Generalized Autoregressive Score Models for Time-varying Parameters in Economics and Finance

Drew Creal[a], Siem Jan Koopman[b], André Lucas[b,c]
plus Francisco Blasques (VU), Pawel Janus (UBS), Bernd Schwaab (ECB)

[a]Booth School of Business University of Chicago
[b]Vrije Universiteit Amsterdam and Tinbergen Institute
[c]Duisenberg school of finance

Research Seminar for National Bank of Serbia, Belgrade, July 6 2012

Motivation and contributions

- Macro Economic and Financial time series often share common features: business cycle dynamics, for example.
- Such time series can be observed at different frequencies.
- Characteristics of the time series can be different.
- Here : observation-driven mixed measurement panel data models.
- The approach allows for non-linear, non-Gaussian models with common factor across different distributions.
- Application I: bivariate model for volatility and dependence with long memory features
- Application II: model for macro time series, credit ratings transitions and loss-given-default (LGD) series. This model incorporates
    1. a time-varying Gaussian model
    2. a time-varying ordered logit
    3. a time-varying beta distribution

Regression model

Regression model :

$$y_t = X_t\beta + \varepsilon_t, \qquad \varepsilon_t \sim \mathrm{NID}(0, \sigma^2), \qquad t = 1, \ldots, n,$$

with usual assumptions:

- parameter vector is $\psi = (\beta; \sigma^2)$.
- estimation by least squares methods.

Let us consider that some coefficients are not constant over time :

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

here we have

- for each t, different value for ψ_t...
- many different unknown values...
- degrees of freedom decreases rapidly !

Time-varying regression model

Time-varying regression model :

$$y_t = X_t\beta_t + \varepsilon_t, \qquad \varepsilon_t \sim \text{NID}(0, \sigma_t^2), \qquad \psi_t = (\beta_t; \sigma_t^2),$$

here ψ_t is allowed to have different values for different t.

Different modelling options for a time-varying ψ_t :

- breaks (different ψ_t in different pre-defined periods)
- Markov-switching structures
- deterministic functions of unknown (exogenous) variables
- functions of time, non-parametric or semi-parametric
- stochastic functions of time

We focus on parametric functions of time, say a random walk :

$$\beta_{t+1} = \beta_t + \eta_t.$$

Time-varying regression model

Time-varying regression parameter model :

$$y_t = X_t \beta_t + \varepsilon_t, \qquad \varepsilon_t \sim \mathrm{NID}(0, \sigma^2), \qquad t = 1, \ldots, n.$$

In case we assume only a stochastic time-varying function for β_t :

$$\beta_{t+1} = \beta_t + \eta_t, \qquad \eta_t \sim \mathrm{NID}(0, \sigma_\eta^2),$$

with some initial condition for β_1.

This is a special case of a state space model.
The Kalman filter and related methods can be used for its analysis.

In a linear Gaussian framework, the methodology for time-varying parameters is well established in econometrics and statistics.

Time Series Analysis by State Space Methods

by J. Durbin and S.J. Koopman (2012, second edition, OUP).

Regression model with time-varying variance

Let us now focus on only a time-varying variance :

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

or

$$y_t = X_t\beta + \sigma_t\varepsilon_t, \qquad \varepsilon_t \sim \text{NID}(0, 1),$$

or

$$y_t = X_t\beta + \exp(\frac{1}{2}h_t)\varepsilon_t, \qquad \varepsilon_t \sim \text{NID}(0, 1),$$

where $h_t = \log \sigma_t^2$ and with time-varying process

$$h_{t+1} = \omega + \phi h_t + \sigma_\eta \eta_t, \qquad \eta_t \sim \text{NID}(0, 1), \qquad t = 1, \dots, n,$$

with some initial condition for h_1.

This is a nonlinear state space model!
Kalman filter and related methods cannot be used for its analysis.

Nonlinear time-varying models

Consider the nonlinear time series model

$$y_t = X_t\beta + \exp(\frac{1}{2}h_t)\varepsilon_t, \qquad \varepsilon_t \sim \mathrm{NID}(0,1),$$

$$h_{t+1} = \omega + \phi h_t + \sigma_\eta \eta_t, \qquad \eta_t \sim \mathrm{NID}(0,1), \qquad t = 1, \ldots, n,$$

with some initial condition for h_1.

This model is known (often $X_t\beta$ is replaced by some constant μ) as the Stochastic Volatility (SV) model which can be used for the modelling of volatility in financial markets.

SV model is related to theoretical models for option pricing.

SV model is often viewed as alternative to the GARCH model: Generalized Autoregressive Conditional Heteroskedasticity

Nonlinear time-varying models

Consider the SV model

$$y_t = \mu + \exp(\frac{1}{2}h_t)\varepsilon_t, \qquad \varepsilon_t \sim \mathrm{NID}(0,1),$$

$$h_{t+1} = \omega + \phi h_t + \sigma_\eta \eta_t, \qquad \eta_t \sim \mathrm{NID}(0,1), \qquad t = 1, \ldots, n,$$

with some initial condition for h_1.

Analysis and parameter estimation for SV models cannot be based on analytical results.
For example, likelihood evaluation is based on

$$p(y) = \int p(y,h)\,\mathrm{d}\,h = \int p(y|h)p(h)\,\mathrm{d}\,h.$$

We may rely on simulation methods, in its crude form

$$\widehat{p}(y) = M^{-1}\sum_{i=1}^{M} p(y|h^i), \qquad h^i \sim p(h).$$

Many developments to report here ...

Time-varying parameter estimation: many approaches

- Extended Kalman filter (Taylor approximations);
  Anderson and Moore (1979).
- Unscented Kalman filter (unscented transformations);
  Julier and Uhlmann (1997).
- Efficient / Simulated Method of Moments;
  Gallant, Hsieh and Tauchen (1997) and Andersen, Chung, and Sorensen (1999).
- Importance sampling (simulated MLE);
  Durbin and Koopman (1997, 2000, 2002), Shephard and Pitt (1997), Danielsson (1997), Richard and Zhang (2007).
- Particle filtering (sequential importance sampling);
  Gordon, Salmond and Smith (1993), Kitagawa (1996).
- Markov chain Monte Carlo (Bayesian estimation);
  Carter and Kohn (1994), de Jong and Shephard (1995).
- Particle MCMC;
  Andrieu, Doucet and Holenstein (2010), Chopin (2012).

Let us return to some basics

Consider model for the data y which we represent as $p(y;\psi)$.
Parameter vector is ψ.
In time series, we evaluate likelihood function via prediction errors

$$p(y;\psi) = p(y_1;\psi)\prod_{t=2}^{n} p(y_t|y_1,\ldots,y_{t-1};\psi).$$

Assume that we want to consider a sub-set of ψ as time-varying :

$$\psi_t = (f_t;\theta),$$

where f_t represents the time-varying parameter and θ the remaining fixed coefficients.

The TV parameter f_t typically represents β_t and/or σ_t.
The TV parameter may be modelled in an autoregressive form

$$f_{t+1} = \omega + Bf_t + A \times \text{" some innovation "}.$$

A new approach

The t-th contribution to the loglikelihood $\ell = \log p(y; \psi)$:

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

where we assume that $f_1, \ldots, f_t$ are known (they are realized).

The parameter value for next period, f_{t+1}, is determined by an autoregressive updating function that has an innovation equal to the score of ℓ_t with respect to f_t.

By determining f_{t+1} in this way, we obtain a recursive algorithm for the estimation of time-varying parameters.

We have labelled this approach as the

generalized autoregressive score model,

or the GAS model. More details are given next.

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 a GAS updating scheme for f_t based on

$$f_{t+1} = \omega + Bf_t + As_t,$$

where the innovation or driving mechanism s_t is given by

$$s_t = S_t \cdot \nabla_t$$

where

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

Volatility modelling

We have

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

The GAS model for f_t can be constructed by considering

$$\begin{aligned} y_t &\sim p(y_t|Y_{t-1}, f_t; \theta), \\ f_{t+1} &= \omega + Bf_t + As_t, \end{aligned}$$

with driving mechanism

$$s_t = S_t \cdot \nabla_t$$

where

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

GAS variance updating reduces to GARCH

Assume $\mu = 0$, we have

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

with variance $f_t = \sigma_t^2$. Score and inverse information matrix are:

$$\begin{aligned}
\ln p(y_t | Y_{t-1}, f_t; \theta) &= -\frac{1}{2}\ln 2\pi - \frac{1}{2}\ln f_t - \frac{y_t^2}{2f_t}, \\
\nabla_t &= \frac{1}{2f_t^2} y_t^2 - \frac{1}{2f_t} \;=\; \frac{1}{2f_t^2}(y_t^2 - f_t), \\
E_{t-1}(\nabla_t) &= 0, \qquad -\mathcal{I}_{t-1} \;=\; -\frac{1}{2f_t^2}, \\
S_t \;=\; \mathcal{I}_{t-1}^{-1} &= 2f_t^2,
\end{aligned}$$

and we have $s_t = S_t \cdot \nabla_t = y_t^2 - f_t$ for the GAS updating

$$f_{t+1} = \omega + Bf_t + A(y_t^2 - f_t).$$

Hence, this GAS update scheme reduces to GARCH for $f_t = \sigma_t^2$:

$$\sigma_{t+1}^2 \;=\; \omega + B\sigma_t^2 + A(y_t^2 - \sigma_t^2) \;=\; \omega + \beta\sigma_t^2 + \alpha y_t^2, \qquad (\beta = B - A).$$

Another example: modelling durations

Consider an exponential ($\mathcal{E}$ is exponential density) model,

$$y_t = \lambda_t \varepsilon_t, \qquad \varepsilon_t \sim \mathcal{E}(1).$$

Let $f_t = \lambda_t$. The score and inverse of the information matrix are:

$$\begin{aligned} \nabla_t &= \frac{y_t}{f_t^2} - \frac{1}{f_t}, \\ S_t = \mathcal{I}_{t-1}^{-1} &= f_t^2. \end{aligned}$$

Here the GAS update scheme reduces to
the E-ACD model of Engle and Russell (1998):

$$f_{t+1} = \omega + A(y_t - f_t) + Bf_t$$

More of such special cases

GAS updating for appropriate observation densities and particular scaling choices reduces to well-known GARCH-type time series models.

- GARCH for $N(0, f_t)$: Engle (1982), Bollerslev (1986)
- EGARCH for $N(0, \exp f_t)$: Nelson (1991)
- Exponential distribution (ACD and ACI): Engle & Russell (1998) and Russell (2001), respectively
- Gamma distribution (MEM): Engle (2002), Engle & Gallo (2006)
- Poisson: Davis, Dunsmuir & Street (2003)
- Multinomial distribution (ACM): Russell & Engle (2005)
- Binomial distribution: Cox (1956), Rydberg & Shephard (2002)

We discuss this general GAS framework in
Creal, Koopman and Lucas (2012, JAE).

What is the use of this ?

- In econometrics, score and Hessian are familiar entities in estimation;
- Using contribution of score at time t only (wrt predictive density) and using it as an innovation in a time-varying parameter scheme is not unreasonable.
- It turns out that many GARCH-type time series models are effectively constructed in this way.
- In case of GARCH (Gaussian), innovation or driver mechanism has an interpretation : $\mathbb{E}(y_t^2) = \sigma^2$.
- In other cases (incl. GARCH with t-densities), choice of driver mechanism is not so clear.
- We then can rely on GAS and still get an appropriate updating scheme.

Statistical properties

The GAS(p, q) model is

$$\begin{aligned} y_t &\sim p(y_t | Y_{t-1}, f_t, f_{t-1}, \ldots, f_{t-q}; \theta), \\ f_{t+1} &= \omega + \sum_{i=0}^{p-1} A_i s_{t-i} + \sum_{j=0}^{q-1} B_j f_{t-j} \\ s_t &= S_t \cdot \nabla_t \end{aligned}$$

- The expectation of the score is zero: $E_{t-1}[\nabla_t] = 0$.
- As a result, s_t is a martingale difference sequence.
- If f_t is stationary, its unconditional expectation is $\mathrm{E}[f_t] = \omega \left(I - B(1)\right)^{-1}$.
- We have established conditions for stationarity and ergodicity (SE): Blasques, Koopman and Lucas (2012).
- It will facilitate derivations for asympotic properties.
- Other scalings S_t, other link functions, will lead to different properties.

Different specifications

$$
\begin{aligned}
y_t &\sim p(y_t | Y_{t-1}, f_t, f_{t-1}, \ldots, f_{t-q}; \theta), \\
f_{t+1} &= \omega + \sum_{i=0}^{p-1} A_i s_{t-i} + \sum_{j=0}^{q-1} B_j f_{t-j} \\
s_t &= S_t \cdot \nabla_t
\end{aligned}
$$

- The default choice for scaling is $S_t = \mathcal{I}_{t-1}^{-1}$ or $S_t = \mathcal{I}_{t-1}^{-1/2}$.
- Alternative: $S_t = I$; "steepest descent" is less stable...
- In case default choice is close to singular, we can do some mild smoothing of past $\mathcal{I}_t$'s using an EWMA scheme:

$$
\mathcal{I}_{t-1}^c = \tilde{\alpha} \mathcal{I}_{t-2}^c + (1 - \tilde{\alpha}) \mathcal{I}_{t-1},
$$

  and $S_t = (\mathcal{I}_{t-1}^c)^{-1}$. This appears to work very effectively.
- Extensions with long-memory: Janus, Koopman and Lucas (2012).

Volatility GAS models

- A class of volatility models is given by

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

$$f_{t+1} = \omega + \beta f_t + \alpha s_t, \qquad (2)$$

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

Introducing FIGAS

The Fractionally Integrated Generalized Autoregressive Score (FIGAS) model is given by

$$y_t \sim p(y_t | Y_{t-1}, f_t; \theta), \qquad t = 1, 2, \dots, T, \tag{3}$$

$$f_t^* = (1 - L)^d f_t, \qquad f_{t+1}^* = \omega + \beta f_t^* + \alpha s_t, \tag{4}$$

where:

- y_t denotes dependent variable; $Y_t = [y_1, \dots, y_t]'$;
- f_t is the time-varying parameter of interest;
- θ collects static parameters;
- d is the fractional integration order;
- $(1 - L)^d = 1 - dL + \frac{d(d-1)}{2!} L^2 - \frac{d(d-1)(d-2)}{3!} L^3 + \dots$
- s_t is the scaled score based on $\partial \log p(y_t | Y_{t-1}, f_t; \theta) \, / \, \partial f_t$.

General FIGAS specification

■ We introduce time-varying parameters with long memory properties in a bivariate heavy-tailed distribution for a set of stock equity returns.

- heavy-tails in returns with different tail properties;
- outliers for marginal and/or joint densities should not dillute volatility and/or correlation processes; especially relevant for long memory features;
- tail dependence is modeled explicitly.

Our approach :

■ We model marginal series by means of conditonal Student's t densities and we model dependence by means of a t copula.

■ The score function in the Student's t class of distributions depends on conditional weights that downweight extreme observations.

■ The degrees of freedom parameter for the Student's t distribution handles the level of robustness for statistical inference.

FIGAS for conditional variance

- Let y_t denote (demeaned) log-return of some asset, assume

$$y_t = \sigma_t \varepsilon_t, \qquad \varepsilon_t \sim \text{Student's } t_\nu(0,1),$$

with loglikelihood function given by

$$\ell_t = c(\nu) - \frac{1}{2}\log(\pi) - \frac{1}{2}\log(\sigma_t^2) - \frac{\nu+1}{2}\log\left(1 + \frac{y_t^2}{(\nu-2)\sigma_t^2}\right),$$

where $c(\nu) = \log\left\{\Gamma\left(\frac{\nu+1}{2}\right)/\Gamma\left(\frac{\nu}{2}\right)\right\} - \frac{1}{2}\log(\nu-2)$ and $\nu > 2$.

- Let $f_t = \log(\sigma_t^2)$, we have

$$\nabla_t = \frac{1}{2\,\sigma_t^2}\left[\omega_t\, y_t^2 - \sigma_t^2\right] \qquad \text{and} \qquad \mathcal{I}_t = \frac{1}{2}\frac{\nu}{\nu+3},$$

where

$$\omega_t = \frac{\nu+1}{\nu-2+y_t^2/\sigma_t^2} \in [0, (\nu+1)/(\nu-2)].$$

Time t weight ω_t attains zero if y_t^2 too large relative to current level of volatility.

FIGAS for conditional variance : the resulting model

The FIGAS model is then given by :

- demeaned log-return of some asset :

$$y_t = \sigma_t \varepsilon_t, \qquad \varepsilon_t \sim \text{ Student's } t_\nu(0,1),$$

  with loglikelihood function given by

$$\ell_t = c(\nu) - \frac{1}{2}\log(\pi) - \frac{1}{2}\log(\sigma_t^2) - \frac{\nu+1}{2}\log\left(1 + \frac{y_t^2}{(\nu-2)\sigma_t^2}\right),$$

  where $\sigma_t^2 = \exp(f_t)$.
- log-variance is updated :

$$f_{t+1}^* = \omega + \beta f_t^* + \alpha s_t, \qquad f_t^* = (1-L)^d f_t,$$

  where the scaled score is given by

$$s_t = \mathcal{I}_t^{-\frac{1}{2}} \nabla_t, \qquad \nabla_t = \frac{1}{2\,\sigma_t^2}\left[\omega_t\, y_t^2 - \sigma_t^2\right] \qquad \text{and} \qquad \mathcal{I}_t = \frac{1}{2}\frac{\nu}{\nu+3}.$$

- FIGAS with leverage (FIGASL) : $\alpha \Rightarrow \alpha + \gamma 1_{(y_t<0)}$.

Conditional volatility

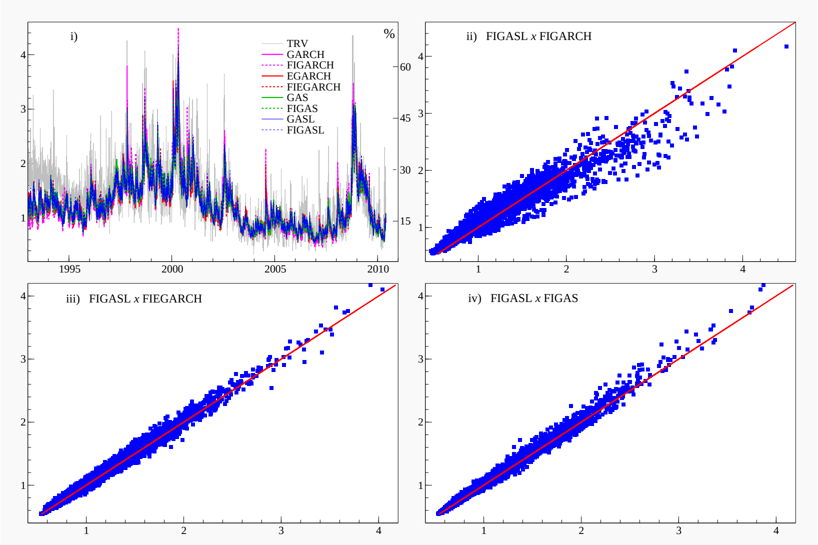


Figure 1: Estimated vol for P&G daily returns over January 4, 1993 to May 28, 2010

Robust filtering of volatility: the role of weight ω_t

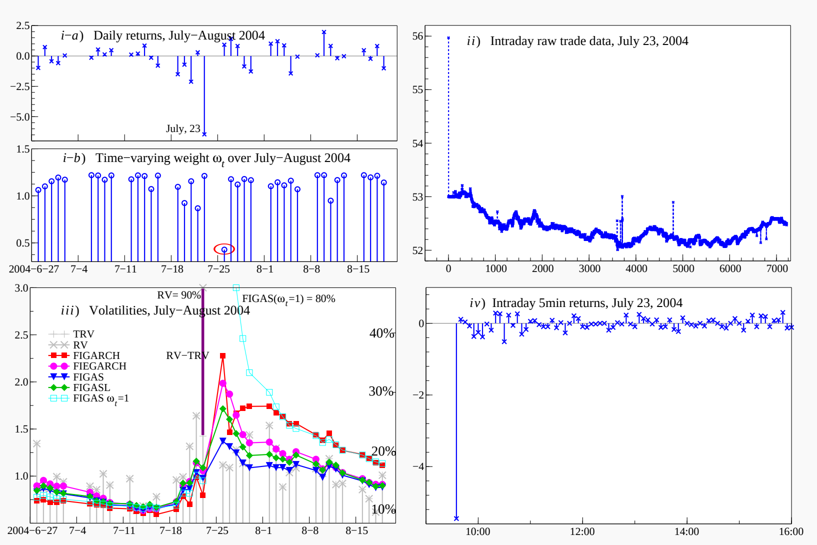


Figure 2: P&G case study

FIGAS for bivariate conditional dependence

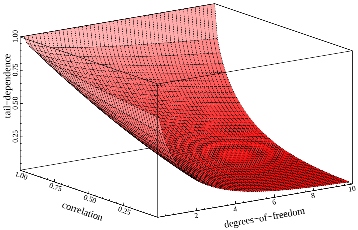
- for dependence between two marginal series : bivariate t copula

$$(1-\rho_t^2)^{-\frac{1}{2}} \frac{\Gamma(\frac{\eta+2}{2})\Gamma(\frac{\eta}{2})}{\left[\Gamma(\frac{\eta+1}{2})\right]^2} \frac{\left(1+\frac{1}{\eta(1-\rho_t^2)}\left(x_{1t}^2+x_{2t}^2-2\rho_t x_{1t}x_{2t}\right)\right)^{-\frac{\eta+2}{2}}}{\prod_{i=1}^{2}\left(1+x_{it}^2/\eta\right)^{-\frac{\eta+1}{2}}},$$

where $x_{it} = t_\eta^{-1}(u_{it})\,, i=1,2,\ u_{it}\in(0,1),\ \rho_t\in(-1,1)$ and $\eta>0$.

- t copula captures tail dependence which is governed by ρ_t and η

- extreme occurences of x_{1t} and/or x_{2t} can be due to heavy-tail nature (low η) of the t copula, not neccessarily due to high ρ_t:



FIGAS for bivariate conditional dependence

Define $f_t = \log\left(1+\rho_t \,/\, 1-\rho_t\right) \in \mathbb{R}$, we have

$$
\begin{aligned}
\nabla_t &= \frac{\dot{\rho}_t}{(1-\rho_t^2)^2}\left[(1+\rho_t^2)\big(\pi_t x_{1t} x_{2t} - \rho_t\big) - \rho_t\big(\pi_t x_{1t}^2 + \pi_t x_{2t}^2 - 2\big)\right],\\
\mathcal{I}_t &= \frac{\dot{\rho}_t^2}{(1-\rho_t^2)^2}\left(1+\rho_t^2 - \frac{2\rho_t^2}{\eta+2}\right)\frac{\eta+2}{\eta+4},
\end{aligned}
$$

where $\dot{\rho}_t$ is derivative of ρ_t wrt f_t, with time-dependent weight defined as

$$\pi_t = \frac{\eta+2}{\eta+m_t} \in [0, (\eta+2)/\eta],$$

where

$$m_t = x_t{}' R_t^{-1} x_t \geq 0, \quad \text{with} \quad x_t = [x_{1t}\ x_{2t}]' \quad \text{and} \quad R_t = \begin{pmatrix} 1 & \rho_t \\ \rho_t & 1 \end{pmatrix}.$$

For a finite η, extreme observations x_{1t} and/or x_{2t} leading to a large Mahalanobis distance m_t will, as the result of downweighting via π_t, have limited impact on the correlation dynamics.

FIGAS for bivariate conditional dependence

The FIGAS model for dependence is then given by :

- The t-copula is given by (27) with

$$\rho_t = \frac{1 - \exp f_t}{1 + \exp f_t},$$

- logit-dependence is updated :

$$f^*_{t+1} = \omega + \beta f^*_t + \alpha s_t, \qquad f^*_t = (1-L)^d f_t,$$

where the scaled score is given by

$$s_t = \mathcal{I}_t^{-\frac{1}{2}} \nabla_t,$$

where

$$\begin{aligned}
\nabla_t &= \frac{\dot{\rho}_t}{(1-\rho_t^2)^2}\left[(1+\rho_t^2)\left(\pi_t x_{1t} x_{2t} - \rho_t\right) - \rho_t\left(\pi_t x_{1t}^2 + \pi_t x_{2t}^2 - 2\right)\right], \\
\mathcal{I}_t &= \frac{\dot{\rho}_t^2}{(1-\rho_t^2)^2}\left(1 + \rho_t^2 - \frac{2\rho_t^2}{\eta+2}\right)\frac{\eta+2}{\eta+4}.
\end{aligned}$$

Conditional dependence

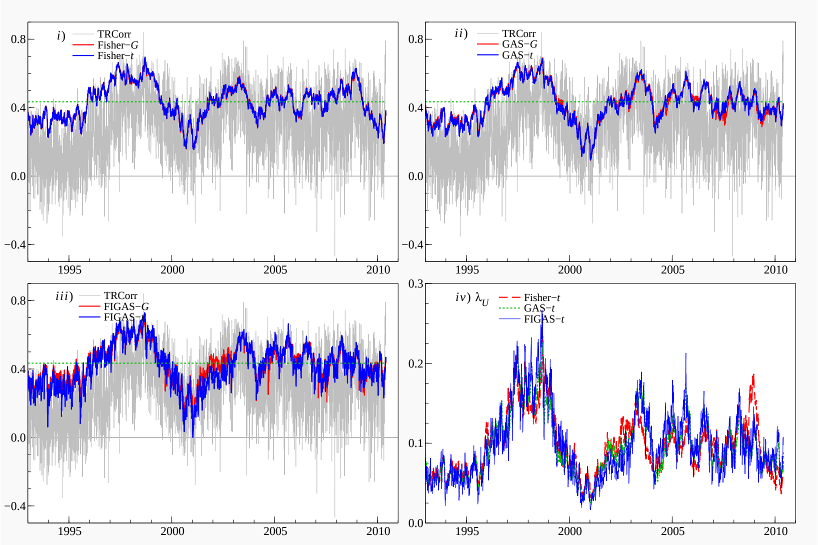


Figure 3: Estimated correlation for GE/KO daily returns over January 4, 1993 to May 28, 2010

Robust filtering of correlation: the role of π_t

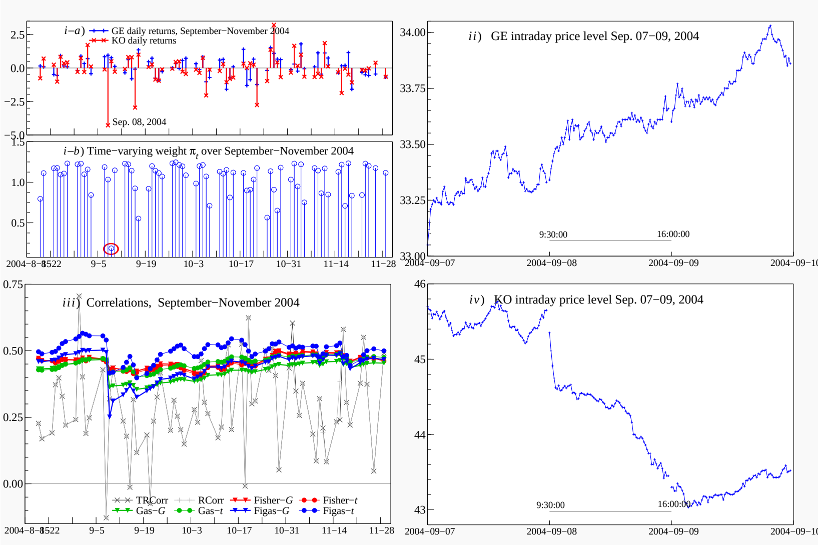


Figure 4: GE/KO case study

Credit Risk Application with Macro and Finance factors

- Economic time series often share common features, e.g. business cycle dynamics.
- Economic time series may be continuous and/or discrete and be observed at different frequencies.
- We introduce
  observation-driven mixed measurement panel data models
- The approach allows for non-linear, non-Gaussian models with common factor across different distributions.
- Application: we develop a models for credit ratings transitions and loss-given-default (LGDs) with macro factors.
- The models include:
  1. Time-varying Gaussian model
  2. Time-varying ordered logit
  3. Time-varying beta distribution

Mixed measurement panel data models

We introduce mixed measurement observation driven models

$$\begin{aligned} y_{it} &\sim p_i(y_{it}|f_t, Y_{t-1};\psi), \quad i=1,\ldots,N, \\ f_{t+1} &= \omega + Bf_t + As_t \end{aligned}$$

The score function is

$$\begin{aligned} s_t &= S_t \nabla_t \\ \nabla_t &= \sum_{i=1}^{N} \delta_{it} \nabla_{i,t} = \sum_{i=1}^{N} \delta_{it} \frac{\partial \log p_i(y_{it}|f_t, Y_{t-1};\psi)}{\partial f_t}, \end{aligned}$$

- The observations y_{it} may come from different distributions.
- The factors f_t may be common across distributions.
- KEY: The score function allows us to pool information from different observations to estimate the common factor f_t.
- δ_{it} is an indicator function equal to 1 if y_{it} is observed and zero otherwise. Missing values are naturally taken into account.

Scaling matrix

Consider the eigenvalue-eigenvector decomposition of Fisher's (conditional) information matrix

$$\mathcal{I}_t \quad = \quad \mathbb{E}_{t-1}[\nabla_t \nabla_t'] \; = \; U_t \Sigma_t U_t',$$

The scaling matrix is then defined as

$$S_t \quad = \quad U_t \Sigma_t^{-1/2} U_t'$$

- S_t is then the "square root" of a generalized inverse.
- The innovations s_t driving f_t have an identity covariance matrix, when the info. matrix is non-singular.
- The conditional information matrix is additive for our models:

$$\mathcal{I}_t \quad = \quad \mathbb{E}_{t-1}[\nabla_t \nabla_t'] \; = \; \sum_{i=1}^{N} \delta_{it} \, \mathbb{E}_{i,t-1}[\nabla_{it} \nabla_{it}'].$$

Log-likehood function and ML estimation

- The log-likelihood function for an observation-driven model can easily be computed.
- The ML estimator is

$$\hat{\psi} \quad = \quad \arg\max_{\psi} \sum_{t=1}^{T} \sum_{i=1}^{N} \delta_{it} \log p_i(y_{it} | f_t, Y_{t-1}; \psi),$$

- Estimation is similar to a GARCH model.

Credit risk

- Growing econometrics literature on models for credit risk: McNeil et al. (2005), Bauwens and Hautsch (JFEct, 2006), Gagliardini and Gourieroux (JFEct, 2005), Koopman Lucas and Monteiro (JEct, 2008), Duffie et al. (JFE, JoF 2008).
- Basic observations:
  1. Probability of default varies over time with the business cycle.
  2. Conditional on default, the loss (recovery rate) varies with the business cycle.
  3. We observe excess clustering of defaults and ratings transitions beyond what can be explained by simply adding covariates.
  4. The literature focuses on a credit risk or frailty factor.
- Industry standard models are too simple to capture these features.
- New models in the literature are parameter driven models requiring simulation methods for estimation.
- We provide observation driven alternatives.

Data: Moody's and FRED

- We observe data from Jan. 1980 to March 2010.
- $7,505$ companies are rated by Moody's.
- We pool these into 5 ratings categories (IG, BB, B, C, D).
- We observe transitions, e.g. IG $\rightarrow$ BB or C $\rightarrow$ D
- There are $J = 16$ total types of transitions.
- 19,450 total credit rating transitions.
- 1,342 transitions are defaults.
- 1,125 measurements of loss-given default (LGD).
- LGD is the fraction of principal an investor loses when a firm defaults.
- We also observe six macroeconomic variables: industrial production growth, credit spread, unemployment, annual S&P500 returns, realized volatility, real GDP growth (qtly).

Models

- Credit ratings can be modeled using the (static) ordered probit model of CreditMetrics; one of the current industry standards, see Gupton Stein (2005).
- LGD's are often modeled by (static) beta distributions.
- GOAL: Build models that improve on current industry standards and are (relatively) easy to implement and estimate.
    1. Time-varying Gaussian model
    2. Time-varying ordered logit
    3. Time-varying beta distribution
- Forecasting credit risk.
- Simulation of loss distributions and scenario analysis.
- Bank executives and regulators and can use them for "stress testing."

Mixed measurement model for credit risk

$$
\begin{aligned}
y_t^m &\sim \mathrm{N}\left(\mu_t, \Sigma_m\right) \\
y_{i,t}^c &\sim \text{Ordered Logit}\left(\pi_{ijt}, j \in \{\text{IG, BB, B, C, D}\}\right), \\
y_{k,t}^r &\sim \text{Beta}\left(a_{kt}, b_{kt}\right), \qquad k = 1, \ldots, K_t,
\end{aligned}
$$

- y_t^m are the macro variables.
- $y_{i,t}^c$ are indicator variables for each credit rating j for firm i.
- $y_{k,t}^r$ are the LGDs for the k-th default.
- K_t are the number of defaults in period t.
- μ_t, π_{ijt}, and (a_{kt}, b_{kt}) are functions of an $M \times 1$ vector of factors f_t.

Time varying Gaussian model for macro data

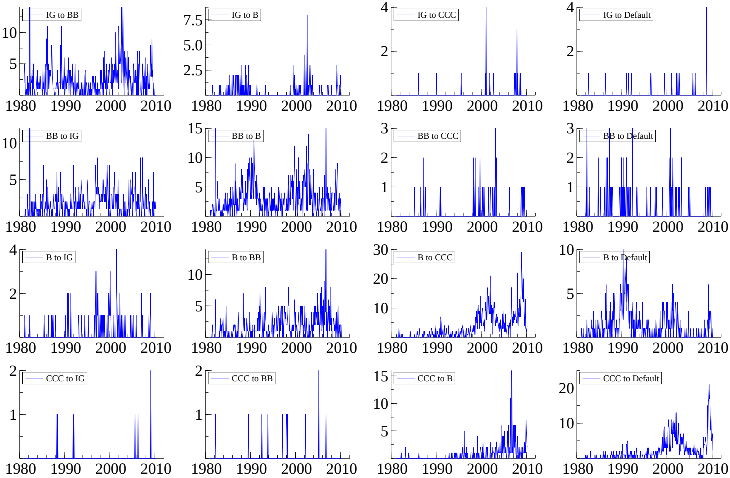
$$
\begin{aligned}
y_t^m &\sim \mathrm{N}\left(\mu_t, \Sigma_m\right), \\
\mu_t &= Z^m f_t.
\end{aligned}
$$

- Z^m is a $(6 \times M)$ matrix of factor loadings.
- Σ_m is a (6×6) diagonal covariance matrix.
- $\tilde{S}_t$ is a selection matrix indicating which macro variables are observed at time t.

$$
\begin{aligned}
\nabla_t^m &= \left(\tilde{S}_t Z^m\right)' \left(\tilde{S}_t \Sigma_m \tilde{S}_t'\right)^{-1} \tilde{S}_t \left(y_t^m - \mu_t\right), \\
\mathcal{I}_t^m &= \left(\tilde{S}_t Z^m\right)' \left(\tilde{S}_t \Sigma_m \tilde{S}_t'\right)^{-1} \tilde{S}_t Z^m.
\end{aligned}
$$

Moody's monthly credit ratings transitions

The data have been pooled together each month.



Time-varying ordered logit

$$\begin{aligned}
y^c_{i,t} &\sim \text{Ordered Logit}\left(\pi_{ijt}, j \in \{\text{IG, BB, B, C, D}\}\right), \\
\pi_{ijt} &= \mathrm{P}\left[R_{i,t+1} = j\right] = \tilde{\pi}_{ijt} - \tilde{\pi}_{i,j-1,t}, \\
\tilde{\pi}_{ijt} &= \mathrm{P}\left[R_{i,t+1} \leq j\right] = \frac{\exp(\theta_{ijt})}{1 + \exp(\theta_{ijt})}, \\
\theta_{ijt} &= z^c_{ijt} - Z^{c\prime}_{it} f_t.
\end{aligned}$$

- $J^c = 5$ categories $j \in \{\text{IG, BB, B, C, D}\}$.
- R_{it} is the rating for firm i at the start of month t.
- y^c_{it} is an indicator variable for each rating type.
- π_{ijt} is the probability that firm i is in category j.
- $\tilde{\pi}_{i,\text{D},t} = 0$ and $\tilde{\pi}_{i,\text{IG},t} = 1$.
- To our knowledge, a time-varying ordered logit model is new.

Time-varying ordered logit

The contribution to the log-likelihood at time t is

$$\ln p_i(y_{it}^c | f_t, Y_{t-1}; \psi) \quad = \quad \sum_{i=1}^{N_t} \sum_{j=1}^{J^c} y_{ijt}^c \log\left(\pi_{ijt}\right)$$

The score and information matrices are

$$\begin{aligned}
\nabla_t^c &= -\sum_{i=1}^{N_t} \sum_{j=1}^{J^c} \frac{y_{ijt}^c}{\pi_{ijt}} \cdot \dot{\pi}_{ijt} \cdot Z_{it}^c, \\
\mathcal{I}_t^c &= \sum_{i=1}^{N_t} n_{it} \left(\sum_j \frac{\dot{\pi}_{ij,t}^2}{\pi_{ij,t}} \right) Z_{it}^c Z_{it}^{c\prime}
\end{aligned}$$

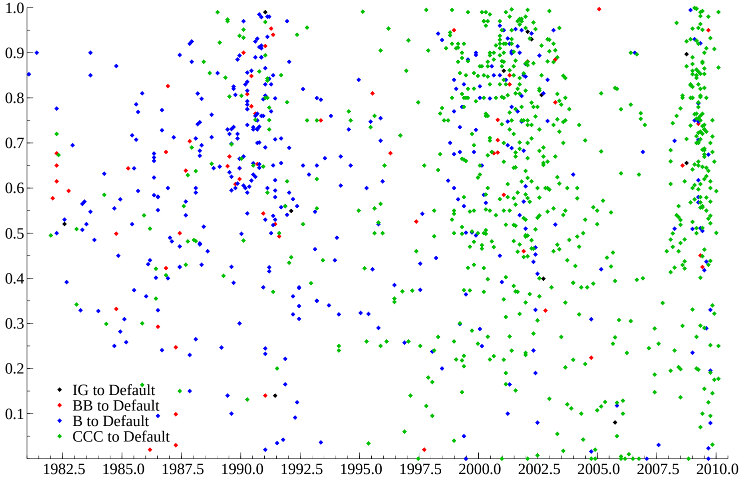
where

$$\dot{\pi}_{ijt} = \tilde{\pi}_{ijt}\left(1 - \tilde{\pi}_{ijt}\right) - \tilde{\pi}_{i,j-1,t}\left(1 - \tilde{\pi}_{i,j-1,t}\right).$$

Loss given default

- When a firm defaults, investors typically lose a fraction of their investment (alternatively, they recover a fraction of their investment).
- The fraction of losses experienced by investors also varies with the business cycle.
- We develop a new model for a time-varying beta distribution.
- See McNeil and Wendin (2007 JEmpFin) for Bayesian inference in a state space model.

Loss given default by transition type



Time-varying beta distribution

$$
\begin{aligned}
y^r_{k,t} &\sim \text{Beta}\left(a_{kt}, b_{kt}\right), \qquad k = 1, \ldots, K_t, \\
a_{kt} &= \beta_r \cdot \mu^r_{kt} \\
b_{kt} &= \beta_r \cdot (1 - \mu^r_{kt}) \\
\log\left(\mu^r_{kt} / \left(1 - \mu^r_{kt}\right)\right) &= z^r + Z^r f_t.
\end{aligned}
$$

- We observe $K_t \geq 0$ defaults at time t.
- $0 < y^r_{k,t} < 1$ is the amount lost conditional on the k-th default.
- μ^r_{kt} is the mean of the beta distribution.
- z^r is the unconditional level of LGDs.
- Z^r is a $(1 \times M)$ vector of factor loadings.
- β_r is a scalar parameter

Time-varying beta distribution

The contribution to the log-likelihood at time t is

$$\begin{aligned} \ln p_i(y_{kt}^r | f_t, Y_{t-1}; \psi) &= \sum_{k=1}^{K_t} (a_{kt} - 1) \log\left(y_{kt}^r\right) + (b_{kt} - 1) \log\left(1 - y_{kt}^r\right) \\ &\quad - \log\left[B\left(a_{kt}, b_{kt}\right)\right] \end{aligned}$$

The score and information matrices are

$$\begin{aligned} \nabla_t^r &= \beta_r \sum_{k=1}^{K_t} \mu_{kt}^r (1 - \mu_{kt}^r) \left(Z^r\right)' (1, -1) \left(\left(\log(y_{kt}^r), \log(1 - y_{kt}^r)\right)' - \dot{B}\left(a_{kt}, b_{kt}\right) \right) \\ \mathcal{I}_t^r &= \beta_r \sum_{k=1}^{K_t} \left(\mu_{kt}^r (1 - \mu_{kt}^r)\right)^2 \left(Z^r\right)' (1, -1) \left(\ddot{B}\left(a_{kt}, b_{kt}\right) \right) (1, -1)' Z^r \end{aligned}$$

where

$$\sigma_{kt}^2 = \mu_{kt}^r \cdot (1 - \mu_{kt}^r)/(1 + \beta_r).$$

Estimation details

- The macro data y_t^m has been standardized.
- We consider models with $p = 1$ and $q = 1$ factor dynamics.
- For identification of the level parameters, we set $\omega = 0$ in the factor recursion:

$$f_{t+1} = A_1 s_t + B_1 f_t$$

- For identification of the factors, we also impose restrictions on Z^m, Z^c, and Z^r.
- Some parameters have been pooled for "rare" transitions; e.g., IG $\rightarrow$ D and BB $\rightarrow$ D.
- Moody's re-defined several categories in April 1982 and Oct. 1999 causing incidental re-ratings (outliers), which we handle via dummy variables for these dates.

AIC, BIC, and log-likelihoods for different models

	(2,0,0)	(2,1,0)	(2,2,0)	(3,0,0)
log-Like	-40447.9	-40199.1	-40162.8	-40056.2
AIC	81005.9	80520.1	80457.0	80242.4
BIC	81640.0	81223.0	81218.0	80991.0
	(3,1,0)	(3,2,0)	(3,1,1)	(3,2,1)
log-Like	-39817.1	-39780.8	-39812.6	**-39780.0**
AIC	79776.2	**79713.6**	79771.2	79716.0
BIC	80594.0	**80589.0**	80612.0	80615.0

The number of factors for each data type are represented by (m, c, r).

Parameter estimates for the (3,2,0) model

Macro loadings Z^m

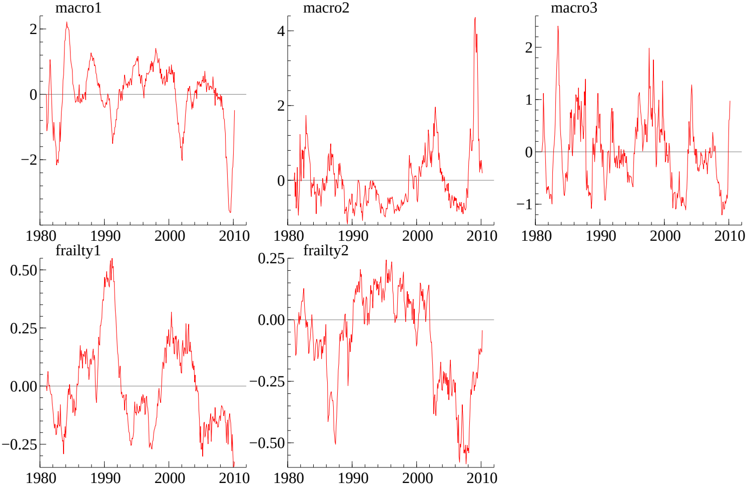
	macro_1	macro_2	macro_3	frailty_1	frailty_2
IP	1.000	0.000	0.000	0.000	0.000
	—	—	—	—	—
UR	-0.892***	0.122***	-0.062*	0.000	0.000
	(0.037)	(0.041)	(0.040)	—	—
RGDP	0.811***	0.072	0.336***	0.000	0.000
	(0.066)	(0.079)	(0.074)	—	—
Cr.Spr.	-0.169**	1.000	0.000	0.000	0.000
	(0.085)	—	—	—	—
$r_{S\&P}$	0.049	-0.268***	1.223***	0.000	0.000
	(0.093)	(0.081)	(0.093)	—	—
$\sigma_{S\&P}$	-0.007	0.648***	1.000	0.000	0.000
	(0.107)	(0.084)	—	—	—

Parameter estimates for the (3,2,0) model

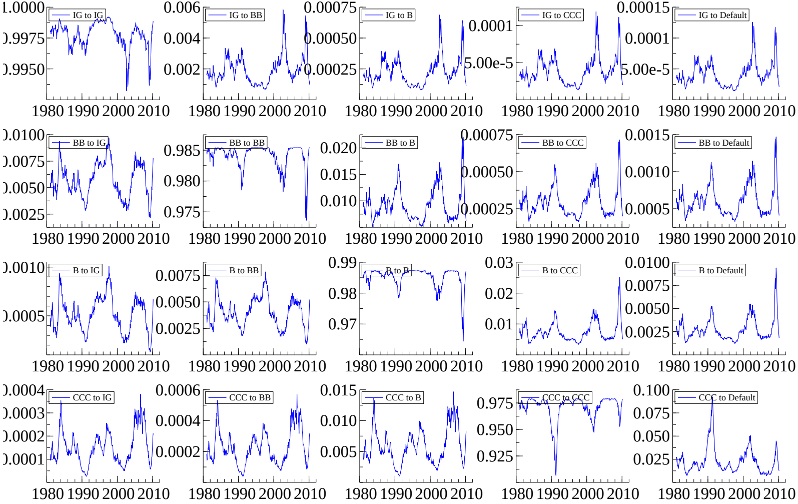
Credit rating and LGD loadings Z^c and Z^r

	macro_1	macro_2	macro_3	frailty_1	frailty_2
Z^c					
IG	-0.052	0.202***	-0.123**	1.475***	-1.165**
	(0.059)	(0.055)	(0.069)	(0.371)	(0.555)
BB	-0.078**	0.172***	-0.102***	1.000	0.000
	(0.037)	(0.037)	(0.040)	—	—
B	-0.184***	0.162***	-0.142***	0.970***	-0.016
	(0.035)	(0.031)	(0.040)	(0.156)	(0.158)
CCC	-0.262***	0.073*	-0.018	1.936***	1.000
	(0.057)	(0.050)	(0.075)	(0.465)	—
Z^r	0.018	0.276***	-0.082*	1.212***	1.065***
	(0.049)	(0.046)	(0.062)	(0.376)	(0.301)

Estimated factors for the (3,2,0) model



Time-varying transition probabilities



observation-driven mixed measurement panel data model

- We have introduced a new class of models :
  observation-driven mixed measurement panel data models
- Time-variation is based on the score :
  the principle of the score is transparent
- Highly flexible and relatively easy to implement :
  in the basics, it is just like GARCH
- We are working on theoretical foundations:
  Blasques, Koopman and Lucas (2012)
- Number of applications are widespread

Thank you !