Workshop4. Autocorrelation

UNAM – FE   Econometrics I

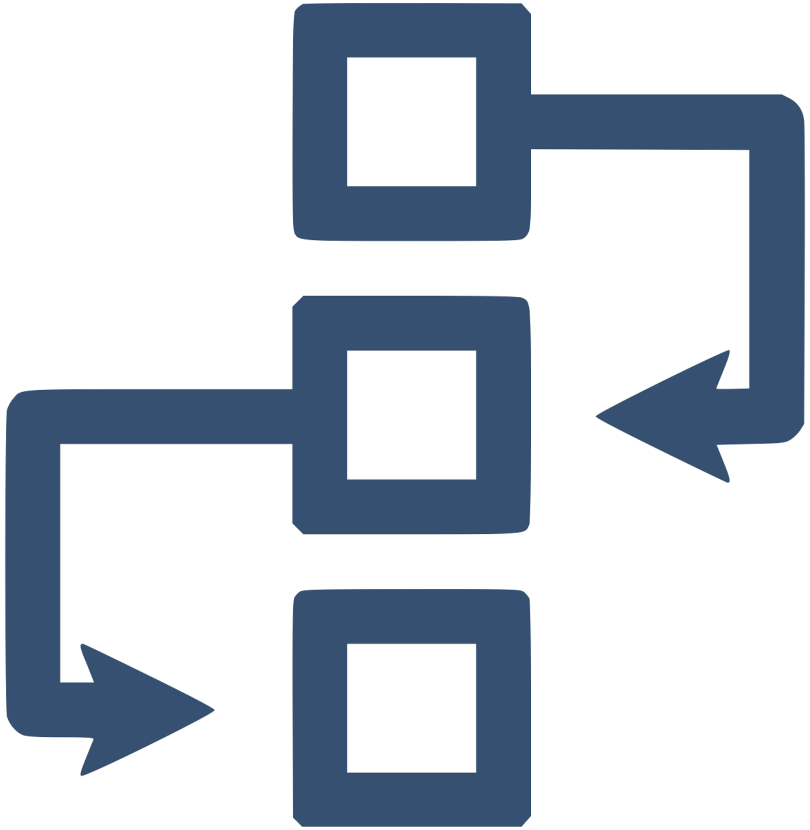
Esp. Humberto Acevedo
Assistant. Emilio Sandoval

March 22, 2022

Introduction

Autocorrelation is defined as *“the correlation among members in time series or cross-sectional data”*

It comes around when error terms in models are not independent from each other



It comes around when error terms in models are not independent from each other

In other words, when

$$E(u_i, u_j) \neq 0 \quad \text{for every } i \neq j$$

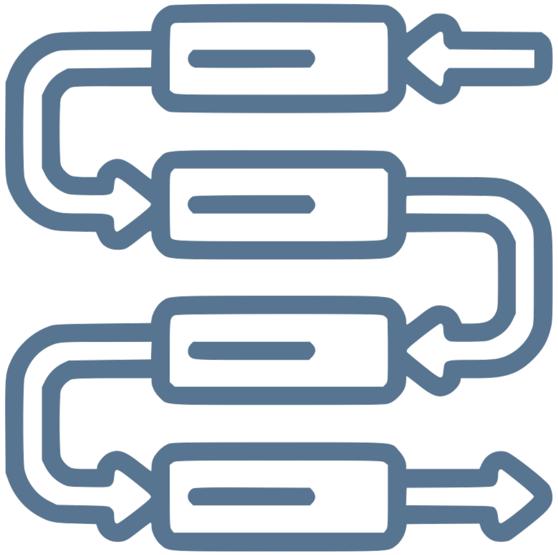
Then, errors are correlated!

Obtained OLS estimators under this circumstance are not efficient anymore.

There are **two types** of serial correlation:

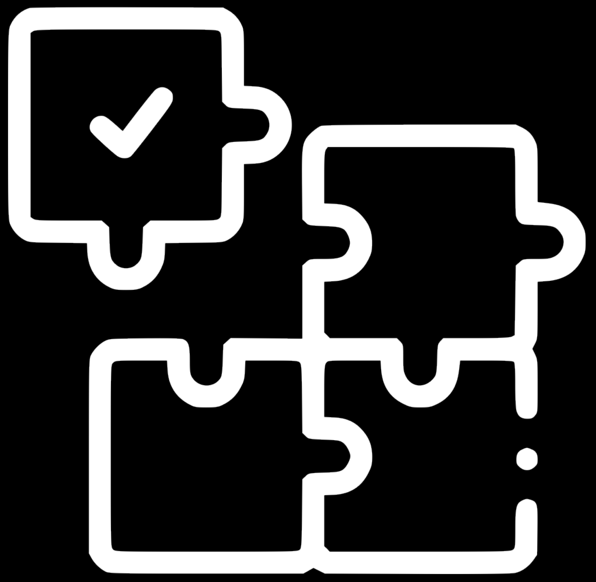
Impure serial correlation

Pure serial correlation



Serial **correlation**

Impure serial correlation



This results from a specification error due to the omission of relevant variables that shows autocorrelaton

Consider the well-specified model for y such that:

$$y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t$$

And if we mistakenly specify the model from this:

$$y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t$$

To this

$$y_t = \beta_1 + \beta_2 X_{2t} + u_t$$

Serial **correlation**

$$y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t$$

$$y_t = \beta_1 + \beta_2 X_{2t} + u_t$$

Serial **correlation**

$$y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t$$

$$y_t = \beta_1 + \beta_2 X_{2t} + u_t$$

Then the error term is:

$$e_t = \beta_3 X_{3t} + u_t$$

Serial **correlation**

$$e_t = \beta_3 X_{3t} + u_t$$

Note: if observations from this variable are dependent over time, then e_t will show autocorrelation

$ARMA(p,q)$ process

Consider the following model:

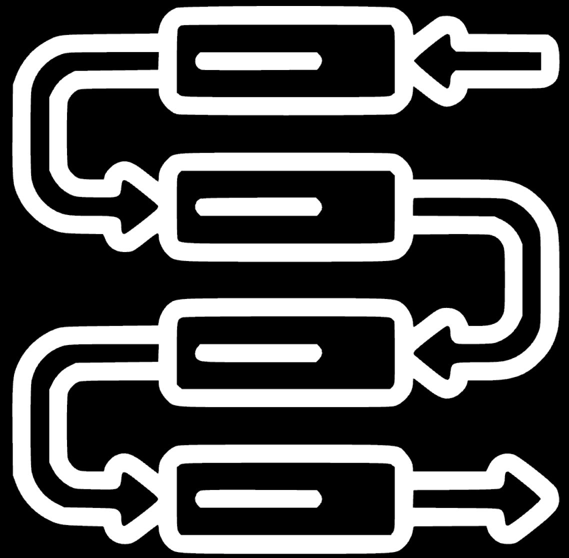
$$y_t = \beta_1 + \beta_2 X_{2t} + \cdots + \beta_k X_{tk} + u_t$$

Errors are strictly exogenous, $E(u) = 0$

Here errors may present an autoregressive process of order 1

$$u_t = \rho u_{t-1} + e_t \qquad t = 1,2,\ldots,n$$

Pure serial correlation

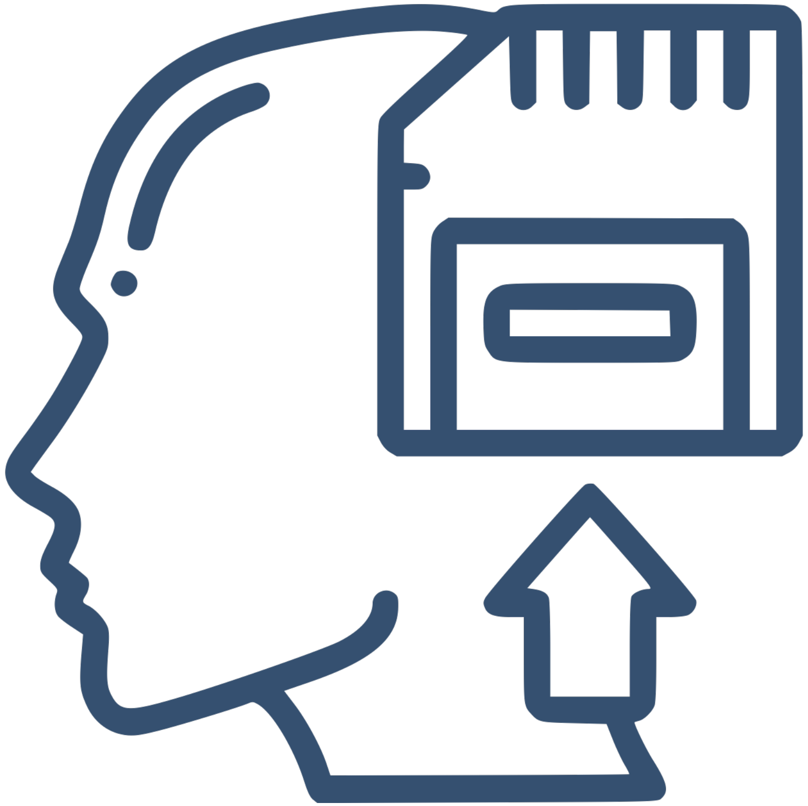


$$u_t = \rho u_{t-1} + e_t \qquad\qquad t = 1,2,\ldots,n$$

Where e_t are not correlated random variables with mean 0 and constant variance

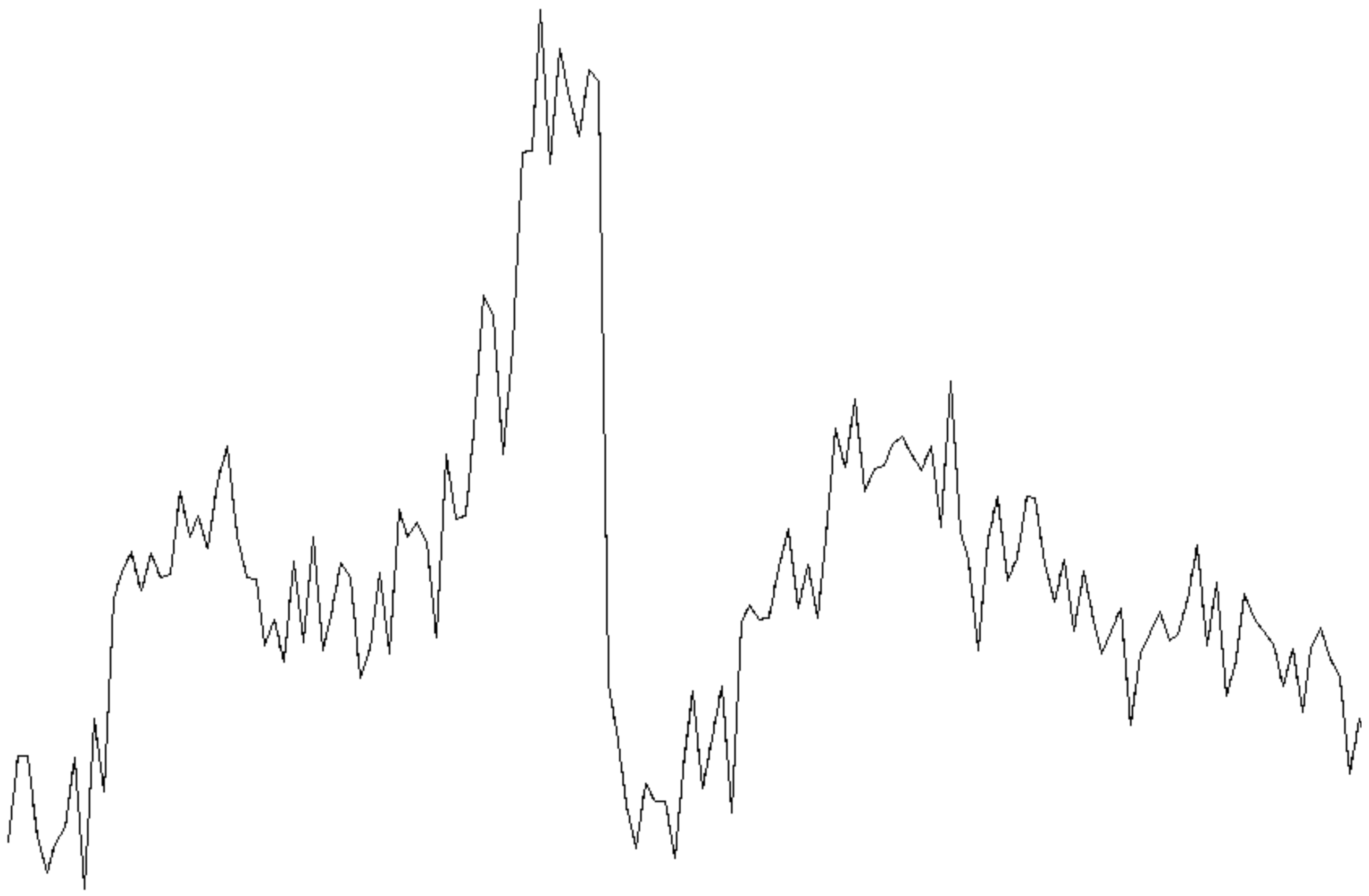
Autocorrelation problem is usually presented in historical data.

Memory is transferred through error term!



The problem relies in shocks

Shocks keep over time



Causes

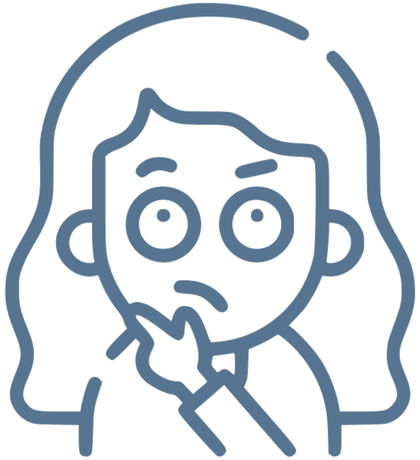
Inertia

There are strong trends that affect future series values

Specification biases

When we did not state correctly the functional form or when there are omitted variables, which generates a systematics behave in stochastic term

Time of adjustment



This implies that there is a time gap for economic individuals to process information, which takes 1 or 2 periods

Due to the fact that the Gauss-Markov theorem demands homoscedasticity as well as non-serial correlated errors, the OLS are not BLUE anymore in presence of this correlation

Consequences

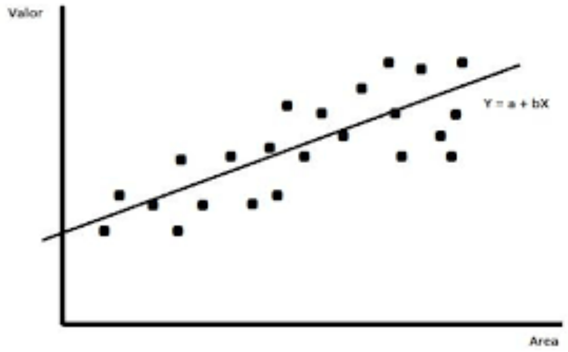
There are two consequence of this:

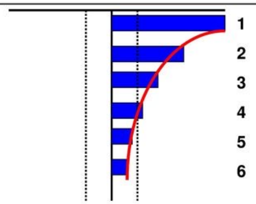
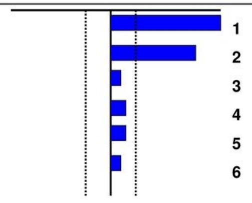
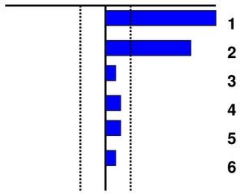
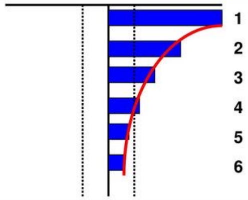
Low efficient estimators

Usual tests are invalid

To detect AR(1) processes we have two ways:

Graph methods



Proceso	FAC	FAP
AR(2)		
MA(2)		

Formal methods

- Durbin-Watson
- Graph test

Durbin-Watson test is based on serial autocorrelation $AR(1)$ in residuals obtained with a OLS regression

$$DW = \frac{\sum_{t=2}^{n}(\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^{n}\hat{u}_t^2}$$

Where t is the number of observations given that DW is approximated to $2(1-\rho_1)$ where ρ_1 is the sampling residual autocorrelation

Detection

Durbin-Watson test is takes values from 0 to 4 such that:

$$\rho_1 = 1 \Longrightarrow DW = 0$$

$$\rho_1 = 0 \Longrightarrow DW = 2$$

There is no autocorrelation in residuals

$$\rho_1 = -1 \Longrightarrow DW = 4$$

Detection

Durbin-Watson test is takes values from 0 to 4 such that:

If $\mathrm{DW} < 2$ then there is evidence of serial positive correlation

If $\mathrm{DW} > 2$ it indicates that values among residuals vary significantly

We use klein.dta which contains information about the level of government spending

To carry these test, it is necessary to declare the temporal variable in order to apply time series in STATA.

STATA COMMANDS

```
1. tsset yr
```

We run a regression of government spending which depends on government proceeds

STATA COMMANDS

```
2. regress consump wagegovt

3. predict u, resid
```

STATA COMMANDS (Durbin-Watson)

1. We calculate the numerator. 1 lag of residuals

```
gen u_1 = L.u
```

2. Calculate the temporal residual deviation and its lag

```
gen u_U1 = (u-u_1)
```

3. To the second power

```
gen u_u1sq = u_u1^2
```

4. Generate the denominator

```
gen usq = u^2
```

5. Generate the summatory

```
tabstat u_u1sq, stat(sum)
tabstat usq, stat(sum)
```

6. Generate the DW statistic

```
display 193.4684/601.2072
```

Consequences

Critical values that are superior to lower and upper d_U y d_L were calculated by Durbin and Watson for different values of k (number of independent variables) and n

If $\mathrm{DW} < d_L$ reject null hypothesis

If $\mathrm{DW} > d_U$ do not reject null hypothesis

If $\mathrm{d}L < DW < d_U$ no conclussion

Durbin-Watson Table

	k'=1		k'=2		k'=3		k'=4		k'=5		k'=6		k'=7		k'=8		k'=9		k'=10	
n	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU
6	0.610	1.400	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
7	0.700	1.356	0.467	1.896	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
8	0.763	1.332	0.559	1.777	0.367	2.287	---	---	---	---	---	---	---	---	---	---	---	---	---	---
9	0.824	1.320	0.629	1.699	0.455	2.128	0.296	2.588	---	---	---	---	---	---	---	---	---	---	---	---
10	0.879	1.320	0.697	1.641	0.525	2.016	0.376	2.414	0.243	2.822	---	---	---	---	---	---	---	---	---	---
11	0.927	1.324	0.758	1.604	0.595	1.928	0.444	2.283	0.315	2.645	0.203	3.004	---	---	---	---	---	---	---	---
12	0.971	1.331	0.812	1.579	0.658	1.864	0.512	2.177	0.380	2.506	0.268	2.832	0.171	3.149	---	---	---	---	---	---
13	1.010	1.340	0.861	1.562	0.715	1.816	0.574	2.094	0.444	2.390	0.328	2.692	0.230	2.985	0.147	3.266	---	---	---	---
14	1.045	1.350	0.905	1.551	0.767	1.779	0.632	2.030	0.505	2.296	0.389	2.572	0.286	2.848	0.200	3.111	0.127	3.360	---	---
15	1.011	1.361	0.946	1.543	0.814	1.750	0.685	1.977	0.562	2.220	0.447	2.471	0.343	2.727	0.251	2.979	0.115	3.216	0.111	3.438
16	1.106	1.371	0.982	1.539	0.857	1.728	0.734	1.935	0.615	2.157	0.502	2.388	0.398	2.624	0.304	2.860	0.222	3.090	0.155	3.304
17	1.133	1.381	1.015	1.536	0.897	1.710	0.779	1.900	0.664	2.104	0.554	2.318	0.451	2.537	0.356	2.757	0.272	2.975	0.198	3.184
18	1.158	1.391	1.046	1.535	0.933	1.696	0.820	1.872	0.710	2.060	0.603	2.258	0.502	2.461	0.407	2.668	0.321	2.873	0.244	3.073
19	1.180	1.401	1.074	1.536	0.967	1.685	0.859	1.848	0.752	2.023	0.649	2.206	0.549	2.396	0.456	2.589	0.369	2.783	0.290	2.974
20	1.201	1.411	1.100	1.537	0.998	1.676	0.894	1.828	0.792	1.991	0.691	2.162	0.595	2.339	0.502	2.521	0.416	2.704	0.336	2.885
21	1.221	1.420	1.125	1.538	1.026	1.669	0.927	1.812	0.829	1.964	0.731	2.124	0.637	2.290	0.546	2.461	0.461	2.633	0.380	2.806
22	1.239	1.429	1.147	1.541	1.053	1.664	0.958	1.797	0.863	1.940	0.769	2.090	0.677	2.246	0.588	2.407	0.504	2.571	0.424	2.735
23	0.126	0.144	0.117	0.154	1.018	0.166	0.986	1.785	0.895	1.920	0.804	2.061	0.715	2.208	0.628	2.360	0.545	2.514	0.465	2.610
24	1.273	1.446	1.188	1.546	1.101	1.656	1.013	1.775	0.925	1.902	0.837	2.035	0.750	2.174	0.666	2.318	0.584	2.464	0.506	2.613
25	1.288	1.454	1.206	1.550	1.123	1.654	1.038	1.767	0.953	1.886	0.868	2.013	0.784	2.144	0.702	2.280	0.621	2.419	0.544	2.560

STATA COMMANDS

We can obtain Durbin Watson by…

```
4. estat dwatson
```

This test is valid if and only if:

- Regression equation contains a constant term
- Regression equation does not contain among independent variables any lag from the dependent variable
- Error autocorrelation are $AR(1)$

Alternative Durbin Watson test incorporates lagged dependent variables that are included as regressors such that past error terms are correlated with those of lagged variables in time t, thus those regressors are not strictly exogenous!

Incorporation of covariances that are not strictly exogenous may cause that statistic DW would be biased provoking to wrongly accept null hypothesis

STATA COMMANDS

To apply this test, it is important to estimate the regression.

Thus, we apply the `estat durbinalt` command with the option small due to the size of sample.

```
5. regress consump wagegovt
6. estat durbinalt, small
```

On the other hand, Breusch-Pagan test stands for those inconvenients presented in DW. It allows:

- Non-Stochastic regressors and lagged values from dependent
- Autoregressive schemes such as $AR(1)$, $AR(2)$, etc..
- Simple Moving Averages or from superior order greater than white noise error terms

To apply this test, it is important to estimate the regression again

Thus, we apply the `estat bgodfrey` command with the option small due to the size of sample.

STATA COMMANDS

```
7. regress consump wagegovt
8. Estat bgodfrey, small
```

How to fix regression?

Both previous test reject null hypothesis of non-serial correlation, so we will readjust the model incorporating two lags over consumption as regressors

We apply alternative Durbin-Watson because we can no longer apply DW

Then we apply Breusch-Godfrey

What happened with statistic significance of the parameter over government income?

STATA COMMANDS

```
9.  regress consump wagegovt L.consump L2.comsump
10.estat durbinalt, small lags(1/2)
11.estat bgodfrey, small lags (1/2)
```

References

- **Salvatore, D., & Sarmiento, J. C.** (1983). *Econometría* (No. HB141 S39). McGraw-Hill.
- **Gujarati, D. N.** (2009). Basic econometrics. Tata McGraw-Hill Education.
- **Wooldridge, J.M.** (2016). *Introductory Econometrics*, Cengage Learning, 6th edition.
- **CFA Institute** (2020), "Level I, Volume 1, 2020, Ethical and Professional Standards and Quantitative Methods; Reading 7: Statistical Concepts and Market Returns", pp. 422-430