Module 9.3: Time Series Analysis
Fall Term 2022

Week 1:

Introduction; Descriptive Time Series Analysis

HSLU Lucerne University of Applied Sciences and Arts

Outline in Weeks

1. Introduction; Descriptive Modelling
2. Returns; Autocorrelation; Stationarity
3. ARMA Models
4. Unit Roots; Regressions between Time Series
5. Volatility Modelling
6. Value at Risk
7. Cointegration

Outline

General Information

- Lectures will be a mix of theory and practice (in EViews).
- Slides and additional materials are available on Ilias.
- 90 min. written exam during exam phase, closed book. Details will be communicated later.
- A mock exam will be made available.

Book

- Course is not explicitly based on any book.
- If you prefer to have a book, then I recommend Brooks (2019)[1]. A reading list follows on the next slide.
- I will also make selected problems and solutions available.

[1] Brooks, C. (2019). Introductory Econometrics for Finance (4th ed.). Cambridge: Cambridge University Press.

Reading List: Brooks (2019)

Pre As a refresher, Sections 1.1–1.6 (mathematical foundations), 2.1–2.7 (statistics and distribution theory).
Week 1 (not covered in book)
Week 2 Sections 6.1 and 6.2
Week 3 Sections 6.3–6.10
Week 4 Section 8.1; Section 5.5
Week 5 Sections 9.1–9.16; 9.17
Week 6 (not covered in book)
Week 7 Sections 8.3 – 8.11
Week 7 Sections 11.1–11.7; Section 14.2

Outline

What is Time Series Analysis

- In Module 9.2, you learned about the classical linear regression model (*CLRM*).
- It is used to estimate linear relationships of the form

$$Y_i = \beta_0 + \beta_1 x_i + U_i, \tag{\dagger}$$

  possibly with more than one regressor.
- Typically, 5 assumption are made about the error term: zero mean, constant variance (homoskedasticity), lack of autocorrelation, no correlation with the regressors (orthogonality), normality.
- These are often justifiable for *cross-sectional* data, where each observation i corresponds to a different entity (e.g., a firm, a country, etc.).

What is Time Series Analysis

- In many areas of financial econometrics (risk models, asset pricing, ...), one deals with *time series data* instead; here, every observation corresponds to a different *time period*. Examples:
  - the price of IBM stock on each trading day since Jan 2nd, 2004;
  - monthly inflation in the EUR area since Jan 2002;
  - US GDP growth in every quarter since 1986Q1, etc.
- As seen above, time series may have different *frequencies* (daily, monthly, quarterly, etc.).
- We will only cover *regular* time series: observations occur at equally spaced time points (e.g., daily closing prices for stocks).

What is Time Series Analysis

- To highlight the fact that we are dealing with time series, we use a subscript t instead of i; thus, a regression model such as † would be written

$$Y_t = \beta_0 + \beta_1 x_t + U_t \qquad (\ddagger)$$

  if $\{Y_t\}$ and $\{x_t\}$ are time series.
- Regression ‡ is unlikely to satisfy the CLRM assumtions; time series usually exhibit *autocorrelation*, and often changes in standard deviation (or in "volatility", for stock returns).
- Time series analysis is the study of methods to deal with these salient features.
- The broader goal (as usual in econometrics) is to empirically *verify* economic theories (e.g., the CAPM).
- Another important aspect is *forecasting* (e.g, GDP forecasts, inflation forecasts, Value at Risk forecasts, etc.)

What is Time Series Analysis

- For most of the course, we will consider *univariate* time series analysis.
- This means that instead of a regression like

$$Y_t = \beta_0 + \beta_1 x_t + U_t$$

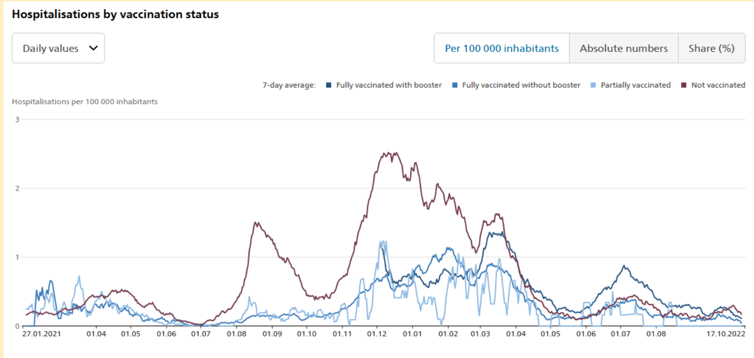
  above, we only have *one* time series $\{Y_t\}$.
- The goal is to describe the (dynamic) behavior of Y_t, e.g., for forecasting.
- We'll start with a purely *descriptive* approach today. Starting next week, we'll move on to actual dynamic models.

Outline

Bitcoin prices

Source: coinmarketcap.com

CoViD19 Hospitalizations

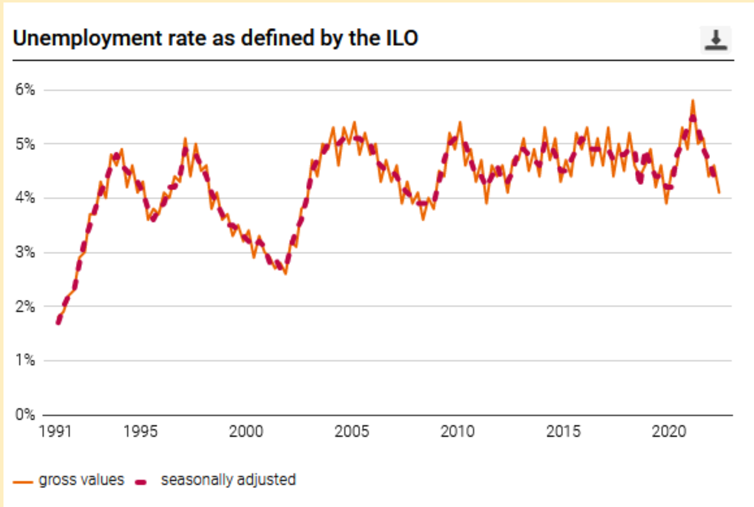


Source: covid19.admin.ch

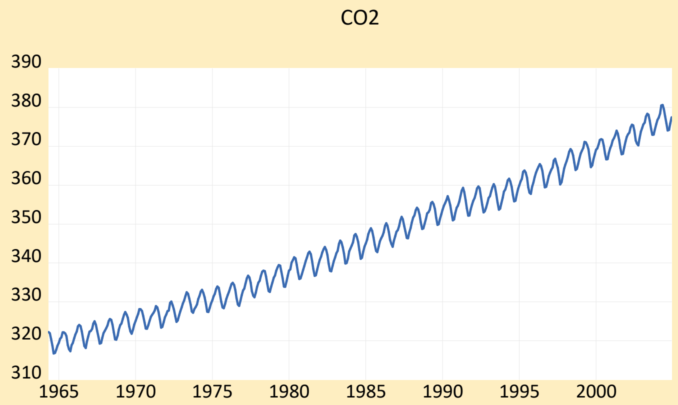
Tesla Stock

Source: Yahoo Finance

Unemployment in Switzerland



Atmospheric CO2 on Mauna Loa, Hawaii



Outline

Time Series Plots

- The above plots were all examples of *time series plots*: plotting the data against time itself.
- This is usually the first thing to do when looking at a new data set.
- We'll see later how to make these plots in EViews.

Decomposing a Time Series

- The main tool of descriptive time series analysis it to decompose it into a *trend*, a *seasonal* component, and a *residual* component, according to the additive model

$$Y_t = F_t + S_t + U_t,$$

where the trend component F_t models long-term movements, the seasonal component S_t measures systematic seasonal patterns, and the residual component U_t contains anything that cannot be explained by the other two[2].
- The Mauna Loa data make the trend and seasonal component very obvious.

[2] Sometimes economic time series also contain a cyclical component stemming from the business cycle, but we will ignore this here.

Outline

Estimating a Linear Trend by OLS

- One way to estimate a *linear trend* is to just regress the data on an intercept and time itself, i.e.,

$$Y_t = \beta_0 + \beta_1 t + U_t.$$

- The estimated trend is then

$$\widehat{F}_t = \widehat{\beta_0} + \widehat{\beta_1} t.$$

Estimating a Quadratic Trend by OLS

- As seen in the exercises, it is also possible to have a nonlinear trend. One example is a *quadratic* trend. This can be estimated via the regression

$$Y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + U_t.$$

- The estimated trend is then

$$\widehat{F}_t = \widehat{\beta_0} + \widehat{\beta_1} t + \widehat{\beta_2} t^2.$$

Estimating an Exponential Trend by OLS

- Another possibility is to use an *exponential* trend. The model is then

$$F_t = \beta_0 \cdot \beta_1^t.$$

- To estimate this by OLS, one takes logs:

$$\log(F_t) = \log(\beta_0) + \log(\beta_1) \cdot t =: c + b \cdot t.$$

- Adding an error term, this exponential trend can be estimated via the regression

$$\log(Y_t) = c + b \cdot t + U_t.$$

- The resulting trend function is

$$F_t = \widehat{\beta_0} \cdot \widehat{\beta_1}^t, \quad \text{where } \widehat{\beta_0} = \exp(\hat{c}), \widehat{\beta_1} = \exp(\hat{b}).$$

Interpreting an Exponential Trend

- If the trend is
$$F_t = \beta_0 \cdot \beta_1^t,$$
then
$$\frac{F_t}{F_{t-1}} = \frac{\beta_0 \cdot \beta_1^t}{\beta_0 \cdot \beta_1^{t-1}} = \beta_1;$$
i.e., Y_t grows by $100 \cdot (\beta_1 - 1)$% per period, on average.
- Example ($\beta_0 = 1, \beta_1 = 1.05$):
$$F_t = 1.05^t,$$
so Y_t grows by 5% a year, on average (cf. compounding interest).

Estimating the Trend via Moving Averages

- Another approach, which has the advantage of adapting to the data automatically, rather than pre-specifying a functional form (linear, quadratic, exponential), is to estimate the trend via a *moving average.*
- E.g., for a third-order moving average ($k = 3$),
$$\widehat{F}_t = (Y_{t-1} + Y_t + Y_{t+1})/3.$$
- Choice of k: the higher, the smoother. If seasonality is present, k should cover at least a full cycle.
- *Downside*: $(k+1)/2$ values at the end points cannot be computed. Thus also not useful for forecasting.
- Note: for a moving average of even order, one averages $k+1$ data points, but the endpoints get half the weight. E.g., with $k = 4$,
$$\widehat{F}_t = \left(\frac{1}{2}Y_{t-2} + Y_{t-1} + Y_t + Y_{t+1} + \frac{1}{2}Y_{t+2}\right)/4.$$

Outline

Dummy Variables

- Simple way to address the seasonality S_t: *seasonal dummies*, which take the value one in one season, and zero in all others.
- Example (see next page): say we have quarterly data. Then we would use four dummies, defined, for $j \in \{1, \ldots, 4\}$, as
$$d_{jt} = \begin{cases} 1, \text{ if observation } t \text{ is in season } j, \\ 0, \text{ otherwise.} \end{cases}$$
- Effectively, every season gets its own intercept.
- Careful: if a full set of dummies is included, then the intercept must be left off, otherwise the regressors are perfectly collinear; this is the *dummy variable trap*.
- Alternatively, keep the intercept, but remove one of the dummies. That season then becomes the baseline, and the other dummies measure the average difference from the baseline, per season.

Example

t	Date	d_1	d_2	d_3	d_4
0	2021Q1	1	0	0	0
1	2021Q2	0	1	0	0
2	2021Q3	0	0	1	0
3	2021Q4	0	0	0	1
4	2022Q1	1	0	0	0
5	2022Q2	0	1	0	0
6	2022Q3	0	0	1	0
7	2022Q4	0	0	0	1
$\vdots$					

Example continued

- If we include a linear trend, then the model becomes

$$\begin{aligned} Y_t &= F_t + S_t + U_t \\ &= \beta_1 \cdot t + \alpha_1 d_{1,t} + \alpha_2 d_{2,t} + \alpha_3 d_{3,t} + \alpha_4 d_{4,t} + U_t, \end{aligned}$$

which can be estimated by OLS.
- If we want to produce a forecast for Y_3 (which is in Season 4; note that in EViews, t starts at zero), then

$$\widehat{Y_3} = \widehat{\beta_1} \cdot 3 + \widehat{\alpha_4}.$$

Example continued

- Alternatively, include an intercept and drop one dummy:

$$Y_t = \beta_0 + \beta_1 \cdot t + \alpha_1 d_{1,t} + \alpha_2 d_{2,t} + \alpha_3 d_{3,t} + U_t.$$

- Now season 4 is the baseline, and we have

$$\widehat{Y_3} = \widehat{\beta_0} + \widehat{\beta_1} \cdot 3.$$

- The other seasons are measured in deviation from the baseline; e.g., α_2 is the average difference between seasons 2 and 3:

$$\widehat{Y_2} = \widehat{\beta_0} + \widehat{\beta_1} \cdot 2 + \widehat{\alpha_3}.$$

Outline

Learning Goals

Students

- Know the difference between cross-sectional and time series data,
- know what a regular time series is, and what its frequency is,
- are able to decompose a time series into trend, seasonality, and the residual component in EViews,
- and are able to produce time series plots in EViews.

Homework

- Freshen up your statistics knowledge, if needed.
- Exercise 1.

Module 9.3: Time Series Analysis
Fall Term 2022

Week 2:

Returns; Autocorrelation; Stationarity

HSLU Lucerne University of Applied Sciences and Arts

Outline in Weeks

1. Introduction; Descriptive Modelling
2. Returns; Autocorrelation; Stationarity
3. ARMA Models
4. Unit Roots; Regressions between Time Series
5. Volatility Modelling
6. Value at Risk
7. Cointegration

Outline

Asset Returns

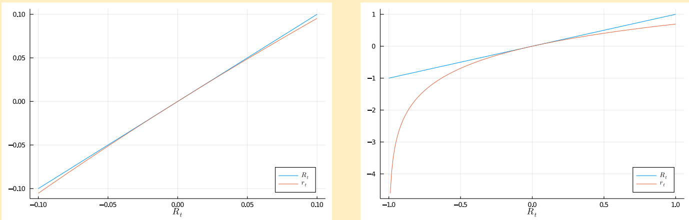
- We consider two definitions of returns:
  1. *Simple* return between dates $t-1$ and t [or: in period t]
     $$R_t = \frac{P_t - P_{t-1}}{P_{t-1}},$$
     where P_t is the asset price at time t.
  2. Continuously compounded return or *log return*
     $$r_t = \log\left(\frac{P_t}{P_{t-1}}\right) = \log(1 + R_t).$$
- They are typically very close for daily returns, as
  $$r_t = \log(1 + R_t) \approx R_t,$$
  when $R_t \approx 0$.

- Log returns and 'simple' returns often are very close, as

$$r_t = \ln(1 + R_t) \approx R_t \text{ when } R_t \approx 0.$$

- Simple returns are bounded below by -1 (100% loss). Log returns live on $(-\infty, \infty)$. Easier to model (e.g., normal distribution).

Simple vs. Log Returns



Log returns: Intuition

- If a one-period interest rate of r is compounded n times, then

$$P_t = (1 + r/n)^n P_{t-1}.$$

- As $n \to \infty$, $(1 + r/n)^n \to e^r$, so

$$P_t = e^r P_{t-1} \Leftrightarrow r = \log(P_t/P_{t-1}) = \ln P_t - \log P_{t-1}.$$

Portfolio Returns

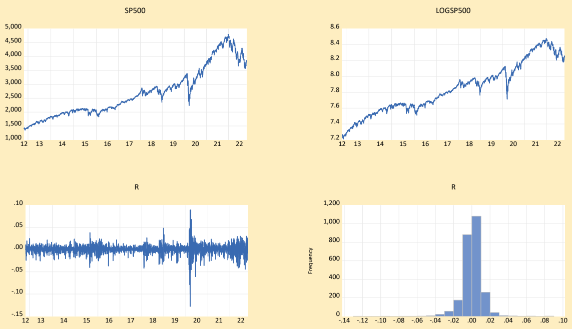
- Advantage of continuously compounded returns: *multi-period return* is sum of single-period returns.
- Advantage of simple returns: *portfolio return* is weighted sum of asset returns.
- Proof: If an investor buys n_i shares in stock i, then the value of the portfolio at time $t-1$ is $V_{t-1} = \sum_{i=1}^n n_i P_{i,t-1}$.
- Ignoring dividends, the payoff is $V_t = \sum_{i=1}^n n_i P_{i,t}$, so the return on the portfolio is

$$\begin{aligned} R_{p,t} = \frac{V_t - V_{t-1}}{V_{t-1}} &= \frac{\sum_{i=1}^n n_i (P_{i,t} - P_{i,t-1})}{V_{t-1}} \\ &= \sum_{i=1}^n \underbrace{\frac{n_i P_{i,t-1}}{V_{t-1}}}_{w_i} \underbrace{\frac{(P_{i,t} - P_{i,t-1})}{P_{i,t-1}}}_{R_{i,t}} = \sum_{i=1}^n w_i R_{i,t}. \end{aligned}$$

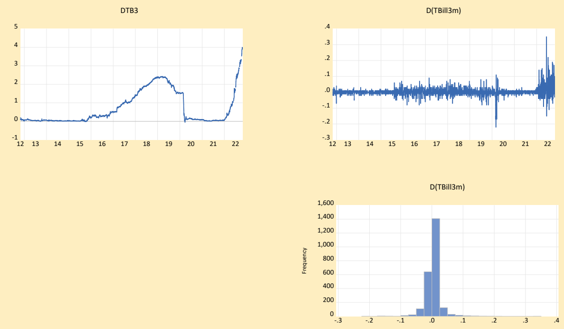
Stylized Facts of Asset Returns

- Prices display (time-varying) *trend*, and variation proportional to price level (motivation for taking logs).
- Returns have constant mean close to zero, and very little autocorrelation.
- Returns display *volatility clustering*: alternating periods of high and low variability.
- Returns have non-Gaussian distribution, *fat tails* (excess kurtosis).
- Interest rates display long swings, very slow *mean-reversion*.
- Interest rate changes have similar characteristics as returns.

Example: S&P 500 index values and returns, 10/29/2012–10/27/2022



Example: 3 Month T-Bill rate, 10/29/2012–10/26/2022



Testing Normality

- Normality can be tested by examining the *skewness* and *kurtosis*.
- Skewness $\mathrm{SK} = m_3/\sqrt{m_2^3}$ and kurtosis $\mathrm{K} = m_4/m_2^2$, where m_j is the j-th centralized moment[1] $m_j = \mathbb{E}[(r_t - \mathbb{E}[r_t])^j]$.
- A normal distribution has SK=0 and K=3.
- *Jarque-Bera normality test*:

$$\mathrm{JB} = \frac{T}{6}\widehat{\mathrm{SK}}^2 + \frac{T}{24}(\widehat{\mathrm{K}} - 3)^2,$$

where the skewness and kurtosis of r_t can be estimated as

$$\widehat{\mathrm{SK}} = \hat{m}_3/\sqrt{\hat{m}_2^3}, \quad \text{and} \quad \widehat{\mathrm{K}} = \hat{m}_4/\hat{m}_2^2, \quad \text{with } \hat{m}_j = \frac{1}{T}\sum_{t=1}^{T}(r_t - \bar{r})^j.$$

- Under the null hypothesis of normality: $\mathrm{JB} \xrightarrow{d} \chi_2^2$.

[1] I.e., the second centralized moment m^2 is just the variance, otherwise known as σ^2.

Outline

Stochastic Processes

- Time series analysis is concerned with modelling, estimating, analyzing and forecasting returns and other financial and economic variables.
- A *time series* $\{y_t, t = 1, 2, \ldots, T\}$ is a collection of subsequent observations on a particular variable. We view such a time series as a *realization* of a *discrete-time stochastic process* $\{Y_t, t = 1, 2, \ldots\}$, which is a collection of (dependent) random variables.
- The goal is to determine which process $\{Y_t\}$ generated the data.
- The distinction between $\{Y_t\}$ (the process) and $\{y_t\}$ (the realization) will usually not be emphasized.
- We will not consider continuous-time stochastic processes here (e.g. Brownian motion).

White Noise

- An important example of a stationary process is the *white noise* process, which has zero mean[2] and zero autocovariances.

$$\begin{aligned}\mathbb{E}[U_t] &= 0,\\ \mathrm{var}(U_t) &= \mathbb{E}\left[U_t^2\right] = \sigma^2,\\ \mathrm{cov}(U_t, U_{t-k}) &= \mathbb{E}[U_t U_{t-k}] = 0, \qquad k = 1, 2, \ldots\end{aligned}$$

- The notation U_t emphasizes the similarity to regression errors.
- White noise is *unpredictable*.
- It is the building block for other processes (which may be predictable).

[2] Brooks allows a white noise process to have a non-zero mean. Usually such a process is called an *uncorrelated* process.

Outline

Excursus: The Efficient Market Hypothesis

- The *weak form EMH*[3] posits that past prices and returns cannot predict future returns.
- This implies that no fund manager can consistently outperform the market, at least based on historical prices alone.
- If weak form EMH holds, then returns should be *uncorrelated*. Since the mean return is small for daily data, they should therefore resemble wight noise.
- An important application of time series analysis ist testing whether the EMH holds.
- The most basic way to do this is to test whether the returns have been generated by a white noise process.

[3] Fama (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *Journal of Finance*, 25(2), pp. 383–417.

Outline

Autocorrelation Function

- Recall that if a process $\{Y_t\}$ is white noise, then it is uncorrelated at all lags (i.e., Y_t should be uncorrelated with Y_{t-1}, with Y_{t-2}, etc.)
- Formally, its *autocorrelation function (ACF)* τ_s is zero for all s, where the ACF is defined from the *autocovariances* γ_s, as

$$\tau_s = \mathrm{Corr}(Y_t, Y_{t-s}) = \frac{\mathrm{Cov}(Y_t, Y_{t-s})}{\mathrm{Var}(Y_t)} = \frac{\gamma_s}{\gamma_0}, \qquad s = 1, 2, \ldots$$

PSA: Population vs. Sample Quantities

- It is important to note that the ACF is a property of a *process*, not of a *sample* (i.e., the observed time series). This makes them *population quantities* or *parameters*.
- The statement on the previous slide says that the random variables $\{Y_1, Y_2, \ldots\}$ generated by a white noise process are uncorrelated.
- Population quantities are *unobserved*. The best we can hope for is to *estimate* them from a sample (a time series).

PSA: Population vs. Sample Quantities

- To use the normal distribution as an analogy: it has two parameters, μ and σ^2. These are *parameters* and thus *unobserved*.
- In a simulation exercise, I can *pretend* to know what μ and σ^2 are.
- E.g., I can set $\mu = 0$ and $\sigma^2 = 4$, simulate 1000 random numbers y_t, and give them to you.
- Unlike me, you won't know what μ and σ^2 are. At best, you can *estimate* them, based on the *sample mean and variance*

$$\bar{y} \equiv \frac{1}{N}\sum_{i=1}^{N} y_i \quad \text{and} \quad s_y^2 \equiv \frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{y})^2.$$

- If the sample is large enough, then these will be *close* to $\mu = \mathbb{E}[Y]$ and $\sigma^2 = \mathrm{Var}(Y)$ by the law of large numbers (*LLN*).

The Correlogram

- Applying the analogy to the ACF, the *Sample ACF* or *correlogram* is defined as

$$\hat{\tau}_s = \frac{\hat{\gamma}_s}{\hat{\gamma}_0} = \frac{\sum_{t=s+1}^{T}(y_t - \bar{y})(y_{t-s} - \bar{y})}{\sum_{t=1}^{T}(y_t - \bar{y})^2}, \qquad \bar{y} = \frac{1}{T}\sum_{t=1}^{T} y_t.$$

- The correlogram is a *sample quantity*, i.e., I can compute it from a given time series.
- If I want to test if the time series is white noise, I can compare my SACF to the ACF of a white noise process.
- If the two are significantly different, then I can reject the null that the time series was generated by a white noise process.
- See exercises and the spreadsheet `simulations.xlsx`.

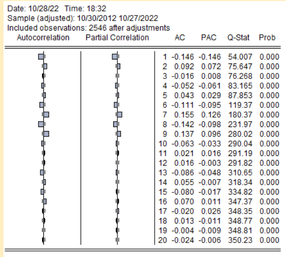
Testing if an Autocorrelation is Zero

- One can show that under the null that the data were generated by a white noise process, the sample autocorrelations are asymptotically[4] normally distributed with zero mean and variance $1/T$.
- This implies that a sample autocorrelation is significantly different from zero if its absolute value is larger than $1.96/\sqrt{T}$.
- We can also test whether the first m autocorrelations are zero jointly: under $H_0 : \tau_s = 0, s \geq 1$, the *Ljung-Box* Q-statistic

$$Q(m) = T(T+2)\sum_{s=1}^{m}\frac{\hat{\tau}_s^2}{T-s} \quad \xrightarrow{d} \quad \chi^2(m).$$

[4] Formally: under $H_0 : \tau_s = 0, s \geq 1$, $\sqrt{T}\hat{\tau}_s \xrightarrow{d} N(0,1)$.

Example: Correlogram of S&P500 returns


Date: 10/28/22 Time: 18:32
Sample (adjusted): 10/30/2012 10/27/2022
Included observations: 2546 after adjustments

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.146	-0.146	54.007	0.000
		2	0.092	0.072	75.647	0.000
		3	-0.016	0.008	76.268	0.000
		4	-0.052	-0.061	83.165	0.000
		5	0.043	0.029	87.853	0.000
		6	-0.111	-0.095	119.37	0.000
		7	0.155	0.126	180.37	0.000
		8	-0.142	-0.098	231.97	0.000
		9	0.137	0.096	280.02	0.000
		10	-0.063	-0.033	290.04	0.000
		11	0.021	0.016	291.19	0.000
		12	0.016	-0.003	291.82	0.000
		13	-0.086	-0.048	310.65	0.000
		14	0.055	-0.007	318.34	0.000
		15	-0.080	-0.017	334.82	0.000
		16	0.070	0.011	347.37	0.000
		17	-0.020	0.026	348.35	0.000
		18	0.013	-0.011	348.77	0.000
		19	-0.004	-0.009	348.81	0.000
		20	-0.024	-0.006	350.23	0.000

- Thin lines represent critical value $1.96/\sqrt{T}$, so autocorrelations at lags 1, 2, 5–10, and 13–17 are significant. Confirmed by the Q-stats, whose p-values are all less than 5%.
- Conclusion: some autocorrelation, and hence predictability, in the returns; returns are not white noise.
- Unclear whether predictability is sufficient to exploit with some trading strategy.

Outline

From Returns to Asset Prices

- We have seen that the EMH suggests that white noise is a reasonable model for stock returns:
  $r_t = U_t, \quad$ where U_t is white noise (not necessarily normal).
- Recall the definition of log returns:

$$r_t = \log P_t - \log P_{t-1}.$$

- Putting the two together implies that

$$\log P_t - \log P_{t-1} = U_t \Leftrightarrow$$
$$\log P_t = \log P_{t-1} + U_t.$$

- This characterizes the asset price as a *random walk*.

Asset Returns Stochastic Processes The EMH The ACF The Random Walk I(0) and I(1) Processes Epilogue

Definition

A *random walk* is the stochastic process

$$Y_t = Y_{t-1} + U_t,$$

where U_t is white noise and Y_0 is some fixed starting value.

Asset Returns Stochastic Processes The EMH The ACF The Random Walk I(0) and I(1) Processes Epilogue

Properties of the Random Walk

- The random walk behaves very differently from white noise.
- A quick calculation shows that

$$Y_t = Y_0 + U_1 + U_2 + \cdots + U_t = Y_0 + \sum_{s=1}^{t} U_s.$$

- From this, it is immediate (see exercises) that

$$\mathbb{E}(Y_t) = Y_0 \quad \text{and}$$
$$\mathrm{var}(Y_t) = \sigma^2 t.$$

- One can also show that

$$\mathrm{corr}(Y_t, Y_{t-k}) = \sqrt{(t-k)/t}.$$

Asset Returns Stochastic Processes The EMH The ACF The Random Walk I(0) and I(1) Processes Epilogue

Properties of the Random Walk

- In words:
  - The effect of a "shock" U_t is permanent; U_t is in all future values $Y_s, s \geq t$, whereas for a white noise process, U_t only affects Y_t.
  - The variance increases over time, because we add up more and more of the U_t, all of which are random.
  - The correlogram decreases slowly, approximately linearly (see also `simulation.xlsx`).
- We say that a random walk is not *mean reverting*; one can show that it will (eventually) hit each and every level L, and its excursions can take arbitrarily long.

Outline

Stationary Processes

- Earlier, we rejected the null that the returns on the S&P500 are white noise (although they are close).
- This also implies that the log stock prices are not (exactly) random walks.
- This means that we need to generalize these concepts to allow for other types of stochastic process.
- Specifically, instead of pure white noise, we will consider *stationary processes*.
- Similarly, we generalize the concept of a random walk to *integrated processes*.
- Specific instances of these processes (ARMA and ARIMA models, respectively) will be considered in Weeks 3 and 4.

Stationarity

Definition

A process $\{y_t\}$ is called *weakly stationary* (or second-order, covariance stationary) if the first two moments are time-invariant:

$$\mathbb{E}[Y_t] = \mu, \quad \mathrm{var}(Y_t) = \gamma_0 = \sigma^2, \quad \mathrm{cov}(Y_t, Y_{t-s}) = \gamma_s, \quad t \in \{0, \dots, T\}, s \geq 1.$$

- This means that the mean, variance, and autocovariances (or autocorrelations) do not change over time; i.e., the autocovariance γ_s depends only on the lag s, not on t.
- Intuitively, there should not be a significant difference if I calculate the mean, variance, and ACF from the first or second half of the sample.
- White noise is one example of a *stationary* process.
- As we saw, the random walk is not stationary; its variance changes over time.

Integrated Processes

- Recall that if returns are white noise, then log prices follow a random walk:

$$\log P_t = \log P_{t-1} + U_t$$

- Alternatively, if log prices follow a random walk, then returns are white noise:

$$r_t = \log P_t - \log P_{t-1} = U_t$$

- We write $\Delta \log P_t$ for $\log P_t - \log P_{t-1}$.
- So a process Y_t is a random walk if ΔY_t is white noise.

Integrated Processes

- We saw above that for the S&P500, we did not get white noise after differencing, but some other stationary process.
- Such processes are called *integrated*.

Definition

A process Y_t is called *integrated* of order 1, or *I(1)*, if it is non-stationary itself, but $\Delta Y_t = Y_t - Y_{t-1}$ is stationary.

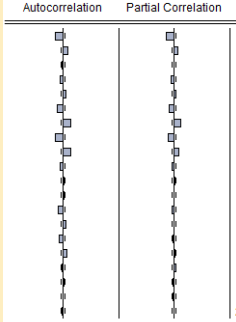
- The random walk is the simplest example of an I(1) process.
- A stationary process is also called *I(0)*.
- An I(2) process would need to be differenced twice to be stationary, but this is rarely necessary in practice.

Properties of Integrated Processes

- Integrated processes have correlograms that stay close to one, which *die out very slowly*.
- An informal way to check whether the stationarity assumption is reasonable is by inspecting the graph and the correlogram of the series. If the graph displays a tendency to revert to a constant mean, with a more or less constant variance, and the correlogram converges to zero *exponentially fast*, then stationarity may be assumed. A formal test will be introduced later.
- Besides prices, many financial and economic time series (e.g., GDP) do not seem to be stationary, because they display a trending mean, and a variance that increases with the level of the process. The latter phenomenon is usually dealt with by a log-transformation, but then quite often the series is still not stationary.

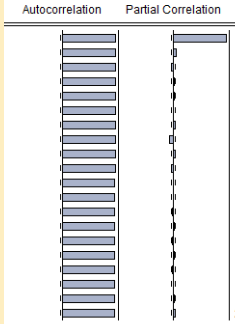
Example: ACF of S&P500 returns and log prices

Date: 10/28/22 Time: 18:32
Sample (adjusted): 10/30/2012 10/27/2022
Included observations: 2546 after adjustments



	AC	PAC	Q-Stat	Prob
1	-0.146	-0.146	54.007	0.000
2	0.092	0.072	75.647	0.000
3	-0.016	0.008	76.268	0.000
4	-0.052	-0.061	83.165	0.000
5	0.043	0.029	87.853	0.000
6	-0.111	-0.095	119.37	0.000
7	0.155	0.126	180.37	0.000
8	-0.142	-0.098	231.97	0.000
9	0.137	0.096	280.02	0.000
10	-0.063	-0.033	290.04	0.000
11	0.021	0.016	291.19	0.000
12	0.016	-0.003	291.82	0.000
13	-0.086	-0.048	310.65	0.000
14	0.055	-0.007	318.34	0.000
15	-0.080	-0.017	334.82	0.000
16	0.070	0.011	347.37	0.000
17	-0.020	0.026	348.35	0.000
18	0.013	-0.011	348.77	0.000
19	-0.004	-0.009	348.81	0.000
20	-0.024	-0.006	350.23	0.000

Date: 11/03/22 Time: 15:29
Sample: 10/29/2012 10/27/2022
Included observations: 2547



	AC	PAC	Q-Stat	Prob
1	0.998	0.998	2541.7	0.000
2	0.997	0.050	5077.1	0.000
3	0.995	-0.037	7605.3	0.000
4	0.994	0.017	10127.	0.000
5	0.992	0.020	12642.	0.000
6	0.991	-0.004	15151.	0.000
7	0.990	0.041	17654.	0.000
8	0.988	-0.075	20150.	0.000
9	0.987	0.039	22640.	0.000
10	0.985	-0.028	25123.	0.000
11	0.984	0.003	27599.	0.000
12	0.982	0.002	30069.	0.000
13	0.980	-0.017	32532.	0.000
14	0.979	0.018	34988.	0.000
15	0.977	-0.010	37438.	0.000
16	0.976	0.017	39881.	0.000
17	0.974	-0.022	42318.	0.000
18	0.973	-0.000	44748.	0.000
19	0.971	0.011	47171.	0.000
20	0.970	0.027	49588.	0.000

Note: Partial Autocorrelation Function

- The EViews output also shows the (sample) *partial autocorrelation function* ((S)PACF) $\hat{\tau}_{kk}, k = 1, 2, \ldots$, where $\hat{\tau}_{kk}$ is the OLS estimator of τ_{kk} in the regression

$$y_t = \alpha + \tau_{k1} y_{t-1} + \ldots + \tau_{kk} y_{t-k} + e_t.$$

Note: this is not the model for y_t, just a regression to estimate τ_{kk}!

- The PACF measures the correlation between y_t and y_{t-k}, *controlling* for the effect of the intermediate lag. I.e., τ_{kk} only measures the *direct* effect of y_{t-k} on y_t.
- For a random walk, it drops to zero after the first lag, because only y_{t-1} has a direct effect.
- For a stationary process, the ACF and PACF converge to zero at a geometric (exponential) rate as k increases.
- If the sample ACF and PACF of a time series do not seem to converge at all, or too slowly (linearly), then this is an indication of nonstationarity.

More Properties of Integrated Series

- *No mean-reversion*. Like the random walk, I(1) processes do not revert to a mean.
- *Persistence of shocks*. Also, the effect of past shocks u_{t-i} does not die out, whereas for stationary series the effect will decay exponentially. Important for economic policy.
- *Increasing forecast intervals*. For I(0) time series, the long-run 95% forecast interval converges to the unconditional mean $\pm$ twice the unconditional standard deviation. For an I(1) process the forecast variance does not converge, so forecasts intervals keep increasing.
- *Spurious regressions*. When regressing two integrated time series onto each other, the R^2 and t-statistic may become very large even if they are totally independent. This is avoided if we regress Δy_t on Δx_t.
- *Asymptotic properties of estimators and tests*. In regressions with I(1) variables, the usual statistical theory breaks down (asymptotic normality of estimators, t-tests, etc).

Outline

Learning Goals

Students

- know the definitions of simple and log returns,
- know the definition of white noise,
- understand the ACF and PACF and their sample analogs,
- are able to use the correlogram and Q-statistics to test if a series was generated by a white noise process, and
- are able to distinguish stationary and integrated processes.

Homework

- Exercise 2
- Questions 9b and 12b from Chapter 6 of Brooks (2019)

Module 9.3: Time Series Analysis
Fall Term 2022

Week 3:

ARMA Models

HSLU Lucerne University of Applied Sciences and Arts

Outline in Weeks

1. Introduction; Descriptive Modelling
2. Returns; Autocorrelation; Stationarity
3. ARMA Models
4. Unit Roots; Regressions between Time Series
5. Volatility Modelling
6. Value at Risk
7. Cointegration

Outline

White Noise and Random Walk

- Last week, we encountered two important stochastic processes: the *white noise* process

$$Y_t = U_t,$$

and the *random walk* process

$$Y_t = Y_{t-1} + U_t.$$

- Observe that we can unify the notation by writing

$$Y_t = \phi_1 Y_{t-1} + U_t,$$

where $\phi_1 = 0$ produces white noise, and $\phi_1 = 1$ results in the random walk.

The AR(1) Process

- In fact, there is no reason to restrict the parameter ϕ_1 to just these two values
- Allowing arbitrary values results in the *autoregressive model of order 1*, or *AR(1)* model.
- Usually, an intercept is also addded. The full model is then

$$Y_t = \alpha + \phi_1 Y_{t-1} + U_t, \quad \text{with } Y_t \text{ white noise.}$$

- We will see soon that the model is stationary if and only if $-1 < \phi_1 < 1$.
- The AR(1) process is a member of a very powerful class of models, the *autoregressive-moving average* (*ARMA*) models, with its special cases *autoregressive* (*AR*) and *moving average* (*MA*) models.
- Goal: find the right model (for forecasting etc.) by matching the correlogram.

Mean of Stationary AR(1)

- A first intuition for the *stationarity condition* is obtained if we try to find the (constant) mean and variance of Y_t. The mean of Y_t is to be solved from

$$\mathbb{E}[Y_t] = \alpha + \phi_1 \mathbb{E}[Y_{t-1}] + \mathbb{E}[U_t] = \alpha + \phi_1 \mathbb{E}[Y_t],$$

which implies

$$\mathbb{E}[Y_t] = \frac{\alpha}{1-\phi_1},$$

and this requires $\phi_1 \neq 1$.

Variance of Stationary AR(1)

- Next, because $\{U_t\}$ is white noise, U_t is uncorrelated with Y_{t-1}, so

$$\mathrm{var}(Y_t) = \phi_1^2 \mathrm{var}(Y_{t-1}) + \mathrm{var}(U_t) + 2\phi_1 \mathrm{cov}(Y_{t-1}, U_t) = \phi_1^2 \mathrm{var}(Y_t) + \sigma^2,$$

so that, if and only if $|\phi_1| < 1$,

$$\mathrm{var}(Y_t) = \frac{\sigma^2}{1-\phi_1^2}.$$

- Note that $\mathrm{var}(Y_t) > \mathrm{var}(Y_{t-1})$ if $|\phi_1| \geq 1$, i.e., the variance grows without bounds in that case.

ACF / PACF of Stationary AR(1)

- It can be shown (see optional exercise) that if $|\phi_1| < 1$, then

$$\tau_k = \frac{\gamma_k}{\gamma_0} = \phi_1^k,$$

i.e., if the process is stationary, then the ACF decays exponentially (or *geometrically*).

- The PACF satisfies

$$\tau_{11} = \phi_1 \quad \text{and} \quad \tau_{kk} = 0, k > 1;$$

i.e., it drops to zero after the first lag.

Summary: Moments of AR(1)

- In summary, we have obtained the following properties for the stationary AR(1) process:

Properties of AR(1)

- $\mathbb{E}(Y_t) = \dfrac{\alpha}{1-\phi_1}$;
- $\mathrm{var}(Y_t) = \dfrac{\sigma^2}{1-\phi_1^2}$;
- $\tau_k = \phi_1^k, \quad k = 1, 2, \ldots$;
- $\tau_{kk} = \begin{cases} \phi_1, & k = 1, \\ 0, & k > 1. \end{cases}$

- Hence, if we find a geometrically declining acf, but a pacf which suddenly cuts off when $k > 1$, then we seem to have an AR(1) process.
- Try playing with this in the sheet “AR(1)“ in `simulation.xlsx`.

The Random Walk

The AR(1) process is non-stationary if:

- $\phi_1 = 1$:
  - now $Y_t = \alpha + Y_{t-1} + U_t$: a *random walk with drift*. It satisfies

$$\begin{aligned} \mathbb{E}[Y_t] &= Y_0 + \alpha t, \\ \mathrm{var}(Y_t) &= \sigma^2 t, \\ \mathrm{corr}(Y_t, Y_{t-k}) &= \sqrt{(t-k)/t}. \end{aligned}$$

  - Like for a random walk without drift, the correlogram typically stays close to 1, and decreases slowly, approximately linearly, while the first difference ΔY_t is stationary (see exercises).
- $\phi_1 > 1$: this is a so-called *explosive* process. The mean and variance increase very fast (exponentially). We usually do not consider this because it is unrealistic (at least for long periods of time).

AR(p) Models

- To fit more complicated ACF patterns, the AR(1) model can be extended to the *AR(p)* model if necessary:

$$Y_t = \alpha + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + U_t.$$

- A *necessary* (but not sufficient) *condition for stationarity* is

$$\phi_1 + \phi_2 + \ldots + \phi_p < 1.$$

- If the model is stationary, then the mean is $\mathbb{E}[Y_t] = \alpha/(1 - \sum_{i=1}^p \phi_i)$.
- The ACF should gradually approach zero, but not necessarily with a clear pattern.
- The PACF satisfies

$$\tau_{kk} = 0, \quad k > p.$$

- Because all autoregressive models are basically regressions (with lagged variables as regressors), they can simply be estimated by ordinary least-squares as usual.

Example

Is the AR(2) process

$$Y_t = .5Y_{t-1} + .5Y_{t-2} + U_t$$

stationary?

Outline

The MA(1) Model

- The first-order *moving average model* — MA(1) — is given by

$$Y_t = \alpha + U_t + \theta_1 U_{t-1},$$

where $\{U_t\}$ is again a white noise process.
- This process is stationary for all values of θ_1. It can be shown that

$$\begin{aligned} \mathbb{E}[Y_t] &= \alpha, \\ \mathrm{var}(Y_t) &= \sigma^2(1+\theta_1^2), \\ \tau_1 = \mathrm{corr}(Y_t, Y_{t-1}) &= \theta_1/(1+\theta_1^2). \end{aligned}$$

ACF and PACF of the MA(1) Model

- The most important property is that, for all $k > 1$,

$$\tau_k = \mathrm{corr}(Y_t, Y_{t-k}) = 0,$$

i.e., the ACF has a *cut-off point*; shocks that happened longer than 1 period ago have no effect on Y_t.
- Also, the PACF of an MA(1) process is $\tau_{kk} = -(-\theta_1)^k \to$ *geometric decay*.
- Note that the patterns of the ACF and PACF are *reversed* compared to the AR(1) process!
- We will use this to *identify* the correct model from the SACF/SPACF.

MA(q) process

- Again we may generalize to the MA(q) model

$$Y_t = \alpha + U_t + \theta_1 U_{t-1} + \ldots + \theta_q U_{t-q}.$$

- Now $\tau_k = 0$ for $k > q$, and τ_{kk} still decays exponentially.
- This will again be used to identify the correct model from the correlogram.

Some more properties of MA(q) Models

- *Estimation*: since the past values of U_t are unobserved, we cannot estimate the coefficient θ_1 by OLS. Eviews has a built-in method, based on *non-linear least-squares*.
- *Stationarity*: All MA(q) processes are stationary, regardless of their parameter values.

ARMA(1, 1) Process

- The most general class of processes is the mixed autoregressive-moving average (*ARMA*) class of models.
- The simplest one is the ARMA(1,1):

$$Y_t = \phi_1 Y_{t-1} + U_t + \theta_1 U_{t-1}.$$

- The model is stationary if $|\phi_1| < 1$.
- The ACF turns out to be

$$\begin{aligned} \tau_1 &= \frac{(1+\phi_1\theta_1)(\phi_1+\theta_1)}{1+\theta_1^2+2\phi_1\theta_1}, \\ \tau_k &= \phi_1^{k-1}\tau_1, \quad k > 1 \rightarrow \textit{geometric decay}. \end{aligned}$$

- The PACF is also gradually declining (no cut-off point).

ARMA(p, q) Process

- The general ARMA(p, q) model is

$$Y_t = \alpha + \phi_1 Y_{t-1} + \ldots + \phi_p Y_{t-p} U_t + \theta_1 U_{t-1} + \ldots + \theta_q U_{t-q}.$$

- Like for the ARMA(1, 1), both its ACF and PACF decay exponentially (provided it is stationary), without a clear pattern.
- *Estimation*: Like pure MA models, ARMA models cannot be estimated by OLS because the past values of U_t are unobserved. Eviews has built-in methods however.
- *Stationarity*: The stationarity or otherwise of an ARMA process depends only on the AR component, i.e., a necessary condition is that

$$\phi_1 + \phi_2 + \ldots + \phi_p < 1.$$

A Theorem

- The following theorem explains why ARMA processes are so important:

Theorem

Any stationary process can be represented as an ARMA process.

- So as long as we can find the right model, we can predict *any* stationary process[1]!
- The properties of the ACF and PACF, summarized on the next slide, will help us do this.

[1] And even integrated processes, by taking the first difference and modelling that.

Summary of Properties

- We have found the following properties of the different processes:
  - AR(p): geometrically decaying ACF, PACF is zero after p lags;
  - MA(q): ACF is zero after q lags, geometrically decaying PACF;
  - ARMA(p, q): geometrically decaying ACF and PACF.
- So we can infer the model type and order from the SACF/SPACF for pure AR and MA models.
- A full ARMA model would be required if both SACF and PACF decline geometrically, but we won't be able to infer the orders then.
- The usual procedure is to try an ARMA(1,1) in that case, and test whether the model needs to be extended.
- The above is called the *Box-Jenkins* approach.

Outline

Box-Jenkins

- In their 1970 textbook, Box and Jenkins proposed an approach to empirical ARMA modelling that soon became and still is the predominant approach to univariate time-series analysis and forecasting. Their procedure consists of three steps:
  1. **Identification**. This refers to the problem of selecting an initial ARMA model, i.e., the choice of the orders p and q. This is based on inspection of the graph and the correlogram of Y_t. Also, so-called *information criteria* may be used for model selection.
  2. **Estimation**. The unknown autoregressive and moving average parameters, as well as the variance σ^2 of the disturbances, need to be estimated.
  3. **Diagnostic checking**. A correctly specified model should not display any autocorrelation in the residuals. Therefore, the main model check is a test for residual autocorrelation. Also other misspecification tests (heteroskedasticity, normality, structural change) may be used.
- If we find some problem with the model in Step 3, then we return to Step 1 and go through the cycle again, until the tests indicate no further problem.

Identification: Stationarity

- Before even choosing p and q, it must be ensured that the data are *stationary*. An integrated time series displays very large autocorrelations, which converge to zero only slowly. In contrast, stationary time series have an autocorrelation function that decays exponentially, or is not significantly different from zero after a few lags.
- Additional information is available from the graph of the time series. Stationary time series should display *mean-reversion*, i.e., they should fluctuate around a constant mean (or a linear trend, in case of trend-stationarity). If a series does not display this property, and behaves more like a random walk, then it may not be stationary.
- A formal procedure to test this (and how to proceed in case of non-stationarity) will be introduced later.

Identification: p and q

- The second step is choosing p and q.
- Recall that the correlograms of AR, MA and ARMA processes are characterized thusly:
  - AR(p): geometrically decaying ACF, PACF is zero after p lags;
  - MA(q): ACF is zero after q lags, geometrically decaying PACF;
  - ARMA(p, q): geometrically decaying ACF and PACF.
- If neither ACF nor PACF have a clear cut-off point, start with an ARMA(1,1), estimate it, and test whether the model needs to be extended.
- General goal: *parsimony*. Find the smallest possible model that describes the data well.

Estimation

- Eviews has built-in routines for estimating ARMA models.
- The specification of, e.g., an ARMA(3, 2) model for a time series $\{y_t\}$ is (see exercises)

  Y c ar(1) ar(2) ar(3) ma(1) ma(2)

- Note: for pure AR models, we may also use (e.g., for AR(2)):

  Y c Y(-1) Y(-2)

- The difference lies in the interpretation of the constant term: Somewhat confusingly, when using the "`ar`" specification, then the estimated coefficient `c` corresponds to $\mathbb{E}(Y_t) = \alpha/(1 - \sum_{i=1}^{p} \phi_i)$ rather than α.
- Yet in practice the "`ar`" specification is preferred, because only in that case does EViews compute the residual Q-tests appropriately.

Diagnostic Testing

- After estimating our favorite ARMA model, we obtain the *residuals* $\hat{u}_t$. These residuals should look like *white noise*.
- If there is significant autocorrelation left in $\hat{u}_t$, we should *extend* the model. E.g.,, if the residuals of an AR(1) model look like an MA(1) process, then we might try an ARMA(1,1) model instead.
- The most often used test for residual autocorrelation in ARMA models is the Ljung-Box Q-statistic, based on the residuals $\hat{u}_t$ instead of the original time series y_t.
- It can be shown that if $\hat{u}_t$ is a residual from an ARMA(p, q) model, then the Q-statistic with m correlations has an approximate χ^2_{m-p-q} distribution under the null hypothesis (for m large enough). This explains why in EViews, the first $(p+q)$ Q-statistics have no p-value.
- The tests cannot reliably be used to find out in which direction the model should be *extended*. This means that we may have to try different alternative specifications before we find a satisfactory choice.

Model Selection Criteria

- If more than one specification passes the diagnostic tests (e.g., both an AR(2) and an ARMA(1,1)), then the decision is often based on the *Akaike information criterion* (AIC) and the *Schwarz criterion* (SC; a.k.a. the *Bayesian* information criterion, BIC):

$$
\begin{aligned}
AIC &= 1 + \log 2\pi + \log\left(\frac{1}{n}\sum_{t=1}^{n} \hat{u}_t^2\right) + \frac{2k}{T}, \\
SC &= 1 + \log 2\pi + \log\left(\frac{1}{n}\sum_{t=1}^{n} \hat{u}_t^2\right) + \frac{k \log T}{T}.
\end{aligned}
$$

- $k =$ number of parameters ($p + q + 2$ including intercept and σ_u^2).
- Constant $1 + \log 2\pi$ is irrelevant in comparisons and sometimes deleted.
- We choose the model with the smallest AIC or SC. Typically AIC leads to higher p and q than SC.
- The idea is similar to the adjusted R^2: the second term is a penalty for including too many parameters. Trade-off between *goodness of fit* and *parsimony*. Smaller models are often better for forecasting.

Outline

Forecasting Terminology

- The main purpose of ARIMA models is forecasting.
- Denote the forecast of y_{t+s} based on information available at time t as $f_{t,s}$.
- We distinguish *in-sample* and *out-of-sample* forecasts.
- In-sample means that the same data are forecasted that were used for estimating the model.
- Out-of sample forecasting means that the sample is divided into two subsamples, say $1, \dots, T_1$ (the *estimation period*) and $T_1 + 1, \dots, T$ (the *holdout period*). The former is used for selecting a model and estimating its parameters. Based on this information at time T_1, we forecast $y_{T_1+1}, \dots, y_T$.

One-Step Ahead Forecasts

- Stationary implies that for, e.g., stationary ARMA$(1, 1)$ model,

$$Y_{t+1} = \alpha + \phi_1 Y_t + U_{t+1} + \theta_1 U_t.$$

- Strategy: replace parameters with estimates, errors U_t with residuals $\hat{u}_t$, and unobserved future U_t with zero. Hence

$$f_{t,1} = \hat{\alpha} + \widehat{\phi_1} y_t + 0 + \theta_1 \hat{u}_t.$$

- Analogous procedure for general ARMA(p, q) models.

Multi-Step Forecasts

- When forecasting multiple steps ahead, not only the future U_t are unknown, but also the future y_t.
- Solution: recursive procedure. Forecast y_{t+1} first, then use this to forecast y_{t+2}, etc. All future errors are replaced with zero.

Remarks

- EViews distinguishes between *static* and *dynamic* forecasts. This is essentially the same as the difference between one-step and multi-step forecasts: “static” uses information up to time $t-1$ to forecast y_t, for any $t > T_1$; this is only possible for in-sample forecasting. “Dynamic”uses only the information up to T (or T_1) to forecast $T+s$ (or T_1+s).

Outline

Learning Goals

Students are able to

- understand the essential characteristics of AR, MA, and ARMA processes;
- identify the correct model type from the correlogram;
- estimate ARMA models with EViews;
- apply diagnostic tests to evaluate a fitted model;
- choose between competing models using information criteria, and
- use the final model for forecasting.

Homework

- Exercise 3
- Questions 3, 6–9, 10a, 11a, 11e, 12 from Chapter 6 of Brooks (2019)

Module 9.3: Time Series Analysis
Fall Term 2022

Week 4:

Unit Roots; Regressions between Time Series

Lucerne University of Applied Sciences and Arts

Outline in Weeks

1. Introduction; Descriptive Modelling
2. Returns; Autocorrelation; Stationarity
3. ARMA Models
4. Unit Roots; Regressions between Time Series
5. Volatility Modelling
6. Value at Risk
7. Cointegration

Outline

Unit Root Testing

- Recall that if a nonstationary series, y_t must be differenced d times before it becomes stationary, then it is said to be *integrated* of order d. Notation: $\mathrm{I}(d)$. When we simply say 'integrated' we mean $\mathrm{I}(1)$. Another synonym is 'the series has a *unit root*'.
- So far, our decision to take differences was based on the correlogram: if autocorrelations decay slowly and approximately linearly, then the series may be integrated and must be differenced.
- This procedure is subjective and unreliable: a *trend-stationary* series, such as $Y_t = \beta_0 + \beta_1 t + U_t$, will also have large and slowly decaying autocorrelations (see `simulation.xlsx`).
- Therefore, it is useful to have a formal testing procedure to distinguish integrated from (trend-) stationary time series. Such tests are called *unit root tests*. The *(Augmented) Dickey-Fuller* test is the most common.

Example: the AR(1) Model

- The simplest example of stationary and integrated time series is the AR(1) model

$$Y_t = \phi_1 Y_{t-1} + U_t,$$

where Y_0 is a constant.
- If $-1 < \phi_1 < 1$, then Y_t is stationary, with mean 0 and variance $\sigma^2/(1-\phi^2)$.
- If $\phi_1 = 1$, then the model becomes a *random walk*, $Y_t = Y_{t-1} + U_t$, with mean Y_0 and variance $\sigma^2 t$.

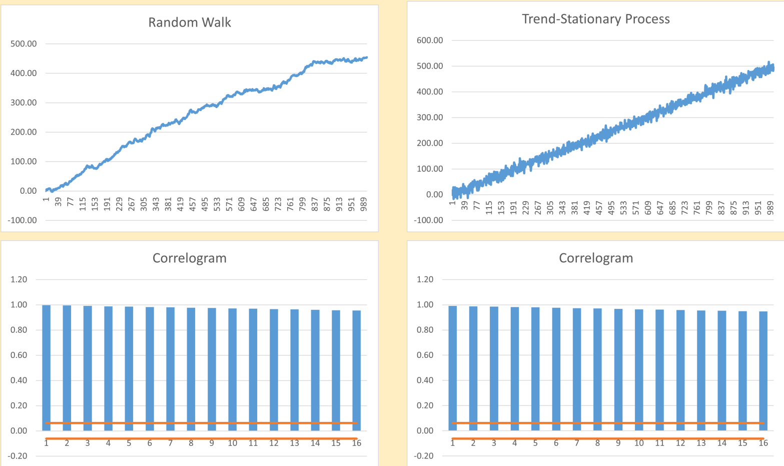
Stochastic vs. Deterministic Trends

- Consider the two models

$$Y_{1,t} = \delta t + U_{1,t} \quad \text{and}$$
$$Y_{2,t} = \delta + Y_{2,t-1} + U_{2,t}.$$

- For both models, $\mathbb{E}[\Delta Y_t] = \delta$, i.e., both series trend upwards by δ each period.
- $Y_{1,t}$ contains a *deterministic trend*: $Y_{1,t} - \delta t = U_{1,t}$ is stationary. In practice, regress y_t on a linear trend like in week 1; the residuals will be stationary.
- $Y_{2,t}$ contains a *stochastic trend*: $Y_{2,t} - \delta t$ is a random walk. It becomes stationary only by differencing, i.e., it is $I(1)$.

Random walk (left) vs. trend stationary process (right).



Why do we care?

1. Inference: OLS estimators have non-standard distribution, so that standard inference is invalid.
2. Forecasting:
   - stationary series display *mean-reversion*; deviations from the mean are corrected in the next period. Shocks U_t have a *transitory*, decreasing effect on future Y_{t+k} (in AR(1) model, the effect is $\phi^k \to 0$);
   - For integrated time series (of order 1): no mean-reversion, shocks U_t have a *persistent* effect on future Y_{t+k}.
3. Spurious regression: regressing two drifting I(1) variables onto each other will spuriously result in significant estimates, because they happen to trend in the same direction.

The Dickey-Fuller Test

- Consider again the AR(1) model
$$Y_t = \phi_1 Y_{t-1} + U_t.$$
- We wish to test the null hypothesis $H_0 : \phi_1 = 1$, against the *one-sided* alternative hypothesis $H_1 : \phi_1 < 1$.
- Under the null, the process is a random walk, and hence integrated. Under the alternative, it is stationary (if $\phi_1 > -1$, which we assume). Therefore, we will be testing $H_0 : Y_t \sim I(1)$ against $H_1 : Y_t \sim I(0)$.
- The *Dickey-Fuller test* is based on the t-statistic for $\phi_1 = 1$ in the AR(1) model, which may be reformulated as the t-statistic for $\psi = 0$ in
$$\Delta y_t = \psi y_{t-1} + u_t,$$
where $\psi := \phi_1 - 1$; note that $\psi < 0$ under the alternative.

The Dickey-Fuller Distribution

- We reject H_0 if the t-statistic is less than the (negative) critical value.
- In classical regressions, the 5% critical value for a one-sided t-test would be -1.645. However, because the regressor Y_{t-1} is non-stationary under the null, a different distribution arises, and the appropriate 5% critical value is $= -1.95$.
- This critical value changes to $= -2.86$ if we add a constant term to the regression:
$$\Delta y_t = \alpha + \psi y_{t-1} + u_t.$$
This is the relevant test if we want to allow for a non-zero mean $\mathbb{E}[Y_t]$ under the alternative.
- If we want to test a random walk *with drift* against a *trend-stationary* alternative, then the relevant regression is
$$\Delta y_t = \alpha + \delta t + \psi y_{t-1} + u_t,$$
and the 5% critical value is $= -3.41$.

Choice of Model

- The difference between the three tests is whether or not a constant and a time trend are included in the test regression.
- In practice, the test without constant and trend is almost never applicable.
- The test with an intercept in the regression test is relevant for series such as interest rates and real exchange rates, where we do not expect a linear trend under either the null or the alternative hypothesis.
- Many other economic and financial time series, such as GDP or (log-) asset prices, clearly display an upward trend, in which case a trend should be included.

The Augmented Dickey-Fuller Test

- The *augmented* Dickey-Fuller (ADF) test is an extension of the procedure to the AR(p) model
$$Y_t = \phi_1 Y_{t-1} + \ldots + \phi_p Y_{t-p} + U_t.$$
- This process is integrated under $H_0 : \phi_1 + \ldots + \phi_p = 1$, and stationary under the alternative hypothesis $H_1 : \phi_1 + \ldots + \phi_p < 1$.
- It can be shown (see exercises) that this is equivalent to testing $H_0 : \psi = 0$ against $H_1 : \psi < 0$ in the regression
$$\Delta y_t = \psi y_{t-1} + \alpha_1 \Delta y_{t-1} + \ldots + \alpha_{p^*} \Delta y_{t-p^*} + u_t,$$
where $\psi = \sum_{i=1}^{p} \phi_i - 1$, and $p^* = p - 1$.
- The interpretation of the null and alternative hypothesis, the role of the constant and trend, and the critical values are the same as in the first-order model.

Choosing p

- In practice, p is, of course, *unknown*.
- The number of lags in the test regression must be chosen large enough, such that the residuals have no autocorrelation, but not too large, because this would decrease test power.
- Often the choice is made based on *model selection criteria* (AIC, BIC). EViews has a built-in option to do this automatically.
- The AIC is preferred in this case, because it tends to include more lags than the BIC. This is preferred because our goal in this case is not to find the "correct" model, but to combat autocorrelation in the test regression.
- So effectively, the ADF test is just the DF test (regress Δy_t on y_{t-1}), but with enough lags of ΔY_t thrown in to remove any autocorrelation.

Outline

ARIMA Models

- As discussed last week, the first step in the Box-Jenkins procedure is to make sure that the data are stationary.
- This is usually decided by the ADF test.
- If the test doesn't reject, then one models the first difference ΔY_t as an ARMA process.
- The model for the levels Y_t is then called an *ARIMA(p,d,q)* model: the data are differenced d times, and the result modeled as an ARMA(p,q) process. Usually, $d = 1$.
- To forecast an ARIMA process (with d=1), first predict ΔY_{t+1}, and then let $\hat{Y}_{t+1} = Y_t + \widehat{\Delta Y_{t+1}}$.
- Longer horizon forecast are obtained recursively as usual.

Outline

Regression with Time Series

Some considerations when analyzing the relation between the time series $\{y_t\}$ and $\{x_t\}$ in the regression model

$$y_t = \alpha + \beta' x_t + u_t :$$

- Model is only reasonable if Y_t and X_t are integrated of same order. If both are $I(1)$, then in general the error will be $I(1)$, except in case of *cointegration* (stationary linear combination of $I(1)$ series).
- In case of non-cointegrated $I(1)$ series, consider regression in first differences: $\Delta y_t = \alpha + \beta \Delta x_t + e_t$.
- Even when non-stationary is not a concern, autocorrelation usually is (most economic/financial series are autocorrelated).
- We will discuss two tests for autocorrelation in a regression: the *Breusch-Godfrey* test and the *Durbin-Watson* test.

Breusch-Godfrey Test

- The *Breusch-Godfrey LM test* is done in four steps:
  1. Estimate the regression, e.g.,
  $$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + u_t,$$
  by OLS, and get the residuals $\widehat{u}_t$.
  2. Estimate the auxiliary regression
  $$\widehat{u}_t = \gamma_1 + \gamma_2 x_{2t} + \gamma_3 x_{3t} + \tau_1 \widehat{u}_{t-1} + \ldots + \tau_r \widehat{u}_{t-r} + v_t.$$
  Denote by R^2_{aux} the R^2 of this regression.
  3. Test the null hypothesis of *no residual autocorrelation*
  $$H_0 : \tau_1 = \ldots = \tau_r = 0,$$
  by the usual F-test, or by the LM test:
  $$T \cdot R^2_{aux} \sim \chi^2(r)$$
  4. The test rejects for large values of $T \cdot R^2_{aux}$.

Remarks

- How do we choose r? Consult the correlogram, and take account of the frequency of the data, e.g., $r = 4$ or 5 for quarterly data.
- This is a general autocorrelation test. It has power against any degree of possible autocorrelation.
- EViews: after estimation, click `View` → `Residual Diagnostics` → `Serial Correlation LM test`

The Durbin-Watson Test

- The *Durbin Watson* test is a test of first-order autocorrelation only.
- This test is performed as follows:
  1. Get the OLS residuals $\widehat{u}_t$ from the regression as before.
  2. Compute the statistic
  $$DW = \frac{\sum_{t=2}^{n} (\widehat{u}_t - \widehat{u}_{t-1})^2}{\sum_{t=2}^{n} \widehat{u}_t^2} \approx 2(1 - \widehat{\tau}).$$
  where $\widehat{\tau}$ is OLS estimator of regression of $\widehat{u}_t$ on $\widehat{u}_{t-1}$. Since $-1 \leq \widehat{\tau} \leq 1$, it follows that $0 \leq DW \leq 4$. If there is no autocorrelation, we expect $DW \approx 2$.
  3. Look up the critical values for the test in a table. The test has a non-standard distribution. There are five different regions that lead to different conclusions (see next slide).
- The DW test is included in EViews estimation output by default.

Test Decision

- The critical values depend on the regressors. However, one can obtain bounds on the critical values.
- The table gives the DW lower (d_l) and upper (d_u) bounds, at the 5% level, for different values of n and k. Test decision:

$$\begin{array}{rcccrcl} 0 & < & DW & < & d_l & \Rightarrow & \text{reject (positive AC)} \\ d_l & < & DW & < & d_u & \Rightarrow & \text{inconclusive} \\ d_u & < & DW & < & 4-d_u & \Rightarrow & \text{do not reject} \\ 4-d_u & < & DW & < & 4-d_l & \Rightarrow & \text{inconclusive} \\ 4-d_l & < & DW & < & 4 & \Rightarrow & \text{reject (negative AC)} \end{array}$$

Critical Values of the DW Test

	$k=2$		$k=3$		$k=4$		$k=6$		$k=10$	
n	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
15	1.08	1.36	0.95	1.54	0.81	1.75	0.56	2.22	0.17	3.22
16	1.11	1.37	0.98	1.54	0.86	1.73	0.61	2.16	0.22	3.09
17	1.13	1.38	1.01	1.54	0.90	1.71	0.66	2.10	0.27	2.97
18	1.16	1.39	1.05	1.53	0.93	1.70	0.71	2.06	0.32	2.87
19	1.18	1.40	1.07	1.54	0.97	1.68	0.75	2.02	0.37	2.78
20	1.20	1.41	1.10	1.54	1.00	1.68	0.79	1.99	0.42	2.70
22	1.24	1.43	1.15	1.54	1.05	1.66	0.86	1.94	0.50	2.57
24	1.27	1.45	1.19	1.55	1.10	1.66	0.92	1.90	0.58	2.46
26	1.30	1.46	1.22	1.55	1.14	1.65	0.98	1.87	0.66	2.38
28	1.33	1.48	1.25	1.56	1.18	1.65	1.03	1.85	0.72	2.31
30	1.35	1.49	1.28	1.57	1.21	1.65	1.07	1.83	0.78	2.25
35	1.40	1.52	1.34	1.58	1.28	1.65	1.16	1.80	0.91	2.14
40	1.44	1.54	1.39	1.60	1.34	1.66	1.23	1.79	1.01	2.07
50	1.50	1.58	1.46	1.63	1.42	1.67	1.33	1.77	1.16	1.99
75	1.58	1.65	1.57	1.68	1.54	1.71	1.45	1.77	1.37	1.90
100	1.65	1.69	1.63	1.71	1.60	1.74	1.57	1.78	1.48	1.87
200	1.76	1.78	1.75	1.79	1.74	1.80	1.72	1.82	1.67	1.86

$\alpha = 5\%$, $k =$ number of regressors (including intercept).

Consequences of Autocorrelation

- Only in a truly static model, the OLS estimator remains unbiased.
- In general, $\widehat{\beta}$ will be biased.
- OLS is no longer BLUE even when it is unbiased. More efficient estimators are available (GLS).
- The OLS standard errors are wrong.
- Inference based on usual t- and F-tests is unreliable.

Solutions

- One should at least use Newey-West *heteroskedasticity and autocorrelation consistent* (HAC) standard errors (in Eviews: Click on the `Options` tab when entering the regression equation)[1].
- The more modern approach is to model the autocorrelation explicitly, by including lags of y_t ($\rightarrow$ *ARX* model) and possibly x_t ($\rightarrow$ *ADL*[2] model). MA terms can also be added ($\rightarrow$*ARMAX* model).

[1] EViews also provides *White standard errors*. These are consistent under heteroskedasticity only, not under autocorrelation.

[2] autoregressive distributed lag

Example

- Consider the ADL model

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \beta_0 X_t + \beta_1 X_{t-1} + u_t. \qquad (1)$$

- The *long-run equilibrium* is the *steady state* of the system (Y_t, X_t) when all future shocks $U_{t+1}, U_{t+2}, \ldots$ are set to zero. If the system is stationary, then

$$Y_t \to Y^* \text{ and } X_t \to X^*.$$

So the long-run equilibrium can be found from the equation

$$Y^* = \alpha_0 + \alpha_1 Y^* + \beta_0 X^* + \beta_1 X^* \Rightarrow$$
$$Y^* = \frac{\alpha_0}{1-\alpha_1} + \frac{\beta_0+\beta_1}{1-\alpha_1} X^*. \qquad (2)$$

Why is a dynamic model needed?

A dynamic model can be necessary for the following reasons:

- Inertia of the dependent variable. E.g., adjustment costs, habits in individual decisions, or market frictions.
- Trending behaviour. E.g., if returns are unpredictable, prices must be trending. Hence, prices are autocorrelated.
- Seasonality; see Lecture 1.
- Over-reactions. Typically the result of some market failure, but think of exchange rate overshooting.

Residual Autocorrelation

- Now assume that instead of the dynamic model (1), the steady state equation (2) is estimated:

$$y_t = a_0 + b_0 x_t + u_t$$

- This is known as *dynamic misspecification*. The effect of the lagged variables is then in the error term. Result: autocorrelated residuals.
- In this case, not only the standard errors will be wrong, but the estimates will be biased: $\hat{b}_0$ will be an estimate of $\frac{\beta_0+\beta_1}{1-\alpha_1}$ in (2), not of β_0 in (1).
- The modern view is that autocorrelation is not a problem, but an *opportunity to improve the model*.
- Thus, instead of correcting the standard errors, one should estimate a dynamic model directly.
- Start with a dynamic model (include lags of Y_t and X_t in the model). Test for autocorrelation, and for the significance of dynamics.

Outline

Learning Goals

Students

- understand the ADF test and can use it to test for unit roots,
- are able to forecast ARIMA models,
- understand the difference between static and dynamic models, and
- know how to test for autocorrelation in a regression

Homework

- Exercise 4
- Questions 2 and 3 from Chapter 8 of Brooks (2019)

Module 9.3: Time Series Analysis
Fall Term 2022

Week 5:

Volatility Modelling

HSLU Lucerne University of Applied Sciences and Arts

Outline in Weeks

1. Introduction; Descriptive Modelling
2. Returns; Autocorrelation; Stationarity
3. ARMA Models
4. Unit Roots; Regressions between Time Series
5. Volatility Modelling
6. Value at Risk
7. Cointegration

Outline

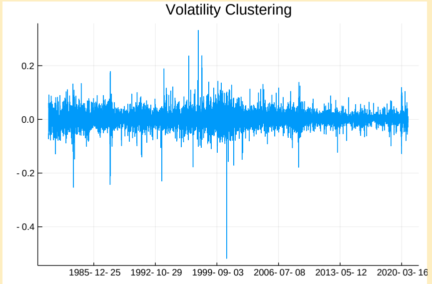
Goal

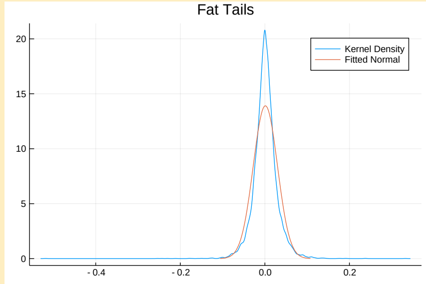
- Recall these stylized facts about asset returns:
  1. Lack of autocorrelation (efficient market hypothesis)
  2. Volatility clustering
  3. Distribution has heavy tails
  4. Leverage effects
- Goal today: model the last 3 of these, starting with the volatility clustering.

Volatility

- The *volatility* of an investment is a measure of its *risk*. Usually defined as the standard deviation of the return on the investment.
- Volatility is an important ingredient in:
  - portfolio selection;
  - risk management;
  - option pricing.
- Daily financial returns display *volatility clustering*: periods of high volatility alternate with more tranquil periods.
- In other words: large (in absolute value) returns tend to be followed by large (in absolute value) returns.
- This forms the basis for the *autoregressive-conditional heteroskedasticity* model (ARCH; Engle, 1982) and the *generalized ARCH* model (GARCH; Bollerslev, 1986).

Example: Daily Returns on Apple Stock





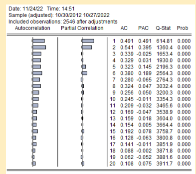
Reminder: Parameters vs. sample values

- We usually write σ for the standard deviation of, e.g., a normally distributed variable.
- σ is a *parameter* and therefore unknown.
- The best we can hope for is to *estimate* it, usually with the *sample standard deviation* s.
- With stock returns, the standard deviation (or *volatility*) changes over time, due to *volatility clustering*.
- We write σ_t for the volatility in period t.
- Note that σ_t is *unobserved*. The best we can do is *estimate* it. We'll write $\widehat{\sigma}_t$ for this estimate.
- Today, we'll mostly discuss different methods of estimation.

Detecting Volatility Clustering (I)

- Since volatility clustering means that large returns tend to be followed by large returns, it is possible to detect it by inspecting the correlogram of the squared returns.

Example: correlogram of squared S&P500 returns.



Date: 11/24/22 Time: 14:51
Sample (adjusted): 10/30/2012 10/27/2022
Included observations: 2546 after adjustments

	AC	PAC	Q-Stat	Prob
1	0.491	0.491	614.81	0.000
2	0.541	0.395	1360.4	0.000
3	0.339	-0.025	1653.4	0.000
4	0.329	0.031	1930.0	0.000
5	0.323	0.145	2196.3	0.000
6	0.380	0.189	2564.3	0.000
7	0.280	-0.065	2764.3	0.000
8	0.324	0.047	3032.4	0.000
9	0.256	0.050	3200.3	0.000
10	0.245	-0.011	3354.3	0.000
11	0.209	-0.032	3465.6	0.000
12	0.169	-0.047	3538.9	0.000
13	0.159	0.018	3604.0	0.000
14	0.154	0.005	3664.4	0.000
15	0.192	0.078	3758.7	0.000
16	0.128	-0.063	3800.8	0.000
17	0.141	-0.011	3851.9	0.000
18	0.088	-0.002	3871.8	0.000
19	0.062	-0.052	3881.6	0.000
20	0.108	0.075	3911.7	0.000

- Clearly, there is a lot of predictability in squared returns (unlike returns themselves).

Detecting Volatility Clustering (II)

- Besides relying on the Q-tests from the correlogram, another formal test is Engle's *ARCH-LM* test (essentially a Breusch-Godfrey test applied to the squared residuals).
- EViews only offers it for residuals, not for a series itself. Hence, we start by regressing the returns on an intercept.
- The ARCH-LM test is based on the auxiliary regression

$$\hat{u}_t^2 = \gamma_0 + \gamma_1 \hat{u}_{t-1}^2 + \ldots + \gamma_m \hat{u}_{t-m}^2 + e_t.$$

- The lag length m is chosen by the user, e.g., 5 for daily data.
- The test statistic is $T \cdot R^2_{aux}$ and has a $\chi^2(m)$ distribution under $H_0 : \gamma_1 = \cdots = \gamma_m = 0$ (no volatility clustering).

Example: ARCH-LM test for the S&P500

Heteroskedasticity Test: ARCH

F-statistic	302.2045	Prob. F(5,2535)	0.0000
Obs*R-squared	948.9587	Prob. Chi-Square(5)	0.0000

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 11/24/22 Time: 15:11
Sample (adjusted): 11/06/2012 10/27/2022
Included observations: 2541 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3.14E-05	8.68E-06	3.614959	0.0003
RESID^2(-1)	0.299355	0.019651	15.23388	0.0000
RESID^2(-2)	0.396998	0.020528	19.33915	0.0000
RESID^2(-3)	-0.087729	0.021921	-4.001999	0.0001
RESID^2(-4)	-0.015277	0.020529	-0.744168	0.4568
RESID^2(-5)	0.145365	0.019653	7.396390	0.0000

R-squared	0.373459	Mean dependent var	0.000120
Adjusted R-squared	0.372223	S.D. dependent var	0.000528
S.E. of regression	0.000419	Akaike info criterion	-12.71671
Sum squared resid	0.000444	Schwarz criterion	-12.70292
Log likelihood	16162.59	Hannan-Quinn criter.	-12.71171
F-statistic	302.2045	Durbin-Watson stat	2.054495
Prob(F-statistic)	0.000000		

The null of no volatility clustering is clearly rejected (p-value is zero, $T \cdot R^2_{aux} = 948.96$ much larger than critical value 11.07).

Outline

Historical Volatility

- A first simple estimator is *historical volatility*, i.e., the sample standard deviation of the most recent m observations (often $m = 250$, one year).
- If $r_t = \ln P_t - \ln P_{t-1}$ denotes the daily log-return, then

$$\hat{\sigma}^2_{t+1,HIST} = \frac{1}{m} \sum_{j=0}^{m-1} r^2_{t-j}.$$

(Typically the average return is relatively close to zero). This is an estimate of the squared volatility over day $t + 1$, made at the end of day t.

- Main disadvantages:
  - either noisy (small m), or reacts slowly to new information (large m);
  - "ghosting" feature: large shock leads to higher volatility for exactly m periods, then drops out.

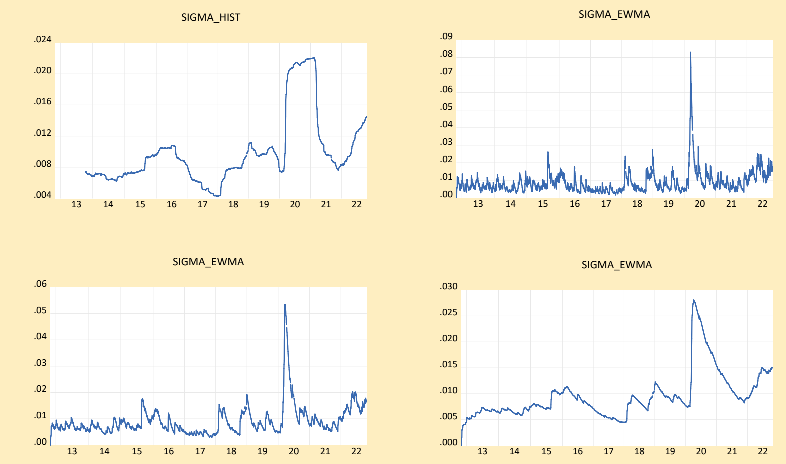
RiskMetrics

- Problems with historical volatility are addressed by replacing equally weighted moving average by an *exponentially* weighted moving average (EWMA), also used in JPMorgan's *RiskMetrics* system:

$$\begin{aligned}\hat{\sigma}^2_{t+1,EWMA} &= (1-\lambda)\sum_{j=0}^{\infty}\lambda^j r^2_{t-j}\\ &= \lambda\hat{\sigma}^2_{t,EWMA} + (1-\lambda)r_t^2, \qquad 0<\lambda<1.\end{aligned}$$

- This means that observations further in the past get a smaller weight.
- In practice we do not have $r_{t-\infty}$, but the second equation can be started up by an initial estimate / guess $\sigma^2_{0,EWMA}$.
- The larger λ, the stronger the persistence of shocks (large returns).
- For daily data, RiskMetrics recommends $\lambda = 0.94$.

Example: S&P500 volatility, historical and EWMA ($\lambda = 0.8, 0.94, 0.99$)



Outline

The ARCH Model

- The first-order *autoregressive-conditional heteroskedasticity* (ARCH(1)) model, due to Engle (1982), for a return r_t with mean zero is

$$\sigma^2_{t+1} = \omega + \alpha r_t^2.$$

- In practice, we need to allow for $\mathbb{E}[r_{t+1}] = \mu_{t+1} \neq 0$. Then $r_{t+1} = \mu_{t+1} + u_{t+1}$, and the model becomes

$$\sigma^2_{t+1} = \omega + \alpha u_t^2.$$

The ARCH Model

- When trying to estimate ARCH models one might find that more lags are needed, leading to ARCH(q):
$$\sigma_{t+1}^2 = \omega + \alpha_1 u_t^2 + \ldots + \alpha_q u_{t-q+1}^2.$$
- *Note*: Variances must be positive, therefore we need to impose $\omega > 0$, $\alpha_i \geq 0$, $i = 1, \ldots, q$.
- It can be shown that an ARCH(q) models corresponds to an AR(q) for the squared returns. Thus, we could determine the order from the correlogram of the squared returns: SPACF should cut off after q lags.
- In the example above, we might conclude that we need an ARCH(6) model.

The GARCH Model

- A simpler structure than ARCH(q) is an ARMA(1,1) for r_t^2 or u_t^2, which leads to the *generalized ARCH* model of orders (1,1) (GARCH(1,1)), due to Bollerslev (1986):
$$\sigma_{t+1}^2 = \omega + \alpha u_t^2 + \beta \sigma_t^2, \qquad \omega > 0, \alpha \geq 0, \beta \geq 0.$$
- *Advantage*: Flexible structure with only 3 parameters to estimate.

The GARCH Model

- The GARCH(1,1) model is stationary if the unconditional ("average") variance $\sigma^2 = \mathbb{E}[\sigma_t^2]$ is positive, constant and finite.
- This requires
$$\begin{aligned} \sigma^2 = \mathbb{E}[\sigma_{t+1}^2] &= \omega + \alpha \mathbb{E}[u_t^2] + \beta \mathbb{E}[\sigma_t^2] \\ &= \omega + \alpha \sigma^2 + \beta \sigma^2. \end{aligned}$$
- Hence, provided that $\alpha + \beta < 1$ (the *stationarity condition*),
$$\sigma^2 = \frac{\omega}{1 - \alpha - \beta}.$$
- The nonstationary model with $\alpha + \beta = 1$ is called *integrated GARCH* (IGARCH): infinite variance, no mean-reversion in volatility.
- Notice that an IGARCH with $r_t = u_t$, $\omega = 0$, $\beta = \lambda$, and $\alpha = (1 - \lambda)$ is just the RiskMetrics model.

The GARCH Model

Some other properties:

- The ACF and PACF of r_t^2 in case of stationary GARCH(1,1) are both exponentially decaying, no cut-off point.
- The *standardized returns*
$$z_{t+1} = \frac{r_{t+1} - \mu_{t+1}}{\sigma_{t+1}}$$
satisfy $E(z_{t+1}) = 0$ and $\mathrm{var}(z_{t+1}) = 1$. Therefore the model may be formulated as
$$\begin{aligned} r_{t+1} &= \mu_{t+1} + u_{t+1} = \mu_{t+1} + \sigma_{t+1} z_{t+1}, \\ \sigma_{t+1}^2 &= \omega + \alpha u_t^2 + \beta \sigma_t^2. \end{aligned}$$
- Often it is assumed that z_t are i.i.d. as $N(0,1)$.
- Even if $z_t \sim N(0,1)$, it can be shown that varying σ_t implies that r_t has non-normal distribution, with higher kurtosis.

The GARCH(p, q) Model

- The GARCH(1, 1) model can be extended to the GARCH(p, q) model

$$\sigma_{t+1}^2 = \omega + \alpha_1 u_t^2 + \cdots + \alpha_q u_{t-q+1}^2 + \beta_1 \sigma_t^2 + \cdots + \beta_p \sigma_{t-p+1}^2$$

  although in practice, this is rarely necessary.
- The model is stationary if $\sum_{i=1}^{p} \beta_i + \sum_{i=1}^{q} \alpha_i < 1$, and the unconditional variance is

$$\frac{\omega}{1 - \sum_{i=1}^{p} \beta_i - \sum_{i=1}^{q} \alpha_i}.$$

Outline

Estimation of GARCH Models

- GARCH cannot be estimated by ordinary least-squares (because σ_t^2 is not observed).
- Such models are estimated by *maximum likelihood*: the joint density of the observations $\{r_1, \ldots, r_T\}$ is maximized with respect to the parameters.
- Maximization of $\log L$ can be done by numerical optimization algorithms. By default, EViews does this under the assumption of normality.
- If we are not sure that the z_t's are normally distributed, then we may still use the same estimation technique. This is called *quasi-maximum likelihood estimator*.
- However, we need to construct standard errors via a more robust method (*Bollerslev-Wooldridge standard errors*).

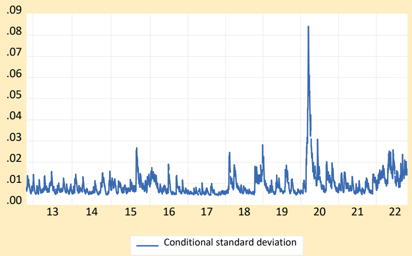
Example: EViews output, estimated GARCH model for S&P500

Dependent Variable: R
Method: ML ARCH - Normal distribution (BFGS / Marquardt steps)
Date: 11/24/22 Time: 15:40
Sample (adjusted): 10/30/2012 10/27/2022
Included observations: 2546 after adjustments
Convergence achieved after 18 iterations
Coefficient covariance computed using outer product of gradients
Presample variance: backcast (parameter = 0.7)
GARCH = C(2) + C(3)*RESID(-1)^2 + C(4)*GARCH(-1)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.000799	0.000135	5.911544	0.0000
Variance Equation				
C	4.50E-06	4.30E-07	10.46617	0.0000
RESID(-1)^2	0.223750	0.016516	13.54710	0.0000
GARCH(-1)	0.741309	0.016652	44.51892	0.0000

R-squared	-0.001398	Mean dependent var	0.000390
Adjusted R-squared	-0.001398	S.D. dependent var	0.010946
S.E. of regression	0.010953	Akaike info criterion	-6.742939
Sum squared resid	0.305331	Schwarz criterion	-6.733760
Log likelihood	8587.761	Hannan-Quinn criter.	-6.739609
Durbin-Watson stat	2.287783		

Example: Estimated GARCH volatility of S&P500



Outline

Testing GARCH Models

- Diagnostic tests are based on the *standardized residuals* $\hat{z}_t := \hat{u}_t/\hat{\sigma}_t$. If μ_t and σ_t are correctly specified, we should find no autocorrelation in $\hat{z}_t$ and $\hat{z}_t^2$.
- Therefore, the model can be tested using Q-statistics for $\hat{z}_t$ or $\hat{z}_t^2$.
- Lagrange-Multiplier (LM) test against ARCH, which is obtained by $T \cdot R^2$ in the regression
$$\hat{z}_t^2 = \gamma_0 + \gamma_1 \hat{z}_{t-1}^2 + \ldots + \gamma_m \hat{z}_{t-m}^2 + e_t.$$
- To test for normality of z_t, we can use the Jarque-Bera test based on the skewness and kurtosis of $\hat{z}_t$.

Example: Correlogram of standardized residuals for the S&P500

Date: 11/24/22 Time: 15:48
Sample (adjusted): 10/30/2012 10/27/2022
Included observations: 2546 after adjustments

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob*
		1	-0.032	-0.032	2.6093	0.106
		2	0.010	0.009	2.8792	0.237
		3	-0.007	-0.007	3.0169	0.389
		4	-0.014	-0.014	3.4899	0.479
		5	-0.022	-0.023	4.7188	0.451
		6	-0.022	-0.023	5.9352	0.430
		7	0.027	0.025	7.7406	0.356
		8	-0.028	-0.027	9.8064	0.279
		9	0.008	0.005	9.9642	0.353
		10	-0.009	-0.009	10.169	0.426
		11	0.021	0.020	11.328	0.416
		12	-0.033	-0.032	14.092	0.295
		13	-0.030	-0.033	16.403	0.228
		14	-0.027	-0.030	18.323	0.192
		15	-0.014	-0.014	18.815	0.222
		16	0.000	-0.002	18.815	0.278
		17	0.022	0.021	20.073	0.271
		18	0.021	0.017	21.170	0.271
		19	-0.021	-0.020	22.254	0.272
		20	0.018	0.015	23.127	0.283

*Probabilities may not be valid for this equation specification.

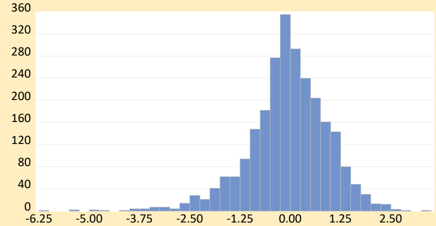
Example: Correlogram of squared standardized residuals for the S&P500

Date: 11/24/22 Time: 15:48
Sample (adjusted): 10/30/2012 10/27/2022
Included observations: 2546 after adjustments

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob*
		1	0.001	0.001	0.0011	0.973
		2	-0.000	-0.000	0.0015	0.999
		3	-0.003	-0.003	0.0182	0.999
		4	0.024	0.024	1.4430	0.837
		5	-0.028	-0.028	3.4029	0.638
		6	-0.008	-0.008	3.5668	0.735
		7	-0.021	-0.021	4.6583	0.702
		8	-0.017	-0.017	5.3624	0.718
		9	-0.008	-0.006	5.5064	0.788
		10	0.038	0.037	9.1425	0.519
		11	-0.004	-0.004	9.1864	0.605
		12	-0.017	-0.017	9.8904	0.626
		13	0.000	-0.001	9.8906	0.703
		14	-0.018	-0.020	10.676	0.711
		15	-0.002	-0.001	10.686	0.774
		16	0.003	0.004	10.712	0.827
		17	0.002	0.002	10.720	0.871
		18	-0.007	-0.005	10.850	0.901
		19	-0.007	-0.009	10.990	0.924
		20	0.020	0.017	11.993	0.916

*Probabilities may not be valid for this equation specification.

Example: Normality test of standardized residuals for the S&P500



Series: Standardized Residuals
Sample 10/30/2012 10/27/2022
Observations 2546

Mean	-0.060225
Median	-0.034548
Maximum	3.439897
Minimum	-6.012431
Std. Dev.	0.998517
Skewness	-0.737406
Kurtosis	5.246133
Jarque-Bera	765.9409
Probability	0.000000

Outline

Asymmetry and the News Impact Curve

- The *news impact curve* (NIC) is the effect of u_t on σ^2_{t+1}, keeping σ^2_t and the past fixed.
- For GARCH(1,1), this is the parabola $NIC(u_t|\sigma^2_t = \sigma^2) = A + \alpha u_t^2$, with $A = \omega + \beta\sigma^2$. This has a minimum at $u_t = 0$, and is symmetric around that minimum.
- For equity, a large negative shock is expected to increase volatility more than a large positive shock, because of *leverage effect*:

  $\downarrow$ value of firm's stock
  $\Rightarrow$ $\downarrow$ equity value of the firm
  $\Rightarrow$ $\uparrow$ debt-to-equity ratio
  $\Rightarrow$ shareholders (as residual claimants) perceive future cashflows as more risky.
- Two popular proposals to deal with this issue:
  - Nelson's exponential GARCH (EGARCH);
  - Glosten, Jagannathan and Runkle's GJR-GARCH.

GJR-GARCH (or TARCH, threshold GARCH)

The GJR-GARCH(1,1) model is

$$\sigma_{t+1}^2 = \omega + \alpha u_t^2 + \gamma u_t^2 \mathrm{I}_t + \beta \sigma_t^2.$$

where

$$\mathrm{I}_t = \begin{cases} 1 & \text{if} \quad u_t < 0 \\ 0 & \text{if} \quad u_t \geq 0 \end{cases},$$

and u_t/σ_t has a symmetric distribution.
Properties:

- NIC is asymmetric if and only if $\gamma \neq 0$; leverage effect if $\gamma > 0$;
- σ_t^2 is positive if $\omega > 0$, $\alpha \geq 0$, $\gamma \geq 0$, $\beta \geq 0$;
- u_t^2 is stationary if $0 \leq \alpha + \frac{1}{2}\gamma + \beta < 1$, with unconditional variance $\sigma^2 = \omega / \left[1 - \alpha - \frac{1}{2}\gamma - \beta\right]$.

EGARCH

The EGARCH(1,1) model is

$$\log \sigma_{t+1}^2 = \omega + \gamma z_t + \alpha(|z_t| - E\,|z_t|) + \beta \log \sigma_t^2,$$

with $z_t = u_t/\sigma_t$ as usual. If $z_t \sim$ i.i.d. $N(0,1)$ then $E\,|z_t| = \sqrt{2/\pi}$.
Properties:

- NIC is asymmetric if and only if $\gamma \neq 0$; leverage effect if $\gamma < 0$;
- σ_t^2 is positive for all parameter values;
- $\gamma z_t + \alpha(|z_t| - E\,|z_t|)$ is an i.i.d. mean-zero shock to log-volatility;
- if $|\beta| < 1$, $\log \sigma_t^2$ is stationary with mean $\omega/(1-\beta)$.

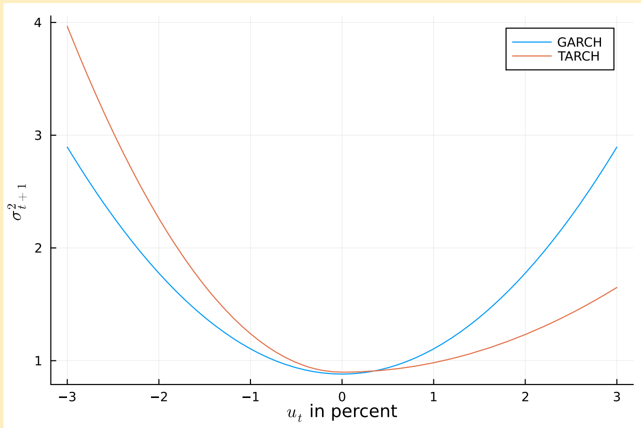
Example: EViews output, estimated TARCH model for S&P500

Dependent Variable: R
Method: ML ARCH - Normal distribution (BFGS / Marquardt steps)
Date: 11/24/22 Time: 15:56
Sample (adjusted): 10/30/2012 10/27/2022
Included observations: 2546 after adjustments
Convergence achieved after 26 iterations
Coefficient covariance computed using outer product of gradients
Presample variance: backcast (parameter = 0.7)
GARCH = C(2) + C(3)*RESID(-1)^2 + C(4)*RESID(-1)^2*(RESID(-1)<0) + C(5)*GARCH(-1)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.000472	0.000146	3.229615	0.0012
Variance Equation				
C	4.18E-06	3.67E-07	11.39238	0.0000
RESID(-1)^2	0.083481	0.008795	9.491794	0.0000
RESID(-1)^2*(RESID(-1)<0)	0.257376	0.028027	9.183082	0.0000
GARCH(-1)	0.755584	0.014405	52.45111	0.0000

R-squared	-0.000056	Mean dependent var	0.000390
Adjusted R-squared	-0.000056	S.D. dependent var	0.010946
S.E. of regression	0.010946	Akaike info criterion	-6.767193
Sum squared resid	0.304922	Schwarz criterion	-6.755719
Log likelihood	8619.636	Hannan-Quinn criter.	-6.763031
Durbin-Watson stat	2.290853		

Example: NIC of GARCH and TARCH models for S&P500



Outline

Volatility Forecasting

- GARCH models directly provide forecasts of next day's volatility:

$$\widehat{\sigma}^2_{t+1} = \hat{\omega} + \hat{\alpha}\hat{u}_t^2 + \hat{\beta}\widehat{\sigma}_t^2.$$

- This can also be expressed as

$$\widehat{\sigma}^2_{t+1} = \widehat{\sigma}^2 + \hat{\alpha}(\hat{u}_t^2 - \widehat{\sigma}^2) + \hat{\beta}(\widehat{\sigma}_t^2 - \widehat{\sigma}^2).$$

with $\widehat{\sigma}^2 = \hat{\omega}/(1-\hat{\alpha}-\hat{\beta})$; see the exercises.

- So the forecast differs from the average variance $\widehat{\sigma}^2$ if $\hat{u}_t^2$ or $\widehat{\sigma}_t^2$ differ from $\widehat{\sigma}^2$.

Multi-Period Forecasts

- Regarding multi-period forecasts of the stationary GARCH(1, 1) model, it can be shown that

$$\widehat{\sigma}^2_{t+s} = \widehat{\sigma}^2 + (\hat{\alpha}+\hat{\beta})^{s-1}(\widehat{\sigma}^2_{t+1} - \widehat{\sigma}^2),$$

see the exercises.

- This implies $\widehat{\sigma}^2_{t+s} \to \widehat{\sigma}^2$ as $s \to \infty$.
- For RiskMetrics ($\alpha + \beta = 1$, $\omega = 0$), this simplifies to $\widehat{\sigma}^2_{t+s} = \widehat{\sigma}^2_{t+1}$ for all s.

Outline

Learning Goals

Students

- can use appropriate tests to detect volatility clustering,
- are able to estimate, interpret, and forecast the various models (historical volatility, RiskMetrics, (G)ARCH, TARCH, EGARCH), and to apply diagnostic tests to the standardized residuals,
- and understand the concept of leverage, and the NIC.

Homework

- Exercise 5
- Questions 1 and 3 from Chapter 9 of Brooks (2019)

Module 9.3: Time Series Analysis
Fall Term 2022

Week 6:

Value at Risk

HSLU Lucerne University of Applied Sciences and Arts

Outline in Weeks

1. Introduction; Descriptive Modelling
2. Returns; Autocorrelation; Stationarity
3. ARMA Models
4. Unit Roots; Regressions between Time Series
5. Volatility Modelling
6. Value at Risk
7. Cointegration

Outline

Value at Risk

- Consider a portfolio with value $V_{PF,t}$ and daily returns $R_{PF,t+1}$.
- Define the one-day Loss on the portfolio as

$$\$Loss_{t+1} = V_{PF,t} - V_{PF,t+1}.$$

- The one-day, $100p\%$, dollar *Value at Risk* ($\$VaR^p_{t+1}$) gives the largest loss on the portfolio that we can expect to incur in the next day with level of confidence $100\,(1-p)\,\%$.
- Mathematically it is given by

$$\Pr(\$Loss_{t+1} \leq \$VaR^p_{t+1}) = 1 - p,$$

or equivalently

$$\Pr(\$Loss_{t+1} > \$VaR^p_{t+1}) = p.$$

Value at Risk

- Usually easier to express the VaR as a percentage of the portfolio value:

$$VaR^p_{t+1} = \frac{\$VaR^p_{t+1}}{V_{PF,t}}.$$

- Hence

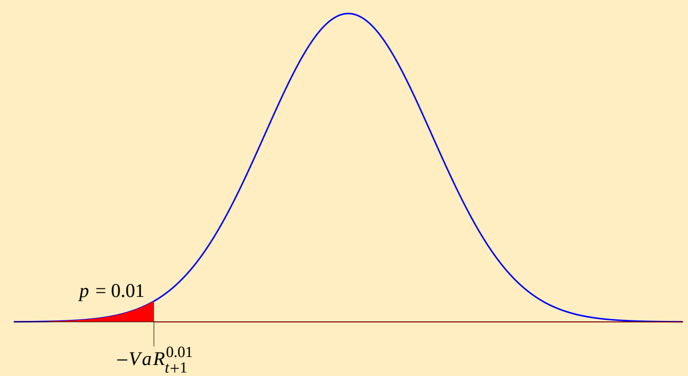
$$\Pr(R_{PF,t+1} < -VaR^p_{t+1}) = p,$$

as

$$R_{PF,t+1} = -\frac{\$Loss_{t+1}}{V_{PF,t}}.$$

- Thus VaR^p_{t+1} is minus the 100pth *percentile* of the return distribution. Usually $p = 0.01$.
- Definition can be naturally extended to K-day VaR, from the distribution of the K-day returns $R_{PF,t+1:t+K}$.

Probability density function of daily returns



Value at Risk

- Value at Risk was proposed as the standard measure of portfolio risk by the Basel Committee of the Bank of International Settlements in 1996.
- The BC imposed that financial institutions should report the Value at Risk on their positions, such that regulators could check the adequacy of the economic capital as a buffer against market risk.
- Banks were allowed to use their own, internal models for the computation of VaR, but the adequacy of these models should be "backtested" using specific criteria.
- A candidate for a standard model is RiskMetrics (developed by J.P.Morgan).
- VaR is scheduled to be replaced by the expected shortfall (ES) with the rollout of Basel 3. The ES is based on the VaR, however.

Outline

VaR Methods: Historical simulation

Historical simulation assumes that the distribution of tomorrow's portfolio returns is well approximated by the empirical distribution (histogram) of the past m observations $\{R_{PF,t}, R_{PF,t-1}, \ldots, R_{PF,t+1-m}\}$.
This is as if we draw, with replacement, from the last m returns and use this to simulate the next day's return distribution.

- The estimator of VaR is given by minus the 100pth percentile of the sequence of past portfolio returns, that is:
  - sort the returns $\{R_{PF,t}, R_{PF,t-1}, ..., R_{PF,t+1-m}\}$ is ascending order;
  - define R^p_{t+1} as the number such that 100p% of the observations are smaller than R^p_{t+1};
  - the estimator for VaR is given by

$$\widehat{VaR}^p_{t+1} = -R^p_{t+1}.$$

- R^p_{t+1} can be computed using EViews' `@quantile` function.

VaR Methods: Historical simulation

Problems / limitations of historical simulation:

- Last year(s) of data not necessarily representative for the next few days (e.g. because of volatility clustering).
- Similar problems as historical volatility (choice of m).
- A large m is required to compute 1% VaR with any degree of precision.
- By focussing on left tails, extreme positive returns are ignored.

Outline

VaR Methods: Normal distribution

- Another simple approach is to assume $R_{t+1} = R_{PF,t+1} \sim N(\mu, \sigma^2)$ and to estimate μ and σ^2 using historical data.
- Denoting the inverse distribution function (quantile function) of the Normal, as Φ^{-1}_p, The VaR becomes

$$VaR^p_{t+1} = -\mu - \sigma\Phi^{-1}_p.$$

  For example, $\Phi^{-1}_{.01} = -2.326$. For daily data we can take $\mu = 0$.

VaR Methods: Normal distribution

- The normal model can be easily extended to a *conditionally* normal model. Assume $R_{t+1} \sim N(0, \sigma^2_{t+1})$ where σ^2_{t+1} may be estimated by:
  - EWMA / RiskMetrics;
  - univariate GARCH;
  - multivariate GARCH.

  The VaR then becomes $VaR^p_{t+1} = -\sigma_{t+1}\Phi^{-1}_p$.

Outline

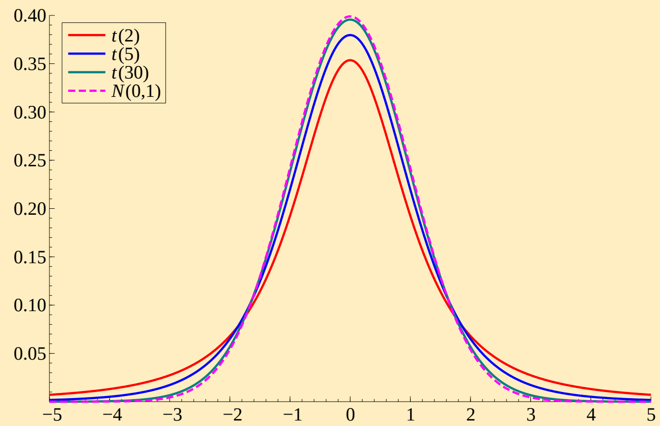
VaR Methods: Standardized *t* distribution

- The VaR methods described on the previous slides require that financial returns are normally distributed.
- This can be tested by the Jarque-Bera test and is usually rejected.
- Solution: use Student's $t(d)$ distribution, where d.o.f. $d > 0$ need not be integer.
- d is just a shape parameter. Small values correspond to fat tails. As $d \to \infty$, we approach the $N(0,1)$ distribution.
- For $d > 2$, the variance of a $t(d)$ random variable x is $d/(d-2)$; the distribution of

$$z = \frac{x}{\sqrt{\mathrm{var}(x)}} = \sqrt{\frac{d-2}{d}}x$$

  is called *standardized* $t(d)$, denoted $\tilde{t}(d)$.
- For $d > 4$ the excess kurtosis is $6/(d-4)$. The distributions are symmetric around 0 (hence mean and skewness are 0).

Student's *t* densities



VaR Methods: Standardized *t* distribution

- The GARCH model $R_{t+1} = \sigma_{t+1} z_{t+1}$, $\sigma^2_{t+1} = \omega + \alpha R_t^2 + \beta \sigma_t^2$, may be extended to $z_t \sim \tilde{t}(d)$, where d is an extra parameter that can be estimated by maximum likelihood.
- In practice this GARCH-t model often gives a substantially better fit than the Gaussian model. The main problem is that the standardized residuals usually have an asymmetric distribution, with a longer left tail than right tail.

Estimation of GARCH-*t* in EViews:

Dependent Variable: R
Method: ML ARCH - Student's t distribution (BFGS / Marquardt steps)
Date: 12/01/22 Time: 17:14
Sample (adjusted): 10/30/2012 10/27/2022
Included observations: 2546 after adjustments
Convergence achieved after 27 iterations
Coefficient covariance computed using outer product of gradients
Presample variance: backcast (parameter = 0.7)
GARCH = C(2) + C(3)*RESID(-1)^2 + C(4)*GARCH(-1)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.000904	0.000126	7.176877	0.0000
		Variance Equation		
C	2.92E-06	6.17E-07	4.743725	0.0000
RESID(-1)^2	0.221296	0.026849	8.242232	0.0000
GARCH(-1)	0.773192	0.023233	33.27982	0.0000
T-DIST. DOF	5.275193	0.568869	9.273121	0.0000

R-squared	-0.002211	Mean dependent var	0.000390
Adjusted R-squared	-0.002211	S.D. dependent var	0.010946
S.E. of regression	0.010958	Akaike info criterion	-6.810961
Sum squared resid	0.305579	Schwarz criterion	-6.799488
Log likelihood	8675.354	Hannan-Quinn criter.	-6.806800
Durbin-Watson stat	2.285926		

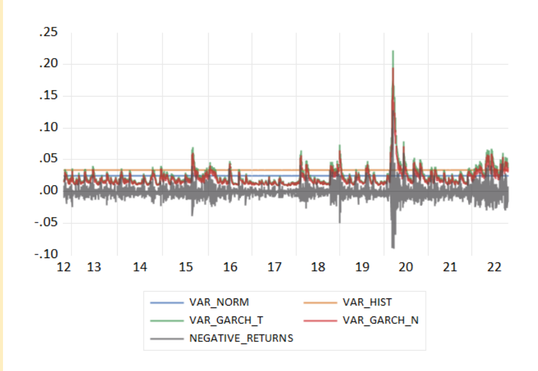
VaR Methods: Standardized *t* distribution

- Let $\tilde{t}_p^{-1}(d)$ be $100p\%$ quantile of the standardized t distribution $\tilde{t}(d)$ and $t_p^{-1}(d)$ the percentile $100p\%$ of the t distribution $t(d)$.
- The implied VaR now is

$$VaR^p_{t+1} = -\sigma_{t+1} \tilde{t}_p^{-1}(d) = -\sigma_{t+1} \sqrt{\frac{d-2}{d}} t_p^{-1}(d),$$

where, e.g., $\tilde{t}^{-1}_{.01}(6) = -2.566$.

Example: minus S&P500 returns, with 1% VaR based on (full-sample) historical simulation, normal distribution, and a GARCH(1, 1) with Normal and *t* errors



Outline

Expected Shortfall

Limitations of Value at Risk:

- VaR is not informative about the magnitude of the losses if they exceed the VaR. Two distributions could have the same 1% VaR, but with different left tails.
- VaR is not *subadditive*: it is not guaranteed that

$$VaR^p_{t+1}(X+Y) \leq VaR^p_{t+1}(X) + VaR^p_{t+1}(Y).$$

  This means that VaR is not a “coherent” risk measure.

Expected Shortfall

- The 1% VaR will be replaced by the 2.5% *expected shortfall* (ES, a.k.a. CVaR), which addresses these problems, on January 1st, 2023:

$$ES^p_{t+1} = -E_t\left[R_{t+1} | R_{t+1} < -VaR^p_{t+1}\right].$$

Outline

Multi-Period VaR

- For the GARCH-$N(0,1)$ and GARCH-$\tilde{t}(d)$ models, the one-day VaR and ES can be determined analytically (when the estimation is based on daily data).
- However, in practice one often needs risk measures for multi-period returns:

$$R_{t+1:t+K} = \sum_{k=1}^{K} R_{t+k}.$$

For example, a horizon of two weeks ($K = 10$ trading days) is common.

Multi-Period VaR

- Problem: even if the distribution of the one-period return is known (e.g., normal), that of $R_{t+1:t+K}$ is not (because the variance is not deterministic).
- *Monte Carlo simulation* is a possible solution: we let the computer generate a large number of scenarios of K daily returns, and compute from this the conditional distribution of the K-day return, and hence the K-day VaR and ES.
- Quick-and dirty practitioner solution: scale the one-day VaR with $\sqrt{K}$ (*square root of time rule*). Only correct under normality.

Outline

Backtesting Value at Risk

- The Basel Committee requires that methods to evaluate VaR be backtested (`http://www.bis.org/publ/bcbsc223.pdf`).
- They recommend constructing the 1% VaR over the last 250 trading days ($\approx$ 1 year), and counting the number of times losses exceed the day's VaR figure (termed *exceptions* or *violations*).
- A method is said to lie in the:
  - *Green zone*, in case of 0–4 exceptions;
  - *Yellow zone*, in case of 5–9 exceptions;
  - *Red zone*, in case of 10 exceptions or more.
- The capital charge for the bank changes according to the zone.

Backtesting Value at Risk

How can we test if a VaR method is accurate?

- Define the *hit sequence*

$$I_{t+1} = \begin{cases} 1, & \text{if } R_{t+1} < -VaR^p_{t+1}, \\ 0, & \text{if } R_{t+1} > -VaR^p_{t+1}. \end{cases}$$

- Consider a test period that covers $t+1 \in \{1, \ldots, T\}$, then the number of exceptions is given by $T_1 = \sum_{t=1}^{T} I_t$.
- The proportion of exceptions is given by $\hat{\pi} = T_1/T$ which is an estimator of $\Pr\left(R_{t+1} < -VaR^p_{t+1}\right)$.
- Recall that if the model that generated VaR^p_{t+1} is correctly specified, then

$$\Pr\left(R_{t+1} < -VaR^p_{t+1}\right) = p,$$

independent of any information at time t.

Backtesting Value at Risk

- Hence, under the null hypothesis of correct specification, the hit sequence $\{I_{t+1}\}$ are independent Bernoulli random variables, and so $T_1 = \sum_{t=1}^{T} I_t$ has a Binomial(T,p) distribution.
- We can test this hypothesis (e.g., with $p = 0.01$) based on the t-statistics

$$t_0 = \frac{\hat{\pi} - p}{\sqrt{p(1-p)/T}} \quad \text{or} \quad t = \frac{\hat{\pi} - p}{\sqrt{\hat{\pi}(1-\hat{\pi})/T}}.$$

- Under H_0 their asymptotic distribution is $N(0,1)$.
- The second t-statistic is equal (up to degrees-of-freedom correction) to OLS-based t-statistic in regression of $I_{t+1} - p$ on a constant.

Backtesting Value at Risk

- The previous test only checks *unconditional* coverage, i.e., $\Pr(I_{t+1} = 1) = p$ *on average*. However, misspecification often is due to the fact that the hits I_{t+1} are not independent over time.
- If exceptions are clustered, then if today there was an exception a risk manager can infer that the probability of occurring another exception tomorrow is higher than p. Hence, there is misspecification.
- We would like to test if the VaR violations are *independent* over time, the null hypothesis is

$$H_0 : \Pr(I_{t+1} = 1 | I_t = 1) = \Pr(I_{t+1} = 1 | I_t = 0),$$

which implies $\Pr(I_{t+1} = 0 | I_t = 0) = \Pr(I_{t+1} = 0 | I_t = 1)$.

Backtesting Value at Risk

- Also of interest is to test if the VaR violations are independent over time and if the number of violations is correct (*conditional coverage*)

$$H_0 : \Pr(I_{t+1} = 1 | I_t = 1) = \Pr(I_{t+1} = 1 | I_t = 0) = p.$$

- A simple approach to test these hypotheses is to consider the linear regression model

$$I_{t+1} - p = b_0 + b_1 I_t + e_{t+1}$$

- The *conditional coverage* hypothesis is equivalent to $H_0 : b_0 = 0$ and $b_1 = 0$ which can be tested using a F-test.
- The *independence* hypothesis is equivalent to $H_0 : b_1 = 0$ which can be tested using a t-statistic.

Backtesting Value at Risk

Results for S&P500 returns, 4 different methods

	Hist	Norm	GARCHn	GARCHt
$\hat{\pi}\quad(\times 100)$	0.98	2.04	2.59	1.65
$t(\pi = 0.01)$	-0.09	3.72	5.06	2.57
$\hat{b}_1$	0.03	0.08	0.00	0.03
$t(b_1 = 0)$	1.54	3.91	0.23	1.60
$F(b_0 = b_1 = 0)$	1.18	14.60	12.80	4.59

The critical values for the t and F tests are, respectively, ± 1.96 and 3.00.
Note: this result is highly unusual. Usually, the GARCHt model fares best.

VaR · Historical Simulation · Normal Distribution · t Distribution · Expected Shortfall · Multi-Period VaR · Backtesting · Epilogue

Outline

VaR · Historical Simulation · Normal Distribution · t Distribution · Expected Shortfall · Multi-Period VaR · Backtesting · Epilogue

Learning Goals

Students

- know the definitions of VaR and Expected Shortfall,
- understand the limitations of the VaR,
- are able to construct VaR forecasts based on various methods,
- and are able to backtest VaR forecasts.

VaR · Historical Simulation · Normal Distribution · t Distribution · Expected Shortfall · Multi-Period VaR · Backtesting · Epilogue

Homework

- Exercise 6

Module 9.3: Time Series Analysis
Fall Term 2022

Week 7:

Cointegration

HSLU Lucerne University of Applied Sciences and Arts

Cointegration and Common Trends ECM and Engle-Granger Example Johansen's Procedure Example continued Epilogue

Outline in Weeks

1. Introduction; Descriptive Modelling
2. Returns; Autocorrelation; Stationarity
3. ARMA Models
4. Unit Roots; Regressions between Time Series
5. Volatility Modelling
6. Value at Risk
7. Cointegration

Cointegration and Common Trends ECM and Engle-Granger Example Johansen's Procedure Example continued Epilogue

Outline

Cointegration and Common Trends ECM and Engle-Granger Example Johansen's Procedure Example continued Epilogue

Cointegration and Common Trends

Suppose we have two time series Y_t and X_t, which are both $I(1)$, and we analyze a regression model of the form

$$Y_t = \beta_1 + \beta_2 X_t + U_t.$$

Here U_t has mean zero but may display autocorrelation. Two cases:

- $U_t \sim I(1)$: if U_t displays no mean-reversion, then Y_t does not revert to the explained part $\beta_1 + \beta_2 X_t$. Even if $\beta_2 = 0$, its t-statistic and R^2 will often seem significant (*spurious regressions*). To avoid this, one should estimate a model in differences, i.e., $\Delta Y_t = a_1 + a_2 \Delta X_t + \Delta U_t$.
- $U_t \sim I(0)$: now Y_t and X_t have a *common* stochastic trend, such that the linear combination $Y_t - \beta_2 X_t$ does not have a trend. This is called *cointegration*.

Example

- Consider the model

$$Y_t = \beta_1 + \beta_2 X_t + U_{1,t}$$
$$X_t = X_{t-1} + U_{2,t}$$

where $\beta_2 \neq 0$, $U_{1,t}, U_{2,t} \overset{\text{iid}}{\sim} (0, \sigma^2)$ independently of each other.
- X_t is a random walk and thus nonstationary. Y_t contains X_t and is thus also nonstationary. But

$$Y_t - \beta_2 X_t = \beta_1 + U_{1,t}$$

is stationary: the RHS is white noise plus a constant.
- $(1, -\beta_2)$ is called the *cointegrating vector*.

Cointegration and Common Trends, contd.

- The concept is easily extended to more than two series: if $X_{2t}, \dots, X_{kt}$ are all $I(1)$ variables, and

$$Y_t = \beta_1 + \beta_2 X_{2t} + \dots + \beta_k X_{kt} + U_t,$$

then this is a spurious regression if $U_t \sim I(1)$ (and $\beta_i = 0$), and a cointegrating relation if U_t is stationary.
- In other words, cointegration between k integrated series means that there exists a *linear combination*[1] of them which is stationary.
- Examples of possibly cointegrated time series:
  - exchange rates and relative prices (*purchasing power parity*);
  - spot and futures prices of assets or exchange rates;
  - short- and long-term interest rates (*term structure models*);
  - stock prices and dividends (*present value relations*).

[1] i.e., a weighted sum

Outline

Testing for Cointegration

- For cointegrated series, one should exploit the long-run equilibrium relationship between variables for estimation rather than differencing. Differencing would *remove* that structure.
- Engle and Granger proposed the following procedure:
  - Conduct individual unit root tests to ensure all series are $I(1)$.
  - Estimate the regression model

$$Y_t = \beta_1 + \beta_2 X_{2t} + \dots + \beta_k X_{kt} + U_t$$

  by ordinary least-squares. Estimates are (super)consistent, but standard errors are wrong because series are $I(1)$.
  - Apply an ADF unit root test (with constant) to the residuals $\hat{u}_t$ from this regression. This yields a test for $H_0 : U_t \sim I(1)$ (spurious regression) against $H_1 : U_t \sim I(0)$ (cointegration). The critical values depend on k. E.g., for $k = 2$ (X_{2t} and an intercept), the 5% c.v. is -3.41.
  - If H_0 is rejected, estimate an *error correction model*.

Engle-Granger critical values

Number of series (w/o constant)	2	3	4	5	6
Critical value	-3.41	-3.80	-4.16	-4.49	-4.74

Error Correction Models

- Cointegration between Y_t and X_t implies that deviations $(Y_{t-1} - \beta_1 - \beta_2 X_{t-1})$ from the equilibrium level should be (partially) corrected in the next period, by Y_t, X_t, or both.
- This leads to a *vector error correction model* (VECM), which in the simplest form is

$$\begin{aligned} \Delta Y_t &= c_1 + \alpha_1 (Y_{t-1} - \beta_1 - \beta_2 X_{t-1}) + e_{1t}, \\ \Delta X_t &= c_2 + \alpha_2 (Y_{t-1} - \beta_1 - \beta_2 X_{t-1}) + e_{2t}, \end{aligned}$$

  where e_{1t} and e_{2t} are two white noise errors (possibly correlated), and where we expect $\alpha_1 < 0$ and/or $\alpha_2 \beta_2 > 0$.
- We might need to add lags of ΔY_t and/or ΔX_t on RHS to combat autocorrelation.
- The *Granger representation theorem* states that cointegration implies an error correction model (possibly with more lags), and vice versa; see exercises.

Engle-Granger Procedure

- The VECM

$$\begin{aligned} \Delta Y_t &= c_1 + \alpha_1 (Y_{t-1} - \beta_1 - \beta_2 X_{t-1}) + e_{1t}, \\ \Delta X_t &= c_2 + \alpha_2 (Y_{t-1} - \beta_1 - \beta_2 X_{t-1}) + e_{2t}. \end{aligned}$$

  is estimated by replacing $Y_{t-1} - \beta_1 - \beta_2 X_{t-1}$ by OLS residual $\hat{u}_{t-1} = Y_{t-1} - \hat{\beta}_1 - \hat{\beta}_2 X_{t-1}$, and estimating α_1 and α_2 by OLS.
- Note that $\hat{u}$ is stationary, so this is a valid regression!
- If $\alpha_2 = 0$, then all correction is done by Y_t, and not by X_t. In that case it makes sense to treat X_t as exogenous and Y_t as endogenous, and consider the "single-equation" error correction model

$$\Delta Y_t = c + \alpha_1 (Y_{t-1} - \beta_1 - \beta_2 X_{t-1}) + e_t.$$

- In general, both Y_t and X_t are endogenous.

Outline

Example

- Until 1971, as part of the Bretton-Woods system of fixed exchange rates, the US dollar was convertible to gold, i.e., it was possible for foreign central banks to redeem US dollars for gold at a fixed rate of 35$ per troy ounce, so that the price of gold was fixed.
- In 1971, US president Nixon unilaterally cancelled the direct convertibility, ultimately ending the Bretton-Woods agreement.
- Gold became a floating asset, and its price increased sharply; in other words, the US$ was massively devalued.

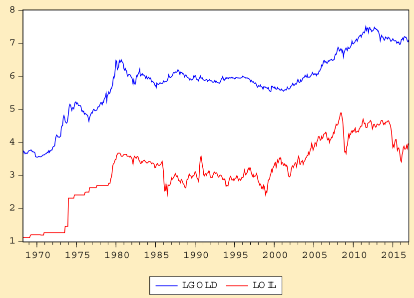
Cointegration and Common Trends | ECM and Engle-Granger | Example | Johansen's Procedure | Example continued | Epilogue

Example continued

- We want to analyze the hypothesis that the increasing price (in US$) of oil is not a consequence of an increased demand for (or a reduced supply of) oil, but rather of a continued devaluation of the US$.
- We have at our disposal monthly data from April 1968 to January 2017 (586 observations) on the following variables:
  - gold_t, the spot price of one troy ounce of gold in US$;
  - oil_t, the spot price of one barrel of WTI crude oil in US$.
- Idea: if the relative price of oil expressed in units of gold $\mathrm{oil}_t/\mathrm{gold}_t$ is stationary, then this implies that $\log(\mathrm{oil}_t) - \log(\mathrm{gold}_t)$ is stationary, so that $\log(\mathrm{oil}_t)$ and $\log(\mathrm{gold}_t)$ must be cointegrated if the individual series are integrated.

Cointegration and Common Trends | ECM and Engle-Granger | Example | Johansen's Procedure | Example continued | Epilogue

$\log(\mathrm{gold}_t)$ and $\log(\mathrm{oil}_t)$



Cointegration and Common Trends | ECM and Engle-Granger | Example | Johansen's Procedure | Example continued | Epilogue

Example continued

Step 0 Take logs: `genr lgold = log(gold), genr loil = log(oil).`

Step 1 Test that the variables are integrated (ADF test with constant and trend)

Null Hypothesis: LOIL has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 1 (Automatic - based on SIC, maxlag=18)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-2.552422	0.3027
Test critical values:	1% level	-3.973847	
	5% level	-3.417533	
	10% level	-3.131184	

*MacKinnon (1996) one-sided p-values.

Null Hypothesis: LGOLD has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 0 (Automatic - based on SIC, maxlag=18)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-1.820648	0.6937
Test critical values:	1% level	-3.973820	
	5% level	-3.417519	
	10% level	-3.131176	

*MacKinnon (1996) one-sided p-values.

Neither test rejects, so the series are I(1).

Example continued

Step 2 Estimate long-run relationship

$$\text{loil}_t = \beta_1 + \beta_2 \text{lgold}_t + U_t$$

and save the residuals (`Proc→Make Residual Series→Ordinary`) as `u`.

Dependent Variable: LOIL
Method: Least Squares
Date: 11/30/20 Time: 16:50
Sample: 1968M04 2017M01
Included observations: 586

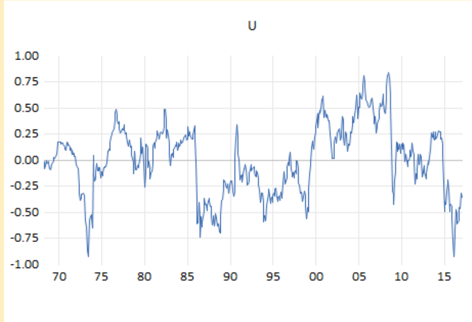
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-2.258963	0.086329	-26.16698	0.0000
LGOLD	0.926232	0.014618	63.36354	0.0000

R-squared	0.873014	Mean dependent var	3.141848
Adjusted R-squared	0.872797	S.D. dependent var	0.929639
S.E. of regression	0.331561	Akaike info criterion	0.633398
Sum squared resid	64.20072	Schwarz criterion	0.648324
Log likelihood	-183.5855	Hannan-Quinn criter.	0.639214
F-statistic	4014.938	Durbin-Watson stat	0.074317
Prob(F-statistic)	0.000000		

The cointegrating vector is $(1, -\beta_2) = (1, -0.926)$ (if the Engle-Granger test rejects).
Careful: standard errors are wrong, because variables are $I(1)$.

Example continued

Residuals (=equilibrium error) $\hat{u}_t$



Example continued

Step 3 Apply ADF test (with intercept) to `u`.

Null Hypothesis: U has a unit root
Exogenous: Constant
Lag Length: 1 (Automatic - based on SIC, maxlag=18)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-3.724771	0.0040
Test critical values:	1% level	-3.441318	
	5% level	-2.866270	
	10% level	-2.569348	

*MacKinnon (1996) one-sided p-values.

Careful: we need to use Engle-Granger 5% critical value of -3.41, not the ones given by EViews. Conclusion: test rejects the null of "no cointegration".

Example continued

Step 4 Estimate VECM. Note: I threw in a lag of the dependent variable in the equation for `d(loil)`, as there was autocorrelation without it (cf. selected lag length in ADF test).

Dependent Variable: D(LOIL)
Method: Least Squares
Date: 11/30/20 Time: 17:09
Sample (adjusted): 1968M06 2017M01
Included observations: 584 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.003641	0.003278	1.110606	0.2672
U(-1)	-0.034971	0.009942	-3.517509	0.0005
D(LOIL(-1))	0.256547	0.040092	6.399038	0.0000

R-squared	0.077666	Mean dependent var	0.004862
Adjusted R-squared	0.074491	S.D. dependent var	0.082210
S.E. of regression	0.079089	Akaike info criterion	-2.231362
Sum squared resid	3.634195	Schwarz criterion	-2.208914
Log likelihood	654.5578	Hannan-Quinn criter.	-2.222613
F-statistic	24.46179	Durbin-Watson stat	2.006622
Prob(F-statistic)	0.000000		

Dependent Variable: D(LGOLD)
Method: Least Squares
Date: 11/30/20 Time: 17:17
Sample (adjusted): 1968M05 2017M01
Included observations: 585 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.005866	0.002342	2.504458	0.0125
U(-1)	0.008997	0.007077	1.271298	0.2041

R-squared	0.002765	Mean dependent var	0.005871
Adjusted R-squared	0.001054	S.D. dependent var	0.056676
S.E. of regression	0.056646	Akaike info criterion	-2.900562
Sum squared resid	1.870743	Schwarz criterion	-2.885616
Log likelihood	850.4143	Hannan-Quinn criter.	-2.894737
F-statistic	1.616198	Durbin-Watson stat	1.932704
Prob(F-statistic)	0.204130		

The adjustment coefficient α_2 in the equation for `d(lgold)` is insignificant. So all the adjustment is done by `loil` $\rightarrow$ single-equation ECM.

Example continued

- The final model is the single-equation ECM

$$\Delta \mathrm{loil}_t = 0.0036 - 0.035(\mathrm{loil}_{t-1} - 0.926\mathrm{lgold}_{t-1} + 2.26) + 0.25\Delta \mathrm{loil}_{t-1} + e_{1t}$$

In our earlier notation,

$$\Delta Y_t = c + \alpha_1(Y_{t-1} - \beta_1 - \beta_2 X_{t-1}) + \gamma \Delta Y_{t-1} + e_{1t},$$

with $c = 0.0036$, $\alpha_1 = -0.035 < 0$ as desired, $\beta_1 = -2.26$, $\beta_2 = 0.926$, and $\gamma = 0.25$.

- Interpretation: there is an equilibrium relationship between loil and lgold. In case of a disequilibrium, loil adjusts towards the equilibrium. The adjustment amounts to 3.5% per period.

Outline

Limitations of Engle-Granger Procedure:

- Choice of the dependent variable Y_t in Engle-Granger test is arbitrary. Selecting another X_{it} as dependent variable should not matter asymptotically (as number of observations $T \to \infty$), but will make a difference in practice.
- Method can only be used if there is only a single unique cointegrating relation, which involves Y_t. So $(X_{2t}, \ldots, X_{kt})$ are not allowed to be cointegrated without Y_t.
- Although the OLS estimators $\hat{\beta}_i$ are consistent under cointegration, they do not have an asymptotic normal distribution, so standard inference (t-tests) fails.

Johansen Procedure

- An alternative is to use *Johansen's procedure*.
- Johansen has derived the MLE of the general VECM with normally distributed errors $\mathbf{u}_t$ and k variables.
- This facilitates likelihood ratio tests for H_{0r}: "r or fewer cointegrating relationships" against the alternative "more than r"
- The test is known as the *Johansen trace test*. The test statistics $\lambda_{trace}(r)$ can be expressed in terms of particular eigenvalues $\hat{\lambda}_i$. Their asymptotic distribution under the null is a multivariate version of the Dickey-Fuller distribution.
- We reject for large positive values of the test statistic; critical values and p-values are built into EViews and similar programs.

Johansen's cointegration test

These tests may be used to estimate the cointegrating rank r in the following way:

1. Start with $r = 0$;
2. Test H_{0r} with $\lambda_{trace}(r)$;
3. If H_{0r} is not rejected, then $\hat{r} = r$; if it is rejected, replace r by $r + 1$ and go back to step 2;
4. If H_{0r} is rejected for all $r = 0, 1, \ldots, k - 1$, then conclude $\hat{r} = k$ (this corresponds to a stationary system).

Cointegration and Common Trends | ECM and Engle-Granger | Example | Johansen's Procedure | Example continued | Epilogue

Treatment of constant and linear trend

Just like with the Dickey-Fuller test, we have to allow for a constant and possibly a linear trend in the VECM. Eviews distinguishes five cases:

1. No constant, no trend. This implies that the variables have mean zero. Rarely applicable.
2. Constant in the cointegrating relations, no trend. Applicable when the data have no trend (interest rates, inflation, real exchange rates).
3. Constant in the VECM equation. This leads to a drift, so applicable for trending series (real gdp, stock prices). This is the default, for good reasons.
4. Constant in VECM, trend in cointegration relations. Similar to 3, but hard to interpret.
5. Trend in the VECM equation. Leads to a quadratic trend. Rarely applicable.

Cointegration and Common Trends | ECM and Engle-Granger | Example | Johansen's Procedure | Example continued | Epilogue

Outline

Cointegration and Common Trends | ECM and Engle-Granger | Example | Johansen's Procedure | Example continued | Epilogue

Example continued

Step 1 Apply Johansen integration test.
Choose Option 3 (unrestricted constant) and one lagged Δ on RHS (lag specification `1 1`), because that's what we found in the Engle-Granger approach.

Johansen test

Unrestricted Cointegration Rank Test (Trace)

Hypothesized No. of CE(s)	Eigenvalue	Trace Statistic	0.05 Critical Value	Prob.**
None *	0.024313	17.92430	15.49471	0.0211
At most 1	0.006060	3.549966	3.841465	0.0595

Trace test indicates 1 cointegrating eqn(s) at the 0.05 level
* denotes rejection of the hypothesis at the 0.05 level
**MacKinnon-Haug-Michelis (1999) p-values

H_{00}: "No Cointegration" is rejected. H_{01}: "One cointegrating relationship" is not.
Conclusion: there is one cointegrating (=equilibrium) relationship.

Example continued

Step 2 Estimate VECM. Choose “1 cointegrating relationship”, and lag specification `1 1`.

Vector Error Correction Estimates
Date: 11/30/20 Time: 17:32
Sample (adjusted): 1968M06 2017M01
Included observations: 584 after adjustments
Standard errors in ( ) & t-statistics in [ ]

Cointegrating Eq:	CointEq1	
LOIL(-1)	1.000000	
LGOLD(-1)	-0.914126	
	(0.09263)	
	[-9.86821]	
C	2.187722	
Error Correction:	**D(LOIL)**	**D(LGOLD)**
CointEq1	-0.034498	0.007463
	(0.00997)	(0.00713)
	[-3.46000]	[1.04641]
D(LOIL(-1))	0.250758	0.048515
	(0.04063)	(0.02906)
	[6.17135]	[1.66924]
D(LGOLD(-1))	0.050704	0.019596
	(0.05870)	(0.04199)
	[0.86374]	[0.46668]
C	0.003354	0.005412
	(0.00329)	(0.00236)
	[1.01839]	[2.29777]

Example continued

- Final model is

$$\Delta \mathrm{loil}_t = 0.0034 - 0.034(\mathrm{loil}_{t-1} - 0.914\mathrm{lgold}_{t-1} + 2.19) + 0.25\Delta\mathrm{loil}_{t-1} + 0.05\Delta\mathrm{lgold}_{t-1} + e_{1t}$$
$$\Delta \mathrm{lgold}_t = 0.0054 - 0.007(\mathrm{loil}_{t-1} - 0.914\mathrm{lgold}_{t-1} + 2.19) + 0.049\Delta\mathrm{loil}_{t-1} + 0.019\Delta\mathrm{lgold}_{t-1} + e_{2t}$$

- Note how similar the first equation is to what we found with Engle-Granger! Here, too, $\alpha_2 = -0.007$ is insignificant, so we could ignore the 2nd equation.
- The cointegrating vector is $(1, -0.914)$.
- Interpretation: there is an equilibrium relationship between loil and lgold. In case of a disequilibrium, loil adjusts towards the equilibrium. The adjustment amounts to 3.4% per period.

Outline

Learning Goals

Students

- Understand the concept of cointegration,
- are able to test for cointegration using both the Engle-Granger procedure and the Johansen test,
- and are able to estimate an error correction model.

Homework

- Exercise 7
- Problem 6 from Chapter 8 of Brooks (2019)