Empirical Methods in Finance (MFE230E)
Week 2: Asymptotic OLS, Estimation of AR processes

Martin Lettau

Spring 2019

Haas School of Business

Overview

1. Review of classical OLS
2. Classical OLS model for AR estimation
3. Non-stationarity
4. Spurious regressions

Estimation of ARMA processes

First: Review of OLS regressions

- First consider AR models:

$$z_t = \mu + \phi_1 z_{t-1} + \ldots + \phi_p z_{t-p} + \epsilon_t, \text{ where } \epsilon_t \sim \text{WN}(0, \sigma_\epsilon)$$

$\longrightarrow$ Linear regression

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{e}$$

with z_t as y-variable and $z_{t-1}, z_{t-2}, \ldots$ as x-variables.

- MA processes

$$z_t = \mu + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \ldots + \theta_q \epsilon_{t-q}$$

Question: Can MA processes be estimated by OLS?

Outline

- Reference: MFE Stats pre-program class, Ruppert ch. 12
- Consider the linear model

$$y_i = \mathbf{x}_i' \boldsymbol{\beta} + e_i$$

$$\mathbf{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}$$

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad \boldsymbol{X} = \begin{pmatrix} \mathbf{x}_1' \\ \mathbf{x}_2' \\ \vdots \\ \mathbf{x}_n' \end{pmatrix}, \quad \boldsymbol{e} = \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix}$$

$$\mathbf{D} = \mathrm{E}(\mathbf{e}\mathbf{e}' | \mathbf{X})$$

- $\mathbf{y}$ and $\boldsymbol{e}$ are $n \times 1$ vectors, $\boldsymbol{X}$ is an $n \times k$ matrix, and $\mathbf{D}$ is a $n \times n$ matrix.

NOTATION FOR TIME AR(p) MODEL

$$y_i = \mathbf{x}_i' \boldsymbol{\beta} + e_i \quad i = 1, \ldots, n$$

$$z_t = \mu + \phi_1 z_{t-1} + \ldots + \phi_p z_{t-p} + \epsilon_t, \quad t = 1, \ldots, T$$

$$y_i = z_t$$

$$\mathbf{x}_i = (1, z_{t-1}, \ldots, z_{t-p})'$$

$$\boldsymbol{\beta} = (\mu, \phi_1, \ldots, \phi_p)'$$

$$\mathbf{y} = \begin{pmatrix} z_{p+1} \\ z_{p+2} \\ \vdots \\ z_T \end{pmatrix}, \quad \boldsymbol{X} = \begin{pmatrix} 1 & z_p & \ldots & z_1 \\ 1 & z_{p+1} & \ldots & z_2 \\ & & \vdots & \\ 1 & z_{T-1} & \ldots & z_{T-p} \end{pmatrix}, \quad \boldsymbol{e} = \begin{pmatrix} \epsilon_{p+1} \\ \epsilon_{p+2} \\ \vdots \\ \epsilon_T \end{pmatrix}$$

Assumption 1 (Classical linear regression model).

1. *The model is linear:* $y_i = \mathbf{x}_i' \boldsymbol{\beta} + e_i$.
2. *The observations* $(y_i, \mathbf{x}_i)$ *come from an* ***iid random sample***.
3. ***Strict exogeneity***, $E(e_i \mid \mathbf{x}_j) = 0 \; \forall i, j \Rightarrow E(e \mid \mathbf{X}) = \mathbf{0}$.
4. *The variables have finite second moments:* $Ey_i^2 < \infty$, $E\left\|\mathbf{x}^2\right\| \infty$.
5. *The second moment matrix of* $\mathbf{x}_i$, $\mathbf{Q_{xx}} = E(\mathbf{x}_i \mathbf{x}_i')$ *is invertible.*

Recall the definition of an iid random sample:

1. The distribution of $(y_i, \mathbf{x}_i)$ is the same for all i.
2. $(y_i, \mathbf{x}_i)$ and $(y_j, \mathbf{x}_j)$ are independent for all i, j.

Additional assumptions

Sometimes, we make additional assumptions:

Assumption 2 (Homoskedastic Linear Regression Model).

Additionally,

$$E(e_i^2 \mid \mathbf{x}_i) = \sigma^2(\mathbf{x}_i) = \sigma^2$$

is independent of $\mathbf{x}_i$*. Therefore*

$$\mathbf{D} = \mathbf{I}_n \sigma^2$$

Assumption 3 (Normality).

The errors are (conditionally) normally distributed:

$$\mathbf{e} \mid \mathbf{X} \sim N(0, \mathbf{I}_n \sigma^2)$$

- Random variables $(y, \mathbf{x})$: Conditional expectation function (CEF) $m(\mathbf{x})$:

$$y = m(\mathbf{x}) + e$$

$$\Rightarrow \; \mathrm{E}(e|\mathbf{x}) = 0$$

- Best linear predictor (BLP)

$$y = \mathbf{x}'\beta + e$$

$$\Rightarrow \; \mathrm{E}(\mathbf{x}e) = 0$$

- Implication: Errors e_i are orthogonal to all x_{jk}

$$\mathrm{E}(e_i x_{jk}) = \mathrm{Cov}(e_i, x_{jk}) = 0$$

- **Strict exogeneity**, $\mathrm{E}(e_i \mid \mathbf{x}_j) = 0 \, \forall i, j$ or $\mathrm{E}(e \mid \mathbf{X}) = \mathbf{0}$
- Implication:

$$\begin{aligned} \mathrm{E}(e_i \mid \mathbf{x}_j) &= \mathbf{0} \quad \forall i, j \\ \Rightarrow \mathrm{E}(e_i x_{jk}) &= 0 \quad \forall i, j, k \\ \Rightarrow \mathrm{Cov}(e_i x_{jk}) &= 0 \quad \forall i, j, k \end{aligned}$$

- Matrix notation:

$$\mathrm{E}(\mathbf{X}' \odot \mathbf{e}) = \mathbf{0},$$

  where $\odot$ means element-by-element multiplication
- Intuition: Each error e_i is uncorrelated with each x_{jk}

OLS: Choose $\boldsymbol{\beta}$ to minimize SSE:

$$SSE_n(\boldsymbol{\beta}) = \sum_{i=1}^{n}(y_i - \mathbf{x}_i'\boldsymbol{\beta})^2$$

$$= \sum_{i=1}^{n} y_i^2 - 2\boldsymbol{\beta}\sum_{i=1}^{n}\mathbf{x}_i y_i + \boldsymbol{\beta}'\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i'\boldsymbol{\beta}$$

$$0 = \frac{\partial}{\partial\boldsymbol{\beta}} SSE_n(\boldsymbol{\beta}) = -2\sum_{i=1}^{n}\mathbf{x}_i y_i + 2\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i'\boldsymbol{\beta}$$

Recall that y_i is a scalar and $\mathbf{x}_i$ is a $k \times 1$ vector.

$$\widehat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i'\right)^{-1}\sum_{i=1}^{n}\mathbf{x}_i y_i = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{y})$$

The OLS estimator

Equivalently, the least squares estimator can can be written as a linear projection:

$$\widehat{\boldsymbol{\beta}} = \left(\mathrm{E}(\mathbf{x}\mathbf{x}')\right)^{-1}\mathrm{E}(\mathbf{x}y) = \boldsymbol{Q}_{\mathbf{xx}}^{-1}\boldsymbol{Q}_{\mathbf{xy}}$$

and moment estimators for $\boldsymbol{Q}_{\mathbf{xx}}$ and $\boldsymbol{Q}_{\mathbf{x}y}$:

$$\widehat{\boldsymbol{Q}}_{\mathbf{x}y} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i\mathbf{y}_i = \frac{1}{n}\mathbf{X}'\mathbf{y}$$

$$\widehat{\boldsymbol{Q}}_{\mathbf{xx}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i' = \frac{1}{n}\mathbf{X}'\mathbf{X}$$

OLS **moment condition**: $\mathrm{E}(e|\mathbf{x}) = 0 \Rightarrow \mathrm{E}(\mathbf{x}e) = 0$.

OLS can be derived from this moment condition:

$$\begin{aligned}
\mathrm{E}(\mathbf{x}e) &= 0 \\
\mathrm{E}(\mathbf{x}(y - \mathbf{x}'\beta)) &= 0 \\
\mathrm{E}(\mathbf{x}y) - E(\mathbf{x}\mathbf{x}')\beta &= 0 \\
\beta &= (E(\mathbf{x}\mathbf{x}'))^{-1}\,\mathrm{E}(\mathbf{x}y)
\end{aligned}$$

Estimation: Replace population means with sample means

$$\text{Moment condition: } \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i\widehat{e}_i = \frac{1}{n}\mathbf{X}'\hat{\mathbf{e}} = 0$$

$$\text{Estimator:} \widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

OLS $\widehat{\boldsymbol{\beta}}$ is unbiased:

$$\begin{aligned}\widehat{\boldsymbol{\beta}} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}\\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\boldsymbol{\beta}+\mathbf{e})\\ &= \boldsymbol{\beta}+(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{e}\end{aligned}$$

$$\begin{aligned}\mathrm{E}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\,|\,\mathbf{X}) &= \mathrm{E}\left((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{e}\,|\,\mathbf{X}\right)\\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\ \mathrm{E}(\mathbf{e}\,|\,\mathbf{X}) = 0\end{aligned}$$

$$\mathrm{E}(\widehat{\boldsymbol{\beta}}) = \mathrm{E}\left(\mathrm{E}(\widehat{\boldsymbol{\beta}}\,|\,\mathbf{X})\right) = \boldsymbol{\beta}$$

Note: Unbiasedness is a direct result of the assumption that the errors are **exogenous**: $\mathrm{E}(\mathbf{e}\,|\,\mathbf{X}) = 0$.

$$\widehat{\boldsymbol{\beta}} = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{e}$$

$$\Rightarrow \mathbf{V} = \mathrm{Var}(\widehat{\boldsymbol{\beta}} \mid \mathbf{X}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{D}\mathbf{X}\,(\mathbf{X}'\mathbf{X})^{-1}$$

Theorem 1 (Variance of the Least-Squares Estimator).

In the linear regression model A1:

$$\begin{aligned} \mathbf{V} &= Var(\widehat{\boldsymbol{\beta}} \mid \mathbf{X}) \\ &= (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{D}\mathbf{X})(\mathbf{X}'\mathbf{X})^{-1} \end{aligned}$$

where $\mathbf{D} = E(\mathbf{e}\mathbf{e}' \mid \mathbf{X})$.

In the homoskedastic case (Assumption 2):

$$\mathbf{V} = (\mathbf{X}'\mathbf{X})^{-1}\sigma^2$$

Theorem 2 (Gauss-Markov Theorem: $\widehat{\beta}$ is BLUE).

- *In the **classical homoskedastic linear regression model (Assumptions 1, 2)**, the best (minimum-variance) unbiased linear estimator is the least-squares estimator*

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

- *In the **classical general linear regression model (Assumption 1)**, the best unbiased linear estimator is*

$$\widetilde{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{D}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{D}^{-1}\mathbf{y}$$

The scaled sample sum of squared residual

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\hat{e}_i^2$$

is an unbiased estimator for σ^2.

Therefore, we can estimate the variance-covariance matrix of $\widehat{\boldsymbol{\beta}}$ by

$$\widehat{\boldsymbol{V}}_{\widehat{\boldsymbol{\beta}}} = (\mathbf{X}'\mathbf{X})^{-1}\hat{\sigma}_e^2$$

in the homoskedastic case and by

$$\widehat{\boldsymbol{V}}_{\widehat{\boldsymbol{\beta}}}^{W} = (\mathbf{X}'\mathbf{X})^{-1}\left(\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i'\hat{e}_i^2\right)(\mathbf{X}'\mathbf{X})^{-1}$$

if errors are heteroskedastic. $\widehat{\boldsymbol{V}}_{\widehat{\boldsymbol{\beta}}}^{W}$ is called the **heteroskedasticity-robust White estimator**.

The results so far hold for any distribution of the errors. Note that we need to derive the distribution of $\widehat{\boldsymbol{\beta}}$ for hypothesis testing.

If errors are assumed to be normal, then $\widehat{\boldsymbol{\beta}}$ is also normally distributed:

Theorem 3 (Normal Regression).

Under the assumptions of the homoskedastic linear regression model (Assumptions 1, 2 3), i.e. $e_i \,|\, \mathbf{x}_i \sim N(0, \sigma^2)$, *then*

$$\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} \sim N(\mathbf{0}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1})$$

$$\frac{\hat{\beta}_j - \beta_j}{s(\hat{\beta}_j)} \sim t_{n-k}$$

$$n\frac{\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-k}$$

Outline

1. Review of classical OLS
2. **Classical OLS model for AR estimation**
3. Non-stationarity
4. Spurious regressions

Back to AR processes

Let's consider AR processes first, MAs later

Let's estimate an AR(1) with **normal shocks** by OLS:

$$z_t = \phi z_{t-1} + \epsilon_t, \text{ where } \epsilon_t \sim \text{NWN}(0, \sigma_\epsilon)$$

Set $\phi = 0, 0.5, 0.95, 0.99$, $\sigma_\epsilon = 1$ and simulate 10,000 samples for $T = 50, 100, 1000, 5000$.

Estimate OLS

$$z_t = \mu + \phi z_{t-1} + \epsilon_t$$

for each of the 10,000 simulation and compute the distribution of $\hat{\phi}$.

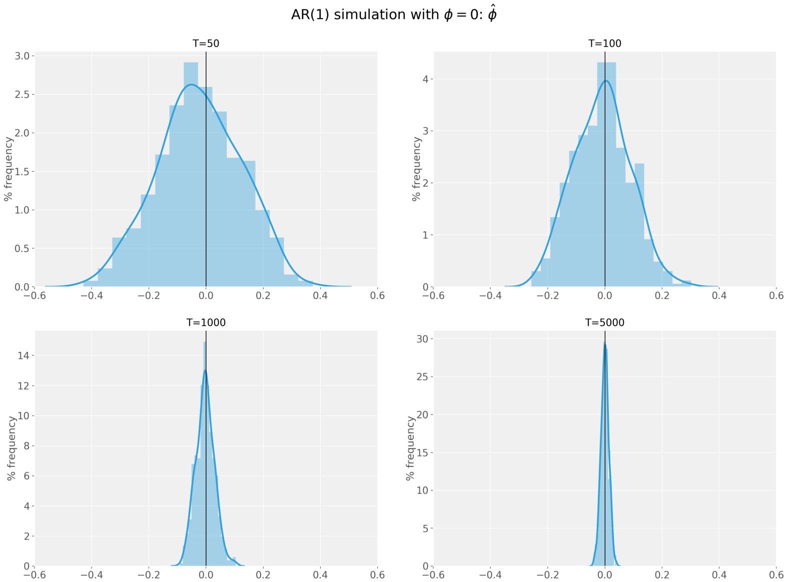
(Note: Inclusion of constant μ is optional but improves small sample behavior.)

Classical OLS with **normal errors**: $\hat{\phi}$ is unbiased and normal:

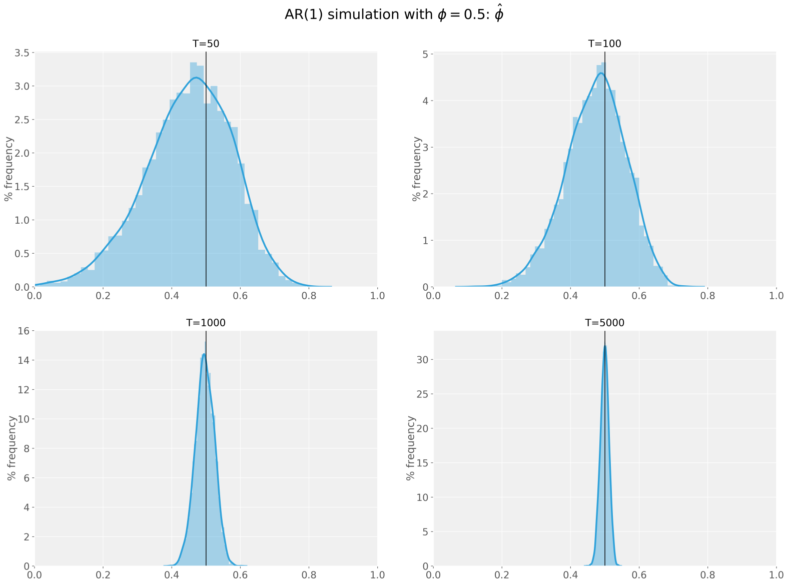
$$\widehat{\phi} - \phi \sim N(0, \sigma_\epsilon^2 (\mathbf{X}'\mathbf{X})^{-1}) = N\left(0, \frac{1-\phi^2}{T}\right)$$

$$\text{since} \quad \mathbf{X}'\mathbf{X} = T\left(\frac{1}{T}\sum_{t=1}^{T} z_t^2\right) = T\,\mathrm{E}(z_t^2) = T\,\frac{\sigma_\epsilon^2}{1-\phi^2}$$

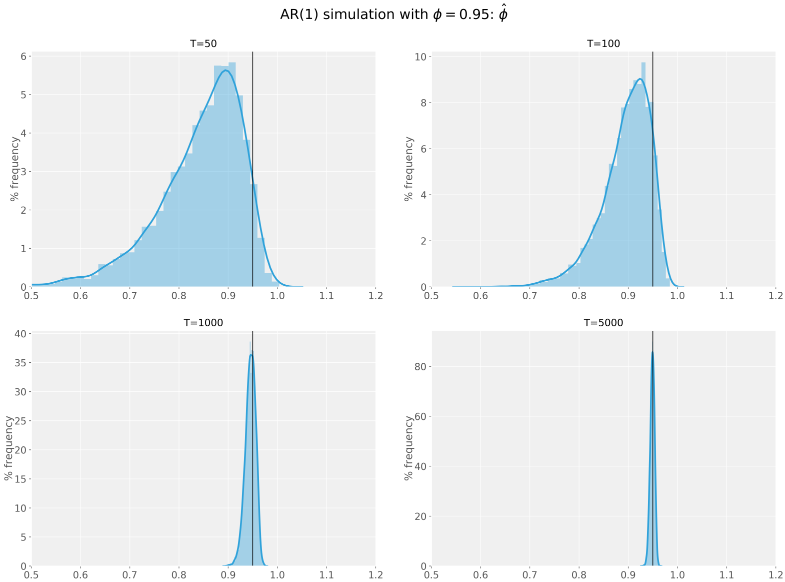
OLS FOR AR(1) WITH $\phi = 0$



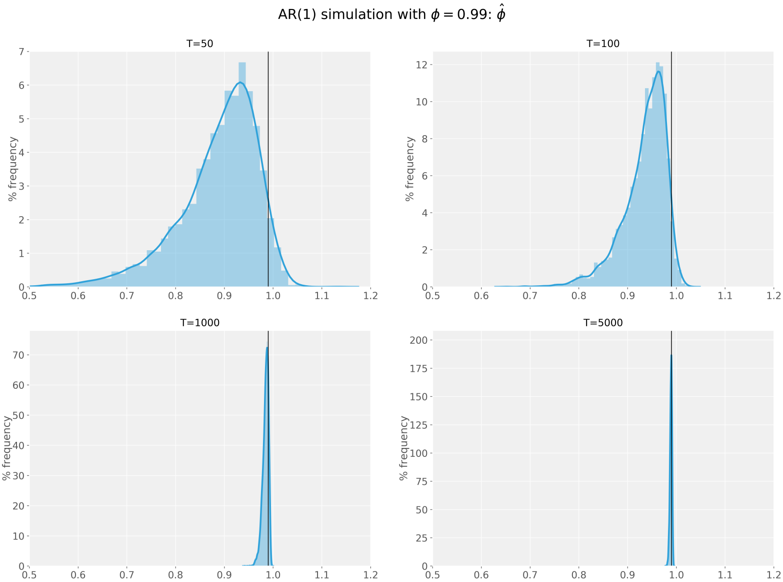
OLS FOR AR(1) WITH $\phi = 0.5$



OLS for AR(1) with $\phi = 0.95$



OLS FOR AR(1) WITH $\phi = 0.99$



Distribution of OLS estimator

The distribution of $\widehat{\boldsymbol{\beta}}$ is not normal even though the errors in the AR(1) process are normal!

- Holding ϕ fixed, $\widehat{\boldsymbol{\beta}}$ looks closer to normal for larger T
- Holding T fixed, $\widehat{\boldsymbol{\beta}}$ looks closer to normal for smaller ϕ
- But didn't we prove that the OLS $\widehat{\boldsymbol{\beta}}$ is normally distributed!?

Mean of OLS estimator

ϕ	$T = 50$	$T = 100$	$T = 1000$	$T = 5000$
0.0	-0.02	-0.01	-0.00	-0.00
0.5	0.45	0.47	0.50	0.50
0.95	0.84	0.90	0.95	0.95
0.99	0.88	0.94	0.98	0.99

- The mean of the estimated $\widehat{\phi}$ is lower than the true ϕ!
- Holding ϕ fixed, the bias is larger for smaller T
- Holding T fixed, the bias is larger for larger ϕ

But didn't we prove that the OLS $\widehat{\boldsymbol{\beta}}$ is unbiased!?

Assumption: The errors are **exogenous**: $\mathrm{E}(\mathbf{e}|\mathbf{X}) = 0$.

Implication: Errors are uncorrelated with $\mathbf{X}$:

$$\mathrm{Cov}(\mathbf{e}, \mathbf{X}) = \mathrm{E}(\mathbf{eX}) - \mathrm{E}(\mathbf{e})\mathrm{E}(\mathbf{X}) = 0$$

$$\mathrm{Cov}(e_i \,|\, x_{jk}) = \mathrm{E}(e_i \,|\, x_{jk}) = 0$$

Thus each error e_i is orthogonal to each x_{jk}

For a time series, $\mathbf{X} = (\ldots, z_{t-1}, z_t, z_{t+1}, \ldots)$, so strict exogeneity implies

$$E(z_t e_s) = 0 \ \ \forall t, s$$

This assumption is violated in time series models!!

$$z_t = \phi z_{t-1} + \epsilon_t$$

$$\begin{aligned} E(z_{t-1}\epsilon_t) &= 0 \\ \text{but} \quad E(z_t\epsilon_t) &= E((\phi z_{t-1} + \epsilon_t)\epsilon_t) \\ &= \phi E(z_{t-1}\epsilon_t) + E(\epsilon_t^2) \\ &= E(\epsilon_t^2) = \sigma^2 \neq 0 \end{aligned}$$

Thus the assumption of strict exogeneity is violated!

Recall: Proof that OLS is unbiased if $\mathrm{E}(\mathbf{e}|\mathbf{X}) = 0$

$$\begin{aligned}
\widehat{\boldsymbol{\beta}} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\boldsymbol{\beta} + \mathbf{e}) \\
&= \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{e}
\end{aligned}$$

Thus

$$\begin{aligned}
\mathrm{E}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}|\mathbf{X}) &= \mathrm{E}\left((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{e}|\mathbf{X}\right) \\
&= \ (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\,\mathrm{E}(\mathbf{e}|\mathbf{X}) = 0 \ \text{ if } \mathrm{E}(\mathbf{e}|\mathbf{X}) = 0
\end{aligned}$$

Important: If $\mathrm{E}(\mathbf{e}|\mathbf{X}) \neq 0$, OLS is no longer unbiased!!

- OLS estimate of AR(1) parameter is biased: **Kendall bias**

$$z_{t+1} = \mu + \phi z_t + e_t$$

$$E\left(\hat{\phi}\right) \approx \phi - \frac{1+3\phi}{T}$$

- OLS is downward biased: understates true persistence

$\phi = 0.95$	$T = 50$	$T = 100$	$T = 1000$	$T = 5000$
Mean of $\hat{\phi}$	0.859	0.906	0.942	0.949
Mean of $\hat{\phi} + \frac{1+3\hat{\phi}}{T}$	0.934	0.943	0.944	0.949

Finite sample Kendall bias

- OLS estimate of AR(1) parameter is biased: **Kendall bias**

$$z_{t+1} = \mu + \phi z_t + e_t$$

$$E\left(\hat{\phi}\right) \approx \phi - \frac{1+3\phi}{T}$$

- OLS is downward biased: understates true persistence

$\phi = 0.95$	$T = 50$	$T = 100$	$T = 1000$	$T = 5000$
Mean of $\hat{\phi}$	0.859	0.906	0.942	0.949
Mean of $\hat{\phi} + \frac{1+3\hat{\phi}}{T}$	0.934	0.943	0.944	0.949

Bias: true mean μ is unknown and is estimated using “future” data, “Proof”: $\mu = \phi = 0$

$$\text{True process: } z_{t+1} = e_{t+1}$$

$$\text{Regression: } z_{t+1} = \mu + \phi z_t + e_{t+1}$$

$$\min_{\hat{\mu},\hat{\phi}} 1/T \sum_t (z_t - [\hat{\mu} + \hat{\phi} z_{t-1}])^2$$

$$\text{FOC: } -1/T \sum_t (z_t - [\hat{\mu} + \hat{\phi} z_{t-1}])(z_{t-1} - \hat{\mu}) = 0$$

$$\Leftrightarrow -1/T \sum_t (e_t - [\hat{\mu} + \hat{\phi} z_{t-1}])(z_{t-1} - \hat{\mu}) = 0$$

$$\Rightarrow \mathrm{E}(\hat{\phi} - \phi) = \mathrm{E}\left[\frac{1/T \sum_t (e_t - \hat{\mu})(z_{t-1} - \hat{\mu})}{1/T \sum_t ((z_{t-1} - \hat{\mu})^2)}\right] < 0$$

$$\text{since } \mathrm{Cov}(e_t - \hat{\mu})(z_{t-1} - \hat{\mu}) < 0$$

Example for $T = 2$:

$$\hat{\mu} = z_1/2 + z_2/ = e_1/2 + e_2/2$$

So if $z_1 < z_2$ then $z_1 - \hat{\mu} < 0$ and $e_2 - \hat{\mu} > 0$

Random sampling in AR(p) processes

- The classical OLS assumption assumes that the sample is i.i.d.
- Independence implies that $\text{Cov}(x_s, x_t) = 0$ for all $s \neq t$.
- This assumption is also violated in (most) financial data.
- Consider again the AR(1) example:

$$z_t = \phi z_{t-1} + e_t$$

$$z_{t+1} = \phi z_t + e_{t+1}$$

$$\Rightarrow \text{Corr}(z_t, z_{t+1}) = \phi$$

- Thus another assumption of the classical OLS model is violated!

OK, we have seen that the classical OLS assumptions are violated in time series models!

What to do?

Next, we will derive an OLS estimator based on weaker assumptions and focus on **asymptotic properties** as the sample size goes to infinity.

Good news: In finance we often have long time series

Not so good news: Some to the material is a bit more technical (this is probably the most difficult part of the class!)

- Classical OLS assumptions are violated in TS models:
  1. Errors are strictly exogenous
  2. $(y_i, \mathbf{x}_i)$ is an i.i.d. random sample
  3. Financial data are usually not normally distributed
- In order to derive properties of the OLS estimator in time series applications, we need to replace these assumptions with weaker assumptions that are satisfied in time series models.

- Recall:

$$\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} = \left(\frac{1}{n}\sum \mathbf{x}_i\mathbf{x}_i'\right)^{-1}\left(\frac{1}{n}\sum \mathbf{x}_i e_i\right)$$

- Instead of deriving the sampling distribution of $\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}$ for a given sample size, we will derive properties of $\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}$ as the sample size increases
- Plan: Under what assumptions are the following limits well-defined?

$$\lim_{n\to\infty} \frac{1}{n}\sum \mathbf{x}_i\mathbf{x}_i' = ?$$

$$\lim_{n\to\infty} \frac{1}{n}\sum \mathbf{x}_i e_i = ?$$

- Focus on the intuition rather than technical details

Weak Law of Large Numbers (WLLN) for random vectors

If the y_i's are *i.i.d.*, and $\mathrm{E}||y|| < \infty$, then

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \xrightarrow{p} \mathrm{E}(y) \quad \text{as} \quad n \longrightarrow \infty$$

Lindeberg-Levy Central Limit Theorem (CLT)

If the y_i's are *i.i.d.* and $\mathrm{E}||y|| < \infty$, then as $n \longrightarrow \infty$,

$$\sqrt{n}(\bar{y} - \mu) \xrightarrow{\mathrm{d}} N(0, V),$$

where $\mu = \mathrm{E}(y)$ and $V = \mathrm{E}(y - \mu)(y - \mu)'$.

Note: These WLLN and CLT apply to i.i.d. samples only, hence we cannot use them for time series models!

Continuous Mapping Theorem (CMT)

If $z_n \xrightarrow{p} c$ and $g(\cdot)$ is continuous at c, then $g(z_n) \xrightarrow{p} g(c)$.

Delta Method

If $\sqrt{n}(\hat{\mu} - \mu) \xrightarrow{d} \eta$, where $g(u)$ is continuously differentiable in a neighbourhood of μ then as $n \longrightarrow \infty$

$$\sqrt{n}(g(\hat{\mu}) - g(\mu)) \xrightarrow{d} G'\eta,$$

where $G = \frac{\partial}{\partial \mu} g(\mu)'$. In particular, if $\eta \sim N(0, V)$, then as $n \longrightarrow \infty$,

$$\sqrt{n}(g(\hat{\mu}) - g(\mu)) \xrightarrow{d} N(0, G'VG).$$

Stationarity: The joint distribution of subsamples does not change over time

Ergodicity: x_t and x_{t-j} are "essentially uncorrelated" if j is large enough

i.i.d.: No serial correlation at all

Stationarity + ergodicity: Some serial correlation (but not "too much")

The ergodic theorem states that the LLN holds even when a stochastic process is serially correlated:

Theorem 4 (Ergodic theorem).

*If $\{z_t\}$ is **stationary** and **ergodic**, then*

$$\bar{z} = \frac{1}{T}\sum_{t=1}^{T} z_t \xrightarrow{p} E[z_t].$$

Therefore, if the stochastic process $(y_i, \mathbf{x}_i)$ is jointly **stationary** and **ergodic**, we can apply the ergodic theorem to guarantee that sample means converge to the population mean.

A more general CLT for time series analysis

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} = \left(\frac{1}{n}\sum \mathbf{x}_i\mathbf{x}_i'\right)^{-1}\left(\frac{1}{n}\sum \mathbf{x}_i e_i\right)$$

$$\Rightarrow \sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \left(\frac{1}{n}\sum \mathbf{x}_i\mathbf{x}_i'\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum \mathbf{x}_i e_i\right)$$

Result: If $\mathbf{x}_i$ is stationary and ergodic:

$$\frac{1}{n}\sum \mathbf{x}_i\mathbf{x}_i' \rightarrow \mathrm{E}(\mathbf{x}_i\mathbf{x}_i')$$

$$\left(\frac{1}{n}\sum \mathbf{x}_i\mathbf{x}_i'\right)^{-1} \rightarrow (\mathrm{E}(\mathbf{x}_i\mathbf{x}_i'))^{-1}$$

The second term $\frac{1}{\sqrt{n}}\sum \mathbf{x}_i e_i$ is more tricky:

Goal: Assumptions so that $\frac{1}{\sqrt{n}}\sum \mathbf{x}_i e_i$ is "well-behaved" as $n \rightarrow \infty$ and converges to a normal distribution with mean of 0.

Note that we cannot apply the Lindeberg-Levy CLT since $\mathbf{x}_i e_i$ are not iid!

Recall: Strict exogeneity $\mathrm{E}(e_i|\mathbf{x}_{jk}) = 0$

Instead, we assume that

$$\mathrm{E}(e_i|\mathbf{x}_i) = 0 \;\wedge\; \mathrm{E}(e_i|e_{i-j}, \mathbf{x}_{i-j}) = 0\,\forall j > 0$$

$$\Leftrightarrow \mathrm{E}(e_i|e_{i-1}, e_{i-2}, \ldots, \mathbf{x}_i, \mathbf{x}_{i-1}, \ldots) = 0$$

$$\Rightarrow \mathrm{E}(\mathbf{x}_i e_i) = 0; \forall i$$

We say that "**regressors are predetermined**" instead of exogenous.

$$\mathrm{E}(e_i|\ldots,\mathbf{x}_{i+2},\mathbf{x}_{i+1},\mathbf{x}_i,\mathbf{x}_{i-1},\mathbf{x}_{i-2},\ldots)=0$$
$$\text{vs.}\quad \mathrm{E}(e_i|e_{i-1},e_{i-2},\ldots,\mathbf{x}_i,\mathbf{x}_{i-1},\ldots)=0$$

and

$$\mathrm{E}(\mathbf{x}_j\, e_i)=0 \quad \text{vs.} \quad \mathrm{E}(\mathbf{x}_i\, e_i)=0$$

Key difference: Error e_i has expectation of zero conditional on

- all past errors $e_{i-1}, e_{i-2}, \ldots$ (i.e. errors are uncorrelated)
- past $\mathbf{x}_{i-1}, \mathbf{x}_{i-2}$
- current $\mathbf{x}_i$

This assumption allows e_t to be correlated with future $\mathbf{x}_{i+j}$ (which is the case in time series models!)

AR(p)

The condition

$$\mathrm{E}(e_i | e_{i-1}, e_{i-2}, \ldots, \mathbf{x}_i, \mathbf{x}_{i-1}) = 0.$$

is clearly satisfied in AR(p) models

$$z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + \ldots + \phi_p z_{t-p} + e_t.$$

Note: $e_i = e_t, \mathbf{x}_i = (z_{t-1}, z_{t-2}, \ldots, z_{t-p})$

The assumption $E(e_i|e_{i-1}, e_{i-2}, \ldots, \mathbf{x}_i, \mathbf{x}_{i-1}) = 0$ allows us to use a more general CLT.

Theorem 5 (Billingsley's CLT).

Assume that

1. $E(e_i|e_{i-1}, e_{i-2}, \ldots, \mathbf{x}_i, \mathbf{x}_{i-1}) = 0$
2. $\mathbf{g}_i \equiv \mathbf{x}_i e_i$ *is stationary and ergodic with* $E(\mathbf{g}_i \mathbf{g}_i') = \mathbf{\Omega}$

Then

$$\sqrt{n}\, \bar{\mathbf{g}}_i = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbf{g}_i \xrightarrow{d} N(0, \mathbf{\Omega})$$

For those with a technical background: Formally $\mathbf{g}_i \equiv \mathbf{x}_i e_i$ is required to be a **martingale difference series**.

Assumption 4 (Asymptotic OLS).

1. *Linearity:* $y_i = \mathbf{x}_i'\boldsymbol{\beta} + e_i, \quad i = 1, \ldots, n.$
2. *The stochastic process* $(y_i, \mathbf{x}_i)$ *is jointly* ***stationary*** *and* ***ergodic***.
3. $Q_{\mathbf{xx}} = E(\mathbf{xx}')$ *is positive definite.*
4. *The regressors* x_i *are* ***predetermined***:
$$E(e_i | e_{i-1}, e_{i-2}, \ldots, \mathbf{x}_i, \mathbf{x}_{i-1}) = 0.$$
   In particular, $E(\mathbf{x}_i e_i) = 0$ *for all* i.
5. $\boldsymbol{\Omega} = E(\mathbf{x}_i \mathbf{x}_i' e_i^2)$ *is nonsingular*

If $(y_i, \mathbf{x}_i)$ are jointly **stationary/ergodic**, then e_i is also stationary/ergodic.

To recap: The following assumptions of the classical regression model are violated in financial time series

1. The data are i.i.d.
2. $\mathrm{E}(\mathbf{e}|\mathbf{x}) = 0$
3. Errors are normal

Implication: OLS $\widehat{\boldsymbol{\beta}}$ is no longer unbiased and normal.

We replace these assumptions with

1. The data are jointly **stationary** and **ergodic**
2. The expectation of the errors is 0 conditional on past errors and past and current $\mathbf{x}$: $\mathrm{E}(e_i|e_{i-1}, e_{i-2}, \ldots, \mathbf{x}_i, \mathbf{x}_{i-1}) = 0$

Next, we will derive asymptotic properties of the OLS estimator.

Consistency of OLS $\widehat{\beta}$

Recall that that we can write,

$$\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} = \left(\frac{1}{n}\sum \mathbf{x}_i\mathbf{x}_i'\right)^{-1}\left(\frac{1}{n}\sum \mathbf{x}_i e_i\right) = \widehat{Q}_{\mathbf{xx}}^{-1}\widehat{Q}_{\mathbf{xe}},$$

where

$$\widehat{Q}_{\mathbf{xx}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i,\ \widehat{Q}_{\mathbf{xe}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i e_i$$

The **Ergodic Theorem** implies that

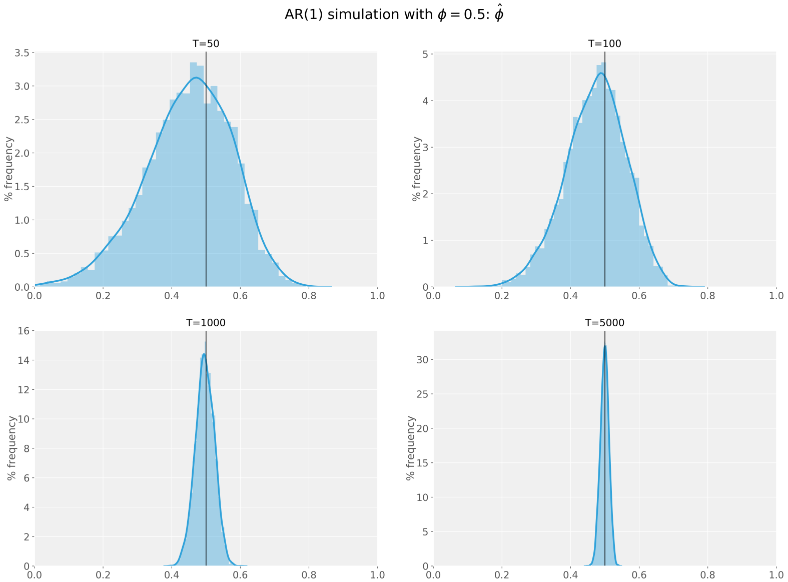
$$\widehat{Q}_{\mathbf{xx}} \xrightarrow{p} Q_{\mathbf{xx}} = \mathrm{E}(\mathbf{x}_i\mathbf{x}_i)$$

$$\widehat{Q}_{\mathbf{x}e} \xrightarrow{p} Q_{\mathbf{x}e} = \mathrm{E}(\mathbf{x}_i e_i) = 0$$

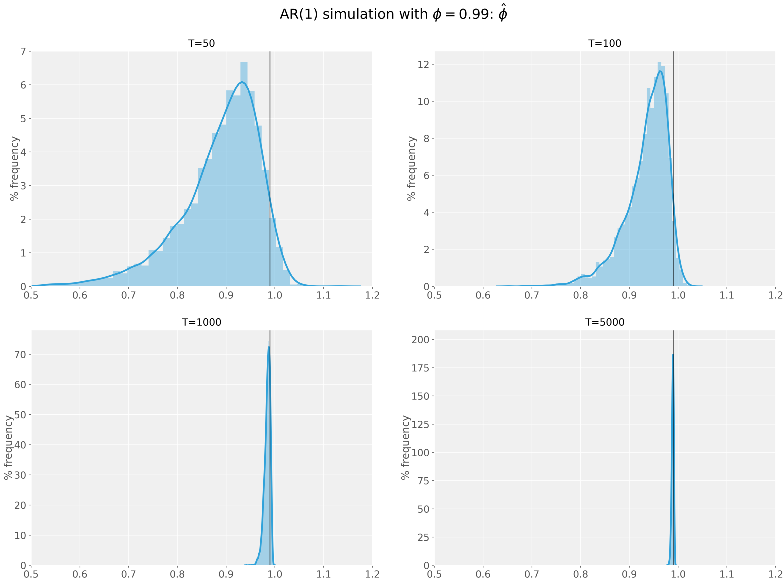
$$\Rightarrow \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} = \widehat{Q}_{\mathbf{xx}}^{-1}\widehat{Q}_e \xrightarrow{p} Q_{\mathbf{xx}}^{-1} \times 0 = 0$$

How about the distribution of $\widehat{\boldsymbol{\beta}}$?

OLS for AR(1) with $\phi = 0.5$



OLS FOR AR(1) WITH $\phi = 0.99$



Some of the distributions of $\widehat{\beta}$ are not normal even though the errors in the AR(1) process are normal

- Holding ϕ fixed, $\widehat{\beta}$ looks closer to normal for larger T
- Holding T fixed, $\widehat{\beta}$ looks closer to normal for smaller ϕ

Next, we derive the asymptotic distribution of the OLS estimator.

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})=\left(\frac{1}{n}\sum \mathbf{x}_i\mathbf{x}_i'\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum \mathbf{x}_i e_i\right)$$

Why do we consider $\frac{1}{\sqrt{n}}\sum \mathbf{x}_i e_i$?

Because the **Billingsley CLT** implies

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{x}_i e_i \xrightarrow{p} N(0,\boldsymbol{\Omega})$$

$$\text{where } \boldsymbol{\Omega}=\mathrm{E}(\mathbf{x}_i\mathbf{x}_i' e_i^2)$$

Therefore, we have the following important result

Theorem 6 (Asymptotic Normality of Least-Squares Estimator).

Under Assumptions 4, as $n \longrightarrow \infty$,

$$\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} N(0, \mathbf{Q}_{xx}^{-1}\, \boldsymbol{\Omega}\, \mathbf{Q}_{xx}^{-1})$$

where $\mathbf{Q}_{xx} = E(\mathbf{x}_i \mathbf{x}_i')$ *and* $\boldsymbol{\Omega} = E(\mathbf{x}_i \mathbf{x}_i' e_i^2)$.

ASYMPTOTIC COVARIANCE MATRIX OF $\widehat{\boldsymbol{\beta}}$

The asymptotic population covariance matrix of $\widehat{\boldsymbol{\beta}}$ is

$$\begin{aligned}
V^{\infty}_{\widehat{\boldsymbol{\beta}}} &= \frac{1}{n}\mathbf{Q}_{xx}^{-1}\boldsymbol{\Omega}\mathbf{Q}_{xx}^{-1} \\
&= \frac{1}{n}\left(\mathrm{E}(\mathbf{x}_i\mathbf{x}_i')\right)^{-1}\mathrm{E}(\mathbf{x}_i\mathbf{x}_i'e_i^2)\left(\mathrm{E}(\mathbf{x}_i\mathbf{x}_i')\right)^{-1}.
\end{aligned}$$

Finite sample: $V^{\infty}_{\widehat{\boldsymbol{\beta}}}$ is estimated using sample moments ($\mathbf{D} = \mathrm{E}(\mathbf{ee}' \mid \mathbf{X})$):

$$\begin{aligned}
\widehat{V}^{\infty}_{\widehat{\boldsymbol{\beta}}} &= \frac{1}{n}\left(\frac{1}{n}\mathbf{X}'\mathbf{X}\right)^{-1}\left(\frac{1}{n}\mathbf{X}'\mathbf{D}\mathbf{X}\right)\left(\frac{1}{n}\mathbf{X}'\mathbf{X}\right)^{-1} \\
&= (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{D}\mathbf{X})(\mathbf{X}'\mathbf{X})^{-1} \\
&= \widehat{V}_{\widehat{\boldsymbol{\beta}}}
\end{aligned}$$

The sample estimate of $\widehat{V}^{\infty}_{\widehat{\boldsymbol{\beta}}}$ is identical to the classical OLS covariance matrix $\widehat{V}_{\widehat{\boldsymbol{\beta}}}$!

Special case: Homoskedasticity

Assume that $\text{Cov}(\mathbf{x}_i\mathbf{x}_i', e_i^2) = 0$ (slightly more general than strict homoskedasticity):

$$\boldsymbol{\Omega} = \text{E}(\mathbf{x}_i\mathbf{x}_i' e_i^2) = \mathbf{Q}_{xx}\sigma^2$$

and therefore

$$\boldsymbol{V}_{\widehat{\boldsymbol{\beta}}}^{\infty} = (\text{E}(\mathbf{x}_i\mathbf{x}_i')^{-1}\sigma^2,$$

which is again identical to the classical case.

Moreover, moment estimators of the residual variance are also consistent:

$$\hat{\sigma}^2 = \frac{1}{n}\sum \hat{e}_i^2 \xrightarrow{p} \sigma^2$$

$$s^2 = \hat{\sigma}^2 = \frac{1}{n}\sum \hat{e}_i^2 \xrightarrow{p} \sigma^2$$

- Theorem 6 states that under quite general assumptions about the joint distribution of $(y_i, \mathbf{x}_i)$, the sampling distribution of the rescaled OLS estimator is approx. normal when the sample size is sufficiently large.
- Consequently, asymptotic normality is routinely used to approximate the finite sample distribution of $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})$.

Disucssion II

- Open question: What does "sufficiently large" sampling size mean?
- Unfortunately, there is no general answer to this question.
- "Sufficiently large" sampling size depends on the particular application.
- Note that for fixed n, the sampling distribution of $\widehat{\boldsymbol{\beta}}$ can be very different from the normal distribution.
- In some cases the normal approximation can be poor even for large n!

Classical vs asymptotic OLS

1. **Classical OLS** with normality and strict exogeneity:
   Stronger assumptions but exact small sample distribution of OLS estimator
2. **Asymptotic OLS** with predetermined regressors, no normality:
   Weaker assumptions but only asymptotic distributions of OLS estimator

Note: The equations for classical and asymptotic OLS are **identical**!!

Difference: The **properties** of the OLS estimator

1. **Classical OLS**: Unbiased and normal if errors are normal
2. **Asymptotic OLS** Consistent and asymptotically normal

Common assumption: **Stationarity and ergodicity!**

- Simulate 10,000 samples for $T = 50, 100, 1000, 5000$ of an AR(1)

$$z_t = \phi z_{t-1} + \epsilon_t, \text{ where } \epsilon_t \sim \text{NWN}(0, \sigma_\epsilon)$$

  for $\phi = 0, 0.5, 0.95, 0.99$, $\sigma_\epsilon = 1$.

- The asymptotic distribution of the OLS estimator is

$$\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} N(0, \sigma_\epsilon^2 \mathbf{Q}_{xx}^{-1}),$$

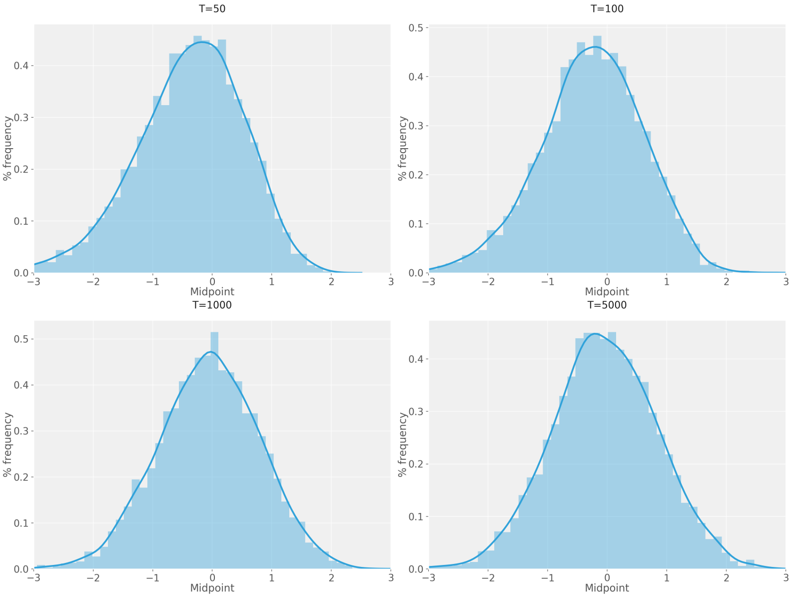
  where $\mathbf{Q}_{xx} = \mathrm{E}(\mathbf{x}_i \mathbf{x}_i')$.

- For the AR(1): $\mathbf{Q}_{xx} = \sigma_\epsilon^2/(1 - \phi^2)$. Hence

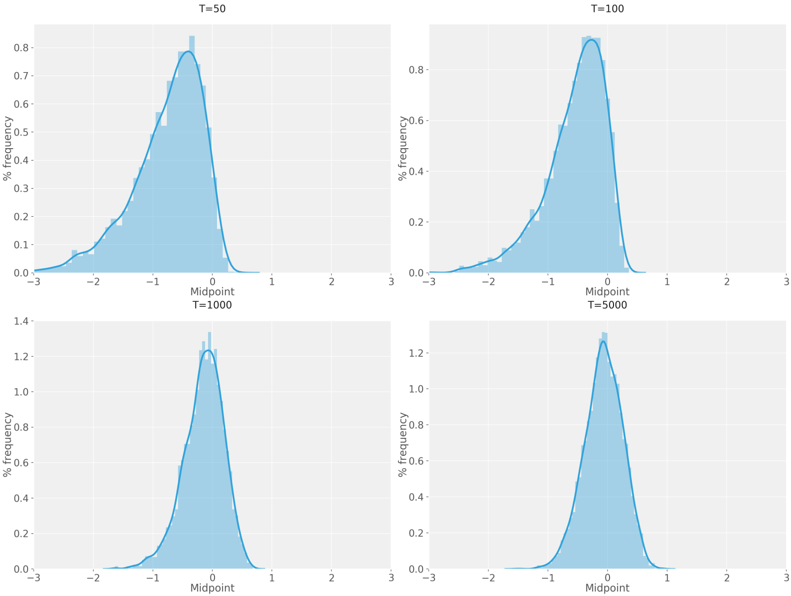
$$\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} N(0, 1 - \phi^2)$$

- Note that the asymptotic variance is decreasing in ϕ!

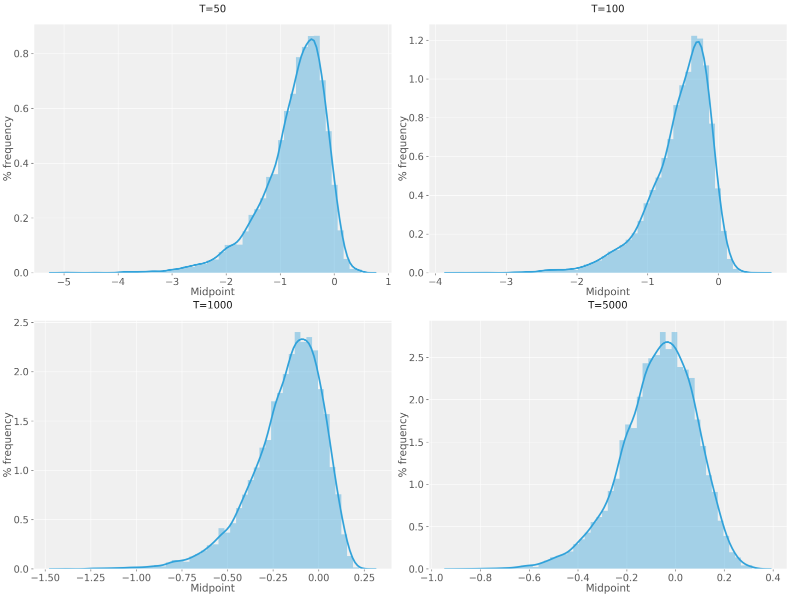
Distribution of $\sqrt{T}(\widehat{\phi} - \phi)$ for $\phi = 0.5$



Distribution of $\sqrt{T}(\widehat{\phi} - \phi)$ for $\phi = 0.95$



Distribution of $\sqrt{T}(\widehat{\phi}-\phi)$ for $\phi = 0.99$



As long as we have "sufficient" data, we can approximate the distribution of $\widehat{\beta}$ by a normal distribution.

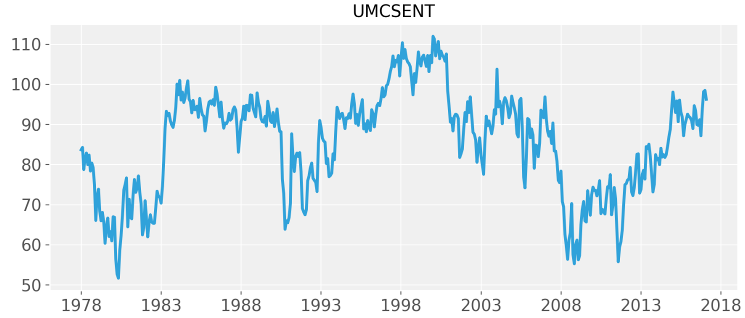
Note that we have made no assumption about the distribution of the errors!

Key assumption: Stationarity and ergodicity

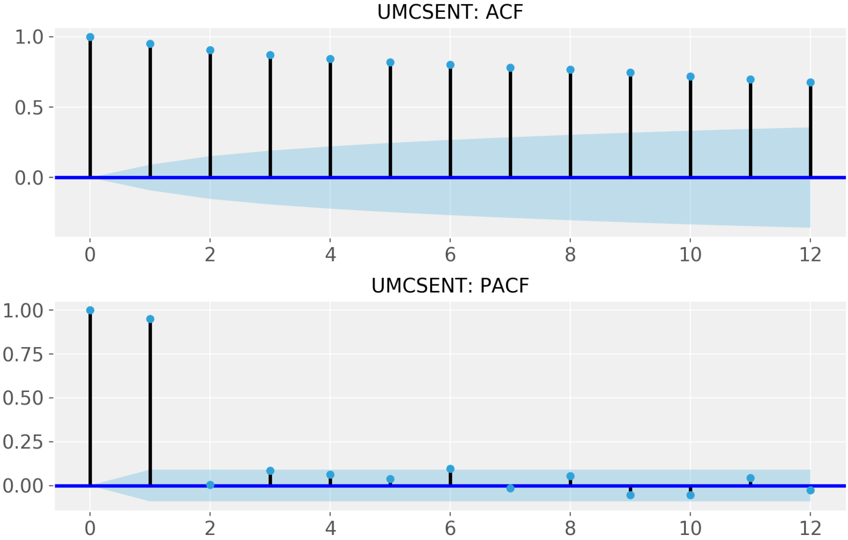
Example: Monthly Consumer Sentiment Index

Data: University of Michigan Consumer Sentiment Index

Source: St. Louis Fred, series: UMCSENT



Example: Monthly Consumer Sentiment Index



	AC	PAC	Q-stat	Prob
lag				
1	0.95	0.95	428.15	0.00
2	0.91	0.01	816.99	0.00
3	0.87	0.09	1176.68	0.00
4	0.84	0.07	1514.01	0.00
5	0.82	0.04	1832.42	0.00
6	0.80	0.10	2138.92	0.00

Estimating an AR(1) process in Python/statsmodels

```
Y = df.UMCSENT
X = df.UMCSENT.shift(periods=1)
X = sm.add_constant(X)
ar1_model = sm.OLS(Y, X, missing="drop").fit()
sum = ar1_model.summary()
```

AR(1) MODEL FOR UMCSENT

Dep. Variable:	CS	**R-squared:**	0.907
Model:	OLS	**Adj. R-squared:**	0.906
Method:	Least Squares	**F-statistic:**	4532.
Date:	Fri, 31 Mar 2017	**Prob (F-statistic):**	1.50e-242
Time:	12:16:37	**Log-Likelihood:**	-1302.4
No. Observations:	469	**AIC:**	2609.
Df Residuals:	467	**BIC:**	2617.
Df Model:	1		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[95.0% Conf. Int.]**
const	4.0583	1.224	3.317	0.001	1.654 6.463
CS	0.9529	0.014	67.323	0.000	0.925 0.981

Omnibus:	16.646	**Durbin-Watson:**	2.018
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	31.533
Skew:	-0.187	**Prob(JB):**	1.42e-07
Kurtosis:	4.214	**Cond. No.**	588.

AR(1) MODEL FOR UMCSENT: ROBUST WHITE COVARIANCE MATRIX

Dep. Variable:	CS	**R-squared:**	0.907
Model:	OLS	**Adj. R-squared:**	0.906
Method:	Least Squares	**F-statistic:**	4583.
Date:	Fri, 31 Mar 2017	**Prob (F-statistic):**	1.45e-243
Time:	12:16:38	**Log-Likelihood:**	-1302.4
No. Observations:	469	**AIC:**	2609.
Df Residuals:	467	**BIC:**	2617.
Df Model:	1		
Covariance Type:	HC0		

	coef	**std err**	**z**	**P>|z|**	**[95.0% Conf. Int.]**
const	4.0583	1.267	3.203	0.001	1.575 6.541
CS	0.9529	0.014	67.696	0.000	0.925 0.980

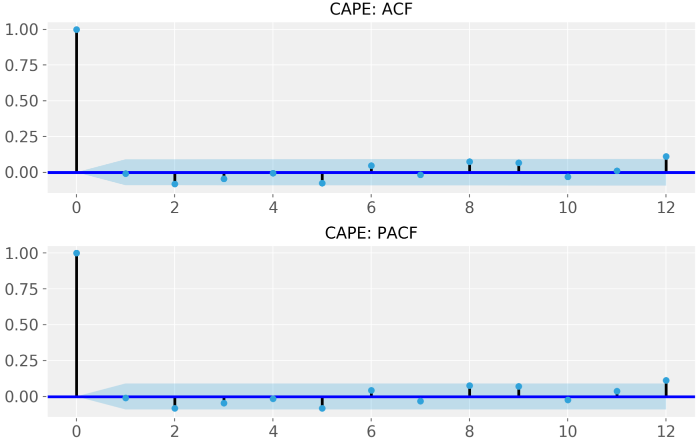
Omnibus:	16.646	**Durbin-Watson:**	2.018
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	31.533
Skew:	-0.187	**Prob(JB):**	1.42e-07
Kurtosis:	4.214	**Cond. No.**	588.

HOW DO WE DETERMINE LAG LENGTH p?

$$z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + \ldots + \phi_p z_{t-p} + \epsilon_t$$

- For fixed p: check for serial correlation in error terms, if there is significant serial correlation, more lags might be needed
- Recursive: start with large p and check for significance of pth lag
- Use information criteria: AIC, BIC
  Intuition: IC “penalize” estimation for additional regressors
- Compute log-likelihood for different p
- Example: Inflation (Ruppert pp. 220-222)
- Here: UMCSENT

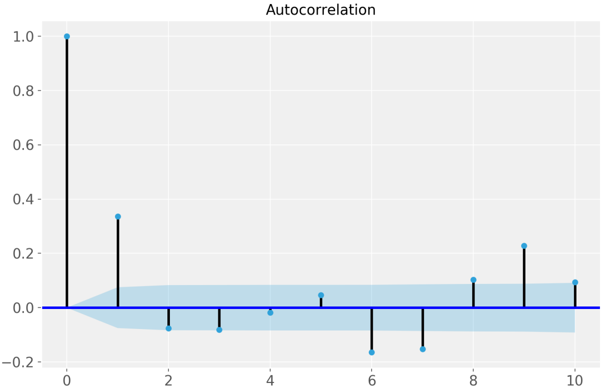
UMCSENT: Autocorrelations of AR(1) errors



If errors are serially correlated: More lags?

UMCSENT: Autocorrelations of AR(1) errors

If the ACF of the regression errors looked like this:



Errors are serially correlated: Probably need more lags!

Testing for serial correlation in errors

Best-known test: Durbin-Watson (1950) test

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

- The DW statistic has a non-standard distribution.
- If DW≈ 2, no evidence for serial correlation in error term.
- Note: The DW statistic is only valid under strict exogeneity!
- “Better” test: Breusch-Pagan-Godfrey (see Stats notes)

Breusch-Pagan LM test

$$H_0 : \gamma_j = 0 \;\; \forall j = 1, \ldots, L$$

lag	BG statistic	p-value
1.00	0.04	0.83
2.00	3.63	0.16
3.00	5.23	0.16
4.00	5.66	0.23
5.00	10.73	0.06
6.00	10.88	0.09

- Likelihood-based criteria:

$$
\begin{aligned}
\text{Log-likelihood} &= \log(L) \\
\text{Akaike AIC} &= -2\log(L)/n + 2k/n \\
\text{Schwarz BIC} &= -2\log(L)/n + 2k\log(n)/n \\
\text{Hannah-Quinn IC} &= -2\log(L)/n + 2k\log(\log(n))/n
\end{aligned}
$$

AR(p) FOR UMCSENT

	AR0	AR1	AR2	AR3	AR4
R2	0.00	0.91	0.91	0.91	0.91
AIC	3697.60	2592.76	2594.73	2593.13	2593.53
BIC	3701.75	2601.05	2607.16	2609.71	2614.25

$\widehat{\phi}_i$	ϕ_0	ϕ_1	ϕ_2	ϕ_3	ϕ_4
AR0	85.57				
AR1	4.05	0.95			
AR2	4.02	0.95	0.01		
AR3	3.66	0.95	-0.08	0.09	
AR4	3.44	0.94	-0.07	0.03	0.06

t-stats	ϕ_0	ϕ_1	ϕ_2	ϕ_3	ϕ_4
AR0	144.76				
AR1	3.19	67.64			
AR2	3.09	18.92	0.15		
AR3	2.78	18.98	-1.19	1.94	
AR4	2.55	18.99	-1.14	0.53	1.23

Outline

1. Review of classical OLS
2. Classical OLS model for AR estimation
3. **Non-stationarity**
4. Spurious regressions

What if the data are not stationary?

Answer: You are in econometric hell!

WHAT IF THE DATA ARE NOT STATIONARY?

- Recall: We assumed that the data is **stationary**
- Let's consider an AR(1)

$$z_t = \phi z_{t-1} + \epsilon_t$$

- If $|\phi| < 1$ standard regression results hold, in particular

$$\sqrt{T}(\hat{\phi} - \phi) \rightarrow N(0, \sigma_\epsilon^2 (\mathbf{X}'\mathbf{X})^{-1})$$

$$\sigma_\epsilon^2 (\mathbf{X}'\mathbf{X})^{-1} = 1 - \phi^2$$

- Note: If $\phi = 1$, then the asymptotic variance converges to 0!?

Random walk: $\phi = 1$

Behavior of an AR(1) is drastically different for $\phi = 1$ than for $|\phi| < 1$:

$$x_t = \phi x_{t-1} + \epsilon_t = \epsilon_t + \phi \epsilon_{t-1} + \phi^2 \epsilon_{t-2} + \dots$$

$$\mathrm{Var}(x_t) = \frac{\sigma_\epsilon^2}{1 - \phi^2} \quad \forall t$$

$$\rho_j = \phi^j$$

Random walk $\phi = 1$:

$$x_t = x_{t-1} + \epsilon_t = \epsilon_t + \epsilon_{t-1} + \epsilon_{t-2} + \dots$$

$$\mathrm{Var}(x_t) = t\,\sigma_\epsilon^2 \to \infty$$

$$\rho_j = 1$$

IRF of a random walk does not converge to 0, effect of shocks never die out!

Implication: A random walk is non-ergodic/non-stationary!

Properties of OLS for non-stationary processes are fundamentally different!

Simulate 10,000 samples for $T = 50, 100, 1000, 5000$ of an AR(1)

$$z_t = \phi z_{t-1} + \epsilon_t, \text{ where } \epsilon_t \sim \text{NWN}(0, \sigma_\epsilon)$$

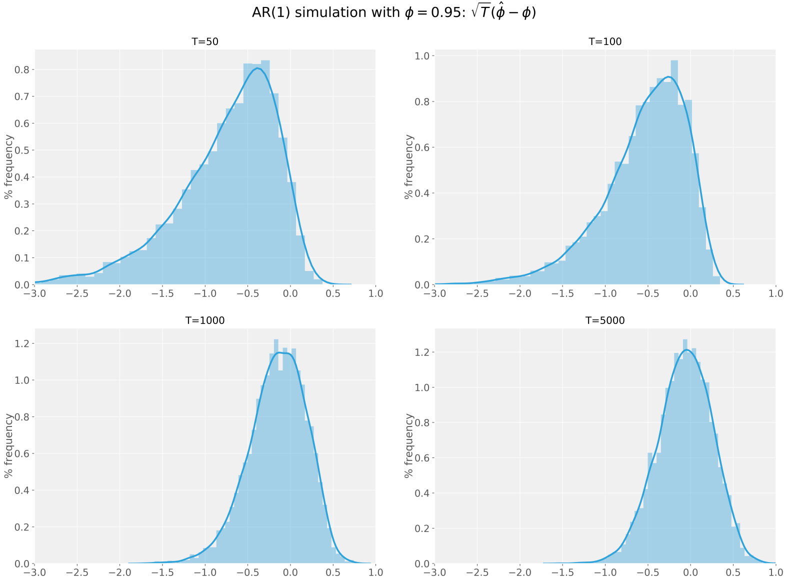
with $\sigma_\epsilon = 1$.

Estimate OLS in 3 cases: $\phi = 0.95, 0.99, 1$

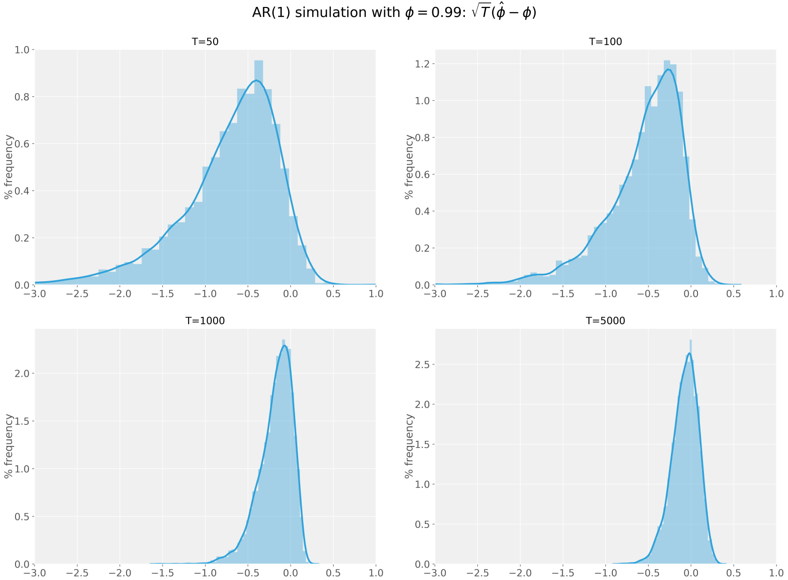
$$z_t = \mu + \phi z_{t-1} + \epsilon_t$$

and compute the distribution of $\hat{\phi}$.

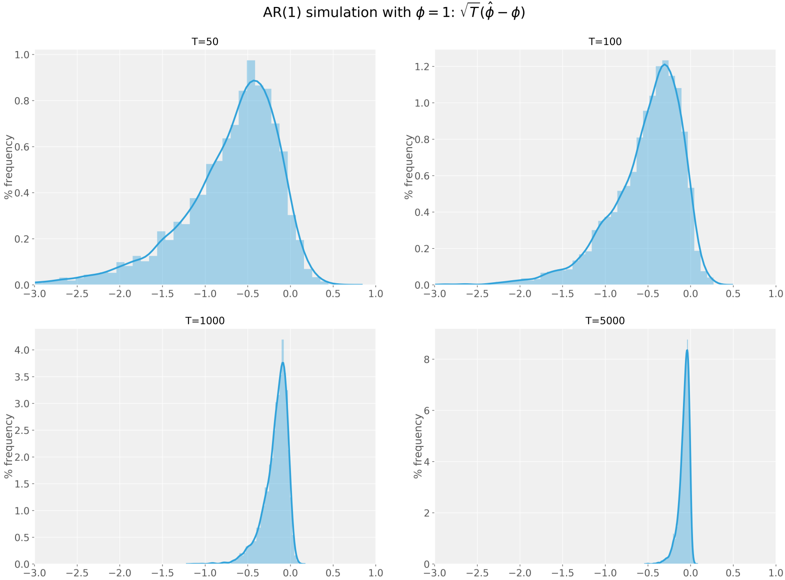
Distribution of $\sqrt{T}(\widehat{\phi} - \phi)$ for $\phi = 0.95$



Distribution of $\sqrt{T}(\widehat{\phi} - \phi)$ for $\phi = 0.99$



Distribution of $\sqrt{T}(\widehat{\phi} - \phi)$ for $\phi = 1$

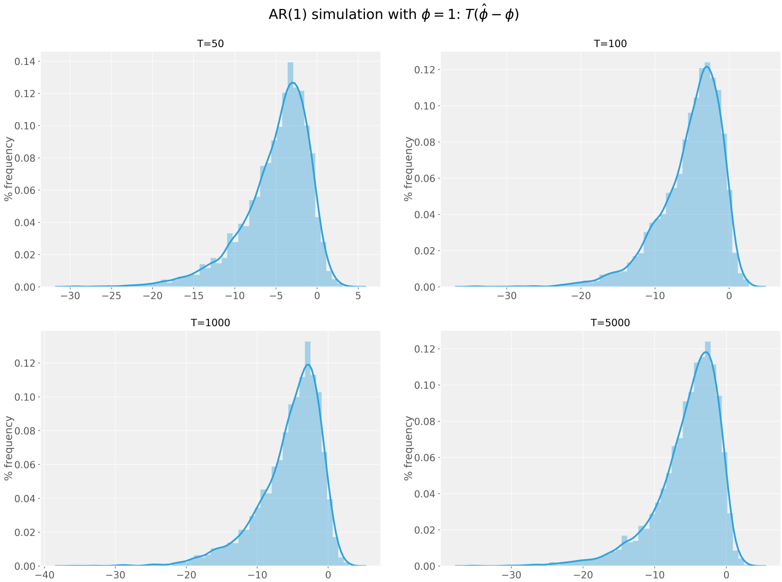


Theorem 7 (Properties of OLS for random walk).

If z_t follows a random walk, $z_t = z_{t-1} + \epsilon_t$, the OLS estimator $\widehat{\phi}$ in the AR(1) model $z_t = z_{t-1} + \epsilon_t$ has the following properties:

- $\sqrt{T}\,(\widehat{\phi} - 1)$ ***does not converge*** *to a normal distribution, nor any other well-defined distribution .*
- $T\,(\widehat{\phi} - 1) \rightarrow DF_{\phi}(\boldsymbol{\theta})$*, where* $DF(\boldsymbol{\theta})$ *is the* ***Dickey-Fuller distribution*** *with nuisance parameters* $\boldsymbol{\theta}$ *that is skewed to the left.*
- $\widehat{\phi}$ *is* ***superconsistent****:* $\widehat{\phi} \rightarrow 1$ *at rate* T*, rather than* $\sqrt{T}$*.*
- $\widehat{\phi}$ *is* ***severely downwards biased****: In two-thirds of the samples generated by a random walk,* $\widehat{\phi}$ *will be less than unity.*
- *The standard* t*-tests of* $H_0 : \phi = 1$ *will* ***reject too often****.*

Distribution of $T(\widehat{\phi} - \phi)$ for $\phi = 1$



- Asymptotic distribution of OLS:

$$|\phi| < 1: \quad \sqrt{T}(\widehat{\phi} - \phi) \xrightarrow{d} N(0, 1 - \phi^2)$$

$$\phi = 1: \quad T(\widehat{\phi} - \phi) \xrightarrow{d} \mathrm{DF}_{\phi}(\boldsymbol{\theta})$$

- Point estimate $\hat{\phi}$ is downward-biased in finite sample
- The bias is more severe the closer the true ϕ is to 1.
- If your estimated $\hat{\phi}$ is “close” to 1, you need to be very careful since the true ϕ is likely to be larger than $\hat{\phi}$!

Every time you run a time series regression, you have to make sure that your data is stationary!!!

$\longrightarrow$ Formal test?

- Ruppert pp. 338-341
- Again, let's start with an AR(1)

$$z_t = \phi z_{t-1} + \epsilon_t$$

- The test $H_o : \phi = 1$ in

$$z_t = \phi z_{t-1} + \epsilon_t$$

  works poorly in practice

- Dickey-Fuller **DF test**: $H_0 : \rho = 0$ in

$$\Delta z_t = (\phi - 1) z_{t-1} + \epsilon_t$$

$$\Delta z_t = \rho z_{t-1} + \epsilon_t$$

$$t_{\mathrm{DF}} = \frac{\hat{\rho}}{s.e.(\hat{\rho})} \longrightarrow \mathrm{DF}_{\rho}(\boldsymbol{\theta})$$

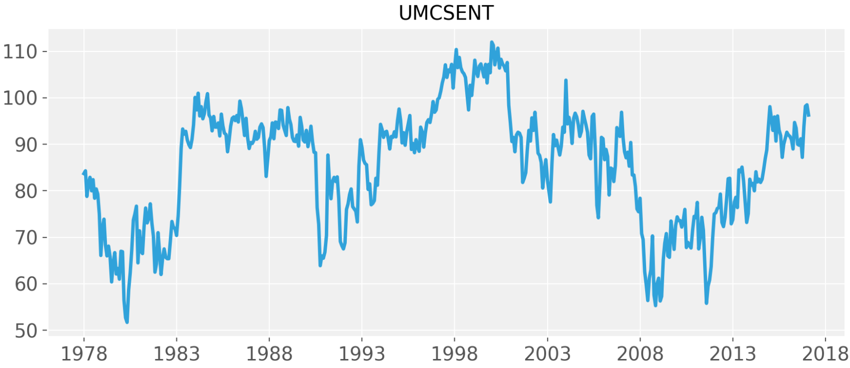
- **Augmented DF test** for $H_0 : \rho = 0$:

$$\Delta z_t = \rho z_{t-1} + \delta_1 \Delta z_{t-1} + \cdots + \delta_p \Delta z_{t-p} + \epsilon_t$$

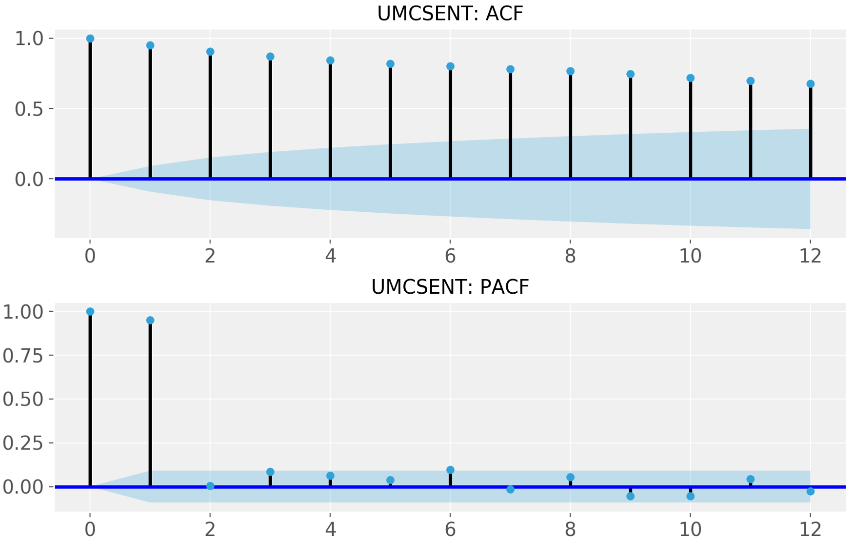
$$t_{\text{ADF}} = \frac{\hat{\rho}}{s.e.(\hat{\rho})} \longrightarrow \text{ADF}_\rho(\boldsymbol{\theta})$$

- ADF test has better small sample properties than DF test
- May include constant and/or linear trend but limiting distribution will change
- Many financial time series are very persistent $\Rightarrow$ always test whether they are non-stationary

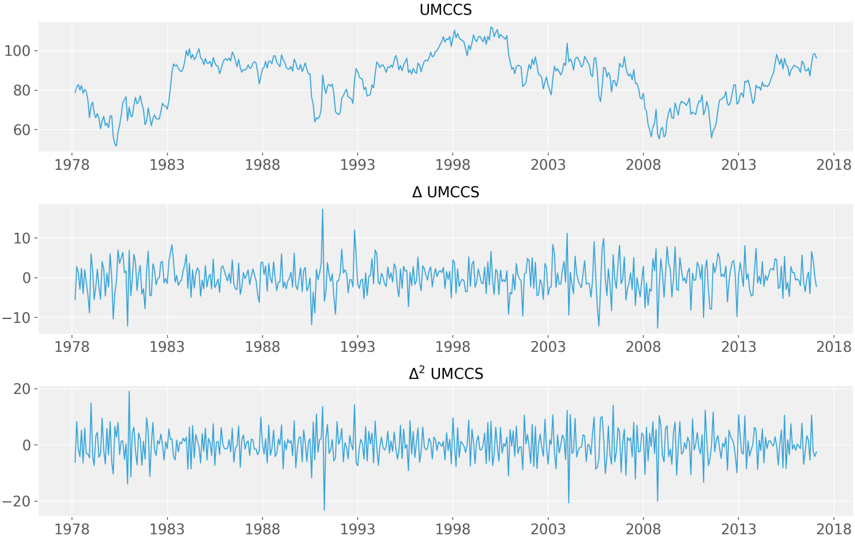
Example: UMCSENT



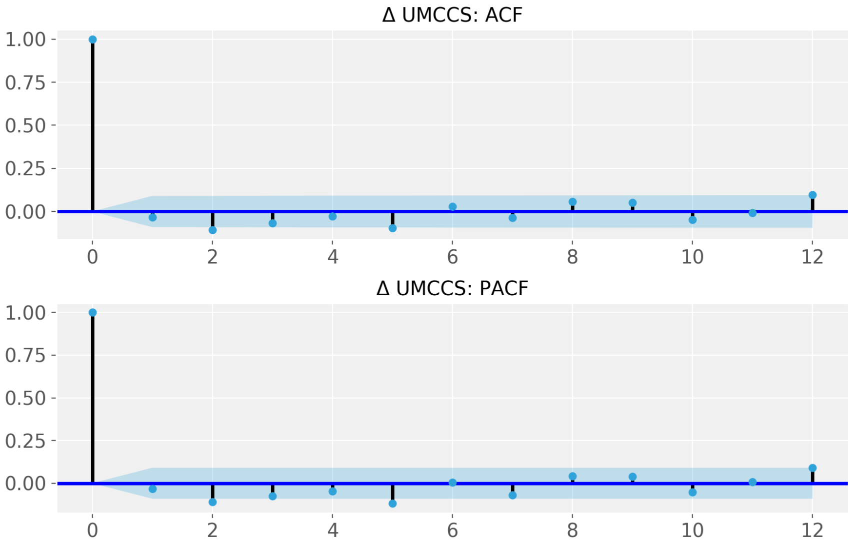
Example: UMCSENT



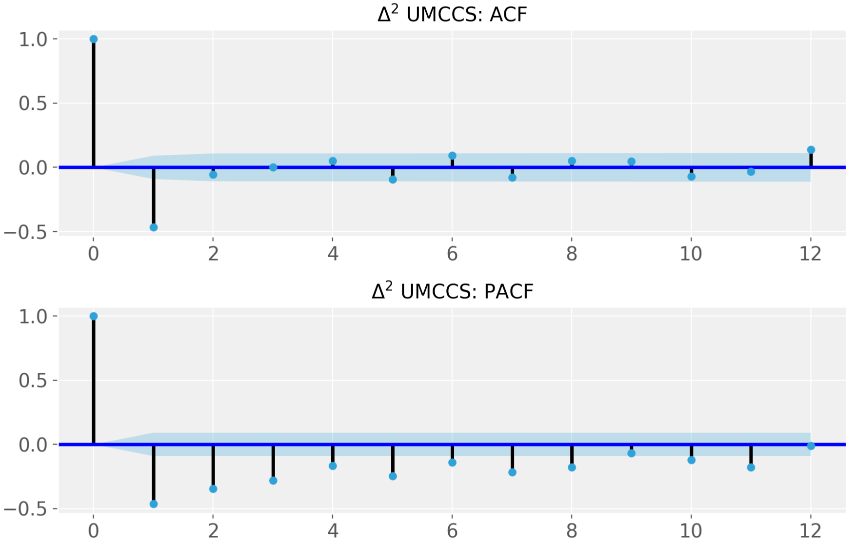
Differencing



Example: Δ UMCSENT



EXAMPLE: Δ^2 UMCENT



Example: UMCSENT - ADF test

lags	ADF	p-value	1%	5%	10%
0.00	-3.34	0.01	-3.44	-2.87	-2.57
1.00	-3.25	0.02	-3.44	-2.87	-2.57
2.00	-2.92	0.04	-3.44	-2.87	-2.57
5.00	-2.33	0.16	-3.44	-2.87	-2.57

lags	ADF	p-value	1%	5%	10%
0.00	-22.33	0.00	-3.44	-2.87	-2.57
1.00	-17.24	0.00	-3.44	-2.87	-2.57
2.00	-14.49	0.00	-3.44	-2.87	-2.57
4.00	-12.15	0.00	-3.44	-2.87	-2.57

△ UMCENT: AR(1) ESTIMATION

Dep. Variable:	CSd1	**R-squared:**	0.001
Model:	OLS	**Adj. R-squared:**	-0.001
Method:	Least Squares	**F-statistic:**	0.4358
Date:	Fri, 31 Mar 2017	**Prob (F-statistic):**	0.509
Time:	13:55:20	**Log-Likelihood:**	-1302.1
No. Observations:	467	**AIC:**	2608.
Df Residuals:	465	**BIC:**	2617.
Df Model:	1		
Covariance Type:	HC0		

	coef	std err	z	P>|z|	[95.0% Conf. Int.]
const	0.0385	0.182	0.212	0.832	-0.318 0.395
CSd1L1	-0.0332	0.050	-0.660	0.509	-0.132 0.065

Omnibus:	12.953	**Durbin-Watson:**	2.002
Prob(Omnibus):	0.002	**Jarque-Bera (JB):**	25.565
Skew:	-0.044	**Prob(JB):**	2.81e-06
Kurtosis:	4.143	**Cond. No.**	3.94

Δ^2 UMCENT: AR(2) ESTIMATION

Dep. Variable:	CSd2	**R-squared:**	0.308
Model:	OLS	**Adj. R-squared:**	0.305
Method:	Least Squares	**F-statistic:**	78.16
Date:	Fri, 31 Mar 2017	**Prob (F-statistic):**	5.67e-30
Time:	13:49:06	**Log-Likelihood:**	-1382.9
No. Observations:	466	**AIC:**	2772.
Df Residuals:	463	**BIC:**	2784.
Df Model:	2		
Covariance Type:	HC0		

	coef	**std err**	**z**	**P>|z|**	**[95.0% Conf. Int.]**
const	-0.0002	0.218	-0.001	0.999	-0.427 0.427
CSd2L1	-0.6223	0.050	-12.461	0.000	-0.720 -0.524
CSd2L2	-0.3442	0.045	-7.704	0.000	-0.432 -0.257

Omnibus:	8.878	**Durbin-Watson:**	2.192
Prob(Omnibus):	0.012	**Jarque-Bera (JB):**	14.308
Skew:	-0.062	**Prob(JB):**	0.000782
Kurtosis:	3.849	**Cond. No.**	6.86

△ UMCSENT: Lag order?

	AR0	AR1	AR2	AR3	AR4
R2	-0.00	0.00	0.01	0.02	0.02
AIC	2591.87	2593.42	2590.06	2589.20	2590.24
BIC	2596.01	2601.70	2602.48	2605.76	2610.94

$\widehat{\phi}$	ϕ_0	ϕ_1	ϕ_2	ϕ_3	ϕ_4
AR0	0.04				
AR1	0.04	-0.03			
AR2	0.04	-0.03	-0.11		
AR3	0.04	-0.04	-0.11	-0.08	
AR4	0.04	-0.05	-0.11	-0.08	-0.05

$t_{\widehat{\phi}}$	ϕ_0	ϕ_1	ϕ_2	ϕ_3	ϕ_4
AR0	0.19				
AR1	0.20	-0.62			
AR2	0.22	-0.69	-2.40		
AR3	0.24	-0.86	-2.46	-1.67	
AR4	0.24	-0.92	-2.53	-1.72	-1.01

Δ^2 UMCSENT: Lag order?

	AR0	AR1	AR2	AR3	AR4
R2	0.00	0.21	0.31	0.36	0.38
AIC	2927.88	2817.44	2761.61	2724.88	2713.75
BIC	2932.02	2825.72	2774.03	2741.44	2734.45

$\widehat{\phi}$	ϕ_0	ϕ_1	ϕ_2	ϕ_3	ϕ_4
AR0	0.00				
AR1	0.00	-0.46			
AR2	0.00	-0.62	-0.34		
AR3	0.01	-0.72	-0.52	-0.28	
AR4	0.01	-0.77	-0.61	-0.40	-0.17

$t_{\widehat{\phi}}$	ϕ_0	ϕ_1	ϕ_2	ϕ_3	ϕ_4
AR0	0.01				
AR1	0.00	-10.26			
AR2	0.01	-12.43	-7.63		
AR3	0.04	-14.27	-9.96	-6.21	
AR4	0.05	-15.19	-10.25	-6.53	-3.14

1. Review of classical OLS
2. Classical OLS model for AR estimation
3. Non-stationarity
4. **Spurious regressions**

- **Non-stationarity can cause major problems**
- Example: **Spurious regressions**
- Suppose x_t, y_t are independent but both are $I(1)$:

$$x_t = x_{t-1} + e_t$$

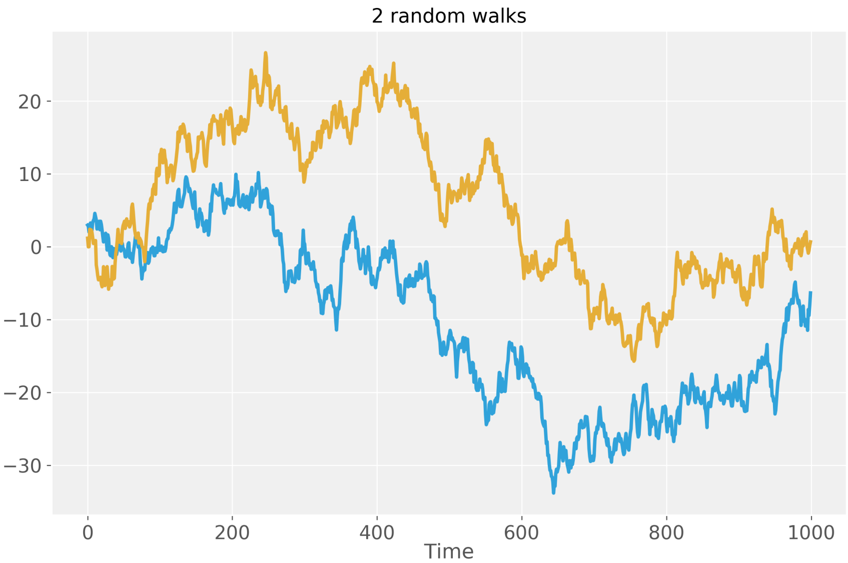
$$y_t = y_{t-1} + w_t$$

- Regression:

$$y_t = \mu + \beta x_t + u_t$$

- Since x_t, y_t are independent: $\mu = \beta = R^2 = 0$
- Hence $\widehat{\mu} \approx \widehat{\beta} \approx R^2 \approx 0$?

SIMULATIONS OF TWO INDEPENDENT RANDOM WALKS



Simulations of Two Independent Random Walks

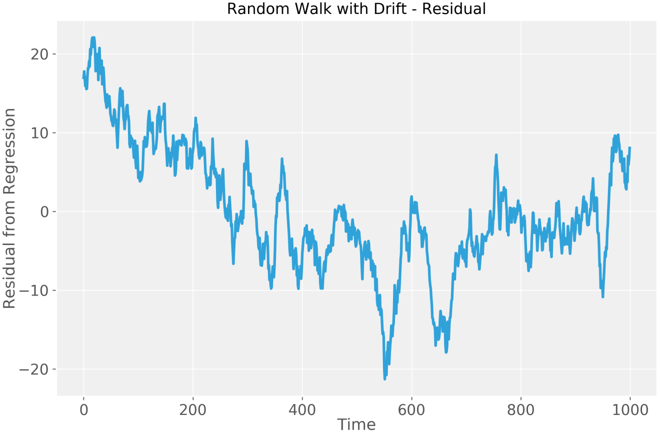
Regression $y_t = \mu + \beta x_t + u_t$

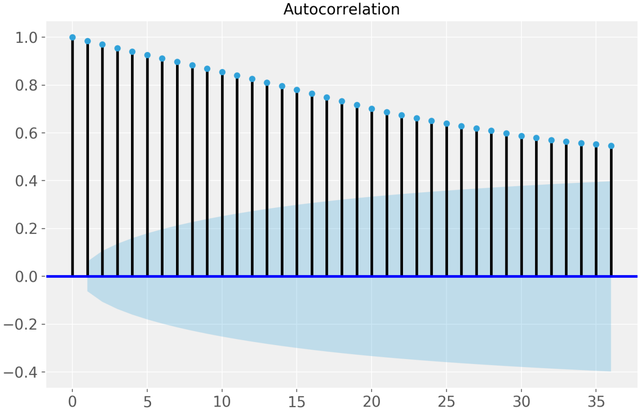
Dep. Variable:	Y	**R-squared:**	0.531
Model:	OLS	**Adj. R-squared:**	0.530
Method:	Least Squares	**F-statistic:**	1129.
Date:	Fri, 15 Mar 2019	**Prob (F-statistic):**	3.90e-166
Time:	11:48:22	**Log-Likelihood:**	-3490.9
No. Observations:	1000	**AIC:**	6986.
Df Residuals:	998	**BIC:**	6996.
Df Model:	1		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
const	-14.9284	0.286	-52.236	0.000	-15.489	-14.368
X	0.8026	0.024	33.595	0.000	0.756	0.849

Omnibus:	12.997	**Durbin-Watson:**	0.024
Prob(Omnibus):	0.002	**Jarque-Bera (JB):**	13.160
Skew:	0.278	**Prob(JB):**	0.00139
Kurtosis:	3.080	**Cond. No.**	13.6

Simulations of two independent random walks: Residuals





Autocorrelation
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
0
5
10
15
20
25
30
35

Simulate two AR(1) processes:

$$x_t = \phi x_{t-1} + e_t$$

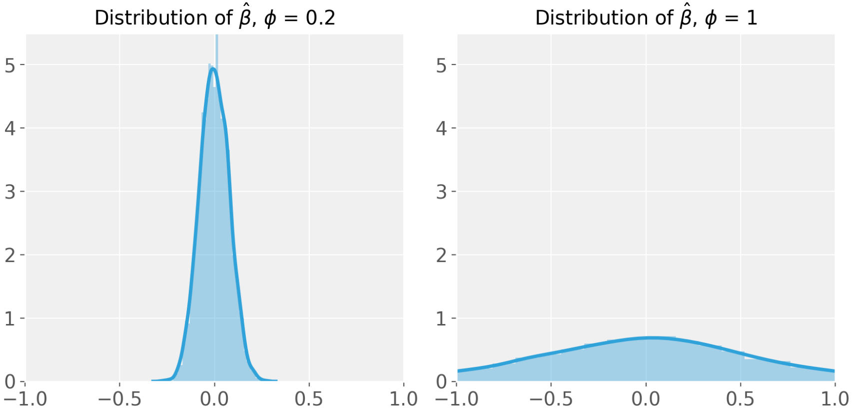
$$y_t = \phi y_{t-1} + w_t$$

$$\text{Regression} \quad y_t = \mu + \beta x_t + u_t$$

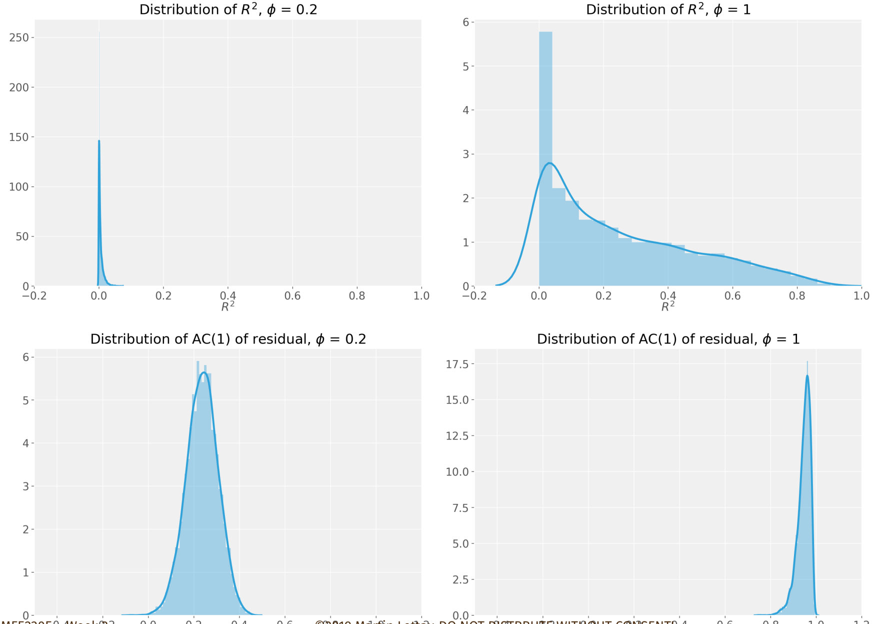
Since x_t and y_t are independent, β and the R^2 should be close to 0.

We set $\phi = 0.2, 1$ and simulate 10,000 runs of length $T = 200$.

Two AR(1) processes



Two AR(1) processes



Everything you have learned about regressions is no longer true!!

- $u_t \sim I(1)$: OLS assumes $u_t \sim I(0)$, usually i.i.d.
- $\widehat{\beta}$ is **not consistent**
- $t_{\hat{\beta}} \to \infty$ as $T \to \infty$
- Thus $H_0 : \beta = 0$ usually rejected
- Instead $t_{\hat{\beta}}/T^{1/2}$ converges to a well-defined distribution
- $R^2 \not\to 0$

One of the most important practical lessons in this class:

Regressing a persistent variable on another persistent variable is likely to lead to detection of a causal relationship where there is none!!!

Spurious regressions: A regression that finds a relationship between two (or more) variables when there is none.

Example: Regression of two independent non-stationary variables

Lesson 1: Always check the stationarity of the y_t, $\mathbf{x}_t$ and the residual. The regression is likely to be spurious if the residual is nonstationary (cannot reject the null hypothesis of the unit root test)!

Lesson 2: Just because two series move together does not mean they are related!

Lesson 3: Use extra caution when you run regression using persistent or nonstationary variables; be aware of the possibility of spurious regressions!

Examples of Spurious Regression

("*Regression that does not make any sense*.")

Typical symptom: "High R^2, t-values, F-value, but **low D/W**"

1. Egyptian infant mortality rate (Y), 1971-1990, annual data,
   on Gross aggregate income of American farmers (I)
   and Total Honduran money supply (M)

   $$\hat{Y} = \underset{(16.63)}{179.9} - \underset{(-2.32)}{.2952}\, I - \underset{(-4.26)}{.0439}\, M, \qquad R^2 = .918,\ D/W = .4752,\ F = 95.17$$

   Corr = .8858, -.9113, -.9445

2. US Export Index (Y), 1960-1990, annual data,
   on Australian males' life expectancy (X)

   $$\hat{Y} = \underset{(-16.70)}{-2943.} + \underset{(17.76)}{45.7974}\, X, \qquad R^2 = .916,\ D/W = .3599,\ F = 315.2$$

   Corr = .9570

3. US Defense Expenditure (Y), 1971-1990, annual data,
   on Population of South African (X)

   $$\hat{Y} = \underset{(-11.34)}{-368.99} + \underset{(16.75)}{.0179}\, X, \qquad R^2 = .940,\ D/W = .4069,\ F = 280.69$$

   Corr = .9694

4. Total Crime Rates in the US (Y), 1971-1991, annual data,
   on Life expectancy of South Africa (X)

   $$\hat{Y} = \underset{(-6.03)}{-24569} + \underset{(9.04)}{628.9}\, X, \qquad R^2 = .811,\ D/W = .5061,\ F = 81.72$$

   Corr = .9008

5. Population of South Africa (Y), 1971-1990, annual data,
   on Total R&D expenditure in the US (X)

   $$\hat{Y} = \underset{(59.44)}{21698.7} + \underset{(26.40)}{111.58}\, X, \qquad R^2 = .974,\ D/W = .3037,\ F = 696.96$$

   Corr = .9873

- Can a regression with two $I(1)$ variables make sense?
- No, if the regression residuals are $I(1)$.
- **YES!** $\longrightarrow$ If the regression residuals are somehow $I(0)$

- Is this possible?

  In the early 1980's Clive Granger of UCSD was visiting David Hendry at Oxford. At afternoon tea, Hendry claimed that this would be possible but offered no proof. Granger replied that he was sure that Hendry was wrong. On the flight back to CA, Granger set out to prove his conjecture. But Granger realized that he was wrong and had an "Eureka" moment:

Birth of a Nobel Prize III

- If both variables share the **same source of the non-stationarity**
  - ⟶ Both variables move together in the long-run (co-move)
  - ⟶ when both variables are **cointegrated**!
- Granger developed this idea with his UCSD colleague Rob Engle and both received the Nobel Prize in Economics in 2003
- Cointegration plays an important role in many areas of finance

Idea:

$$z_t = z_{t-1} + e_t \ \sim I(1)$$

$$y_t = \boxed{z_t} + u_t \ \sim I(1)$$

$$x_t = \boxed{z_t} + w_t \ \sim I(1)$$

$$\Rightarrow x_t - y_t = v_t \ \sim I(0)!!$$

More general:

$$z_t \sim \text{ARIMA}(p, 1, q)$$

$$a_t, b_t \sim \text{ARIMA}(p, 0, q)$$

$$x_t = \alpha_x \boxed{z_t} + b_t \sim I(1)$$

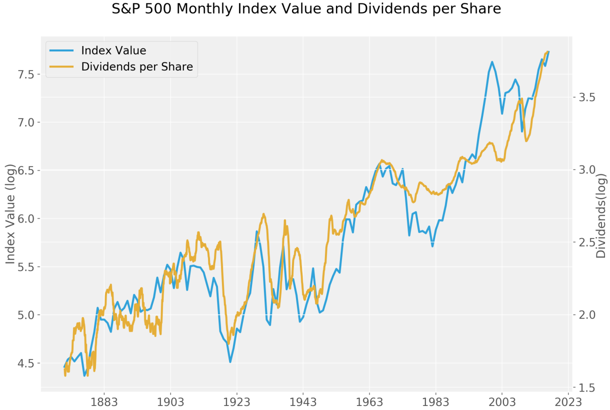
$$y_t = \alpha_y \boxed{z_t} + a_t \sim I(1)$$

$$\Rightarrow \frac{x_t}{\alpha_x} - \frac{y_t}{\alpha_y} = v_t \sim \text{ARIMA}(p, 0, q)$$

Regression $y_t = \mu + \beta x_t + \epsilon_t$ is OK since $\epsilon_t \sim I(0)$

Example: S&P 500 dividend and prices

Stock prices and dividends are both $I(1)$ but are linked together



Example: S&P 500 dividend and prices

Future topic:

Investigate further how and why dividends (or more generally fundamentals, such as earnings, book values, sales, ...) are linked to prices. This will lead us to ask what "market efficiency" means. Stay tuned!

- OLS is well behaved as long as the time series is stationary and ergodic
- The properties of the OLS estimator are completely different for nonstationary time series: Non-standard limiting distribution, superconsistent, biased t-stats, nuisance parameters
- In practice, of AR coefficient is “close” to 1, be careful!
- DF and ADF tests for nonstationarity
- In practice: Differencing is often a cure for nonstationarity
- Spurious regressions

SUMMARY OF LINEAR TIME SERIES MODELS

- White noise process
- AR(1) model: autocorrelations, forecasting, role of ϕ
- AR(p) models: more complex autocorrelation patterns
- MA(q) and ARMA(p, q) models
- Lag operators, inversion of AR and MA, Wold representation
- Estimation of AR(1) : Kendall bias, lag length
- Multivariate ARMA processes
- Stationarity, random walk, unit root
- Non-stationarity and unit roots, testing for unit roots, DF test
- Spurious regressions