Univariate Time Series

Time Series, Econ 311: Topic 5

Prof. Harald Uhlig[1]

[1]University of Chicago
Department of Economics
huhlig@uchicago.edu

Winter 2019

Outline

Outline

Shocks

Assumption

$\epsilon_t \in \mathbb{R}$ is a martingale difference sequence with constant, finite variances,

$$E_{t-1}[\epsilon_t] = 0,\ E\left[\epsilon_t^2\right] \equiv \sigma^2,\ 0 < \sigma^2 < \infty.$$

plus additional assumptions, such that the law of large numbers and the central limit theorem hold.

"Shocks". Occasionally, and alternatively:

Assumption

$\epsilon_t \in \mathbb{R}^n$ is a vector martingale difference sequence with constant, finite, positive definite variance-covariance matrix, s.t. LLN and CLT hold and

$$E_{t-1}[\epsilon_t] = 0,\ E\left[\epsilon_t \epsilon_t'\right] = \Omega,\ 0 < \Omega < \infty.$$

Autoregressive Processes, AR

- AR(1):

$$y_t = \rho y_{t-1} + \epsilon_t$$

  Include a constant, if you wish. Notation more messy.

- AR(m):

$$y_t = \sum_{j=1}^{m} \rho_j y_{t-j} + \epsilon_t$$

- Lag operator notation:

$$(1 - \rho(L))y_t = \epsilon_t, \; \rho(L) = \sum_{j=1}^{m} \rho_j L^j$$

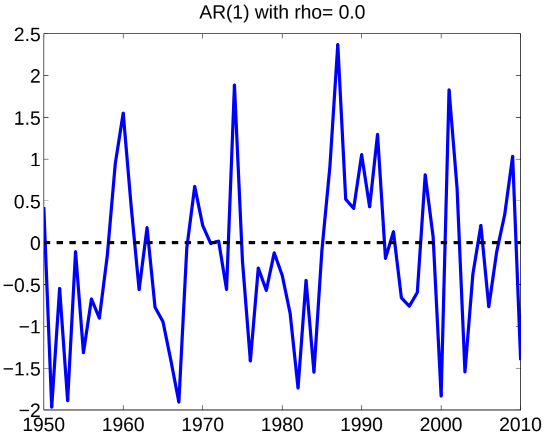
Stacking: VAR(1)

$$x_t = \begin{bmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-m+1} \end{bmatrix} = \begin{bmatrix} \rho_1 & \rho_2 & \cdots & \rho_{m-1} & \rho_m \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix} x_{t-1} + A\epsilon_t, \; A = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

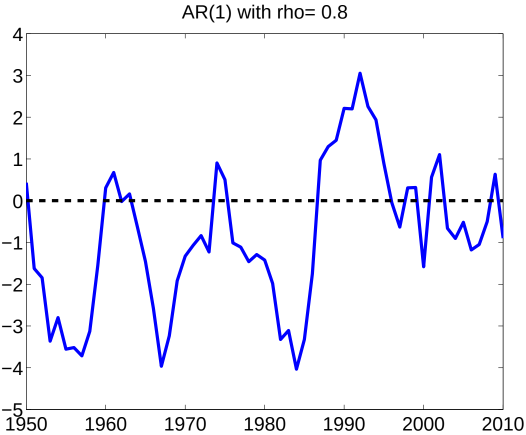
or

$$x_t = Bx_{t-1} + A\epsilon_t$$

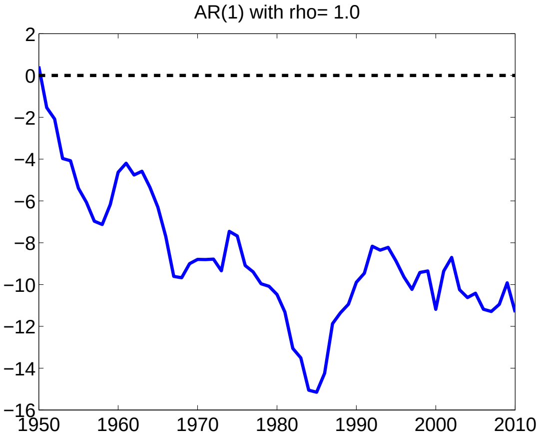
AR(1) with $\rho = 0$.



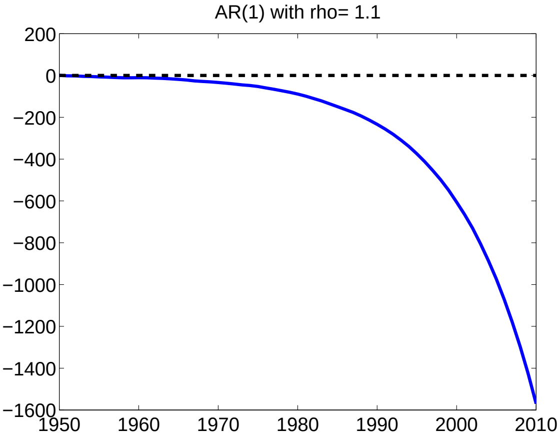
AR(1) with $\rho = 0.8$.



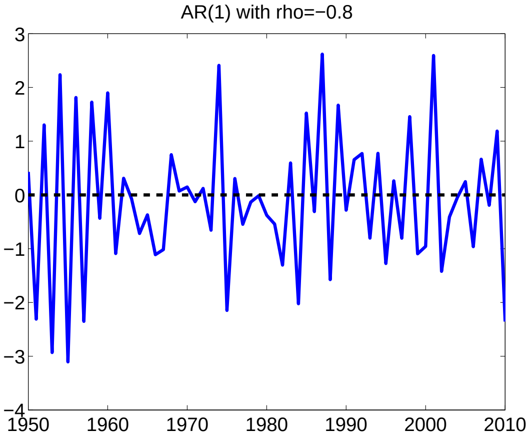
AR(1) with $\rho = 1.0$.



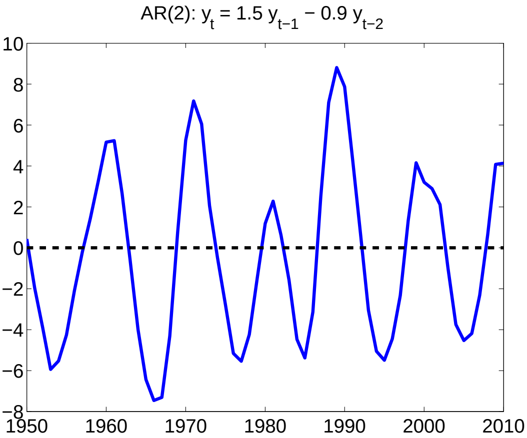
AR(1) with $\rho = 1.1$.



AR(1) with $\rho = -0.8$.



AR(2): $y_t = 1.5y_{t-1} - 0.9y_{t-2}$.



Moving-Average Processes, MA

- MA(1):

$$y_t = \theta_0 \epsilon_t + \theta_1 \epsilon_{t-1}$$

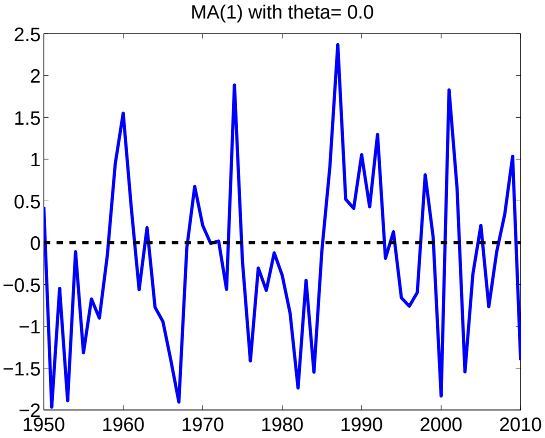
- MA(n):

$$y_t = \sum_{j=0}^{n} \theta_j \epsilon_{t-j}$$

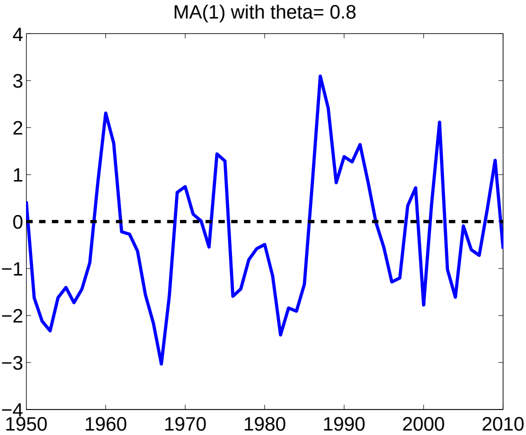
- Lag operator notation:

$$y_t = \theta(L)\epsilon_t, \; \theta(L) = \sum_{j=0}^{n} \theta_j L^j$$

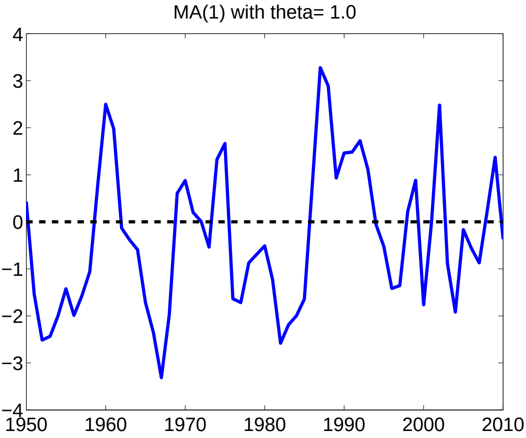
MA(1) with $\theta_1 = 0$.



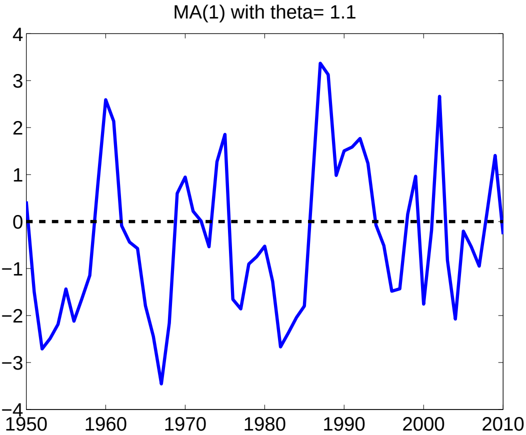
MA(1) with $\theta_1 = 0.8$.



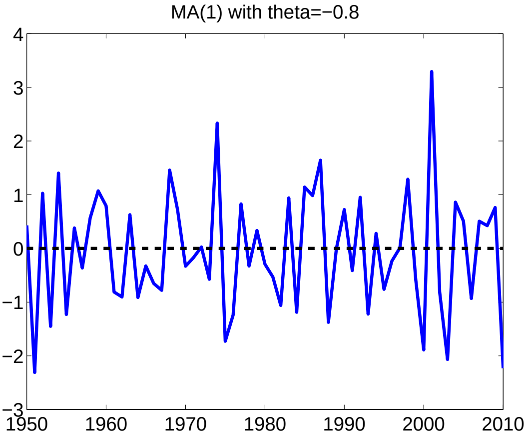
MA(1) with $\theta_1 = 1.0$.



MA(1) with $\theta_1 = 1.1$.



MA(1) with $\theta_1 = -0.8$.



ARMAs

- ARMA(m,n):

$$y_t - \sum_{j=1}^{m} \rho_j y_{t-j} = \sum_{j=0}^{n} \theta_j \epsilon_{t-j}$$

- Lag operator notation:

$$(1 - \rho(L))y_t = \theta(L)\epsilon_t$$

Outline

Autocovariances

Definition

Let y_t be a vector-valued time series.

- The k-th autocovariance $\Gamma_k = E[y_t y_{t-k}']$ at time t is defined as the covariance between y_t and y_{t-k}.
- If the series is univariate, we write γ_k instead of Γ_k.

Note:

$$\Gamma_k = \Gamma_{-k}'$$

AR(1)

- AR(1): $y_t = \rho y_{t-1} + \epsilon_t$
- Variance:

$$\begin{aligned}\gamma_0 = E[y_t y_t] &= E[(\rho y_{t-1} + \epsilon_t)(\rho y_{t-1} + \epsilon_t)] \\ &= \rho^2 E[y_{t-1} y_{t-1}] + \sigma^2\end{aligned}$$

- If $| \rho |< 1$,

$$\gamma_0 = E[y_t y_t] = \frac{\sigma^2}{1-\rho^2}$$

- Information in initial observation. Unconditional likelihood function.
- If $| \rho |\geq 1$,

$$\gamma_0 = E[y_t y_t] = \infty$$

AR(1): Covariance

- Assume $\mid \rho \mid < 1$. Autocovariance:

$$\begin{aligned}
\gamma_k = E[y_t y_{t-k}] &= E[\epsilon_t y_{t-k}] \\
&\quad + \rho E[\epsilon_{t-1} y_{t-k}] \\
&\quad + \ldots \\
&\quad + \rho^k E[y_{t-k} y_{t-k}] \\
&= \rho^k E[y_t y_t] \\
&= \frac{\rho^k \sigma^2}{1-\rho^2}
\end{aligned}$$

- k-th autocorrelation: ρ^k.
- For AR(m): use stacked VAR(1) instead ...

AR(m), VAR(1)

- Assume

$$x_t = Bx_{t-1} + A\epsilon_t,\; E[\epsilon_t \epsilon_t'] = \Omega$$

- Yule-Walker equation:

$$\Gamma_k = E[x_t x_{t-k}'] = BE[x_{t-1}x_{t-k}'] = B\Gamma_{k-1}$$

- Calculate

$$\begin{aligned}
\Gamma_0 = E[x_t x_t'] &= BE[x_{t-1}x_{t-1}']B' + A\Omega A' \\
&= B\Gamma_0 B' + A\Omega A' \\
\mathrm{vec}(\Gamma_0) &= (B \otimes B)\,\mathrm{vec}(\Gamma_0) + \mathrm{vec}(A\Omega A')
\end{aligned}$$

If B only has eigenvalues smaller than unity in absolute value:

$$\mathrm{vec}(\Gamma_0) = (I_{m^2} - B \otimes B)^{-1}\,\mathrm{vec}(A\Omega A')$$

MA(n)

- MA(n), with $\theta_j \in \mathbb{R}^{m \times r}$, $E[\epsilon_t \epsilon_t'] = \Omega$:

$$
\begin{aligned}
y_t &= \sum_{j=0}^{n} \theta_j \epsilon_{t-j} \\
y_{t-k} &= \sum_{i=0}^{n} \theta_i \epsilon_{t-(i+k)} \\
&= \sum_{j=k}^{n+k} \theta_{j-k} \epsilon_{t-j}
\end{aligned}
$$

- Thus,

$$
\Gamma_k = \sum_{j=\max\{0,k\}}^{\min\{n,n+k\}} \theta_j \Omega \theta_{j-k}'
$$

Stationarity

Definition

A stochastic sequence $y_t, t = -\infty, \ldots, \infty$ is called covariance stationary, if the mean of y_t as well as all autocovariances are finite and do not depend on t.

Remark

An AR(1) is covariance stationary, if and only if $\mid \rho \mid < 1$.

(Note: maintained assumption is that (ϵ_t) is a martingale difference sequence with constant variance.)

Outline

The characteristic polynomial

Definition

- For an AR(m) process in $\mathbb{R}$,

$$(1 - \rho(L))y_t = \epsilon_t$$

  define the characteristic polynomial $p(\lambda)$ as

$$\begin{aligned} p(\lambda) &= \lambda^m(1 - \rho(\lambda^{-1})) \\ &= \lambda^m - \rho_1\lambda^{m-1} - \rho_2\lambda^{m-2} - \ldots - \rho_m \end{aligned}$$

- The complex-valued solutions $\lambda_1 \in \mathbb{C}, \ldots, \lambda_m \in \mathbb{C}$ of $p(\lambda) = 0$ are called the roots of the characteristic polynomial.

Stacking

Lemma

The roots of the characteristic polynomial are the eigenvalues of the stacked matrix B and vice versa.

Proof.

Calculate the determinant by "developing" it along the first row,

$$det(\lambda I - B) = det\left(\begin{bmatrix} \lambda - \rho_1 & -\rho_2 & \cdots & -\rho_{m-1} & -\rho_m \\ -1 & \lambda & \cdots & 0 & 0 \\ 0 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & -1 & \lambda \end{bmatrix}\right)$$

$$= p(\lambda)$$

□

The fundamental theorem of algebra

Theorem

The fundamental theorem of algebra: In complex numbers, a polynomial $p(\lambda)$ of $m-$th degree has always exactly m roots $\lambda_1, \ldots, \lambda_m$, provided one permits multiplicity of roots of the same value.

Rewriting the characteristic polynomial

Remark

Let $\lambda_1 \in \mathbb{C}, \ldots, \lambda_m \in \mathbb{C}$ *be the solutions to* $p(\lambda) = 0$*. Then,*

$$1 - \rho(L) \quad = \quad (1 - \lambda_1 L)(1 - \lambda_2 L) \ldots (1 - \lambda_m L)$$

Stationarity

Remark

An AR(m) process is covariance stationary, if all roots $\lambda_i, i = 1, \ldots, m$ *of the characteristic polynomial* $p(\lambda)$ *are smaller than 1 in absolute value,* $\mid \lambda_i \mid < 1$.

Inverting Lag Polynomials, part 1

- An AR(1), $\mid \rho \mid < 1$:

$$\begin{aligned} y_t &= \rho y_{t-1} + \epsilon_t \\ &= \rho^k y_{t-k} + \sum_{j=0}^{k-1} \rho^j \epsilon_{t-j} \\ &\to \sum_{j=0}^{\infty} \rho^j \epsilon_{t-j} \end{aligned}$$

- Or:

$$\begin{aligned} (1 - \rho L) y_t &= \epsilon_t \\ y_t &= \left(\sum_{j=0}^{\infty} (\rho L)^j \right) \epsilon_t \\ &= \frac{1}{1 - \rho L} \epsilon_t \end{aligned}$$

Inverting Lag Polynomials, part 2

- An AR(m), $\mid \lambda_j \mid < 1, j = 1, \ldots, m$.
-

$$\begin{aligned} (1 - \rho(L)) y_t &= \epsilon_t \\ y_t &= \frac{1}{1 - \lambda_1 L} \cdots \frac{1}{1 - \lambda_m L} \epsilon_t \\ y_t &= \sum_{j=0}^{\infty} \theta_j \epsilon_{t-j} \end{aligned}$$

- An AR(m) is an MA(∞).
- Calculate these geometric sums successively. Convolution.
- Note: $\theta_0 = 1$.
- ... or: use VAR(1): next slide.

Inverting Lag Polynomials, part 3

- Assume a VAR(1) with stable eigenvalues,

$$(1 - BL)x_t = A\epsilon_t, \; E[\epsilon_t \epsilon_t'] = \Omega$$

- Thus

$$\begin{aligned} x_t &= B^k x_{t-k} + \sum_{j=0}^{k-1} B^j A \epsilon_{t-j} \\ &\rightarrow \sum_{j=0}^{\infty} B^j A \epsilon_{t-j} \\ &= (1 - BL)^{-1} A \epsilon_t \end{aligned}$$

- (For y_t: extract θ_j for MA(∞) from first row: $\theta_j = (B^j A)_{11}$)
- Note: if B is diagonalizable, $B = VDV^{-1}$, then

$$B^j = VD^jV^{-1}$$

Outline

Wold decomposition (or: Wold representation)

Theorem

Wold Decomposition Theorem (or: Wold Representation Theorem):
Any covariance stationary time series can be represented as

$$y_t = \mu_t + \sum_{j=0}^{\infty} c_j u_{t-j}$$

where $c_0 = 1$ *and where* u_t *are the one-step ahead linear forecast errors for* y_t*, given information on lagged values of* y_{t-j}*,* $j = 1, 2, \ldots$*.*

Some remarks

- $\mu_t = E[y_t]$: unconditional mean. Until now: $= 0$.
- If y_t is covariance stationary, $E[u_t^2] \equiv \sigma_u^2$.
- Linear forecast means

$$u_t = y_t - P(y_t \mid y_{t-1}, y_{t-2}, \ldots)$$

where $P(\cdot \mid \cdot)$ denotes linear projection or best linear prediction on $y_{t-1}, y_{t-2}, \ldots$. I.e., a linear regression of y_t on $y_{t-1}, \ldots, y_{t-q}$ and $q \to \infty$. Regression coefficients: functions of autocovariances γ_j.

Remark

Two processes with the same autocovariances $\gamma_j, j = -\infty, \ldots, \infty$ have the same coefficients c_j in their Wold decomposition and vice versa.

Wold decomposition for AR(m)

Assume stable roots. Recall

$$
\begin{aligned}
(1-\rho(L))y_t &= \epsilon_t \qquad (1)\\
y_t &= \sum_{j=0}^{\infty} \theta_j \epsilon_{t-j} = \epsilon_t + \left(\sum_{j=1}^{\infty} \theta_j L^{j-1} \right) \epsilon_{t-1}
\end{aligned}
$$

(per (1): ϵ_{t-1} is lin. funct. of y_{t-1}, y_{t-2} etc.. So:)

$$
\begin{aligned}
&= \epsilon_t + \left(\sum_{j=1}^{\infty} \theta_j L^{j-1} (1-\rho(L)) \right) y_{t-1}\\
&= \epsilon_t + Q(y_{t-1}, y_{t-2}, \ldots)
\end{aligned}
$$

since $\theta_0 = 1$, where $Q(\cdot)$ is a linear function of $y_{t-1}, y_{t-2}, \ldots$. Since we cannot do better than predicting y_t up to ϵ_t, we must have $\epsilon_t = u_t$, $c_j = \theta_j$ and

$$
P(y_t \mid y_{t-1}, y_{t-2}, \ldots) = Q(y_{t-1}, y_{t-2}, \ldots)
$$

Wold decomposition for AR(m)

Proposition

The Wold decomposition for an AR(m) with stable roots is given by

$$y_t = \sum_{j=0}^{\infty} \theta_j \epsilon_{t-j}$$

i.e. $u_t = \epsilon_t$, $c_j = \theta_j$, *as constructed above.*

Wold decomposition for MA(1): Example 1

- Example 1:

$$y_t = \epsilon_t + 0.5\epsilon_{t-1}$$

  with $E[\epsilon_t^2] = 4$.

- Wold decomposition: fundamental decomposition

$$y_t = \epsilon_t + 0.5\epsilon_{t-1}$$

  with $E[\epsilon_t^2] = 4$, i.e. $u_t = \epsilon_t$, $c_j = \theta_j$.

Wold decomposition for MA(1): Example 2

- Example 2:

$$y_t = \epsilon_t + 2\epsilon_{t-1}$$

  with $E[\epsilon_t^2] = 1$.
- Claim: Wold decomposition

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

  with $E[u_t^2] = 4$ and $\epsilon_t \neq u_t$.

Example 2: a comparison

	$y_t = \epsilon_t + 2\epsilon_{t-1}$ $E[\epsilon_t^2] = 1$	$y_t = u_t + 0.5u_{t-1}$ $E[u_t^2] = 4$
γ_0	5	5
γ_1	2	2
$\gamma_j, j > 1$	0	0

- Same autocovariances, thus same Wold decomposition.
- Needed: deeper insights into the relationship between MA(m)'s and their autocovariances.
- Will do per detour: spectral densities.

Outline

Forecasting with a known Wold decomposition

- Suppose, the coefficients μ_t, c_j in the Wold decomposition are known. E.g. for y_{t+k},

$$y_{t+k} = \mu_{t+k} + \sum_{j=0}^{\infty} c_j u_{t+k-j}$$

- Assume a finite-order AR, finite-order MA, so that one-step ahead prediction errors u_t can be calculated from available data.
- Best linear forecast for y_{t+k}, given $y_t, y_{t-1}, \ldots$:

$$P(y_{t+k} \mid y_t, y_{t-1}, \ldots) = \mu_{t+k} + \sum_{j=k}^{\infty} c_j u_{t+k-j}$$

Forecasting with an unknown Wold decomposition

- In general, c_j have to be estimated.
- What is the appropriate manner to express uncertainty, regarding the forecast?
- Example: AR(1)

$$\begin{aligned} y_t &= \rho y_{t-1} + \epsilon_t \\ P(y_{t+k} \mid y_t, y_{t-1}, \ldots) &= \rho^k y_t \end{aligned}$$

  Needed: an estimator for ρ^k and its distribution.
- MLE$(\rho^k) = ($MLE$(\rho))^k$.
- More to come.

Impulse responses with a known Wold decomposition

- Impulse response: the best forecast for $y_{t+j} - \mu_{t+j}, j \geq 0$, given $u_t = 1$ (or $u_t = \sigma$) and everything else zero.
- Answer:

$$P(y_{t+j} \mid u_t = 1) = c_j$$

- I.e., the coefficients of the Wold decomposition provide the impulse responses to one-step ahead prediction errors.
- Remark: in MA(m), one can also calculate impulse responses to ϵ_t. They may not be the impulse responses from the Wold decomposition, see example 2 above.
- Note also: in general, c_j have to be estimated.

Outline

Time Domain versus Frequency Domain

- Time domain: represent a time series y_t as a function of dates $t = 1, 2, ..., T$. Suppose: t counts quarters.
- Frequency domain: represent a time series as a function of frequencies $\omega \in [-\pi, \pi]$.
  1. ω: fluctuations proportional to $y_t = \cos(\omega t + \phi)$.
  2. $\omega = \pi$: alternations every quarter. For $\phi = 0$: $y_t = -1, 1, -1, 1, \ldots$.
  3. $\omega = \pi/2$: annually recurring events. For $\phi = 0$: $y_t = 0, -1, 0, 1, 0, \ldots$.
  4. $\omega = \pi/20 = 0.158..$: recurring every ten years.
- One can change from one representation to the other (using Fourier transformations).

The exponential function

$$\exp(i\omega) = \cos(\omega) + i \sin(\omega)$$
$$\text{where} \quad i = \sqrt{-1} \in \mathbb{C}$$

Fourier Transformations

- Given a (non-stochastic) sequence $x_j, j = \ldots, -1, 0, 1, \ldots$ with $\sum_{j=-\infty}^{\infty} \mid x_j \mid < \infty$, define

$$\tilde{x}(\omega) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} x_j e^{-ij\omega}$$

- Inverse Fourier transform:

$$x_j = \int_{-\pi}^{\pi} \tilde{x}(\omega) e^{i\omega j} d\omega$$

- One can show: this also works, if $\sum_{j=-\infty}^{\infty} \mid x_j \mid^2 < \infty$.
- This can be extended to covariance-stationary stochastic processes x_t (t instead of j). We will use this heuristically.

White noise

- Suppose, $x_t = \epsilon_t$, where ϵ_t is a martingale-difference sequence with constant variance σ^2.
- In that case, $\tilde{\epsilon}(\omega)$ can be thought of as a collection of independent random variables with constant variance.
- If ϵ_t is normally distributed: Brownian motion increments.

Lag Operator Calculus and Fourier Transforms

- Consider

$$y_t = h(L)x_t = \sum_{j=-\infty}^{\infty} h_j x_{t-j}$$

-

$$\tilde{y}(\omega) = h(e^{-i\omega})\tilde{x}(\omega) = 2\pi\tilde{h}(\omega)\tilde{x}(\omega)$$

- Show this for $h(L) = L^k$. Use linearity.
- Convolution becomes multiplication.

Two special cases

- AR(m):

$$
\begin{aligned}
(1-\rho(L))y_t &= \epsilon_t \\
(1-\rho(e^{-i\omega}))\tilde{y}(\omega) &= \tilde{\epsilon}(\omega)
\end{aligned}
$$

- MA(n):

$$
\begin{aligned}
y_t &= \theta(L)\epsilon_t \\
\tilde{y}(\omega) &= \theta(e^{-i\omega})\tilde{\epsilon}(\omega)
\end{aligned}
$$

Outline

Spectrum

- Let $x_t \in \mathbb{R}$, $t = \ldots, -1, 0, 1 \ldots$ be covariance stationary, mean zero. Recall $\gamma_j = E[x_t x_{t-j}] = \gamma_{-j}$
- Population spectrum per Fourier transformation of γ:
  is defined as

$$\begin{aligned} s_x(\omega) &= \tilde{\gamma}(\omega) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \gamma_j e^{-ij\omega} \\ &\cong E[\tilde{x}(\omega)\overline{\tilde{x}(\omega)}] \end{aligned}$$

- Note: $s_x(\omega) = s_x(-\omega)$. Multivariate: $s_x(\omega) = s_x(-\omega)'$.
- Inverse Fourier transformation:

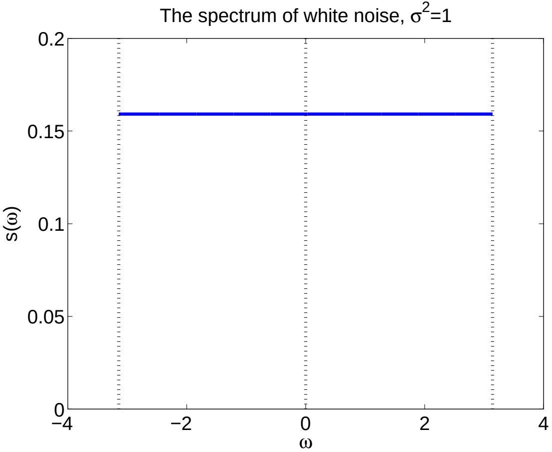
$$\gamma_j = \int_{-\pi}^{\pi} s_x(\omega) e^{i\omega j} d\omega$$

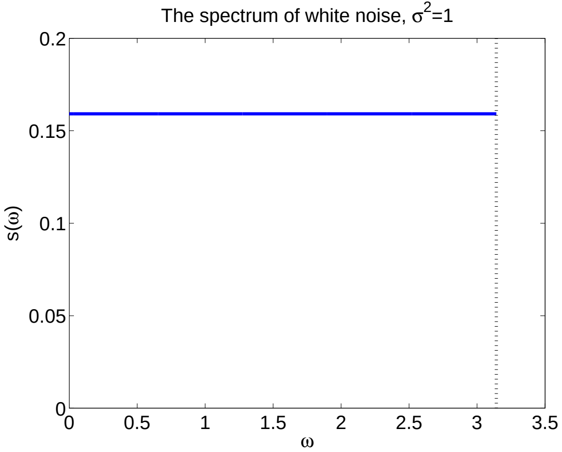
- Related: moment-generating function.

The spectrum of white noise

- Suppose, $x_t = \epsilon_t$, where ϵ_t is a martingale-difference sequence with constant variance σ^2.
- $\gamma_j = 0$, except $\gamma_0 = \sigma^2$. Thus

$$s_x(\omega) = \frac{\sigma^2}{2\pi}$$


The spectrum of white noise, σ²=1
s(ω)
ω
0.2
0.15
0.1
0.05
0
−4
−2
0
2
4


The spectrum of white noise, σ²=1
s(ω)
ω
0
0.05
0.1
0.15
0.2
0
0.5
1
1.5
2
2.5
3
3.5

Lag Operator Calculus and the Spectrum

- Consider

$$y_t = h(L)x_t = \sum_{j=-\infty}^{\infty} h_j x_{t-j}$$

- Then,

$$s_y(\omega) = h(e^{-i\omega})h(e^{i\omega})s_x(\omega) = \mid h(e^{-i\omega}) \mid^2 s_x(\omega)$$

- Heuristic proof:

$$\begin{aligned} s_y(\omega) &\cong E[\tilde{y}(\omega)\overline{\tilde{y}(\omega)}] \\ &\cong h(e^{-i\omega})E[\tilde{x}(\omega)\overline{\tilde{x}(\omega)}]h(e^{i\omega}) \\ &= h(e^{-i\omega})s_x(\omega)h(e^{i\omega}) \end{aligned}$$

- Multivariate (with ' denoting complex-conjugate transpose):

$$s_y(\omega) = h(e^{-i\omega})s_x(\omega)\left(h(e^{-i\omega})\right)'$$

AR(1)

$$\begin{aligned}
\epsilon_t &= (1 - \rho L) y_t \\
\frac{\sigma^2}{2\pi} &= (1 - \rho e^{i\omega})(1 - \rho e^{-i\omega}) s_y(\omega) \\
&= (1 - 2\rho\cos(\omega) + \rho^2) s_y(\omega)
\end{aligned}$$

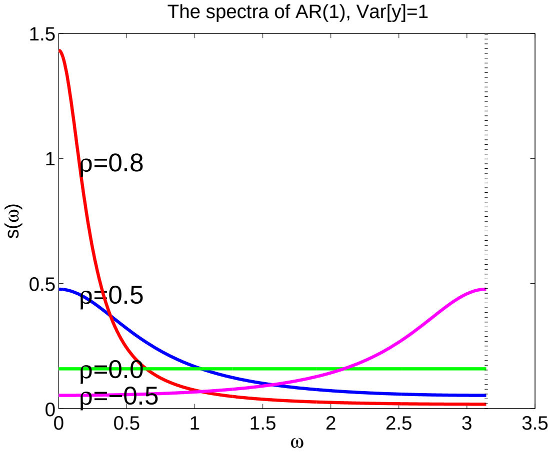
Therefore

Remark

The spectrum for an AR(1), $(1 - \rho L) y_t = \epsilon_t$, $E[\epsilon_t^2] = \sigma^2$ is given by

$$s_y(\omega) = \frac{1}{1 - 2\rho\cos(\omega) + \rho^2} \frac{\sigma^2}{2\pi}$$

The spectra of an AR(1).



AR(m)

AR(m), all roots stable:

$$\begin{aligned}(1-\rho(L))y_t &= \epsilon_t \\ (1-\lambda_1 L)\ldots(1-\lambda_m L)y_t &= \epsilon_t\end{aligned}$$

Suppose, all roots are real-valued. Then,

$$s_y(\omega) = \frac{\sigma^2}{2\pi}\prod_{j=1}^{m}\frac{1}{1-2\lambda_j\cos(\omega)+\lambda_j^2}$$

AR(m)

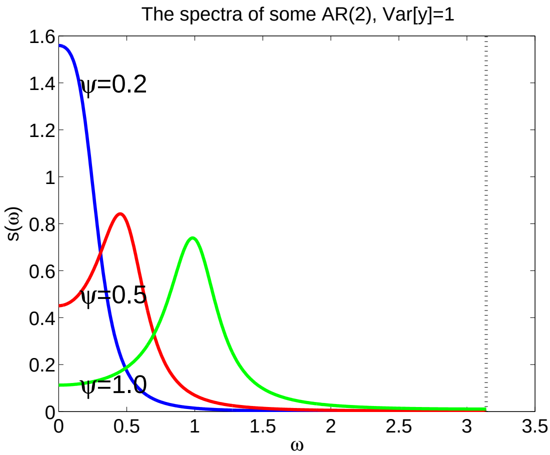
Generally (and observing that complex roots come in complex-conjugate pairs):

$$\begin{aligned} s_y(\omega) &= \frac{\sigma^2}{2\pi} \prod_{j=1}^{m} \left| \frac{1}{(1-\lambda_j e^{-i\omega})(1-\lambda_j e^{i\omega})} \right| \\ &= \frac{\sigma^2}{2\pi} \prod_{j=1}^{m} \frac{1}{(1-\lambda_j e^{-i\omega})(1-\overline{\lambda_j} e^{i\omega})} \end{aligned}$$

Example: AR(2)

$$(1 - 2\rho\cos(\psi)L + \rho^2 L^2)y_t \quad = \quad (1 - \rho e^{i\psi}L)(1 - \rho e^{-i\psi}L)y_t = \epsilon_t$$

AR(2), $(1 - 2\rho\cos(\psi)L + \rho^2 L^2)y_t = \epsilon_t$, $\rho = 0.8$



MA(n)

$$y_t = \theta(L)\epsilon_t$$

The spectrum is given by

$$s_y(\omega) = \frac{\sigma^2}{2\pi}\theta(e^{i\omega})\theta(e^{-i\omega})$$

MA(1)

$$
\begin{aligned}
y_t &= (\theta_0 + \theta_1 L)\epsilon_t \\
s_y(\omega) &= (\theta_0 + \theta_1 e^{i\omega})(\theta_0 + \theta_1 e^{-i\omega})\frac{\sigma^2}{2\pi} \\
&= (\theta_0^2 + 2\theta_0\theta_1 \cos(\omega) + \theta_1^2)\frac{\sigma^2}{2\pi}
\end{aligned}
$$

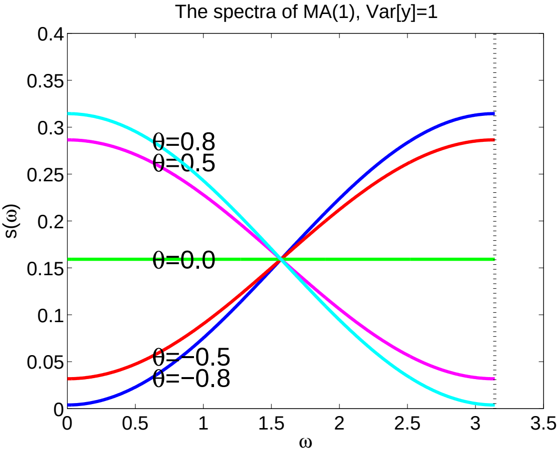
Therefore

Remark

The spectrum for an MA(1), $y_t = (\theta_0 + \theta_1 L)\epsilon_t$, $E[\epsilon_t^2] = \sigma^2$ *is given by*

$$
s_y(\omega) = (\theta_0^2 + 2\theta_0\theta_1 \cos(\omega) + \theta_1^2)\frac{\sigma^2}{2\pi}
$$

The spectra of an MA(1), $y_t = (1 + \theta L)\epsilon_t$.



Blaschke factor, root flipping

- Blaschke factor: for $0 \neq \lambda \in \textit{complex}$, define

$$B_\lambda(z) = \frac{z - \lambda}{1 - \lambda z} = -\lambda \frac{1 - \lambda^{-1} z}{1 - \lambda z}$$

- Note:

$$B_\lambda(e^{-i\omega}) = \frac{e^{-i\omega} - \lambda}{1 - \lambda e^{-i\omega}} = \frac{1 - \lambda e^{i\omega}}{e^{i\omega} - \lambda} = \left(B_\lambda(e^{i\omega})\right)^{-1}$$

- Root flipping: Let $y_t = \theta(L)\epsilon_t$. Similar to characteristic polynomial for AR's, let $p(\lambda) = \lambda^n \theta(\lambda^{-1})$. Suppose λ is a root of $p(\lambda)$. Then, $\theta(L) = (1 - \lambda L) * \ldots$. Let $\check{y}_t = \check{\theta}(L)\epsilon_t$, where

$$\check{\theta}(L) = B_\lambda(L)\theta(L)$$

  Then

$$s_{\check{y}}(\omega) = \frac{\sigma^2}{2\pi} B_\lambda(e^{-i\omega}) B_\lambda(e^{i\omega}) \theta(e^{i\omega}) \theta(e^{-i\omega}) = s_y(\omega)$$

Remarks

- Same spectral density,

$$s_{\check{y}}(\omega) = s_y(\omega)$$

- ... therefore, same autocorrelations,

$$\check{\gamma}_j = \gamma_j$$

- ... therefore, same Wold decomposition as the original representation.

MA(n), Fundamental Representation

- Consider

$$
\begin{aligned}
y_t &= \theta(L)\epsilon_t \\
&= \theta_0 \left(1 - \lambda_1 L\right) \ldots \left(1 - \lambda_n L\right) \epsilon_t
\end{aligned}
$$

- Suppose that

$$
\mid \lambda_1 \mid > \ldots \mid \lambda_r \mid > 1 > \mid \lambda_{r+1} \mid > \ldots \mid \lambda_n \mid
$$

- Flip explosive roots.

MA(n), Fundamental Representation

Definition

Define the fundamental representation

$$y_t = C(L)u_t \tag{2}$$

where

$$\begin{aligned} C(L) &= \frac{1}{(-\lambda_1)\dots(-\lambda_r)\theta_0} B_{\lambda_1}(L)\dots B_{\lambda_r}(L)\theta(L) \\ &= \left(1-\lambda_1^{-1}L\right)\dots\left(1-\lambda_r^{-1}L\right)(1-\lambda_{r+1}L)\dots(1-\lambda_n L) \\ \text{Var}\,(u_t) &= (\lambda_1\dots\lambda_r\theta_0)^2\,\text{Var}(\epsilon_t) \end{aligned}$$

Properties of the fundamental representation, part 2

- The fundamental representation is invertible. Thus, find u_t per

$$\begin{aligned} u_t &= \left(1 - \lambda_1^{-1} L\right)^{-1} \cdots \left(1 - \lambda_r^{-1} L\right)^{-1} \cdots \\ &\qquad (1 - \lambda_{r+1} L)^{-1} \cdots (1 - \lambda_n L)^{-1} y_t \\ &= y_t - P(y_t \mid y_{t-1}, y_{t-2}, \ldots) \end{aligned}$$

- Therefore, (2) is the Wold decomposition of y_t.
- Relationship between ϵ_t and u_t? Treat ϵ_t as a hidden state and apply the Kalman Filter.

Recall Wold decomposition for MA(1), example 2

- Example 2:

$$\begin{aligned} y_t &= \epsilon_t + 2\epsilon_{t-1}, E[\epsilon_t^2] = 1 \\ &= (1 - \lambda L)\epsilon_t, \; \lambda = -2 \end{aligned}$$

- $\mid \lambda \mid > 1$. Flip it.
- Fundamental representation / Wold decomposition:

$$\begin{aligned} y_t &= (1 - \lambda^{-1} L) u_t \\ &= u_t + 0.5 u_{t-1} \\ E[u_t^2] &= \lambda^2 E[\epsilon_t^2] = 4 \end{aligned}$$

Outline

Unit roots

- We always assumed that $\mid \lambda \mid > 1$ or $\mid \lambda \mid < 1$.
- What about $\mid \lambda \mid = 1$?

Integration

Definition

- An AR(m) process, for which all roots $\lambda_i, i = 1, \ldots, m$ of the characteristic polynomial $p(\lambda)$ are smaller than 1 in absolute value or exactly equal to 1, with at least one root exactly equal to 1, is called integrated or is said to have a unit root.
- Suppose all roots $\lambda_i, i = 1, \ldots, m$ of the characteristic polynomial $p(\lambda)$ are smaller than 1 in absolute value or exactly equal to 1. Let r be the number of roots exactly equal to one. The process is then said to be integrated of order r or I(r).

Stationary and Integration

Remark

For an AR(m) process, the following two statements are equivalent:

- *The process is covariance stationary*
- *The process is I(0)*

Differencing

Definition

Define the difference operator Δ per

$$\Delta y_t = y_t - y_{t-1} = (1 - L)y_t$$

Multiple differencing

- Multiple differencing is indicated by powers,

$$\Delta^r y_t = \Delta(\Delta(\ldots\Delta y_t)\ldots)$$

- Example:

$$\begin{aligned}\Delta^2 y_t &= \Delta(\Delta y_t) \\ &= \Delta(y_t - y_{t-1}) \\ &= y_t - 2y_{t-1} + y_{t-2}\end{aligned}$$

Differencing an I(r) process

Remark

Let y_t an AR(m) process be I(r). Then, $\Delta^r y_t$ is a covariance stationary AR(m-r) process.

Example: AR(4), which is I(2):

$$
\begin{aligned}
(1-\rho(L))y_t &= \epsilon_t \\
(1-2L+.75L^2+.5L^3-.25L^4)y_t &= \epsilon_t \\
(1-.5L)(1+.5L)(1-L)(1-L)y_t &= \epsilon_t \\
(1-\rho^*(L))\,x_t &= \epsilon_t \\
\text{where } x_t &= \Delta^2 y_t \text{ is stationary} \\
1-\rho^*(L) &= (1-.5L)(1+.5L)
\end{aligned}
$$

Integration

Definition

- An AR(m) process, for which all roots $\lambda_i, i = 1, \ldots, k$ of the characteristic polynomial $p(\lambda)$ are smaller than or equal to 1 in absolute value, with at least one root equal to 1 in absolute value, is called seasonally integrated.

This is used and investigated much more rarely than the case of a root exactly equal to 1.

Outline

Estimating an AR(1)

- AR(1):

$$y_t = \rho y_{t-1} + \epsilon_t, t = 1, \ldots, T$$

  where ϵ_t is a martingale-difference, $E[\epsilon_t^2] = \sigma^2$, LLN, CLT.

- OLSE:

$$\hat{\rho} = \frac{\sum_{t=1}^{T} y_t y_{t-1}}{\sum_{t=1}^{T} y_{t-1}^2} \tag{3}$$

- If $\mid \rho \mid < 1$,

$$\hat{\rho} = \rho + \frac{1}{\sqrt{T}} \frac{\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \epsilon_t y_{t-1}}{\frac{1}{T} \sum_{t=1}^{T} y_{t-1}^2}$$

  so

$$\sqrt{T}(\hat{\rho} - \rho) \xrightarrow{d} \mathcal{N}(0, 1 - \rho^2)$$

- What, if $\mid \rho \mid = 1$?

Brownian motion

Definition

A Brownian motion $W(s)$ on $s \in [0,1]$ is a continuous-time stochastic process in $\mathbb{R}$ such that

- $W(0) = 0$.
- For any dates $0 \leq s_1 < s_2 < \ldots < s_k \leq 1$, the changes $W_{s_j} - W_{s_{j-1}}, j = 2, \ldots, k$ are independent with

$$W_{s_j} - W_{s_{j-1}} \sim \mathcal{N}(0, s_j - s_{j-1})$$

- $W(s)$ is almost surely continuous.

Motivation

- Recall central limit theorem:

$$\frac{1}{\sqrt{T}\sigma}\sum_{t=1}^{T}\epsilon_t \xrightarrow{d} \mathcal{N}(0,1)$$

- For $0 \leq s \leq 1$, define

$$X_T(s) = \frac{1}{\sqrt{T}\sigma}\sum_{t=1}^{[sT]}\epsilon_t$$

- Let $0 \leq s_1 < s_2 < \ldots < s_k \leq 1$. Observe that

$$X_T(s_j) - X_T(s_{j-1}) \xrightarrow{d} \mathcal{N}(0, s_j - s_{j-1})$$

  and that these increments are independent.

Functional Central Limit Theorem

Theorem

The random functions $X_T : [0,1] \times \Omega \to \mathbb{R}$ *converge to a Brownian motion in distribution,*

$$X_T \xrightarrow{d} W, \; T \to \infty$$

(Other names: Donsker's theorem, invariance principle)

$C([0,1])$ and continuous functionals

- The space of continuous functions:

$$C([0,1]) = \{f : [0,1] \to \mathbb{R} \mid f \text{ is continuous } \}$$

- Norm:

$$\| f \|_\infty = \max_{s \in [0,1]} | f(s) |$$

- $(C([0,1]), \| \cdot \|_\infty)$ is a Banach space, i.e. a complete normed vector space.
- A functional is a mapping $g : C([0,1]) \to \mathbb{R}$.
- Example:

$$g(f) = \int_{[0,1]} f(s)^2 ds$$

- $g : C([0,1]) \to \mathbb{R}$ is continuous functional, if, for every $\epsilon > 0$ and every $f_0 \in C([0,1])$, there is a $\delta > 0$, so that $\| f_1 - f_0 \|_\infty < \delta$ implies $| g(f_1) - g(f_0) | < \epsilon$.

Continuous Mapping Theorem

Theorem

Suppose that the random functions $X_T : [0,1] \times \Omega \rightarrow \mathbb{R}$ *converge to the continuous random function* $X : [0,1] \times \Omega \rightarrow \mathbb{R}$ *in distribution. Suppose* $g : C([0,1]) \rightarrow \mathbb{R}$ *is a continuous functional on the space of continuous functions on* $[0,1]$*. Then,*

$$g(X_T) \xrightarrow{d} g(X), \; T \rightarrow \infty$$

Estimating an AR(1), again

- $y_t = \rho y_{t-1} + \epsilon_t$.
- OLSE:

$$\hat{\rho}_T = \frac{\sum_{t=1}^{T} y_t y_{t-1}}{\sum_{t=1}^{T} y_{t-1}^2} \tag{4}$$

- Recall

$$X_T(s) = \frac{1}{\sqrt{T}\sigma} \sum_{t=1}^{[sT]} \epsilon_t$$

- Assume $y_0 = 0, \rho = 1$. Let $s = t/T$ and $\Delta s = 1/T$.

$$y_t = \sum_{\tau=1}^{[sT]} \epsilon_\tau = \sigma\sqrt{T} X_T(s)$$

The pieces of $\hat{\rho}$, part 1

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T} y_{t-1}\epsilon_t &= \frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{t-1}\epsilon_j\epsilon_t \\
&= \frac{1}{2T}\left(\sum_{t=1}^{T}\sum_{j=1}^{T}\epsilon_j\epsilon_t\right) - \frac{1}{2T}\sum_{t=1}^{T}\epsilon_t^2 \\
&= \frac{1}{2T}y_T^2 - \frac{1}{2T}\sum_{t=1}^{T}\epsilon_t^2 \\
&= \frac{\sigma^2}{2}X_T(1)^2 - \frac{1}{2T}\sum_{t=1}^{T}\epsilon_t^2 \\
&\xrightarrow{d} \frac{\sigma^2}{2}\left(W(1)^2 - 1\right)
\end{aligned}$$

The pieces of $\hat{\rho}$, part 2

$$
\begin{aligned}
\frac{1}{T^2}\sum_{t=1}^{T} y_{t-1}^2 &= \sigma^2 \sum_{j=0}^{T-1} X_T(j\Delta s)^2 \Delta s \\
&\approx \sigma^2 \int_{[0,1]} X_T(s)^2 ds \\
&\xrightarrow{d} \sigma^2 \int W(s)^2 ds
\end{aligned}
$$

The limit distribution for the OLSE

Continuous mapping theorem: take the ratio and then the limit.

Proposition

For $y_t = \rho y_{t-1} + \epsilon_t$ and under the null hypothesis $\rho = 1$,

$$T(\hat{\rho}_T - 1) \xrightarrow{d} \frac{1}{2} \frac{W(1)^2 - 1}{\int W(s)^2 ds}$$

- The right-hand side is some random variable, defined as a continuous functional of a Brownian motion. The distribution is skewed to the left, i.e. the limit distribution is not normal. Tables etc are available.
- Superconsistency: the rate of convergence is higher than $\sqrt{T}$.

Remarks

- Huge literature, developed in 80s and early 90s. Now, part of standard econometrics packages.
- Unit root tests: look them up, if needed.
  - augmented Dickey-Fuller
  - Phillips-Perron
- Multivariate context: Johansen procedure.
- It matters, whether constants and/or time trends are included, when doing these tests.
- Later developments: allow for breaks in the series.
- Often: pretest for unit roots and then proceed, as if it is known that there is a unit root (or not). Conditionality of results! Not a full classical procedure. See the conundrum of the experimenter.

Example: forecasting with an AR(1)

- $y_t = \rho y_{t-1} + \epsilon_t$
- Test $H_0 : \rho = 1$.
- If not rejected, produce forecasts, imposing $\rho = 1$.
- Appropriate confidence bounds for forecasts?

Outline

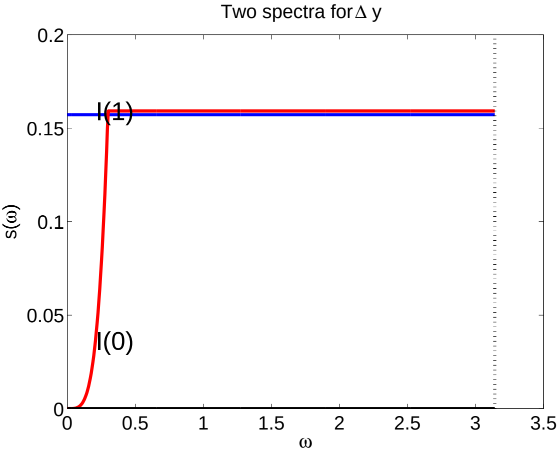
Comparing spectra of I(0) vs I(1)

- y_t is $I(1)$ iff $x_t = \Delta y_t$ has nonzero, finite spectrum at $\omega = 0$.
- If the spectra of two series are very close in the L_1-distance,

$$\int \mid s_x(\omega) - s_{\check{x}}(\omega) \mid d\omega$$

  then so are their autocovariances.
- For any series with a nonzero spectrum at frequency zero, there is another one with a spectrum close to it, but vanishing at frequency zero, and vice versa.

Comparing spectra of I(0) vs I(1)



Power = size for unit root tests

- Consider a test of H_0:"y is I(0)" against a nonempty alternative of I(1) processes. Suppose the test has power above $1-\beta$, i.e. if truth is an element $\tilde{y}$ of the alternative, the null hypothesis is rejected with probability above $1-\beta$. Find y, which is I(0) and close to $\tilde{y}$ in the L_1-distance. If it is close enough, then the test will reject the null hypothesis with probability above $1-\beta$, if y is true.
- One can switch the role of I(0) and I(1) in this argument.
- Therefore, to obtain power of, say, 95%, when size is 5%, the null hypothesis needs to be restricted further.

Remark

Tests for or against unit roots have "power = size", unless the null hypothesis is restricted further. Therefore, the power of unit root tests comes from restrictions on the transitory dynamics.

Estimating the spectrum at frequency zero

- Infinite data: (multivariate notation)

$$2\pi S(0) = \sum_{j=-\infty}^{\infty} \Gamma_j$$

- Finite data, $t = 1, \ldots, T, j = -(T-1), \ldots, 0, \ldots, T-1$:

$$\hat{\Gamma}_j = \frac{1}{T} \sum_{t=\max\{j+1,1\}}^{\min\{T,T+j\}} y_t y'_{t-j}$$

- Newey-West, Bartlett

$$2\pi \hat{S}_T(0) = \sum_{j=-q}^{q} \left(1 - \frac{|\, j \,|}{q+1}\right) \hat{\Gamma}_j$$

  Pos. semidef.. $\hat{S}(0) \to s(0)$ if $T, q \to \infty$, $q/T^{1/4} \to 0$.
- Kernel-density estimation. Other kernels, other estimators.
- Picture: power per restricting shape/slope near zero.

Application: HAC

- HAC: heteroskedasticity and autocorrelation-consistent.
- Linear regression:

$$y_t = x_t'\beta + \epsilon_t,\ E[\epsilon_t \mid x_t] = 0$$

- OLSE b_T:

$$\sqrt{T}\,(b_T - \beta) = Q_T^{-1} g_T, \text{where } Q_T = \frac{1}{T}\sum_{t=1}^{T} x_t x_t',\ g_T = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} x_t \epsilon_t$$

- Let $z_t = x_t \epsilon_t$ and let $S = 2\pi S_z(0)$. Under mild conditions,

$$\sqrt{T}\,(b_T - \beta) \xrightarrow{d} \mathcal{N}(0, Q^{-1} S Q^{-1})$$

- HAC: Estimate S per Newey-West S_T. Use $Q_T^{-1} S_T Q_T^{-1}$

Outline

The Bayesian perspective

- Christopher A. Sims and Harald Uhlig, "Understanding Unit Rooters: A Helicopter Tour.", Econometrica, vol. 59, no. 6, Nov. 1991, 1591-1599.
- The nonstationarity is in the data, not in the parameters.
- At the time of inference, the data is given.
- The likelihood-function in the parameters is Normal-Wishart in shape.
- Standard F- and t-statistics summarize the shape of the likelihood function.

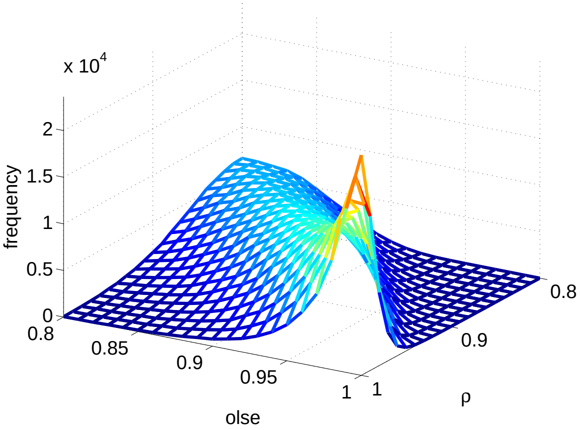
The AR(1) case

- Likelihood function, conditional on y_0:

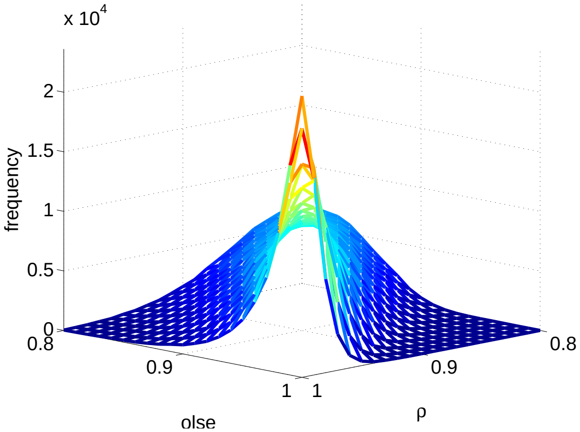
$$
\begin{aligned}
\log L &= -\frac{T}{2}\log(2\pi\sigma^2) - \sum_{t=1}^{T}\frac{(y_t - \rho y_{t-1})^2}{2\sigma^2} \\
&= -\frac{T}{2}\log(2\pi\sigma^2) - \sum_{t=1}^{T}\frac{(y_t - \hat{\rho} y_{t-1})^2}{2\sigma^2} - (\hat{\rho} - \rho)^2 \sum_{t=1}^{T}\frac{y_{t-1}^2}{2\sigma^2}
\end{aligned}
$$

- $\hat{\rho}$ is a function of the data $y_0, \ldots, y_T$. ρ is a parameter.
- Thus: "Strange" shape in $\hat{\rho}$, given ρ. Classical perspective.
- Normal shape (quadratic) in ρ, given data: Bayesian perspective.
- Log-Posterior = Log Prior + Log Likelihood + Integrating Constant.
- Assume flat prior. Figures: reconstructed, according to Sims-Uhlig (1991).

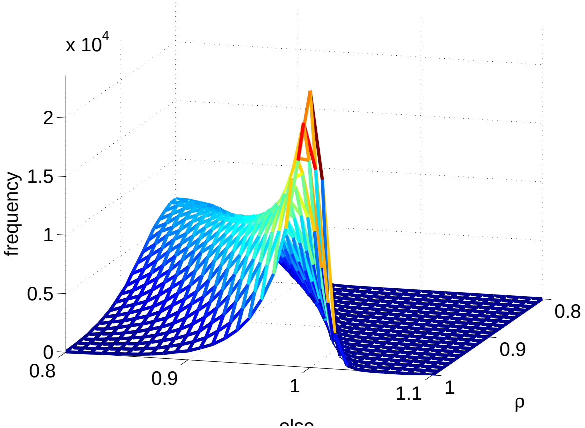
The joint distribution for ρ and $\hat{\rho}$



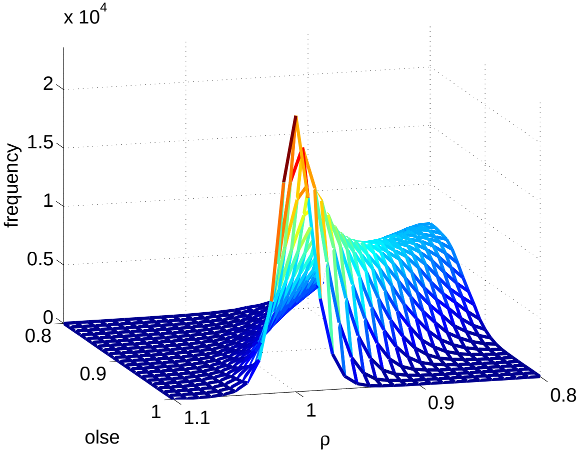
The joint distribution for ρ and $\hat{\rho}$



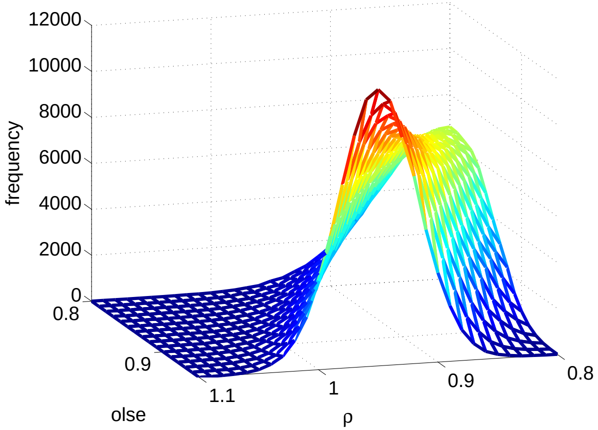
The joint distribution for ρ and $\hat{\rho}$



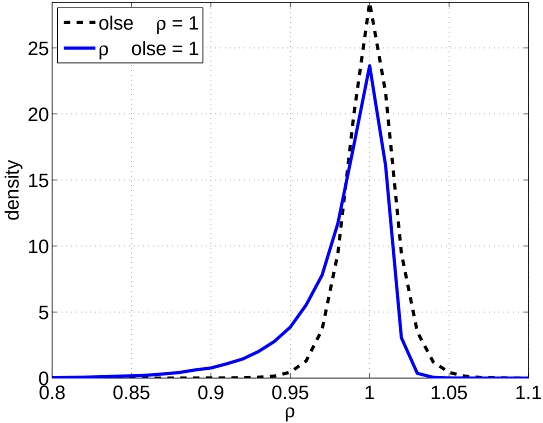
The joint distribution for ρ and $\hat{\rho}$



The joint distribution for ρ and $\hat{\rho}$



Comparing marginals: $\rho \mid \hat{\rho} = 1$ and $\hat{\rho} \mid \rho = 1$



Forecasting an AR(1)

- Forecasting with Bayesian methods.
- Find posterior distribution for y_{t+k}, given $y_0, \ldots, y_t$.
- AR(1):

$$y_{t+k} \mid y_0, \ldots, y_t \sim \rho^k y_t + \sum_{j=0}^{k-1} \rho^j \epsilon_{t+k-j}, \; \epsilon_s \sim \mathcal{N}(0, \sigma^2)$$

- Draw (ρ, σ) from the posterior. Draw $\epsilon_{t+1}, \ldots, \epsilon_{t+k}$. Combine to generate a draw for y_{t+k}.
- Note: no special distinction for $\rho = 1$, unless imposed per prior.

Priors and such

- Peter Phillips critique: do not use the flat prior. Use Jeffreys prior.
- Debate in 1993, 1994: various priors.
- Role of information in initial observation y_0.
- Route 1: continue to use the convenient prior: Normal-Wishart.
- Route 2: treat nonstationarity issues or issues of cointegration in more sophisticated ways (Jeffreys prior, dummy observations,...).
- Uhlig, Harald, “What Macroeconomists Should Know About Unit Roots: A Bayesian Perspective,” Econometric Theory, Vol. 10, Nos. 3/4, 1994, pp. 645-671.