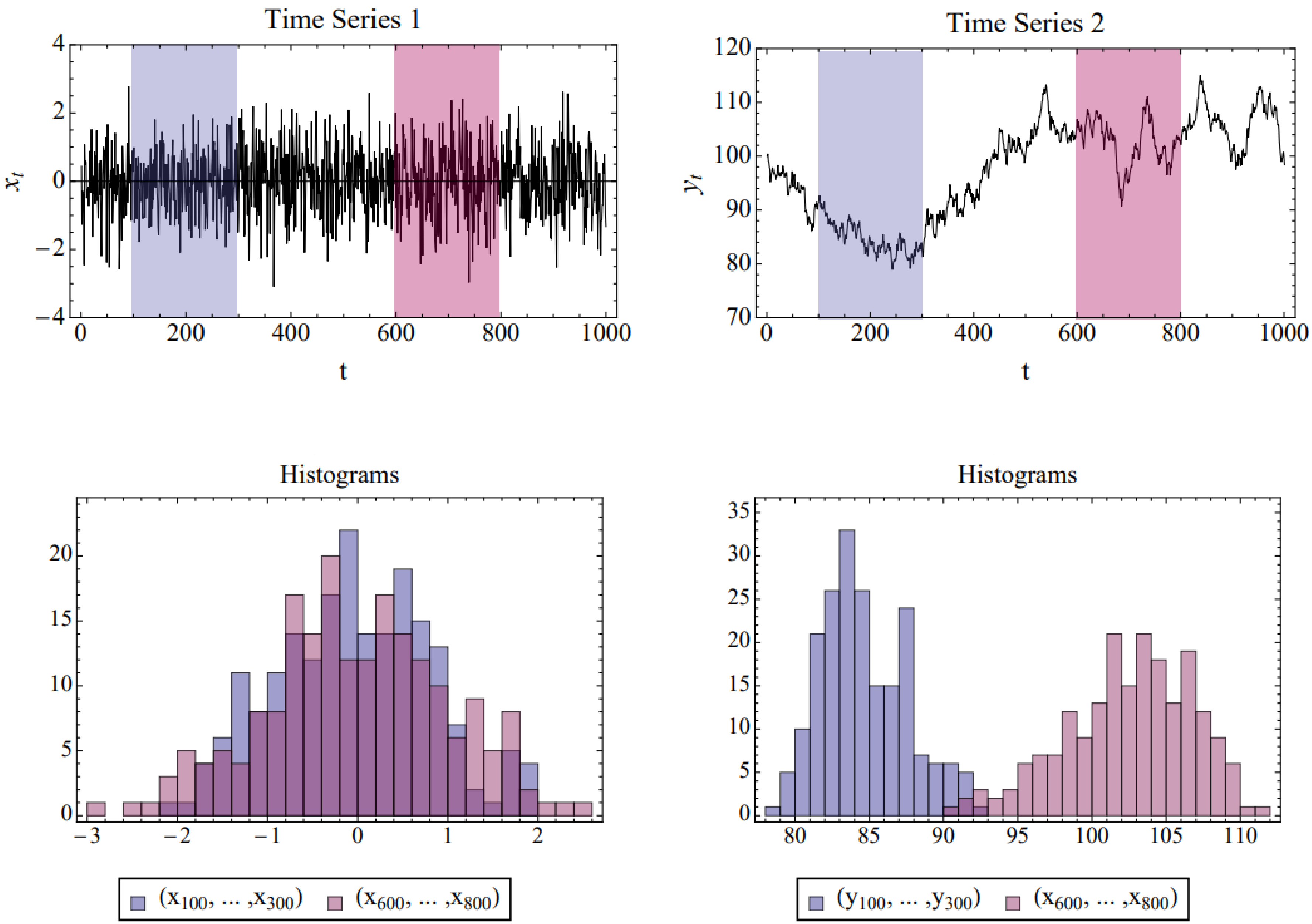
Lecture 10: More on random processes and serial correlation

Jan Zouhar

Introductory Econometrics

Classification of random processes

- **stationary** vs. **non-stationary** processes
  - stationary = distribution does not change over time
  - more precisely, the joint distribution of $(x_{t_1}, x_{t_2}, \ldots, x_{t_m})$ is the same as that of $(x_{t_1+h}, x_{t_2+h}, \ldots, x_{t_m+h})$
  - if distribution is stable around a (linear) time trend, we have a **trend-stationary** process
- **weakly** vs. **strongly dependent** processes:
  - in a weakly dependent process, the dependence between x_t and x_{t+h} vanishes if h grows without bound
  - for instance, weak dependence implies that corr(x_t, x_{t+h}) tends to zero if $h \to \infty$, and does so rapidly enough
  - we need weak dependence for the law of large number and the central limit theorem to work
  - this means that with strongly dependent time series, all our theory collapses (std. errors, hypothesis tests, p-values)
  - moreover, *spurious regression* will likely occur


Time Series 1
x_t
t
Time Series 2
y_t
t
Histograms
$(x_{100}, \ldots, x_{300})$
$(x_{600}, \ldots, x_{800})$
Histograms
$(y_{100}, \ldots, y_{300})$
$(x_{600}, \ldots, x_{800})$

Příklady náhodných procesů s diskrétním časem

Šum

- Řada stejně rozdělených nezávislých (iid) náhodných veličin, např. z $N(0, \sigma^2)$

Náhodná procházka

- $y_t = y_{t-1} + šum$

Autoregresní proces prvního řádu – AR(1)

- $y_t = c + \rho y_{t-1} + šum$, kde $c \in \mathbf{R}$, $\rho \in (-1, 1)$. Platí

$$E(y_t) = \frac{c}{1-\rho}, \qquad \operatorname{cov}(y_s, y_t) = \frac{\sigma^2_{šum}}{1-\rho^2}\rho^{|t-s|}$$

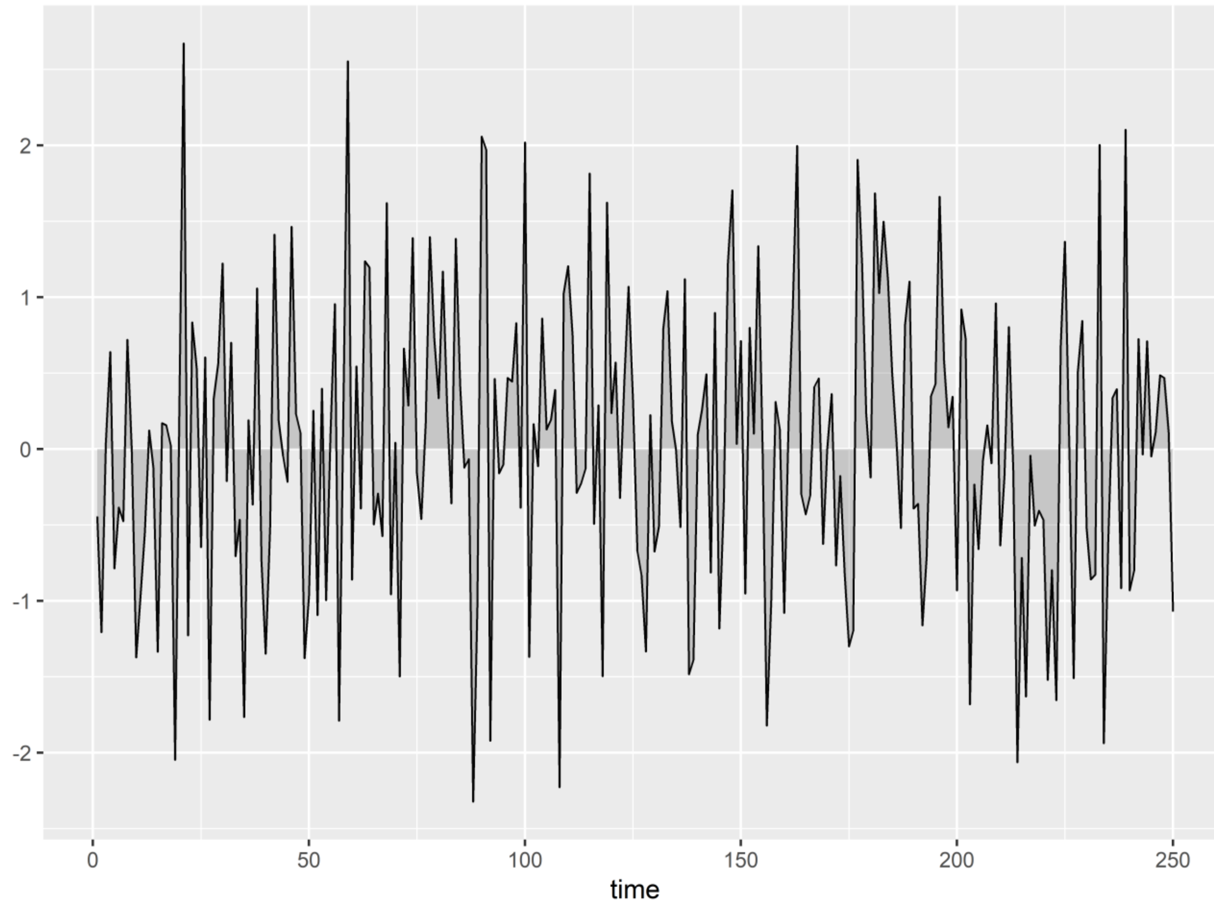
Proces klouzavých průměrů prvního řádu – MA(1)

- $y_t = \mu + \varepsilon_t + \theta\varepsilon_{t-1}$, kde ε_t představuje hodnoty šumu s rozptylem σ^2. Platí:

$$E(y_t) = \mu, \quad \operatorname{var}(y_t) = (1+\theta^2)\sigma^2, \quad \operatorname{cov}(y_t, y_s) = \begin{cases} \theta\sigma^2, & \text{pokud } s = t-1, \\ 0 & \text{jinak.} \end{cases}$$

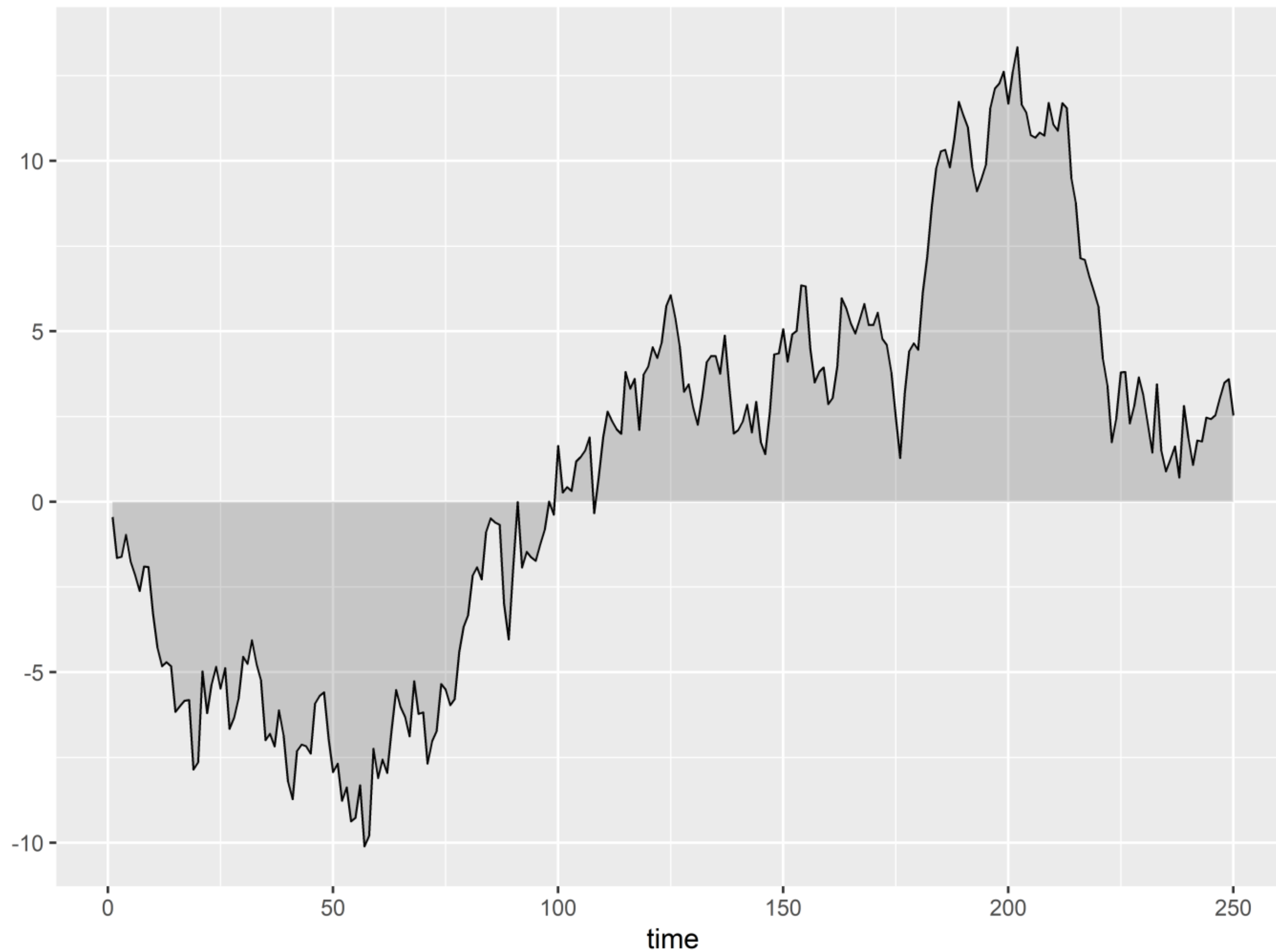
Noise

- serial independence, same distribution
- stationary, weakly dependent



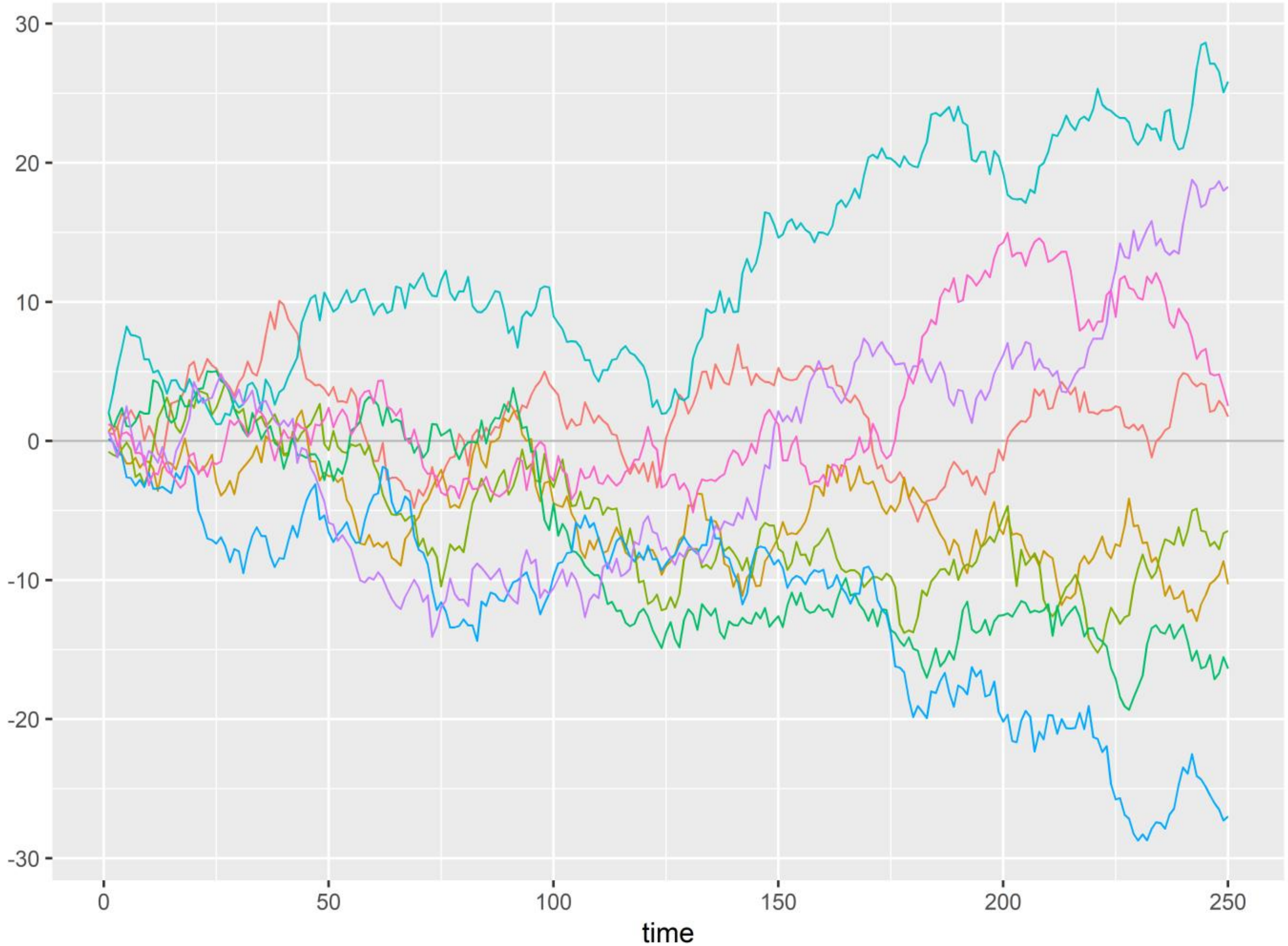
Random walk

- aggregated noise: $y_t = \rho y_{t-1}$ + noise
- non-stationary, strongly dependent (*unit-root process*)



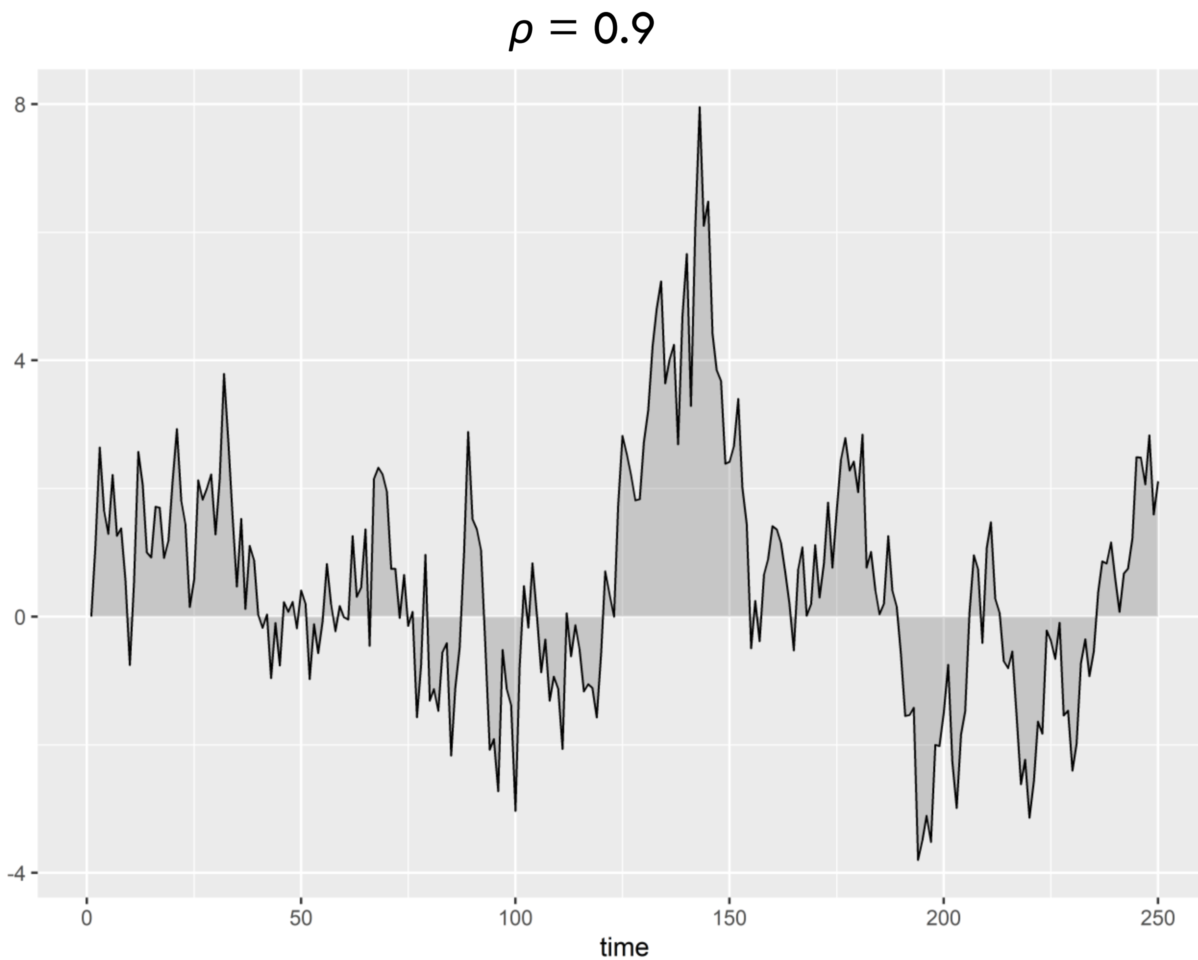
Random walk

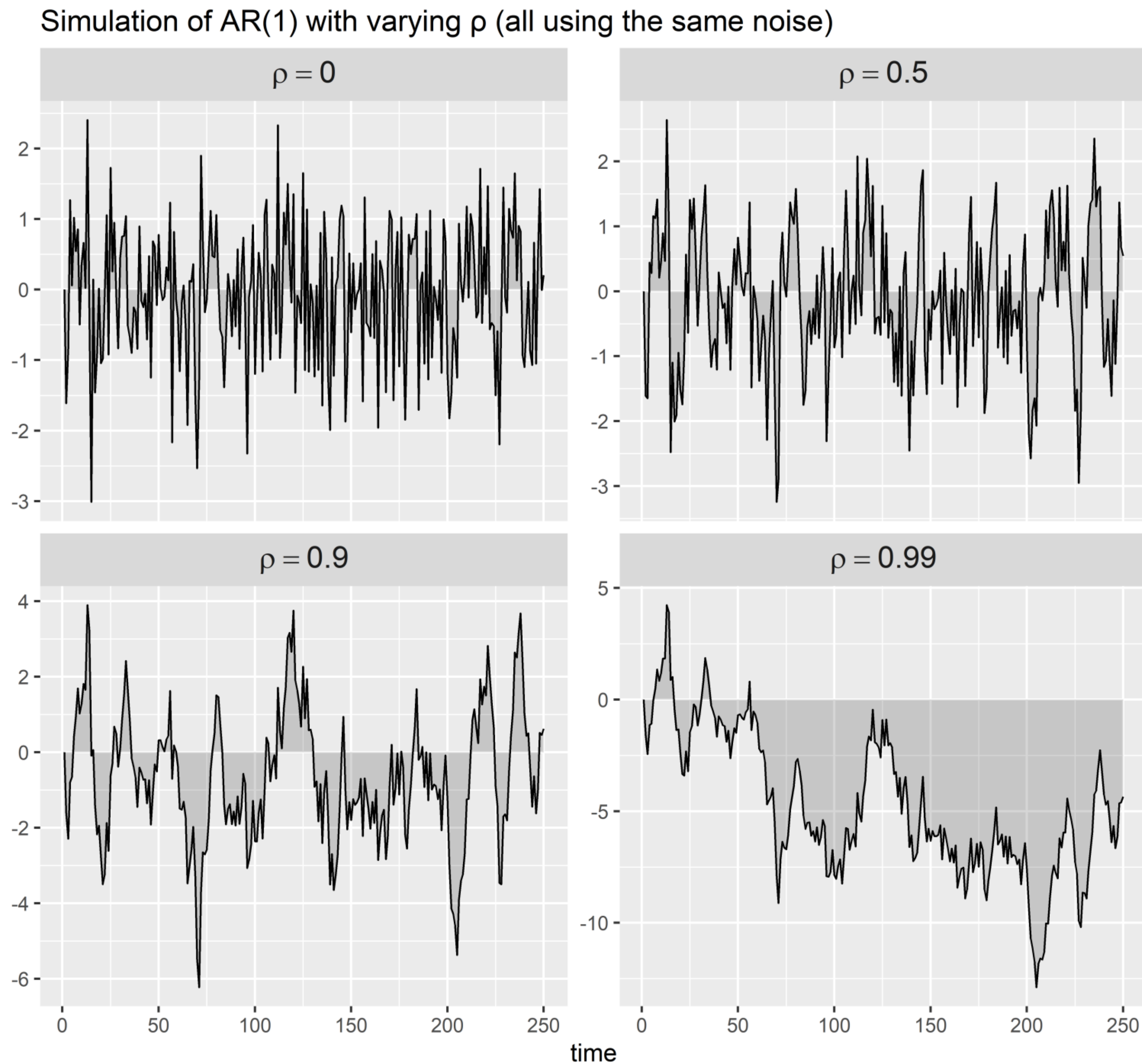
- aggregated noise: $y_t = \rho y_{t-1}$ + noise
- non-stationary, strongly dependent (*unit-root process*)



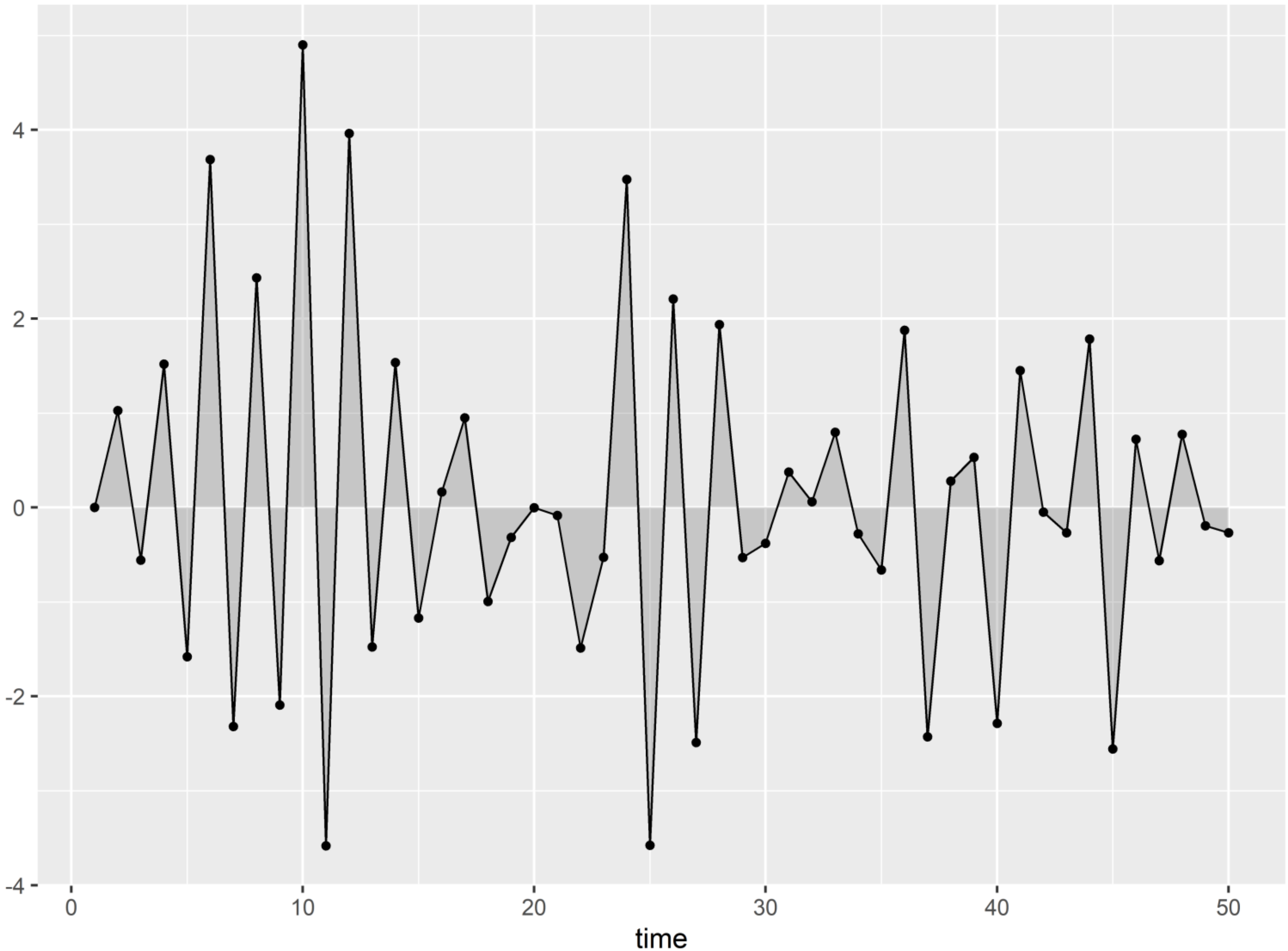
Autoregressive process of order one – AR(1)

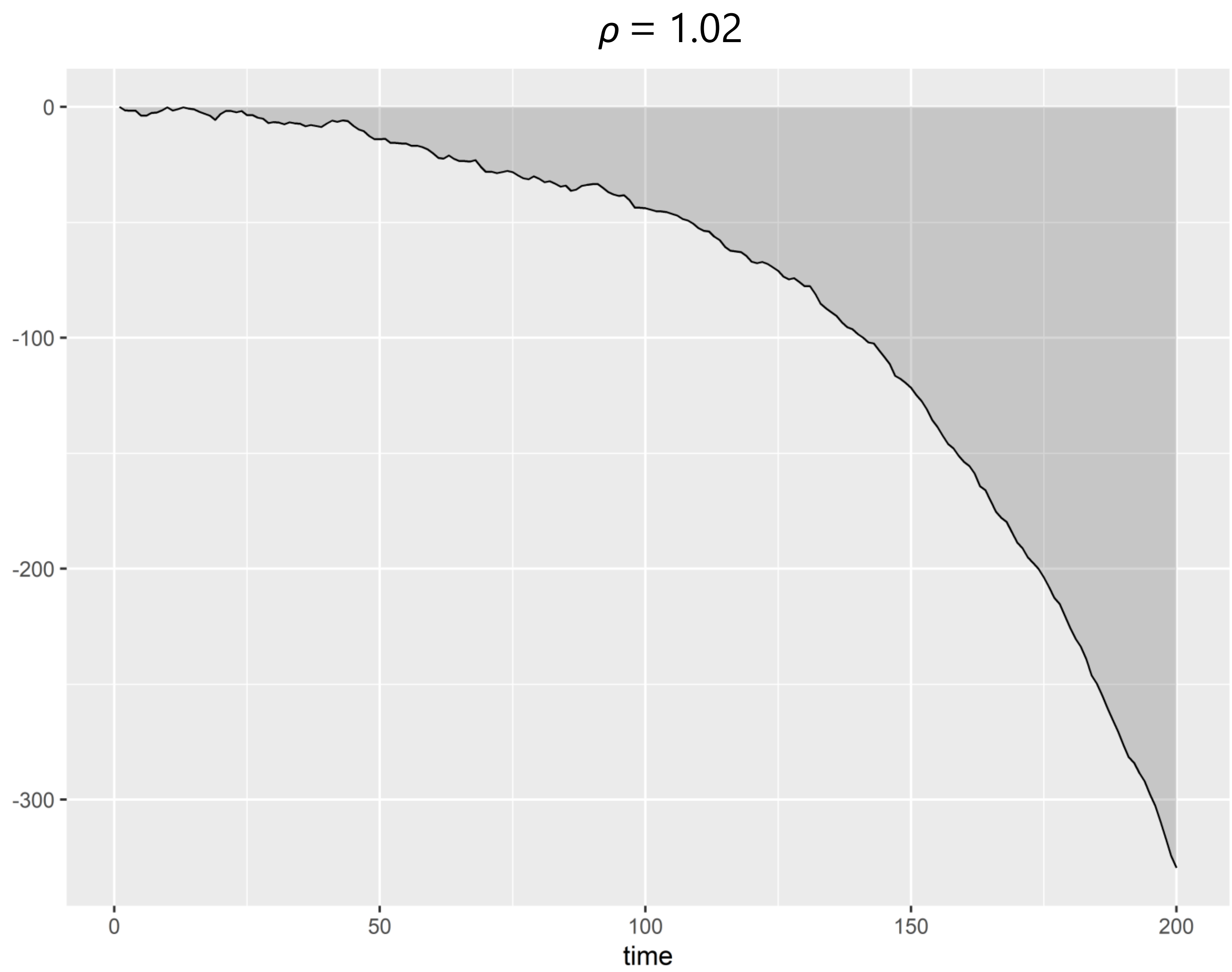
- definition: $y_t = \rho y_{t-1}$ + noise
- stationary, weakly dependent only if **stable:** $|\rho| < 1$

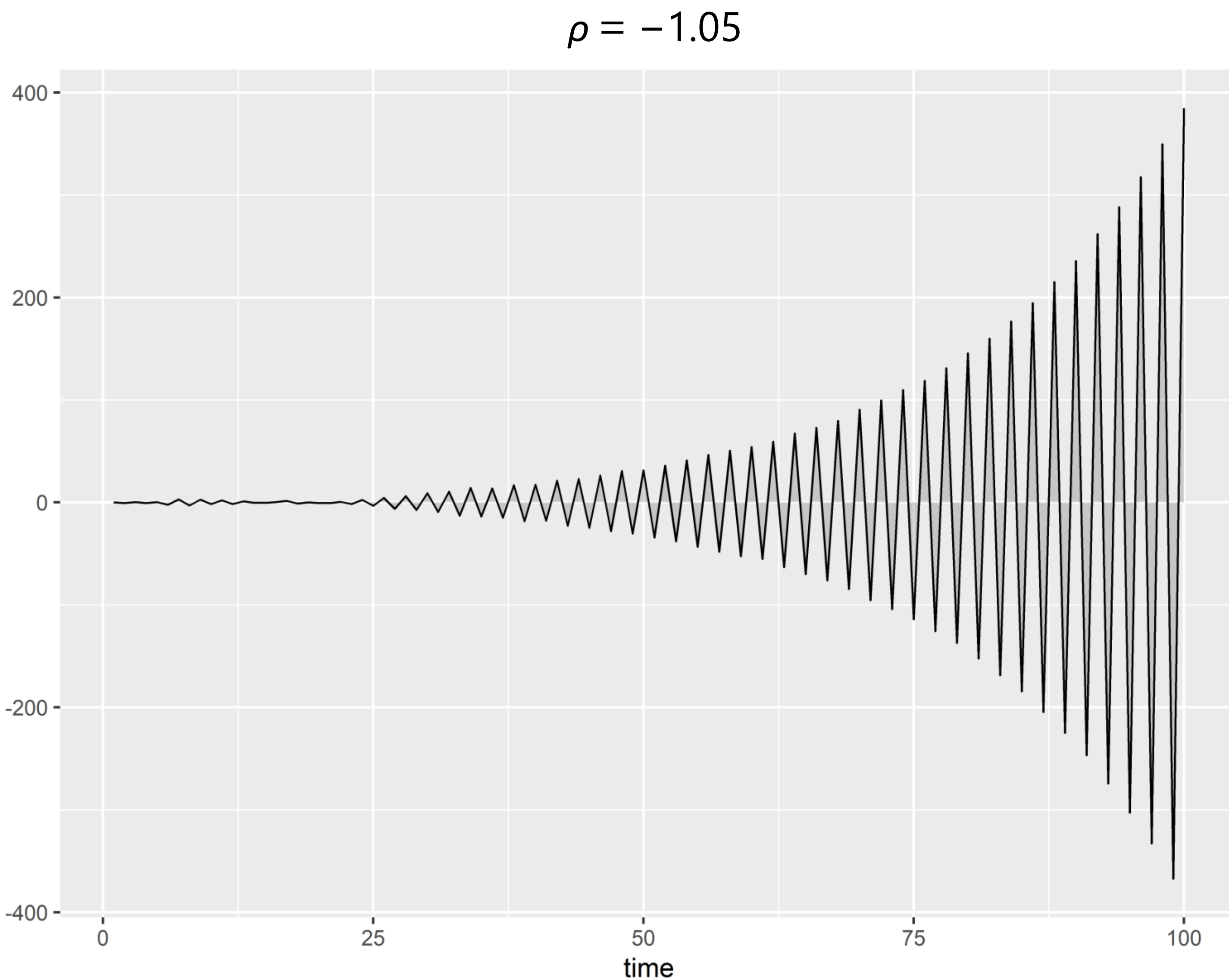



Simulation of AR(1) with varying ρ (all using the same noise)
ρ = 0
ρ = 0.5
ρ = 0.9
ρ = 0.99
time

AR(1) process, $\rho = -0.9$



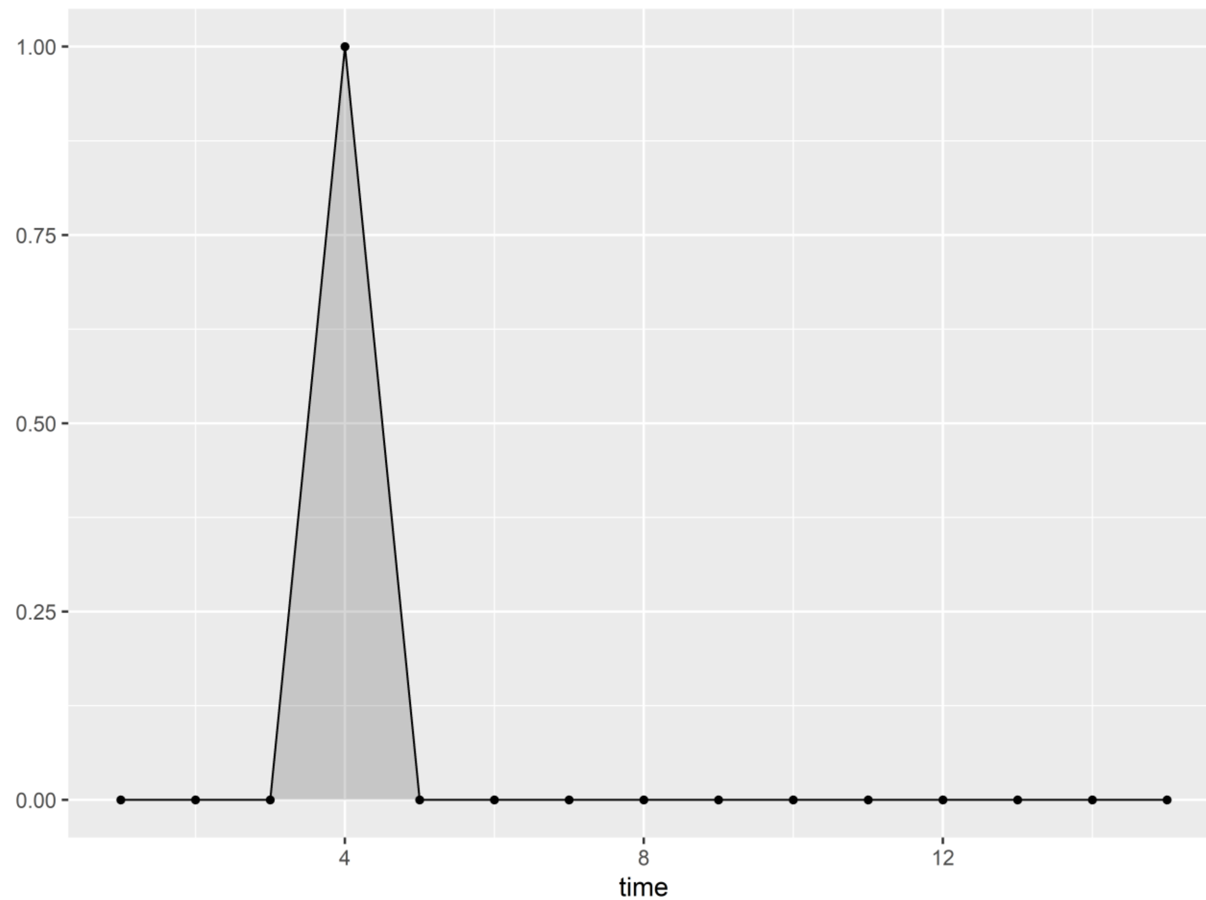

ρ = 1.02
0
-100
-200
-300
0
50
100
150
200
time



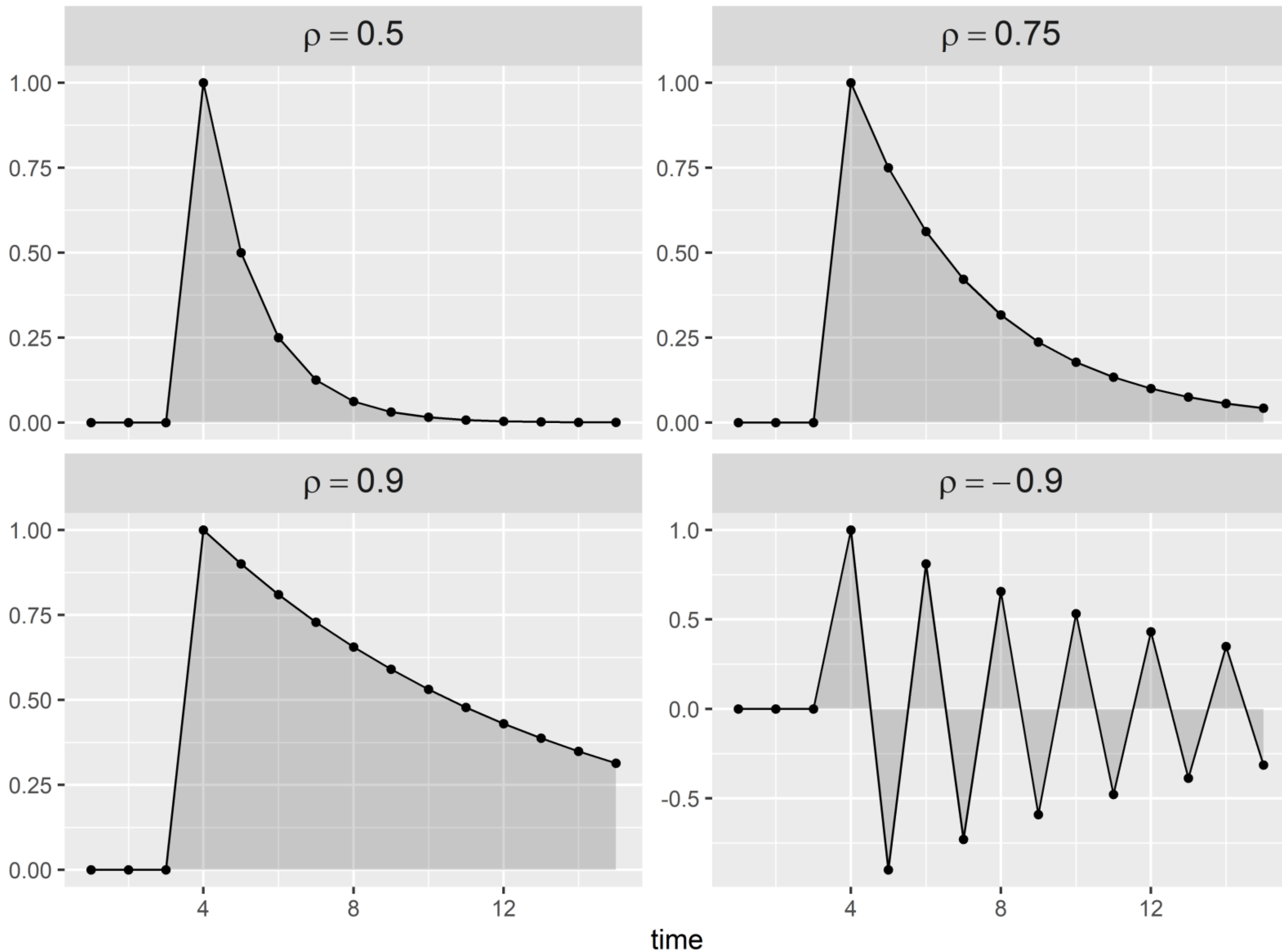
$\rho = -1.05$
400
200
0
-200
-400
0
25
50
75
100
time

Dirac impulse

- non-stochastic time series
- used instead of noise as input to e.g. the AR(1) formula to study its properties



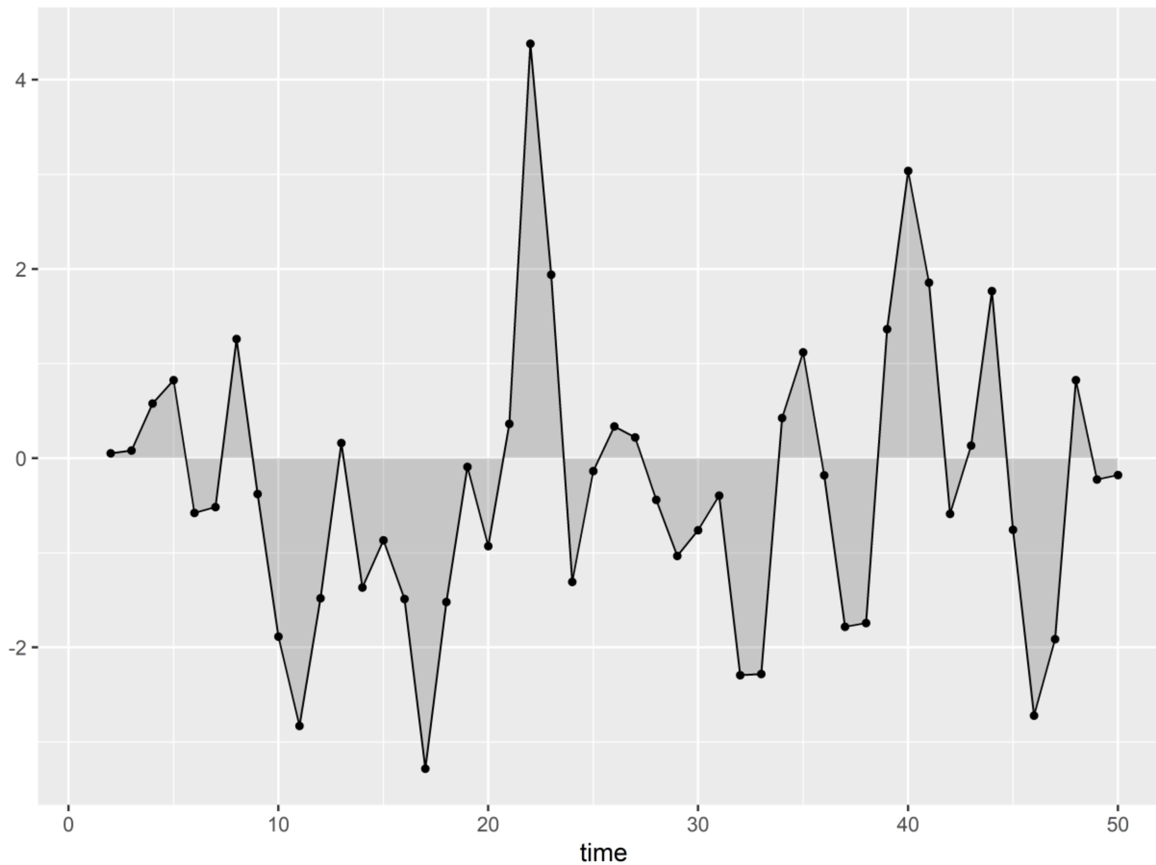
AR(1) impulse response



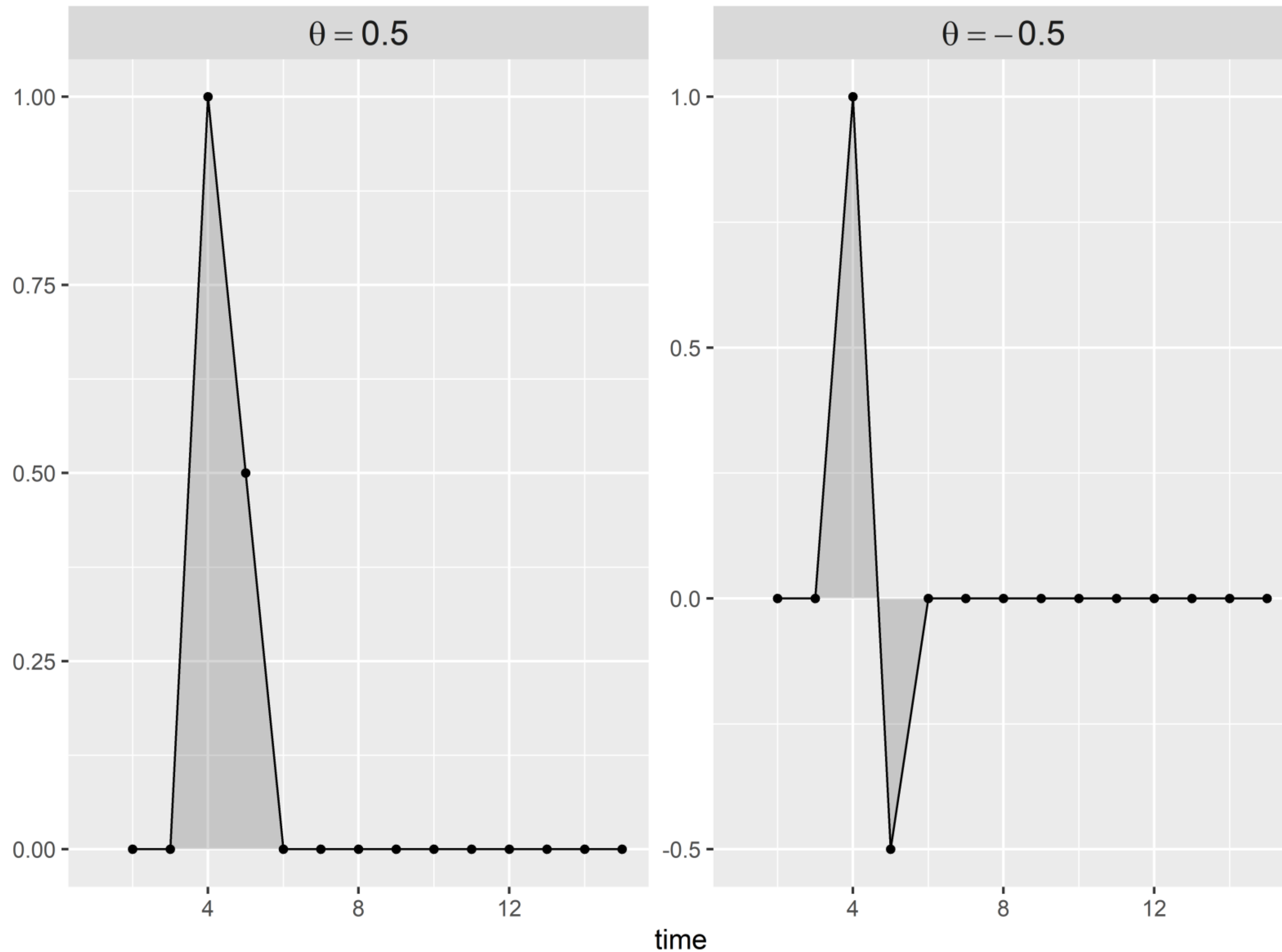
Moving average process of order one – MA(1)

- "mild" serial dependence, observations two or more periods apart are independent
- stationary, weakly dependent
- $y_t = \mu + \varepsilon_t + \theta\varepsilon_{t-1}$,   where ε_t is noise

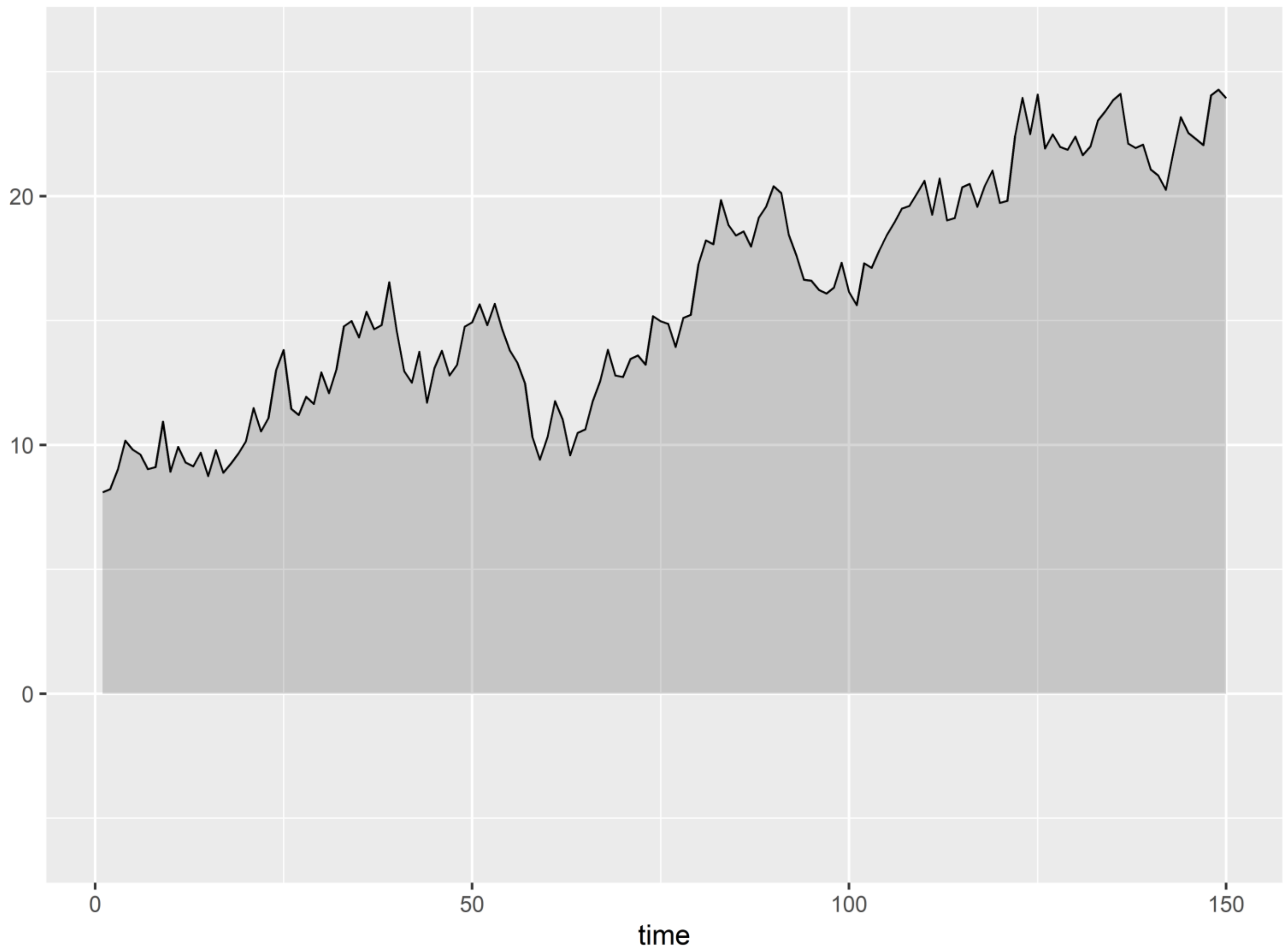
MA(1) process, $\mu = 0$, $\theta = 1$:



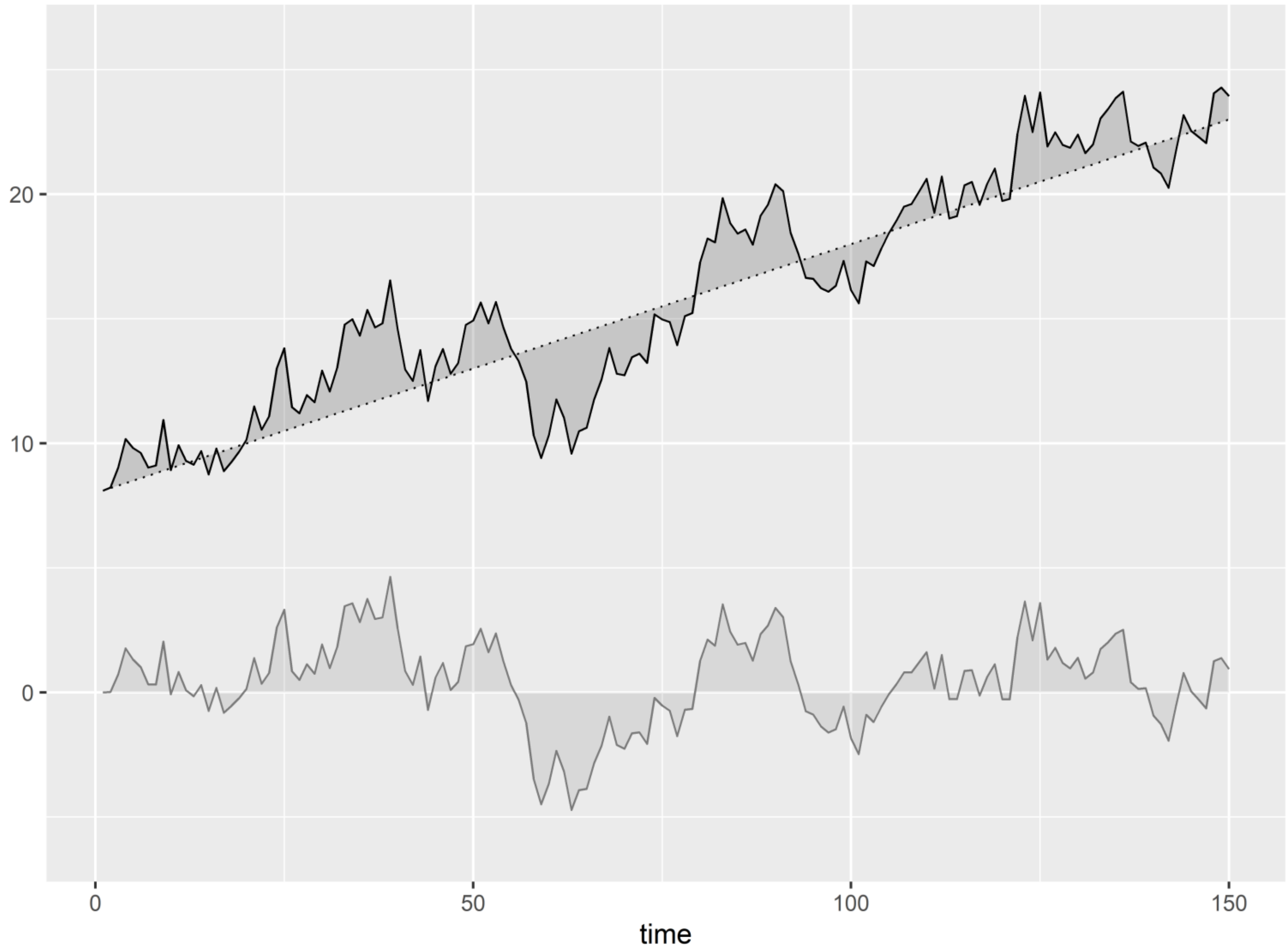
MA(1) impulse response



This is surely non-stationary, …



…but also trend-stationary.



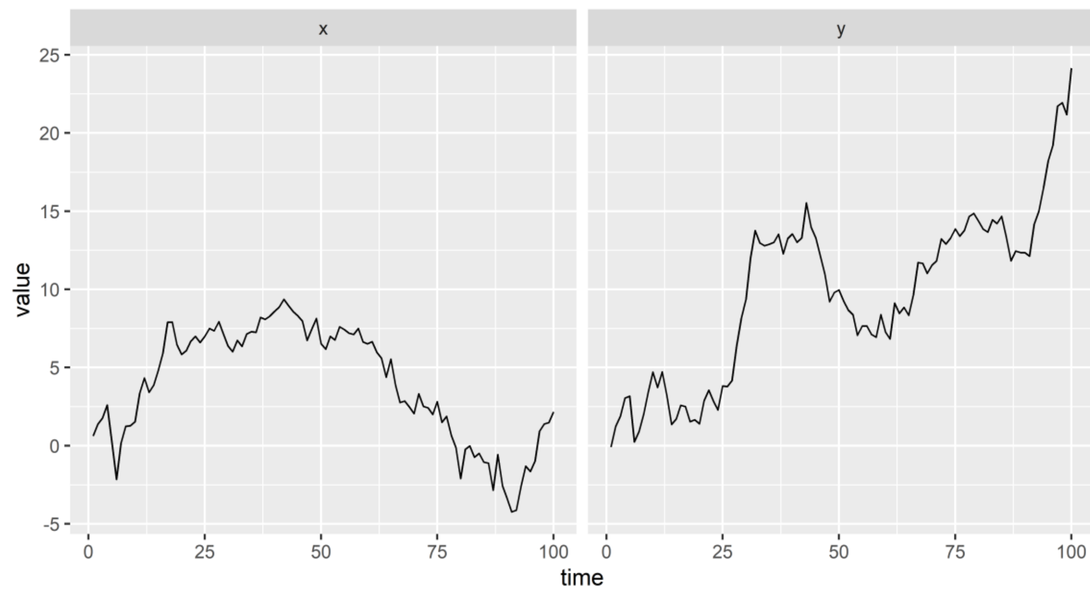
More on random walks

- compare the dependence in AR(1) and random walk:

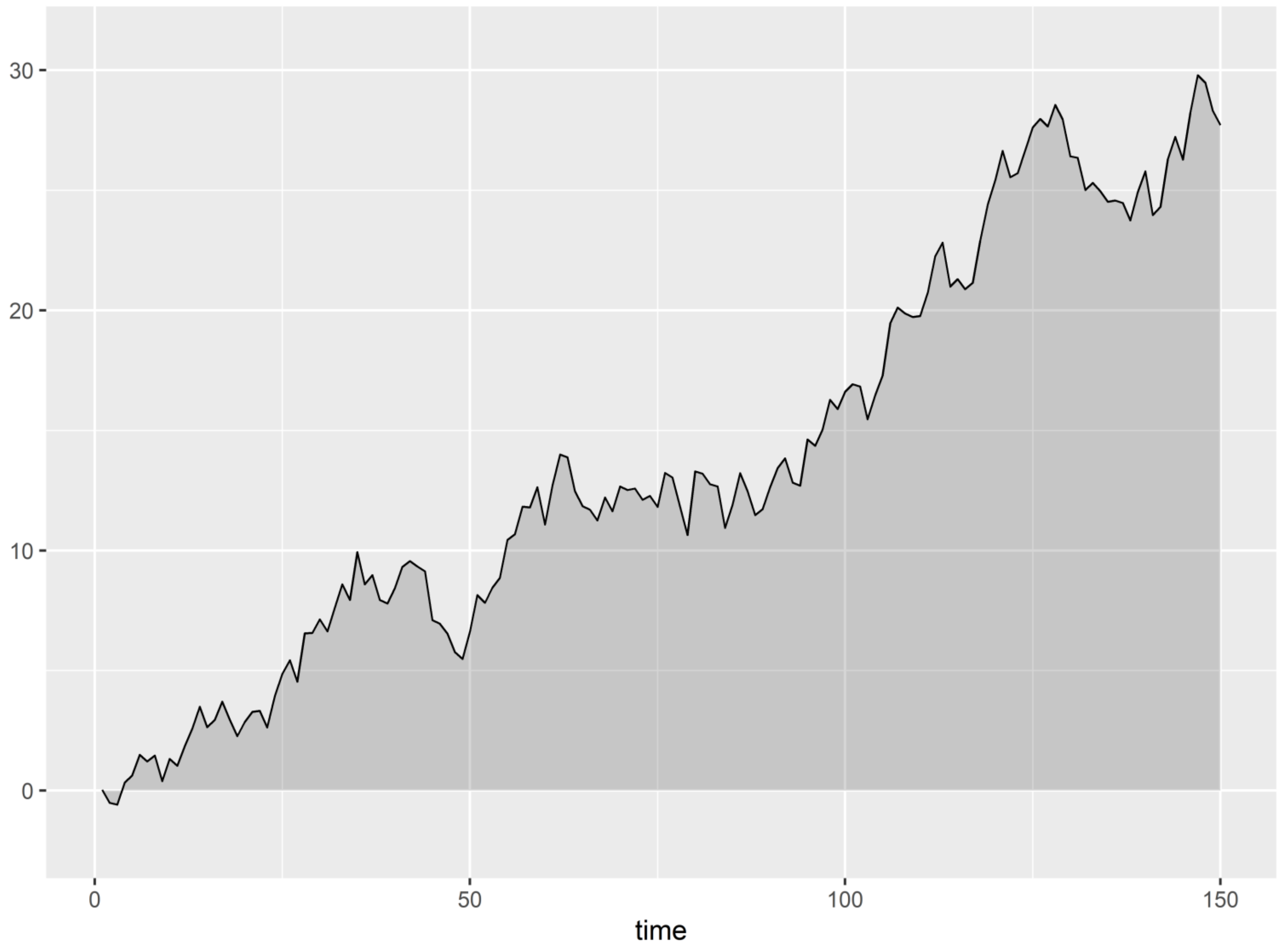
$$\text{AR(1):} \quad E(y_{t+h} \mid y_t) = \rho^h y_t \quad \text{corr}(y_{t+h}, y_t) = \rho^h$$

$$\text{random walk:} \quad E(y_{t+h} \mid y_t) = y_t \quad \text{corr}(y_{t+h}, y_t) = \sqrt{\frac{t}{t+h}}$$

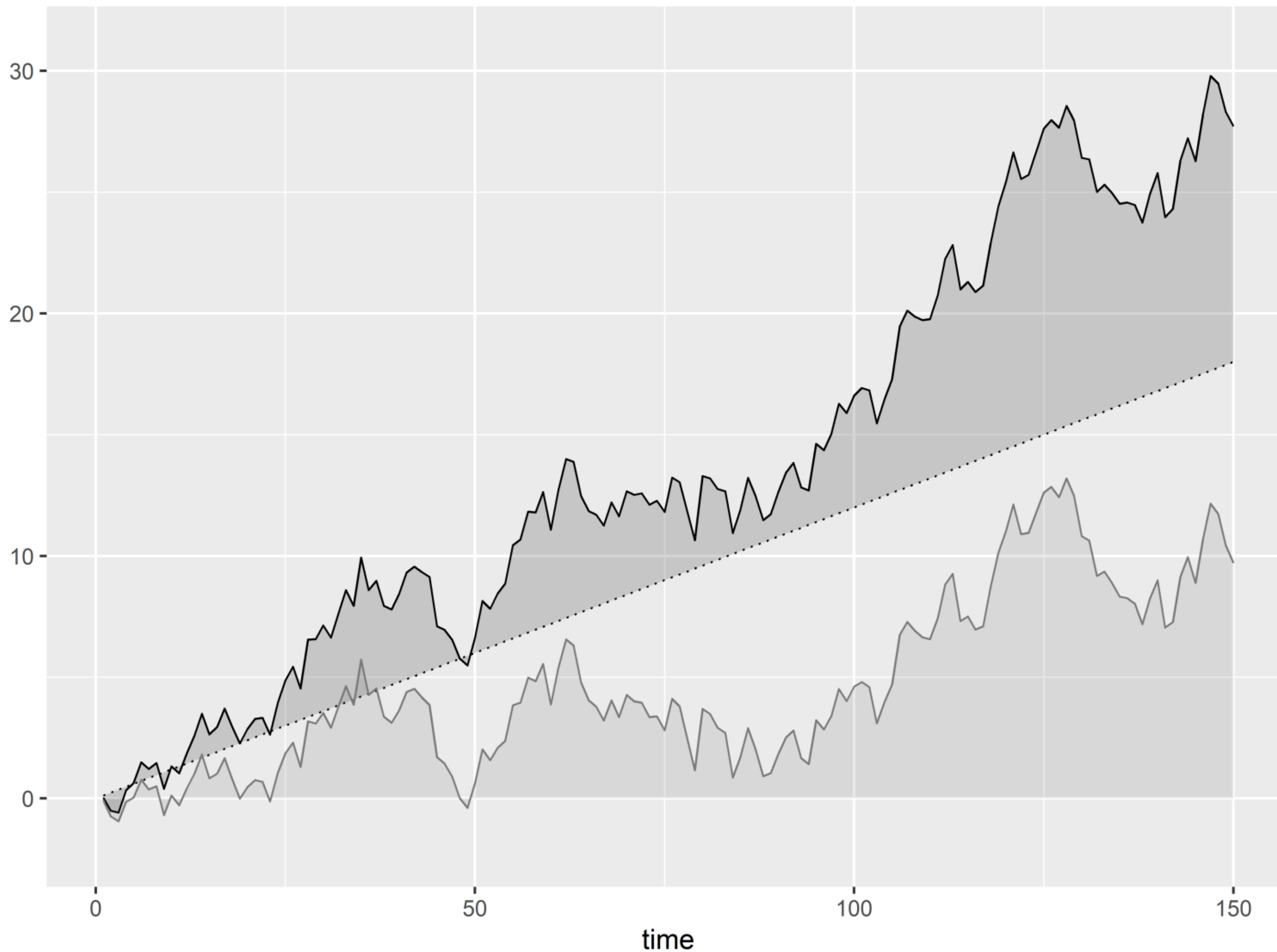
- spurious regression with random walks:
  - 2 independent RWs
  - in a regression of y on x, the effect of x will be significant ($p = 0.0108$)
- RWs can have a drift:
  - $y_t = a_0 + y_{t-1} + e_t$, i.e. $y_t = a_0 t +$ accumulated noise



Random walk with a drift



Drift and “random walk minus drift”



Assumptions needed for regressions with time series I

Assumption **TS.1** (linear in parameters)

The random process $\{(y_t, x_{t1}, x_{t2}, \ldots, x_{tk})\}_{t=1,\ldots,n}$ follows the linear model

$$y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \ldots + \beta_k x_{tk} + u_t ,$$

where $\beta_0, \beta_1, \ldots, \beta_k$ are constant parameters and $\{u_t\}_{t=1,\ldots,n}$ is a series of random errors (disturbances).

Assumption **TS.2** (no perfect collinearity)

In the sample (and therefore in the underlying process), no independent variable is constant nor a perfect linear combination of the others..

Assumption **TS.3** (strict exogeneity)

For each t, the expected value of the error u_t, given the explanatory variables for *all* time periods, is zero: $E(u_t | \mathbf{X}) = 0$, $t = 1, \ldots, n$.

Assumptions needed for regressions with time series II

Assumption **TS.1'** (linear in parameters)

TS.1 + the assumption that $\{(y_t, x_{t1}, x_{t2}, \ldots, x_{tk})\}_{t=1,\ldots,n}$ is stationary and weakly dependent. In particular, the law of large numbers and the central limit theorem can be applied to sample averages.

Assumption **TS.2'** (no perfect collinearity)

Same as TS.2.

Assumption **TS.3'** (contemporaneous exogeneity)

The explanatory variables $\mathbf{x}_t = (y_t, x_{t1}, x_{t2}, \ldots, x_{tk})$ are **contemporaneously exogenous**: $E(u_t | \mathbf{x}_t) = 0, \ \ t = 1, \ldots, n$.

Assumptions needed for regressions with time series III

Assumption **TS.4’** (homoskedasticity)

Conditional on **X**, the variance of u_t is the same for all t: var(u_t|**X**) = var(u_t) = 0 pro t = 1, …, n..

Assumption **TS.5’** (no serial correlation)

Conditional on **X**, the errors in two different time periods are uncorrelated corr(u_s, u_t|**X**) = corr(u_s, u_t) = 0 for any $s \neq t$.

Statistical properties of OLS with time series

- as with cross-sectional data, we can show that the OLS estimator has some favourable properties in time-series regressions
- again, we need some assumptions to show this
- Wooldridge gives 2 alternative sets of assumptions, useful in different settings: TS.1, TS2, … vs. TS.1', TS2', …
  - the first set ("no prime" version), requires **strictly exogenous regressors** (a rather limiting assumption, but needed for small-sample inference)
    - rules out (i)"feedback loops from y_t to x_{t+1} and (ii) an inclusion of the lagged dependent variable among regressors
  - the second set ("prime") instead requires **weak dependence** of the (multivariate) random process $(y_t, x_{t1}, x_{t2}, \ldots, x_{tk})$
    - only asymptotic inference, but much more flexible

Serial correlation of random errors

- a violation of this assumption has similar consequences as heteroskedasticity:
  - the OLS estimator (of $\beta_0,\ldots, \beta_k$) is still *unbiased* and *consistent*
  - however, it is not BLUE; there are other estimators that are, on average, more accurate (asymptotically)
  - the usual statistical inference is not valid (std. errors, t-statistics, p-values are not usable)
- with heteroskedasticity, we mostly just used OLS with robust standard errors
- this is also an option here, however the accuracy of OLS is very limited if random errors exhibit substantial persistence
- in other words, the consequences are typically more severe than under heteroskedasticity
- therefore, we will discuss a method that is tailored for autocorrelation

Durbin-Watson test for autocorrelation

- printed in most regression packages after a time-series regression
- tests for a presence of AR(1) process in the random errors; in fact, as usual, residuals are used for the test instead of the unknown u
- the test requires strictly exogenous regressors; e.g. it rules out equations with a lagged dependent variable among the regressors, such as

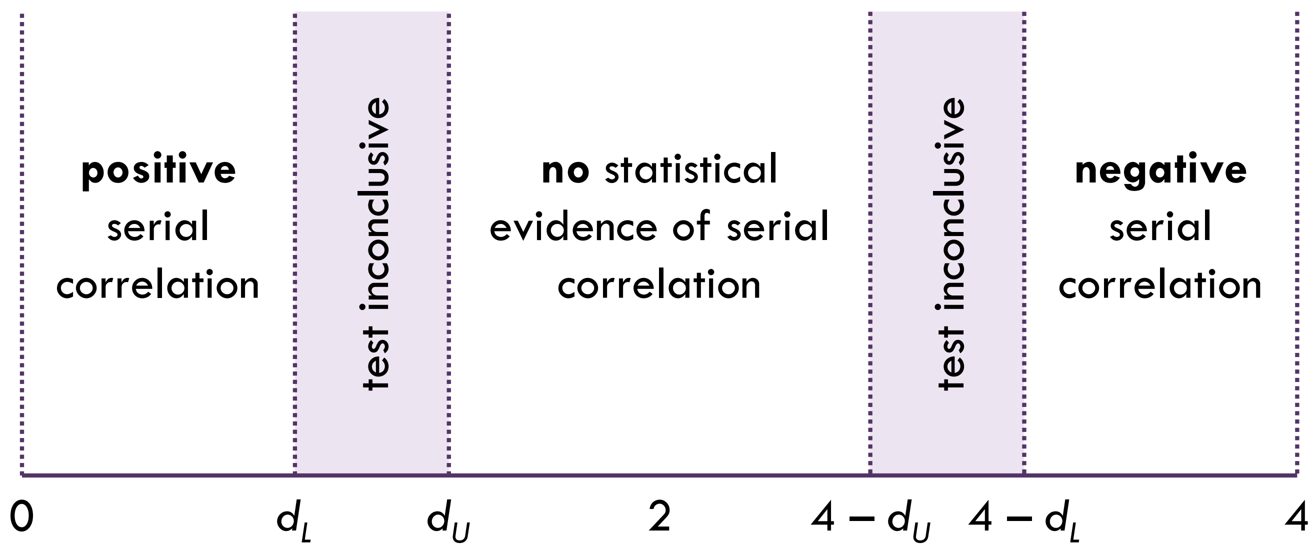
$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 x_t + u_t$$

- moreover, it requires homoskedasticity and normality of random errors
- the test statistic of the test (denoted either d or DW) is closely related to the OLS estimate of ρ in the equation $\hat{u}_t = \rho \hat{u}_{t-1} + error$
- the exact formula is $d = DW = \dfrac{\sum_{t=2}^{n} \left(\hat{u}_t - \hat{u}_{t-1}\right)^2}{\sum_{t=1}^{n} \hat{u}_t^2}$

  and $d \simeq 2\left(1 - \hat{\rho}\right)$

Durbin-Watson test for autocorrelation (cont'd)

- possible values for d are between 0 and 4
- values of d close to 0 indicate positive autocorrelation, values of d close to 4 indicate negative autocorrelation
- statistical tables contain critical values for given n and k
- two critical values given, d_L and d_U, as the D-W test has a region of inconclusiveness (see below)



Breusch-Godfrey test for autocorrelation

- fewer assumptions → should generally preferred to D-W
- procedure to test for the presence of AR(1) in random errors:
  1. After your original OLS regression, save residuals.
  2. Regress $\hat{u}_t$ on $\hat{u}_{t-1}$ and all regressors from your original regression.
  3. Test the null hypothesis that the coefficient on $\hat{u}_{t-1}$ equals zero. (Use the usual t-test.) A rejection means significant evidence of autocorrelation.
- can easily be made robust to heteroskedasticity (just use robust std. errors in step 3)
- can also be modified to higher lags – AR(2), AR(3) etc. – just add more lags of the residuals in step 2 and test for joint significance of all lags
- built in Gretl: Tests → Autocorrelation after OLS regression

Cochrane-Orcutt method

- with serial correlation, OLS is no longer BLUE
- asymptotically more efficient (= accurate) methods exist
- C-O is a simple, widely used alternative

… to be continued …

Lecture 10:
More on random processes and serial correlation

Jan Zouhar

Introductory Econometrics