

# Chapter 5

## Application

To illustrate the above techniques, several experiments has been performed. For most of them we used daily observations of plane co-ordinates of a point gathered 2 years by means of NAVSTAR Global Positioning System (GPS) on several permanent stations that takes part in European Reference Network. Establishment of such a network serves for various geodetic and geophysical purposes, e.g., for regular monitoring of recent kinematics of the Earth's crust (local, regional and global). For further details about principles and applications of GPS see [28].

Organization of the chapter follows ordering of theoretic chapters above. The first section deals with testing for linearity, modelling and forecasting with nonlinear models. Next, we model the same time series by linear models considering also the common trend feature and compare the best models (nonlinear, ordinary linear and linear with cointegration) with respect to their out-of-sample performance. Testing for common deterministic trends and common deterministic seasonality takes part yet in the second section. Finally we provide some experimental study of copulas.

As another asset of this thesis we demonstrate the model specification for simulated nonlinear bivariate time series on algorithm coded in computer algebra system Mathematica (currently in version 5.2). The system was chosen for its vast potential for research use, moreover the code can be exported to commonly used lower-level programming language *C*, or implemented as web based application freely available through the Internet. The demonstration procedure together with all necessary functions (including our "flagship" – the function `fConditionalRegimeSwitching`) are placed in Appendix, a brief description is given at the end of this chapter.

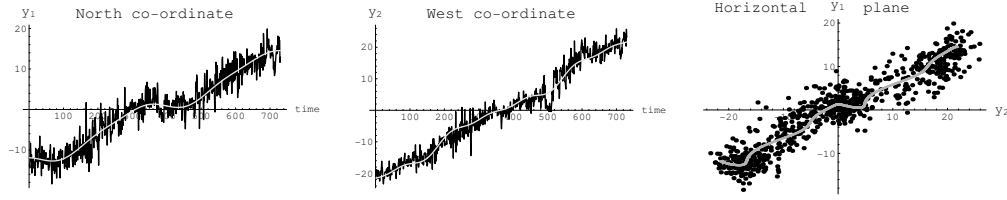


Figure 5.1: Observations at station BOR1 fitted by linear and seasonal pattern, scaled in [mm/day].

## 5.1 Threshold and smooth transition nonlinearity

Time series subjected to nonlinearity analysis was chosen to contain also significant trend and seasonal pattern in order to compare out-of-sample performance with those of the common-feature-modification of linear models described in the next section. Such time series occur very often in geodetic practice. We used 2 years daily observations on permanent station BOR1 (Poland), in particular the north coordinate as  $y_{1t}$  and east coordinate of horizontal topocentric system as  $y_{2t}$  variable.

### Model specification

The experiment examined two regression setups: regressing variables on deterministic (trend and seasonal) and stochastic (lagged self variables) either separately in 2 consequent or jointly in single OLS regression. The orders of AR (or VAR) models were found by minimising AIC, BIC and HQIC<sup>1</sup> information criteria to be  $p = 4$  ( $p = 2$  or  $4$ ) for both series. Although there exist suspicion for several outlying sequences in data (also in forecasted part), due to lack of metadata this was left explicitly untreated and hoped for being described by regime-switching mechanism. Spectral analysis confirmed the seasonal pattern present in both annual and semi-annual frequency, i.e. period  $S = 365, 365/2$ . Other potentially significant cycles were neglected. The residua of decomposition was tested for linearity (against regime-switching nonlinearity) by Tsay's test and LM-type test. Both of them, taken over  $p \times d$  grid up to value 7 (days), rejected the linearity in most cases (assuming the usual  $\alpha = 0.05$ ), yet the correct value of delay  $d$  cannot be selected with uniqueness. However, p-value displayed minimal values mostly for the

<sup>1</sup>Hannan-Quin information criterion HQIC is defined similarly to BIC in (1.56) as  $HQIC(p) = n \ln |\hat{\Sigma}_p| + 2 \ln(\ln(n))k^2p$ .

delays up to 4 days which is also more acceptable than larger values, from the practical experience point of view.

Due to sizeable computational intensity and certain instability of nonlinear least square fit routine we put our attention on ordinary least square fit conditional on fixed values of unknown parameters. Thus a three dimensional grid was made from approximate discrete values of delay ( $d = 1, 2, 3, 4$ ), threshold ( $r_1 = -2.3, -2.0, \dots, 2.2$  or  $r_2 = -2.2, -1.9, \dots, 2.0$ ) and smoothness parameter ( $\gamma = 1, 2, 10, 100$ ). Then the triplet that minimized trace of variance-covariance matrix of residuals was considered to be the estimate of parameters  $d, r, \gamma$ . The approximate values of threshold was chosen to preserve at least 15% of observations for each regime, whereas the smoothness parameter was to reflect only illustrational purposes. For  $\gamma = 100$ , LSTAR model can be regarded as TAR, therefore no extra estimation was performed for pure threshold setup counting only two parameters. We also applied few aggregation functions ( $\mathcal{MIN}$ ,  $\mathcal{MAX}$ ,  $\mathcal{M}$ ) to find out whether better results can be achieved comparing to basic setup ( $z_t = \xi_{t-d}$ , which corresponds to aggregation function **Last** in the appended program). The improvement was testified in case of arithmetic mean. The two-regimes model were estimated for a bivariate ( $\mathbf{y}_t = (y_{1t}, y_{2t})$ ) and then for two univariate time series, with the variable  $\xi_t$  being one of the modelled variable but freed of trend and seasonal component. To confront the in-sample fit with out-of-sample performance the same model, the mean squared prediction error was computed taking 30 one-step-ahead forrecasts into account. That means that as the length of series is  $n = 730$ , every model was fitted to first 700 observations and the last 30 (approximately  $\sqrt{n}$ ) served as etalon to forecasts. Results of the experiment can be seen in Table 5.1.

What is obvious at the first glance, is that transition between regimes is rather crisp than fuzzy. At other glance, the improvement brought by aggregation operator is too rare to make any optimistic conclusions.

Since this thesis does not set itself a task of rigorous examination of given data, neither additional transition type, nor three or more regimes, nor other aggregation operators were implemented. The focus is put on showing theoretical possibilities of STAR type models and we provide algorithm that can be helpful for practitioners in carrying out their practical experiments. The demonstration on simulated data given in Appendix can be thought of as user's manual additional to the self descriptions of constituent functions. All of them, especially the mentioned **fConditionalRegimeSwitching**, which is also the largest one, are built up in as most general way as possible in given time, yet with the least impact on computational efficiency.

Table 5.1: Conditional estimation of STAR and TAR model parameters. Models preferred from in-sample fit ( $\text{Tr}(\Sigma)$  minimized) with corresponding out-of-sample fit measure (MSPE).

*Autoregression of trend- and season-free series*

modelled var.	transition var.	estimated parameters			$\text{Tr}(\Sigma)$	MSPE
$\mathbf{u}_t$	$u_{i,t-d}$	$d$ [day]	$r$ [mm]	$\gamma$	[mm <sup>2</sup> ]	[mm <sup>2</sup> ]
$(u_1, u_2)$	$u_1$	2	-0.8	100	9.44	8.55
	$u_2$	1	-0.7	100	9.48	8.13
$(u_1)$	$u_1$	4	-2.3	10	5.14	5.59
$(u_2)$	$u_2$	4	-2.2	100	4.41	3.10
$\mathcal{M}(u_{i,t-1}, \dots, u_{i,t-d})$						
$(u_1, u_2)$	$u_1$	1	-0.8	100	9.47	8.98
	$u_2$	4	-1.6	100	9.45	8.09
$(u_1)$	$u_1$	4	1.9	10	5.13	5.37
$(u_2)$	$u_2$	4	-1.6	100	4.40	3.21

*Autoregression with both deterministics*

modelled var.	transition var.	estimated parameters			$\text{Tr}(\Sigma)$	MSPE
$\mathbf{y}_t$	$u_{i,t-d}$	$d$ [day]	$r$ [mm]	$\gamma$	[mm <sup>2</sup> ]	[mm <sup>2</sup> ]
$(y_1, y_2)$	$u_1$	2	-0.8	100	9.35	7.90
	$u_2$	1	-0.7	100	9.37	7.85
$(y_1)$	$u_1$	2	-0.2	10	5.01	4.87
$(y_2)$	$u_2$	1	-2.2	100	4.31	3.54
$\mathcal{M}(u_{i,t-1}, \dots, u_{i,t-d})$						
$(y_1, y_2)$	$u_1$	1	1.6	100	9.39	8.26
	$u_2$	4	-1.6	100	9.34	7.93
$(y_1)$	$u_1$	4	1.9	10	5.08	4.94
$(y_2)$	$u_2$	4	-1.9	100	4.30	3.82

$\mathcal{M}$  denotes arithmetic mean and  $u_i$  are residual series from static regression

$$y_i = \mu_i + \delta_i t + \beta_{i1} \cos\left(\frac{2\pi}{365}\right) + \beta_{i2} \sin\left(\frac{2\pi}{365}\right) + \beta_{i3} \cos\left(\frac{4\pi}{365}\right) + \beta_{i4} \sin\left(\frac{4\pi}{365}\right) + u_i$$

## 5.2 Common trend and seasonality in linear models

In this section, we first focus on testing for the presence of stochastic and deterministic trend in time series modelled already above which are consequently transformed into 2 new variables – one containing full information about trend whilst the another being trend-free. The same decomposition is done with (deterministic) seasonality and the new systems of variables are exposed to modelling, forecasting and backward transformation into original system wherein the out-of-sample performance is quantified. Next we deal with testing for common deterministic trend applied to observations from 5 permanent stations.

### Transformation into common trend and common seasonality system

Again we took the time series from BOR1 ( $n = 730$ , daily solutions) under subjection. The first thing when interested in common trend behaviour is to check whether each of the time series are nonstationary (trend-stationarity or unit root). Augmented Dickey-Fuller test for unit root labelled the series as trend-stationary therefore containing only deterministic trend. The next step is usually<sup>2</sup> to verify the series to have the trend in common. This was done by testing both in single equation by Engle-Granger test and in the system of equations as proposed by Johansen. Results of all above tests can be found in Table 5.2.

In the system of two variables  $(y_{1t}, y_{2t})$ , one cointegration relation was found represented by cointegrating vector  $\mathbf{v}_1 = (-0.855, 0.511)'$  and supplemented by exactly one common stochastic trend to be obtained from vector  $\mathbf{w}_2 = (0.500, 0.866)'$  (both vectors are normed to unity length). Thus transformation matrix  $\mathbf{T}_T = (\mathbf{w}_2, \mathbf{v}_1)$  get us from system, say,  $Y(y_{1t}, y_{2t})$  to "common trend" system  $Y^T(y_{1t}^T, y_{2t}^T)$ . The geometric interpretation (3.43) illustrated in Figure 3.1 (wherein the axes  $u_1, u_2$  represent system  $Y^T$ ) corresponds very closely to the above transformation for rotation angle  $\alpha \doteq 59^\circ$ . Observations transformed into  $Y^T$  (see Figure 5.2) were then modelled in an ordinary way, that is by AR(4) model with linear trend ( $y_{1t}^T$  only) and trigonometric polynomial of annual and semi-annual period. After computing 1-step-ahead forecasts these were transformed back into  $Y$ , i.e.  $\mathbf{y}_t = \mathbf{T}_T^{-1} \mathbf{y}_t^T$  for  $t = 701, \dots, 730$ , and compared to real data with MSPE as the final outcome.

---

<sup>2</sup>If more general - stochastic - trend is found.

Table 5.2: Testing for unit roots and cointegration (at  $\alpha = 0.05$ )*Augmented Dickey-Fuller test for unit roots*

variable	deterministic component	$t(\hat{\rho})$	critical value	inference
$y_1$	none	-1.70	-1.95	$\rho = 0 \Rightarrow$ ST
	constant	-1.69	-2.86	$\rho = 0 \Rightarrow$ ST
	constant & trend	-7.59	-3.41	$\rho < 0$ , accept DT
$y_2$	none	-0.95	-1.95	$\rho = 0 \Rightarrow$ ST
	constant	-0.96	-2.86	$\rho = 0 \Rightarrow$ ST
	constant & trend	-7.19	-3.41	$\rho < 0$ , accept DT

*Engle-Granger test for cointegration*

regression of	deterministic component	$t(\hat{\rho})$	critical value	inference
$y_2$ on $y_1$	constant	-8.69	-3.34	cointegration
	constant & trend	-9.56	-3.78	cointegration
$y_1$ on $y_2$	constant	-9.71	-3.34	cointegration
	constant & trend	-9.90	-3.78	cointegration

*Johansen's test for cointegration ( $k = 2$ , case 3)*

r	test	statistic	critical	inference
0	Trace	77.18	17.95	$H_0$ rejected
1		0.44	8.18	$H_2$ rejected, $r = 1$
0	$\lambda$ -max	76.73	14.90	$H_0$ rejected
1		0.44	8.18	$H_2$ rejected, $r = 1$

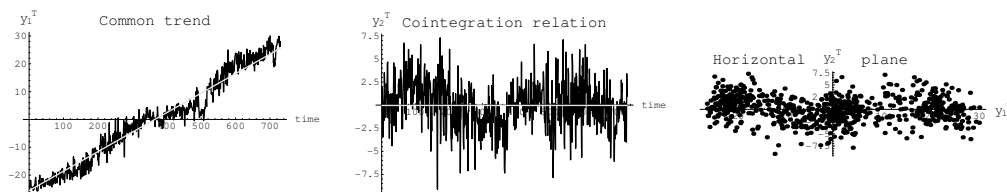


Figure 5.2: Observations at BOR1 in common trend coordinate system.

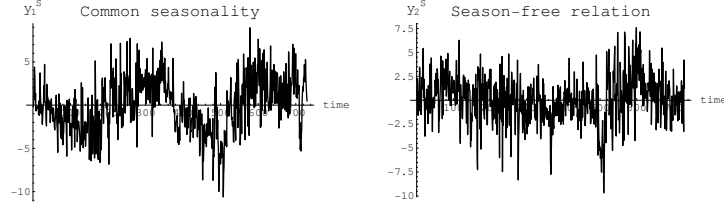


Figure 5.3: Trend-free common seasonality decomposition of BOR1 data.

Further on we were interested in testing for common deterministic seasonality at annual and semi-annual frequency. The test statistics (3.62) indeed fell into 95% confidence interval of  $\chi^2(r[2(4-1)+r])$  distribution for  $r = 1$  season-free relations. As we are interested in what improvement can bring such decomposition, we constructed transformation matrix  $\mathbf{T}_S$  from eigenvectors  $\mathbf{w}_1, \mathbf{w}_2$  brought by in the test such that  $\mathbf{T}_S = (\mathbf{w}_2, \mathbf{w}_1)'$ , and used the matrix for transformation into system  $Y^S(y_{1t}^S, y_{2t}^S)$  of which one axis contains full information about seasonal fluctuations while the another is free of the seasonal effects. As a matter of experiment the transformation was imposed on both to original series and to those decomposed according to common trend priority, however in either case, the linear trend had been removed in advance to be added back later (after backward projection) to forecasts. The series in  $Y^S$  were modelled by AR(4) model with seasonal trigonometric terms ( $y_{1t}^S$  only).

As a benchmark we used univariate as well as vector autoregression approach, with or without deterministic terms in the same OLS procedure, applied to original time series. To retain comparability, the deterministic part is restricted to linear and trigonometric polynomial with period  $S = 365$  and  $S = 365/2$  while the autoregression has the constant order  $p = 4$  with one exception granted for VAR model once in order to preserve the number of parameters. Results can be found in Table 5.3.

From the results it is immediately apparent that both transformation ideas bring significant improvement in out-of-sample performance of a linear model. The amount to which every approach contribute to better description of particular variable depends on how much information about the certain nonstationarity-driving process is involved in other variables.

Alternatively to daily solutions at single GPS permanent station, we checked the performance of season-transformation also on subdaily observations (3 hours solutions) gathered for 43 days at permanent stations GANP (Ganovce-Poprad) and LOMS (The Lomnický peak) situated in the same region but different altitude (elevation difference approx. 2km). We focused on north-

Table 5.3: Out-of-sample performance of linear models considering common-feature-based transformations.

*Autoregression of deterministic-terms-free series*

common feature transformation	AR/VAR	MSPE [mm <sup>2</sup> ]		
		$y_{1t}$	$y_{2t}$	$\mathbf{y}_t$
—	AR(4)	5.27	3.17	8.44
—	VAR(4)	5.31	3.09	8.40
—	VAR(2)	5.54	3.04	8.58
trend		4.51	3.12	7.63
trend, season	AR(4)	4.27	3.39	7.66
season		4.29	3.38	7.67

*Autoregression with relevant deterministics*

common feature transformation	AR/VAR	MSPE [mm <sup>2</sup> ]		
		$y_{1t}$	$y_{2t}$	$\mathbf{y}_t$
—	AR(4)	5.17	3.20	8.37
—	VAR(4)	5.20	3.11	8.32
—	VAR(2)	5.41	3.06	8.47
trend		4.47	3.16	7.63
trend, season	AR(4)	4.26	3.39	7.65
season		4.28	3.38	7.66

ward seasonal co-movement in trend-adjusted time series, this time a dummy variables (as in (1.45) with  $S = 8$ ) were used instead of sines and cosines. As a result, the presence of common deterministic seasonality was found with the test, albeit for significance level  $\alpha = 0.01$ . One-step-ahead forecasting through transformed series again brought better MSPE ( $5.58 + 4.36 = 9.94$ ) compared to univariate AR ( $5.85 + 4.64 = 10.49$ ) or multivariate VAR model ( $5.74 + 4.77 = 10.51$ ). Summands in the brackets represent MSPE of  $y_{1t}$  and  $y_{2t}$  in respective order.

### Testing for common deterministic trend slopes

For this test we used 5 time series (subsequently north  $y_{1t}$ , east  $y_{2t}$  co-ordinate observations and common trend direction  $y_{1t}^T$ ) from permanent stations denoted in international framework as BOR1 (Poland), GOPE (Bohemia), POTS (Germany), HFLK (Austria), PENC (Hungary) stacked in vector  $\mathbf{z}_t$  in the same order as listed above. Table 5.4 contains OLS estimates of para-



meter  $\delta$  in units mm/year.

Table 5.4: Trend parameter estimates [mm/year]

point:	BOR1	GOPE	POTS	HFLK	PENC
$\hat{\delta}_{y_{1t}}$	14.2	14.7	14.4	13.4	12.2
$\hat{\delta}_{y_{2t}}$	22.3	23.1	21.1	21.5	23.9
$\hat{\delta}_{y_{1t}^T}$	26.4	27.4	25.5	25.3	26.9

By defining the  $q \times k$  matrix  $\mathbf{R}$  and  $q \times 1$  vector  $\mathbf{r}$

$$\mathbf{R} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \quad \mathbf{r} = \begin{pmatrix} \hat{\delta}_1 \\ \hat{\delta}_1 \\ \hat{\delta}_1 \\ \hat{\delta}_1 \end{pmatrix},$$

the null hypothesis (3.51) says that the first (reference) time series has the same slope as the rest of time series. Moving the zero column in  $\mathbf{R}$  and filling  $\mathbf{r}$  with corresponding parameter, slopes of other series are tested.

We also consider a joint test, when  $q = k$ ,  $\mathbf{R}$  is  $k$ -dimensional identity matrix and  $\mathbf{r}$  contains trend parameter of an average time series,  $\bar{n}_t = \frac{1}{k} \sum_{j=1}^k (\mathbf{n}_t)_j$  for instance. Test results are collected in Table 5.5 and 5.6, critical values for  $F$ -tests are given in a separate line bellow. Every Wald test statistic is provided with  $p$ -value, which (given null hypothesis is valid) represents a probability that we get the particular value or even the more contradictory to tested hypothesis, in other words it expresses a probability the trend slopes are statistically the same.

First to conclude, all the tests confirmed common deterministic trend in north direction coordinate for all the series except for PENC. Although the joint test for  $q = 5$  did not reject the null hypothesis due to quite a large set of variables, from the five tests within equations it is obvious that this last slope does not correspond to the others. As for other two variables,  $y_{2t}$  and  $y_{1t}^T$ , no significant deterministic relation was found. Therefore next we took foursomes into account (Table 5.6) removing seemingly most troublesome series, and observed how it changed the test statistics. Removing PENC accentuated the common trend in  $y_{1t}$ , and pointed slightly at promises in  $y_{1t}^T$ . It is useful now to repeat that slope of  $y_{1t}^T$  represents a velocity in the direction a particular point moves, and does not contain information about direction.

[illegible]

Honestly, we don't know the cause of this minor effects in tested parameters, there may be some speculations uttered about local instability of given points in particular direction or some residual systematic components in time series

[illegible]

that weren't taken into account in model specification and which might cause spurious estimates. Nevertheless, although the data visibly show significant linear trend behaviour caused by tectonic plate drift, the tests rejected common deterministic trend for two of the observed points. It is subject to study, why this happened.

### 5.3 Comparison

As comparison by simple ordering the MSPE values can be slightly misleading, we adopt here the modified Diebold-Mariano test to statistically distinguish the predictive accuracy of models found in the first and second section as the most appropriate ones.

Table 5.7: Testing the equality of predictive accuracy (at  $\alpha = 0.10$ )

reference model		compared to model No.						sum	hit parade
description	No.	1	2	3	4	5	6		
LST VAR(4) $u_2$ (1,-0.7,100)	1	·	1	1	-1	-1	-1	-1	2
AR(4)	2	-1	·	0	-1	-1	-1	-4	3
VAR(4)	3	-1	0	·	-1	-1	-1	-4	3
AR(4) in $Y^T$	4	1	1	1	·	0	0	3	1
AR(4) in $Y^{TS}$	5	1	1	1	0	·	0	3	1
AR(4) in $Y^S$	6	1	1	1	0	0	·	3	1

For every 2-tuple of the six selected models, the MDM test statistic (1.32) had been checked whether it falls into 90% confidence interval of Student t-distribution with  $(30 - 1)$  degrees of freedom, and if so, zero value was assigned to the pair denoting their equality in accuracy. Otherwise if MDM is smaller (larger) than 0.05 (0.95) quantile, the first model in the pair is significantly more (less) accurate and value 1 (-1) is assigned. The test results are shown in Table 5.7. According to the "hit parade" column, the linear models applied in common-feature-transformed systems are the winners of our competition.

### 5.4 Fitting copulas

We employed bivariate time series - daily observations of plane co-ordinates of a point gathered 2 years (728 realizations). Observations were made by

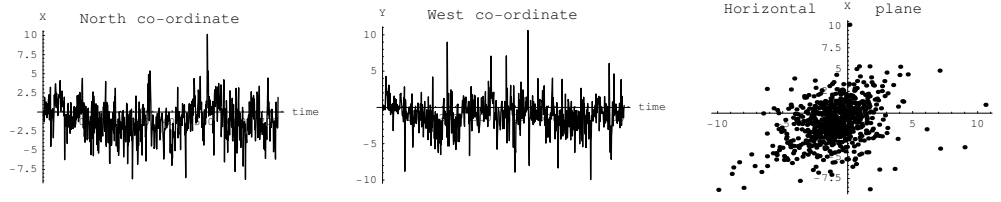


Figure 5.4: Two univariate time-series linked together to form bivariate random vector of a point location

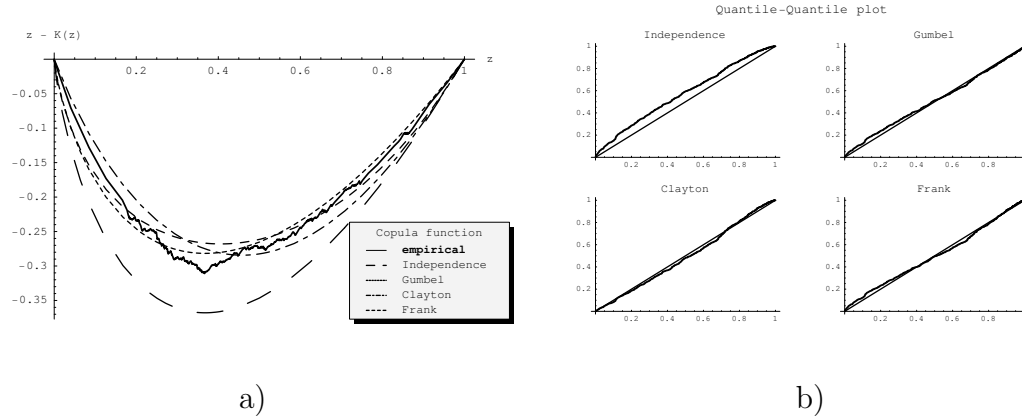


Figure 5.5: Graphical evaluation of nonparametric method:  
a) Empirical function  $K_n$  fitted by  $K_\phi$  of corresponding copula function  
b) Quantile-quantile plots

means of GPS on permanent station MOPI taking part in European Reference Network too. The two random variables that make our bivariate observations thus share common physical phenomenon through the geometry and time reference. Indeed, from visual check on Figure 5.4, we may expect some dependence.

The data was processed as follows. Firstly, we examined the two individual univariate time-series. Interestingly, both of them follow logistic distribution rather than normal. The logistic distribution with *mean* and *scale* parameter is frequently used in place of the normal distribution when a distribution with longer tails is desired. Nevertheless, further on we worked solely with the empirical marginal distribution function (4.18) to avoid any influence of a biased marginal model upon estimation of dependence structure. Next we computed scalar representatives of this structure, that is, measures of dependence

Correlation coef.	Spearman's $\rho$	Kendall's $\tau$
0.3670	0.3314	0.2346 .

Table 5.8: Nonparametric and semi-parametric estimates of copula dependence parameters

a)			
Family:	Gumbel	Clayton	Frank
Nonparametric procedure			
$\theta$	1.3066	0.6131	2.2117
$d(K_\phi, K_n)$	0.446	0.543	0.493
Log-Likelihood procedure (semi-parametric)			
$\theta$	1.3044	0.5638	2.3153
AIC	-106.2	-109.0	-90.7
$d(C_\theta, C_n)$	5.665	6.370	6.003
Nonlinear Fit procedure (semi-parametric)			
$\theta$	1.2914	0.5597	2.0245
$d(C_\theta, C_n)$	5.612	6.369	5.219

b)			
Linear convex combination:	Clayton-Gumbel	Clayton-Frank	Frank-Gumbel
Nonlinear Fit procedure (semi-parametric)			
$\alpha$	0.4437	0.3185	0.6345
$d\left(\alpha C_{\theta_1} + (\alpha - 1)C_{\theta_2}, C_n\right)$	3.953	4.850	5.013

Note that, if the data were nonstationary and required some variance stabilizing such as logarithmic transformation (which is strictly increasing), the pre-processing would have biased only the correlation coefficient, and none of the others.

Within a semi-parametric procedure, (a) we firstly applied the second procedure outlined in section 4.4, (b) then as an alternative (and as a backup too) we utilized nonlinear parametric least-square fit to empirical copula. For linking both (a) and (b) approaches, we computed  $L_2$ -norm distance between estimated and empirical copula. As seen from Table 5.8, the differences are nonsignificant and in preferring Gumbel family to Frank and Clayton both methods agree with the nonparametric one. However, there seems to be a disharmony with AIC criterion of maximum likelihood estimate goodness, which surprisingly promotes the Clayton. On that account we performed some computations under different input conditions and figured out, that log-likelihood function of Clayton copula density (see Figure 5.6) is pretty sensitive to lower tail dependencies, namely to "perfect" extremes in data

(notice the lower tail protruders in the very right-hand plot of Figure 5.5). Even just one (the most extremal) outlier chopped off from the lower tail of the data pushed the *AIC* of Clayton to between Frank and Gumbel. Dropping the other two degraded Clayton into "least appropriate" position among copulas under consideration. Upper tail extremes have no evident impact to Clayton likelihood estimate.

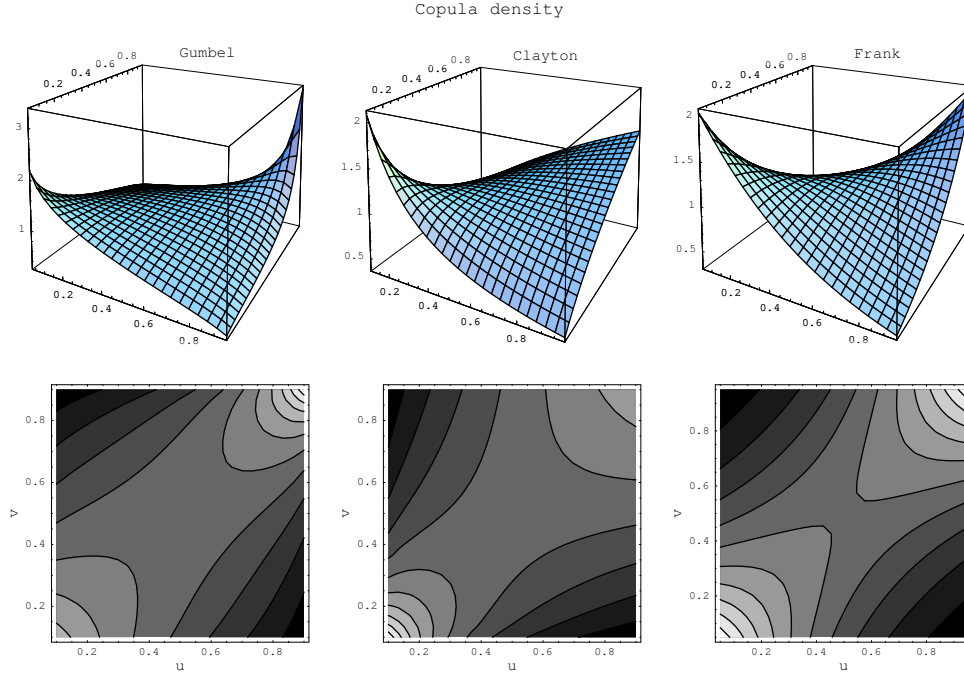


Figure 5.6: Copula density for three Archimedean families

This kind of "revelations" appears to be quite important when choosing the best copula. Since nonlinear least-square fit demands a much more CPU time and memory, discussion of the nonparametric and semi-parametric (pseudo log-likelihood) is surely in order. As mentioned in [1], neither method is generally more convenient, but if there are outliers or if the marginal distributions are heavy tailed, it seems reasonable to choose the nonparametric approach. If we work with large data set, the likelihood estimator may be more precise.

There are many families of copula, that could be estimated by above procedures and, if necessary, should be considered as the alternatives to the three above but mainly to most used Gaussian distribution, which - by its nature - cannot be satisfactory in numerous applications. In that of ours, the sum of squares of residuals unambiguously refused the appropriateness of bi-normal

distribution.

Finally, as we have estimated the copula parameters by particular method and chose "the best" of them, we contemplated a possibility to improve the nonlinear fit of parametric copulas by simply fitting their linear convex combinations to empirical copula and compare the  $L_2$  distances. It can be shown, that the linear convex combination  $\alpha C_1 + (1 - \alpha)C_2$  of any two copulas  $C_1$  and  $C_2$  is also a copula with parameter  $\alpha \in [0, 1]$ . Such a copula may possess benefits of both parents when fitting empirical copula. And indeed, Table 5.8 supports this assumption. The best combination is given by Clayton and Gumbel, with a slight dominance of Gumbel.

## 5.5 Program outline

The source code provided for Mathematica users in Appendix and on the website [www.math.sk/bacigal/homepage/research\\_en.html](http://www.math.sk/bacigal/homepage/research_en.html) spreads in four calculation sheets (notebooks) with the following content:

1. Regime-switching modelling of the artificial, TAR generated bivariate stationary time series with detailed guidance to data preliminary analysis, testing, model specification and evaluation. Afterwards a survey on real data modelling is given in a more compact way.
2. Linear modelling of nonstationary bivariate time series with significant deterministic trend and seasonal component. Three main approaches are compared in term of out-of-sample performance: univariate AR and multivariate VAR model of given data, and multivariate VAR model applied to data transformed such that the deterministic components are segregated in single variable.
3. Definition of all functions utilized in the foregoing notebooks. In such a form, functions need to be evaluated manually at the beginning after all the necessary built-in Mathematica packages has been loaded. An alternative way of initialisation is to load a package provided on our website, where functions are encapsulated and can be easily summoned with automatic call for other required packages.
4. Analysis of dependence structure of a vector variable. In three subsections three one-parameter copulas are fitted to real data, then a linear extension to three-parameters copula is considered and additionally standard errors for copula parameters are computed.

In the following paragraphs we try to outline the potential of our routines in particular stage of time series analysis. Note that their naming differs from that of Mathematica built-in functions in starting character `f`.

For more exhaustive details we recommend reading the description provided within the notebooks.

### Preliminaries and linear modelling

Generally the notebooks begin with initial settings such as loading of packages and data followed by preliminary analysis by plotting correlations or spectral densities. To find order of linear AR or VAR model the selection criteria AIC, BIC and HQIC are computed via Levinson-Durbin procedure provided by the `TimeSeries` package. Alternatively we generalise implementation of this procedure to account for exogenous time series. If data are nonstationary because of deterministic components, namely polynomial trend (constant, linear, quadratic...) or sine waves, and stationarity is required for further processing, `fDeterministicsRemoval` uses OLS to return deterministic-free residuals. Irregular cyclic patterns and lagged data may enter the linear model through design matrix and `fΦ`.

### Regime-switching nonlinearity

Two tests of linearity are available, `fTsayLTest` for testing against threshold and `fLMtypeLTest` against smooth switching (either logistic or exponential) alternative. Both functions allow to use aggregation operators (some are predefined in `fAggregationOperator`) and exogenous variables, they return test statistic and degrees of freedom that are converted into p-value by `fChiSquarePValue`. If applied over grid of input parameters (order and max. delay), `fArgMinTensor` finds the best configuration which minimizes the p-value.

Estimation of regime-switching model is the job of `fConditionalRegimeSwitching`. As the name betrays, OLS (conditional on delay  $d$ , threshold  $r$  and smooth parameter  $\gamma$ ) applies instead of NLS. The function has seven arguments, it allows both endogenous (response) and exogenous variables to enter the regression, also with lags that can be arbitrarily chosen either by their maximum values (*orders*) or listed, jointly or separately for every variable, jointly or separately for every regime. Parameters  $(d, r, \gamma)$  constitute the fifth argument, either as lists of discrete values or uniform sequence given by its boundary values and step. Here `fThresholdRange` comes useful since it divides central bandwidth of transition variable to get sequence of thresholds. If we prefer abrupt transition (threshold model), parameter  $\gamma$  must be



excluded from the list. The last argument determines the general purpose of function and may contain additional input data. Its first element serves as switch across modes (see below), between logistic and exponential transition function, and over the options for threshold variable construction. Second element carries a name of aggregation operator and the third element covers the needs of individual mode. The modes of `fConditionalRegimeSwitching` are nothing more than particular tasks to be executed and outputs to be returned. At present, five modes (kinds of output) are available: 1) sum of squared residuals as trace of covariance matrix, 2) estimates of parameter matrix,  $\hat{\Phi}$ , covariance matrix,  $\hat{\Sigma}$ , and residuals,  $\hat{\epsilon}_t$ , 3) AIC and BIC selection criteria, 4) forecast errors and finally 5) forecasts. The forecast errors are computed over the sample data tail, the length of which have to be specified together with horizon and number of Monte Carlo cycles. If forecasts are desired from regression on exogenous variables, these must be generated in advance and put in *specification* argument.

The results are structured as nested lists according to parameters  $d, r_1, \dots, r_{m-1}, \gamma_1, \dots, \gamma_{m-1}$  (in case of  $m$ -regimes). Values of the parameters are situated in the second part of output. For modes 1 and 3 the function `fArgMinTensor` searches the (first part of) output heap and returns the parameters corresponding to the global minimum. They then enter the mode 2 and the result is subjected to test for residual correlations (`fPortmanteauTest`) and linearity (`fResidualTest`).

### Comparison of predictive accuracy

As an evaluation of out-of-sample fit, two kinds of "hit parades" come available. By `fMeanXPredictionErrorHitparade` we may directly compare overall MSPE, MAPE or other measures of fit, while `fDieboldMarianoHitparade` at first tests the equality of predictive accuracy and then returns ranking which may award one position to multiple entries. Both functions are fed with forecast errors and a loss function, Diebold-Mariano test additionally requires horizon of forecasts and significance level.

### Estimation of dependence structure

Routines in the last notebook need not many comments as they consist of simple commands and built-in functions. It might be interesting to notice computing times by each estimation method provided by `Timing`. The most consuming procedure at all is the generation of empirical copula which takes half an hour on a 1GHz machine. Obviously this is one of the main drawbacks of such computer algebra system as Mathematica which stays far behind

(e.g.) C-compiled programs in the sense of data handling speed.