Lecture 3

Autocorrelation

Lars A. Lochstoer

UCLA Anderson School of Management

Winter 2025

Overview of Lecture 3

Autocorrelation

1. Introduction to autocorrelations
   - How past observations affect future observations in a linear model
   - The autocorrelation function
2. The Ljung-Box Q-test

Correlation

Definition

The **correlation** between two random variables X and Y is defined as

$$\rho_{x,y} = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$$

- $\rho_{x,y}$ is known as Pearson's correlation
- measures linear dependence
- bounded between -1 and 1
- two variables are uncorrelated if $\rho_{x,y} = 0$, perfectly (negatively) correlated if $\rho_{x,y} = 1$ ($\rho_{x,y} = -1$)
- if X and Y are random normal variables, then $\rho_{x,y} = 0$ if and only if X and Y are independent

Sample Correlation

Definition

The **sample correlation** between two random variables X and Y is:

$$\widehat{\rho}_{x,y} = \frac{\sum_{t=1}^{T}(x_t - \overline{x})(y_t - \overline{y})}{\sqrt{\sum_{t=1}^{T}(x_t - \overline{x})^2 \sum_{t=1}^{T}(y_t - \overline{y})^2}}$$

where $\overline{x}$ and $\overline{y}$ are the sample means.

- this is **not** a regression coefficient
- $\widehat{\rho}_{x,y}$ consistently estimates $\rho_{x,y}$
- $\widehat{\rho}_{x,y}$ is built from **method of moments** estimators

Autocorrelation

Definition

The **autocorrelation** for a series $\{r_t\}$ is defined as:

$$\rho_j = \frac{\mathrm{Cov}(r_t, r_{t-j})}{\sqrt{Var(r_t)Var(r_{t-j})}} = \frac{\mathrm{Cov}(r_t, r_{t-j})}{Var(r_t)} = \frac{\gamma_j}{\gamma_0}.$$

- a covariance-stationary series r_t is not serially correlated if $\rho_j = 0$ for all j
- autocorrelations are a key signature of the dynamics of the time series you're interested in modeling
- l'th autocorrelation *is* a regression coefficient in univariate regression of r_t and r_{t-j}

Autocorrelation: why important?

First, from covariance-stationarity

$$Cov(r_t, r_{t-j}) = Cov(r_{t+j}, r_t)$$

Any time $\text{Cov}(r_{t+j}, r_t) \neq 0$, we have that current value of the series, r_t, can predict future realizations.

Patterns in $\text{Cov}(r_{t+j}, r_t)$ vs j tell you a lot about the nature of predictability within the series you are looking at.

- Note: auto*correlation*, as opposed to auto*covariance*, is convenient for intuition as the scale is easy to understand
- In fact, autocorrelations tell you about which model you need for capturing the predictability of the series at any horizon

1st Order Autocorrelation

Definition

The **sample autocorrelation** for a series $\{x_t\}$ is:

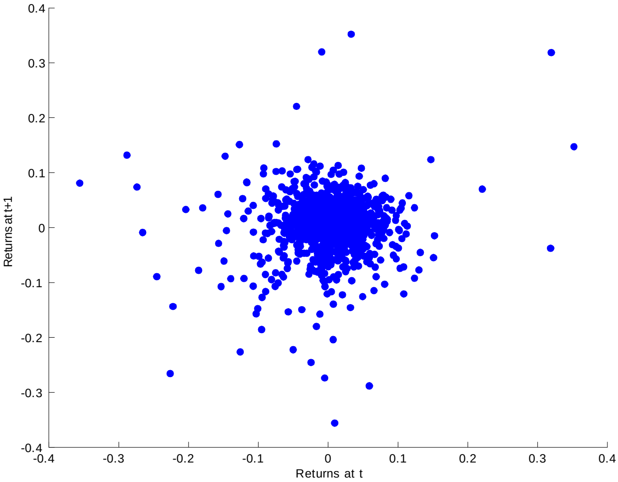
$$\widehat{\rho}_1 = \frac{\sum_{t=2}^{T}(x_t - \overline{x})(x_{t-1} - \overline{x})}{\sum_{t=1}^{T}(x_t - \overline{x})^2}, 0 \leq j \leq T - 1$$

where $\overline{x}$ are the sample means.

- under some conditions, $\widehat{\rho}_1$ is a consistent estimator of ρ_1
- $\widehat{\rho}_1$ is asymptotically normal with mean zero and variance $(1/T)$ if $\{x_t\}$ are independently and identically distributed over time.
- to test $H_0 : \rho_1 = 0$, use t-stat :

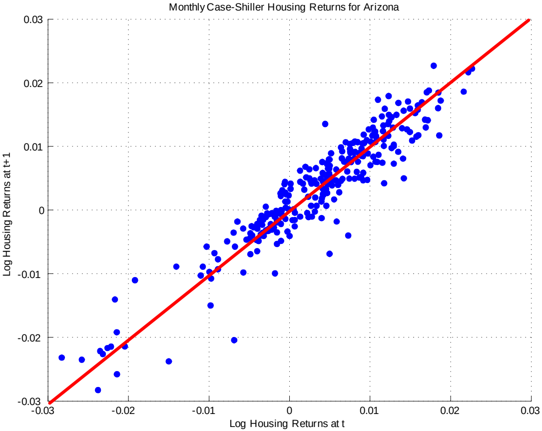
$$t = \sqrt{T}\widehat{\rho}_1$$

Autocorrelation in stock returns?



Scatter plot of monthly log returns (VW-CRSP) 1925-2013.

Autocorrelation in Real Estate Returns?



Scatter plot for Monthly log House Price Changes in AZ. Case-Shiller Index. 1987.1-2013.10

Higher-order Autocorrelations

Definition

The **sample autocorrelation** for a series $\{x_t\}$ at lag j is:

$$\widehat{\rho}_j = \frac{\sum_{t=j+1}^{T}(x_t - \overline{x})(x_{t-j} - \overline{x})}{\sum_{t=1}^{T}(x_t - \overline{x})^2}, 0 \leq j \leq T - 1$$

where $\overline{x}$ are the sample means.

Autocorrelation

1. financial time series: e.g. stock returns and housing returns
   - financial returns tend to be only very weakly autocorrelated [if markets are fairly efficient and liquid]
   - strong autocorrelations in returns would create huge profit opportunities!
2. macroeconomic time series: e.g. GDP growth rates
   - macroeconomic time series have growth rates that are highly autocorrelated
   - macroeconomic shocks tend to have very persistent effects (e.g. think about the effect of the subprime crisis on GDP growth rates)

Autocorrelation vs. "Persistence"

Persistence (or persistent) is not a precisely defined term, but refers to how long-lasting shocks are in terms of their impact on the series at hand

- A "persistent" time series refers to a time series with high autocorrelation at some lag (in absolute value, really, but in economics typically high and positive)

Persistence is typically used in qualitative description/discussion

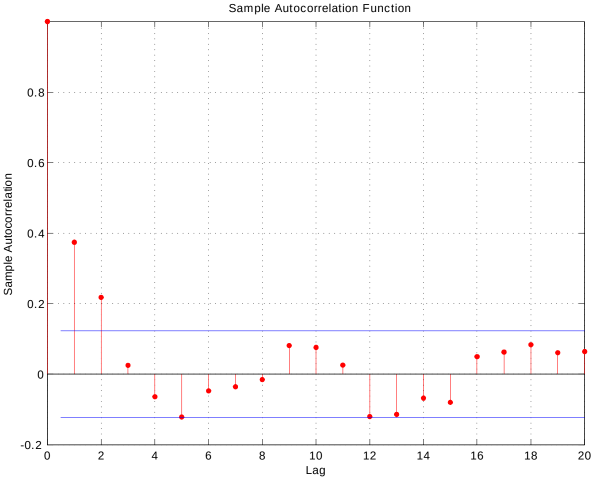
- Autocorrelations and the autocorrelation function (to be defined) are a little more technical

The Autocorrelation Function (ACF)

Definition

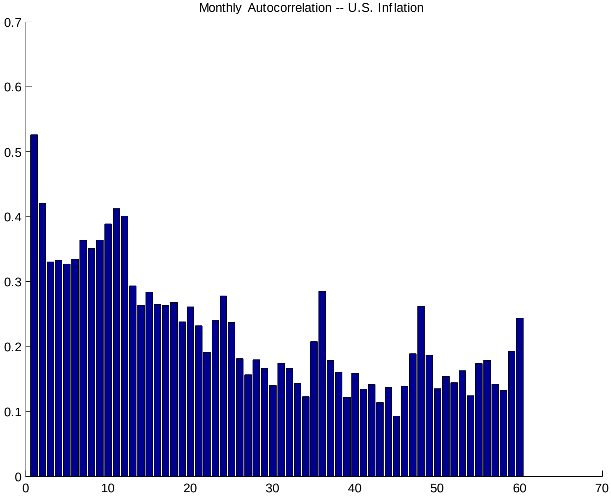
The **sample autocorrelation function** (ACF) of a time series is defined as $\widehat{\rho}_1, \widehat{\rho}_2, \ldots, \widehat{\rho}_k, \ldots,$

Autocorrelation in Quarterly GDP Growth



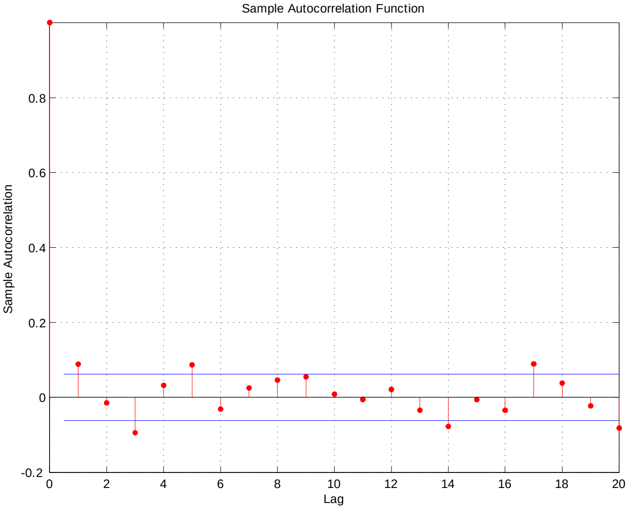
Autocorrelation Function for Quarterly U.S. GDP growth. Two standard error bands around zero. 1947.I-2012.IV

Autocorrelation in Monthly Inflation



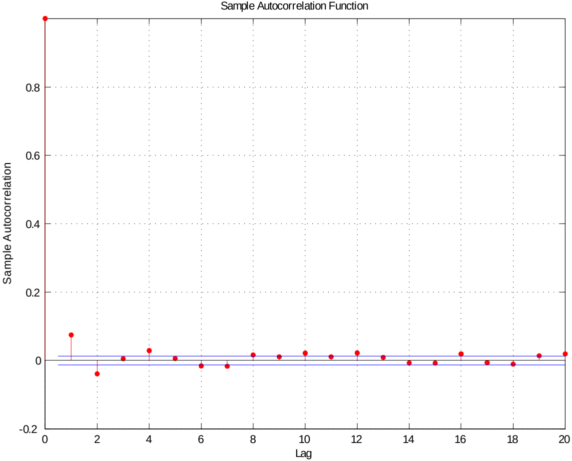
Autocorrelation Function for Monthly U.S. Inflation. 1950-2007. $\hat{\rho}_1 = 0.52$.

Autocorrelation of Monthly Log Stock Returns



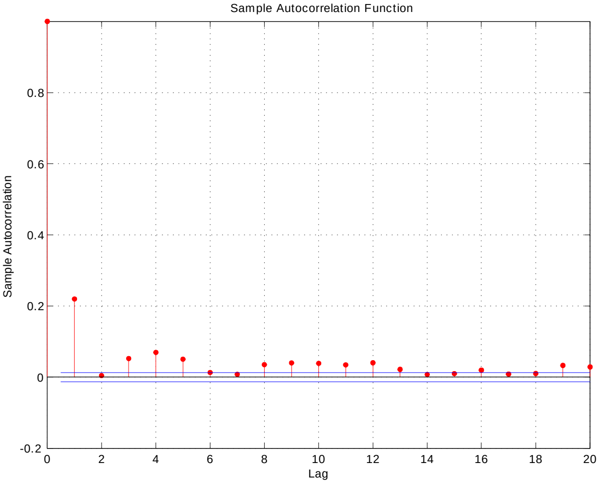
Autocorrelation Function for Monthly log Returns on VW-CRSP Index. Two standard error bands around zero. 1926-2012 . $\hat{\rho}_1 = 0.088$.

Autocorrelation of Daily Stock Returns



Autocorrelation Function for Daily log Returns on VW-CRSP Index (value-weighted). Two standard error bands around zero. 1926-2012. $\hat{\rho}_1 = 0.07$.

Autocorrelation of Daily Stock Returns



Autocorrelation Function for Daily log Returns on EW-CRSP Index (equal-weighted). Two standard error bands around zero. 1926-2012. $\hat{\rho}_1 = 0.21$.

ACF for the Stock Market

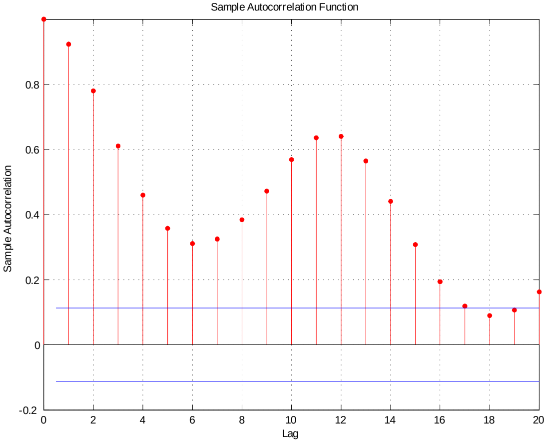
some evidence of mainly positive autocorrelation in stock market returns, more so for smaller stocks

- Thus, lagged stock market returns may slightly predict future returns, but R^2 very very small
- Consistent with an highly competitive market
  - That is, it is not easy to make money from trading against other investors

Benchmark Model of Portfolio Theory

- In the benchmark model of portfolio theory, returns are assumed to be independently and identically distributed (i.i.d.) over time.
  - If returns are i.i.d., the variance grows linearly in the investment horizon
  - The investor's horizon turns out to be irrelevant for optimal portfolio allocation.
- Not quite true in the data

Autocorrelation of Monthly Log House Price Changes



Autocorrelation Function for Monthly log House Price Changes. Two standard error bands around zero. 1987-2013

Testing for autocorrelation: Ljung and Box (1978)

Definition

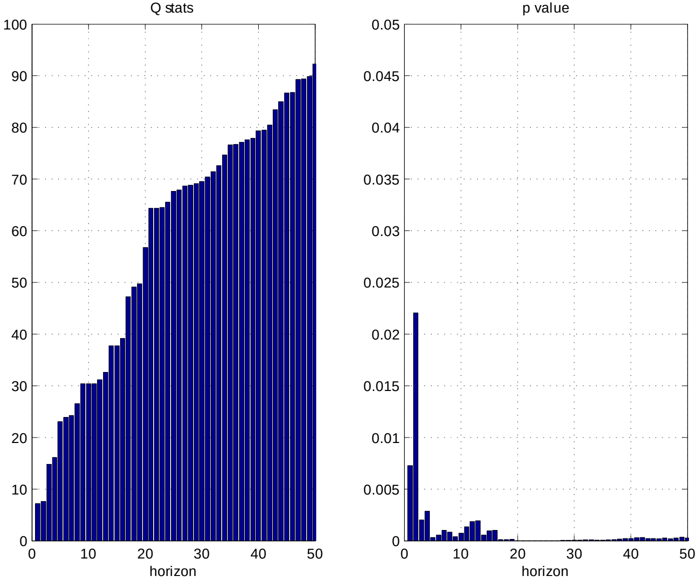
The **Ljung-Box** statistic tests the null that $H_0 : \rho_1 = \ldots = \rho_m = 0$

$$Q(m) = T(T+2) \sum_{i=1}^{m} \frac{\widehat{\rho}_i^2}{T-i}$$

$Q(m)$ is asymptotically χ^2 with m degrees of freedom.

- reject the null if $Q(m) > \chi^2(\alpha)$ where $\chi^2(\alpha)$ denotes the $(1-\alpha) \times 100$-th percentile

Q-test on Monthly Returns



Q-test for Monthly log Returns on VW-CRSP Index. 1926-2012

Monthly Stock Returns

- choice of m matters: rule of thumb $m = \ln(T)$
- if we use this rule of thumb, $m = 6$ and $Q(6) = 26$ and p-value is $1.9e-4$
- in any case, we reject the null that there is no autocorrelation in monthly U.S. stock returns for all holding periods considered..
- even though these autocorrelations are small, they're measured rather precisely, allowing us to reject the null.

White Noise

Definition

A time series ε_t is said to be **white noise** if $\{\varepsilon_t\}$ is a sequence of independent and identically distributed random variables.

Notation: $\varepsilon_t \sim \text{WN}\left(0, \sigma_\varepsilon^2\right)$

If ε_t is white noise + normally distributed with mean zero and variance σ^2, then it is called **Gaussian white noise**.

Notation: $\varepsilon_t \sim \text{GWN}\left(0, \sigma_\varepsilon^2\right)$

- There is no autocorrelation. ACF's are all zero