On multi-zoom autoregressive time series models

Piotr Fryzlewicz

In classic linear time series autoregression (AR), the univariate time series X_t under consideration is modelled as a linear but otherwise unconstrained function of its own past values $X_{t-1}, X_{t-2}, \ldots$, plus white-noise-like innovation ε_t. That is,

$$X_t = a_1 X_{t-1} + \ldots + a_p X_{t-p} + \varepsilon_t. \tag{1}$$

In some situations, it appears to be a good idea to model X_t as depending explicitly on some other features of its own past, rather than on the individual variables $X_{t-1}, \ldots, X_{t-p}$.

As an example, consider the problem of modelling mid- and high-frequency financial returns, where X_t represents a fine-scale, e.g. one-minute, return on a financial instrument. In the hope of improving the predictive power, the analyst may wish to model X_t as depending not only on the past few one-minute returns, but also perhaps on past returns on lower frequencies, such as one hour or one day. Representing this in an unconstrained way as in (1) with a large value of p would lead to obvious over-parameterisation.

Our proposed way to resolve this issue is to adopt what we call a "multi-zoom" approach to time series analysis. The main idea of the approach is to include as regressors for X_t features of the path $X_1, \ldots, X_{t-1}$ which "live" on multiple time-scales, and hence correspond to considering the time series at different zoom levels.

For example, in the financial time series context described above, we could entertain a multi-zoom AR model of the form

$$X_t = \alpha_1 \frac{1}{\tau_1}(X_{t-1} + \ldots + X_{t-\tau_1}) + \ldots + \alpha_p \frac{1}{\tau_p}(X_{t-1} + \ldots + X_{t-\tau_p}) + \varepsilon_t, \tag{2}$$

where the time scales τ_k are such that $1 = \tau_1 < \tau_2 < \ldots < \tau_p$. Note that $X_{t-1} + \ldots + X_{t-\tau_k}$ represents the most recent τ_k-minute return. There is nothing to stop τ_k, $k > 1$, from being large, e.g. of the order of tens or hundreds. The number of scales p would typically be much smaller than the longest time scale τ_p (note that the standard AR model (1) can always be rewritten in the form (2) if we take $\tau_p = p$). Including the regressors $X_{t-1} + \ldots + X_{t-\tau_k}$, rather than the individual variables X_{t-s}, corresponds to "zooming out" of the original time scale on which the data were collected, and explicitly incorporating information from coarser time scales. In this instance, the returns $X_{t-1} + \ldots + X_{t-\tau_k}$ represent the multi-zoom "features" that we believe have some predictive power with respect to X_t.

The following questions are of immediate methodological interest:

- *Model identification and stationarity.* We note that the multi-zoom AR model in equation (2) is a particular, sparsely parameterised, instance of the AR(τ_p) model. Therefore, stationarity (or otherwise) of multi-zoom AR can be established via the usual route for AR processes.

- *Estimation of p, τ_k and α_k.* In the simplest case, the values of p and $\{\tau_k\}_{k=1}^p$ are chosen by the analyst, and only the coefficients $\{\alpha_k\}_{k=1}^p$ need to be estimated. This can be done e.g. via OLS, or by performing an unconstrained estimation for AR(τ_p) and then grouping the estimated coefficients into sections of piecewise constancy. If $\{\tau_k\}_{k=1}^p$ are unknown, the grouping can be achieved via change-point detection techniques. If p is also unknown, change-point detection needs to be coupled with devices for model choice based e.g. on thresholding or on the use of information criteria.
- *Use of other multi-zoom features.* It is of interest to generalise model (2) to other multi-zoom features, for example the wavelet coefficients of the original price process at different scales, or nonlinear breakout-type statistics (the latter being of interest in e.g. algorithmic trading). The introduction of non-linearity introduces particularly challenging methodological questions of model identifiability and estimation. Note that the linear dependence on non-linear features that this induces goes in the opposite direction to the non-linear dependence on linear features seen, for example, in Generalised Linear Models.
- *Applicability in financial statistics.* Preliminary results suggest that multi-zoom AR processes are good at explaining the apparent lack of serial dependence in time series of financial returns, when measured via the sample autocorrelation, which can be blind to multi-scale dependencies such as those in (2) due to its single-scale nature. Moreover, empirically, multi-zoom AR processes appear to have relatively good predictive power for forecasting high- and mid-frequency financial returns.

The fact that multi-zoom AR processes can "mask" as white noise from the point of view of the sample autocorrelation (and hence be potentially be attractive from the point of view of modelling financial returns, which tend to exhibit this empirical feature) is illustrated in Figure 1. Despite the model being far from white noise, the sample autocorrelation fails to detect the serial dependence in the process, which is in part due to the fact that this measure takes no account of the multi-zoom structure of the model.

We are grateful to the workshop participants for pointing us to some other related literature, and in particular to the models described in [3], [1], [2]. We emphasize again that in contrast to these, our approach enables, in particular, automatic selection of the relevant time-scales τ_k. This also sets it apart from the autoregressive index models in [4].

References

[1] F. Corsi, *A simple approximate long-memory model of realized volatility*, Journal of Financial Econometrics **7** (2009), 174-196.

[2] S. Heston, R. A. Korajczyk and R. Sadka, *Intraday patterns in the cross-section of stock returns*, Journal of Finance **65** (2010), 1369–1407.

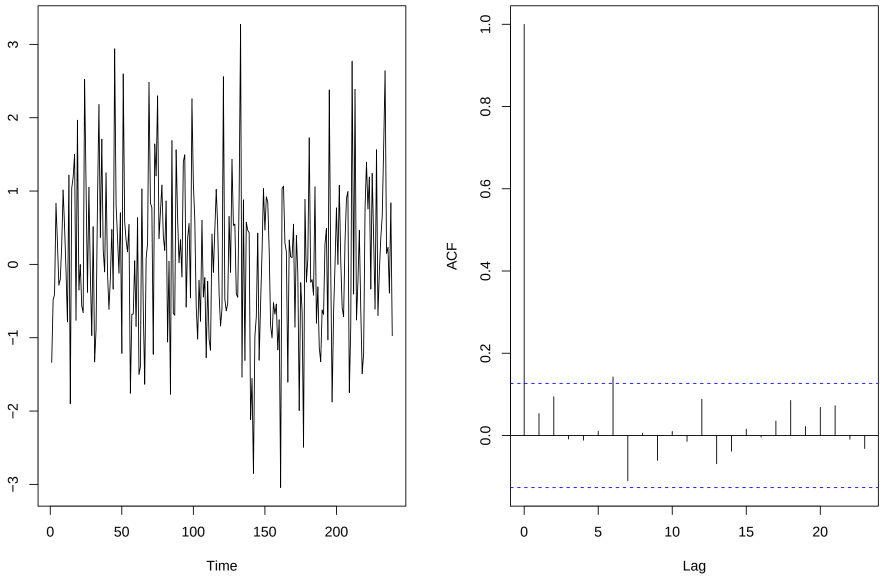

FIGURE 1. Left: sample path simulated from model (2) with length $n = 250$, $p = 2$, $\tau_1 = 1$, $\tau_2 = 10$, $\alpha_1 = 0.1$, $\alpha_2 = 0.5$, ε_t iid standard normal. Right: the sample autocorrelation of the simulated sample path.

[3] U. Müller, M. Dacorogna, R. Dav, R. Olsen, O. Pictet, and J. von Weizsacker, *Volatilities of different time resolutions - Analysing the dynamics of market components*, Journal of Empirical Finance **4** (1997), 213-239.

[4] G. Reinsel, *Some results on multivariate autoregressive index models*, Biometrika **70** (1983), 145–156.