Lecture 12

Time series

Lectures in SDPE: *Econometrics I* on March 6, 2024

Markus Jäntti
Swedish Institute For Social Research
Stockholm University

Purpose of lecture

- To introduce basic concepts when data consist of time series.
- Discuss central issues for univariate and multivariate time series, including non-stationarity.

Introduction

- A time series Y_t, $t = 1, \ldots, T$ is observed in discrete time.
- Y_t and Y_s for $s \neq t$ can not be considered independent. The series $(\ldots, Y_1, \ldots, Y_T, \ldots)$ should be considered a multivariate random variable but can be reduced to manageable dimensions by imposing more structure.
- If Y_t is scalar ($\mathbb{R}$), it is a *univariate* time series.
  - The primary model for these are autoregressive (AR) models.
  - Moving average (MA) and autoregressive moving average (ARMA) models are also used.
  - A major concern is if such series are stationary or if they have a *unit root*.
  - Univariate time series relevant also for *panel* (longitudinal) data.
- If an m-dimensional vector ($\mathbb{R}^m$), it is a multivariate times series.
  - The primary model for such time series is vector autoregression (VAR).
  - *Cointegration* is a major concern for multivariate time series.

Stationarity and Ergodicity – Definitions

- The time series $\{Y_t\}$ is *covariance stationary* (weakly stationary) if

$$\mathrm{E}[Y_t] = \mu \text{ and } \mathrm{Cov}[Y_t, Y_{t-k}] = \gamma(k) \tag{1}$$

  do not depend on t.
- $\gamma(k)$ is called the *autocovariance function* with $\gamma(0) = \mathrm{Cov}[Y_t, Y_{t-0}] = \mathrm{Var}[Y_t]$. (Note that $\gamma(k) \equiv \gamma_k$; the latter is used for notational convenience at times.)
- The *autocorrelation function* is

$$\rho(k) = \frac{\gamma(k)}{\gamma(0)} = \mathrm{Corr}[Y_t, Y_{t-k}] \tag{2}$$

- The series is said to be *strictly stationary* if the joint distribution of $(Y_t, \ldots, Y_{t-k})$ is independent of t for all k.
- A stationary time series is *ergodic* if $\gamma(k) \to 0$ as $k \to \infty$.

Stationarity and Ergodicity – Theorems

- For a strictly stationary and ergodic Y_t, a function $\boldsymbol{X}_t = f(Y_t, Y_{t-1}, \ldots)$ is strictly stationary and ergodic.
- (Ergodic theorem) For a strictly stationary and ergodic Y_t that is absolutely convergent (i.e., $\mathrm{E}[|Y_t|] < \infty$), as $T \to \infty$,

$$\frac{1}{T}\sum_{t=1}^{T} Y_t \xrightarrow{p} \mathrm{E}[Y_t]. \tag{3}$$

  Sample analogs of the moments:
  - Mean: $\hat{\mu} = T^{-1}\sum_{t=1}^{T} Y_t$
  - Sample autocovariance: $\hat{\gamma}(k) = T^{-1}\sum_{t=1}^{T}(Y_t - \hat{\mu})(Y_{t-k} - \hat{\mu})$
  - Sample autocorrelation: $\hat{\rho}(k) = \hat{\gamma}(k)/\hat{\gamma}(0)$.
- For a strictly stationary and ergodic Y_t with finite second moment $\mathrm{E}[{Y_t}^2] < \infty$,
  1. $\hat{\mu} \xrightarrow{p} \mathrm{E}[Y_t] = \mu$
  2. $\hat{\gamma}(k) \xrightarrow{p} \gamma(k)$
  3. $\hat{\rho}(k) \xrightarrow{p} \rho(k)$

Autoregressions

- Recall that $(\ldots, Y_1, \ldots, Y_T, \ldots)$ should be considered jointly random.
- Consider the expectation of Y_t conditional on its past history $\mathcal{F}_{t-1}$

$$\mathrm{E}[Y_t|\mathcal{F}_{t-1}] = \mathrm{E}[Y_t|Y_{t-1}, Y_{t-2}, \ldots]. \tag{4}$$

- An autoregressive model of order p ($AR(p)$) limits the relevant history to only p lags (imposing linearity):

$$\begin{aligned} \mathrm{E}[Y_t|\mathcal{F}_{t-1}] =& \mathrm{E}[Y_t|Y_{t-1}, \ldots, Y_{t-p}] \\ =& \alpha + \rho_1 Y_{t-1} + \rho_2 Y_{t-2} + \cdots + \rho_k Y_{t-p}. \end{aligned} \tag{5}$$

- Using

$$e_t = Y_t - \mathrm{E}[Y_t|\mathcal{F}_{t-1}] \tag{6}$$

the autoregressive model can be written as

$$\begin{aligned} Y_t =& \alpha + \rho_1 Y_{t-1} + \rho_2 Y_{t-2} + \cdots + \rho_k Y_{t-p} + e_t \\ \mathrm{E}[e_t|\mathcal{F}_{t-1}] =& 0. \end{aligned} \tag{7}$$

- Because $\mathrm{E}[e_t|\mathcal{F}_{t-1}] = 0$, e_t is a *martingale difference sequence* (MDS).
- MDS is analogous to the conditional mean zero assumption of cross-section regression error.
- MDS:s are central to time series, including the distributional theory of estimators.
- A property of MDS is that e_t is uncorrelated with any function of $\mathcal{F}_{t-1}$, so for $p > 0$, $\mathrm{E}[Y_{t-p}e_t] = 0$.

Stationarity of AR(1) Process

- Consider the simplest AR process, the zero mean AR(1)

$$Y_t = \rho Y_{t-1} + e_t \tag{8}$$

with iid e_t, $\mathrm{E}[e_t] = 0$ and $\mathrm{Var}[e_t] = \sigma^2 < \infty$.
- Re-write using back substitution as

$$Y_t = \rho(\underbrace{\rho\, Y_{t-2} + e_{t-1}}_{Y_{t-1}}) + e_t = \sum_{p=0}^{\infty} \rho^p e_{t-p}. \tag{9}$$

- The above infinite series converges if $\rho^k e_{t-k}$ becomes small as $k \to \infty$, which takes place if $|\rho| < 1$. If so, then the AR(1) is strictly stationary and ergodic.
- The moments on the AR(1) can be computed from eq 9 as

$$\begin{aligned} \mathrm{E}[Y_t] =& \mathrm{E}[\sum_{p=0}^{\infty} \rho^p e_{t-p}] = \sum_{p=0}^{\infty} \rho^p \mathrm{E}[e_{t-p}] = 0 \\ \mathrm{Var}[Y_t] =& \sum_{p=0}^{\infty} \rho^{2p} \mathrm{Var}[e_t] = \frac{\sigma^2}{1-\rho^2} \end{aligned} \tag{10}$$

Stationarity of AR(1) Process

- If the AR(1) is not zero mean, the equation has an intercept and

$$\mathrm{E}[Y_t] = \mathrm{E}[\alpha + \rho Y_{t-1} + e_t] = \alpha + \rho \mathrm{E}[Y_{t-1}] \tag{11}$$

- As the AR(1) is stationary, $\mathrm{E}[Y_t] \equiv \mathrm{E}[Y_{t-1}]$ and

$$\mathrm{E}[Y_t] = \frac{\alpha}{1 - \rho} \tag{12}$$

Stationarity of AR(p)

- A mean zero AR(p) model is

$$Y_t = \rho_1 Y_{t-1} + \rho_2 Y_{t-2} + \cdots + \rho_p Y_{t-p} + e_t \tag{13}$$

  for iid and well-behaved e_t.
- Using the *lag* operator L ($LY_t = Y_{t-1}$), this can be re-expressed as

$$(1 - \rho_1 L - \rho_2 L^2 - \cdots - \rho_p L^p) Y_t = e_t \text{ or } \rho(L) = e_t \tag{14}$$

  $\rho(L)$ is the autoregressive polynomial of Y_t.
- The AR polynomial can be factored as

$$\rho(z) = (1 - \lambda_1^{-1} z)(1 - \lambda_2^{-1} z) \ldots (1 - \lambda_p^{-1} z) \tag{15}$$

- The $(\lambda_1, \lambda_2, \ldots, \lambda_p)$ are the complex roots of $\rho(z)$ which satisfy $\rho(\lambda_j) = 0$.
- For an AR(1), the $|\rho| < 1$ ensures stationarity. Analogously, for an AR(p), all roots must be larger then one, for all moduli $|\lambda_j| > 1$.
- An AR(p) is strictly stationary and ergodic if all moduli $|\lambda_j| > 1$ ("all roots are outside the unit circle").
- If any root is one, Y_t is said to have a unit root, an important special case of non-stationarity.

- If an AR(p) process is stationary, it can be estimated using LS, with $\boldsymbol{X}_t = (1, Y_{t-1}, Y_{t-2}, \ldots, Y_{t-p})'$.
- The key to asymptotic properties is $u_t = \boldsymbol{X}_t e_t$ which is a MDS since

$$\mathrm{E}[u_t|\mathcal{F}_{t-1}] = \mathrm{E}[\boldsymbol{X}_t e_t|\mathcal{F}_{t-1}] = \boldsymbol{X}_t \mathrm{E}[e_t|\mathcal{F}_{t-1}] = 0. \tag{16}$$

- It is also strictly stationary and ergodic so

$$\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{X}_t e_t = \frac{1}{T}\sum_{t=1}^{T}u_t \xrightarrow{p} \mathrm{E}[u_t] = 0. \tag{17}$$

- As $\boldsymbol{X}_t$ is strictly stationary and ergodic, so is $\boldsymbol{X}_t\boldsymbol{X}_t{}'$ so

$$\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{X}_t\boldsymbol{X}_t{}' \xrightarrow{p} \mathrm{E}[\boldsymbol{X}_t\boldsymbol{X}_t{}'] = \boldsymbol{Q} \tag{18}$$

- Then we know that the LS estimator

$$\widehat{\beta} = \boldsymbol{\beta} + \left(\sum_{t=1}^{T} \boldsymbol{X}_t \boldsymbol{X}_t{}'\right)^{-1} \left(\sum_{t=1}^{T} \boldsymbol{X}_t e_t\right) \xrightarrow{p} \boldsymbol{Q}^{-1}\mathbf{0} \tag{19}$$

- Thus, for an AR(p) Y_t that is strictly stationary and ergodic and $\mathrm{E}[Y_t] < \infty$,

$$\widehat{\beta} \xrightarrow{p} \beta \text{ as } T \to \infty. \tag{20}$$

Asymptotic Distribution

- For u_t that is a strictly stationary and ergodic MDS with $\mathrm{E}[u_t u_t'] = \boldsymbol{\Omega} < \infty$, the martingale CLT says

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} u_t \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Omega}). \tag{21}$$

- This can be used to show that $\widehat{\beta}$

$$\sqrt{T}(\widehat{\beta} - \boldsymbol{\beta}) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{Q}^{-1} \boldsymbol{\Omega} \boldsymbol{Q}^{-1}) \tag{22}$$

where $\boldsymbol{\Omega} = \mathrm{E}[\boldsymbol{X}_t \boldsymbol{X}_t{}' e_t^2]$.

Moving average processes

- A moving average process of order 1 ($MA(1)$) process is

$$Y_t = e_t + \theta e_{t-1}, \tag{23}$$

  with iid e_t, $\mathrm{E}[e_t] = 0$ and $\mathrm{Var}[e_t] = \sigma^2 < \infty$.
- Its expectation is

$$\begin{aligned} \mathrm{E}[Y_t] &= \mathrm{E}[e_t + \theta e_{t-1}] \\ &= \mathrm{E}[e_t] + \mathrm{E}[\theta e_{t-1}] \\ &= 0. \end{aligned} \tag{24}$$

- And its variance is

$$\begin{aligned} \mathrm{Var}[Y_t] =& \mathrm{E}[(Y_t)^2] \\ &= \mathrm{E}[(e_t + \theta e_{t-1})^2] \\ &= \mathrm{E}[e_t^2 + 2\theta e_t e_{t-1} + \theta^2 e_{t-1}^2] \\ &= \mathrm{E}[e_t^2] + 2\theta \mathrm{E}[e_t e_{t-1}] + \theta^2 \mathrm{E}[e_{t-1}^2] \\ &= \mathrm{E}[e_t^2] + \theta^2 \mathrm{E}[e_{t-1}^2] \\ &= (1 + \theta^2)\sigma^2. \end{aligned} \tag{25}$$

Moving average processes

- The autocovariance between Y_t and Y_{t-1} is

$$\begin{aligned}\gamma(1) &= \mathrm{Cov}[Y_t, Y_{t-1}] = \mathrm{E}[(Y_t)(Y_{t-1})] \\ &= \mathrm{E}[(e_t + \theta e_{t-1})(e_{t-1} + \theta e_{t-2})] \\ &= \theta \mathrm{E}[e_{t-1}^2] \\ &= \theta\sigma^2.\end{aligned} \tag{26}$$

$$\tag{27}$$

- The autocovariance between Y_t and Y_{t-2} is

$$\begin{aligned}\gamma(2) &= \mathrm{Cov}[Y_t, Y_{t-2}] = \mathrm{E}[(Y_t)(Y_{t-2})] \\ &= \mathrm{E}[(e_t + \theta e_{t-1})(e_{t-2} + \theta e_{t-3})] \\ &= 0.\end{aligned} \tag{28}$$

  In general, $\mathrm{Cov}[Y_t, Y_{t-p}] = 0$ for $p = 2, 3, 4, \ldots$

- If $|\theta| < 1$, then we can transform (invert) the $MA(1)$ process into an $AR(\infty)$ process.

Moving average processes

- An example of a $MA(q)$ process

$$Y_t = e_t + \theta_1 e_{t-1} + \ldots + \theta_q e_{t-q} \tag{29}$$

with iid e_t, $\mathrm{E}[e_t] = 0$ and $\mathrm{Var}[e_t] = \sigma^2 < \infty$.

- The *invertability* condition for general $MA(q)$ processes is that the inverted roots of the lag polynomial lie inside the unit circle.
- Denoting the autocovariance at lag k with $\gamma(k)$, we have

$$\begin{aligned}
\gamma(0) =& \mathrm{E}[{Y_t}^2] = \mathrm{E}[(e_t + \theta_1 e_{t-1} + \ldots + \theta_q e_{t-q})^2] \\
=& \sigma^2(1 + \theta_1^2 + \cdots + \theta_q^2) \\
\gamma(1) =& \mathrm{E}[Y_t Y_{t-1}] = \mathrm{E}[(e_t + \theta_1 e_{t-1} + \ldots + \theta_q e_{t-q})(e_{t-1} + \theta_1 e_{t-2} + \ldots + \theta_q e_{t-1-q})] \\
=& \theta_1 \mathrm{E}[e_{t-1} e_{t-1}] \ldots \theta_{q-1}\theta_q \mathrm{E}[e_{t-q} e_{t-q}] \\
=& \sigma^2(\theta_1 + \sum_{j=1}^{q-1} \theta_j \theta_{j+1}) \\
\gamma(k) =& \sigma^2(\theta_1 + \sum_{j=1}^{q-k} \theta_j \theta_{j+k}), \quad k \leq q
\end{aligned} \tag{30}$$

A General ARMA Process

- We can now formulate a more general autoregressive moving average process, $ARMA(p, q)$

$$Y_t = \rho_1 Y_{t-1} + \ldots \rho_p Y_{t-p} + e_t + \theta_1 e_{t-1} + \ldots + \theta_q e_{t-q} \tag{31}$$

  with iid e_t, $\mathrm{E}[e_t] = 0$ and $\mathrm{Var}[e_t] = \sigma^2 < \infty$.
- For nonstationary series, we formulate an $ARIMA(p, d, q)$ model.

- The $ARMA(p,q)$ process is

$$Y_t = \rho_1 Y_{t-1} + \ldots \rho_p Y_{t-p} + e_t + \theta_1 e_{t-1} + \ldots + \theta_q e_{t-q} \tag{32}$$

  with iid e_t, $\mathrm{E}[e_t] = 0$ and $\mathrm{Var}[e_t] = \sigma^2 < \infty$.
- We assume stationarity; weak stationarity is enough, as it gives us $\mathrm{E}[Y_t Y_{t-s}] = \mathrm{E}[Y_{t-k} Y_{t-k-s}] = \gamma_s \; \forall s$, which is used a lot below.

The autocovariance function for ARMA(1,1)

- We focus on the $ARMA(1,1)$,

$$Y_t = \rho Y_{t-1} + e_t + \theta e_{t-1}. \tag{33}$$

- Set up the Yule-Walker equations for k as

$$\mathrm{E}[Y_t Y_{t-k}] = \ldots, k = 0, 1, \ldots \tag{34}$$

- It is sufficient to set these up for $k = 0, 1, 2$:

$$\begin{aligned} \mathrm{E}[Y_t Y_t] =& \mathrm{E}[(\rho Y_{t-1} + e_t + \theta e_{t-1}) Y_t] \\ \mathrm{E}[Y_t Y_{t-1}] =& \mathrm{E}[(\rho Y_{t-1} + e_t + \theta e_{t-1}) Y_{t-1}] \\ \mathrm{E}[Y_t Y_{t-2}] =& \mathrm{E}[(\rho Y_{t-1} + e_t + \theta e_{t-1}) Y_{t-2}] \end{aligned} \tag{35}$$

The autocovariance function for ARMA(1,1)

- For $k = 0$ we get

$$\begin{aligned}\gamma_0 = \mathrm{E}[Y_t Y_t] &= \mathrm{E}[Y_t(\rho Y_{t-1} + e_t + \theta e_{t-1})] \\ &= \rho \underbrace{\mathrm{E}[Y_t Y_{t-1}]}_{=\gamma_1} + \underbrace{\mathrm{E}[Y_t e_t]}_{=\sigma^2} + \theta \underbrace{\mathrm{E}[Y_t e_{t-1}]}_{=(\rho+\theta)\sigma^2} \\ &= \rho\gamma_1 + \sigma^2 + \theta(\rho + \theta)\sigma^2.\end{aligned} \tag{36}$$

- The result that $\mathrm{E}[Y_t e_{t-1}] = (\rho + \theta)\sigma^2$ requires the substitution of the expression for the $ARMA(1, 1)$ twice as in

$$\begin{aligned}\mathrm{E}[Y_t e_{t-1}] &= \mathrm{E}[(\rho Y_{t-1} + e_t + \theta e_{t-1})e_{t-1}] \\ &= \mathrm{E}[(\rho(\rho Y_{t-2} + e_{t-1} + \theta e_{t-2}) + e_t + \theta e_{t-1})e_{t-1}] \\ &= \rho^2 \underbrace{\mathrm{E}[Y_{t-2} e_{t-1}]}_{=0} + \rho \underbrace{\mathrm{E}[e_{t-1}^2]}_{=\sigma^2} + \rho \underbrace{\mathrm{E}[e_{t-2} e_{t-1}]}_{=0} + \\ &\quad \underbrace{\mathrm{E}[e_t e_{t-1}]}_{=0} + \theta \underbrace{\mathrm{E}[e_{t-1}^2]}_{=\sigma^2} \\ &= (\rho + \theta)\sigma^2.\end{aligned} \tag{37}$$

The autocovariance function for ARMA(1,1)

- For $k = 1$ we have

$$\begin{aligned}\gamma_1 = \mathrm{E}[Y_t Y_{t-1}] &= \mathrm{E}[(\rho Y_{t-1} + e_t + \theta e_{t-1}) Y_{t-1}] \\ &= \rho \underbrace{\mathrm{E}[Y_{t-1} Y_{t-1}]}_{=\gamma_0} + \underbrace{\mathrm{E}[e_t Y_{t-1}]}_{=0} + \theta \underbrace{\mathrm{E}[e_{t-1} Y_{t-1}]}_{=\sigma^2} \\ &= \rho\gamma_0 + \theta\sigma^2\end{aligned} \tag{38}$$

- For $k = 2$ we have

$$\begin{aligned}\gamma_2 = \mathrm{E}[Y_t Y_{t-2}] &= \mathrm{E}[(\rho Y_{t-1} + e_t + \theta e_{t-1}) Y_{t-2}] \\ &= \rho \underbrace{\mathrm{E}[Y_{t-1} Y_{t-2}]}_{=\gamma_1} + \underbrace{\mathrm{E}[e_t Y_{t-2}]}_{=0} + \theta \underbrace{\mathrm{E}[e_{t-1} Y_{t-2}]}_{=0} \\ &= \rho\gamma_1.\end{aligned} \tag{39}$$

The autocovariance function for ARMA(1,1)

- We now have three equations, in three unknowns:

$$\begin{aligned}\gamma_0 =& \rho\gamma_1 + \sigma^2 + \theta(\rho+\theta)\sigma^2 \\ \gamma_1 =& \rho\gamma_0 + \theta\sigma^2 \\ \gamma_2 =& \rho\gamma_1 .\end{aligned} \tag{40}$$

The autocovariance function for ARMA(1,1)

- We want to express $\gamma_k,\ k = 0, 1, 2$ in terms of the "primitive" parameters (ρ, θ, σ^2).
- We accomplish this by solving (γ_0, γ_1) from the first two (and $k = 2, 3, \ldots$ follow from the form of $\gamma_k, k > 1$).
- It should be relatively easy to see that the solution (to use the notation that $\gamma_k = \gamma(k)$) is:

$$\begin{aligned}
\gamma(0) &= \frac{1 + \theta^2 + 2\rho\theta}{1 - \rho^2}\sigma^2 \\
\gamma(1) &= \frac{(1 + \rho\theta)(\rho + \theta)}{(1 - \rho^2)}\sigma^2 \\
\gamma(2) &= \rho\gamma_1 \\
\gamma(s) &= \rho\gamma_{s-1}, \quad s \geq 2
\end{aligned} \tag{41}$$

These equations are derived with discussion in e.g. Enders (2010).

- Consider the AR(p) model

$$\rho(L)Y_t = \mu + e_t, \rho(L) = 1 - \rho_1 L - \rho_2 L^2 - \cdots - \rho_p L^p. \tag{42}$$

- If Y_t has a unit root,

$$\rho(1) = 0 \Leftrightarrow \rho_1 + \rho_2 + \cdots + \rho_p = 1 \tag{43}$$

- In that case, Y_t is non-stationary and the MDS CLT and ergodic theorems do not apply. Importantly, the estimators and test statistics that are based on those fail to apply and asymptotic distribution results do not hold.

- The Dickey-Fuller (re)parameterization, using $\Delta = 1 - L$, of eq 42 is

$$\Delta Y_t = \mu + \alpha_0 Y_{t-1} + \alpha_1 \Delta Y_{t-1} + \cdots + \alpha_{p-1} \Delta Y_{t-p} + e_t \tag{44}$$

- The lag polynomial of this is

$$(1 - L) - \alpha_0 L - \alpha_1 (L - L^2) - \cdots - \alpha_{p-1} (L^{p-1} - L^p) = \rho(L) \tag{45}$$

  so in the unit root case

$$\rho(1) = -\alpha_0. \tag{46}$$

- The null hypothesis of a unit root, against the alternative of stationarity is

$$\mathbb{H}_0 : \alpha_0 = 0 \text{ against } \mathbb{H}_1 : \alpha_0 < 0. \tag{47}$$

- Under $\mathbb{H}_0$, we have the AR(p-1) process for Y_t:

$$\Delta Y_t = \mu + \alpha_1 \Delta Y_{t-1} + \cdots + \alpha_{p-1} \Delta Y_{t-(p-1)} + e_t \tag{48}$$

  If Y_t has a unit root, ΔY_t is stationary. We say Y_t is $I(1)$. A series is $I(d)$ if it needs to be differenced d times to be stationary.

- A "natural" test for $\mathbb{H}_0$ is to estimate α_0 by LS. Called a "augmented Dickey-Fuller" test statistic, it has a non-standard distribution.
- The Dickey-Fuller theorem has it that

$$T\hat{\alpha}_0 \xrightarrow{d} (1 - \alpha_1 - \alpha_2 - \cdots - \alpha_{p-1}) DF_\alpha \text{ and } ADF = \frac{\hat{\alpha}_0}{s(\hat{\alpha}_0)} \to DF_t \tag{49}$$

  DF_α and DF_t are limiting non-normal distributions.

- There are DF distributions for models w/o an intercept as well as with time trend.

Spurious regression

- Consider the regression equation(with $(Y_{t-1}, \boldsymbol{Z}_t, 1) = \boldsymbol{X}_t'$):

$$
\begin{aligned}
Y_t =& \beta_1 Y_{t-1} + \beta_2 \boldsymbol{Z}_t + \beta_3 + e_t = \boldsymbol{X}_t'\beta + e_t, \\
&\mathrm{E}[e_t | \boldsymbol{X}_t] = 0, \ \mathrm{E}[e_t^2 | \boldsymbol{X}_t] = \sigma^2.
\end{aligned}
\tag{50}
$$

- We shall now examine two endemic problems with time series, namely what happens to estimators when $\beta_1 = 1$ (so Y_t has a unit root) and when $\boldsymbol{Z}_t$ has a unit root.

- First, set $\beta_3 = 0$ (no intercept), $\beta_2 = 0$ (eliminate $\boldsymbol{Z}_t$) and let $\beta_1 = 1$ (which, of course, we do not know).
- Such an Y_t is a unit root process known as a “random walk”.
- We can re-write eq. 50 as

$$
\begin{aligned}
Y_t &= \underbrace{Y_{t-1}}_{Y_{t-2}+e_{t-1}} + e_t = Y_{t-2} + e_{t-1} + e_t \\
&\vdots \\
&= \sum_{s=0}^{t} e_{t-s}.
\end{aligned}
\tag{51}
$$

- Suppose we now try to estimate

$$Y_t = \rho Y_{t-1} + e_t. \tag{52}$$

- The LS estimator is

$$\widehat{\rho} = \frac{\sum_{t=1}^{T} Y_t Y_{t-1}}{\sum_{t=1}^{T} {Y_{t-1}}^2}. \tag{53}$$

- We would normally assume its asymptotic distribution to be

$$\sqrt{T}(\widehat{\rho} - \rho) \xrightarrow{d} N(0, \sigma^2 \boldsymbol{Q}^{-1}), \tag{54}$$

with

$$\boldsymbol{Q} = \mathrm{E}[{Y_t}^2]. \tag{55}$$

Spurious regression

- To understand what happens, examine

$$\frac{1}{T}\sum_{t=1}^{T} {Y_t}^2 = \frac{1}{T}\sum_{t=1}^{T}\left(\sum_{s=0}^{t} e_{t-s}\right)^2. \tag{56}$$

- The e are iid, so we have

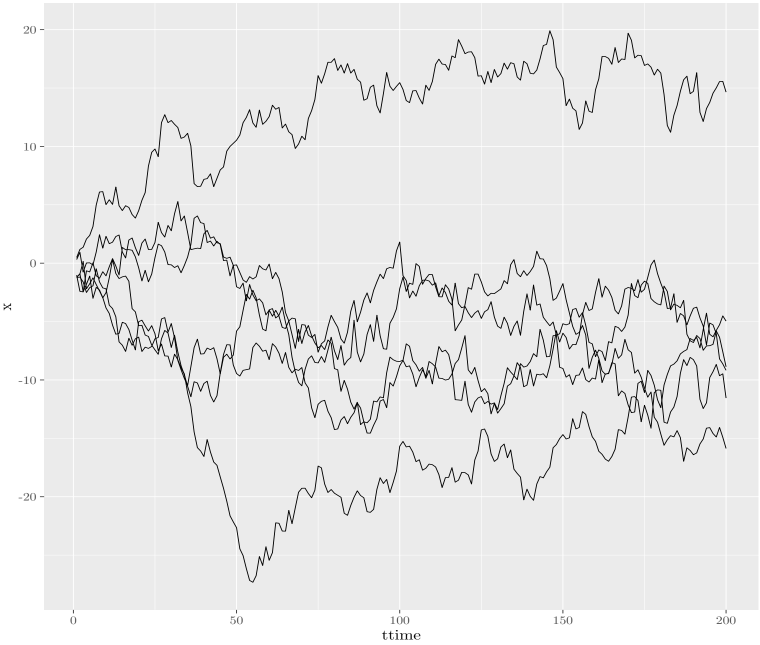
$$\begin{aligned} \mathrm{E}\left[\frac{1}{T}\sum_{t=1}^{T} {Y_t}^2\right] &= \frac{1}{T}\sum_{t=1}^{T}\sum_{s=0}^{t}\mathrm{E}[e_{t-s}^2] = \frac{1}{T}\sum_{t=1}^{T} t\sigma^2 \\ &= \sigma^2\frac{1}{T}T(T+1)/2 = \sigma^2(T+1)/2. \end{aligned} \tag{57}$$

- Thus, the standard assumption of an invertible $\boldsymbol{Q}$ does not hold. . .
- . . . and if the "plug-in" estimator $\widehat{\boldsymbol{Q}}$ is used as a basis for statistical inference, the variance estimator of the asymptotic distribution goes to zero as sample size increases.
- The situation is, in fact, worse in that the true asymptotic distribution of $\widehat{\rho}$ is not centered on ρ (we have not developed the theory of examining the true distribution).

Spurious regression

- The term "spurious regression" refers to the case (starting from eq. 50) when $\beta_2 \neq 0$ but $\boldsymbol{Z}_t$ has a unit root.
- What happens can be seen e.g. by studying $\boldsymbol{Q} = \mathrm{E}[\boldsymbol{Z}_t{}^2]$ when it is a random walk – it is not finite and therefore is not invertible.
- Using the plug-in estimator from standard least squares estimation techniques results in an approximation of an asymptotic sampling distribution that in fact converges onto a point so you end up rejecting the null with probability one.
- For more discussion, see Hendry (1995), Davidson and MacKinnon (2004) or Enders (2010).

Demo of spurious regression: 6 random walks



The spurious regression

```
summary(reg1 <- lm(x1 ~ x2 + x3 + x4 + x5 + x6, subset=dtime>1))

##
## Call:
## lm(formula = x1 ~ x2 + x3 + x4 + x5 + x6, subset = dtime > 1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -12.190  -1.853   0.247   2.806   7.520
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -2.9139     1.3634   -2.14    0.034
## x2           -0.4002     0.1202   -3.33    0.001
## x3           -0.1824     0.0778   -2.34    0.020
## x4           -0.1511     0.0944   -1.60    0.111
## x5            0.3140     0.1266    2.48    0.014
## x6            0.9963     0.1140    8.74  1.1e-15
##
## Residual standard error: 4.2 on 193 degrees of freedom
## Multiple R-squared:  0.51,Adjusted R-squared:  0.497
## F-statistic: 40.1 on 5 and 193 DF,  p-value: <2e-16
```

Use t-ratios to reduce model

```
tx <- summary(reg1)$coefficients[, "t value"]
these.xs <- names(tx)[abs(tx)>2]
these.xs <- these.xs[-1]
ff2 <- as.formula(paste("x1", paste(these.xs, collapse="+"), sep="~"))
```

Use t-ratios to reduce model

```
summary(reg2 <- lm(ff2, subset=dtime>1))

##
## Call:
## lm(formula = ff2, subset = dtime > 1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -12.634  -1.817   0.331   2.790   7.160
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -3.897      1.222   -3.19   0.0017
## x2            -0.312      0.107   -2.91   0.0041
## x3            -0.162      0.077   -2.10   0.0371
## x5             0.278      0.125    2.22   0.0273
## x6             0.924      0.105    8.79  7.7e-16
##
## Residual standard error: 4.2 on 194 degrees of freedom
## Multiple R-squared:  0.503,Adjusted R-squared:  0.493
## F-statistic: 49.1 on 4 and 194 DF,  p-value: <2e-16
```

Adding lags often dramatically changes the results

```
lx1 <- stats::lag(x1, 1)
summary(reg3 <- update(reg2,  ~ . + lx1, subset=dtime>1))
```

```
##
## Call:
## lm(formula = x1 ~ x2 + x3 + x5 + x6 + lx1, subset = dtime > 1)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -6.97e-15 -3.15e-16 -2.10e-17  2.90e-16  6.45e-15
##
## Coefficients:
##              Estimate Std. Error   t value Pr(>|t|)
## (Intercept) -1.47e-15   2.66e-16 -5.50e+00  1.2e-07
## x2           2.21e-16   2.33e-17  9.49e+00  < 2e-16
## x3          -5.21e-17   1.66e-17 -3.15e+00   0.0019
## x5           1.13e-16   2.69e-17  4.20e+00  4.1e-05
## x6           1.78e-16   2.64e-17  6.75e+00  1.7e-10
## lx1          1.00e+00   1.53e-17  6.55e+16  < 2e-16
##
## Residual standard error: 8.9e-16 on 193 degrees of freedom
## Multiple R-squared:      1,Adjusted R-squared:      1
## F-statistic: 1.73e+33 on 5 and 193 DF,  p-value: <2e-16
```

References

- Davidson, Russell and MacKinnon, James G (2004). *Econometric Theory and Methods*. Oxford: Oxford University Press.
- Enders, Walter (2010). *Applied Econometric Time Series*. 3rd. John Wiley & Sons.
- Hendry, David (1995). *Dynamic Econometrics*. Oxford: Oxford University Press.