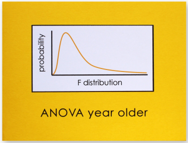
Modelling time-varying nonlinear dependence

Bård Støve
joint work with
Kristian Gundersen   Timothee Bacri   Jan Bulla   Sondre Hølleland

UNIVERSITY OF BERGEN

Congratulations, Nils!



- Best wishes from the stats group at UiB!
- Significant contributions in the field of statistics
- Nils always shows interest and creates a pleasant environment, being seminars/conferences/conversations.... → source of inspiration!
- Scientific connection:
  - I started out in nonparametric density/regression estimation[1]
  - This talk selected due to the inspiration from Nils' work[2]

[1] Glad & Hjort (1995). Nonparametric density estimation with a parametric start, *AoS*, inspiration for my PhD thesis

[2] Hjort & Jones (1996). Locally parametric nonparametric density estimation, *AoS*

Goal

- Investigate the possible time-varying nonlinear dependence structures of a pair of stochastic variables (motivated by returns from financial markets)
- Combine a regime-switching model with the local Gaussian correlation (LGC)
  1. Describe the regimes by a Hidden Markov Model (HMM)
  2. Test whether the dependence structure between the variables differs across the regimes

Existing literature:
Regime-switching copulas e.g. Okimoto (2008), Chollete et al (2009) and BenSaïda et al. (2018),...

Empirical analysis S&P500 vs. FTSE100

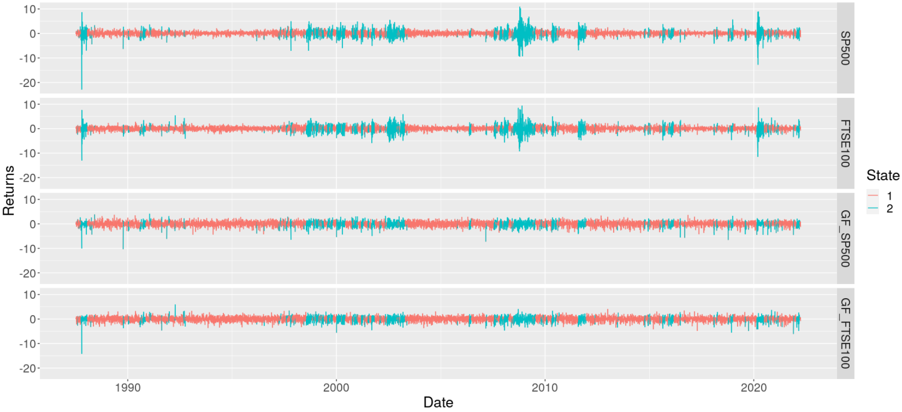


Figure: Daily log-returns and GARCH-filtrated log-returns (lower two plots) of S&P500 and FTSE100 (based on closing index prices) with classification in two regimes using a Gaussian bivariate HMM.

Local Gaussian Correlation (LGC)

Introduced by Tjøstheim & Hufthammer (2013). The central idea:

- Approximate a bivariate density f of $R = (R_1, R_2)$ at **a point** $x = (x_1, y_2)$ by a bivariate Gaussian density $\psi(v, \mu(x), \Sigma(x))$. We take the correlation $\rho(x)$ parameter of that Gaussian density as our measure of local dependence, and we call it the ***local Gaussian correlation***.
- As we move to another point $x' = (x'_1, y'_2)$ of f, another Gaussian $\psi(v, \mu(x'), \Sigma(x'))$ is required to approximate f in a neighbourhood A' of x'.
- In this way the dependence in f is described by the family of Gaussian distributions $\{\psi(v, \mu(x), \Sigma(x))\}$ and the associated correlations $\{\rho(x)\}$.

Note: $\mu(x) = (\mu_1(x), \mu_2(x))^T$ is the local mean vector and $\Sigma(x) = (\sigma_{ij}(x))$ is the local covariance matrix. The local correlation at the point x is $\rho(x) = \frac{\sigma_{12}(x)}{\sigma_{11}(x)\sigma_{12}(x)}$.

Local Gaussian approximation

Tjøstheim & Hufthammer (2013) demonstrated that for a given neighbourhood characterized by a **bandwidth parameter** b the local population parameters $\gamma(x) = (\mu_1(x), \mu_2(x), \sigma_1^2(x), \sigma_2^2(x), \rho(x))$ can be defined by minimizing a likelihood related **penalty function** q given by

$$q = \int K_b(v - x)\,[\psi(v, \gamma(x)) - \log \psi(v, \gamma(x)) f(v)]\,dv, \qquad (1)$$

where $K_b(v - x) = b^{-1}K(b^{-1}(v - x))$ with K being a kernel function. Such a penalty function was used in **Hjort & Jones (1996)** for density estimation purposes.

Define the population value $\gamma(x) = \gamma_b(x)$ as the the minimizers of this penalty function, as it value depends on b.

Numerical maximization of the accompanying **local likelihood**, leads to local likelihood estimates $\gamma_{T,b}(x)$.

Example: Clayton copula

Simulated observations from a Clayton copula with parameter equal 2, standard normal marginals, T=784

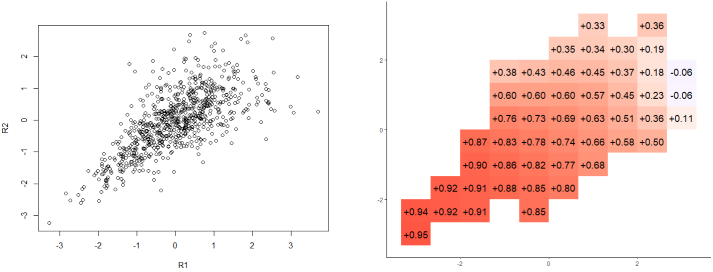


Figure: The observations (left) and estimated LGC map (right)

Extensions

The point of using the **local Gaussian approximation** is that one can move away from global Gaussian distributions and describe much more general situations, also multivariate thick tailed distributions like those met in finance. At the same time one can **exploit** much of the multivariate **Gaussian theory locally**.

- multivariate and conditional density estimation, Otneim & Tjøstheim (2017,2018)
- Various independence tests for iid and time series data, Berentsen & Tjøstheim (2014), Lacal & Tjøstheim (2017, 2019)
- Copula GOF tests, Berentsen, Støve, Tjøstheim and Nordbø (2014)
- Nonlinear spectral analysis: A local Gaussian approach, Jordanger & Tjøstheim (2022)
- The local Gaussian Partial correlation, Otneim & Tjøstheim (2019)
- Applications in finance, Støve & Tjøstheim (2014), Sleire et al (2022)
- R package: 'lg', Otneim (2021)

Hidden Markov Models

Observed time series $\{R_t : t = 1, \ldots, T\}$. A mixture of conditional distributions is assumed to be driven by an unobserved (hidden) homogeneous Markov chain, with regimes $\{C_t : t = 1, \ldots, T\}$.
m-regime Gaussian HMMs with conditional distribution specified by

$$p_k(r) = \mathrm{P}(R_t = r | C_t = k) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{r-\mu_k}{\sigma_k}\right)^2},$$

with parameters $(\mu_k, \sigma_k), k = 1, \ldots, m$.
Use the R-package TMB for the estimation of the HMM parameters:

- Relies on Automatic Differentiation and Laplace approximation
- Approximates log-likelihood and produces gradient and Hessian

See Bacri et al (2022,23) - avoid the commonly used EM-algorithm.
Use **local decoding** for determining the **most probable regime at time** t, cfr. Zucchini et al (2016).

Regime-switching and LGC

Two-step procedure:

1. Fit bivariate m (two)-regime Gaussian* HMM to the observations
   - Classify each pair of observations in one of the regimes by local decoding
2. Estimate the LGC separately for the m (two) regimes
   - Estimate the LGC maps on a pre-defined grid using the (GARCH-filtrated**) observations

Allows us to test whether the LGC maps are equal or not.

* or your favorite distribution

** if needed

Test of equal dependence in two regimes

Let ρ_1 and ρ_2 be the LGC map from regime 1 and 2, respectively, estimated in the gridpoints (x_i, y_j).

$$H_0: \quad \rho_1(x_i, y_j) = \rho_2(x_i, y_j) \quad \text{for} \quad i, j = 1, \cdots, n$$
$$H_1: \quad \rho_1(x_i, y_j) \neq \rho_2(x_i, y_j) \quad \text{for} \quad i, j = 1, \cdots, n$$

Test statistic:

$$D_2 = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=i}^{n} \left[\hat{\rho}_1(x_i, y_j) - \hat{\rho}_2(x_i, y_j)\right]^2 w(x_i, y_j)$$

w is a weight function, which permits to screen off parts of the local correlation or to concentrate on a certain region.

We **approximate the sampling distribution** of D_2 by **bootstrap** samples, as the asymptotic theory for similar test statistics (e.g. Berentsen & Tjøstheim (2014)) is not very accurate unless the number of observations is very large.

Level study of the proposed test

- The same data generating process (DGP) is used for both regime 1 and 2, hence H_0 is true.
- For every DGP we use two Gaussian marginal distributions, each with a mean equal to zero and a standard deviation equal to four, but with a different copula.
- Each table entry is based on 1000 replications, each with 400 observations.

	Nominal level (α)		
(r)1-4 Model	0.01	0.05	0.1
1. Clayton copula, $\theta = 1$	0.017	0.056	0.102
2. Clayton copula, $\theta = 2$	0.012	0.059	0.116
3. Gaussian copula, $\rho = -0.5$	0.011	0.045	0.094
4. Gaussian copula, $\rho = 0.3$	0.007	0.058	0.116
5. Gumbel copula, $\theta = 2$	0.019	0.063	0.110
6. Gumbel copula, $\theta = 3$	0.011	0.054	0.102

Power study of the proposed test

- Empirical power (times 100) of the bootstrap test in the Monte Carlo study.
- The DGP in the first regime is a Gaussian copula with $\rho = 0.5$ and with Gaussian marginals.
- Each table entry is based on 1000 replications, each with 400 observations.

	Nominal level (α)		
(r)1-4 Model under H_1	0.01	0.05	0.1
1. Clayton copula, $\theta = 2$	31.9	68	82.2
2. Clayton copula, $\theta = 3$	82.1	97.8	99.7
3. Gaussian copula, $\rho = -0.5$	100	100	100
4. Gaussian copula, $\rho = 0.8$	73.6	93.5	96.8
5. Gumbel copula, $\theta = 2$	24.1	53.7	68.9
6. Gumbel copula, $\theta = 3$	96.3	99.4	99.9

Study of power with HMM classification

- Gaussian bivariate HMM model to classify observations into two regimes.
- Is the power still acceptable when we use a HMM model to determine the regimes?

	Predicted regime 1	Predicted regime 2
True regime 1	68.5%	6.8%
True regime 2	14.2%	10.5%

Figure: Confusion matrix of aggregated predictions of all the 500 HMM classifiers

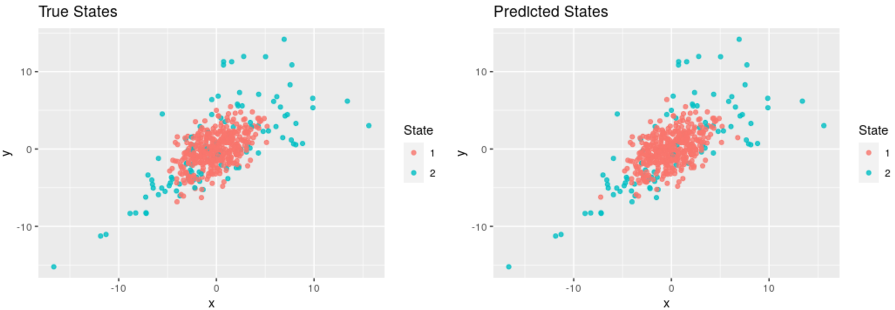


Figure: True vs. predicted regimes for one realization

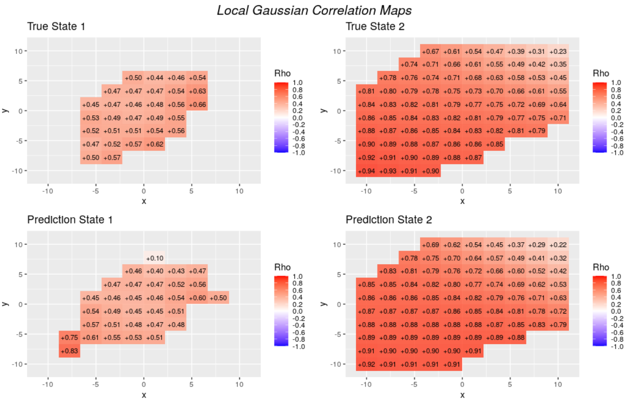


Figure: LGC-map of true vs. predicted states

Resulting power for two models (500 realizations):

	Nominal level (α)					
		True			Prediction	
Model under H_1	0.01	0.05	0.1	0.01	0.05	0.1
1. Gaussian copula, $\rho = -0.5$	100	100	100	100	100	100
2. Clayton copula, $\theta = 3$	96.2	99.4	99.8	64.6	81.8	87.2

Empirical analysis S&P500 vs. FTSE100

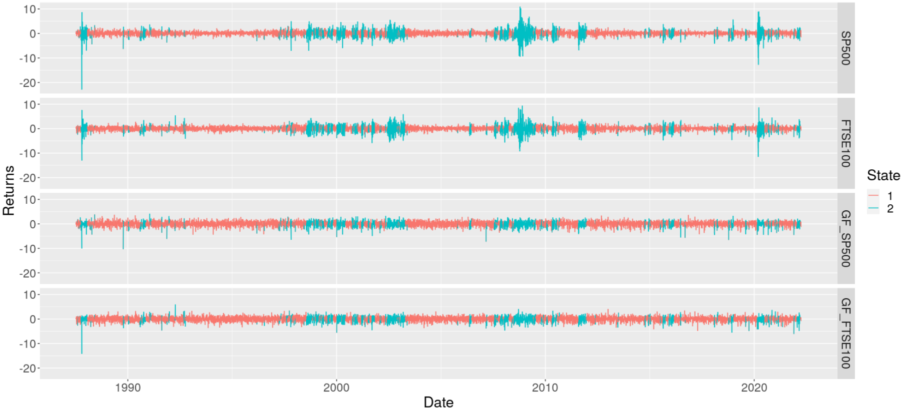


Figure: Daily log-returns and GARCH filtrated log-returns of S&P500 and FTSE100 with classification in two regimes with a Gaussian bivariat hidden Markov model.

Empirical analysis S&P500 vs. FTSE100

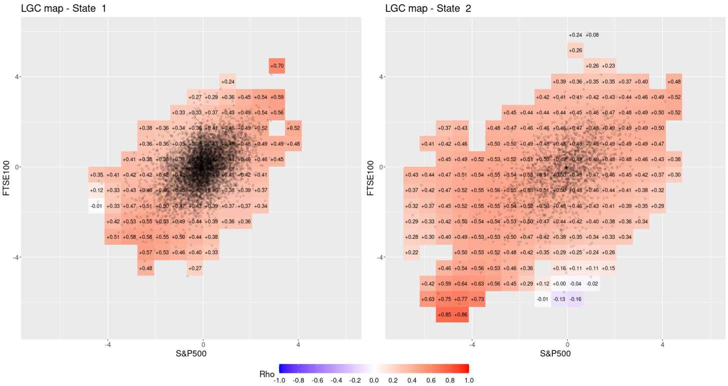


Figure: LGC regime 1 (left plot) and regime 2 (right plot)

- We performed the test with 1000 bootstrap replications $\rightarrow$ p-value $= 0.009$ (null of equal dependence rejected)

Concluding remarks

- Proposed a new approach (RS-LGC + test) for studying time-varying nonlinear dependence
- Simulation study shows that test behaves well, and indicates validity
- Several empirical examples reveals nonlinear and time-varying relationships between financial returns (confirm previous studies)
- Advantages:
  - intuitive
  - test - can be extended to many regimes (and in principle to higher dimensions)
  - can be used to help select a parametric dependence model (say, copula)
- Disadvantage:
  - Two-step procedure. A simultaneous estimation of the HMM and the LGC would be better, but currently not feasible due to the semi-parametric nature of the LGC

References

Tjøstheim & Hufthammer (2013). Local Gaussian correlation: A new measure of dependence. *Journal of Econometrics*
Bacri, Berentsen, Bulla & Hølleland (2022). A gentle tutorial on accelerated parameter and confidence interval estimation for hidden Markov models using Template Model Builder. *Biometrical Journal*.
Bacri, Berentsen, Bulla & Støve (2023). Computational issues in parameter estimation for hidden Markov models with Template Model Builder. *Journal of Statistical Computation and Simulation.*
Berentsen & Tjøstheim (2014). Recognizing and visualizing departures from independence in bivariate data using local Gaussian correlation. *Statistics & Computing*.
Berentsen, Støve, Tjøstheim & Nordbø (2014). Recognizing and visualizing copulas: an approach using local Gaussian approximation. *Insurance: Mathematics & Economics*.
Støve & Tjøstheim (2014). Asymmetric dependence patterns in financial returns: An empirical investigation using local Gaussian correlation. *Nonlinear Time Series Econometrics*, OUP.
Støve, Tjøstheim & Hufthammer (2014). Using local Gaussian correlation in a nonlinear re-examination of financial contagion. *Journal of Empirical Finance*.
Lacal & Tjøstheim (2017). Local Gaussian autocorrelation and tests for serial independence. *Journal of Time Series Analysis*
Lacal & Tjøstheim (2019). Estimating and testing nonlinear local dependence between two time series. *Journal of Business & Economic Statistics*.
Otneim & Tjøstheim (2017). The locally gaussian density estimator for multivariate data. *Statistics & Computing*.
Otneim & Tjøstheim (2018). Conditional density estimation using the local Gaussian correlation *Statistics & Computing*.
Tjøstheim , Otneim & Støve (2022). Statistical dependence: Beyond Pearson's ρ. *Statistical Science*.
Hjort & Jones (1996). Locally parametric nonparametric density estimation. *The Annals of Statistics*.
BenSaïda, Boubaker, & Nguyen (2018). The shifting dependence dynamics between the G7 stock markets. *Quantitative Finance*.
Holland & Wang (1987). Dependence function for continous bivariate densities. *Communications in Statistics*
Okimoto (2008). New evidence of asymmetric dependence structures in international equity markets. *J. Financ. Quant. Anal.*
Chollete, Heinen & Valdesogo (2009). Modeling international financial returns with a multivariate regime switching copula. *Journal of Financial Econometrics*.
Sleire, Støve, Berentsen, Otneim, Tjøstheim, Haugen (2022). Portfolio allocation under asymmetric dependence in asset returns using local Gaussian correlations. *Finance Research Letters*.
Jordanger & Tjøstheim (2022). Nonlinear Spectral Analysis: A Local Gaussian Approach. *JASA*.
Zucchini, MacDonald & Langrock (2016). Hidden Markov Models for time series: An introduction using R, second edition. *CRC Press.*

Local Gaussian approximation II

Tjøstheim & Hufthammer (2013) show that once a unique population vector $\gamma_b(x)$ exists, under weak regularity conditions one can let $b \to 0$ to obtain a local population vector $\gamma(x)$ defined at a point x. The population vectors $\gamma_b(x)$ and $\gamma(x)$ are both consistent with a local log-likelihood function defined by

$$L\big(R_1, \ldots, R_T, \gamma_b(x)\big) = T^{-1} \sum_i K_b(R_i - x) \log \psi(R_i, \gamma_b(x)) - \int K_b(v - x) \psi(v, \gamma_b(x)) v, \quad (2)$$

for given pairwise observations $R_1, \ldots, R_T$ iid or an ergodic time series. The numerical maximization of the local likelihood (2) leads to local likelihood estimates $\gamma_{T,b}(x)$, including estimates $\rho_{T,b}(x)$ of the local correlation. **Choice of bandwidths**!!

Example: bivariate t-distribution

Simulated observations from a bivariate t-distribution with 4 degrees of freedom and global correlation equal to 0, T=784

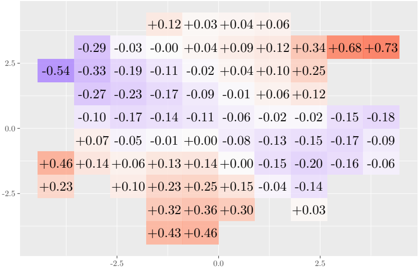


Figure: The estimated LGC map

Note: the bivariate t-distribution with correlation equal to 0, is in fact *not* independent

Hidden Markov Models: local decoding

- $\boldsymbol{\theta}$: distribution and Markov chain parameters
- C_t: Markov chain regime at time t

Smoothing probabilities: the probability of being in regime k at time t for $k = 1, ..., m, t = 1, ..., T$ given all observations, i.e.

$$\forall k = 1, \ldots, m, p_{kt}(\boldsymbol{\theta}) = \mathrm{P}_{\boldsymbol{\theta}}(C_t = k | \boldsymbol{R}^{(T)} = \boldsymbol{r}^{(T)})$$

The derived smoothing probabilities then serve for determining the **most probable regime at time** t:

$$C_t = \arg\max_k p_{kt}(\boldsymbol{\theta})$$

cfr. Zucchini et al (2016).