Solution to Exercise 2

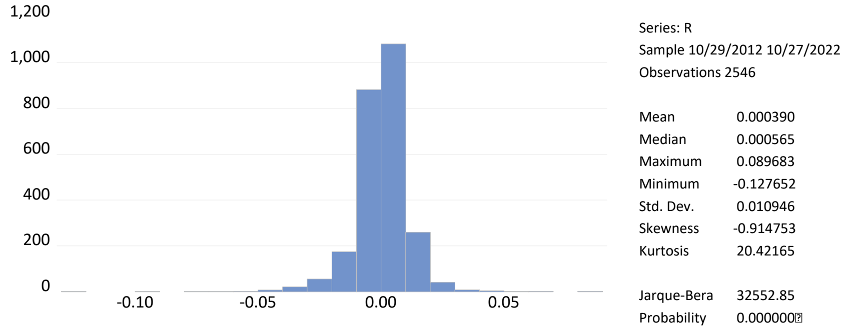
Simon A. Broda

1. (a) The first thing to observe is the difference between the population quantities (parameters) μ and σ, and their estimates $\bar{y}$ and s_y, which are sample quantities. The latter will generally be close to the former because of the law of large numbers, but not the same. The estimates also change every time new random numbers are drawn. The time series plot shows that the observations are randomly scattered around 0. The autocorrelations in the correlogram will mostly be insignificant, i.e., statistically indistinguishable from zero. Formally, the hypothesis being tested is $H_0 : \tau_s = 0$ vs. $H_a : \tau_s \neq 0$. Even if H_0 is true here, there is still a 5% probability that any given autocorrelation will be significant; this is called the type-1 error: the probability of rejecting the null even though it is true. A similar statement concerns the Q-statistics, which test whether the first m autocorrelations are all equal to zero[1]. **Important**: make sure to look at the formulas behind the cells and make sure you understand how they work; specifically, the correlogram, the Q-stats, and their respective critical values. You don't need to understand how the random numbers u_t themselves are generated (they use a trick called inverse transform sampling).

   (b) The time series plot looks very different from that in the other sheet, because a random walk is not mean reverting. Also, the correlogram and the Q-stats are now highly significant, so that we (correctly) reject the null that the data were generated from a white noise process. In fact, the slow and almost linear decay of the correlogram suggests (correctly) that the data were generated by an integrated process. **Important**: make sure you understand how the simulated random walk y_t is constructed, by always taking "yesterday's" value y_{t-1} and adding u_t to it.

2. (a) The command `genr logsp500 = log(sp500)` generates the log prices. From these, the returns can be obtained using the command `genr r = d(logsp500)`; the `d` stands for the first difference Δ. Alternatively, the returns can be obtained from the prices directly using the command `genr r = dlog(sp500)`. The time series plots can be generated like in the exercises for Week 1, resulting in the plots shown in the slides. It is immediately evident that the plot for the returns is very different from that of the (log) prices. This is because the (log) prices are integrated, while returns are stationary. The histogram is obtained by double-clicking the return series, and then clicking `View`→`Descriptive Statistics and Tests`→`Histogram and Stats`. This results in the plot below.

[1]Formally, $Q(m)$ can be used to test $H_0 : \tau_1 = \tau_2 = \ldots = \tau_m = 0$.



(b) The output shows a skewness of SK $= -0.914753$, and a kurtosis of K $= 20.42165$, whereas for a normal distribution, one would expect these values to be close to 0 and 3. This shows that the returns are left-skewed and have heavy tails, a very common finding. Inserting these values, together with $T = 2,546$, into the formula

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

yields a Jarque-Bera statistic of JB $= 32,552.84$ (up to rounding error, this value is of course already given in the output). This value is much larger than the 5% critical value of the χ^2_2 distribution, 5.991, so we strongly reject the null that the data were drawn from a normal distribution. We could also have concluded this directly from the p-value given in the output, but that might not be available in an exam...

(c) The correlogram is obtained by double-clicking the return series, and then clicking `View→Correlogram...`. The window that appears allows you to choose the number of lags. Arbitrarily choosing 20 produces the plot below.

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

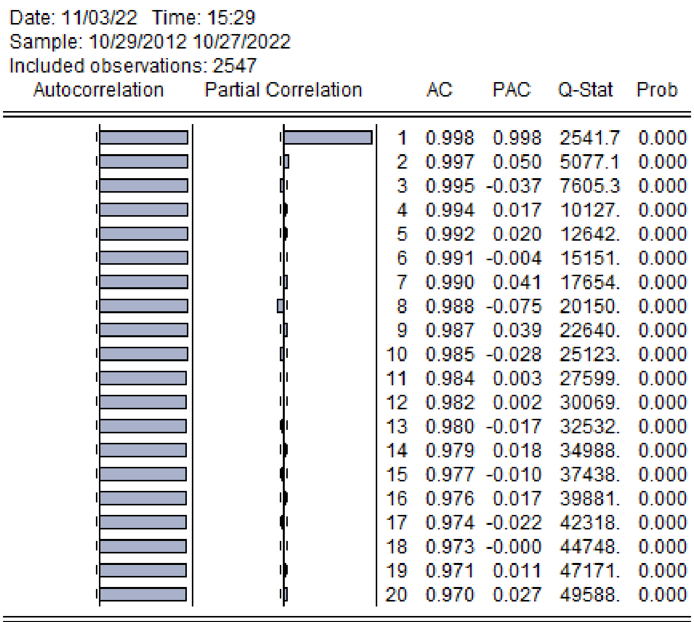
The correlogram doesn't show any indication of non-stationarity, so the data are likely stationary. They don't appear to have been generated by a pure white noise process though, because the autocorrelations and lags 1, 2, 5–10, and 13–17 are significant at the 5% level; their absolute values exceed the little line representing the critical value $1.96/\sqrt{2546} = 0.0388$.

(d) The test statistic for testing $H_0 : \tau_1 = \cdots = \tau_{10} = 0$, vs. the alternative that at least one of them is non-zero, is

$$Q(10) = T(T+2)\sum_{s=1}^{10} \frac{\hat{\tau}_s^2}{T-s} = 290.04 \sim \chi^2_{10}.$$

The observed value is 290.04, much larger than the critical value 18.307. Thus, we reject the null and conclude that at least one of the first 10 autocorrelations is different from zero.

(e) The correlogram of `logsp500` looks as follows.



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

Autocorrelation	Partial Correlation		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

The autocorrelations are very high, and appear to decay linearly, not exponentially. This is a clear sign the process that generated the data is integrated. It is not a pure random walk though; for a random walk, we would expect that the sample PACF on the right becomes insignificant after the first lag; here, some partial autocorrelations are significant. This, of course, corresponds to the fact that the returns are not pure white noise, as we determined earlier.

3. (a) By repeatedly plugging in,

$$\begin{aligned} Y_t &= Y_{t-1} + U_t \\ &= Y_{t-2} + U_{t-1} + U_t \\ &= Y_{t-3} + U_{t-2} + U_{t-1} + U_t \\ &\vdots \\ &= Y_0 + \sum_{s=1}^{t} U_s \end{aligned}$$

as claimed.

(b) The result from the previous question implies that

$$
\begin{aligned}
\mathbb{E}\left[Y_t\right] &= \mathbb{E}\left[Y_0 + \sum_{s=1}^{t} U_s\right] \\
&= Y_0 + \sum_{s=1}^{t} \mathbb{E}[U_s] = Y_0.
\end{aligned}
$$

Here, we have used that the expectation of a sum is the sum of the expectations, together with the fact that Y_0 is assumed to be a constant, and that $\mathbb{E}[U_t] = 0$ because U_t is white noise. For the variance, recall that in general, $\text{var}(X+Y) = \text{var}(X) + \text{var}(Y) + 2\,\text{cov}(X,Y)$. But since the U_t are all independent, we have that $\text{var}(U_s + U_t) = \text{var}(U_s) + \text{var}(U_t) + 0 = 2\sigma^2$. Thus

$$
\begin{aligned}
\text{var}\left[Y_t\right] &= \text{var}\left[Y_0 + \sum_{s=1}^{t} U_s\right] \\
&= \text{var}\left[\sum_{s=1}^{t} U_s\right] \\
&= \sum_{s=1}^{t} \text{var}(U_s) \\
&= \sum_{s=1}^{t} \sigma^2 \\
&= t \cdot \sigma^2.
\end{aligned}
$$