Rebecca Killick
r.killick@lancs.ac.uk

Joint work with
Idris Eckley (Lancaster), Marina Knight (York) & Guy Nason (Bristol)

June 30, 2014
What do I mean by Nonstationary Time Series?

- I mean *NOT* second-order stationary.
- So, unconditional variance *changes* with time.
- Autocovariance, spectrum, etc. *change* with time.
- Typically assume $\mathbb{E}X_t = 0$ (assume mean removed).
What do I mean by Nonstationary Time Series?

- I mean *NOT* second-order stationary.
- So, unconditional variance *changes* with time.
- Autocovariance, spectrum, etc. *change* with time.
- Typically assume $\mathbb{E}X_t = 0$ (assume mean removed).

More interested in Variability and CI
Structure of Presentation

- Motivation
- Nonstationary forecasting
- The local partial autocorrelation function
- Forecasting using the lpacf
Motivation
Motivation - ABML

- ABML consists of gross value added amounts
- Component in the estimate of GDP
- 223 observations from Q1 1955 to Q3 2010
- We use second differences to remove trend
- Tests of stationarity reject H0.
Motivation - ABML

- ONS currently use ARIMA models to forecast this data
- What is the danger in doing this?

- Red - Full series forecast
- Blue - Last 30 obs forecast

- Full fits ARMA(1,1) non-zero mean
- Last 30 obs fits AR(2)
ONS currently use ARIMA models to forecast this data

What is the danger in doing this?

- Red - Full series forecast
- Blue - Last 30 obs forecast
- Full fits ARMA(1,1) non-zero mean
- Last 30 obs fits AR(2)
- Overconfident in forecast?
Suppose have data $x_1, \ldots, x_T$ from *stationary* series.

Want to make forecast $\hat{x}_T(h)$ made at time $T$ for horizon $h$.

Want forecast using linear combination of past:

$$\hat{x}_T(1) = \sum_{i=0}^{T-1} \varphi_i x_{T-i}$$

$$= \varphi_0 x_T + \varphi_1 x_{T-1} + \varphi_2 x_{T-2} + \cdots$$

Note: $\varphi$ sequence DOES NOT depend on time (stationary $x_t$)

Theory can tell us optimal least-squares forecast (Box-Jenkins).
Extension to Nonstationary Time Series
Modelling nonstationary time series

Modelling in the face of non-stationarity is no easy task!

- Various approaches have been explored, built on models that fit particular types of non-stationarity:
  - assume piecewise stationarity;
  - use parametric models with time-changing coefficients, e.g. Time Varying AR (tvAR).

- Processes with a slowly time-varying second order structure are known as locally stationary (LS).
  - Advanced LS models (ARCH) (Dahlhaus and Subba Rao, 2006).
  - Locally stationary fourier processes (Dahlhaus, 1997).
  - Locally stationary wavelet processes (Nason et al., 2000).
Locally stationary models

If you take a small enough region, it will appear stationary as the structure varies slowly over time.

Application areas include;

- medicine, finance
- environmental processes, e.g. wind speeds.
Locally stationary wavelet (LSW) processes

LSW processes (Nason et al., 2000):

\[
X_{t,T} = \sum_{j=-J(T)}^{J(T)} \sum_{k \in \mathbb{Z}} w_{j,k;T} \psi_{j,k}(t) \xi_{j,k}, \quad t = 1, \ldots, T.
\]

- \(\{\psi_{j,k}\}\) is a collection of discrete non-decimated wavelets.
- \(\{\xi_{j,k}\}_{j,k}\) is a sequence of zero-mean, orthonormal random variables.
- Smoothness of wavelet amplitudes \(w_{j,k;T}\) as a function of \(k\) controls the degree of non-stationarity.
- LSW processes encapsulate other models and represent processes whose variance and autocorrelation function vary over time.
- This leads to a localised measure of autocovariance \(c(t, \tau)\).
Forecasting in LSW framework

Given observations $x_0, \ldots, x_{t-1}$, we:

- Predict $\hat{x}_t = \sum_{s=t-p}^{t-1} b_{t-1-i} X_s$, where
  
  $p$ is the number of latest observations used for prediction and
  
  $b$ is the solution to localized Yule-Walker equations.

$$
\begin{bmatrix}
  c(t, 1) \\
  c(t, 2) \\
  \vdots \\
  c(t, p)
\end{bmatrix}
= 
\begin{bmatrix}
  c(t - 1, 0) & c(t - 2, -1) & \cdots \\
  c(t - 1, 1) & c(t - 2, 0) & \cdots \\
  \vdots & \vdots & \ddots \\
  c(t - 1, p) & c(t - 2, p - 1) & \cdots
\end{bmatrix}
\begin{bmatrix}
  b_{t-1} \\
  b_{t-2} \\
  \vdots \\
  b_{t-1-p}
\end{bmatrix}
$$
Given observations $x_0, \ldots, x_{t-1}$, we:

- Predict $\hat{x}_t = \sum_{s=t-p}^{t-1} b_{t-1-i, T} X_s$, where

- $p$ is the number of latest observations used for prediction and
- $b$ is the solution to localized Yule-Walker equations.

\[
\begin{bmatrix}
  c(t, 1) \\
  c(t, 2) \\
  \vdots \\
  c(t, p)
\end{bmatrix}
= 
\begin{bmatrix}
  c(t - 1, 0) & c(t - 2, -1) & \cdots \\
  c(t - 1, 1) & c(t - 2, 0) & \cdots \\
  \vdots & \vdots & \ddots \\
  c(t - 1, p) & c(t - 2, p - 1) & \cdots
\end{bmatrix}
\begin{bmatrix}
  b_{t-1} \\
  b_{t-2} \\
  \vdots \\
  b_{t-1-p}
\end{bmatrix}
\]

These require knowledge of the covariance structure at time $t$ which is the point we are trying to predict.

- The covariance at time $t$ is similar to that at time $t - 1$.
- We extrapolate at the smoothing step.
Fryzlewicz et al. (2003) propose a method for LSW forecasting:
- smooth and extrapolate the covariances directly by kernel smoothing;
- choose $p$ (and bandwidth) by in sample optimization.

For a practitioner this leaves many questions.
- How much data do I train on?
- Do I update $p$ and bandwidth simultaneously or another method?
- When updating, how many alternative options do I consider?
- What kernel smoother should I use?

Ultimately this $p$ is hard to choose.
Our approach

Our work:

- introduces \textit{localised partial} acf
- shows that local partial ACF is an interesting tool in its own right
- gives encouraging forecasting results

Thus we,

- mirror the stationary process and
- produce a data driven approach for practitioners.
We define the lpacf at time $t$ and lag $\tau$ as:

$$q(t, \tau) = \text{Corr}(X_t, X_{t-\tau} | \{X_{t-1} \ldots, X_{t-\tau+1}\})$$

$$q_T(t, \tau) = \begin{cases} 
\frac{c(t-1,1)}{\sqrt{c(t,0)c(t-1,0)}}, & \text{for } \tau = 1 \\
\sqrt{\frac{\text{MSPE}(\hat{X}_t, X_t | X_{t-1} \ldots, X_{t-\tau+1})}{\text{MSPE}(\hat{X}_{t-\tau}, X_{t-\tau} | X_{t-1} \ldots, X_{t-\tau+1})}}, & \text{for } \tau \geq 2.
\end{cases}$$

For stationary models the square root equals 1 and the $\varphi_{t,\tau,\tau}$ is the usual estimate of the pacf.
ABML lpacf

LPACF of ABML second differences

Year


Rebecca Killick (Lancaster University) Forecasting locally stationary time series June 30, 2014 16 / 20
Forecasting Simulations
Comparisons with ARMA

Range of models considered:
- TVAR(1)
- TVAR(2)
- TVAR(12)
- TVMA(1)
- TVMA(2)
- Uniformly modulated white noise
- LSW process

Summary

lpacf greatly improves on ARMA in 2/3 of cases
comparable results (ratio $\pm 0.05$) in 1/3.
AMBML - Ipacf forecasting

RMSE: Ipacf=3430m, B-J=4290m, Fryzlewicz=8830m, 11/15, 8/15, 6/15

Forecast and Prediction Intervals
Motivated why forecasting nonstationary time series is important.

Proposed a new measure – the local partial autocorrelation function – and associated theoretical justification.

Used the lpacf to choose $p$ for the localised Yule-Walker equations.

Showed increased forecasting performance when using the lpacf.

We have used the lpacf as a tool for forecasting but it can be used in a variety of settings.
Statistical inference for time-varying ARCH processes.  

R. Dahlhaus.  
Fitting Time Series Models to Nonstationary Processes.  

Wavelet Processes and Adaptive Estimation of the Evolutionary Wavelet Spectrum.  

Forecasting non-stationary time series by wavelet process modelling.  