TU Dortmund University
Department of Statistics
Summer 2023

M.Sc. Karsten Reichold
M.Sc. Sven Pappert

Case Studies – Project 1
Forecasting Bitcoin

Deadline: May 16, 2023, at 2:00 pm
Submission: Submit your report (together with corresponding `R` code in a separate file) electronically via Moodle **and** as a printed version (without `R` code) in the course.

This report is concerned with forecasting Bitcoin, to be more precise, with forecasting the growth rate of Bitcoin. Let y_t denote the value of Bitcoin (in U.S. dollar) at the end of quarter t. The quarterly growth rate of y_t (expressed in percentages) is defined as

$$g_{y,t} := \frac{y_t - y_{t-1}}{y_{t-1}} \times 100, \tag{1}$$

$t = 2, \ldots, T$.[1]

Tasks (First Part):

(a) Discuss why the US unemployment rate, the US inflation rate, the federal funds rate and the growth rate of the S&P 500 could affect Bitcoin growth.

(b) Generate a plot that shows Bitcoin's growth rate and the four potential drivers (at quarterly frequency) over the longest time period available.

You may use the following data sources: https://coinmarketcap.com/currencies/bitcoin/ for Bitcoin and https://research.stlouisfed.org/econ/mccracken/fred-databases/ for the other four variables.

(c) Create one-quarter-ahead forecasts of Bitcoin growth for this time period using an autoregressive model of order one (i. e., an AR(1) model). Illustrate the forecasts together with the actual growth rates in one plot. Calculate the root mean squared forecasting error over the whole period.[2] Here and in the following you can set all required starting values equal to zero.

(d) Create one-quarter-ahead forecasts of Bitcoin growth for the same period using a vector autoregressive model of order one (i. e., a VAR(1) model) containing the four additional variables. Illustrate the forecasts together with the actual growth rates and the forecasts based on the AR(1) model in one plot, potentially extending the plot in (c). Calculate the root mean squared forecasting error over the whole period.

(e) Explain the concept of Granger causality (compare the discussion in Kilian and Lütkepohl, 2017, Chapters 2.5 and 7) and identify the variables in the VAR(1) model from (d) that Granger cause Bitcoin growth.

---

[1] This definition uses the previous quarter (i. e., y_{t-1}) as a reference point. An alternative definition uses the same quarter in the previous year (i. e., y_{t-4}) as a reference point. You could discuss advantages and disadvantages of either definition in your report.

[2] Using generic notation, let $\hat{z}_{t+1|t}$ be a forecast of the variable z at time $t+1$ given the information up to time t and let z_{t+1} denote the actual value of z at time $t+1$. The mean squared forecasting error over the period $t = 1, \ldots, T$ is then defined as $\text{MSFE}_z := \sqrt{T^{-1} \sum_{t=1}^{T-1} (z_{t+1} - \hat{z}_{t+1|t})^2}$.

(f) Create one-quarter-ahead forecasts of Bitcoin growth using a VAR(p) model containing the four additional variables. To determine p use Akaike's (1974) information criterion as described in Kilian and Lütkepohl (2017, p. 55) exploiting the whole sample.[3] Illustrate the forecasts together with the actual growth rates and the forecasts based on the AR(1) and VAR(1) models in one plot, potentially extending the plot in (d). Calculate the root mean squared forecasting error over the whole period.

(g) Compare the performance of the different forecasting approaches. Does it suffice to merely fit an AR(1) model? In which periods is it particularly advantageous to exploit information in other variables? Are there any periods in which the forecasts fit poorly to the actual observations? Are these periods potentially related to "economic events"? Is there any structure in the data not captured by the methods used? How could you account for these features?

Familiarize yourself with the concept of "unrestricted mixed-data sampling" (unrestricted MIDAS) as proposed in Foroni *et al.* (2015).[4] Download the US unemployment rate, the US inflation rate, the federal funds rate and the growth rate of the S&P 500 at monthly frequency. Let X_t denote a vector of the four variables observed at monthly frequency $t = 1, 2, 3, \ldots, 3T-2, 3T-1, 3T$ (potentially calculating monthly growth rates, similarly as in (1))and let $g_{y,t}$ denote Bitcoin growth observed at quarterly frequency, i. e., every third month (at $t = 3, 6, \ldots, 3T$). This notation implies that the first quarter in the sample contains the months $t = 1, 2, 3$, the second quarter contains the months $t = 4, 5, 6$ and so on. T denotes the number of full quarters in the sample. We are interest in forecasting Bitcoin growth in quarters $1 \leq s^* \leq T$ using the information at monthly frequency up to time $t = 3s^* - 1$, i. e., until one month prior to the end of quarter s^*.[5]

To this end, consider the model

$$g_{y,3s} = \mu_0 + \mu_1 g_{y,3(s-1)} + \Phi_0' X_{3s-1} + \Phi_1' X_{3s-2} + \ldots + \Phi_K' X_{3s-(K+1)} + \varepsilon_{3s}, \tag{2}$$

for $s = 1, \ldots, s^*-1$ and estimate the coefficients μ_0 and μ_1 and the four-dimensional coefficient vectors $\Phi_0, \Phi_1, \ldots, \Phi_K$ by OLS to obtain the estimates $\hat{\mu}_{0|3(s^*-1)-1}$, $\hat{\mu}_{1|3(s^*-1)-1}$, $\hat{\Phi}_{0|3(s^*-1)-1}$, $\hat{\Phi}_{1|3(s^*-1)-1}$, $\ldots$, $\hat{\Phi}_{K|3(s^*-1)-1}$. The forecast of g_{y,s^*} based on model (2) is then defined as

$$\begin{aligned}\hat{g}_{y,3s^*} = {} & \hat{\mu}_{0|3(s^*-1)-1} + \hat{\mu}_{1|3(s^*-1)-1} g_{y,3(s^*-1)} \\ & + \hat{\Phi}'_{0|3(s^*-1)-1} X_{3s^*-1} + \hat{\Phi}'_{1|3(s^*-1)-1} X_{3s^*-2} + \ldots + \hat{\Phi}'_{K|3(s^*-1)-1} X_{3s^*-(K+1)}.\end{aligned}$$

Tasks (Second Part):

(h) Forecast Bitcoin growth for the whole period using model (2). Determine K, with $1 \leq K \leq 3$, using AIC exploiting the whole sample.

(i) Remove those lags of the variables in X_t whose coefficients are not significantly different from zero and repeat task (h) with the reduced model.

(j) Compare the forecast adequately by calculating RMSFEs for the whole period and creating meaningful graphical illustrations.

(k) Does employing monthly observations help to improve the forecasts of Bitcoin growth?

(l) Finally, consider also Bitcoin at monthly frequency and and repeat task (f) with all variables now at monthly frequency. Restrict the maximum number of lags to twelve. What do you notice?

[3] Restrict the maximum number of lags to three.

[4] If you do not have access to the paper, you can download the working paper version here.

[5] This is one example of "nowcasting". Please note that Bitcoin is available at even higher frequencies (e. g., daily). However, for didactic reasons, we assume in here that Bitcoin is only available at quarterly frequency and will be released directly at the beginning of each quarter.

(m) Pick one variable in X_t that seems to help forecasting Bitcoin growth and provide a brief evaluation why the connection between this variable and Bitcoin growth is plausible also from an economic or financial perspective. You should use at least one reference from the economic or financial literature on cryptocurrencies. (Compared to task (a), here you have a posteriori knowledge.)

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control* **19**, 716 – 723.

Foroni, C., Marcellino, M., Schumacher, C. (2015). Unrestricted Mixed Data Sampling (MIDAS): MIDAS regressions with unrestricted lag polynomials. *Journal of the Royal Statistical Society A* **178**, 57–82.

Kilian, L., Lütkepohl, H. (2017). Structural Vector Autoregressive Analysis. Cambridge University Press, Cambridge.