Inaugurale lezing
Belgische Francqui-leerstoel 2019-2020
Universiteit Antwerpen

Siem Jan Koopman

Vrije Universiteit Amsterdam, Tinbergen Institute

Monday 30 November 2020

http://sjkoopman.net

The Econometrics of Time-Varying Parameters

- Presentation is NOT a comprehensive review of literature !!
- It is just my limited view of the subject ...
- We start with models in the class of dynamic *parameter-driven models*
- Later we move on to models in the class of dynamic *observation-driven models*
- Along the way we discuss different approaches to estimation
- Most focus is on univariate models, despite the fact that the multivariate models are the way forward

- Much work to do . . .
- on this dark, late afternoon of the last November day in 2020.
- So much to discuss, there is the danger that it will all end in a disaster at the end of the presentation
- But in the first three December Tuesdays, starting tomorrow,
- we have our lectures in the afternoon with much more time for the detail !
- You are all invited and very welcome.
- In any case,

Let's start !

We start with the mean

We denote the observed values of a time series variable by

$$y_1, \quad y_2, \quad \ldots \quad , \quad y_{n-1}, \quad y_n,$$

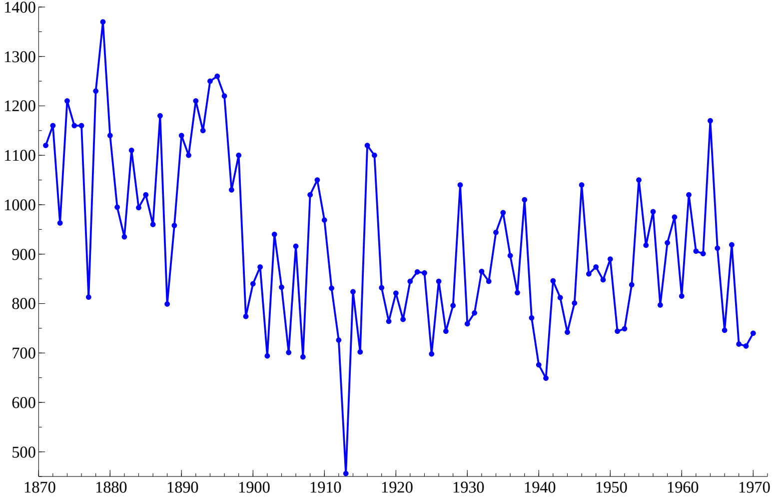
where index n indicates the number of observations (or length of the time series).

The statistical model for y_t with a fixed mean μ is given by

$$y_t = \mu + \varepsilon_t, \qquad \varepsilon_t \sim p(\varepsilon_t), \qquad t = 1, \ldots, n.$$

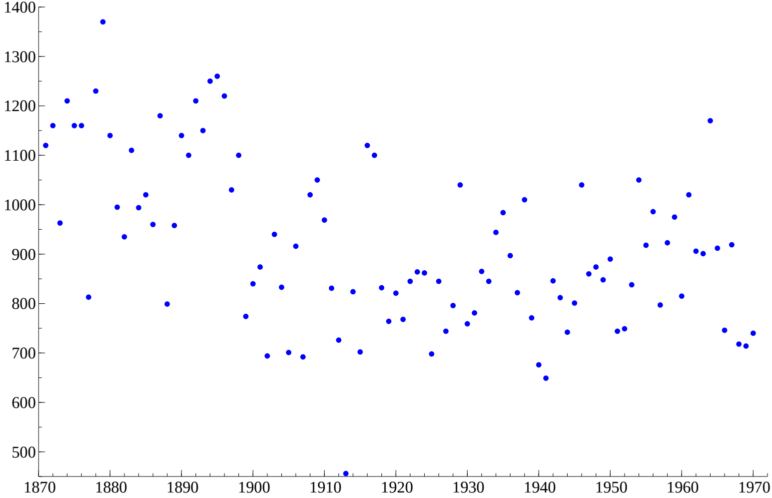
where μ is the unknown constant and ε_t is the remainder that is assumed to be a stochastic variable with a probability density function $p(\cdot)$. We further assume that $\mathrm{E}(\varepsilon_t) = 0$.

Nile Data

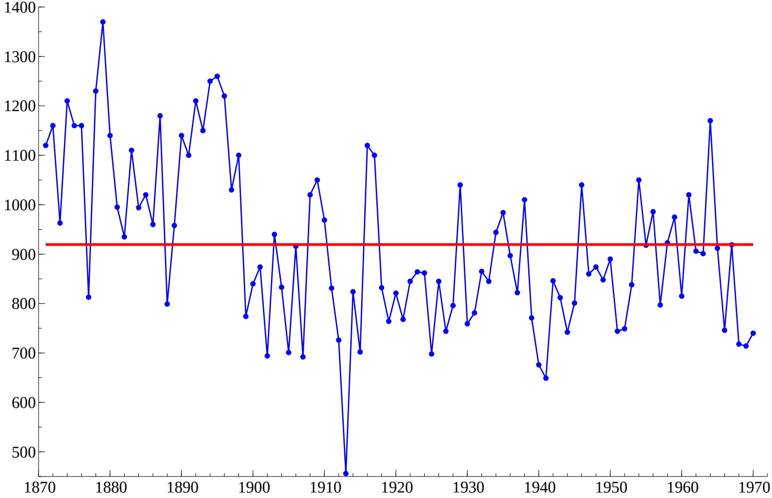


1400
1300
1200
1100
1000
900
800
700
600
500
1870
1880
1890
1900
1910
1920
1930
1940
1950
1960
1970

Nile Data: what is signal, what is noise ?



Constant: $y_t = \mu + \varepsilon_t$



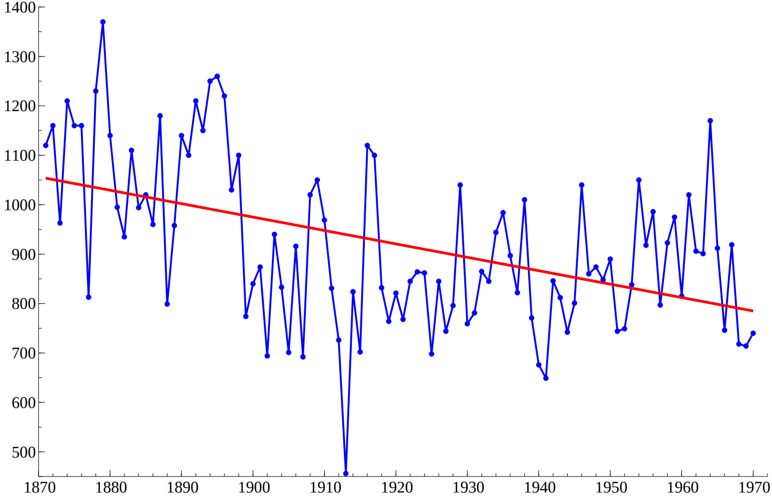
Is the constant mean a good signal ?

We can extend the model with time trend functions:

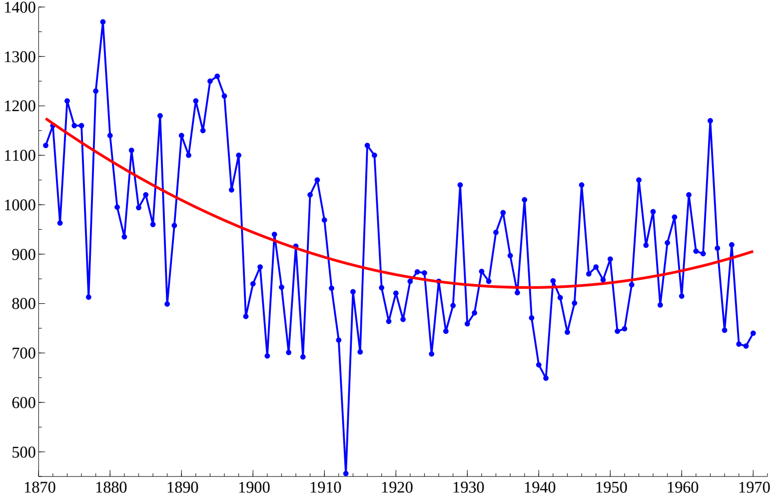
$$y_t = \mu + \beta\, t + \delta\; t^2 + \gamma\, t^3 + \;\dots\; + \varepsilon_t, \qquad t = 1, \dots, n.$$

This is a time-varying signal : $\mu + \beta\, t + \gamma\, t^2 + \;\dots$

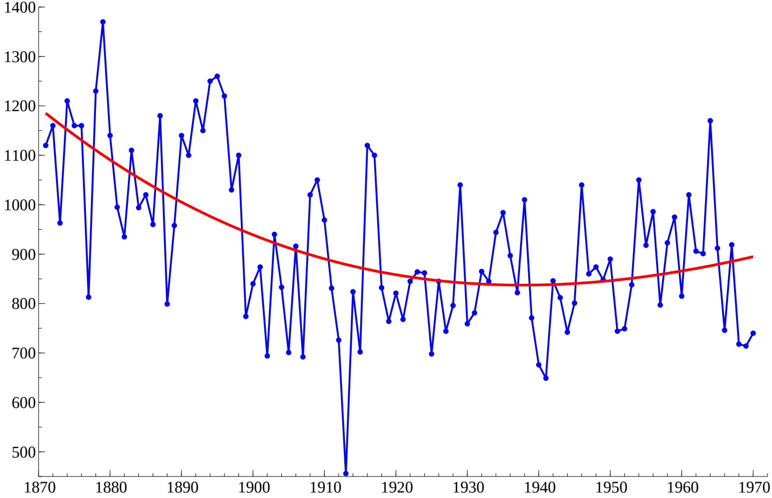
Trend t: $y_t = \mu + \beta t + \epsilon_t$



Trend t^2: $y_t = \mu + \beta t + \delta t^2 + \epsilon_t$



Trend t^3: $y_t = \mu + \beta t + \delta t^2 + \gamma t^3 + \epsilon_t$



Constant vs Time-Varying Mean Model

- Constant mean:
  - fixed level μ :

$$y_t = \mu + \varepsilon_t, \qquad \varepsilon_t \sim \text{NID}(0, \sigma_\varepsilon^2)$$

- Time Varying mean: replace μ by μ_t with
  - deterministic function of time:

$$\mu_t = a + b\,t + c\,t^2 + \dots$$

  - stochastic function of time, for example:

$$\mu_t = \mu_{t-1} + \eta_t, \qquad \eta_t \sim \text{NID}(0, \sigma_\eta^2)$$

Local Level Model

The local level model is given by

$$\begin{aligned} y_t &= \mu_t + \varepsilon_t, \qquad & \varepsilon_t &\sim \text{NID}(0, \sigma_\varepsilon^2) \\ \mu_{t+1} &= \mu_t + \eta_t, \qquad & \eta_t &\sim \text{NID}(0, \sigma_\eta^2) \end{aligned}$$

- Time-varying level is modelled as a random walk process;
- The disturbances ε_t, η_s are independent for all s, t;
- The model is incomplete without initial specification for μ_1;
- The process μ_t is nonstationary and y_t is nonstationary.

The local level model or random walk plus noise model :

$$\begin{aligned} y_t &= \mu_t + \varepsilon_t, \qquad \varepsilon_t \sim \mathrm{NID}(0, \sigma_\varepsilon^2) \\ \mu_{t+1} &= \mu_t + \eta_t, \qquad \eta_t \sim \mathrm{NID}(0, \sigma_\eta^2) \end{aligned}$$

- The level μ_t and irregular ε_t are both unobserved;
- Parameters σ_ε^2 and σ_η^2 are unknown;
- Define q as the *signal-to-noise ratio* : $q = \sigma_\eta^2 \,/\, \sigma_\varepsilon^2$.

The local level model or random walk plus noise model :

$$\begin{aligned} y_t &= \mu_t + \varepsilon_t, \qquad \varepsilon_t \sim \text{NID}(0, \sigma_\varepsilon^2) \\ \mu_{t+1} &= \mu_t + \eta_t, \qquad \eta_t \sim \text{NID}(0, \sigma_\eta^2) \end{aligned}$$

- Trivial special cases:
  - $\sigma_\eta^2 = 0 \implies y_t \sim \text{NID}(\mu_1, \sigma_\varepsilon^2)$ (IID, global level);
  - $\sigma_\varepsilon^2 = 0 \implies y_{t+1} = y_t + \eta_t$ (random walk);
- Local Level model is basic illustration of state space model.

Kalman filter for Local Level Model

Estimation of Time-Varying Level μ_t is via Kalman filter.
We have Local Level model given by

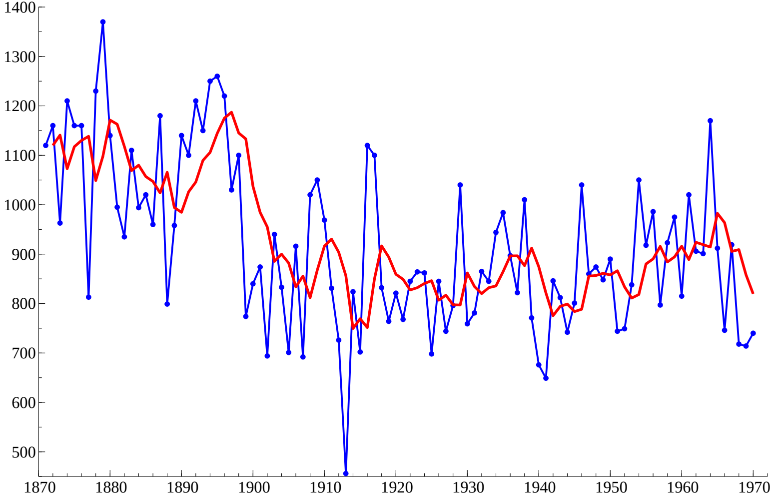
$$y_t = \mu_t + \varepsilon_t, \qquad \mu_{t+1} = \mu_t + \eta_t.$$

We denote $a_t = \mathrm{E}(\mu_t | Y_{t-1})$ and $p_t = \mathrm{Var}(\mu_t | Y_{t-1})$.
The Kalman filter is a recursion from $t = 1$ to $t = n$:

$$\begin{aligned} v_t &= y_t - a_t, & k_t &= p_t \,/\, f_t, \\ a_{t+1} &= a_t + k_t v_t, & p_{t+1} &= k_t\, \sigma_\varepsilon^2 + \sigma_\eta^2. \end{aligned}$$

where v_t is the prediction error with its variance $f_t = p_t + \sigma_\varepsilon^2$.
Kalman filter only operational with initial values for a_1 and p_1.

Signal Extraction: predicted estimates



Time-Varying Regression Model

- Standard time series Regression Model:

$$y_t = X_t\beta + \varepsilon_t, \qquad \varepsilon_t \sim \text{NID}(0, \sigma_\varepsilon^2)$$

  where X_t is a row vector of explanatory variables and β is a row-vector of unknown regression coefficients.

- Time-Varying Regression Model is obtained by replacing β by β_t which is modelled as a "stochastic function of time", for example, the vector random walk process

$$\beta_{t+1} = \beta_t + \eta_t, \qquad \eta_t \sim \text{NID}(0, \Sigma_\eta),$$

  where η_t is a vector of disturbances which have mean zero, the variance matrix of η_t is Σ_η.

The Time-Varying Regression Model is given by

$$y_t = X_t \beta_t + \varepsilon_t, \qquad \varepsilon_t \sim \mathrm{NID}(0, \sigma_\varepsilon^2)$$

with time-Varying coefficient vector

$$\beta_{t+1} = \beta_t + \eta_t, \qquad \eta_t \sim \mathrm{NID}(0, \Sigma_\eta),$$

for $t = 1, \ldots, n$.

The initial coefficient vector β_1 needs to be provided (this is known as the initialization problem).

We are effectively looking at the

Linear Gaussian State Space Model.

State Space Model

Linear Gaussian State Space Model is defined in three parts:

→ Observation equation:

$$y_t = Z_t \alpha_t + \varepsilon_t, \qquad \varepsilon_t \sim \text{NID}(0, H_t),$$

→ State transition equation:

$$\alpha_{t+1} = T_t \alpha_t + R_t \zeta_t, \qquad \zeta_t \sim \text{NID}(0, Q_t),$$

→ Initial condition: $\alpha_1 \sim \mathcal{N}(a_1, P_1)$.

- ε_s and ζ_t independent for all t, s, and independent from α_1;
- observation variable y_t (scalar or vector);
- state vector α_t is unobserved;
- system matrices T_t, Z_t, R_t, Q_t, H_t are fixed at time t.
- system matrices determine (dynamic) structure of model.
- initial conditions : specify a_1 and P_1.

- The important key variables in a state space model are the observation variable y_t and the state vector α_t;
- State vector α_t contains all the dynamic features in the model;
- State vector α_t can also represent other effects in the model (fixed and regression effects);
- State space model is a convenient representation or formulation for almost all linear Gaussian time series models;
- It is mostly used for the purpose of using the Kalman filter and its related algorithms;

Given the state space model $\alpha_{t+1} = T_t\alpha_t + R_t\zeta_t, \quad y_t = Z_t\alpha_t + \varepsilon_t$ with known system matrices :

Define

$$a_{t+1} = \mathbb{E}(\alpha_{t+1}|y_1, \ldots, y_t), \qquad P_{t+1} = \mathbb{V}\text{ar}(\alpha_{t+1}|y_1, \ldots, y_t).$$

The next-period state vector α_{t+1} can be "estimated" using past observations through the *Kalman filter*:

$$\begin{aligned} v_t &= y_t - Z_t a_t, \\ F_t &= Z_t P_t Z_t' + H_t, \\ K_t &= T_t P_t Z_t' F_t^{-1}, \\ a_{t+1} &= T_t a_t + K_t v_t, \\ P_{t+1} &= T_t P_t T_t' + R_t Q_t R_t' - K_t F_t K_t', \end{aligned}$$

for $t = 1, \ldots, n$ and starting with given values for a_1 and P_1.

Unobserved Components

The unobserved components time series model provides a model-based approach to the decomposition of a time series into, for example, trend and seasonal:

$$y_t = \mu_t + \gamma_t + \varepsilon_t.$$

where μ_t is trend, γ_t is seasonal and ε_t is remainder.
The seasonal component γ_t is subject to number of 'seasons' in the data:

- $s = 12$ for monthly data,
- $s = 4$ for quarterly data,
- $s = 7$ for daily data when modelling a weekly pattern.

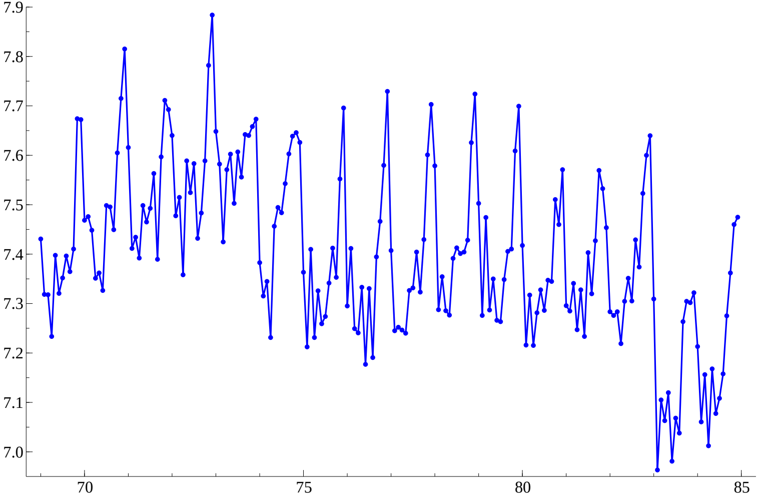
Unobserved Components

The unobserved components time series model (UCTSM) is

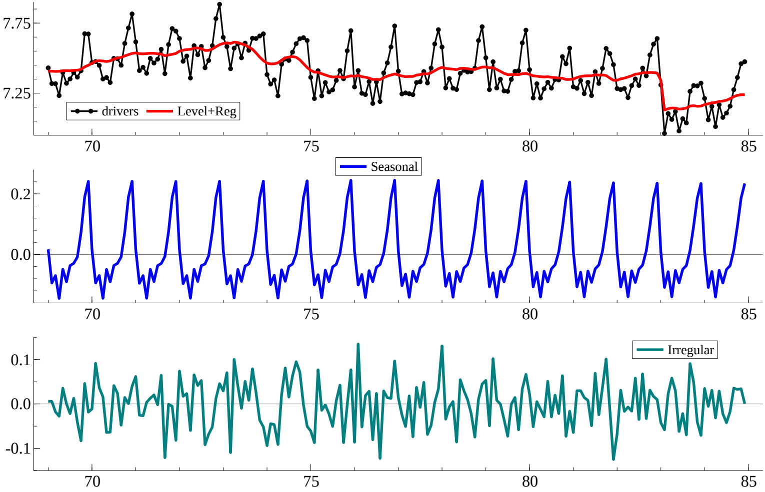
$$y_t = \mu_t + \gamma_t + \varepsilon_t.$$

- Different specifications for the trend and the seasonal components can be considered.
- Other components of interest, like cycles, explanatory variables, interventions effects, outliers, are easily added.
- For UCTSM, we model non-stationarity directly.
- For UCTSM, components have an explicit interpretation: the model is not just a forecasting device.
- Estimation is done by maximum likelihood, Kalman filter provides the likelihood function, see Harvey (1989)

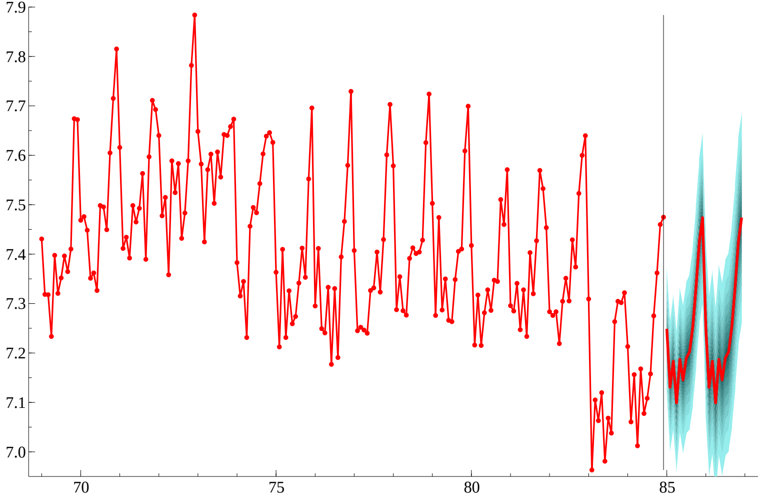
Seatbelt Law



Seatbelt Law: decomposition



Seatbelt Law: forecasting



Other Time Series Models

There are many !!

- Autoregressive models
- Autoregressive moving average (ARMA) models
- ... *unobserved components time series models* ...
- Long memory models, fractional integration (ARFIMA) models
- Dynamic regression models, error correction models
- Vector autoregressive models, cointegration, vector error correction models
- ... *state space models* ...
- Regime-switching, Markov-switching, treshold autoregression, smooth transitions models
- Generalized autoregressive conditional heteroskedasticity (GARCH) models
- Autoregressive conditional duration models and related models
- ... *stochastic volatility models* ...

But we are discussing

Time-Varying Parameters

- Parameter-driven models
  - Local level model
  - Unobserved components time series models
  - General state space models
  - Stochastic volatility models
- Observation-driven models
  - Single-source of error models
  - Count time series models, ARMA-type
  - Generalized autoregressive conditional heteroskedasticity (GARCH) models
  - Generalized autoregressive score (GAS) models
  - Score-driven models

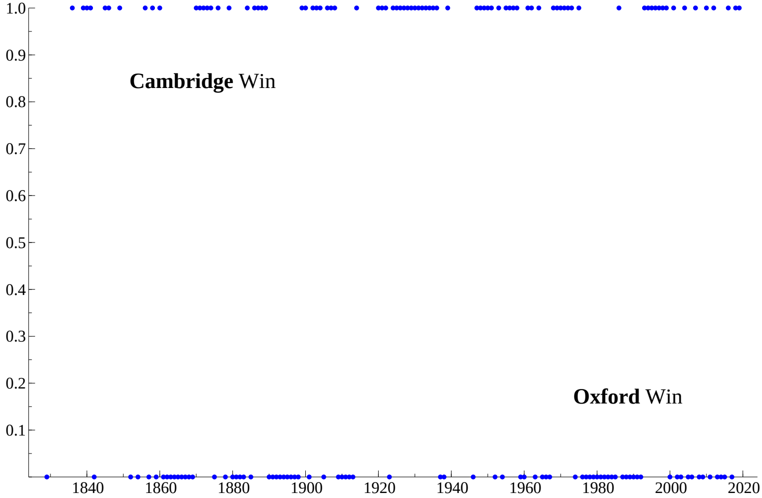
Motivation Score-driven

Why Observation-driven or Score-driven models ?

- In a linear Gaussian world, classical and Bayesian inference procedures are well in place;
- There are many important challenging cases where time series clearly posses nonlinear and non-Gaussian features . . .
- State space analysis is numerically involved in such cases . . .
- Key example is Stochastic Volatility (SV) model versus GARCH model
- Is there an analogue to general state space ?
- For this purpose we explore score-driven models.

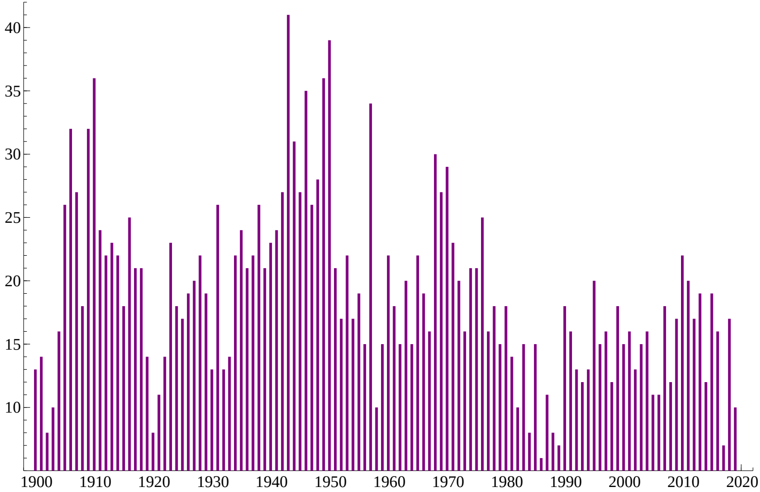
Outcomes of Boat Race: Cambridge – Oxford

Wins: 84 Cambridge vs 80 Oxford, next race is Sunday 29th March 2020, 4:44 PM (UK time), what do you predict ?

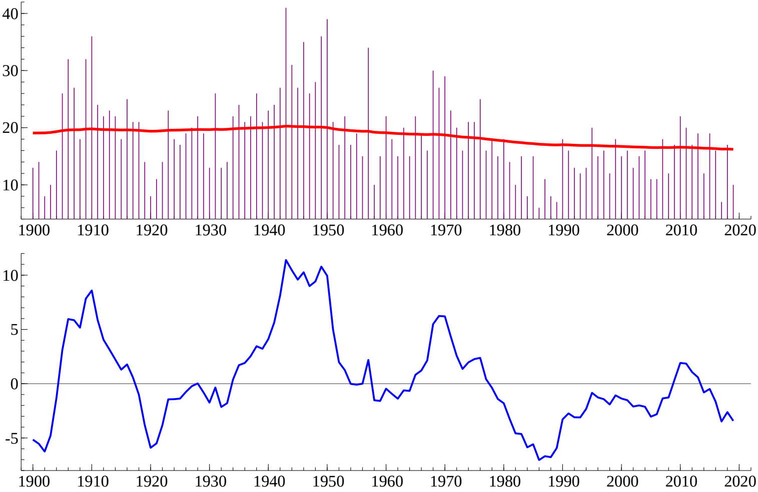


Number of Earthquakes in a year, since 1900

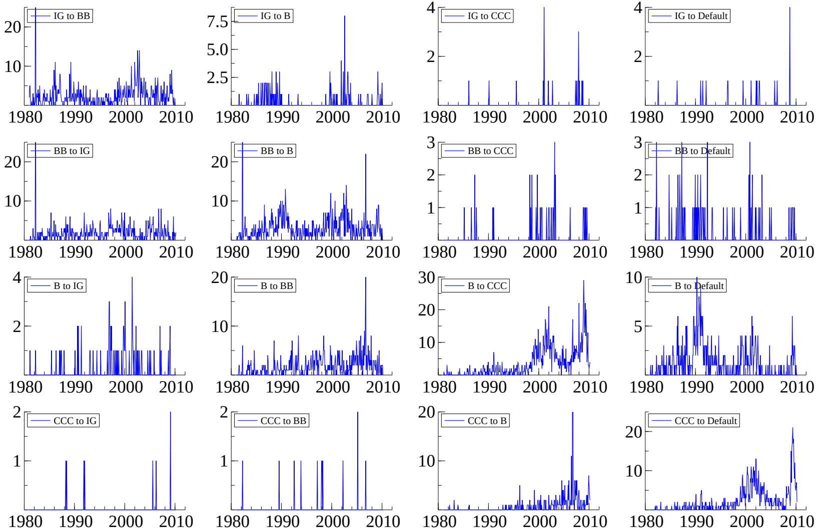
https://en.wikipedia.org/wiki/Lists_of_21st-century_earthquakes



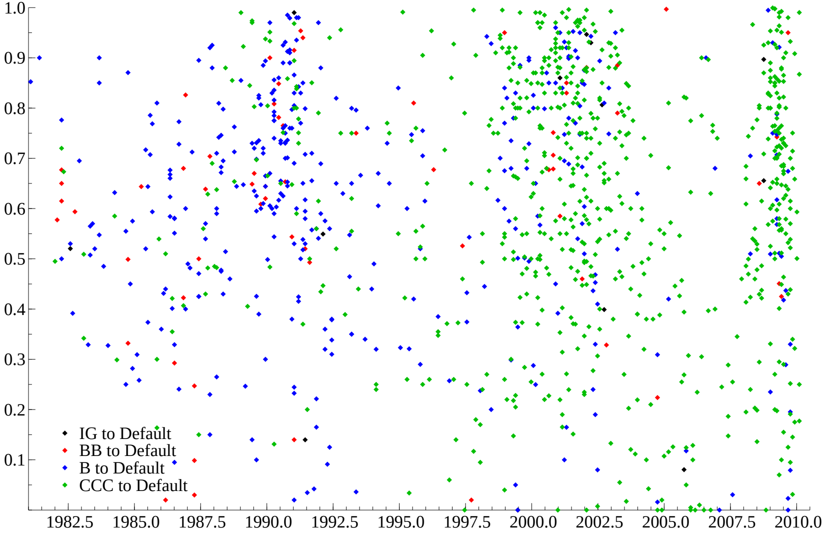
Decomposition of Dynamics in Earthquakes



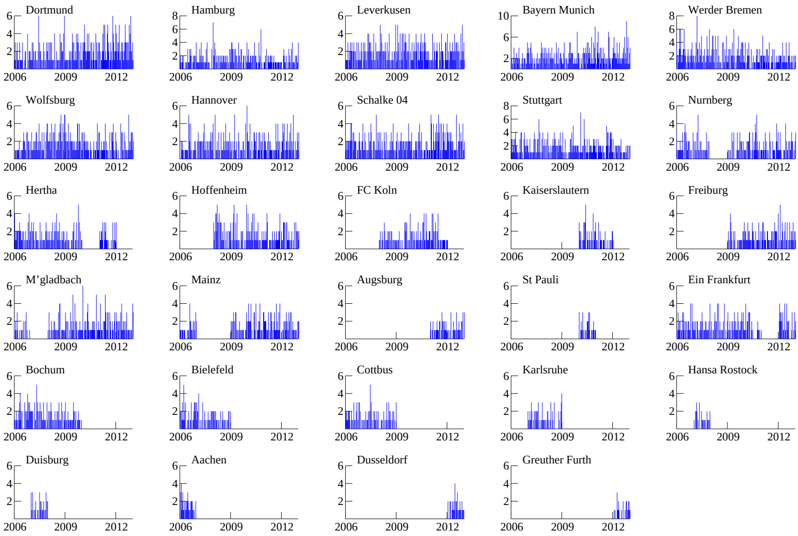
Credit Rating Transitions for US companies



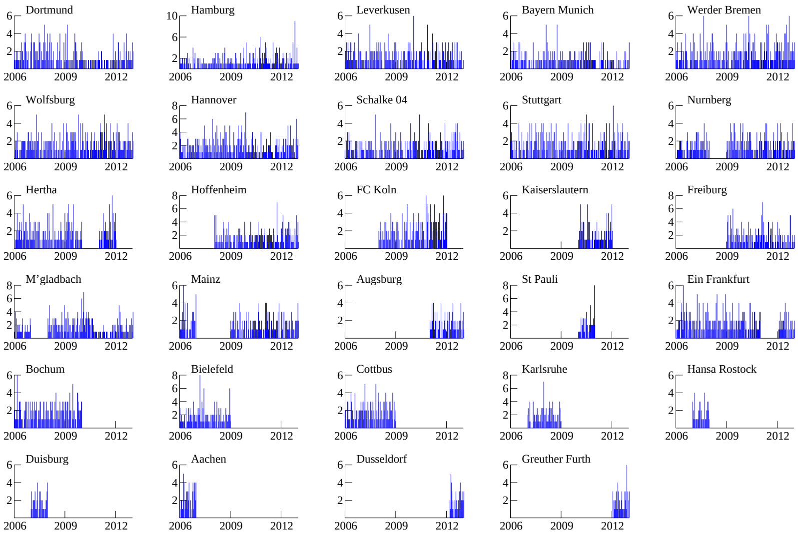
Loss Given Defaults for US companies



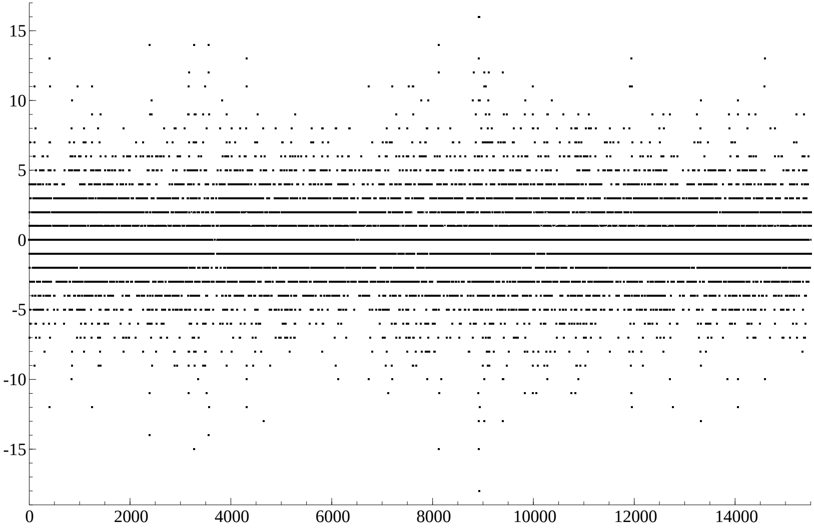
Goals scored by football teams in Bundesliga



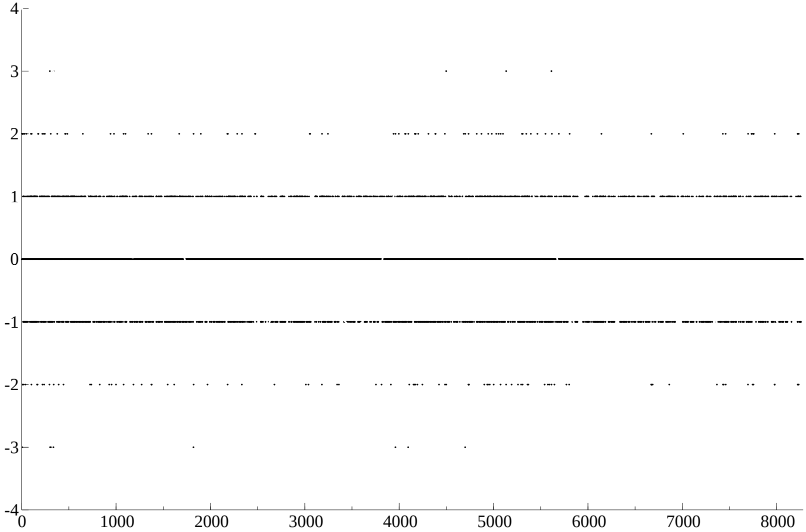
Goals conceded by football teams in Bundesliga



A Day of tick-by-tick returns for IBM



A Day of tick-by-tick returns for Verizon



A Restart of Time-Varying Parameters

We can represent a statistical model with (partially) time-varying parameters as

$$y_t \sim p(y_t | f_t; \theta),$$

with time-varying parameter

$$f_{t+1} = \omega + \beta f_t + \alpha \times \text{" some innovation "},$$

it is the "some innovation" that determines whether we are having a parameter-driven model or an observation-driven model:

- Parameter-driven : "some innovation" is random noise, that is $\eta_t \sim \text{NID}(0, \sigma_\eta^2)$, Kalman filter provides the "estimate"
- Observation-driven : "some innovation" is function of data $y_t, y_{t-1}, y_{t-2}, \ldots$

Given the p.e.d., contribution to loglikelihood $\ell = \log p(y|f;\theta)$ at time t is

$$\ell_t = \log p(y_t|y_1, \ldots, y_{t-1}, f_1, \ldots, f_t; \theta).$$

In an observation-driven model, $f_1, \ldots, f_t$ are functions of past data.

In many cases, ℓ_t reduces to

$$\ell_t = \log p(y_t|f_t;\theta).$$

The parameter value for next period is then updated by

$$f_{t+1} = \omega + \beta f_t + \alpha \times \text{"some function of } y_t \text{ and possibly past data"},$$

Key proposal : set this innovation equal to **score of** ℓ_t with respect to f_t.

We label this approach as the

Generalized Autoregressive Score model

or the GAS model, see Creal, Koopman and Lucas (2013).

Generalized Autoregressive Score model

For the observation equation,

$$y_t \sim p(y_t | Y_{t-1}, f_t; \theta), \qquad Y_t = \{y_1, \ldots, y_t\},$$

we propose the score updating scheme for f_t based on

$$f_{t+1} = \omega + \beta f_t + \alpha s_t,$$

where the innovation or driving mechanism s_t is given by

$$s_t = S_t \cdot \nabla_t$$

where

$$\begin{aligned} \nabla_t &= \frac{\partial \log p(y_t | Y_{t-1}, f_t; \theta)}{\partial f_t}, \\ S_t &= \mathcal{I}_{t-1}^{-1} = -\,\mathbb{E}_{t-1}\left[\frac{\partial^2 \log p(y_t | Y_{t-1}, f_t; \theta)}{\partial f_t \partial f_t'}\right]^{-1}. \end{aligned}$$

Illustration 1 : time-varying mean

Consider time-varying mean model $y_t \sim \text{NID}(f_t, \sigma^2)$ or

$$y_t = f_t + \varepsilon_t, \qquad \varepsilon_t \sim \text{NID}(0, \sigma^2),$$

for $t = 1, \ldots, n$. For the score-driven solution, we have

$$f_{t+1} = \omega + \beta f_t + \alpha s_t, \qquad s_t = S_t \cdot \nabla_t,$$

where

$$\nabla_t = \frac{\partial \ell_t}{\partial f_t}, \qquad S_t = -\,\mathbb{E}_{t-1}\left[\frac{\partial^2 \ell_t}{\partial f_t \partial f_t}\right]^{-1}$$

with

$$\ell_t = -\frac{1}{2}\log 2\pi - \frac{1}{2}\log \sigma^2 - \frac{1}{2\sigma^2}(y_t - f_t)^2.$$

We obtain

$$\nabla_t = \frac{1}{\sigma^2}(y_t - f_t), \qquad S_t = \sigma^2,$$

and we have simply, $s_t = y_t - f_t$ as the prediction error.

Illustration 1 : time-varying mean is ARMA model

For the time-varying mean model

$$y_t = f_t + \varepsilon_t, \qquad \varepsilon_t \sim \mathrm{NID}(0, \sigma^2),$$

the score updating is

$$f_{t+1} = \omega + \beta f_t + \alpha(y_t - f_t), \quad \text{since} \quad s_t = y_t - f_t.$$

We can replace f_t by $y_t - \varepsilon_t$, then the score updating becomes

$$y_{t+1} = \omega + \beta y_t + \varepsilon_{t+1} + (\alpha - \beta)\varepsilon_t,$$

and hence score updating implies the ARMA(1,1) model for y_t

$$y_t = \omega + \phi y_{t-1} + \varepsilon_t + \theta\varepsilon_{t-1},$$

where $\phi \equiv \beta$ and $\theta = \alpha - \beta$.

If $\alpha = \beta$, we obtain the AR(1) model.

Illustration 2: time-varying variance

For the time-varying variance model $y_t \sim \text{NID}(\mu, f_t)$ or

$$y_t = \mu + \varepsilon_t, \qquad \varepsilon_t \sim \text{NID}(0, f_t),$$

with $t = 1, \ldots, n$, we set $\mu = 0$. The predictive logdensity is

$$\begin{aligned}
\ell_t &= -\frac{1}{2}\log 2\pi - \frac{1}{2}\log f_t - \frac{y_t^2}{2f_t}, \\
\nabla_t &= \frac{1}{2f_t^2}y_t^2 - \frac{1}{2f_t} \;=\; \frac{1}{2f_t^2}(y_t^2 - f_t), \\
\mathbb{E}_{t-1}(\nabla_t) &= 0, \qquad \mathbb{E}_{t-1}\left(-\frac{\partial \nabla_t}{\partial f_t}\right) = \mathcal{I}_{t-1} = \frac{1}{2f_t^2}, \\
S_t \;=\; \mathcal{I}_{t-1}^{-1} &= 2f_t^2,
\end{aligned}$$

we simply obtain

$$s_t = S_t \cdot \nabla_t = y_t^2 - f_t.$$

Illustration 2: time-varying variance is GARCH model

For the time-varying variance model

$$y_t = \mu + \varepsilon_t, \qquad \varepsilon_t \sim \mathrm{NID}(0, f_t),$$

with $t = 1, \ldots, n$, we set $\mu = 0$. The score updating is

$$f_{t+1} = \omega + \beta f_t + \alpha(y_t^2 - f_t), \quad \text{since} \quad s_t = y_t^2 - f_t,$$

and hence score updating implies the GARCH(1,1) model:

$$f_{t+1} = \omega + \beta^* f_t + \alpha^* y_t^2,$$

where $\beta^* = \beta - \alpha$ and $\alpha^* \equiv \alpha$.

If $\alpha = \beta$, we obtain the ARCH model. Generalizations to GARCH(p, q) are obvious.
Generalizations to fat-tailed distributions, are they obvious ?

Volatility model is typically given by

$$y_t = \mu + \sigma(f_t)u_t, \qquad u_t \sim p_u(u_t;\theta), \qquad t = 1, 2, \ldots, n,$$

$$f_{t+1} = \omega + \beta f_t + \alpha s_t,$$

where:

- $\sigma()$ is some continuous function;
- $p_u(u_t;\theta)$ is a standardized disturbance density;
- s_t is the scaled score based on $\partial \log p(y_t | Y_{t-1}, f_t;\theta) \,/\, \partial f_t$.

Some special cases

- $\sigma(f_t) = f_t$ and p_u is Gaussian : score $\Rightarrow$ GARCH;
- $\sigma(f_t) = \exp(f_t)$ and p_u is Gaussian : score $\Rightarrow$ EGARCH;
- $\sigma(f_t) = \exp(f_t)$ and p_u is Student's t : score $\Rightarrow$ t-GAS.

Score-driven models

- The same principle of introducing time-varying parameters in any model for nonlinear and non-Gaussian time series can be applied: e.g. count data.
- Many applications of this have been developed, see `http://www.gasmodel.com`
- Issues on inference require more attention.
- Multivariate versions also need further development.

Thank you

- Still, much work to do !
- Thank you for providing me with the Francqui Chair.
- I am looking forward to work together on research !!

Selected literature

- A.C.Harvey (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press
- G.Kitagawa & W.Gersch (1996). *Smoothness Priors Analysis of Time Series*. Springer-Verlag
- J.Harrison & M.West (1997). *Bayesian Forecasting and Dynamic Models*. Springer-Verlag
- J.Durbin & S.J.Koopman (2012). *Time Series Analysis by State Space Methods, Second Edition*. Oxford University Press
- J.J.F.Commandeur & S.J.Koopman (2007). *An Introduction to State Space Time Series Analysis*. Oxford University Press
- D.Creal, S.J. Koopman and A. Lucas (2013): *Generalized Autoregressive Score Models with Applications*, Journal of Applied Econometrics, 28(5), p.777-795.
- A.C.Harvey (2013) *Dynamic Models for Volatility and Heavy Tails: With Applications to Financial and Economic Time Series*. Cambridge University Press.