Methods To Predict the Price of Cryptocurrencies

TCF

0. Introductory Definitions

A **time series** is a sequence of numerical data points taken in successive order on any variable that changes over time. For example, in investing, a time series is used to track the movement of the price of assets. There is no minimum or maximum amount of time that must be included neither is there a fixed step length. This allows for a large flexibility in how the data can be gathered.

The **lag operator** of a time series is a function which offsets a time series such that the 'lagged' values align with the actual time series. An example of the lag operator acting on a time series with lags of 1 and 2 is shown below:

Date	V_t	V_{t-1}	V_{t-1}
01/01/2020	0	NA	NA
02/01/2020	1	0	NA
03/01/2020	2	1	0
04/01/2020	3	2	1
05/01/2020	4	3	2

In economic models, an **exogenous variable** is one whose value is determined outside the model and is imposed on the model, an exogenous change is a change in an exogenous variable. Whereas, an **endogenous variable** is a variable whose value is determined by the model. An endogenous change is a change in an endogenous variable in response to an exogenous change that is imposed upon the model.

If a time series is **stationary**, intuitively, the statistical properties of a process generating the time series do not change over time. It does not mean that the time series foes not change over time, just that the way it change does not itself change over time.

1. Vector Autoregression

Intuition

A model, such as VAR, is said to be **autoregressive** if the predictions for future values of a time series are calculated using a linear function of previous values. That is, when using autoregression, time series are modelled as linear combinations of their own lags. Hence, a typical equation for an autoregression model of order p, AR[p], would be

$$Y_t = \alpha + \beta_1 Y_{t-1} + ... + \beta_p Y_{t-p} + \epsilon_t,$$

where α is the intercept, the β_i are the coefficients of the lags (up to order p) and ϵ_t is the error, modelled as white noise (to help the model account for volatility). [1]

Vector autoregression is a multivariate forecasting algorithm which is used when two or more time series influence one another. This means that time series are modelled as linear combinations of past values of themselves and other time series in the system. Each time series has a separate equation used to model predictions. For example, a system of equations for a VAR[2] model (a model of order 1) with two time series given by Y_1 and Y_2 is

$$Y_{1,t} = \alpha_1 + \beta_{11,1} Y_{1,t-1} + \beta_{12,1} Y_{2,t-1} + \epsilon_{1,t},$$
$$Y_{2,t} = \alpha_2 + \beta_{21,1} Y_{1,t-1} + \beta_{22,1} Y_{2,t-1} + \epsilon_{2,t}.$$

Since the terms in the equations are interrelated, the values are considered to be endogenous variables, rather than exogenous predictors.

Building

The following steps describe how to build a VAR model: [2]

1. Analyse time series characteristics - visualise the time series by plotting them, generally we are looking to check the time series have similar characteristics.
2. Test for causation amongst the time series - the time series must have causation or VAR will not be an appropriate model. To test for causation use Granger's Causality test and the cointegration test.
3. Test for stationarity - to test for stationarity we can use Augmented Dickey-Fuller test (ADF). If a series is found to be non-stationary we must transform it by differencing the series until it becomes stationary.
4. Find optimal order [p] - iteratively fit increasing orders of VAR models and pick the model with the least AIC (Akaikes Information Criteria). We can also use other best fit comparison estimates of BIC, FPE and HQIC.
5. Prepare training and test datasets.
6. Train the model.
7. Check for serial correlation of residuals - if there is any we have some pattern in the time series yet to be accounted for by the model so the order of the model should be increased and we should retrain.
8. Roll back transformations, if any, to obtain real forecast.
9. Evaluate the model using test set - plot actual values vs predicted and use MAPE, ME, MAE, MPE, RMSE, corr and minmax to determine the accuracy of model.

[1]