beta) &= \frac{c}{T} + O(T^{-2}). \end{aligned} \tag{9}$$

- The idea is that differencing (8) and (9) can remove the term $\frac{c}{T}$, thus we can estimate β (scaled by some constant) with a bias only of order $O(T^{-2})$.

- Combining (8) and (9),

$$\begin{aligned} E(\widehat{\beta}-\beta) - m^{-2}\sum_{j=1}^{m} E(\widehat{\beta}^{(j)}-\beta) &= O(T^{-2}), \\ E(\widehat{\beta} - m^{-2}\sum_{j=1}^{m}\widehat{\beta}^{(j)}) - (\beta - m^{-1}\beta) &= O(T^{-2}), \\ E(\widehat{\beta} - m^{-2}\sum_{j=1}^{m}\widehat{\beta}^{(j)}) - \frac{m-1}{m}\beta &= O(T^{-2}). \end{aligned}$$

- Thus we can obtain an unbiased estimator of $\frac{m-1}{m}\beta$:

$$E(\widehat{\beta} - m^{-2}\sum_{j=1}^{m}\widehat{\beta}^{(j)}) = \frac{m-1}{m}\beta + O(T^{-2}).$$

So

$$E\underbrace{\left[\tfrac{m}{m-1}(\widehat{\beta} - m^{-2}\textstyle\sum_{j=1}^{m}\widehat{\beta}^{(j)})\right]}_{\widehat{\beta}_{jack}} = \beta + O(T^{-2}).$$

- So (7) holds. $\square$

- Note that an important feature of the jackknife is that, it achieves the bias correction as long as the bias is expressed $E(\widehat{\beta} - \beta) = \frac{c}{T} + O(T^{-2})$, for any c.
- The jackknife method does not require the expression $c = -\phi(1 + 3\rho)$ (which holds under strong assumptions, as shown above for Stambaugh estimator).
- In practice, $m = 2$ works very well. If the time series is reasonably long (e.g. monthly data), we can also use $m = 3$.
- From the derivation above, (approximately) equally division is important (otherwise the jackknife estimator needs to be adjusted).

A simulation study

- The DGP (data generating process), $t = 1, 2, ..., T$,

$$\begin{aligned} y_t &= \mu_y + \beta x_{t-1} + u_t \\ x_t &= \mu_x + \rho x_{t-1} + v_t \end{aligned}$$

  where

$$\begin{pmatrix} u_t \\ v_t \end{pmatrix} \overset{iid}{\sim} N(0, \Sigma), \text{ with } \Sigma = \begin{pmatrix} 1 & \delta \\ \delta & 1 \end{pmatrix}_{2\times 2}.$$

- Set $\mu_y = \mu_x = 0$.
- Set $\beta = 0$.
- $\rho \in \{0.5, 0.9\}$.
- $\delta \in \{-0.5, -0.9\}$.
- $T = 100$.
- Initialization: $x_0 = 0$. I generate 600 observations, and the first 500 observations are dropped.

The Matlab function which runs the simulations:

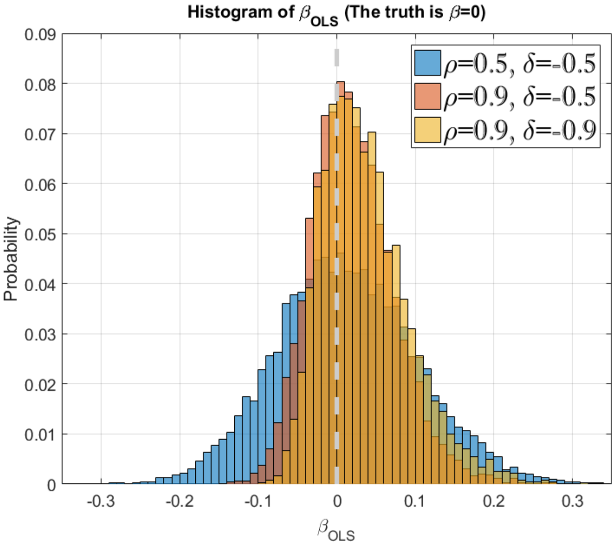
```
function [b_ols,b_bc,b_jk2,t_ols]=dgp1(b,rho,del,n,rep)
```

$\rho = 0.5$ $\delta = -0.5$	OLS	OLS_bc	OLS_jk2
mean	0.0127	0.0006	-0.0002
STD	0.0896	0.0903	0.0945

$\rho = 0.9$ $\delta = -0.5$	OLS	OLS_bc	OLS_jk2
mean	0.0201	0.0022	-0.0008
STD	0.0529	0.0534	0.0646

$\rho = 0.9$ $\delta = -0.9$	OLS	OLS_bc	OLS_jk2
mean	0.0366	0.0044	-0.0023
STD	0.0558	0.0572	0.0723

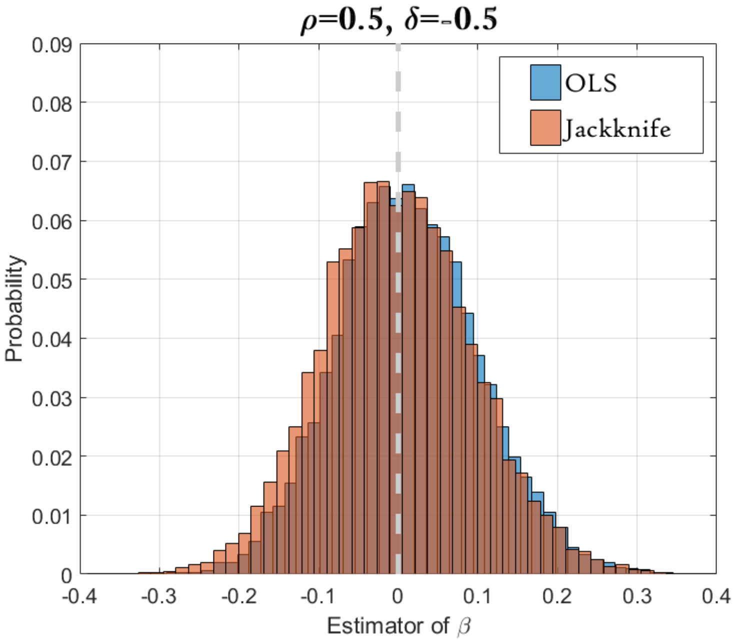
The true value $\beta = 0$.



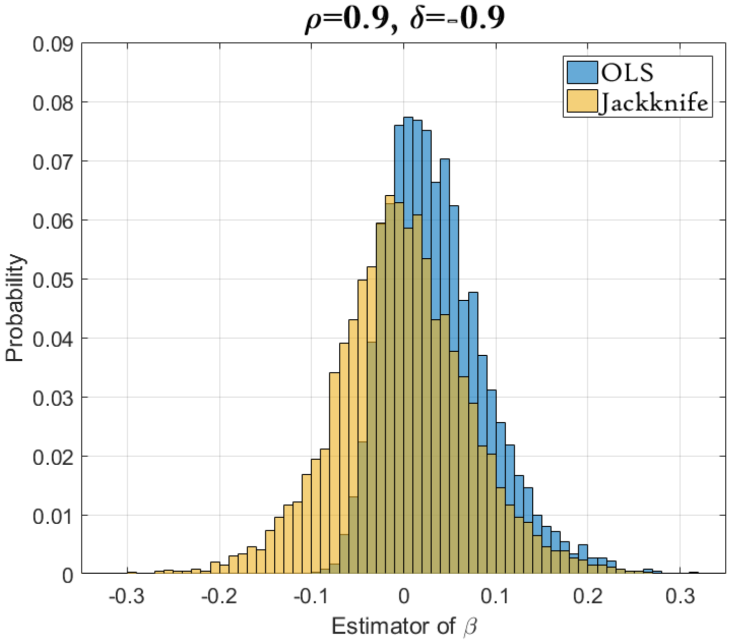
`histogram(b_ols_05,'Normalization','probability')`

10k replications

- In the jargon of asymptotic theory, the convergence rate is faster when ρ is closer to one.
- But in finite samples, does the quality of $\widehat{\beta}$ really get better when ρ increases?
- There is a trade-off.
  - The SE gets smaller (good).
  - The bias gets larger (bad).
  - Asymptotic theory says the bias (of order T^{-1}) is less important than the SE (of order $T^{-1/2}$).



ρ=0.5, δ=-0.5
OLS
Jackknife
Probability
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
-0.4
-0.3
-0.2
-0.1
0
0.1
0.2
0.3
0.4
Estimator of β



ρ=0.9, δ=-0.9
OLS
Jackknife
Probability
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
-0.3
-0.2
-0.1
0
0.1
0.2
0.3
Estimator of β

Inference: Bootstrap method

- We now consider hypothesis testing in the predictive regression model

$$y_t = \mu_y + \beta x_{t-1} + u_t.$$

- The standard t-statistic for $H_0 : \beta = 0$ is

$$t_0 = \frac{\widehat{\beta}}{\widehat{\sigma}_u \left(\sum_{t=2}^{T} \widetilde{x}_{t-1}^2 \right)^{-1/2}}, \tag{10}$$

where $\widehat{\sigma}_u^2 = (T-1)^{-1} \sum_{t=2}^{T} \widehat{u}_t^2$.

- Standard t-test still works (no need to adjust for serial correlation; Newey-West type t-test is not needed), since $\{x_{t-1} u_t\}$ is an uncorrelated sequence.
- E.g.

$$E(x_{t-2} u_{t-1} \cdot x_{t-1} u_t) = E u_t \cdot E(x_{t-2} u_{t-1} x_{t-1}),$$

since (assuming $\mu_x = 0$), $x_t = v_t + \rho v_{t-1} + \rho^2 v_{t-2} + \ldots$

- Thus $x_{t-2} u_{t-1} x_{t-1}$ only depends on the information up to time $t-1$ (i.e. depends on $u_{t-1}, v_{t-1}, u_{t-2}, v_{t-2}, \ldots$).

Bootstrap method

- Generate pseudo data $\{x_t^*, y_t^* : t = 1, ..., T\}$ many times (B times, say $B = 500$).
- For each pseudo dataset ($b = 1, ..., B$), compute the OLS estimator $\widehat{\beta}^{(b)}$, and the t-stat

$$t_0^{(b)} = \frac{\widehat{\beta}^{(b)} - \widehat{\beta}}{\widehat{\sigma}_u^* (\sum_{t=2}^T \widetilde{x}_{t-1}^{*2})^{-1/2}}, \tag{11}$$

  where $\widehat{\sigma}_u^{*2} = (T-1)^{-1} \sum_{t=2}^T \widehat{u}_t^{*2}$.
- Then we have B bootstrap slope coefficients $\{\widehat{\beta}^{(b)} : b = 1, ..., B\}$, and B bootstrap t-stats $\{t_0^{(b)} : b = 1, ..., B\}$.
- Be careful that $t_0^{(b)}$ is centered at $\widehat{\beta}$ (not zero), which is the original OLS.

Two purposes of bootstrapping

- 1. Bootstrap bias correction: Use $B^{-1}\sum_{b=1}^{B}\widehat{\beta}^{(b)}-\widehat{\beta}$ to estimate the bias.
  - Thus the bias-corrected estimator of β is

$$\widehat{\beta}-[B^{-1}\sum_{b=1}^{B}\widehat{\beta}^{(b)}-\widehat{\beta}].$$

- 2. Bootstrap test: Use the distribution of $\{t_0^{(b)}:b=1,...,B\}$ to approximate the distribution of t_0 (e.g. compute the critical values and p-values), *instead of N(0,1)*, as in the standard approach.
  - 5% right-tail critical value is the 95% quantile of $\{t_0^{(b)}:b=1,...,B\}$.
  - The right-tail p-value is

$$\frac{\#\{b:t_0^{(b)}>t_0\}}{B},$$

    where t_0 is the original t-stat in (10).
  - 5% two-tailed critical value is the 95% quantile of $\{|t_0^{(b)}|:b=1,...,B\}$.
  - The two-tailed p-value is

$$\frac{\#\{b:|t_0^{(b)}|>|t_0|\}}{B}.$$

How to generate pseudo data?

Pseudo data $\{x_t^*, y_t^* : t = 1, ..., T\}$ are generated using the model (1)-(2). Let $\widehat{\mu}_x, \widehat{\rho}, \widehat{\mu}_y, \widehat{\beta}$ be OLS estimators, and $\widehat{u}_t$ and $\widehat{v}_t$ be OLS residuals, from the original data.

- Step 1: For $t = 2, ..., T$, generate

$$\begin{pmatrix} u_t^* \\ v_t^* \end{pmatrix} = \begin{pmatrix} \widehat{u}_t \\ \widehat{v}_t \end{pmatrix} \varepsilon_t^*,$$

  where ε_t^* is a random variable with mean 0 and variance 1 (e.g. N(0,1)).

- Step 2: For $t = 2, ..., T$, generate

$$x_t^* = \widehat{\mu}_x + \widehat{\rho} x_{t-1}^* + v_t^*,$$

  where $x_1^* = x_1$, and

$$y_t^* = \widehat{\mu}_y + \widehat{\beta} x_{t-1}^* + u_t^*, \tag{12}$$

  where $y_1^* = y_1$.

A simulation study

The setting is the same as before.

It shows how the bootstrap bias correction works. It has less variance than the jackknife. This can be understood since the bootstrap utilizes the AR(1) structure for the predictor.

$\rho = 0.5$ $\delta = -0.5$	OLS	OLS_bc	OLS_jk2	OLS_boot
mean	0.0127	0.0006	-0.0002	-0.0006
STD	0.0896	0.0903	0.0945	0.0913

$\rho = 0.9$ $\delta = -0.5$	OLS	OLS_bc	OLS_jk2	OLS_boot
mean	0.0201	0.0022	-0.0008	0.0018
STD	0.0529	0.0534	0.0646	0.0545

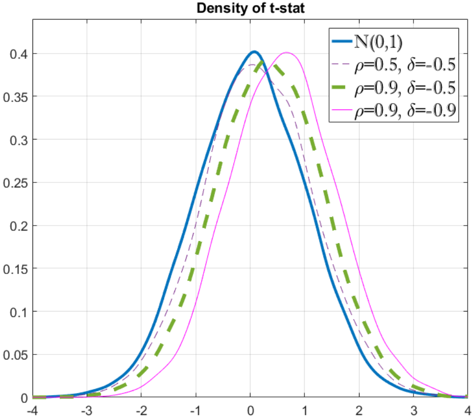
$\rho = 0.9$ $\delta = -0.9$	OLS	OLS_bc	OLS_jk2	OLS_boot
mean	0.0366	0.0044	-0.0023	0.0013
STD	0.0558	0.0572	0.0723	0.0582

The true value $\beta = 0$.

A simulation study

The graph shows the distribution (over 10k replications) of t_0.

N(0,1) would not be a good approximation in some cases. Over-rejection may be expected.




```
>>pltdens(t_ols);
```

A simulation study

The setting is the same as before.

Obvious over-rejection for t_0 (using N(0,1) critical value) on the right tail.

Bootstrap works better.

Rej.Rate 5% (10k replications)	$\rho = 0.5$ $\delta = -0.5$	$\rho = 0.9$ $\delta = -0.5$	$\rho = 0.9$ $\delta = -0.9$
t-test (right)	6.8%	9.5%	12.4%
t-test (left)	4.0%	2.6%	1.3%
Bootstrap (right)	5.7%	6.1%	7.0%
Bootstrap (left)	5.2%	5.3%	6.6%

Application (US stock market 1945-2015): Tests

- The predictive regression

$$
\begin{aligned}
ret_t &= 0.384 + \underset{[2.03]}{0.092} dp_{t-1} + u_t, \\
dp_t &= -0.268 + 0.925 dp_{t-1} + v_t,
\end{aligned}
$$

$$
Var \begin{pmatrix} u_t \\ v_t \end{pmatrix} = \begin{pmatrix} 0.027 & -0.026 \\ -0.026 & 0.029 \end{pmatrix}.
$$

- The t-stat for $\beta = 0$ is 2.03.

Is 2.03 significant?

- The t-stat for $\beta = 0$ is 2.03.
- The following table gives 5% critical values:

	5% Left	5% Right	5% two-sided
N(0,1)	-1.645	1.645	1.96
Bootstrap	-0.78	2.25	2.25

- The following table gives p-values:

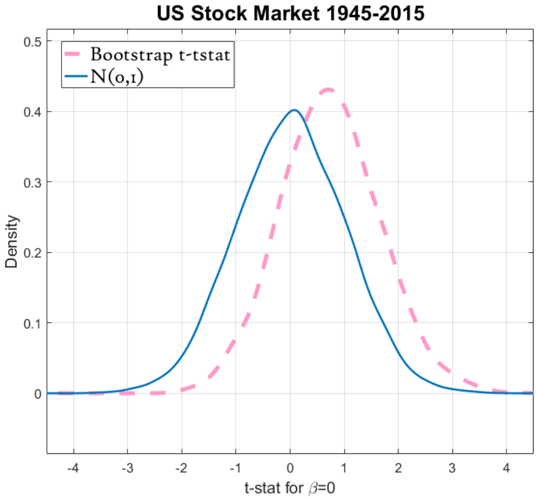
	Left	Right	two-sided
N(0,1)	97.9%	2.1%	4.2%
Bootstrap	91.9%	8.06%	8.12%

	# out of 5000 bootstrap rep.
t_boot<-2.03	3
t_boot>2.03	399
|t_boot|>2.03	402

- MATLAB commend:

```
>>[b_ols,b_boot_bc,t_ols,cv_boot,pv_boot]
=boot_naive_2([0;ret_4],dp_4,0.05);
```

- Be careful of the data input of `boot_naive_2(r,x,level)`: The same row of `r` and `x` represent the same time period (e.g. the first rows are observations of the return and the predictor at period one).
- The first element of `ret_4` is the return *at period two* (the return at period one is not computable).




```
>>pltdens(t_boot);
```

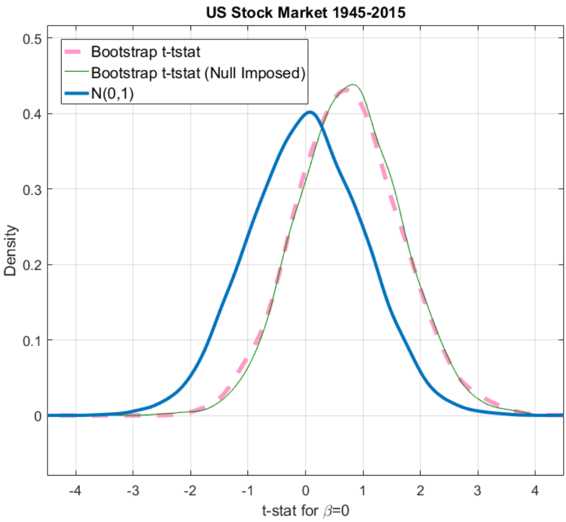
5k bootstrap replications are used.

The pink and blue lines provide two different approximations of the distribution of t_0. Of course, they can yield (very) different conclusion.

- In conclusion, at the usual significance level 5%, the standard t-test *is significant*, but the bootstrap t-test *is not significant*.
- Another interesting finding is that, the bootstrap right-tail p-value (8.06%) does not improve much the bootstrap two-tailed p-value (8.12%).
- In contrast, for the standard t-test, the right-tail p-value (2.1%) is always half of the two-tailed p-value (4.2%).

- Lastly, we mention a different type of bootstrap (*null-imposition*), which is also quite widely used.
- This method imposes the null hypothesis in generating the bootstrap data (thus the construction of bootstrap t-stat is also changed).
- The implementation is very simple, which only needs two changes in the previous algorithm
  - Change (12) to be
$$y_t^* = \widehat{\mu}_y + 0 \cdot x_{t-1}^* + u_t^*. \tag{13}$$
  - Change (11) to be
$$t_0^{(b)} = \frac{\widehat{\beta}^{(b)} - 0}{\widehat{\sigma}_u^* (\sum_{t=2}^T \widetilde{x}_{t-1}^{*2})^{-1/2}}. \tag{14}$$
- Some people like the null-imposition bootstrap because it's simpler.
- Some people like the naive bootstrap because it can also perform bias correction (and it's more general).
- The figure below shows that they provide very similar approximation to the distribution of t-stat.

The naive bootstrap and the null-imposition bootstrap perform similarly.



Application 2: Manager sentiment

- Behavioral finance: speculative market sentiment can lead prices to diverge from their fundamental values.
- Textual disclosure tone-based *manager sentiment index*: it is constructed based on the aggregated textual tone in firm financial statements and conference calls.
- Rationale: a qualitative description of the firm's business and financial performance can reflect managers' subjective opinions and beliefs about recent firm performance, and their expectations for future performance.
- Measure: textual tone is measured as the difference between the number of positive and negative words in the disclosure scaled by the total word count of the disclosure.
- It is an aggregate index (not firm-level), which gauges the overall manager sentiment in the market.

- Monthly data: 2003/01 - 2014/12 ($T = 144$)
- Optimistic behavioral bias from corporate managers *negatively* (and significantly) predicts future market returns.
- Hence, corporate managers (as a whole) tend to be overly optimistic when the economy and the market peak, and the manager sentiment index is a *contrarian* return predictor.
- A high sentiment index leads to market-wide over-valuation, which leads to subsequent low stock returns in the future.
- The dependent variable (xr): *excess market return* (monthly return on the S&P 500 index (including dividends) minus the risk-free rate). Not log returns.
- Predictor (MS): manager sentiment (standardized to have mean zero and variance one).

- The predictive regression

$$\begin{aligned} xr_t &= 0.0076 - \underset{[-3.9]}{0.0126} MS_{t-1} + u_t, \quad \mathrm{R}^2 = 9.74\% \\ MS_t &= 0.0238 + 0.905 MS_{t-1} + v_t, \end{aligned}$$

$$Var \begin{pmatrix} u_t \\ v_t \end{pmatrix} = \begin{pmatrix} 0.0015 & 0.0015 \\ 0.0015 & 0.1549 \end{pmatrix}.$$

- $\delta = cor(u_t, v_t) = 0.097$.
- $\phi = 0.009$. (OLS of u_t on v_t)
- A one-standard-deviation increase in manager sentiment is associated with a -1.26% decrease in the expected excess market return for the next month.
- It is economically significant, comparing with the intercept 0.76%.

- The estimator of β

$$\begin{aligned}
\widehat{\beta} &= -1.26\% \\
\widehat{\beta}_{bc} &= -1.23\% \text{ (Stambaugh)} \\
\widehat{\beta}_{jk2} &= -1.42\% \text{ (Jackknife with } m = 2\text{)} \\
\widehat{\beta}_{boot-bc} &= -1.21\% \text{ (5k bootstraps)}
\end{aligned}$$

- OLS has a negative bias (OLS is biased downwards).
- All corrections are minor.

```
>> [b_ols,b_bc,b_jk2]=pred_bias_correction(er,MS)
%running time:  0.003 seconds

>> [b_ols,b_boot_bc,t_ols,cv_boot,pv_boot]
=boot_naive_2(er,MS,0.05)
%running time:  10 seconds
```

Is -3.9 significant?

- The t-stat for $\beta = 0$ is -3.9.
- The following table gives 5% critical values:

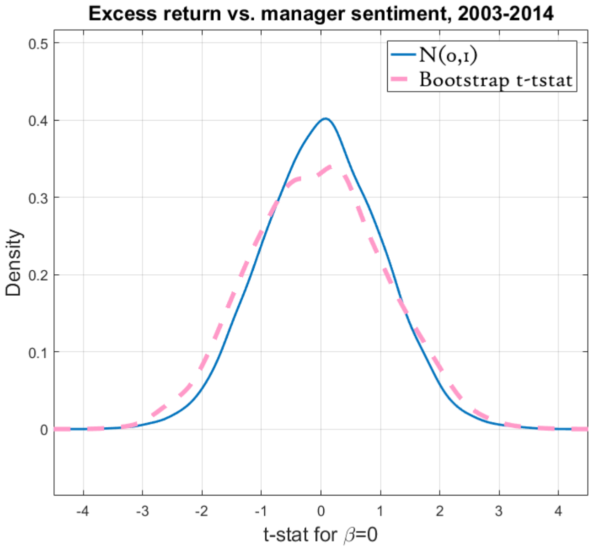
	5% Left	5% Right	**5% two-sided**
N(0,1)	-1.645	1.645	1.96
Bootstrap	-1.881	1.800	2.15

- The following table gives p-values:

	Left	Right	**two-sided**
N(0,1)	0.005%	99.995%	0.01%
Bootstrap	0.00%	100%	0.00%

- Out of 5000 bootstrap replications, there is no single t-tstat such that $|$t_boot$|>$3.9.
- I thus increase the number of bootstrap replications to 50k

	# out of 50k bootstrap rep.	Prob.
t_boot<-3.9	4	0.01%
t_boot>3.9	10	99.99%
$|$t_boot$|>$3.9	14	0.03%




```
>> pltdens(t_boot);
```

- In conclusion, manager sentiment is a highly (negatively) significant predictor for market return.
- Both standard tests and bootstrap tests yield the very similar results. (Thus no sophisticated econometrics is needed.)
- The reason for such similarity is the very low endogeneity of the predictor (*despite its strong persistence*).