DEPENDENCE STRUCTURES FOR MULTIVARIATE HIGH–FREQUENCY DATA IN FINANCE

Wolfgang Breymann

RiskLab, Department of Mathematics, ETH Zurich, Switzerland

Alexandra Dias and Paul Embrechts

www.math.ethz.ch/~embrechts

Department of Mathematics, ETH Zurich, Switzerland

ACKNOWLEDGEMENT

We thank Olsen Data for having provided us with the high-frequency data used in this study

OVERVIEW

- Motivation
- The data
- Deseasonalisation
- Dependence structure modelling
  for the whole dataset
- Tail dependence analysis
- Conclusion

For TECHNICAL DETAILS we refer to our PAPER!

MOTIVATION

- The Goal:
  Studying the dependence structure across time scales

- Why?
  - The change of the behavior as a function of the time horizon may contain important information
  - Improves extrapolation from small to large time horizons

- Requires:
  Characterising dependence for horizons from minutes to months

- Here:
  Restriction to high-frequency region (1 hour – 1 day)

- Pecularities of high-frequency data are taken into account

THE DATA

- Tick–by–tick bid and ask quotes
- Period: Febr 1986–Dec 1998
- Collected and filtered by Olsen Data
- Irregularly spaced
- About 10 million data points for a single series
- Regularisation to 5 min. time series by linear interpolation
- Reduction to logarithmic middle prices:

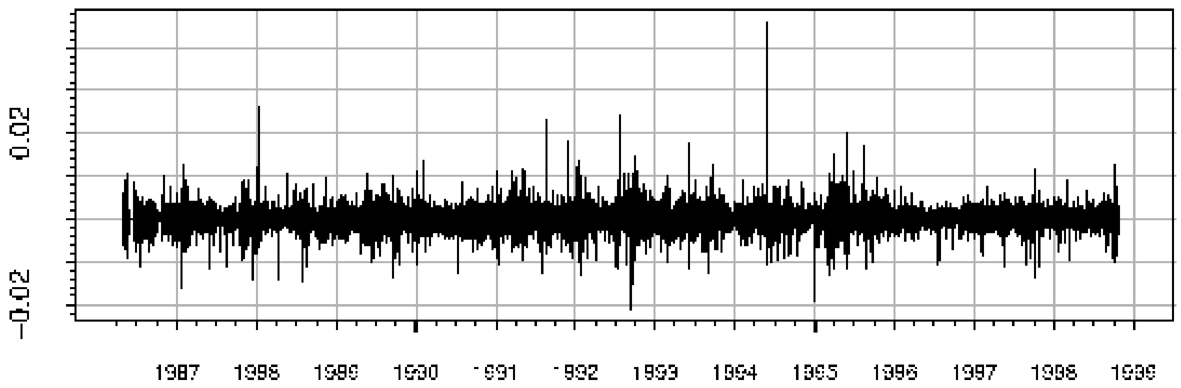
$$\xi_{\alpha,t} = \frac{\log\left(p_{\alpha,t}^{Bid} \cdot p_{\alpha,t}^{Ask}\right)}{2}$$

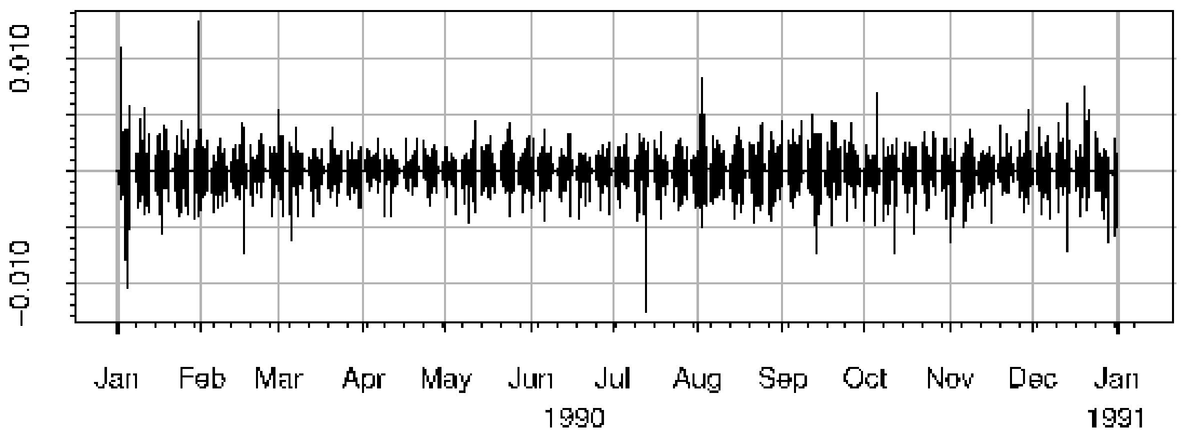
FX PRICES FOR USD/DEM AND USD/JPY

DESEASONALISATION OF FINANCIAL DATA

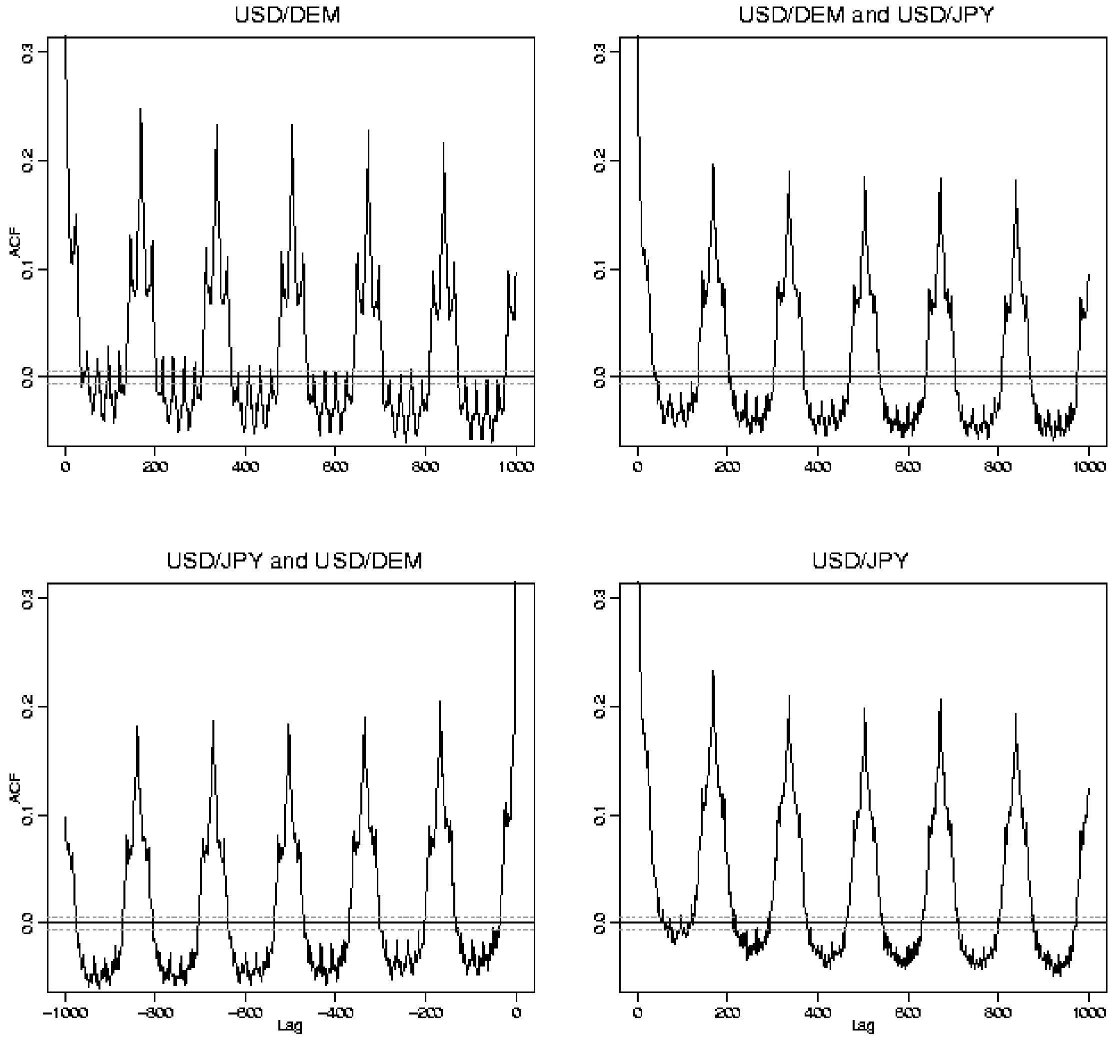
- High frequency financial data present strong seasonalities
- Main periodicities: daily and weekly
- Seasonalities cover more subtle statistical properties
- Affected by Daylight Saving Time (DST)
- Theory of stochastic processes favors time transformation to an activity-based time scale, but:
  $\rightarrow$ Loss of synchronicity in the multivariate case
- Instead:
  $\rightarrow$ Volatility weighting based on weekly activity pattern
- Drawback: $\rightarrow$ Aggregation property of returns has to be replaced by a more complicated relationship

HOURLY RETURNS

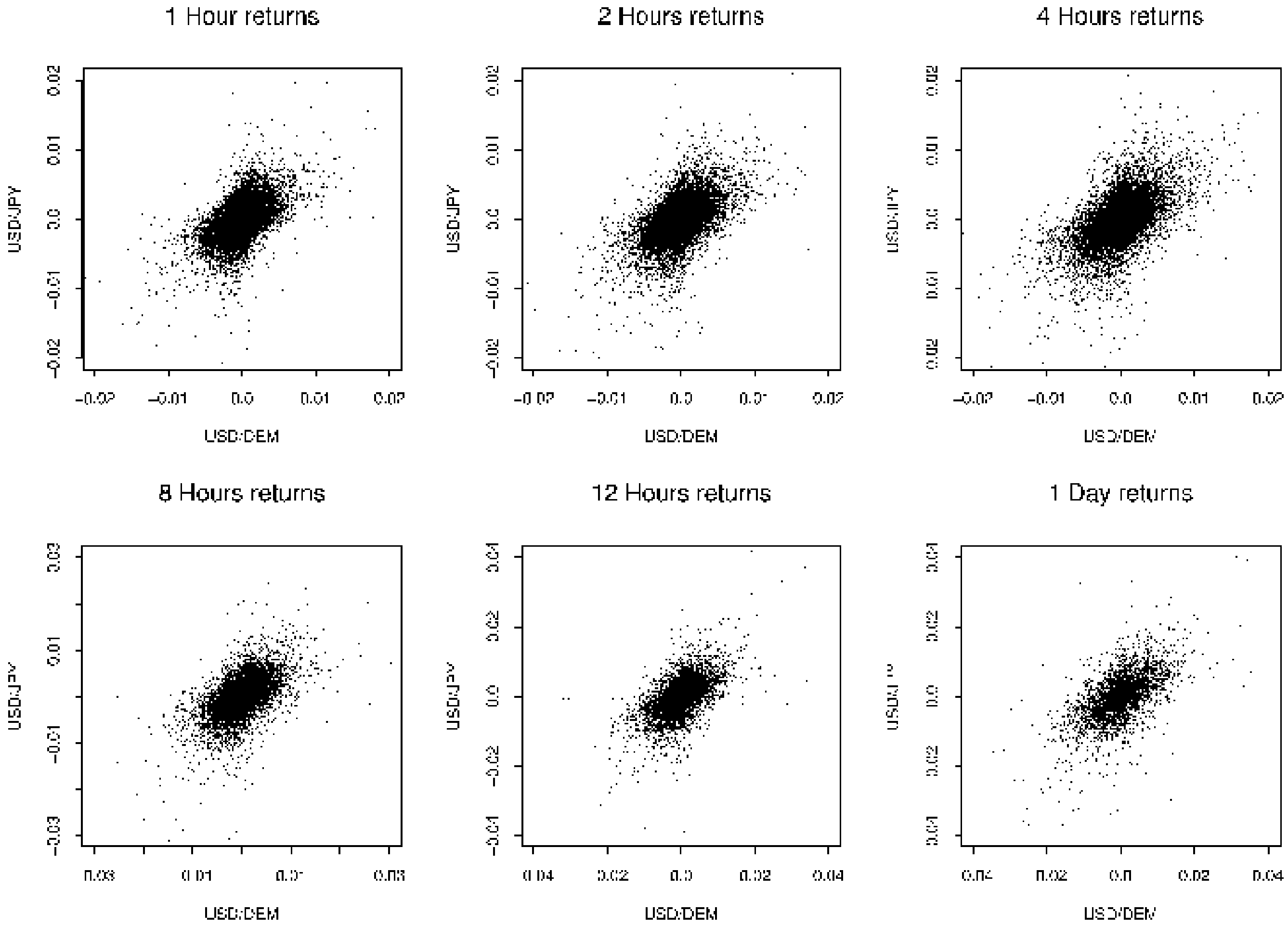




AUTOCORRELATION FUNCTIONS OF ABSOLUTE RETURNS



RETURN SCATTER PLOTS



MODELLING REQUIREMENTS

- The flexibility of modelling arbitrary patterns that display abrupt volatility changes ($\rightarrow$ Japanese lunch break)

- Taking into account slow temporal changes in the habits of the market participants, institutional changes, etc

- Keeping track of Daylight Saving Time (DST) to take into account the one hour displacement between DST and non–DST periods

- The modelling of the geographical decomposition of market activity to take into account local public holidays and other irregularities

THE VOLATILITY PATTERN

- Integrated squared volatility: $V_t^2 = \sum_{t' \leq t} (v_{t'}[\delta])^2$

- Volatility wrt horizon ΔT:

  $\Delta V_t^2[\Delta T] \equiv V_t^2 - V_{t-\Delta T}^2 = \sum_{i=0}^{n-1} (v_{t-i\delta}[\delta])^2 = (v_t[\Delta T])^2$

- Deseasonalised returns:

$$x_t[\Delta T] = \frac{\xi_t - \xi_{t-\Delta T}}{\sqrt{\Delta V_t^2[\Delta T]}}.$$

- $\delta = 5$ minutes: elementary time step; $n = \Delta T/\delta$

- Aggregation property:

$$x_t[\Delta T] = \frac{x_{t-\Delta T_2}[\Delta T_1]\ \sqrt{\Delta V_{t-\Delta T_2}^2[\Delta T_1]} + x_t[\Delta T_2]\ \sqrt{\Delta V_t^2[\Delta T_2]}}{\sqrt{\Delta V_t^2[\Delta T]}}$$

COMPUTING THE VOLATILITY PATTERN

- Decomposition of the volatility:

$$v_t^2[\delta] = a_t \left(v_\tau^{(d)}[\delta] \right)^2$$

  - with relative market activity factor a_t and
  - volatility averaged over DST period d conditional to the time in the week, $\tau = t \mod (1 \text{ week})$:

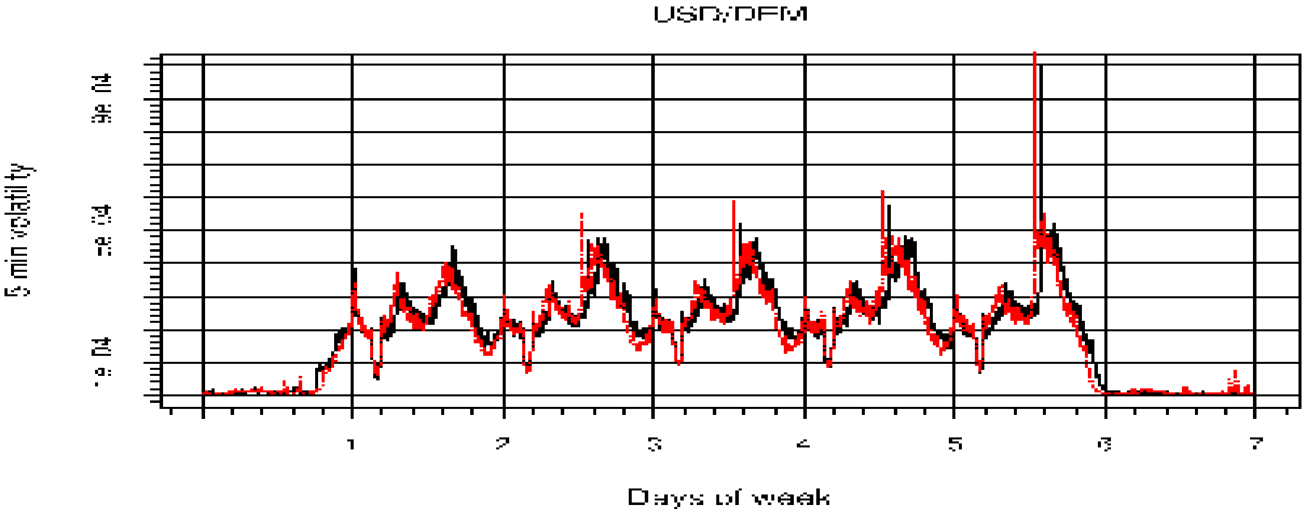
$$\left(v_\tau^{(d)}[\delta] \right)^2 = \frac{1}{N_d} \sum_{i=1}^{N_d} (r_{t_i+\tau}[\delta])^2$$

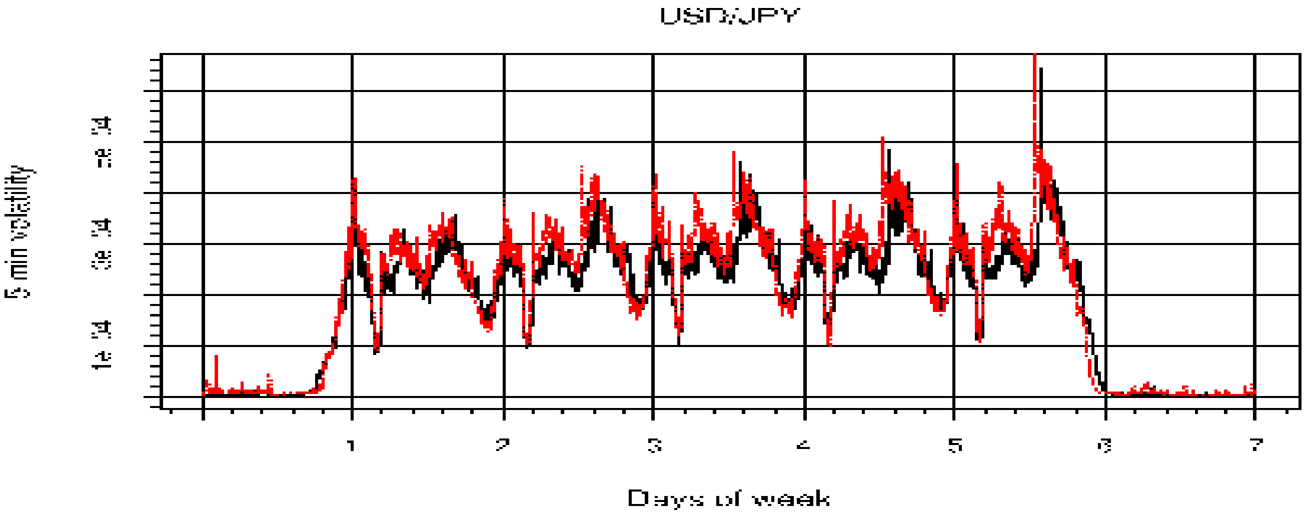
- Weekend volatility:

$$v^{(w)}[\delta] = \left| r_{t_w^{(end)}}[\Delta T_w] \right| \sqrt{\frac{\delta}{\Delta T_w}},$$

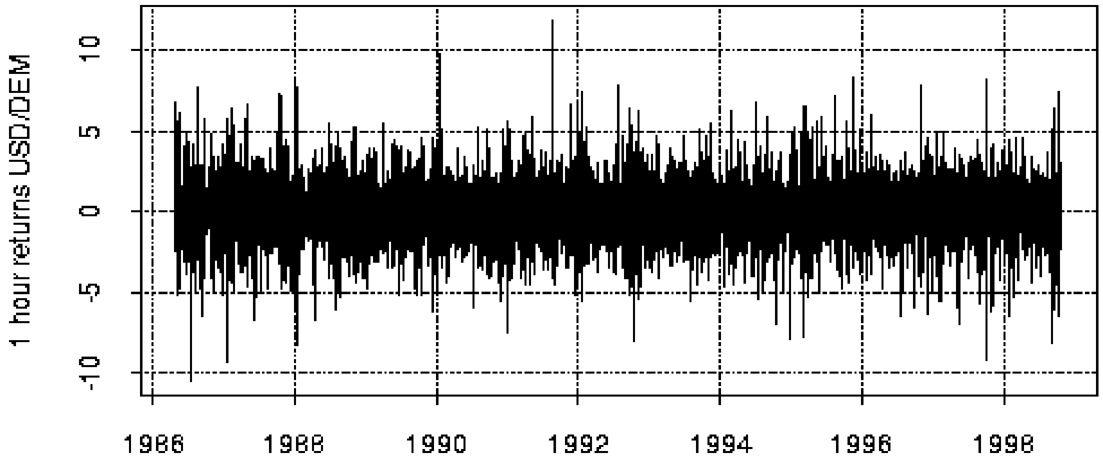
  - with weekend length $\Delta T_w = t_w^{(end)} - t_w^{(start)}$

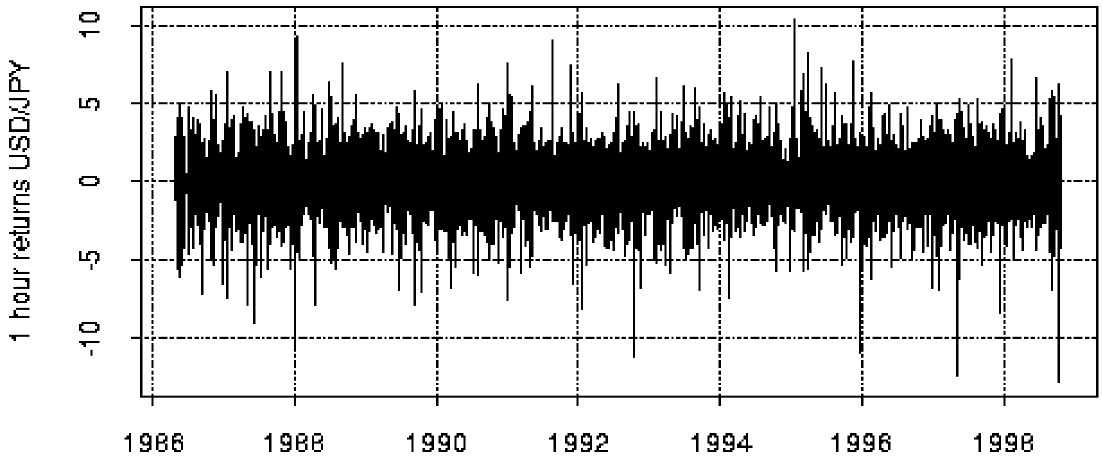
THE WEEKLY VOLATILITY PATTERN



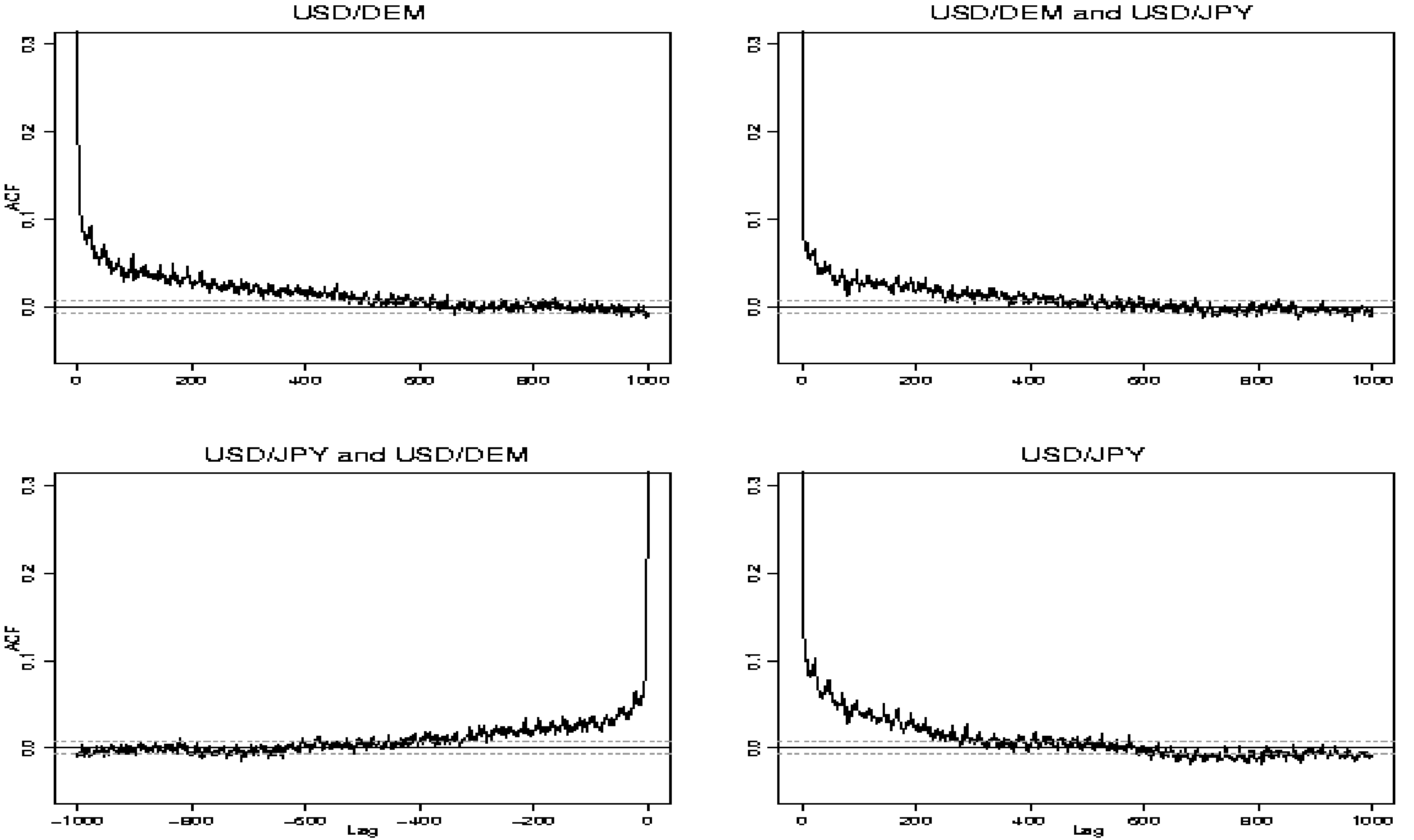


DESEASONALISED RETURNS

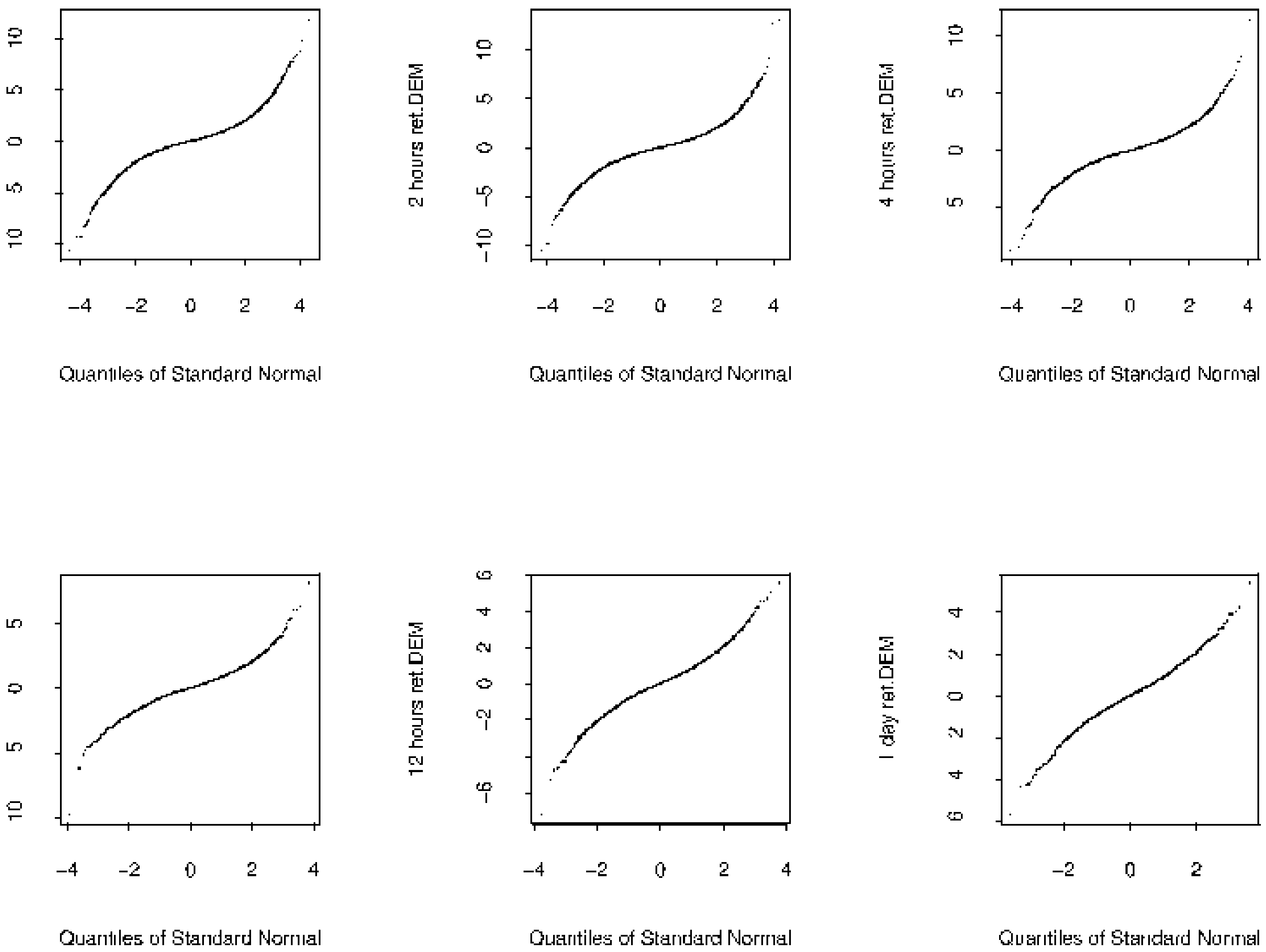




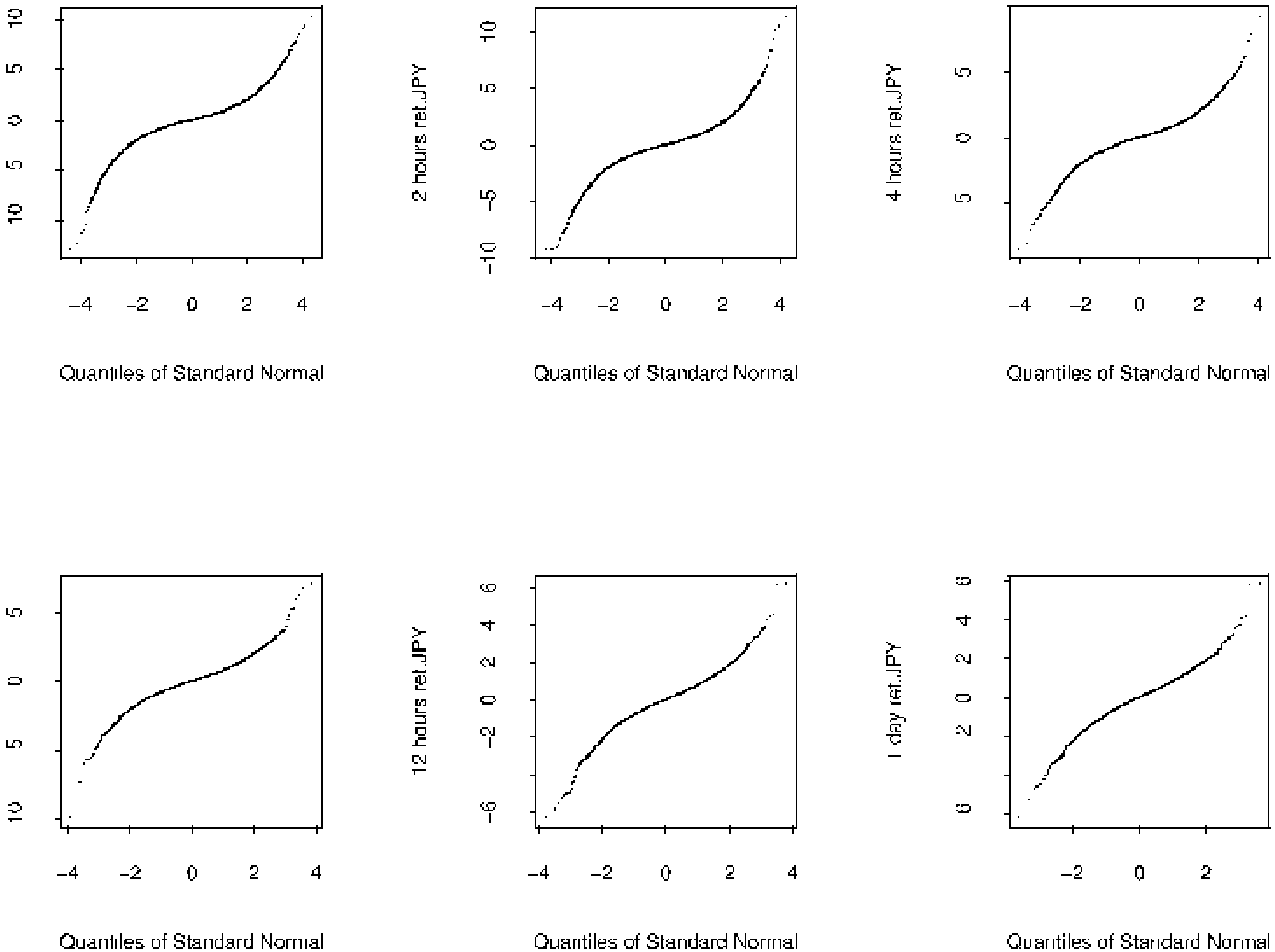
ACF OF DESEASONALISED HOURLY ABSOLUTE RETURNS



QQ-PLOTS FOR USD/DEM



QQ-PLOTS FOR USD/JPY



DEPENDENCE STRUCTURE MODELLING FOR THE WHOLE DATASET

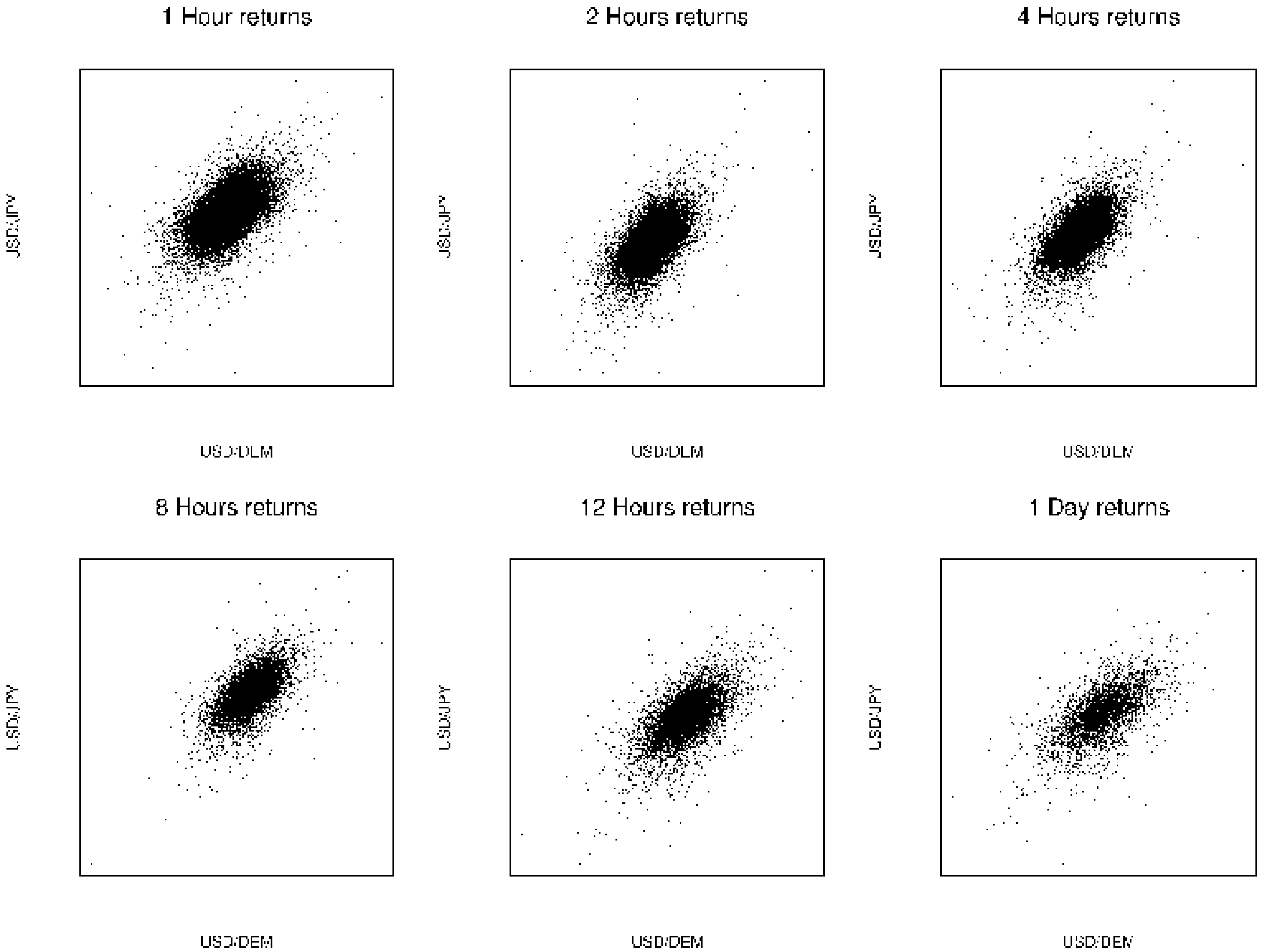
- Exploratory analysis (scatter plots)

- Families of copulas across time scales

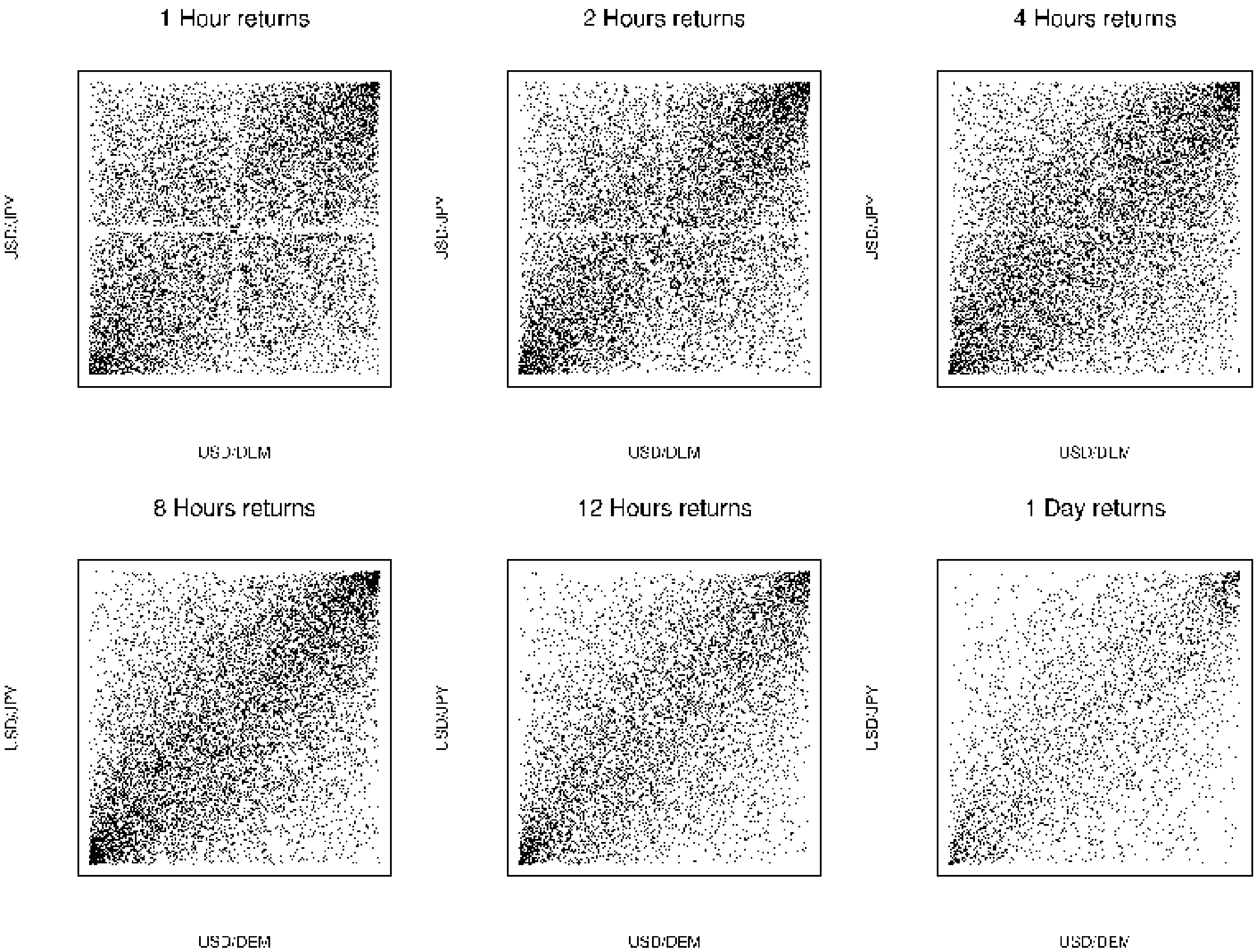
- Tail coefficient estimates

- Goodness of fit and ellipticality test

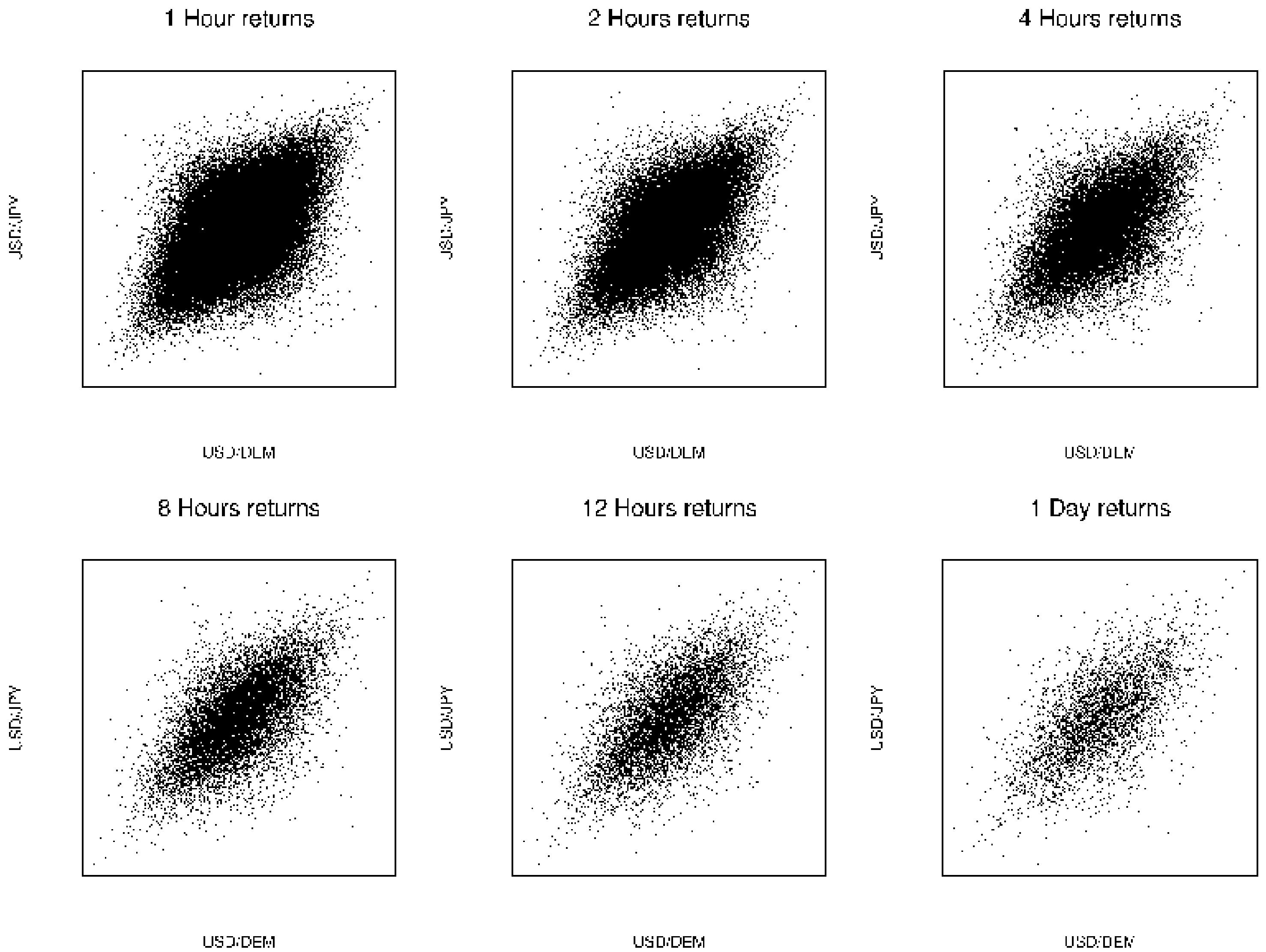
SCATTER PLOTS OF DESEASONALISED RETURNS



COPULA DENSITY OF DESEASONALISED RETURNS



PSEUDO OBSERVATIONS WITH NORMAL MARGINS



FAMILIES OF COPULAS

- Gaussian copula for correlation ρ:

$$C_\rho^{Ga}(u,v) = \int_\infty^{\Phi^{-1}(u)} \int_\infty^{\Phi^{-1}(v)} \frac{1}{2\pi\sqrt{(1-\rho^2)}} \exp\left\{-\frac{(s^2 - 2\,\rho s\,t + t^2)}{2(1-\rho^2)}\right\} \mathrm{d}s\,\mathrm{d}t$$

- t–copula for ν degrees of freedom and correlation ρ:

$$C_{\nu,\rho}^{t}(u,v) = \int_\infty^{t_\nu^{-1}(u)} \int_\infty^{t_\nu^{-1}(v)} \frac{1}{2\pi\sqrt{(1-\rho^2)}} \left\{1 + \frac{(s^2 - 2\,\rho s\,t + t^2)}{\nu(1-\rho^2)}\right\}^{-\frac{(\nu+1)}{2}} \mathrm{d}s\,\mathrm{d}t$$

FAMILIES OF COPULAS (CONT.)

- Gumbel copula:

$$C_\beta^{Gu}(u,v) = \exp\left[-\left\{(-\log u)^{1/\beta} + (-\log v)^{1/\beta}\right\}^\beta\right]$$

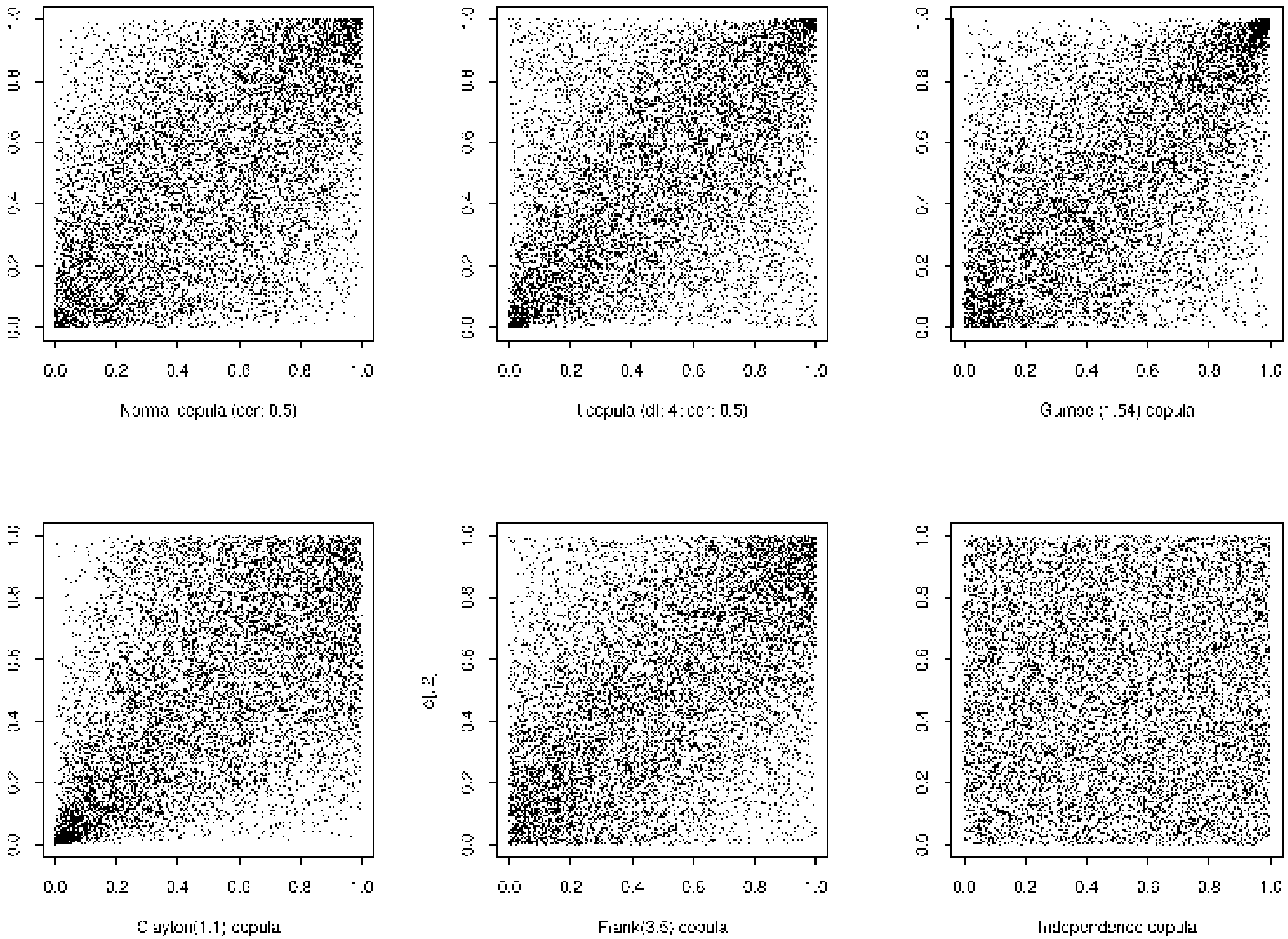
- Clayton copula:

$$C_\beta^{Cl}(u,v) \;=\; \max\left[-\left\{(-\log u)^{1/\beta} + (-\log v)^{1/\beta}\right\}^\beta, 0\right]$$

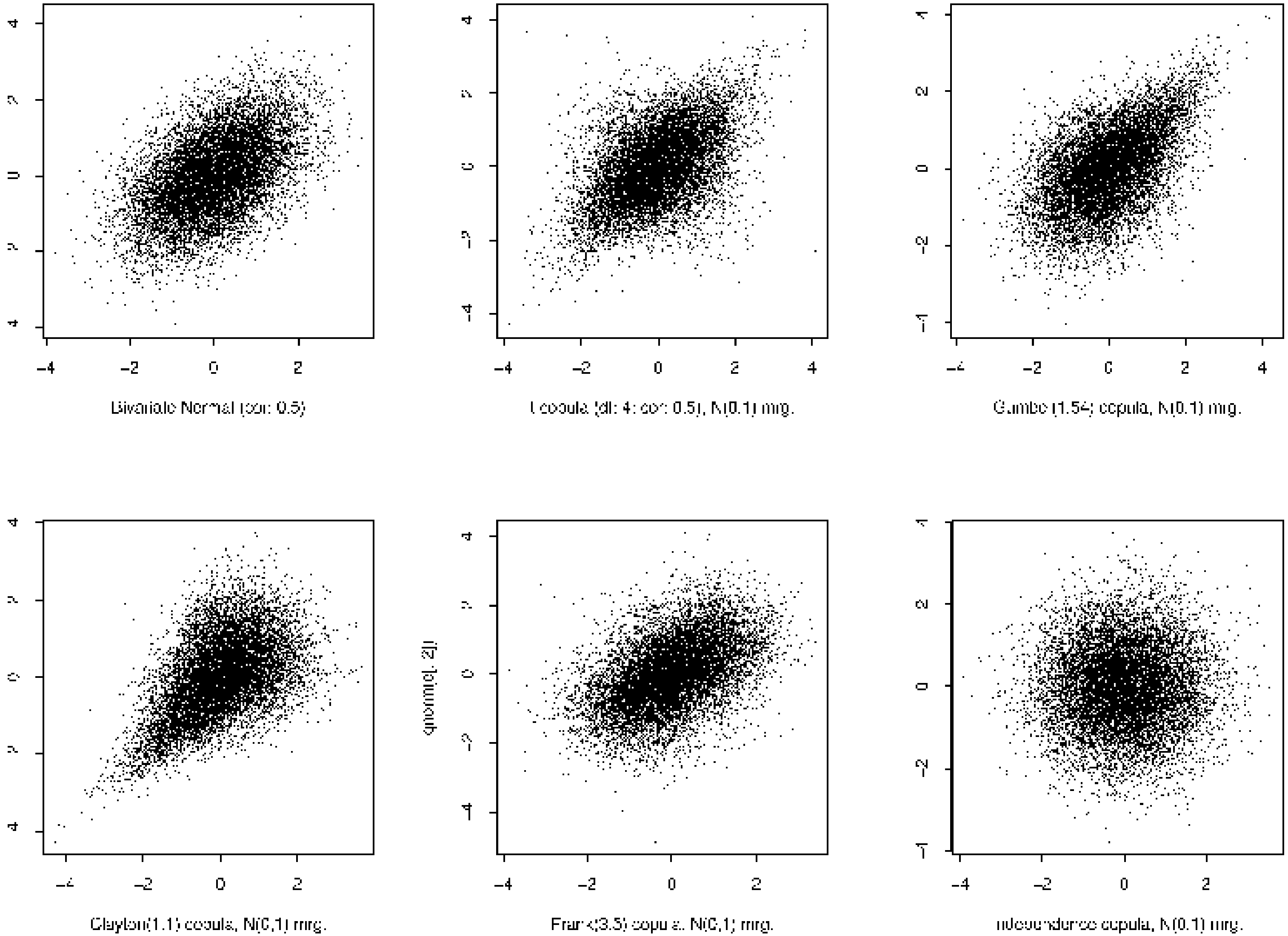
- Frank copula:

$$C_\beta^{Fr}(u,v) = -\frac{1}{\beta}\log\left[1 + \frac{(e^{-\beta u}-1)(e^{-\beta v}-1)}{e^{-\beta}-1}\right]$$

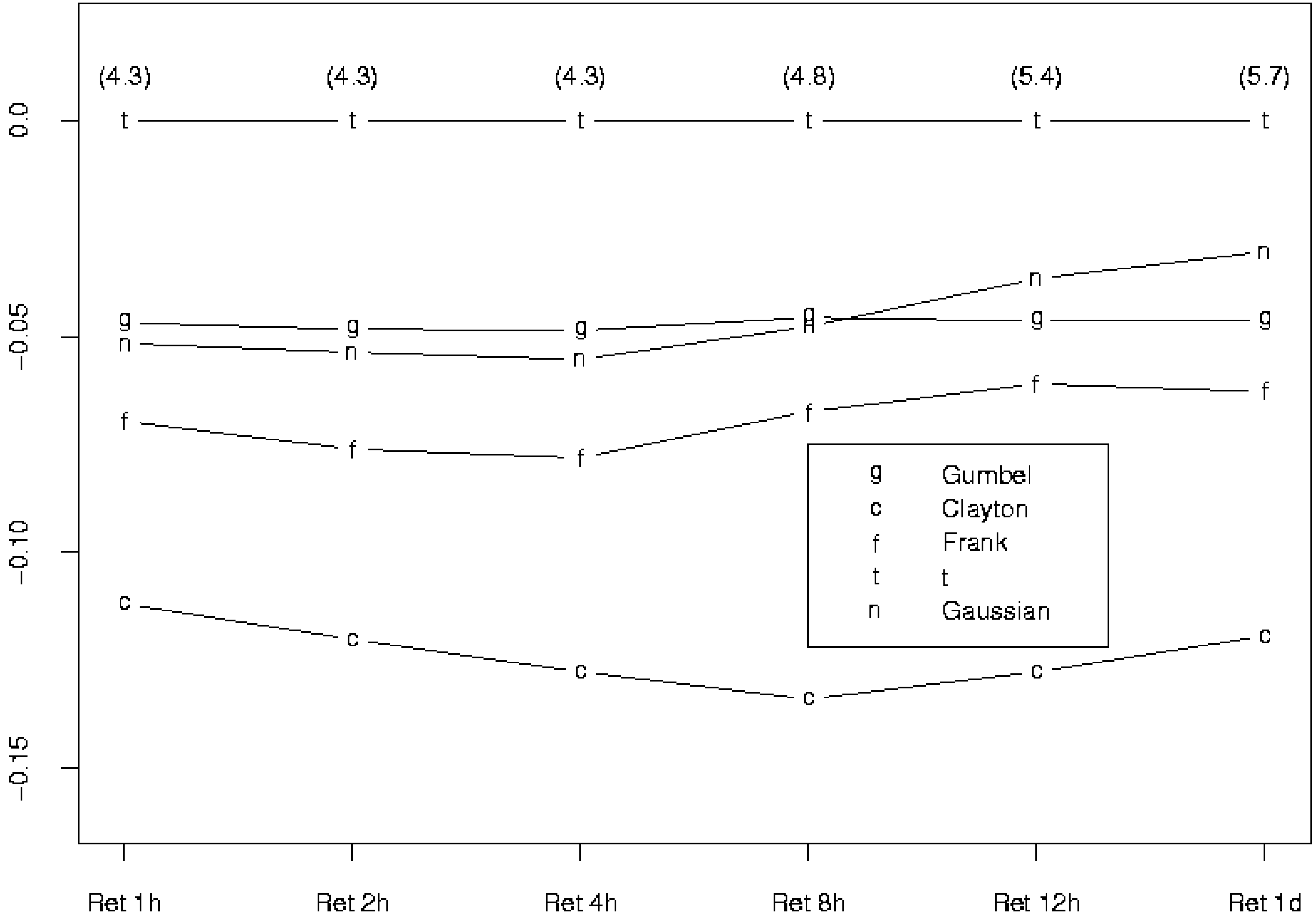
COPULA DENSITIES FOR SELECTED COPULAS



COPULA DENSITIES WITH NORMAL MARGINS



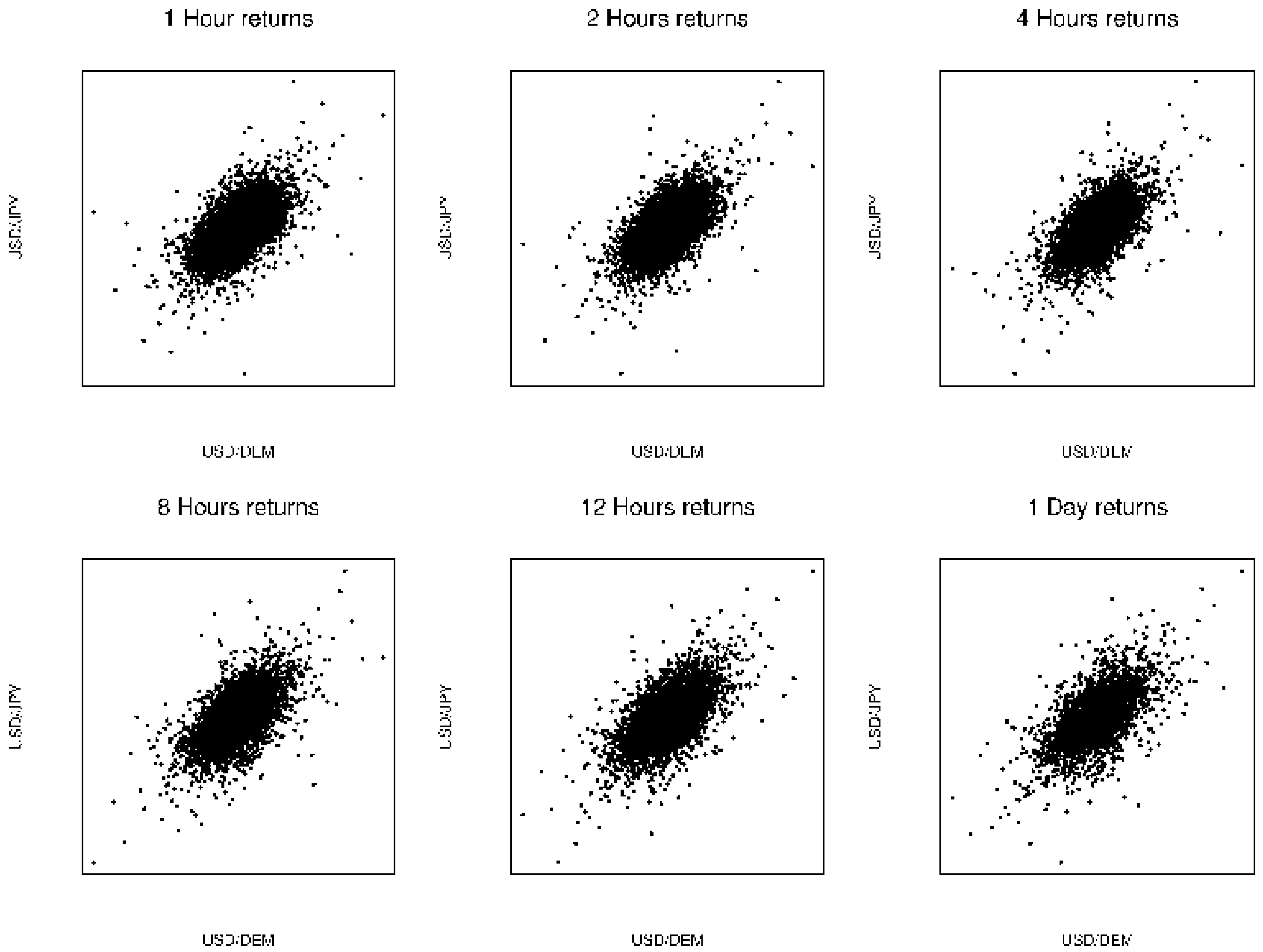
GOODNESS OF FIT FOR DIFFERENT FREQUENCIES



GOODNESS OF FIT AND ELLIPTICALITY TEST

Time Freq.	Prob. Integral test		P–values for ellipticality test	
	Model	p–value	original margins	t margins
1 hour	t	0	0	0
2 hours	t	0	0	0
4 hours	t	0.01	0	0.092
8 hours	t	0.27	0	0.231
12 hours	t	0.19	0.034	0.369
1 day	t	0.74	0.821	0.675

PSEUDO OBSERVATIONS WITH t MARGINS



TAIL DEPENDENCE ANALYSIS

- Tail coefficient estimates
- Spectral measure estimation
- Multivariate excesses modelled by copulas

TAIL COEFFICIENT ESTIMATES

Frequency	d.f. $\widehat{\nu}$	correl. $\widehat{\rho}$	tail dep. coeff. $\widehat{\lambda}$
1 hour	4.339	0.563	0.273
2 hours	4.269	0.585	0.291
4 hours	4.282	0.599	0.299
8 hours	4.833	0.619	0.287
12 hours	5.438	0.623	0.264
1 day	5.712	0.624	0.254

Tail coefficient estimates for the DEM and JPY bivariate returns for the different time frequencies considered

$$\lambda = \lim_{\alpha \to 1^-} P(X_2 > F_2^{-1}(\alpha) | X_1 > F_1^{-1}(\alpha))$$

SPECTRAL MEASURE ESTIMATION

- Suppose that the d-dimensional random vector $\mathbf{X}$ has a regularly varying tail with tail index α

- Limit behaviour of $\mathbf{X}$ (vague convergence):

$$\frac{P(\| \mathbf{X} \| > tx, \mathbf{X} / \| \mathbf{X} \| \in \cdot)}{P(\| \mathbf{X} \| > t)} \quad \xrightarrow{v} \quad x^{-\alpha} P(\Theta \in \cdot),$$

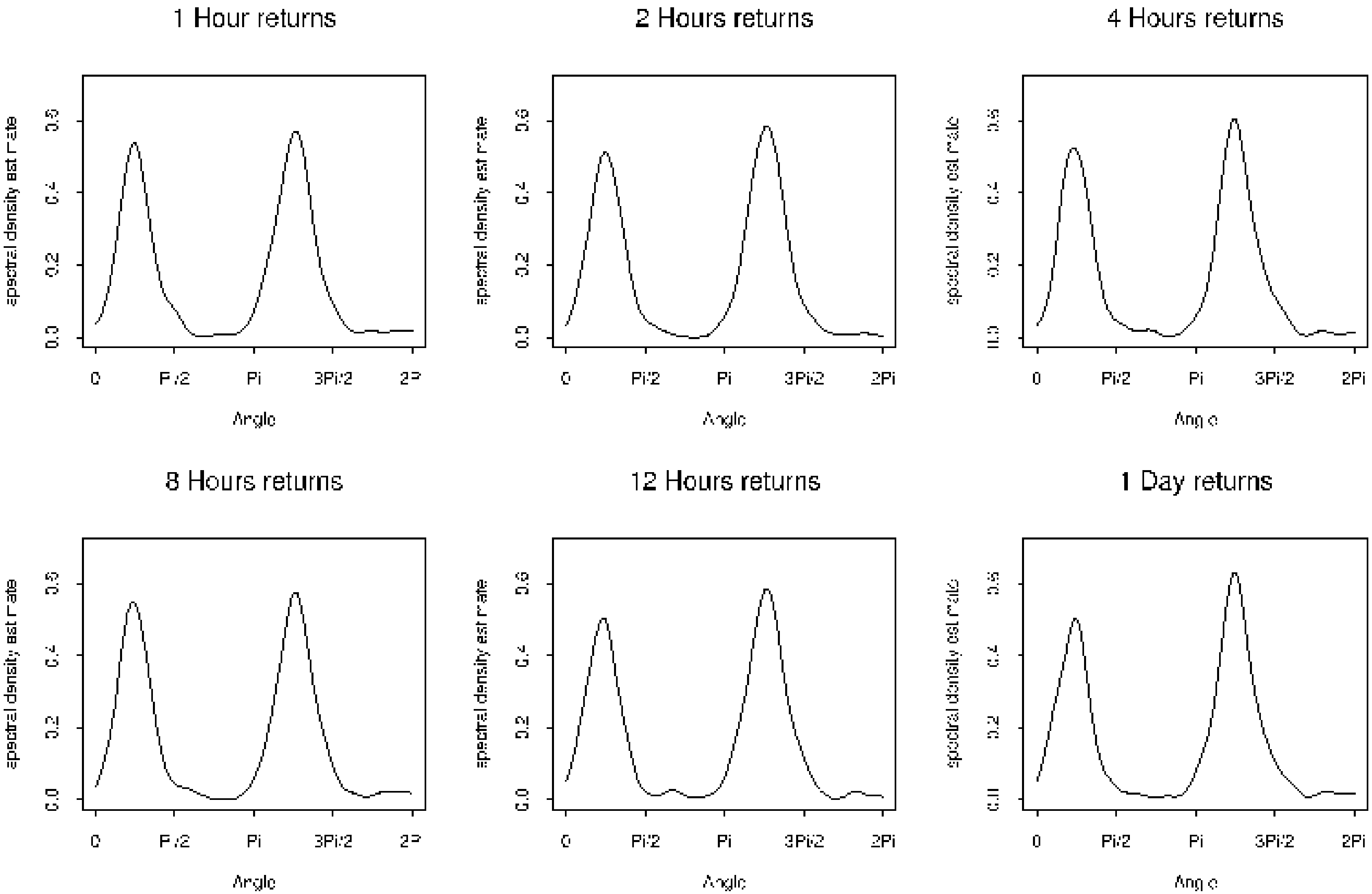
  with $x > 0$, $t \to \infty$, and Θ random vector on the space $(\mathbb{S}^{d-1}, \mathcal{B}(\mathbb{S}^{d-1}))$

- Distribution function of Θ is SPECTRAL DISTRIBUTION of $\mathbf{X}$.

- Alternatively:
  $\exists$ measure ν and positive sequence (a_n), $a_n \to \infty$, such that

$$nP(a_n^{-1}\mathbf{X} \in \cdot) \quad \xrightarrow{v} \quad \nu(\cdot) \quad \text{for } n \to \infty$$

SPECTRAL MEASURE FOR DIFFERENT HORIZONS



MULTIVARIATE EXCESSES MODELLED BY COPULAS

- Extreme tail dependence copula relative to a threshold t:

$$C_t(u,v) = P(U \leq F_t^{-1}(u), V \leq F_t^{-1}(v) | U \leq t, V \leq t)$$

with conditional distribution function

$$F_t(u) := P(U \leq u | U \leq t, V \leq t),\ 0 \leq u \leq 1$$

- Archimedean copulas: $\exists$ cont., strictly decreasing function $\psi : [0,1] \mapsto [0,\infty]$ with $\psi(1) = 0$, s.t.

$$C(u,v) = \psi^{[-1]}(\psi(u) + \psi(v))$$

- For "sufficiently regular" Archimedean copulas (Juri and Wüthrich (2002)):

$$\lim_{t \to 0^+} C_t(u,v) = C_\alpha^{Clayton}(u,v)$$

MULTIVARIATE EXCESSES (CONT:)

- Data:
  1 hour pseudo–returns of DEM and JPY, $(\widehat{F}_{1n}(x_{1i}), \widehat{F}_{2n}(x_{2i}))$

- For several thresholds t for hourly pseudo-returns we modelled

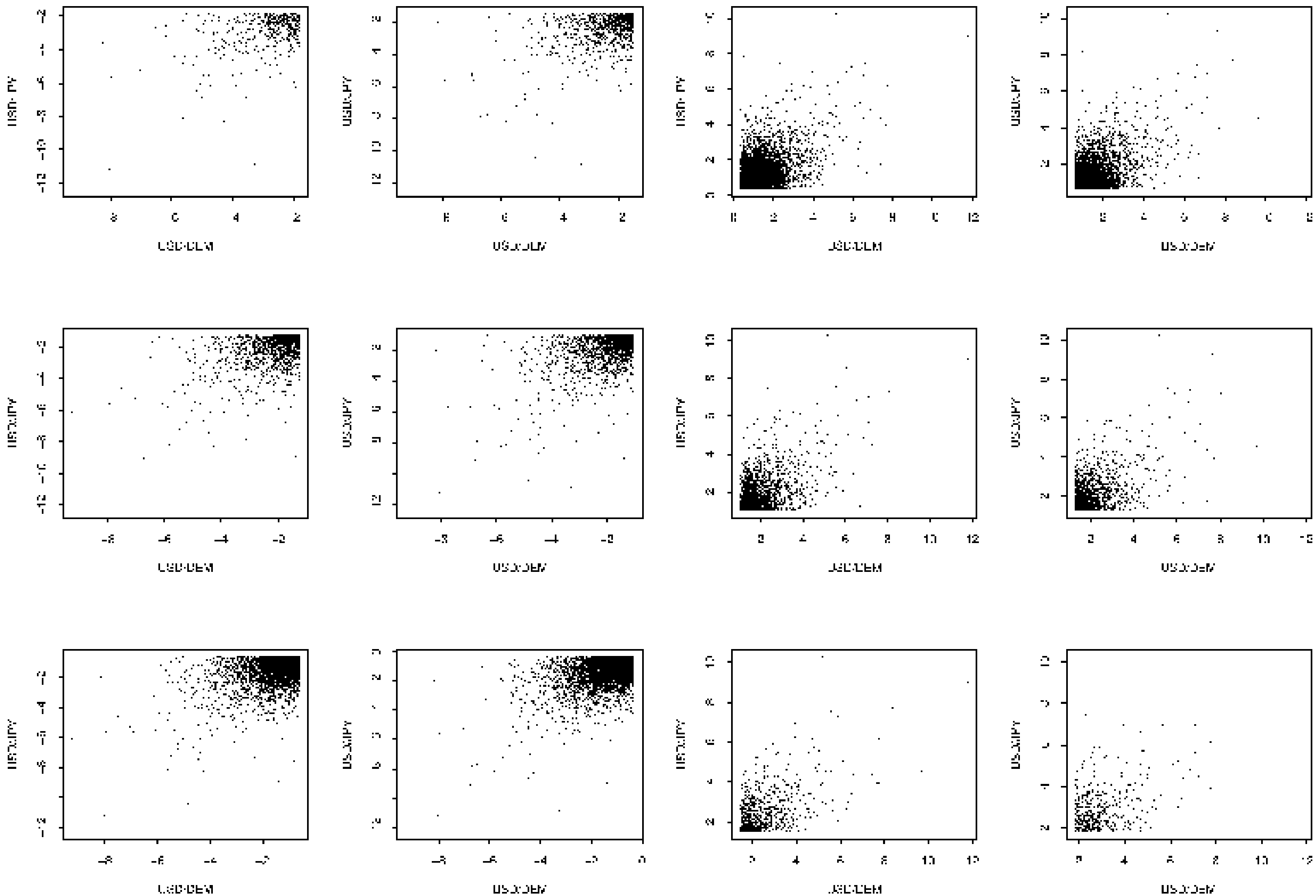
$$C_{t^-}(u,v) = P(U \leq F_t^{-1}(u), V \leq F_t^{-1}(v) | U \leq t, V \leq t)$$

  and

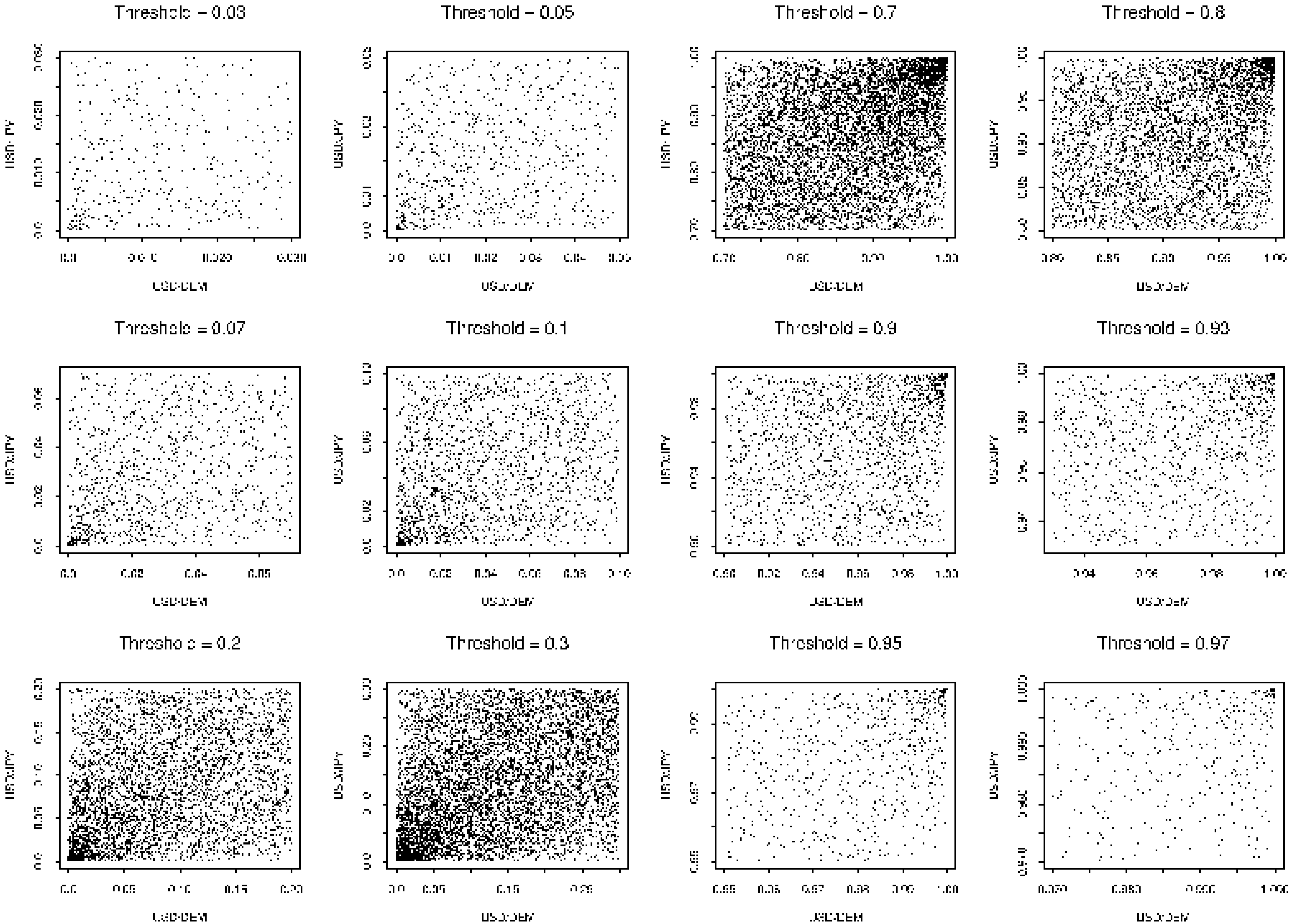
$$C_{t^+}(u,v) = P(U \leq F_t^{-1}(u), V \leq F_t^{-1}(v) | U \geq t, V \geq t)$$

- Thresholds:
  0.03, 0.05, 0.07, 0.1, 0.2, 0.3, 0.7, 0.8, 0.9, 0.93, 0.95, 0.97

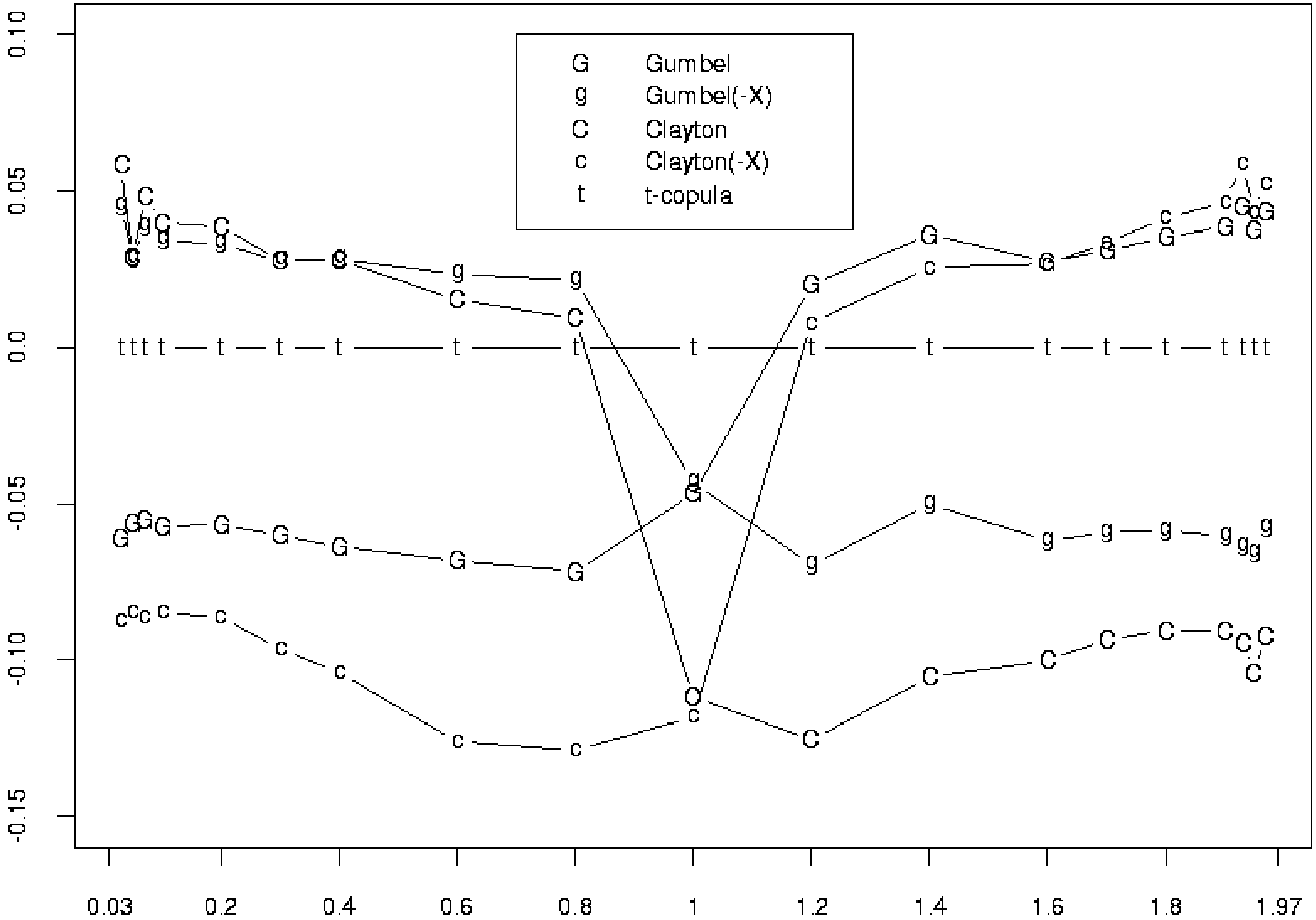
BIVARIATE EXCESSES FOR DIFFERENT THRESHOLDS



COPULA DENSITY OF BIVARIATE EXCESSES



AIC VALUES FOR DIFFERENT THRESHOLDS



CONCLUSION

- 2-dim. high-frequency data have been deseasonalised by means of volatility weighting
- The dependence structure for 2-dim., high-frequency FX return data has been analysed
- Methods used:
  - Copula modelling
  - Statistical techniques for extremal clustering

CONCLUSION: THE OVERALL PICTURE

- t–copulas with successively higher degrees of freedom work best for the whole dataset

- However, t–copulas have not enough structure for the shortest time horizons

- Test for ellipticality only rejected for 1 hour and 2 hours returns if the margins are transformed to t-distributions with the number of degrees of freedom adjusted to the result of the copula fit

- With the empirical margins, ellipticality rejected for horizons of 8 hours and shorter

- Extreme tails best described by Clayton resp. survival Clayton copula, as predicted by theory

CONCLUSION: FINAL

- This is a first analysis of the bivariate case
- The paper raises a lot of questions

  For example:

  - Further details on the two-dimensional stylized facts
  - What about temporal interdependence
  - High-dimensional data, beyond two
  - Multivariate deseasonalisation