Modelling multivariate financial returns using changepoint-induced multiscale bases

Piotr Fryzlewicz   Anna Schröder
{p.fryzlewicz,a.m.schroeder}@lse.ac.uk

Department of Statistics, London School of Economics, UK

BIRS, 1 May 2014

'Adaptive' bases in TSA – motivation no. 1

'Classical' time series analysis typically uses fixed decomposition bases for the spectral analysis of time series. This also applies to locally stationary models. Some examples are below.

- Fourier-Cramer representation:

$$X_t = \int_{-1/2}^{1/2} A(\omega) \exp(i2\pi\omega t) dZ(\omega)$$

- 'Fourier' local stationarity:

$$X_{t,T} = \int_{-1/2}^{1/2} A_T(\omega, t) \exp(i2\pi\omega t) dZ(\omega)$$

- 'Wavelet' local stationarity:

$$X_{t,T} = \sum_{j,k} W_j(k/T) \psi_{j,t-k} \xi_{j,k}$$

'Adaptive' bases in TSA – motivation no. 1

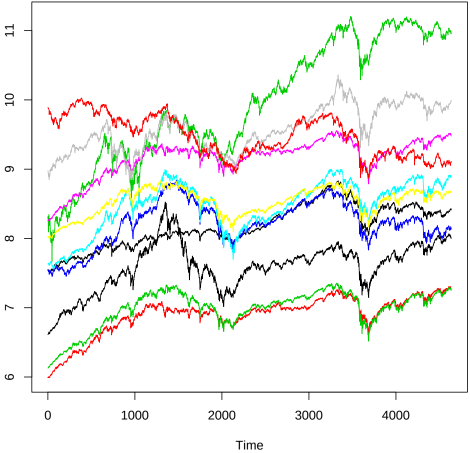
It is tempting to entertain the thought of selecting decomposition bases from the data. We refer to them as 'adaptive' bases in this talk.

Questions:

- What useful purpose would this serve, and for what types of data?
- What dictionaries of bases are worth considering, and how to select the basis to be used?

'Adaptive' bases in TSA – motivation no. 2

Logged daily closing values of 11 major stock indices, from 3 January 1995 to 26 October 2012.



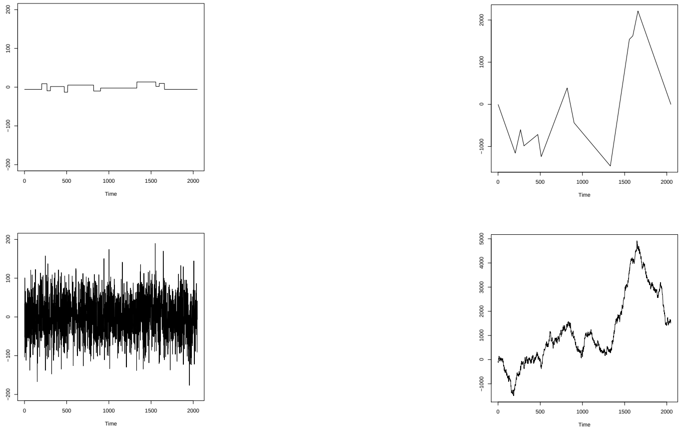
‘Adaptive’ bases in TSA – motivation no. 2

Some observations and modelling ideas:

- Many of the time series exhibit common trends, which change between positive and negative (and flat?).
- The similarities are sometimes less, sometimes more strong (probably the strongest during the recent financial crisis).
- Can we try modelling via e.g. linear trends, changing their slope (possibly in different ways for different series) at certain time points, plus some “idiosyncratic” noise?
- Modelling should be easier in the returns domain (i.e. after differencing), where we do not have the (near-)unit-root behaviour.
- Piecewise-linear trends in the log-price domain get transformed into piecewise-constant trends in the log-return domain (plus noise).
- Therefore we will be attempting to build a multivariate time series model with a piecewise-constant trend, change-points at possibly the same locations across the panel, plus its own noise for each time series component.

‘Cartoon’ illustration

Left column: log-returns (simulated noiseless trend, same trend + noise). Right: log-prices (cumsum of left column).



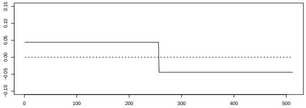
Further thoughts on modelling

Some further thoughts on the potential use of an oscillatory basis for the modelling of the (multivariate) returns.

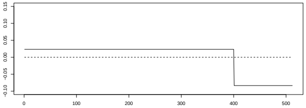
- The basis would need to be suitable for piecewise-constant functions, hence basis functions would perhaps need to be piecewise-constant themselves, with their supports (?) determined by the change-points (common to all time series).
- Corresponding basis coefficients would be dependent across the time series panel,
- but possibly independent across the basis (as in the classical Cramer-Fourier representation).
- Change-points unknown, hence basis functions would have to be recovered from the data.
- Given the above thoughts, let us focus attention on the Unbalanced Haar wavelets.

Unbalanced Haar wavelets

Balanced, scale 0:

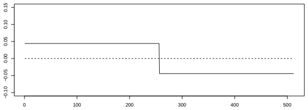


Unbalanced, scale 0:

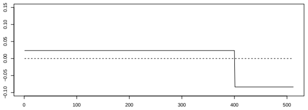


Unbalanced Haar wavelets

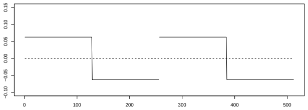
Balanced, scale 0:



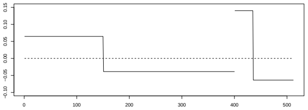
Unbalanced, scale 0:



Scale 1:



Scale 1:



Unbalanced Haar wavelets contd.

Example: Complete set of Unbalanced Haar vectors with $n = 6$.

$$\begin{pmatrix} 6^{-1/2} & 6^{-1/2} & 6^{-1/2} & 6^{-1/2} & 6^{-1/2} & 6^{-1/2} \\ \{5/6\}^{1/2} & -30^{-1/2} & -30^{-1/2} & -30^{-1/2} & -30^{-1/2} & -30^{-1/2} \\ 0 & \{3/10\}^{1/2} & \{3/10\}^{1/2} & -\{2/15\}^{1/2} & -\{2/15\}^{1/2} & -\{2/15\}^{1/2} \\ 0 & 2^{-1/2} & -2^{-1/2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 6^{-1/2} & 6^{-1/2} & -\{2/3\}^{1/2} \\ 0 & 0 & 0 & 2^{-1/2} & -2^{-1/2} & 0 \end{pmatrix}$$

Fact

A set of Unbalanced Haar vectors, constructed as in this example, is an orthonormal basis for $\mathbb{R}^n$.

The vectors are indexed by scale j (from -1 up) and location k (from 0 up).

The new model (univariate)

A stochastic process X_t, $t = 1, 2, \dots, T$ is called the Unbalanced Haar process if is has a representation

$$X_t = T^{1/2} \sum_{(j,k)\in\mathcal{I}} A_{j,k} \psi_t^{b_{j,k}} + \sigma_t \varepsilon_t, \quad t \in [1, T], \qquad (1)$$

where $\mathcal{I}$ is a set of indices, of finite dimensionality $|\mathcal{I}| = N + 1 < \infty$, such that $(-1, 0) \in \mathcal{I}$, and connected. The random variables $\{A_{j,k}\}_{(j,k)\in\mathcal{I}}$ are mutually independent, drawn from continuous distributions and satisfy $E(A_{j,k}) = 0$ and $E(A_{j,k}^2) < \infty$. The vector $\psi^{b_{j,k}}$ is a UH vector with a change-point at $b_{j,k}$. The constants σ_t are such that $0 < \underline{\sigma} < \sigma_t < \overline{\sigma} < \infty$, and $\{\varepsilon_t\}_t$ is a sequence of independent standard normal variables, also independent of $A_{j,k}$.

The new model (univariate) contd

Some points to note regarding the new model.

- The $T^{1/2} \sum_{(j,k)\in\mathcal{I}} A_{j,k} \psi_t^{b_{j,k}}$ part models the noiseless piecewise-constant trend; $\sigma_t \varepsilon_t$ adds the noise.
- Since each $\psi^{b_{j,k}}$ introduces one change-point, we need as many vectors $\psi^{b_{j,k}}$ as there are change-points. In this talk, we assume a finite number N of change-points, all fixed in 'rescaled time'.
- One immediate problem is that the representation is not unique: there are many ways in which we can represent a piecewise-constant function via UH wavelets.
- All representations are recoverable as long as we can identify the change-points (as all possible bases are induced by change-points).

'Canonical' basis via binary segmentation

Suppose f_t is a piecewise-constant function. One way of selecting an UH basis for such a signal is via Binary Segmentation (BS). Generic algorithm for BS:

1. Find $\bar{f}_i$, a step function with one change-point, minimising

$$\sum_{t=1}^{T} (f_t - \bar{f}_t)^2.$$

2. Denote the location of the change-point in $\bar{f}_t$ by b.
3. Perform similar fitting on $1, \ldots, b$ and $b+1, \ldots, T$.
4. Continue in the same manner until all change-points in f_t have been accounted for.

‘Canonical’ basis via binary segmentation contd

What does this have to do with UH wavelets?

Denote by $\bar{f}_t^{s,b,e}$ a step function (vector) starting at index s, with a change-point at b, ending at e. We have

$$b_0 := \arg\min_b \sum_{t=s}^{e} (f_t - \bar{f}_t^{s,b,e})^2 = \arg\max_b |\langle f, \psi^{s,b,e} \rangle|,$$

where $\psi^{s,b,e}$ is the UH vector starting at s, with jump at b, ending at e.

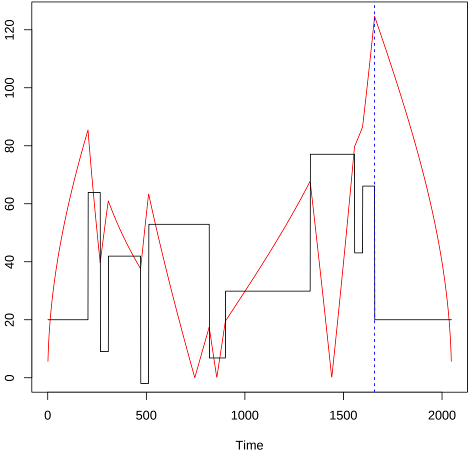
Thus, change-points in f_t are located by inspecting the maxima of $|\langle f, \psi^{s,b,e} \rangle|$ over b.

The thus chosen UH basis is now (a) unique and (b) interpretable in the sense that it orders the change-points hierarchically.

THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

BS UH basis – example of construction

First iteration of BS. Black: f_t. Red: $|\langle f, \psi^{1,t,T}\rangle|$. Blue: the first detected change-point b. Hence the first basis vector is $\psi^{1,b,T}$.



It is obviously possible to have other ‘unique’ and ‘interpretable’ UH bases. The key is to have some recipe for arranging the change-points hierarchically from the most to the least ‘important’, on the underlying noiseless function f_t, so that this ordering is then recoverable from its noisy version.

Once we have fixed the basis, we can e.g. define the Unbalanced Haar spectrum, analogous to the Cramer-Fourier case:

$$\alpha_{j,k} = E(A^2_{j,k})$$

Some further quantities of interest:

$$\begin{aligned} \mathrm{Var}(X_t) &= T \sum_{(j,k)\in\mathcal{I}} E(A^2_{j,k})(\psi_t^{b_{j,k}})^2 + \sigma_t^2 \\ \mathrm{Cov}(X_t, X_{t+\tau}) &= T \sum_{(j,k)\in\mathcal{I}} E(A^2_{j,k})\psi_t^{b_{j,k}}\psi_{t+\tau}^{b_{j,k}}, \quad \tau \neq 0. \end{aligned}$$

Conditional properties of UH processes

UH processes have mean zero and are covariance-nonstationary.

However, conditioning on the $A_{j,k}$'s, the process is also mean-nonstationary (exhibits jumps).

$$\begin{aligned} E(X_t | A_{j,k} = a_{j,k}) &= g_t \\ \mathrm{Var}(X_t | A_{j,k} = a_{j,k}) &= \sigma_t^2 \\ \mathrm{Cov}(X_t, X_{t+\tau} | A_{j,k} = a_{j,k}) &= 0 \quad \text{for} \quad \tau \neq 0, \end{aligned}$$

where

$$g_t = T^{1/2} \sum_{(j,k) \in \mathcal{I}} a_{j,k} \psi_t^{b_{j,k}}.$$

Recovering the basis

The change-point ordering, and hence the basis, is obviously unknown and needs to be recovered from the data.

Theorem (see paper for precise formulation). The probability of drawing values $a_{j,k}$ of $A_{j,k}$ for which the change-point number, their locations, as well as the canonical basis $\psi^{b_{j,k}}$ and the basis coefficients $a_{j,k}$ cannot be estimated consistently via Binary Segmentation as $T \to \infty$ when one conditions on these values, is zero.

In brief, this means that we can recover the basis from noisy data X_t. Having recovered the basis, we can then estimate the $a_{j,k}$'s.

This is one difference compared to traditional 'spectral' modelling in time series: here, we do not know the basis and need to estimate it.

Suppose now that we wish to forecast the conditional mean g_{T+h}, having observed $X_1, \ldots, X_T$, where h is "small". Our assumptions on change-points mean that $g_{T+h} = g_T$ if $h = o(T)$, which means the asset modelled is assumed to be trend-following in the short run.

We use

$$\hat{g}_T = T^{1/2} \sum_{(j,k)\in\hat{\mathcal{I}}} \hat{a}_{j,k} \psi_t^{\hat{b}_{j,k}}.$$

This gives a very interesting 'multiscale' view of forecasting financial returns.

For example, if one estimates all $a_{j,k}$'s as zero from scale $j_0 + 1$ onwards, the forecast is the sample mean of the data going back as far as the corresponding change-point in the most recent $\hat{\psi}^{b_{j_0,\cdot}}$.

This yields a natural and flexible stochastic framework for 'adaptive' forecasting of financial returns.

Forecasting contd

Brief discussion of forecasting results:

- We tried a number of assets in 4 asset classes: equity indices, equities, foreign exchange, commodities. We compared our forecast, for a range of *h*, against the 'benchmark' (non-adaptive moving average), both with optimised parameters. The error criterion was sign predictability.
- Overall, our adaptive forecast outperformed the benchmark: perofmance was better in commodities, and similar in other asset classes. Breaking up into different forecast horizons *h*, it was better in short- and medium-term, and not as good in long-term prediction.
- It was much better in periods of strong movements, when adaptivity to changing trends was particularly important.

Multivariate and high-dimensional modelling

Consider a panel of time series $X_t^{(i)}$. Our model is now

$$X_t^{(i)} = T^{1/2} \sum_{(j,k)\in\mathcal{I}} A_{j,k}^{(i)} \psi_t^{b_{j,k}} + \sigma_t^{(i)} \varepsilon_t^{(i)},$$

where $\mathrm{Cov}(A_{j,k}^{(i)}, A_{j,k}^{(l)}) = \alpha_{j,k}^{(i,l)}$, otherwise zero, and $\varepsilon_t^{(i)}$ is 'idiosyncratic' to $X_t^{(i)}$, for $i, l = 1, \ldots, p$.

'Factor modelling' interpretation: the jumps in $\psi_t^{b_{j,k}}$ can be interpreted as 'factors' affecting the whole system. Each time series react with a possibly different 'loading' $A_{j,k}^{(i)}$.

The dimensionality of $A_{j,k}^{(i)}$ itself (for each j, k) can be reduced, e.g. via factor modelling.

For $i \neq l$, we have $\mathrm{Cov}(X_t^{(i)}, X_s^{(l)}) = T \sum_{j,k} \alpha_{j,k}^{(i,l)} \psi_t^{b_{j,k}} \psi_s^{b_{j,k}}$.

Final discussion

Some final remarks:

- In estimating the cross-covariance matrix

$$\text{Cov}(X_t^{(i)}, X_t^{(l)}) = T \sum_{j,k} \alpha_{j,k}^{(i,l)} (\psi_t^{b_{j,k}})^2 + \sigma_t^2 I\{i = l\},$$

  sparsity or structure can be imposed on the estimated $\alpha_{j,k}^{(i,l)}$ so that the estimated cross-covariance is stably invertible.
- Change-points in this problem can also be estimated in other ways, not necessarily via Binary Segmentation.
- Results (e.g. to do with forecasting) can be averaged over a variety of possible bases.
- Model also useful in simulating multivariate returns.

Reference

A. L. Schroeder, P. Fryzlewicz (2013) Adaptive trend estimation in financial time series via multiscale change-point-induced basis recovery, *Statistics and Its Interface*, **6**, 449-461.

Available from

http://stats.lse.ac.uk/fryzlewicz/