CAWTHRON

Report No. 1135

An approach to the evaluation of temporal trends in Taranaki state of the environment macroinvertebrate data

Prepared for

TARANAKI REGIONAL COUNCIL

April 2006

An approach to the evaluation of temporal trends in Taranaki state of the environment macroinvertebrate data

Prepared for

Taranaki Regional Council

by

John D. Stark
Chris R. Fowles[1]

Cawthron Institute
98 Halifax Street East
Private Bag 2
NELSON
NEW ZEALAND

Phone: +64.3.548.2319
Fax: +64.3.546.9464
Email: info@cawthron.org.nz

[1] Taranaki Regional Council, Stratford

Cover Photo: Upper reaches of the Waiwakaiho River (Photo: TRC)

EXECUTIVE SUMMARY

In this report we examine several simple statistical approaches for the detection of significant trends in stream macroinvertebrate biological indices. These include non-parameter tests based on the Mann-Kendall or Spearman rank correlations and a parametric approach using linear regressions.

We recommend the following cost-effective method for examining trends in macroinvertebrate biological data.

1. Visualise the trend using scatterplots of biological index vs time with LOWESS fit.
2. Test for significance using the Mann-Kendall test at the 5% significance level followed by Benjamini-Hochberg False Discovery Rate (FDR) analysis. The trends remaining significant following this procedure should be clear trends, although the final decision on ecological significance rests with the best professional judgement (BPJ) of an experienced freshwater macroinvertebrate ecologist.

Although this approach is appropriate for trends analysis of a range of biological indices, we recommend that high-level SEM reporting should focus on the MCI, since this index has a strong track record as an indicator of river health in New Zealand. There are real problems in using indices such as taxa richness and SQMCI or QMCI for reporting SEM results. Besides, reporting using multiple indices often raises some anomalies that a biologist can explain, but which only serve to confuse laypersons. We believe that the KISS principle should apply.

When a trend has been identified there are two vital questions that arise (Rutherford 1985). First, “What caused the trend?” and second “Will it continue into the future?” Unfortunately trend analysis alone cannot provide certain answers to these very closely related questions. If the cause/s of the trend can be determined, then one may be able to predict the future with some confidence. Conversely, if one cannot determine why a trend has occurred then extrapolation into the future may be most unwise because there remains a chance that the trend is simply an artefact of natural variability or sampling error.

Determining the reasons for any significant trends detected inevitably will require additional information, which could include information such as stream flows, weather patterns, catastrophic erosion events in the catchment, physico-chemical water quality, or changes to land or water management practices that may have resulted from TRC water management initiatives, and industry or farming activities. Discussion of reasons for trends was beyond the scope of this report. It is worth emphasising that the ultimate decision on whether or not any trend should be considered ecologically significant is reliant on the BPJ of an experienced freshwater ecologist. We caution water managers about the interpretation of statistically significant trends in stream biological health particularly if these trends are only marginally significant, cannot be explained and/or may be unrelated to initiatives aimed at improving stream condition.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

Report reviewed and approved for release by:

Rowan Strickland, Freshwater Group Manager

1. INTRODUCTION

Regional Councils are responsible for undertaking long term biological monitoring of freshwater systems in order to assess the "health" or state of the environment within their regions. One question that arises after state of the environment monitoring (SEM) has been undertaken for several years is whether or not conditions have got worse, improved, or stayed the same. This is a typical application for a class of statistics called time series or trends analysis.

There are a variety of different techniques for time series analyses that vary in their data requirements and complexity. Linear regression-based parametric methods should only be applied when the trend is expected to be linear (which it often is not) unless data transformations are used. Furthermore, there are problems with parametric methods like linear regression if there is heteroscedasticity in the data (*i.e.*, variance differs with time). Non-parametric techniques for trend analysis are much better able to handle non-normal data with censored, tied, and missing values, so they have found favour for analysing trends, particularly in water quality data.

A popular non-parametric trend test for water quality data is the seasonal Kendall trend test – a technique described by McBride (2005) and implemented by Bill Vant[1] (Environment Waikato) (see Vant & Smith 2004). However, this technique requires monthly data collected for at least three, and preferably five, years or more (*i.e.*, 36 - 60 data points). This seldom is the case for biotic data where sampling may be seasonal at best, and is more often undertaken only once or twice per year. Although the seasonal Kendall test *could* be adapted for detecting trends in biotic data collected much less frequently, it is doubtful whether it would be worth the effort compared with less complicated methods that are likely to give a similar result.

This report outlines a practical approach to visualising and assessing the significance of temporal trends in biotic and biological indices.

2. LONG-TERM MONITORING SITES

Taranaki Regional Council (TRC) established a SEM macroinvertebrate monitoring programme in mid 1995 (TRC 2005). This programme has continued unchanged since then. Following the collection of ten years data (in 2005) from 51 sites (Table 1) it was considered timely to assess the significance of temporal trends in these data. Twelve of the sites (bold site code in Table 1) are regarded as reference sites since they are located near the Egmont National Park boundary and do not experience significant anthropogenic disturbance. Data from a further nine sites (Table 2) that have been monitored routinely over time have also been extracted from the TRC biological database. Standard TRC sampling protocols (TRC 2004, 2005) were used throughout the ten year period with SEM sampling undertaken primarily in spring (October to mid December) and summer (late January to March) each year (Table 3). The field sampling technique was very similar to Protocol C1 of the national protocols (Stark et al, 2001). No sampling was undertaken within seven days of a stream fresh in excess of three times median flow or within ten days of a fresh in excess of seven times median flow. All samples were processed by the TRC following Protocol P1 (Stark et al. 2001) and included regular quality control procedures (Protocol QC1 (Stark et al. 2001)) instigated by the TRC.

---

[1] http://www.ew.govt.nz/enviroinfo/water/healthyrivers/waikato/documents/SeasonalKendallTrendAnalysis.xls

Table 1 Summary of biological monitoring sites included in the SEM programme. Bold site codes indicate reference sites.

Type		Sites		
		River/Stream	**Location**	**Site Code**
High conservation		Hangatahua R (Stony)	Mangatete Road	**STY000300**
			SH45	STY000400
		Maketawa S	Denby Rd	MKW000200
			Tarata Rd	MKW000300
Large catchments/Multiple impacts		Waitara R	Mamaku Road	WTR000850
		Manganui R	SH3	MGN000195
			Bristol Road	MGN000427
		Patea R	Barclay Road	**PAT000200**
			Swansea Road	PAT000315
			Skinner Road	PAT000360
		Mangaehu R	Raupuha Road	MGH000950
		Waiwhakaiho R	National Park	**WKH000100**
			SH3 (Egmont Village)	WKH000500
			Constance St, NP	WKH000920
			near mouth	WKH000950
		Mangorei S	SH 3	MGE000970
Intensive usage		Waingongoro R	Adjacent to N Park boundary	**WGG000115**
			Opunake Road	WGG000150
			Eltham Road	WGG000500
			Stuart Road	WGG000665
			SH45	WGG000895
			Ohawe Beach	WGG000995
Primary Agricultural		Punehu S	Wiremu Road	**PNH000200**
			SH45	PNH000900
		Mangaoraka S	Corbett Road	MRK000420
		Timaru S	Carrington Road	**TMR000150**
			SH45	TMR000375
(& Erosion)		Waiaua R	Wiremu Road	**WAA000200**
			SH45	WAA000447
Riparian	Western	Waimoku S	Lucy's Gully	**WMK000100**
			Beach	WMK000298
		Katikara S	Carrington Road	**KTK000150**
			Coast	KTK000248
		Kapoaiaia S	Wiremu Road	**KPA000250**
			Wataroa Road	KPA000700
			Cape Egmont	KPA000950
	Southern	Kaupokonui R	Opunake Rd	**KPK000250**
			u/s Kaponga oxi ponds	KPK000500
			u/s Lactose Co	KPK000660
			Upper Glenn Rd	KPK000880
			near mouth	KPK000990
Small Degraded ("poor") Catchments		Mangawhero S	u/s Eltham ox. ponds	MWH000380
			d/s Mangawharawhara S	MWH000490
		Mangati S	d/s railway line	MGT000488
			Te Rima Pl, Bell Block	MGT000520
Urbanisation		Huatoki S	Hadley Drive	**HTK000350**
			Huatoki Domain	HTK000425
			Near coast	HTK000745
Northern lowland catchment		Waiau S	Inland North Road	WAI000110
Major Abstraction		Waiongana S	SH3a	WGA000260
			Devon Road	WGA000450

Table 2 Selected non-SEM monitoring sites for which time-series biological data are available.

Site Code	River name	Location
KPN000275	Kapuni Stream	d/s Hawera Water Treatment Plant intake
KPN000360	Kapuni Stream	Kokiri Road
KRP000300	Kurapete Stream	u/s Inglewood Waste Water Treatment Plant
KRP000660	Kurapete Stream	6km d/s Inglewood WWTP
INH000400	Inaha Stream	Kohiti Road
MGO000050	Mangaone Stream	Egmont Road
MGO000190	Mangaone Stream	Rifle Range Road
MHW000060	Mangahewa Stream	u/s McKee Production Station
WGG000540	Waingongoro River	400m d/s Riverlands meatworks

Table 3 Monthly distribution of samples for all time-series data and SEM monitoring sites.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
All data (1125 samples)	61	285	187	22	22	4	17	11	29	208	199	80
SEM data (902 samples)	40	249	159	2	6	2	5	2	14	175	172	76

TRC provided pre-calculated taxa richness, MCI, SQMCI and % EPT$_{taxa}$ values for all available sampling dates from the sites listed in Tables 1 and 2. The analyses undertaken in this report are based upon data as received (subject to data handling required to import data from Excel into STATSTICA 7.1 and SYSTAT 10).

3. PURPOSE OF TREND TESTS

When a series of observations of a random variable (such as a chemical concentration or a biological index) have been collected over a period of time it often is of interest to determine whether there has been an increase or decrease over time. For a biotic index such as the MCI, which is known to be indicative of enrichment or "stream health", the simple question we wish to answer is "Have conditions become better or worse"? In statistical terms this is a determination of whether the probability distribution from which these data arise has changed over time. We may also like some measure of the amount or rate of change, which could be expressed in terms of changes to some central value such as the median or mean.

It is worth pointing out also that there are important implications if we wish to predict trends beyond the period of record (*i.e.*, extrapolate) compared with simply examining any trends the might appear within the sampling period. If we only wish to estimate or test trends within the scale of our data, then we need fewer data if residuals are serially correlated. In other words, it is safe to ignore serial correlation in a model that assumes independence of residuals (McBride 2005). In this report, we are concerned with detecting trends within the period of sampling and not predicting the nature and significance of trends that may occur into the future.

Furthermore, the more data one has over a given time period, the more likely it is that a trend will be detected within that period. In other words, the power of the test depends upon sample size.

4. EXAMINING TRENDS

A helpful first step in trends analysis is to plot a graph (scatterplot) of the variable of interest versus time. Let's take, for example, the Huatoki Stream at Hadley Drive (Site HTK000350) – a catchment that is rural upstream of the site and urban (new Plymouth) below the site - and plot MCI vs time. The graph suggests that there may have been an overall increase in MCI since SEM monitoring commenced in late 1996 (Figure 1)

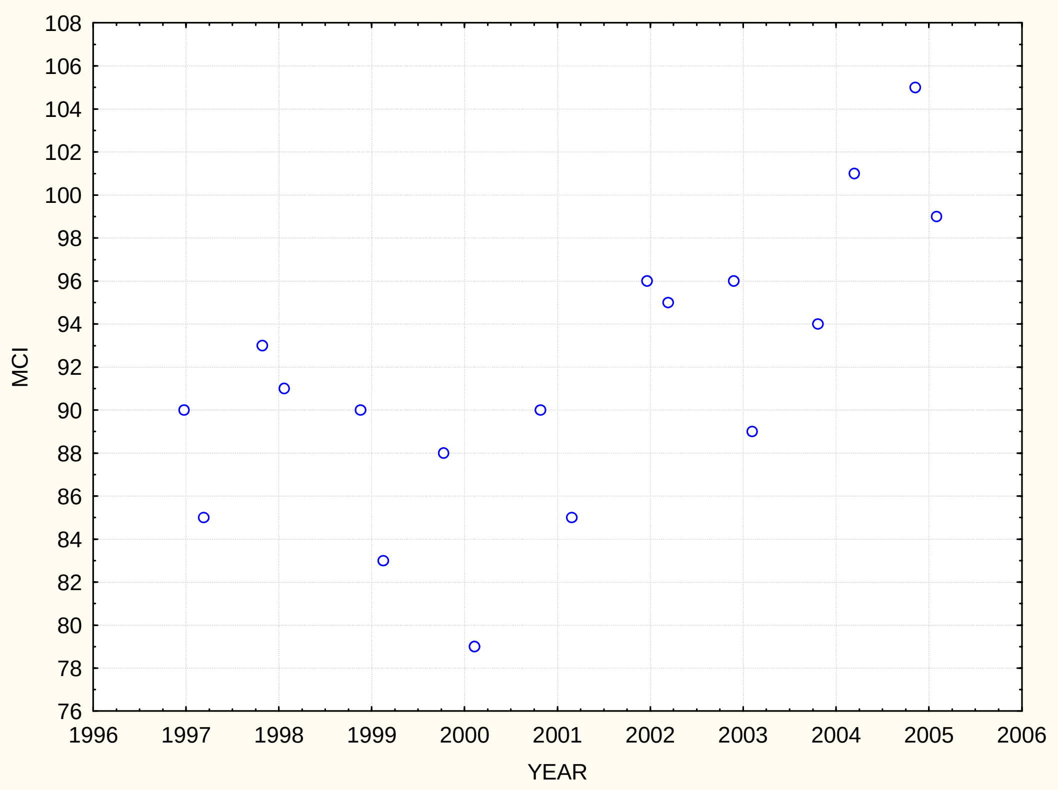


Figure 1 Scatterplot of MCI vs time for the Huatoki Stream at Hadley Drive (Site HTK000350).

Figure 1 is a simple scatterplot with no attempt to track how MCI values may have changed from one sampling occasion to another. At one extreme we could join the points (Figure 2a) and at the other fit a linear regression (Figure 2b). In terms of tracking the trend over time, however, both of these approaches often are somewhat deficient (unless there is an extremely strong linear trend in which case either method would illustrate it well). Simply tracking from data value to data value (Figure 2a) takes no account of the "error" associated with each data value. In the case of the MCI calculated from single hand-net samples each value has an error of around ±11% (Stark 1998). On the other hand, a trend line based on a fitted linear regression implies that the trend was constant – in this case increasing – for the entire sampling period. This may or may not be the case.

A useful technique for visualising the trend in a data series like this is to fit a LOWESS (**LO**cally, **WE**ighted **S**catterplot **S**moothing) line to the data. LOWESS describes the relationship between Y and X without assuming linearity or normality of residuals. It is a robust description of data pattern. Changes to the smoothing coefficient (aka stiffness or tension) result in different smooth patterns, so the fit can be altered as desired. The tension ranges from 0 to 1 and dictates how closely the fitted line follows the data values (Figure 3). LOWESS with tension of 0 fits the data values perfectly (*i.e.*, a line joining each data point to the next) and at the other extreme a tension of 1.0 produces a near-linear fit to the data points (Figure 3). The tension should be chosen so that the fitted line does not have several local minima or maxima, but not so smooth as to eliminate true changes in slope. From experience, we have found that a tension of 0.4 seems most suitable for depicting trends in biotic indices, although values of 0.3 to 0.6 could be considered equally suitable in most respects. In this report all LOWESS fits have used a tension of 0.4.


(a) (b)

Figure 2 Scatterplot of MCI vs time for the Huatoki Stream at Hadley Drive (Site HTK000350) with (a) individual data points joined and (b) a fitted linear regression.

For a LOWESS fit to provide a useful way to visualise trends in data, there must be sufficient data points in the time series. For example, if LOWESS is fitted using a tension of 0.4 and there are only six data points in the time series, then there is no smoothing – the fitted line simply joins the data values. With only six data points in the times series the tension must exceed 0.66 for there to be any smoothing at all. Fitting a smooth to such a short data series is a little pointless. It would be preferable to wait until there were at least 10-12 data points (equivalent, for example to five or six years of spring and summer monitoring) before undertaking trends analyses. By contrast, when assessing trends in water quality data, using the seasonal Kendall trend test, a minimum of three years' of monthly sampling (*i.e.*, 36 data points) is required, with five years' data (*i.e.*, 60 data points) preferred.

Figure 4 presents the LOWESS (tension = 0.4) fit to the Huatoki (Site HTK000350) MCI data.

The LOWESS fitted line suggests that MCI values in the Huatoki Stream at Hadley drive showed a slight decrease from late 1996 until early 2000 followed by a marked increase.

It now remains to be determined whether or not the trend is likely to be statistically significant.

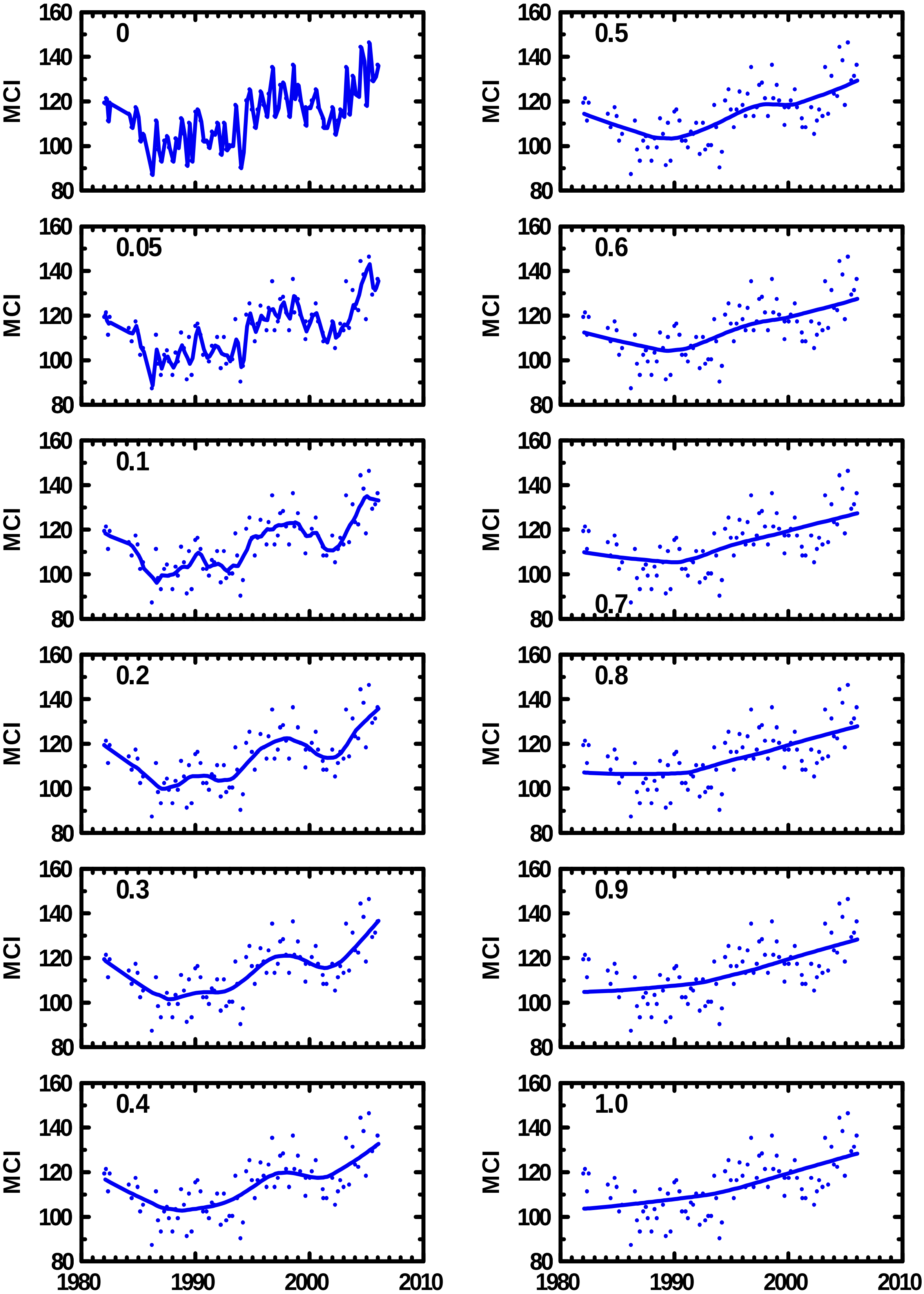


Figure 3 The effect of tension (0 – 1.0) on the fit of the LOWESS smooth.

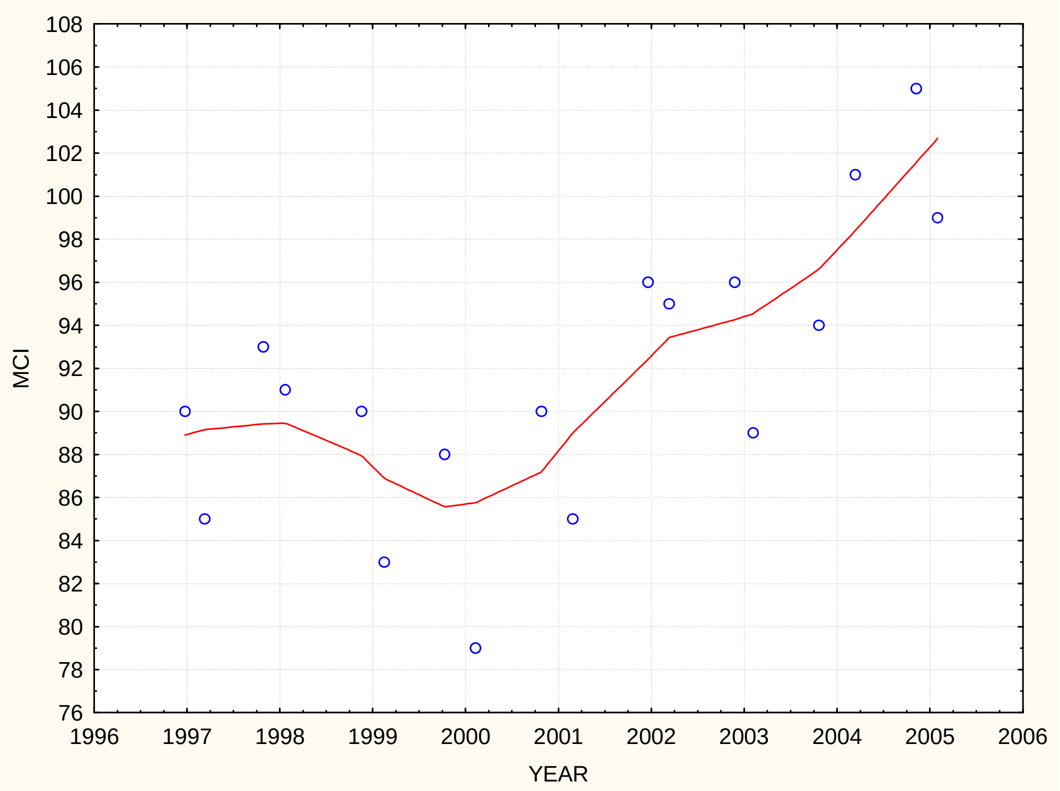


Figure 4 Scatterplot of MCI vs time for the Huatoki Stream at Hadley Drive (Site HTK000350) with a LOWESS (tension = 0.4) fitted line.

4.1 Detection of significant trends

There are a number of approaches that can be used for testing the statistical significance of trends. Five types presented in Table 4 are classified based on two factors. The first, shown in rows of the table, is whether the test is parametric, non-parametric or a mixture of parametric and non-parametric procedures. The second factor (columns) is whether or not there is an attempt to remove variation caused by other associated variables.

Table 4 Approaches to trend testing (Helsel & Hirsch 1992).

	Not adjusted for X	**Adjusted for X**
Non-parametric	Mann-Kendall trend test on Y	Mann-Kendall trend test on residuals R from LOWESS of Y on X
Mixed	-	Mann-Kendall trend test on residuals R from regression of Y on X
Parametric	Regression of Y on T	Regression of Y on X and T

The table uses the following notation:-

Y = the random response variable of interest in the trend test (*e.g.*, MCI)
X = an exogenous variable expected to affect the value of Y (*e.g.*, season, river flow)
R = the residuals from a regression or LOWESS of Y versus X, and
T = time (often expressed in years).

In this report only simple trend tests (*i.e.*, not adjusted for X) are discussed because they are appropriate for the task of detecting trends in biological data. More sophisticated tests are available (*e.g.*, seasonal Kendall test), which are computationally more demanding, but don't necessarily provide a much better answer to the simple question: "Have conditions improved, deteriorated or stayed the same?"

4.1.1 *Mann-Kendall trend test*

Testing the significance of Kendall's tau where the X variable is time was first suggested as a test for trends by Mann (1945). This non-parametric test is directly analogous to regression, where the test for significance of the correlation coefficient r is also the significance test for a simple linear regression. The Mann-Kendall test effectively is a test for whether Y values (MCI etc.) tend to increase or decrease with T (time). No assumption of normality is required but there must be no serial correlation in the Y values for the resulting p-values to be valid. This should be a fairly safe assumption for biotic indices collected on a seasonal (or less frequent) basis.

Running the Mann-Kendall test on the MCI data from Huatoki Stream at Hadley Drive (Site HTK000350) produces the following results using STATISTICA 7.1 (Table 5).

Table 5 Mann-Kendall test results for MCI vs time Huatoki Stream at Hadley Drive (Site HTK000350).

	N	**Kendall**	**Z**	**p-level**	**p-exact**
YEAR & MCI	18	0.425308	2.464782	0.013710	----

This test suggests that the overall trend in MCI from late 1996 through to early 2005 at the Hadley Drive site on the Huatoki Stream is likely to be statistically significant at the 5% level (*i.e.*, P <0.05), although not significant at the 1% level (because in Table 5 P > 0.01). STATISTICA output automatically highlights significant results in red. Examination of the graph (Figure 4) and the sign of the Kendall tau value (*i.e.*, +0.425308: Table 5) indicates that the significant trend is positive. In other words, between 1996 and 2005, MCI has increased or stream health has improved overall in the Huatoki Stream at Hadley Drive.

4.1.2 *Seasonal Kendall test*

A comment on the seasonal Kendall test is warranted here. If seasonal cycles are present in the data, then tests for trend that remove these cycles, or are not affected by them, should be used (Gilbert 1987). Like the Mann-Kendall test, the non-parametric seasonal Kendall test can be used even though there are missing, tied, censored or non-determined values and the data do not have to be normally distributed.

The seasonal Kendall test was first proposed by Hirsch et al. (1982) for use with monthly data and three years of data (36 data values) are the minimum requirement. The test can be used for other "seasons" (*e.g.*, annual or quarterly sampling) but the computed Z statistic cannot then be compared with standard normal tables and the degree of approximation when there are fewer than 12 "seasons" has not be given in the literature. Hirsch et al. (1982) provided a computerised procedure for obtaining the exact distribution of the Kendall test statistic for any combination of seasons and years.

However, given that around 36 data points are the minimum requirement for applying the seasonal Kendall test, and few biotic data sets contain that many, opportunities to use this test are limited. Since there are simpler alternatives that perform almost as well, the seasonal Kendall test is not discussed further here, suffice to say that the seasonal Kendall test could be worth considering when sufficient data have been collected (*i.e.*, 36, or more, data points).

4.1.3 Linear regression

The simple linear regression of Y (*e.g.*, MCI) on T (time) is essentially a trend test. The null hypothesis is that the slope is zero (*i.e.*, there is no significant trend or increase/decrease of Y with time), so a statistically significant linear regression (*i.e.*, $P < 0.05$ or $P < 0.01$) rejects the null hypothesis and indicates that there is a significant (+ve or –ve) slope (or trend).

Regression, which is a parametric procedure, makes stronger assumptions about the distribution of Y over time than does Mann-Kendall. Strictly speaking, normality of residuals, constancy of variance and linearity of relationship should all be checked using residuals plots prior to regression analysis. If Y is not linear over time a transformation may be required. Testing data to ensure that these assumptions are met can become tedious especially when many trends need to be examined.

Returning to the Huatoki Stream (Site HTK000350) data, the scatterplot (Figure 1) suggests that the relationship between MCI and time may not be linear over the entire sampling period. One can imagine, for example, a period from 1996 to 2000 when MCI values are either stable or decreasing, and a period from 2000 to 2005 of marked increase in MCI over time. This, in fact, is what the LOWESS plot shows (Figure 4). In this case, therefore, it may be best not to use the linear regression test for trend significant testing, or, if we do, to transform the data first.

However, to keep matters simple, let's assume that it is acceptable to apply the linear regression test. We now need to test the residuals for normality. This can be done in the Multiple Regression module of STATISTICA because it saves the residuals and predicted values for subsequent examination.

Below we see that the linear regression is statistically significant ($P = 0.003234$).

```
 Multiple Regression Results

Dependent: MCI              Multiple R =  .65405685     F = 11.96178
                                     R²=  .42779037    df =   1,16
No. of cases: 18            adjusted R²=  .39202727     p =  .003234
             Standard error of estimate: 5.103476509
Intercept: -3135.955601  Std.Error: 933.2064  t(   16) = -3.360  p =  .0040

         YEAR beta=.654

(significant betas are highlighted)
```

We should then plot the residuals versus the predicted (Figure 5) to look for two possible problems: curvature and heteroscedasticity. The ideal residuals plot should be symmetrical about the horizontal (Y = 0) line with data points equidistant above and below the line throughout the range of predicted values. Heteroscedasticity occurs when the variance alters along the range of predicted values. In other words, the data points are closer to the zero line for some predictions than others. Curvature is evident when the best fit line on the residuals vs predicted plot is not the straight line at Y = zero, but rather a curved line of some form or other. Heteroscedasticity and curvature can often (but not always) be fixed by transformations or by using more robust procedures (see Helsel & Hirsch 1992) which are not discussed here. Alternatively, a simpler option is to revert to the non-parametric Mann-Kendall approach described above.

In the present case (Site HTK000350) the residuals plot (Figure 5) suggests that there could be a U-shaped curvature (from +ve to –ve to +ve) across the range of predicted values, and also some heteroscedasticity with lower variance towards the extremes of predicted values than nearer the middle. While we are cautioned not to read too much into these plots (Helsel & Hirsch 1992), there do not appear to be any robust guidelines or objective methods for interpreting them. We consider this to be somewhat unsatisfactory, especially, if trends testing is to be undertaken routinely by ecologists with limited statistical experience.

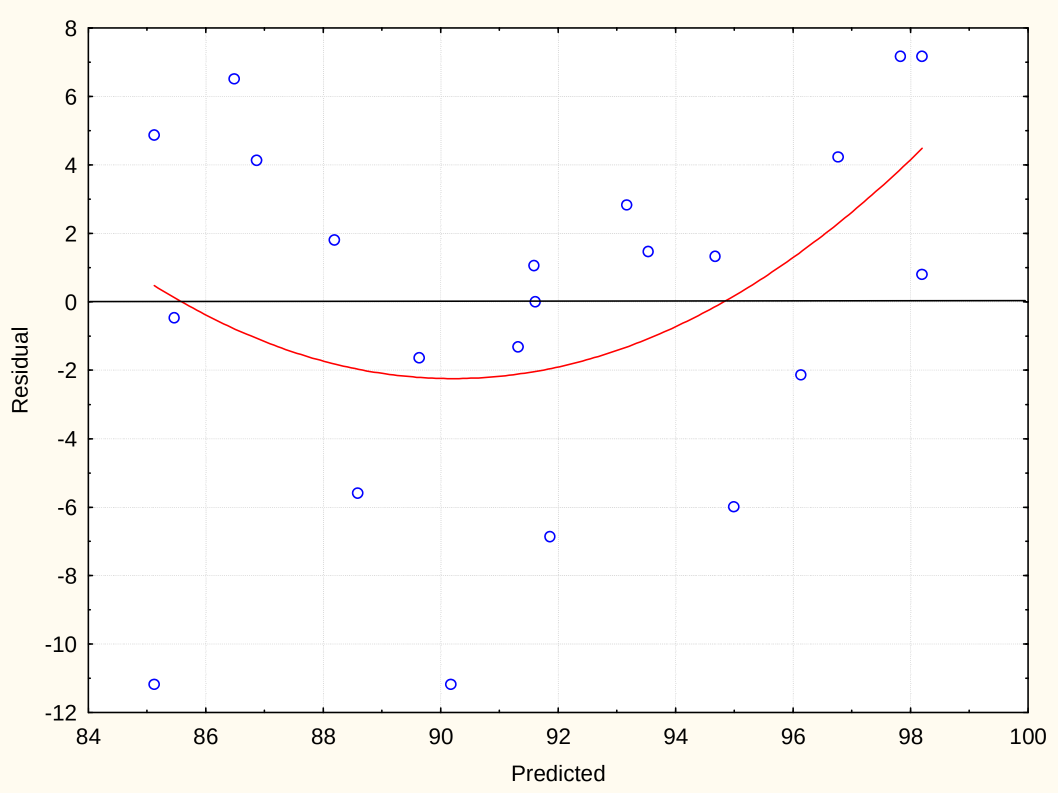


Figure 5 Residuals plot for MCI vs time for the Huatoki Stream at Hadley Drive (Site HTK000350) with a LOWESS (tension = 0.4) fitted line.

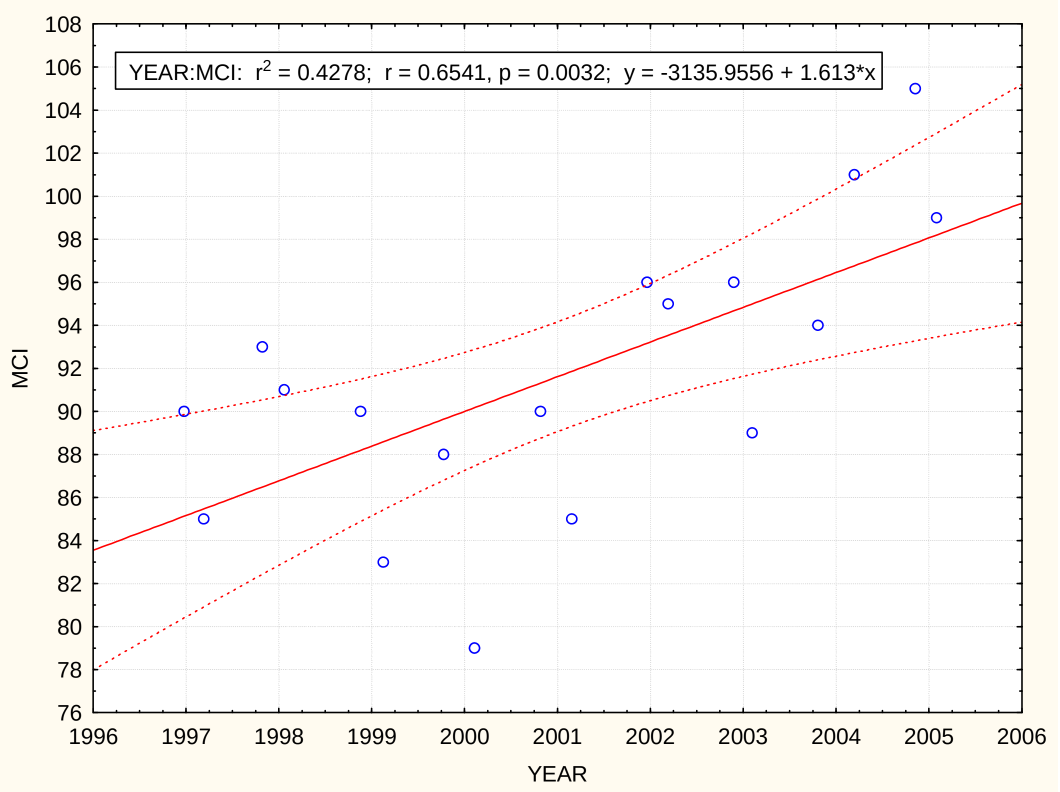


Figure 6 Linear regression (with 95% confidence limits about the slope) for MCI vs time for the Huatoki Stream at Hadley Drive (Site HTK000350).

The LOWESS plot (Figure 4) depicts the trend in a realistic visual manner, and the residuals plot (Figure 5) also suggests that a simple linear relationship may not be the best fit for these data and/ or the variance may not be constant across the range of data values. If we ignore these warnings and apply the linear regression anyway we see that there is a significant positive trend in MCI over time at the Huatoki Stream (Hadley Drive). The p-level (P = 0.0032) (Figure 5), which may be misleading due to the assumptions that have been violated, suggests that the relationship is someone stronger than that indicated by the Mann-Kendall test (P = 0.0137) (Table 5). Both the Mann-Kendall and linear regression tests, however, indicate a significant positive trend in MCI at this site at the 5% level, although the linear regression method suggests that the trend is significant at the 1% level also. In the event that we need a p-value to indicate the strength of the trend, the value from the Mann-Kendall test is likely to be more reliable.

4.1.4 Equivalence testing

Some have argued that the classical null hypothesis (which in the case of trends analysis means that there is no trend or that the slope of any trend line is zero) is inappropriate if the intent is to prove that there is no trend present (Dixon 1998). If there is no trend, then one expects a statistical test to accept the null hypothesis, but accepting the null hypothesis (that there is no trend) does not prove that there is no trend. The null hypothesis may be accepted because the slope is near zero (so the trend is only very slight), because the number of data points in the time series is too small, or because there is a large amount of random variability in the data. Equivalence tests have been proposed as an alternative to classical tests of significance to overcome these problems. In terms of trends analysis, the hypothesis then becomes the non-equivalence hypothesis – that the true slope of the trend line lies outside some equivalence region. The upper and lower bounds of the equivalence region normally are chosen *a priori* to reflect the difference that is considered biologically or ecologically equivalent to zero.

Using equivalence tests for trends testing instead of parametric linear regression models is comparatively straightforward. The test determines whether the slope (or trend) in a regression model is equivalent to zero (*cf.*, the regression model which assumes that it equals zero). This can be done by defining a range of slopes that are considered equivalent to zero and doing an equivalence test for the slope (Dixon 1998). However, parametric equivalence testing still requires checking for normality of residuals, constancy of variance and linearity of relationships using residuals plots prior to analysis. If Y is not linear over time a transformation may be required. Testing data to ensure that these assumptions are met can become tedious especially when many trends need to be examined.

Unlike the classical non-parametric tests such as Mann-Kendall, non-parametric equivalence testing is problematical. Non-parametric tests rely on randomisation, which is easy to do when the hypothesis is the usual one of no difference, but much harder with equivalence hypotheses. At present, possible ways to deal with this problem are being developed (Dixon 1998).

4.1.5 Spearman rank correlations

Collier & Kelly (2006) determined the significance of temporal trends over 8-10 years in the Waikato region by examining Spearman rank correlations of years versus a variety of biological indices. Spearman coefficient values were used to define four trend classes using different levels of certainty based upon "professional judgment of ecological significance and a defined level of statistical significance." The False Discovery Rate (McBride 2005) was used to adjust for Type 1 errors when making multiple comparisons to distinguish "clear trends".

Table 6 Trend classes used to define ecological and statistical significance of relationships for different sample sizes (from Collier & Kelly 2006).

	Trend Class			
n	**Stable**	**Possible**	**Probable**	**Clear**
5 – 9	$R_s \leq 0.50$	$0.50 > R_s < 0.70$	$0.70 \geq R_s \leq R_{s(FDR)}$	$R_s > R_{s(FDR)}$
10 – 16	$R_s \leq 0.50$	$0.50 > R_s < R_{s(\alpha=0.05)}$	$R_{s(\alpha=0.05)} \geq R_s \leq R_{s(FDR)}$	$R_s > R_{s(FDR)}$
> 16	$R_s \leq R_{s(FDR)}$	NA	NA	$R_s > R_{s(FDR)}$

R_s = Spearman rank correlation coefficient; FDR – False Discovery Rate; NA = not applicable; n = number of samples.

Collier & Kelly's (2006) approach attempts to add best professional judgment (BPJ) about the ecological significance of detected trends to the results of statistical significance testing (Table 6). Their suggested interpretation (Table 6) also recognises a universal problem with point-null hypothesis testing, which is that the outcome (P-value) is strongly dependent upon the number of samples (n). With small numbers of samples one can get ecologically significant changes that do not give rise to $P < 0.05$, and vice versa – ecologically insignificant changes that turn up as statistically significant when there are many data. Clearly, as ecological monitoring data series lengthen over time, this increasingly will become a problem.

4.2 Benjamini-Hochberg False Discovery Rate (FDR)

A statistical problem arises when undertaking multiple comparisons. Put simply, when multiple correlations are undertaken there is a chance that some will be found to be significant purely by chance. Put another way, testing a hypothesis for each correlation at a level α (say 0.05) will inflate the overall Type I error[2] rate (McBride 2005). The most recent and most powerful method for dealing with this problem was advocated by Benjamini & Hochberg (1995). The False Discovery Rate (FDR) is defined as the expected proportion of true hypotheses rejected out of the total number of rejections. It considers how many of the α-level rejections may be in error. McBride (2005) describes the method and it can be applied using a computer programme written by Ian Jowett (NIWA, Hamilton) or on an Excel Spreadsheet available from the first author of this report or the following URL.

http://www2.tltc.ttu.edu/Westfall/images/6348/p-vals.xls

The overall effect of applying the Benjamini-Hochberg FDR method to a table of multiple correlations is that FDR is controlled at the α level. This means that the expected proportion of rejections that are in error is less than α. Thus the number of correlations deemed to be statistically significant is reduced.

4.3 Comparing the relative strength of trends

When the number of data values in the time series is similar or when there are 16 or more data values, then the P-values are likely to provide a reliable means for ranking the strength of trends. This is the case for most of the Taranaki data we have analysed: 77% of the sites have 16 or more data values in the time series. We recommend working down the list from the strongest significant trend until PBJ suggests that the trend may not have ecological significance or cannot easily be explained. In our view it is preferable to identify a smaller number of trends with confidence than to assemble a long list that includes many that are dubious or unexplainable. If we can’t explain trends that are marginally significant (in a statistical sense) now, how do we explain them in later years when additional data are added and they become non-significant? Besides, weak trends are unlikely to be ecologically significant.

5. A PRACTICAL STRATEGY FOR TRENDS TESTING

Before testing for the statistical significance of trends it is recommended that biological indices and plotted against time with a LOWESS fit (tension = 0.4) to enable the trend to be visualised easily. Of the simple methods for testing the statistical significance of trends, the linear regression method is most powerful and provides the lowest error variance for the slope. However, strictly this method should be applied only when it is known that the relationship between the variable of interest and time is linear, and the residuals are truly normal. In the real-world situation we can never know this for sure.

In situations where the assumptions for applying the test based on linear regression are not met (or we don’t consider the time it would take to test them would be well-spent), the Mann-Kendall procedure will perform either as well or better (Helsel & Hirsch 1992). Furthermore, if many analyses are required to be performed, to the extent that

---

[2] A Type I error is made if we reject a hypothesis when it is true and the risk of this happening is the probability level (normally 0.05 or 0.01).

case-by-case checking of the assumptions of linearity and normality of residuals could become tedious, the non-parametric Mann-Kendall procedure is an appealing and more cost-effective alternative to the parametric regression method. The Mann-Kendall test is nearly always almost as powerful as regression, and the failure to edit out or correctly transform a small percentage of outlying data will not have a substantial effect on the results. The sign (positive or negative) of the Kendall tau statistic indicates the direction of the trend and the P-value provides an indication of the relative strength of the trends when there are similar numbers of samples in each test[3]. In using BPJ to determine whether or not trends are likely to be ecologically significant one should start with the strongest trend and work down until there is doubt concerning the explanation or ecological significance.

Alternatively, the approach used by Collier & Kelly (2006) based upon Spearman rank correlations could be used. This approach denotes “clear” trends as those that remain significant after the FDR method has been applied. At this stage, the results are very similar to the alternative Mann-Kendall method. In addition, however, Collier & Kelly’s (2006) approach identifies less-strong trends that are considered to be “probable” or “possible” trends. They use the term “stable” rather than NS (not significant) when a significant trend has not been detected.

Collier & Kelly’s (2006) method is somewhat more complex that the Mann-Kendall method we advocate because they aim also to detect “possible” and “probable” trends. However, as the criteria in Table 6 indicate, the added complexity applies only when the number of data in the time series are 16 or fewer. Their criteria (Table 6) aim “to define four trend classes using different levels of certainty based on professional judgement of ecological significance and a defined level of statistical significance.” For N > 16, there is no difference in interpretation between Collier & Kelly’s (2006) approach and ours’.

We are not convinced that this added complexity of interpretation is necessary or that it is desirable to add to the list of significant trends that remain after application of the FDR, even if they are described as “probable” or “possible” to denote less certainty. Neither are we convinced that trends testing should be undertaken at all when n is low (say < 10), although it is possible and we appreciate that, at times, water managers may desire it.

We believe that a conservative approach is warranted. In other words, one should err on the side of identifying too few significant trends than too many. Based upon our examination of the Taranaki data, we believe that the true number of ecologically significant trends is likely to be less (rather than more) than the number remaining after applying the FDR method following statistical testing at the 5% level. In fact, given that most sites in the Taranaki SEM macroinvertebrate data series now have more than 16 data values in the time series (and will only increase in future), this discussion is somewhat academic - when there are more than 16 data values in the time series, the statistical test indicates either a clear trend or a non-significant trend (Table 6).

---

[3] In cases where there are disparate numbers of data values in the time series (*e.g.*, when one site has been sampled twice per year for 10 years and another on a seasonal basis for a longer period), a subset of data from the latter site (including only those data collected in the same years and seasons as the former), could be selected to make the data more comparable. This would enable the resulting P-values to be used as an indication of the relative strengths of the trends.

5.1 Trends in long-term biological data from Taranaki streams

5.1.1 Mann-Kendall approach

Appendix 1 presents LOWESS plots and Mann-Kendall tests of significance for trends in taxa richness, MCI, SQMCI, and $\%EPT_{taxa}$ for 51 SEM sites and nine other sites in Taranaki rivers.

In hypothesis testing, the significance level is the criterion used for rejecting the null hypothesis. The significance level is used in hypothesis testing as follows: First, the difference between the results of the experiment and the null hypothesis is determined. In this case, the null hypothesis is that there is no trend. Then, assuming the null hypothesis is true, the probability of a difference that is as large or larger is computed. Finally, this probability is compared to the significance level. If the probability is less than or equal to the significance level, then the null hypothesis is rejected and the outcome is said to be statistically significant. Traditionally, either the 5% level ($P < 0.05$) or the 1% level ($P < 0.01$) has been used, although the choice of levels is largely subjective. The lower the significance level, the more the data must diverge from the null hypothesis to be significant and fewer significant results will be returned. Therefore, the 1% level is more conservative than the 5% level.

After the Benjamini-Hochberg FDR analysis has been applied to the table of multiple correlations, the number remaining significant is reduced irrespective of whether the 5% or 1% level was used. Determining whether or not the 5% or 1% level, followed by FDR analysis, is most appropriate is an issue that we believe is best left to professional judgment, bearing in mind how many trends are detected and whether or not they can be explained.

Table 7 Summary of Mann-Kendall test results for taxa richness, MCI, SQMCI, and %EPT$_{taxa}$ vs time for 60 Taranaki rivers. See Tables 1 and 2 for key to site codes and locations. NS = trend not statistically significant (P > Benjamini-Hochberg FDR of 0.05), -ve = negative trend, +ve = positive trend. * denotes significant trends with FDR of 0.01.

Site	Taxa richness	MCI	SQMCI	%EPT$_{taxa}$
HTK000350	NS	NS	NS	+ve
HTK000425	NS	+ve	+ve	NS
HTK000745	+ve	NS	NS	NS
INH000400	NS	NS	NS	NS
KPA000250	NS	NS	NS	NS
KPA000700	NS	+ve	NS	+ve
KPA000950	NS	NS	NS	NS
KPK000250	NS	NS	NS	NS
KPK000500	NS	NS	NS	NS
KPK000660	NS	+ve	NS	NS
KPK000880	NS	NS	NS	NS
KPK000990	NS	NS	+ve	NS
KPN000275	-ve	NS	+ve	NS
KPN000360	NS	NS	+ve	NS
KRP000300	NS	+ve	NS	NS
KRP000660	+ve*	+ve*	NS	+ve*
KTK000150	NS	NS	NS	NS
KTK000248	NS	+ve	+ve	NS
MGE000970	NS	NS	NS	NS
MGH000950	+ve	+ve*	+ve	NS
MGN000195	NS	NS	NS	NS
MGN000427	NS	NS	NS	NS
MGO000050	+ve	NS	NS	NS
MGO000190	NS	NS	NS	NS
MGT000488	NS	NS	NS	NS
MGT000520	+ve*	NS	NS	NS
MHW000060	NS	NS	NS	NS
MKW000200	NS	NS	NS	NS
MKW000300	NS	NS	NS	NS
MRK000420	NS	+ve	NS	+ve
MWH000380	NS	NS	NS	NS
MWH000490	NS	NS	NS	NS
PAT000200	NS	NS	NS	NS
PAT000315	NS	NS	NS	NS
PAT000360	NS	NS	NS	NS
PNH000200	NS	NS	NS	NS
PNH000900	NS	NS	NS	NS
STY000300	NS	NS	NS	NS
STY000400	NS	NS	NS	NS
TMR000150	NS	NS	NS	NS
TMR000375	+ve	+ve	NS	+ve
WAA000200	NS	NS	NS	NS
WAA000447	NS	NS	NS	NS
WAI000110	NS	NS	NS	NS
WGA000260	NS	NS	NS	+ve
WGA000450	+ve	+ve*	NS	+ve*
WGG000115	NS	NS	NS	+ve
WGG000150	NS	NS	NS	NS
WGG000500	NS	NS	NS	NS
WGG000540	NS	NS	NS	NS
WGG000665	NS	NS	NS	NS
WGG000895	NS	NS	NS	NS
WGG000995	+ve	NS	NS	NS
WKH000100	-ve	NS	NS	NS
WKH000500	NS	NS	NS	NS
WKH000920	NS	NS	NS	+ve
WKH000950	NS	NS	NS	+ve
WMK000100	NS	NS	NS	NS
WMK000298	NS	NS	NS	NS
WTR000850	+ve	+ve	NS	+ve*
No. of +ve trends @ 5%FDR level (1%)	9 (2)	11 (3)	6 (0)	11 (3)
No. of NS trends @ 5%FDR level (1%)	49 (58)	49 (58)	54 (60)	49 (57)
No. of –ve trends @ 5%FDR level (1%)	2 (0)	0 (0)	0 (0)	0 (0)

Table 7 summarises the Mann-Kendall test results for each site with statistically significant results at the 5% level. All correlations are statistically significant at cut-off P values corresponding to a False Discovery Rate (FDR) of 0.05 (Benjamini & Hochberg 1995). Asterisks on Table 7 indicate trends that remain significant at a FDR of 0.01.

Twenty-four sites (with bold site codes in column 1 of Table 7) exhibited statistically significant trends at the 5% level for one or more of the biological indices tested. This reduced to only five sites (KRP000660, MGH000950, MGT000520, WGA000450, WTR000850) when the 1% level was used.

Eleven sites showed trends in taxa richness (9 positive and 2 negative) at the 5% level with only two positive trends at the 1% level (Table 7).

Trends in taxa richness should be interpreted with caution because kick-samples are not collected from a precisely defined area of streambed and taxa richness can be highly variable depending not only upon the effects of flow variability but also variations in sample size resulting from different personnel collecting the samples. Variability in taxa richness can also be introduced if different personnel with variable expertise are used for sample processing (although that is unlikely to be a problem with the TRC data because all processing has been done by the same personnel).

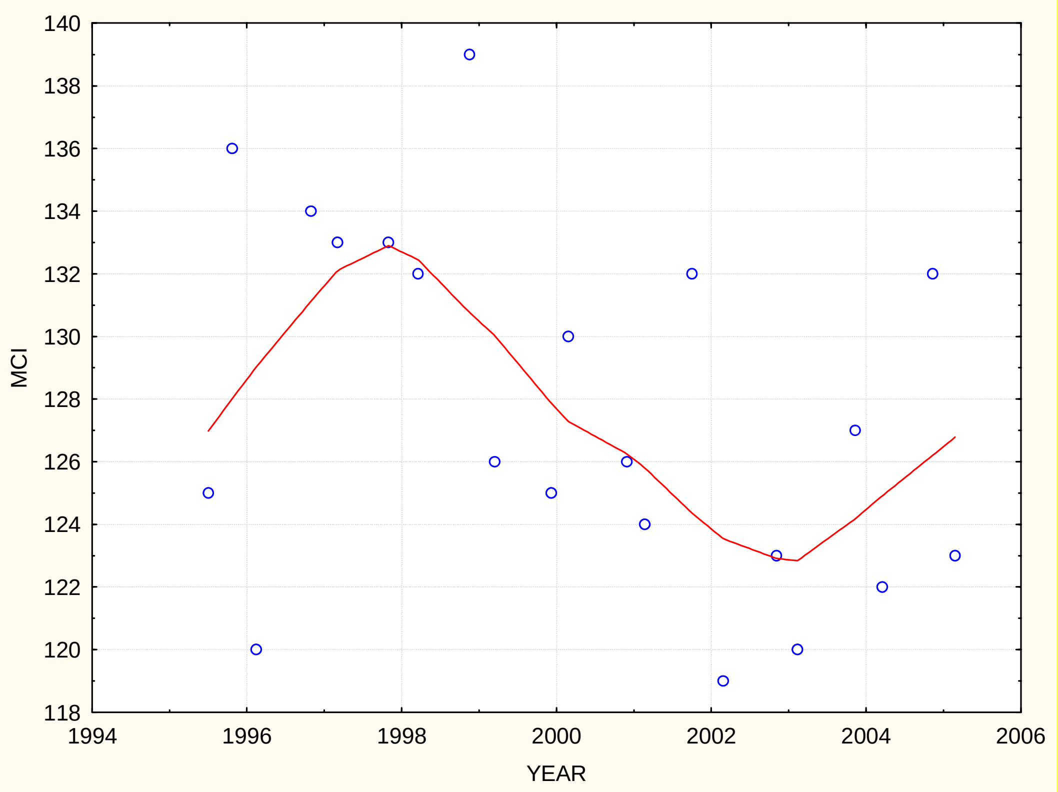


Figure 7 LOWESS trend for MCI vs time for the Waingongoro River at Opunake Road (Site WGG000150).

Eleven sites showed significant trends in MCI at the 5% level with all of these positive (Table 7). Interestingly, a significant (at the 5% level prior to FDR analysis) negative trend in MCI was identified for the Waingongoro River at Opunake Road (Site WGG000150) (Figure 7 & Appendix 1). The LOWESS fit suggests that MCI at this site increased between 1995 and 1998, decreased for the next eight years, and has been increasing again since. Note that the trend is not especially strong ($P = 0.027$) and that the range of MCI on the LOWESS line from the high in 1998 (133) to the low in 2003 (123) is only about 10 MCI units, which is within the "error" of ±10.83 MCI units associated with estimating an MCI from a single hand-net sample (Stark 1998). Given the above, and the fact that most spot MCI values recorded from this site have exceeded 120, indicating excellent stream health, it is unlikely that the trend in MCI detected here should be of any concern to water managers. This trend was never likely to have any ecological significance, so it is comforting to note that it was deemed non-significant following FDR analysis.

Positive trends at three sites (MWH000490, WGG000115, & WGG000150), although significant at the 5% level prior to FDR analysis (Appendix 1), were within the "error" associated with estimating MCI values from single kick samples. In each of these cases, the range in MCI from the minimum point on the LOWESS fit to the maximum was less than 10.8 MCI units. This suggests that these trends *could* be an artefact of the data and *may* not have much ecological significance, suggesting, perhaps, that a more conservative approach might be warranted to ensure than any trends detected are likely to be "real". Once again, these trends were deemed non-significant following FDR analysis.

Application of the FDR procedure to statistical testing undertaken at the 5% level seems to eliminate trends that are unlikely to have any ecological significance. However, if a more conservative approach is desired, then the P-level for assessing significance can be set at the 1% level.

Only three sites (KRP000660, MGH000950, WGA000450) showed significant trends in MCI at the 1% level (Table 7). All of these trends were positive. This compares with 11 positive significant trends at the 5% level.

It is of interest to note that all 12 "reference" sites (bold site codes in Table 1) situated either within, or a short distance downstream of the National Park boundary did not show statistically significant trends in MCI values at the 5% level after FDR analysis. The lack of trends in MCI at reference sites is an expected result when, as is the case here, there are no known reason/s why an improvement or deterioration in stream condition should be expected. One reference site (WGG000115, see Appendix 1) did, however, show a significant (positive) trend in MCI score at the 5% level, but was eliminated by FDR analysis, providing additional support for a conservative approach.

Six significant positive trends in SQMCI were detected at the 5% level with FDR analysis (Table 7). There were no significant trends in SQMCI when significance was assessed at the 1% level with FDR analysis.

There were 11 significant positive trends at the 5% level (after FDR) in $\%EPT_{taxa}$, reducing to 3 at the 1% level with FDR analysis (Table 7).

Applying the FDR analysis makes a considerable difference to the number of trends that remain significant. At the 5% level the number of significant trends reduced from 75 to 39 (including two negative trends) and at the 1% level FDR causes a reduction from 43 to eight (Appendix 1 *cf*. Table 7). Overall at the 5% level following FDR analysis there were a total of 201 non-significant trends from 240 tests (60 sites x 4 indices), so 16.25%

(39) of the trends were considered significant. At the 1% level following FDR analysis there were 232 non-significant tests leaving 3.3% (8) that were considered statistically significant.

The choice of significance level is somewhat arbitrary - both the 5% and 1% levels are equally valid. However, we believe that it may be prudent to adopt a conservative approach to detecting trends in biological indices. When multiple trends tests are undertaken the FDR correction must be applied and assessing trends at the 5% level seems to produce a reasonable number of significant trends, most of which should also have ecological significance. Using this approach, which is statistically defensible, the trends that are detected will be the strongest ones, and those that are unlikely to have any ecological significance will not be deemed to be statistically significant.

Where testing the significance of a single trend is required we recommend that the significance test should be interpreted at the 1% level. This is because the cutoff p-value obtained by testing at the 5% level followed by FDR analysis is very close to $P = 0.01$. When a single test is undertaken the FDR is not applicable because it can be used only when multiple comparisons have been made.

Irrespective of these statistical guidelines, somewhat less-strong trends may be accepted as significant if, in the BPJ of an experienced macroinvertebrate ecologist, they can be explained or are considered to have ecological significance.

5.2 Spearman rank correlation approach of Collier & Kelly (2006)

Appendix 2 presents the results of trends testing using Collier & Kelly's (2006) method followed by FDR analysis to eliminate significant results that could have arisen by chance because multiple comparisons were done. Appendix 2 is the output from a computer programme (FDR.EXE) written by Ian Jowett (NIWA, Hamilton) that applies the FDR analysis to a "coefficients file" containing (in this case) Spearman rank correlation coefficients calculated from years vs index data. The other input file for the FDR programme contains the number of data values. Appendix 2 provides examples of the data file format for the two input files. They can be separate pages in an Excel spreadsheet.

Table 8 summarises the results of significance testing using the Mann-Kendall procedure described in this report and the Spearman rank correlation-based method described by Collier & Kelly (2006) by ranking the strongest trends that remained significant at the 5% level following FDR analysis.

All of the sites listed in Table 8 showed "clear" temporal trends in the various indices (using the terminology of Collier & Kelly 2006). Bold entries for each index indicate sites where both methods identified clear trends. For taxa richness and the SQMCI both methods ranked the same sites in the top four (although not always in exactly the same order), with the top five and top eight in common for the MCI and %EPT_{taxa} (shaded cells in Table 8). These results suggest that either method does an acceptable job at identifying the strongest trends.

Table 8 Ranking of trends in taxa richness, MCI, SQMCI, and %EPT$_{taxa}$ vs time in Taranaki rivers identified by (a) the Mann-Kendall procedure described in this report and (b) Collier & Kelly's (2006) Spearman rank correlation procedure. See Tables 1 and 2 for key to site codes and locations. Bold site codes indicate sites at which significant trends were detected by both methods.

	Taxa		**MCI**		**SQMCI**		**%EPT$_{taxa}$**	
	(a)	**(b)**	**(a)**	**(b)**	**(a)**	**(b)**	**(a)**	**(b)**
1	**MGT000520**	**KRP000660**	**KRP000660**	**MGH000950**	**KTK000248**	**KTK000248**	**WGA000450**	**KRP000660**
2	**KRP000660**	**MGT000520**	**MGH000950**	**KRP000660**	**HTK000425**	**HTK000425**	**KRP000660**	**WGA000450**
3	**HTK000745**	**TMR000375**	**WGA000450**	**WGA000450**	**KPN000360**	**KPN000360**	**WTR000850**	**WTR000850**
4	**TMR000375**	**HTK000745**	**KPK000660**	**KPK000660**	**KPN000275**	**KPN000275**	**TMR000375**	**KPA000700**
5	WGG000995	**WTR000850**	**KRP000300**	**KRP000300**	**MGH000950**	**KPK000990**	**KPA000700**	**WKH000950**
6	**MGO000050**	**MGO000050**	**MRK000420**	**KTK000248**	**KPK000990**	WAI000110	**WKH000950**	**WKH000920**
7	**KPN000275**	**WGA000450**	KPA000700	**MRK000420**		KPK000880	**WKH000920**	**TMR000375**
8	**WGA000450**	**MGH000950**	**KTK000248**	**TMR000375**		WGG000665	**WGA000260**	**WGA000260**
9	**WTR000850**	**KPN000275**	**TMR000375**	HTK000350		**MGH000950**	MRK000420	
10	WKH000100		HTK000425			WGG000500	WGG000115	
11	**MGH000950**		WTR000850				HTK000350	

The most notable difference between our approach and that of Collier & Kelly (2006) is that the latter also identify "probable" and "possible" trends. These are not as strong as "clear" trends and we considered them to be non-significant.

6. WHICH INDICES ARE BEST SUITED FOR ASSESSING TRENDS IN SEM DATA?

Taxa richness is not strictly a measure of stream condition or stream "health". In fact, highest taxa richness often is associated with slightly enriched streams (*e.g.*, those experiencing diffuse-source nutrient enrichment from farmland) rather than pristine streams in reference condition. Low taxa richness can be associated with quite "sterile" environments with extremely pure water, perhaps where torrential water velocities and lack of nutrients result in low productivity. Such places are naturally unproductive. In other words, there is no valid basis to consider that high taxa richness is "good" and low taxa richness is "bad". Furthermore, estimates of taxa richness are highly dependent on sample size, which, in turn, can be influenced by sampling or processing effort (which can vary markedly with different personnel). For these reasons, use of taxa richness for SEM reporting is problematical and is not recommended.

We do not recommend the SQMCI (or QMCI) for SEM reporting either because community percentage composition (more so than taxonomic composition) can change during the SEM sampling period (which can be several weeks) in response to small freshes or as a flow recession lengthens. Consequently, differences between sites arise as a result of when samples were collected and, as such, are an artefact of the sampling regime rather than a true measure of stream health. The SQMCI and QMCI are more suited to compliance monitoring and synoptic surveys (where all samples are collected on the same day under similar conditions) and not for SEM where samples may be collected over a month or more and yet need to be compared on a common basis.

The MCI and %EPT_{taxa} are both reliable indices for assessing stream health. The MCI is not affected by changes in percentage composition, and, consequently, is affected much less than the SQMCI and QMCI by flow-related or seasonal factors. The MCI is essentially a scaled average score per taxon, so (being an average) is relatively unaffected by sample size (unlike taxa richness). MCI and %EPT_{taxa} are highly correlated (using data from the 60 sites discussed in this report $r^2 = 0.836$), so provide essentially the same assessment of river health. We suggest that both the MCI and %EPT_{taxa} need not be reported in SEM reports for the general public or for Regional Councillors. Using two indices that tell much the same story seems somewhat superfluous, and could be confusing for laypersons, so we recommend that the MCI should be the index of choice for SEM reporting.

7. CAUSES OF TRENDS

When a trend has been identified there are two vital questions that arise (Rutherford 1985). First, "What caused the trend?" and second "Will it continue into the future?" Unfortunately trend analysis alone cannot provide certain answers to these very closely related questions. If the cause/s of the trend can be determined, then one may be able to predict the future with some confidence. Conversely, if one cannot determine why a trend has occurred then extrapolation into the future may be most unwise because there remains a chance that the trend is simply an artefact of natural variability or sampling error.

It is worth emphasising that the methods for trends testing discussed in this report are concerned with determining the significance of trends that have already occurred (*i.e.*, within the period of data collection). Forecasting is quite another matter and the disclaimer often quoted by financial advisors with respect to the performance of a portfolio of investments applies – "Past performance is no guarantee of future performance". In other words, there is no guarantee than a trend observed in existing data will continue into the future. In fact, we can be absolutely sure, for example, that a linear trend of increasing MCI will not continue in perpetuity because it is impossible for the MCI to exceed 200 and, in practical terms, MCI values over 160 are rare. By implication also, the notion that weak negative trends might provide early warning of further environmental degradation is dubious unless one knows the cause and has reason to believe the trend will continue as a result. The statistical testing alone is not sufficient.

Any explanations for trends are likely to require additional information, which could include information such as stream flows, weather patterns, catastrophic erosion events in the catchment, physico-chemical water quality, or changes to land or water management practices that may have resulted from TRC water management initiatives, and industry or farming activities. Discussion of the reasons for trends is beyond the scope of this report.

We re-iterate that if significant positive or negative trends are identified, then it is prudent to know why these trends are occurring, and, preferably, to know that they have occurred due to water management initiatives if they are to be used to indicate that the Council has played an important role in maintaining or improving stream health in the region. There may be other reasons for trends, beyond the influence of water management initiatives.

Furthermore, there is a difference between statistical significance and ecological significance. A trend may be statistically significant but have no ecological significance and *vice versa*. The ultimate decision on whether or not any trend should be considered ecologically significant must depend on the BPJ of an experienced freshwater ecologist. Indications are that the strongest trends (lowest p-values) are most likely to have the greatest ecological significance.

8. SUMMARY AND RECOMMENDATIONS

The following procedure is recommended as a most cost-effective method for examining trends in macroinvertebrate biological data.

1. Visualise the trend: Scatterplots of biological index vs time with LOWESS fit (tension of 0.4).
2. Test for significance using the Mann-Kendall test. When multiple tests are required we recommend the 5% significance level followed by FDR analysis. The trends remaining significant following this procedure should be clear trends. If the number of data values (n) for each site is similar then the p-value resulting from the test will provide a reliable indication of the strength of trends. Given the fact that statistical significance is not the same as ecologically significance, explanations should be sought for these clear trends starting with the one with the lowest p-value and working down the list.
3. We recommend testing for significance of a trend from a single site (as opposed to testing multiple sites) using the 1% level.
4. We do not believe that weaker trends than those identified using the above procedure (termed “possible” or “probable” trends by Collier & Kelly 2006) need to be considered although there is no reason why they should not be included if explanations for them are apparent.

Although this approach is appropriate for trends analysis of a range of biological indices, we recommend that high-level reporting of temporal trends in SEM macroinvertebrate data should focus on the MCI, since this index has a strong track record as an indicator of river health in New Zealand. There are real problems in using indices such as taxa richness and SQMCI or QMCI for reporting SEM results. Taxa richness is not a good measure of stream health (because highest taxa richness often is associated with slightly enriched conditions and lowest taxa richness with pristine water quality with torrential flows) and is influenced markedly by sampling effort. Differences in SQMCI or QMCI can be an artefact of sampling date and $\%EPT_{taxa}$ is highly correlated with MCI so “tells the same story”. Use of other indices, as well as the MCI, for SEM reporting often raises some anomalies that a biologist can explain, but which only serve to confuse laypersons. We believe that the KISS principle should apply! [Technical reports prepared by Council staff on SEM monitoring programmes could, however, include indices in addition to the MCI.]

Determining the reasons for any significant trends detected inevitably will require additional information, which could include information such as stream flows, weather patterns, catastrophic erosion events in the catchment, physico-chemical water quality, or changes to land or water management practices that may have resulted from TRC water management initiatives, and industry or farming activities. Discussion of reasons for trends was beyond the scope of this report. It is worth emphasising that the ultimate decision on whether or not any trend should be considered ecologically significant is reliant on the BPJ of an experienced freshwater ecologist.

We caution against extrapolation of trends based upon the results of statistical testing alone. However, if explanations for trends can be determined and there is evidence to suggest that the cause/s might be ongoing, then there could be good reason to suggest that the trend might continue in the same direction (albeit for an unknown duration and not necessarily at the same rate).

9. ACKNOWLEDGMENTS

The authors wish to thank Peter Nolly (Computer Systems Manager, TRC) for provision of data from TRC's macroinvertebrate database and the various biologists who have contributed to the TRC SEM programme over the ten-year period since its inception. We also thank Graham McBride (NIWA, Hamilton) for his helpful comments on the statistical methods, and Kevin Collier (Environment Waikato) for helpful email dialogue and for providing details of Collier & Kelly's (2006) trend detection method.

10. REFERENCES

Benjamini Y, Hochberg Y 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society B (57)*: 289-300.

Collier, K.J.; Kelly, J. 2006: Patterns and trends in the ecological condition of Waikato streams based on the monitoring of aquatic invertebrates from 1994 to 2005. Environment Waikato Technical Report 2006/04. 28p.

Dixon, P.M. 1998: Assessing effect and no effect with equivalence tests. Chapter 12 p275 - 301 *in* Newman, M.C. & Strojan, C.L. (eds), Risk Assessment: Logic and Measurement. Ann Arbor Press, Michigan.

Gilbert, R.O. 1987: Statistical methods for environmental pollution monitoring. Van Nostrand Reinhold, New York.

Helsel, D.R.; Hirsch, R.M. 1992: Statistical methods in water resources. Studies in Environmental Sciences 49. Elsevier, Amsterdam. 522p.

Hirsch, R.M.; Slack, J.R.; Smith, R.A. 1982: Techniques of trend analysis for monthly water quality data. *Water Resources Research 18(1):* 107-121.

Mann, H.B. 1945: Nonparametric test against trend. *Econometrica 13*: 245-259.

McBride, G.B. 2005: Using statistical methods for water quality management: issues, problems and solutions. John Wiley & Sons Inc., Hoboken, New Jersey, USA. 313p.

Rutherford, J.C. 1985: Some thoughts on trend analysis. Biological monitoring in freshwaters: proceedings of a seminar. p39-53 *in* R. D. Pridmore and A. B. Cooper (eds). Water and Soil Miscellaneous Publication No. 82. National Water and Soil Conservation Authority, Wellington, New Zealand.

Stark, J. D. 1998: SQMCI: a biotic index for freshwater macroinvertebrate coded-abundance data. *New Zealand Journal of Marine and Freshwater Research 32*: 55-66.

Stark, J. D.; Boothroyd, I. K. G; Harding, J. S.; Maxted, J. R.; Scarsbrook, M. R. 2001: Protocols for sampling macroinvertebrates in wadeable streams. New Zealand Macroinvertebrate Working Group Report No. 1. Prepared for the Ministry for the Environment. Sustainable Management Fund Project No. 5103. 57p.

TRC 2004: State of the Environment Regional water quality network of Taranaki Regional Council, macroinvertebrate biological sampling techniques for freshwater rivers and streams. TRC Internal Report.

TRC 2005: Fresh water macroinvertebrate fauna biological monitoring programme annual state of the environment monitoring report. Taranaki Regional Council Technical Report 2005-72. 179p.

Vant, B.; Smith, P. 2004: Trends in river water quality in the Waikato Region, 1987 – 2002. Environment Waikato Technical Report 2004/02.

http://www.ew.govt.nz/publications/technicalreports/documents/TR04-02.pdf

Appendix 1 LOWESS (tension = 0.4) plots of taxa richness, MCI, SQMCI, and % EPT_{taxa} versus time for 60 Taranaki streams and rivers. Results of the Mann-Kendall non-parametric test for assessing the statistical significance of trends are given also.

Results in red type are considered statistically significant at the 1% level, and results in orange type at the 5% level. The sign of the Kendall tau statistic (*i.e.*, ±) indicates whether the trend is positive (increasing with time) or negative (decreasing with time).[4] See Tables 1 and 2 for key to site codes and locations.

The histogram above each LOWESS plot shows the numbers of data point in ten time periods throughout the sampling period. For most sites, sampling has been undertaken for 10 years, so each bar represents 1 year and normally represents two sampling occasions. For example, for site HTK000350 most time periods had two data values (except the fifth and sixth, which had one each).

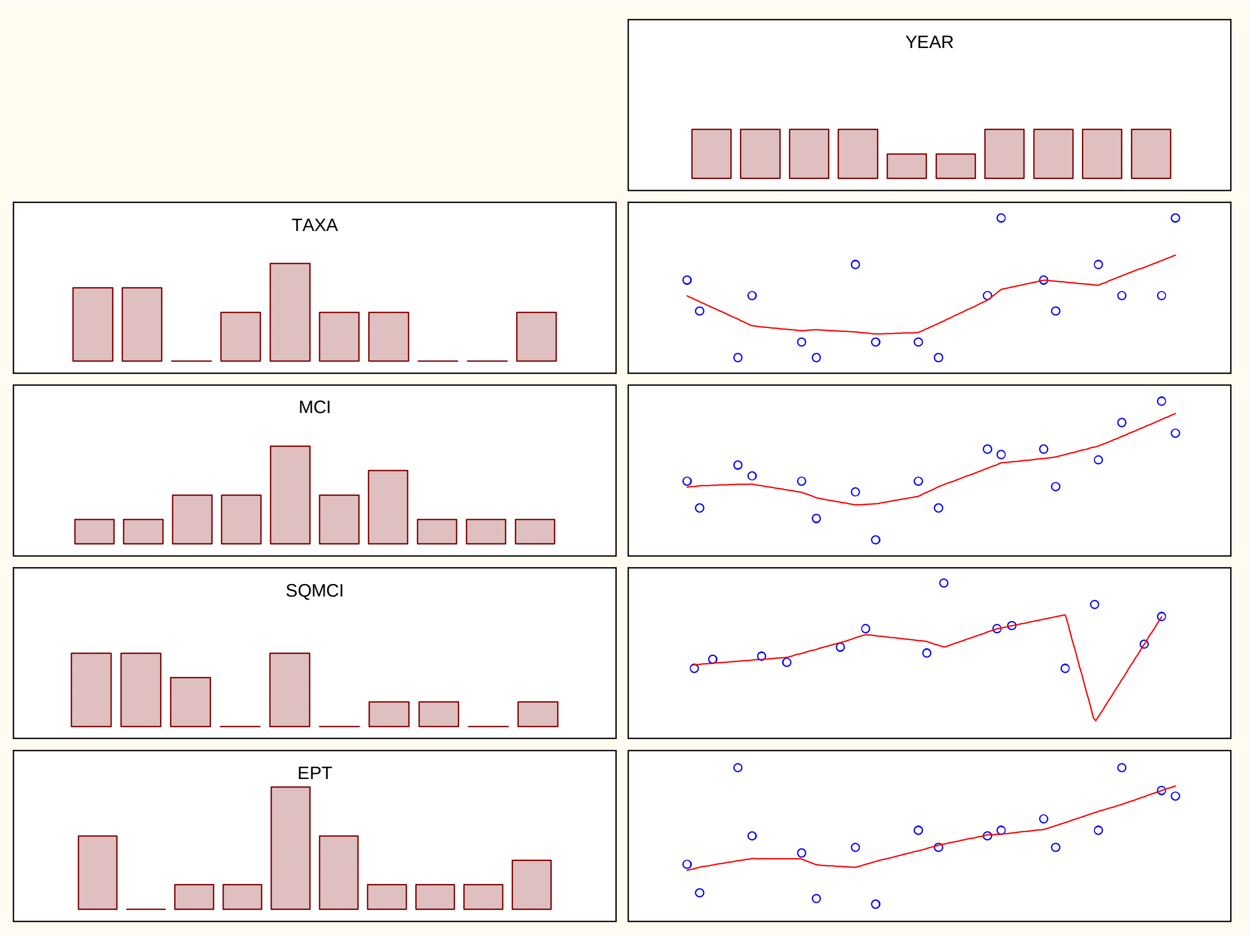
The histograms to the left of the LOWESS plot show the distribution of data values for each of the biological indices, again in ten bins. For example, for site HTK000350 there were 3 taxa richness values in the lowest bin and 2 in the highest. The total of all the bars equals the N in the table below the graph (18 in this case).

There is no way to label scales on the axes of the LOWESS plots produced by STATISTICA using the procedure outline in the footnote below. However, it is a very quick procedure for obtaining an overall picture of trends. The scatterplot procedure should be used if fully labelled graphs are required.

---

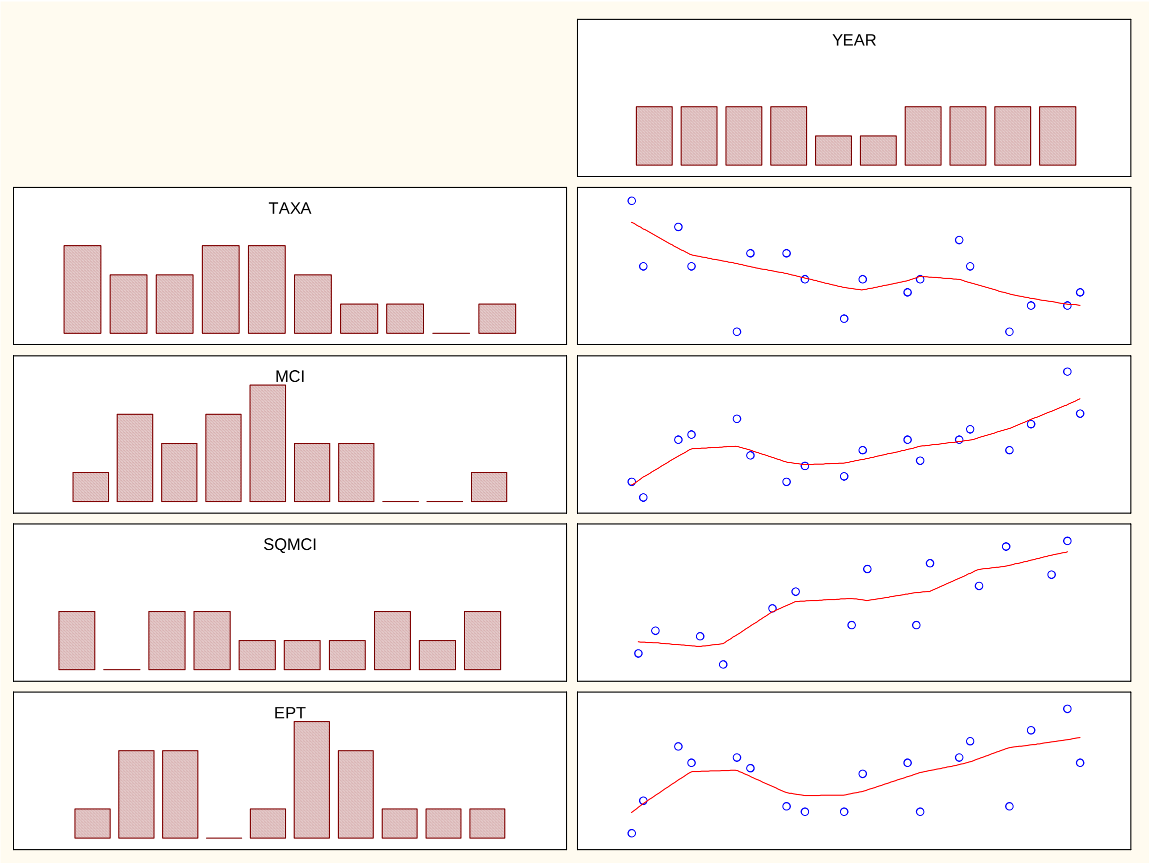
[4] These results were obtained with STATISTICA 7.1 using the Statistics By-Group Analyses, non-parametric correlations with TAXA, MCI, SQMCI and EPT as the first variables and YEAR (expressed as a decimal *e.g.*, 2002.211 = 19-Mar-2002) as the second variable. The grouping- or by-variable was SITE$ (*i.e.*, the site code HTK000350 etc.). The radio button labelled “create Kendall tau” should be checked and “all results” and “detailed results” should be selected on the General tab of the analysis. Once the graph was displayed double-clicking on the fitted linear regression line enables the fit to be changed to LOWESS and click on the ”More” button enables the tension (called stiffness in this module) to be changed from the default of 0.25 to 0.4.

HTK000350



Kendall Tau Correlations MD pairwise deleted Marked correlations are significant at p <.05000					
	N	**Kendall**	**Z**	**p-level**	**p-exact**
TAXA & YEAR	18	0.297004	1.721224	0.085210	----
MCI & YEAR	18	0.425308	2.464782	0.013710	----
SQMCI & YEAR	14	0.477807	2.380333	0.017297	----
EPT & YEAR	18	0.449827	2.606875	0.009137	----

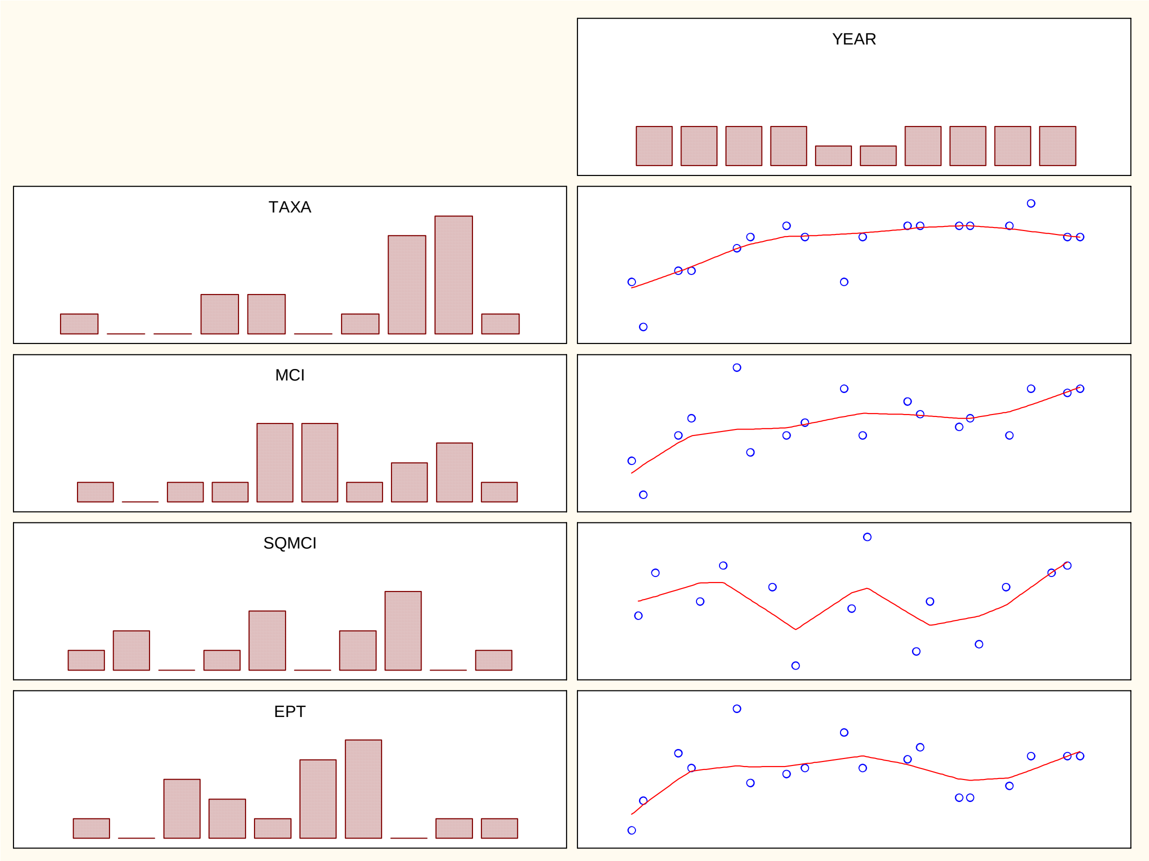
HTK000425



Kendall Tau Correlations MD pairwise deleted Marked correlations are significant at p <.05000

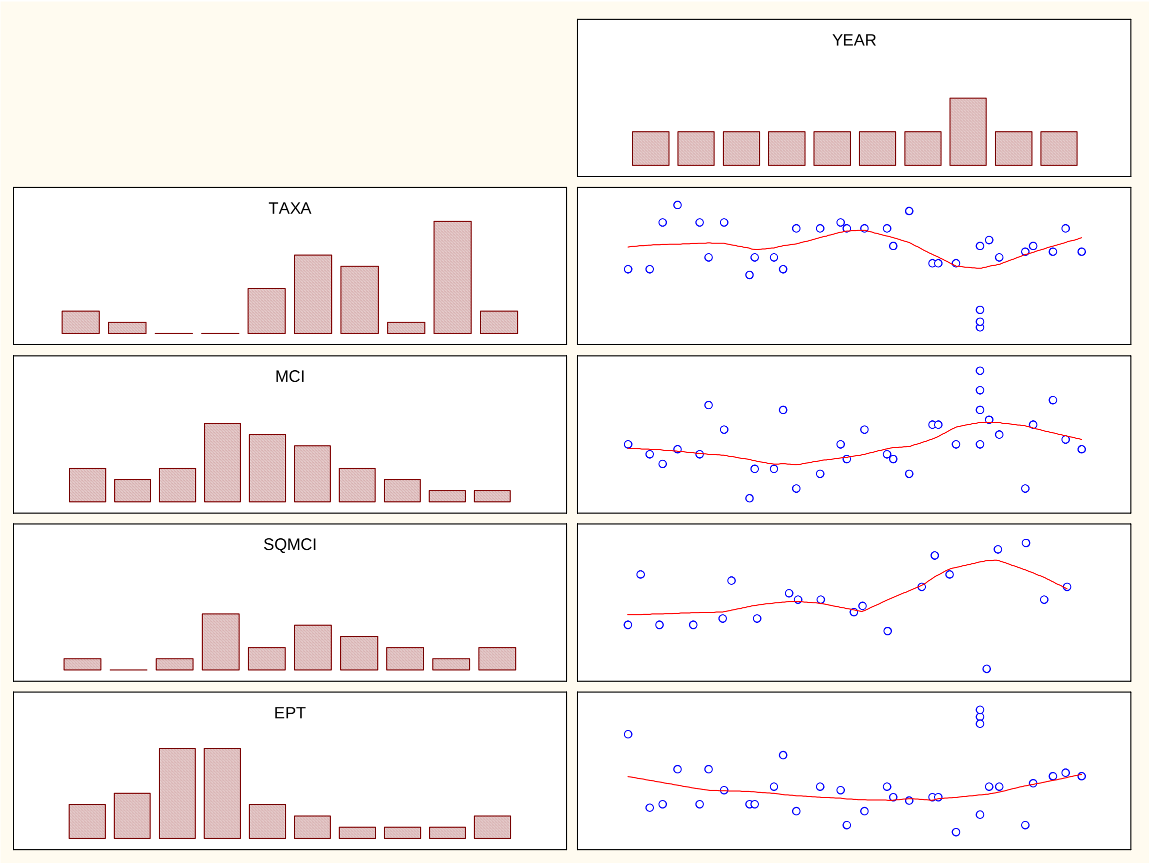
	N	Kendall	Z	p-level	p-exact
TAXA & YEAR	18	-0.385355	-2.23324	0.025533	----
MCI & YEAR	18	0.465180	2.69585	0.007021	----
SQMCI & YEAR	14	0.685093	3.41299	0.000643	----
EPT & YEAR	18	0.275267	1.59525	0.110656	----

HTK000745



Kendall Tau Correlations MD pairwise deleted Marked correlations are significant at p <.05000					
	N	**Kendall**	**Z**	**p-level**	**p-exact**
TAXA & YEAR	18	0.576181	3.339135	0.000840	----
MCI & YEAR	18	0.412398	2.389963	0.016850	----
SQMCI & YEAR	14	0.146104	0.727860	0.466699	----
EPT & YEAR	18	0.173961	1.008153	0.313381	----

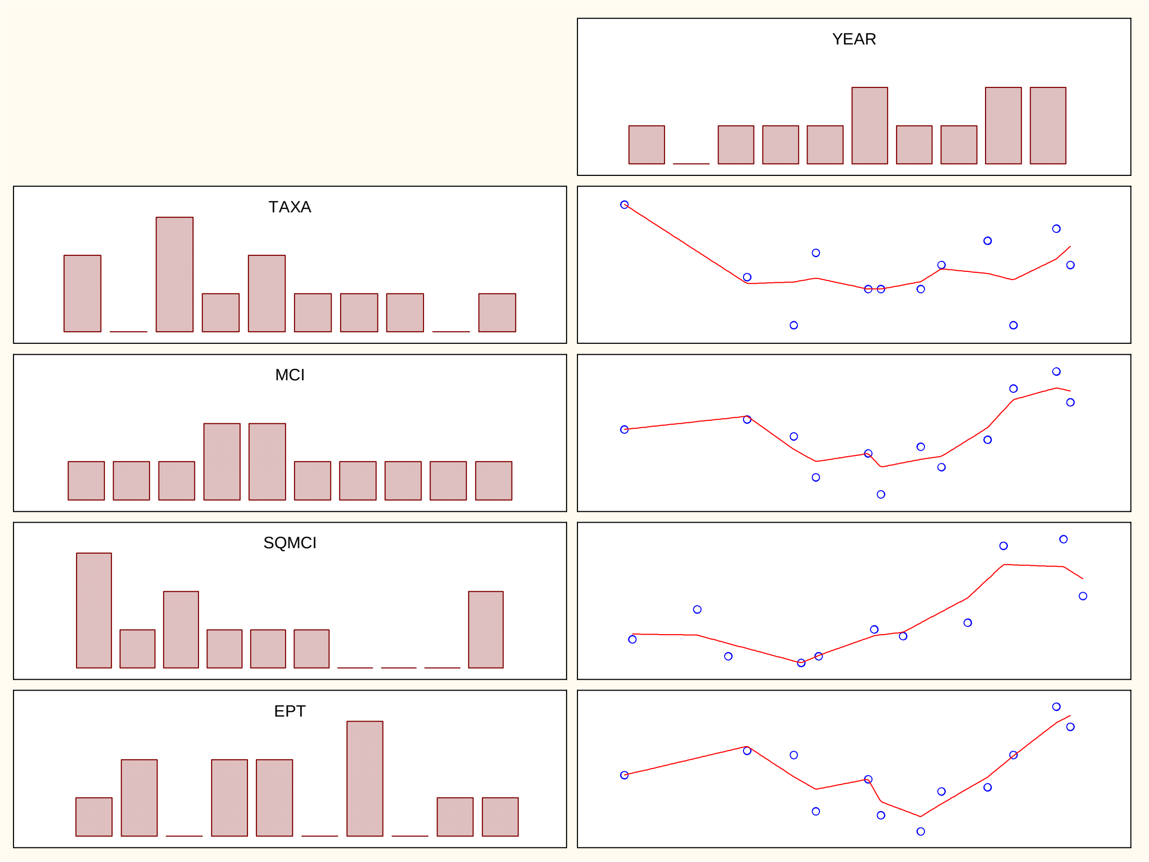
INH000400



Kendall Tau Correlations MD pairwise deleted Marked correlations are significant at p <.05000

	N	Kendall	Z	p-level	p-exact
TAXA & YEAR	33	-0.120737	-0.987755	0.323273	----
MCI & YEAR	33	0.234742	1.920438	0.054803	----
SQMCI & YEAR	21	0.306643	1.944531	0.051831	----
EPT & YEAR	33	0.113078	0.925095	0.354916	----

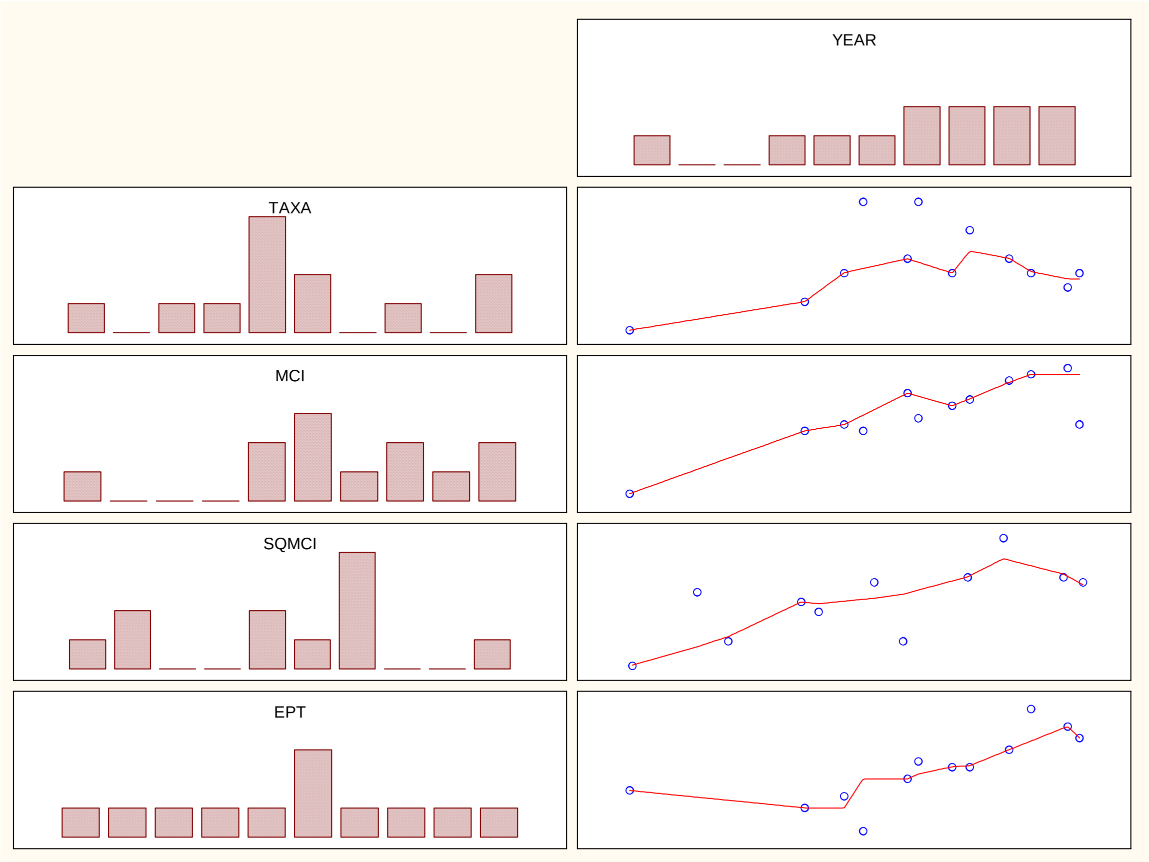
KPA000250



Kendall Tau Correlations MD pairwise deleted Marked correlations are significant at p <.05000

	N	Kendall	Z	p-level	p-exact
TAXA & YEAR	12	0.047281	0.213983	0.830561	----
MCI & YEAR	12	0.212121	0.960016	0.337047	----
SQMCI & YEAR	11	0.513783	2.199888	0.027815	----
EPT & YEAR	12	0.137409	0.621882	0.534019	----

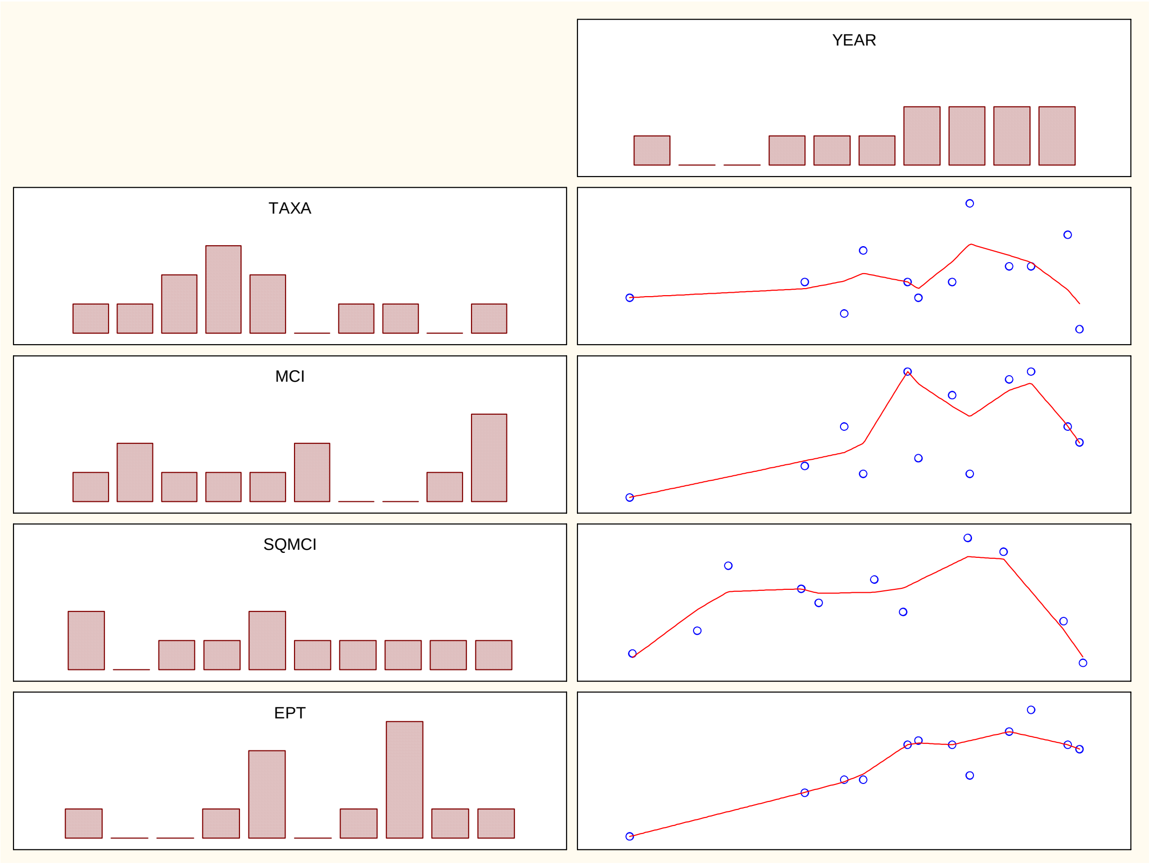
KPA000700



Kendall Tau Correlations MD pairwise deleted Marked correlations are significant at p <.05000

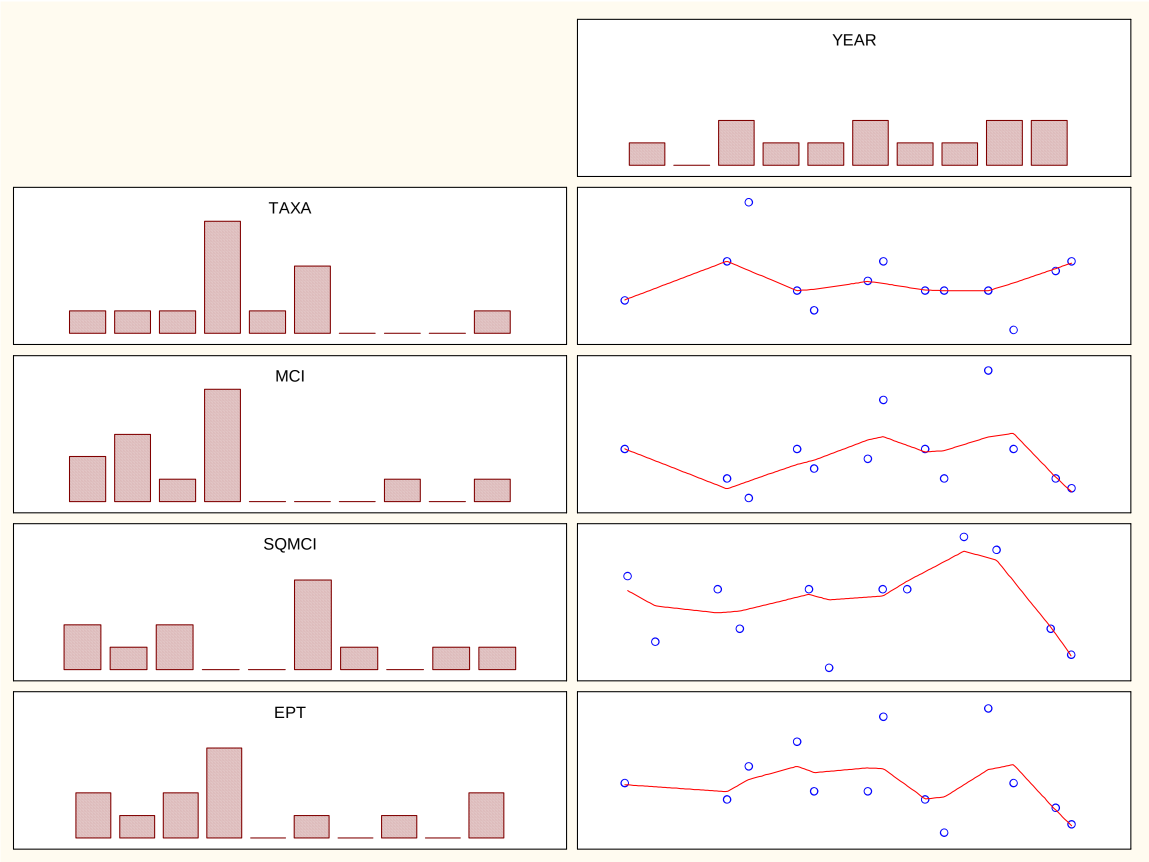
	N	Kendall	Z	p-level	p-exact
TAXA & YEAR	12	0.064651	0.292596	0.769831	----
MCI & YEAR	12	0.646230	2.924702	0.003448	----
SQMCI & YEAR	11	0.523570	2.241794	0.024975	----
EPT & YEAR	12	0.687043	3.109411	0.001875	----

KPA000950



Kendall Tau Correlations MD pairwise deleted Marked correlations are significant at p <.05000					
	N	**Kendall**	**Z**	**p-level**	**p-exact**
TAXA & YEAR	12	0.236404	1.069913	0.284658	----
MCI & YEAR	12	0.325669	1.473911	0.140506	----
SQMCI & YEAR	11	0.054545	0.233550	0.815335	----
EPT & YEAR	12	0.562775	2.547000	0.010865	----

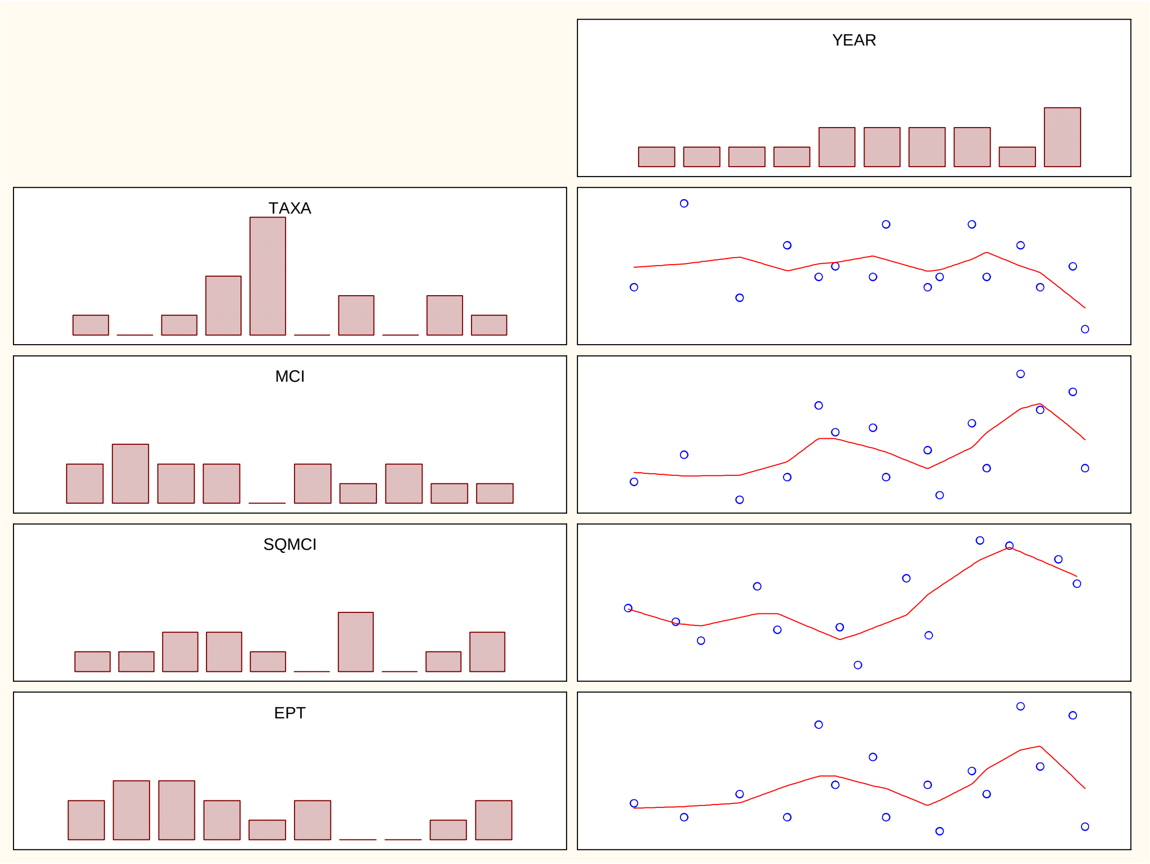
KPK000250



Kendall Tau Correlations MD pairwise deleted Marked correlations are significant at p <.05000

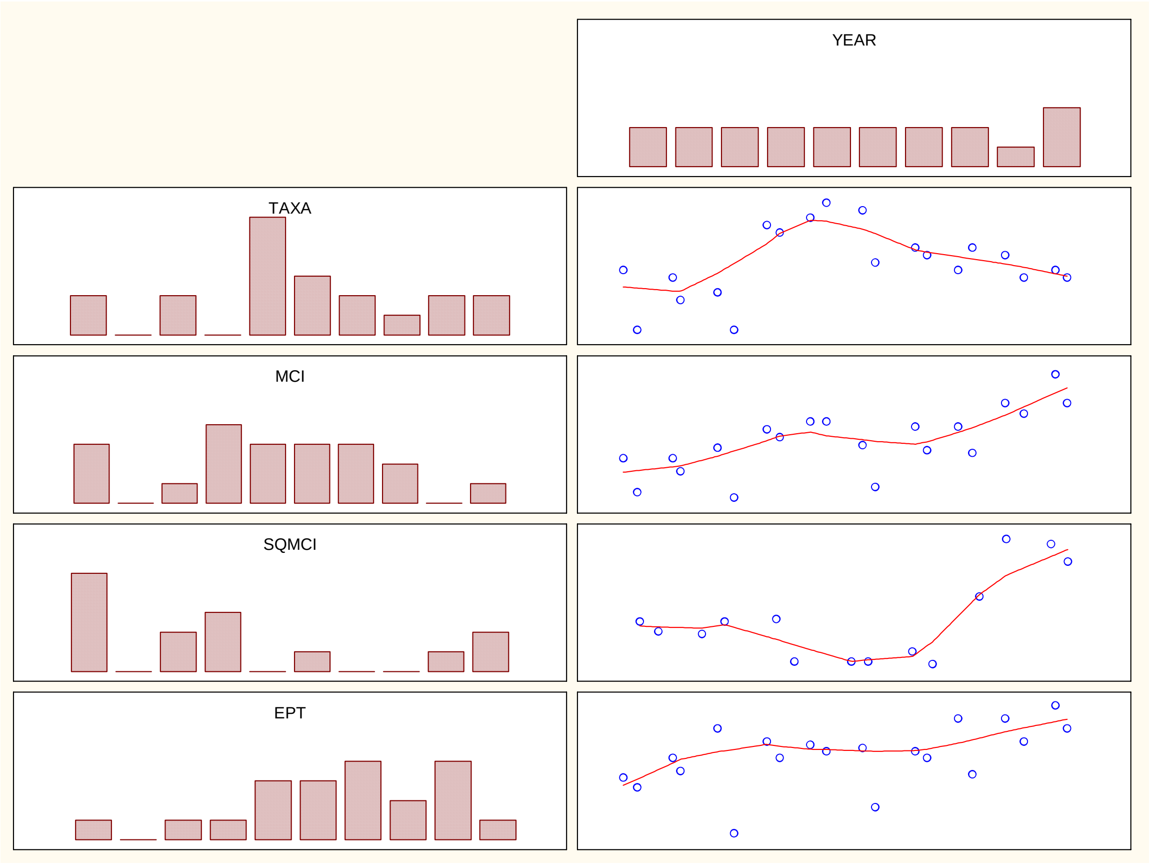
	N	Kendall	Z	p-level	p-exact
TAXA & YEAR	13	-0.013631	-0.06487	0.948281	----
MCI & YEAR	13	-0.013631	-0.06487	0.948281	----
SQMCI & YEAR	12	-0.016025	-0.07253	0.942183	----
EPT & YEAR	13	-0.222265	-1.05769	0.290196	----

KPK000500



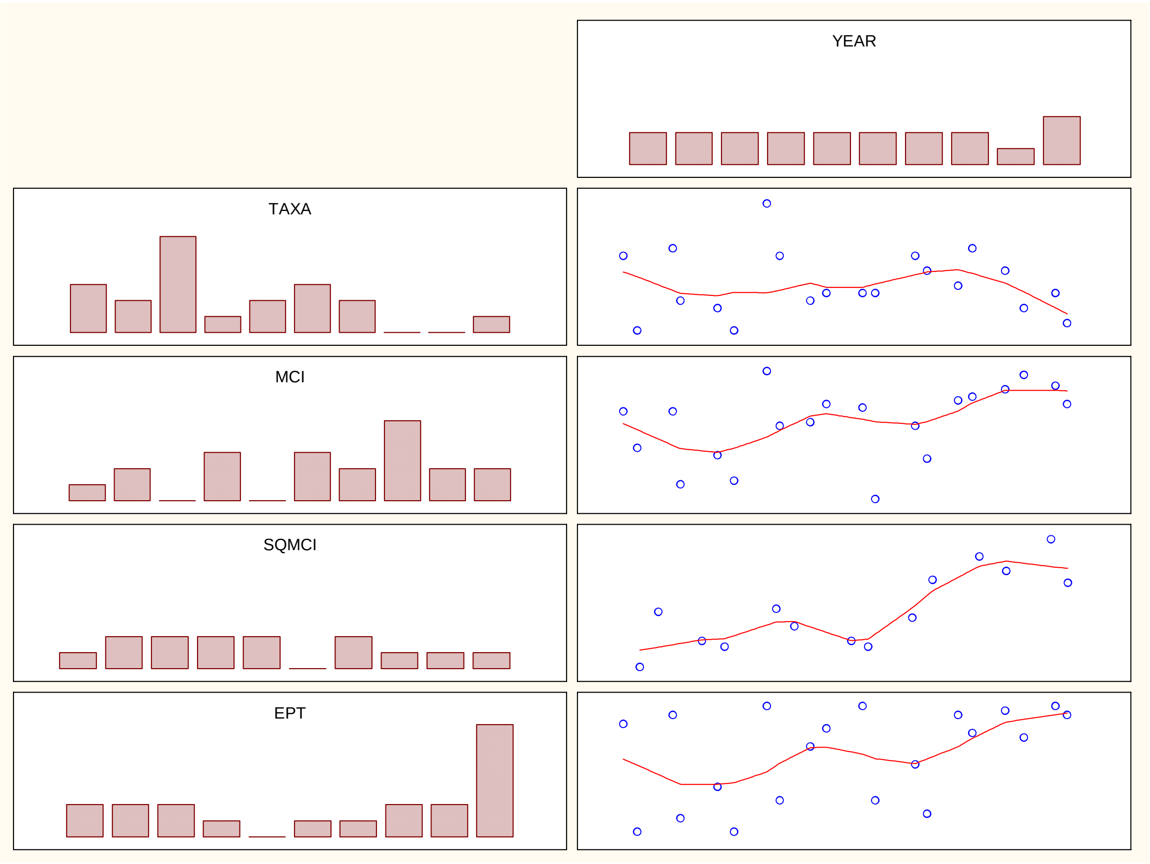
Kendall Tau Correlations MD pairwise deleted Marked correlations are significant at p <.05000					
	N	**Kendall**	**Z**	**p-level**	**p-exact**
TAXA & YEAR	16	-0.105409	-0.569495	0.569020	----
MCI & YEAR	16	0.336146	1.816098	0.069355	----
SQMCI & YEAR	13	0.282051	1.342196	0.179533	----
EPT & YEAR	16	0.195789	1.057790	0.290151	----

KPK000660



Kendall Tau Correlations MD pairwise deleted Marked correlations are significant at p <.05000					
	N	**Kendall**	**Z**	**p-level**	**p-exact**
TAXA & YEAR	20	0.080886	0.498617	0.618049	----
MCI & YEAR	20	0.521306	3.213547	0.001311	----
SQMCI & YEAR	14	0.191059	0.951817	0.341190	----
EPT & YEAR	20	0.391490	2.413306	0.015809	----

KPK000880



Kendall Tau Correlations MD pairwise deleted Marked correlations are significant at p <.05000

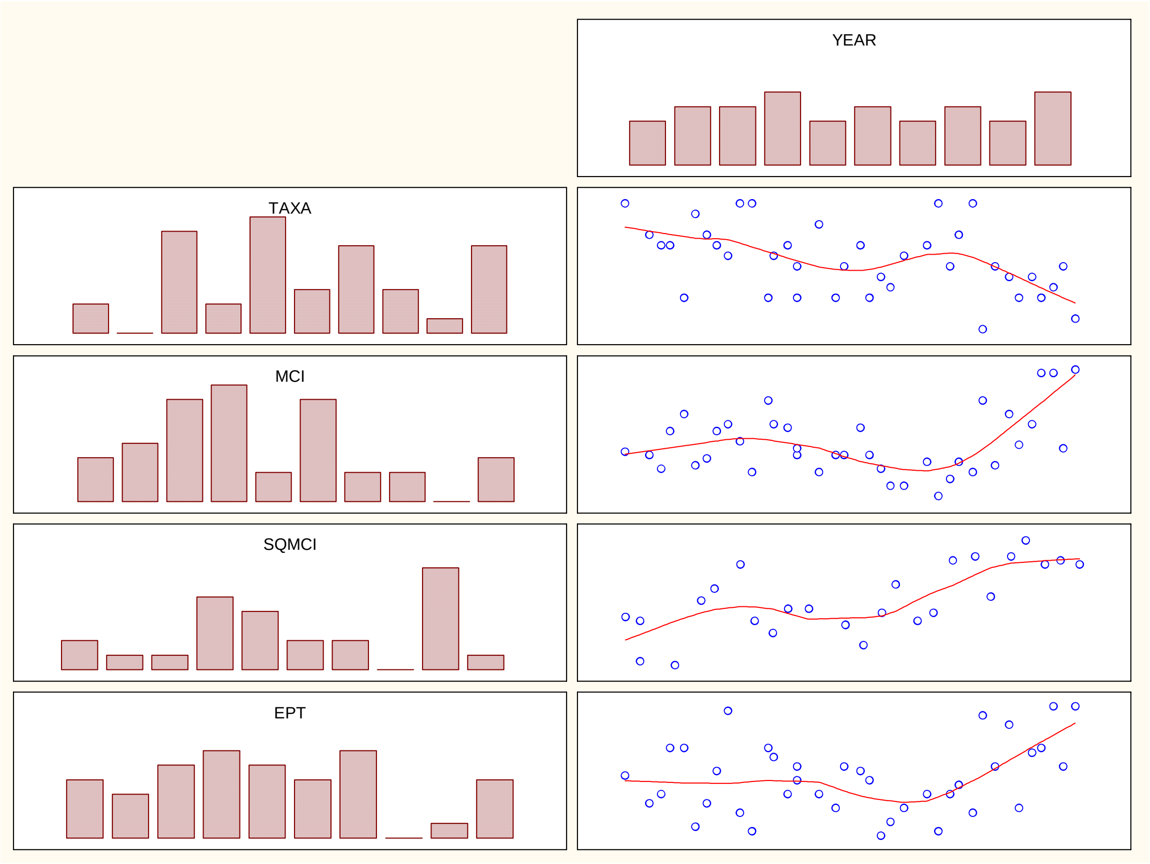
	N	Kendall	Z	p-level	p-exact
TAXA & YEAR	20	0.000000	0.000000	1.000000	----
MCI & YEAR	20	0.366059	2.256541	0.024037	----
SQMCI & YEAR	14	0.544478	2.712473	0.006678	----
EPT & YEAR	20	0.290390	1.790083	0.073441	----

KPK000990

Kendall Tau Correlations MD pairwise deleted Marked correlations are significant at p <.05000

	N	Kendall	Z	p-level	p-exact
TAXA & YEAR	12	-0.250122	-1.13200	0.257635	----
MCI & YEAR	12	0.338502	1.53199	0.125526	----
SQMCI & YEAR	12	0.635831	2.87764	0.004007	----
EPT & YEAR	12	0.461593	2.08907	0.036701	----

KPN000275



Kendall Tau Correlations MD pairwise deleted Marked correlations are significant at p <.05000

	N	Kendall	Z	p-level	p-exact
TAXA & YEAR	38	-0.331982	-2.93407	0.003345	----
MCI & YEAR	38	0.058023	0.51281	0.608087	----
SQMCI & YEAR	25	0.465250	3.25977	0.001115	----
EPT & YEAR	38	0.135403	1.19670	0.231423	----

KPN000360

Error! Objects cannot be created from editing field codes.

Kendall Tau Correlations MD pairwise deleted Marked correlations are significant at p <.05000

	N	Kendall	Z	p-level	p-exact
TAXA & YEAR	28	0.133482	0.996841	0.318842	----
MCI & YEAR	28	0.180865	1.350690	0.176795	----
SQMCI & YEAR	25	0.480678	3.367868	0.000758	----
EPT & YEAR	28	0.154722	1.155461	0.247902	----

KRP000300

Error! Objects cannot be created from editing field codes.

Kendall Tau Correlations MD pairwise deleted Marked correlations are significant at p <.05000

	N	Kendall	Z	p-level	p-exact
TAXA & YEAR	21	0.330413	2.095267	0.036147	----
MCI & YEAR	21	0.498861	3.163454	0.001559	----
SQMCI & YEAR	13	-0.039757	-0.189191	0.849943	----
EPT & YEAR	21	0.259262	1.644075	0.100161	----

KRP000660

Error! Objects cannot be created from editing field codes.

Kendall Tau Correlations MD pairwise deleted Marked correlations are significant at p <.05000

	N	Kendall	Z	p-level	p-exact
TAXA & YEAR	21	0.568874	3.607428	0.000309	----
MCI & YEAR	21	0.582634	3.694689	0.000220	----
SQMCI & YEAR	13	0.065372	0.311086	0.755736	----
EPT & YEAR	21	0.637906	4.045189	0.000052	----

KTK000150

Error! Objects cannot be created from editing field codes.

Kendall Tau Correlations MD pairwise deleted Marked correlations are significant at p <.05000

	N	Kendall	Z	p-level	p-exact
TAXA & YEAR	12	-0.254257	-1.15071	0.249851	----
MCI & YEAR	12	-0.170589	-0.77205	0.440086	----
SQMCI & YEAR	12	0.077540	0.35093	0.725640	----
EPT & YEAR	12	0.046524	0.21056	0.833232	----

KTK000248

Error! Objects cannot be created from editing field codes.

Kendall Tau Correlations MD pairwise deleted Marked correlations are significant at p <.05000

	N	Kendall	Z	p-level	p-exact
TAXA & YEAR	11	-0.094407	-0.404226	0.686046	----
MCI & YEAR	11	0.672727	2.880446	0.003971	----
SQMCI & YEAR	11	0.844072	3.614101	0.000301	----
EPT & YEAR	11	0.440386	1.885618	0.059346	----

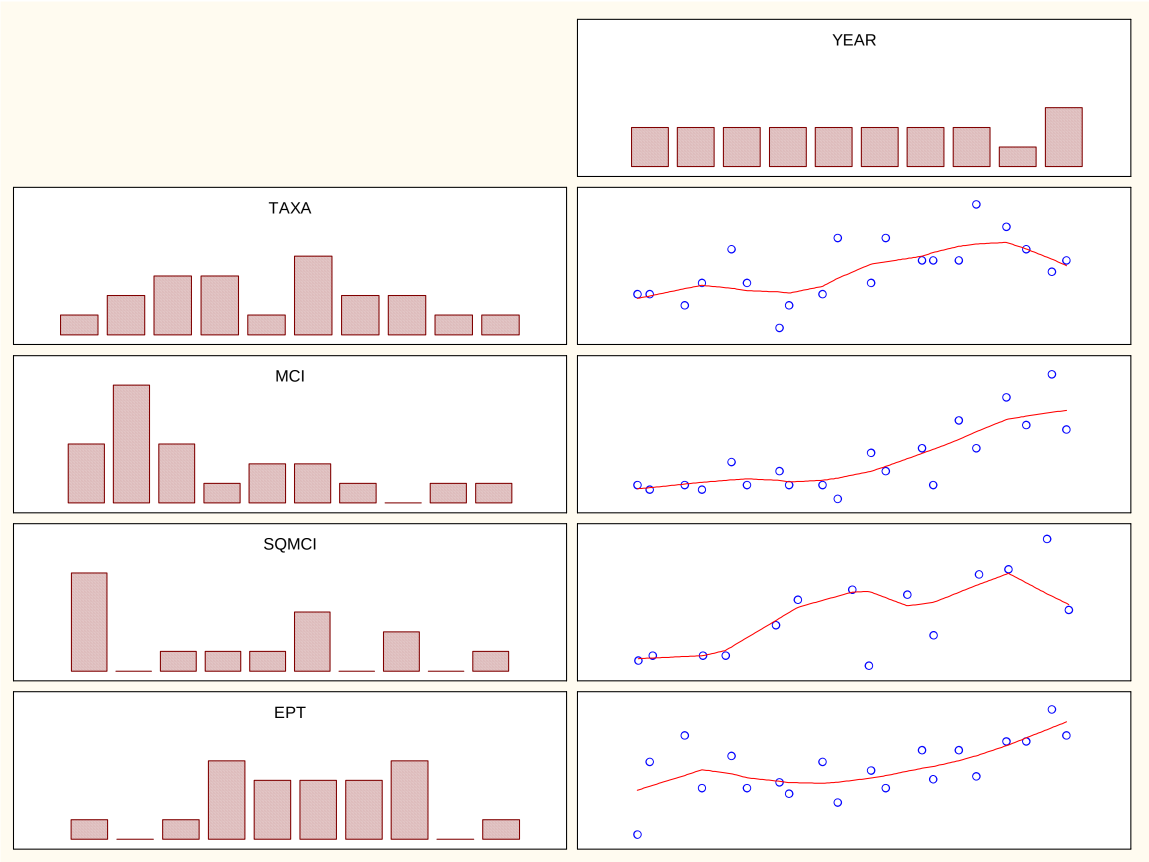
MGE000970

Error! Objects cannot be created from editing field codes.

Kendall Tau Correlations MD pairwise deleted Marked correlations are significant at p <.05000

	N	Kendall	Z	p-level	p-exact
TAXA & YEAR	6	-0.690066	-1.94461	0.051822	.068
MCI & YEAR	6	0.200000	0.56360	0.573025	.360
SQMCI & YEAR	6	0.200000	0.56360	0.573025	.360
EPT & YEAR	6	0.690066	1.94461	0.051822	.068

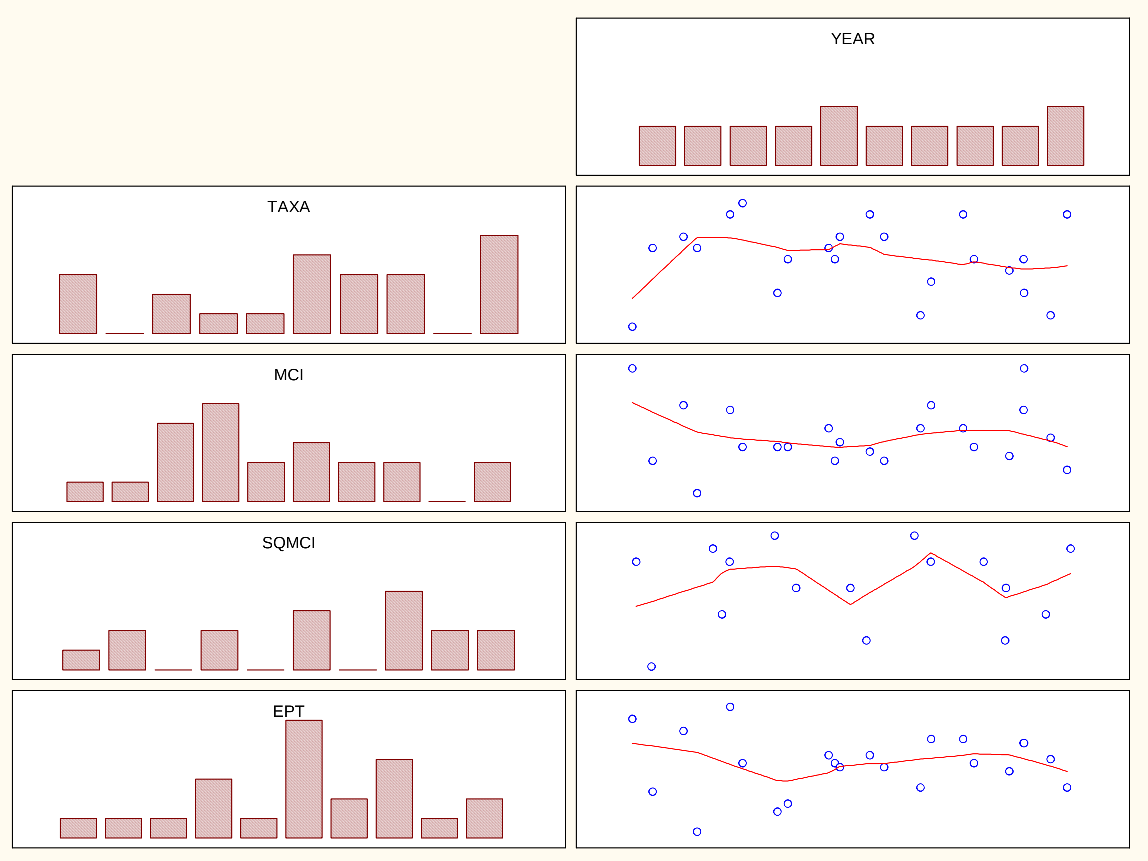
MGH000950



Kendall Tau Correlations MD pairwise deleted Marked correlations are significant at p <.05000

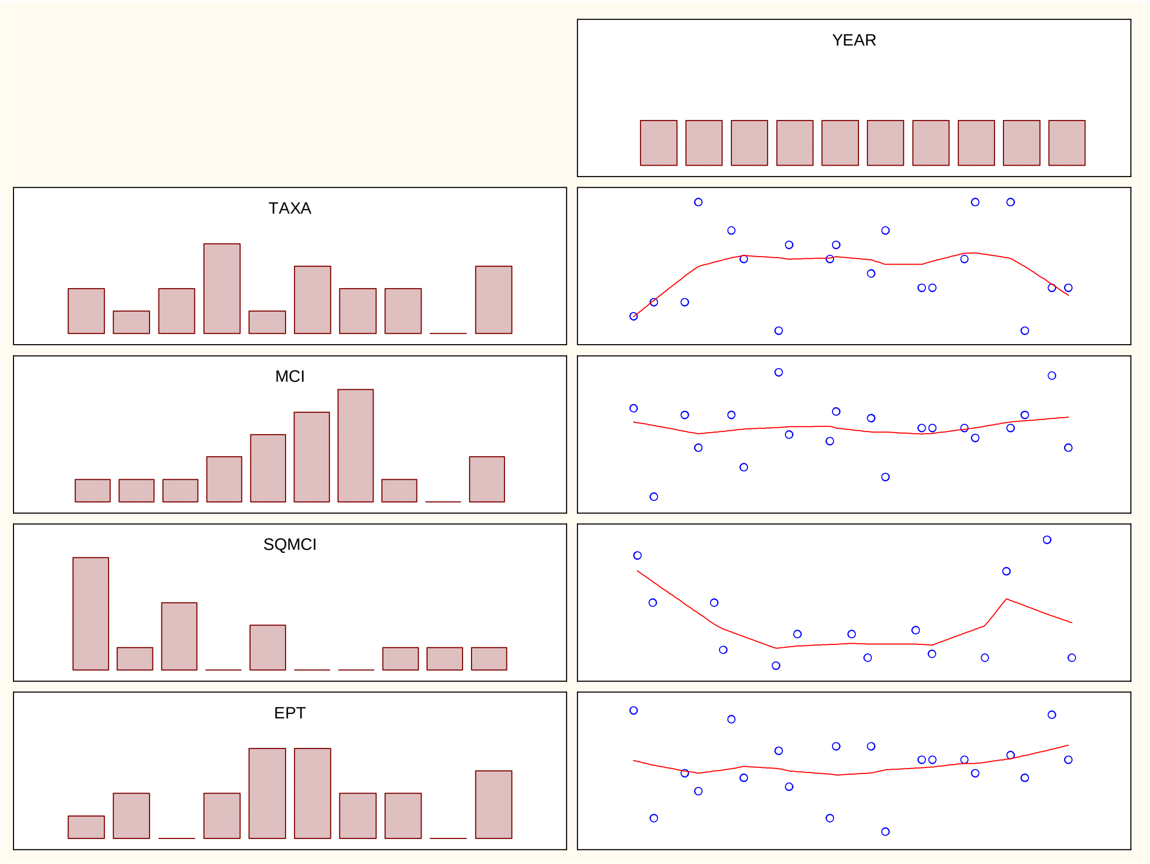
	N	Kendall	Z	p-level	p-exact
TAXA & YEAR	20	0.433243	2.670687	0.007570	----
MCI & YEAR	20	0.597424	3.682769	0.000231	----
SQMCI & YEAR	14	0.581087	2.894852	0.003793	----
EPT & YEAR	20	0.412941	2.545541	0.010911	----

MGN000195



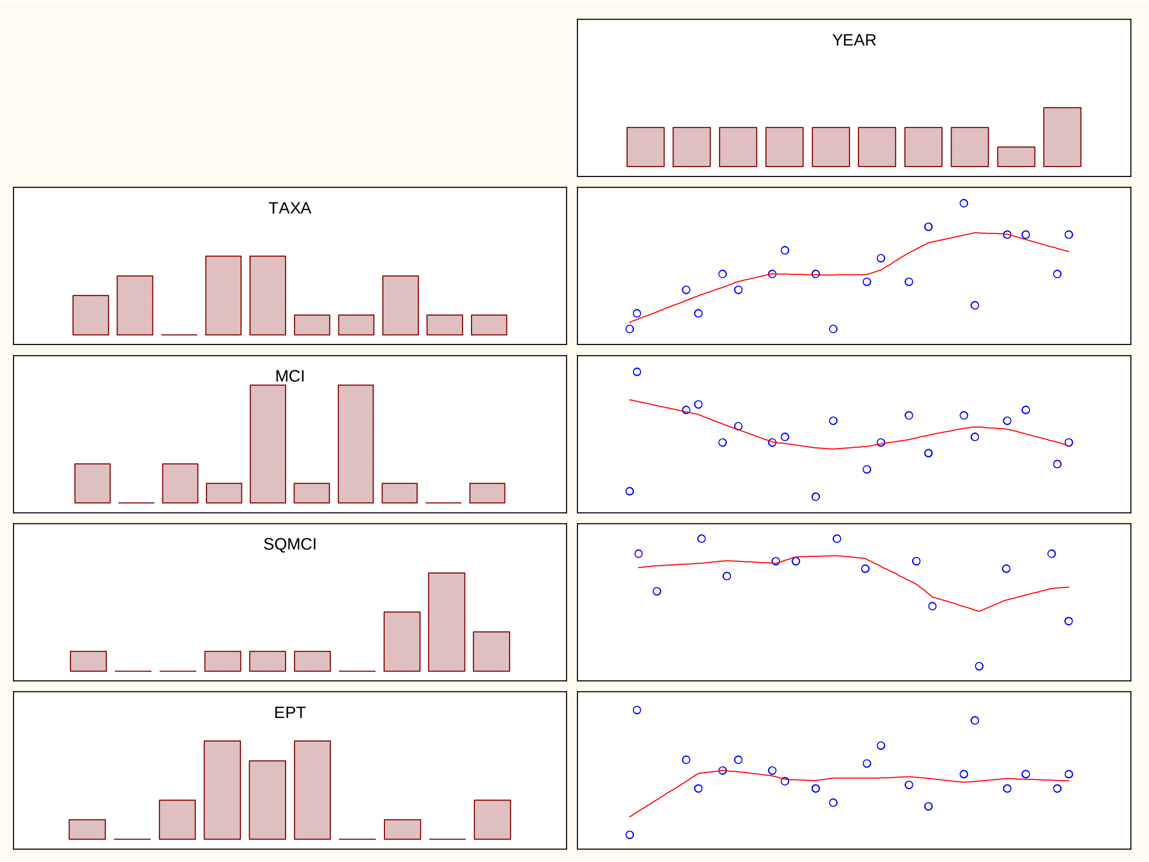
Kendall Tau Correlations MD pairwise deleted Marked correlations are significant at p <.05000					
	N	**Kendall**	**Z**	**p-level**	**p-exact**
TAXA & YEAR	22	-0.113238	-0.737602	0.460756	----
MCI & YEAR	22	0.000000	0.000000	1.000000	----
SQMCI & YEAR	16	-0.008825	-0.047679	0.961972	----
EPT & YEAR	22	0.013218	0.086098	0.931389	----

MGN000427



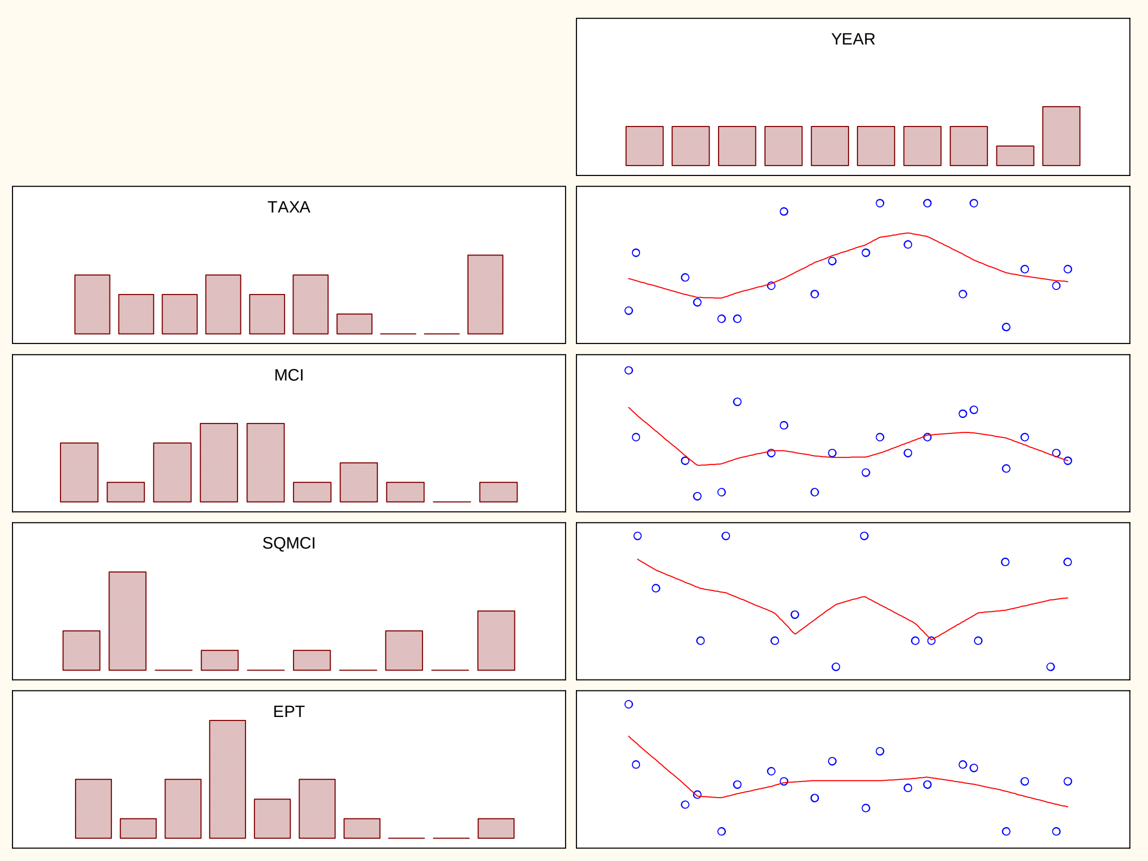
Kendall Tau Correlations	MD pairwise deleted	Marked correlations are significant at p <.05000			
	N	**Kendall**	**Z**	**p-level**	**p-exact**
TAXA & YEAR	20	0.054998	0.339032	0.734586	----
MCI & YEAR	20	0.021630	0.133333	0.893930	----
SQMCI & YEAR	14	-0.158255	-0.788394	0.430467	----
EPT & YEAR	20	0.064889	0.400000	0.689157	----

MGO000050



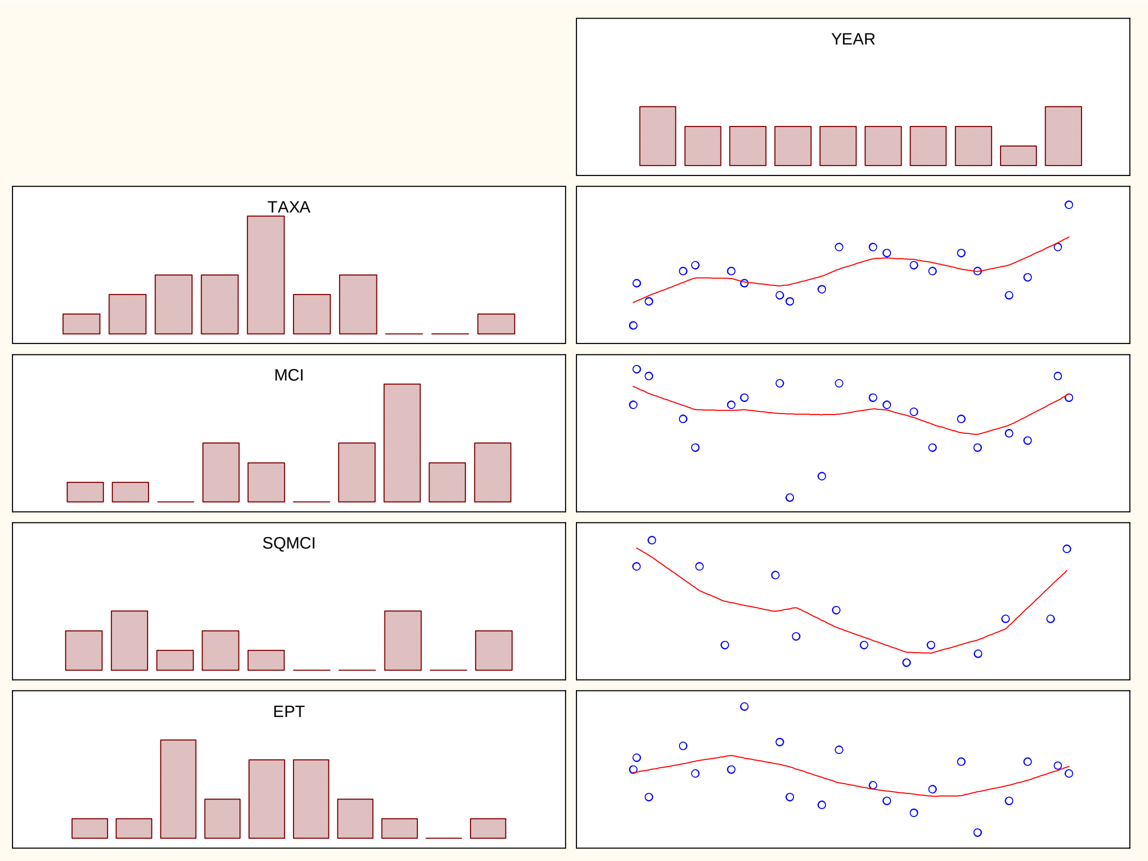
Kendall Tau Correlations MD pairwise deleted Marked correlations are significant at p <.05000					
	N	**Kendall**	**Z**	**p-level**	**p-exact**
TAXA & YEAR	20	0.485318	2.99170	0.002774	----
MCI & YEAR	20	-0.075703	-0.46667	0.640738	----
SQMCI & YEAR	14	-0.261516	-1.30281	0.192638	----
EPT & YEAR	20	-0.103027	-0.63510	0.525363	----

MGO000190



Kendall Tau Correlations MD pairwise deleted Marked correlations are significant at p <.05000					
	N	**Kendall**	**Z**	**p-level**	**p-exact**
TAXA & YEAR	20	0.193593	1.19339	0.232717	----
MCI & YEAR	20	0.021874	0.13484	0.892738	----
SQMCI & YEAR	14	-0.216444	-1.07828	0.280911	----
EPT & YEAR	20	-0.086041	-0.53039	0.595838	----

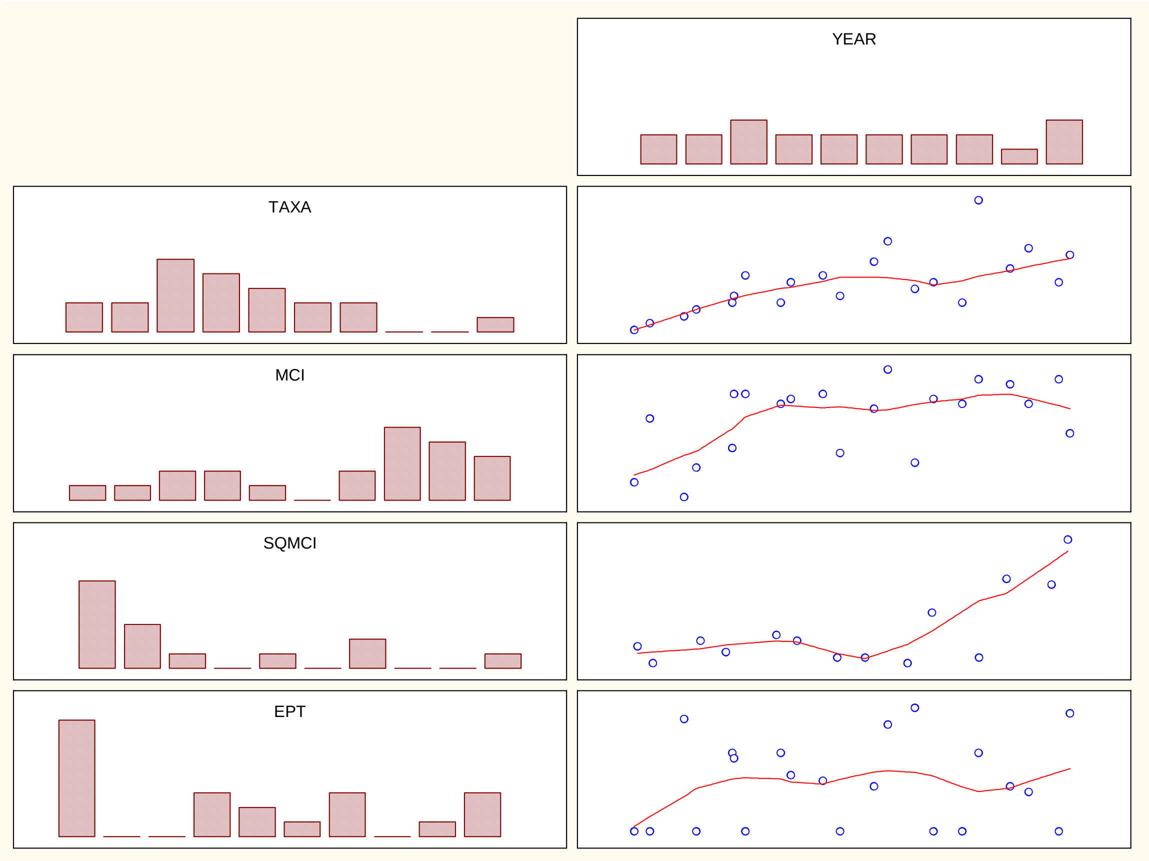
MGT000488



Kendall Tau Correlations MD pairwise deleted Marked correlations are significant at p <.05000

	N	Kendall	Z	p-level	p-exact
TAXA & YEAR	21	0.354891	2.25049	0.024418	----
MCI & YEAR	21	-0.156931	-0.99515	0.319662	----
SQMCI & YEAR	14	-0.271295	-1.35153	0.176525	----
EPT & YEAR	21	-0.207244	-1.31421	0.188776	----

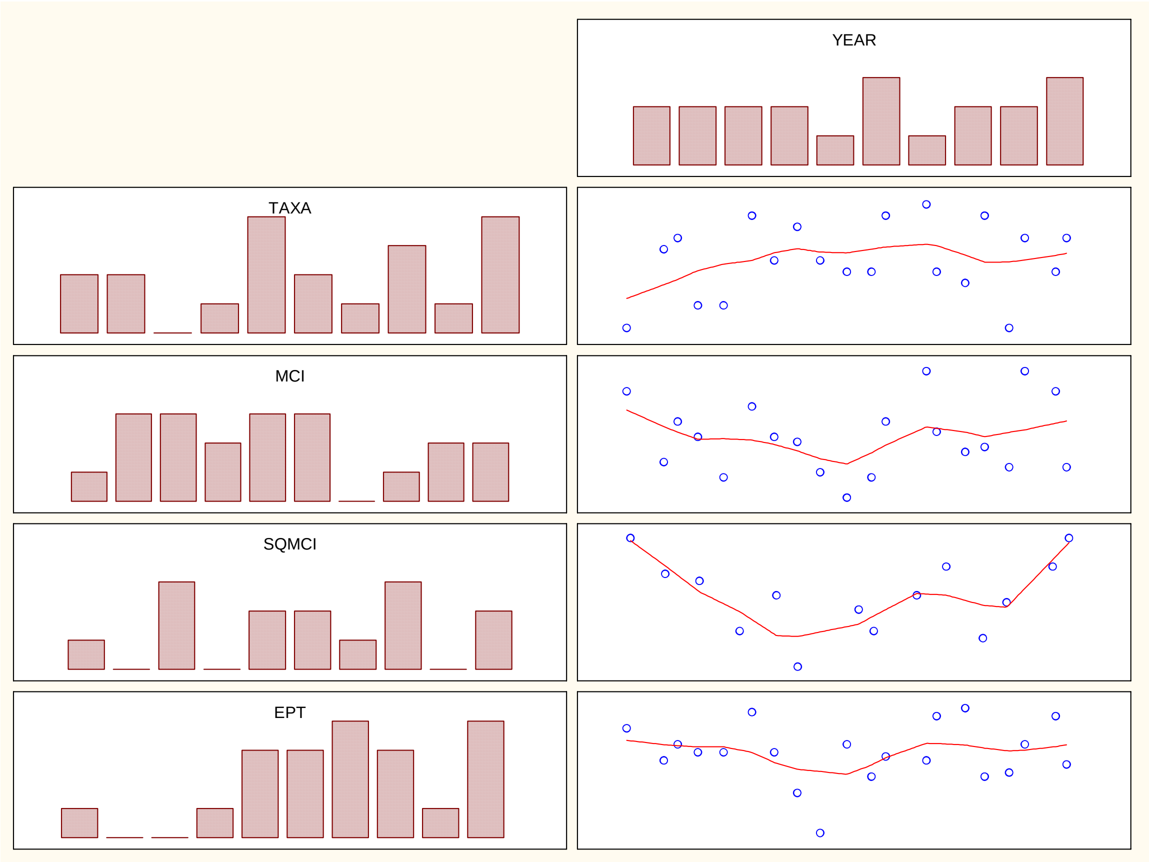
MGT000520



Kendall Tau Correlations MD pairwise deleted Marked correlations are significant at p <.05000

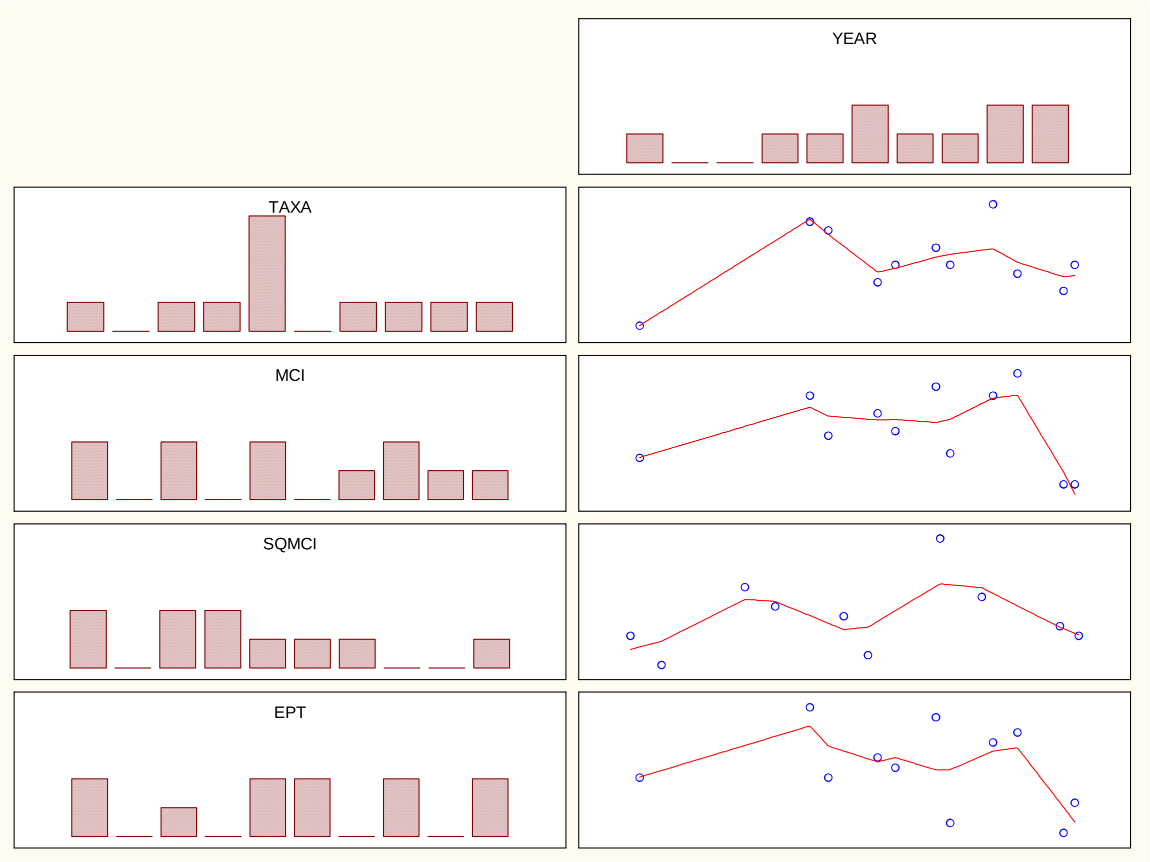
	N	Kendall	Z	p-level	p-exact
TAXA & YEAR	21	0.602055	3.817845	0.000135	----
MCI & YEAR	21	0.339870	2.155235	0.031143	----
SQMCI & YEAR	14	0.339118	1.689415	0.091140	----
EPT & YEAR	21	0.051723	0.327991	0.742918	----

MHW000060



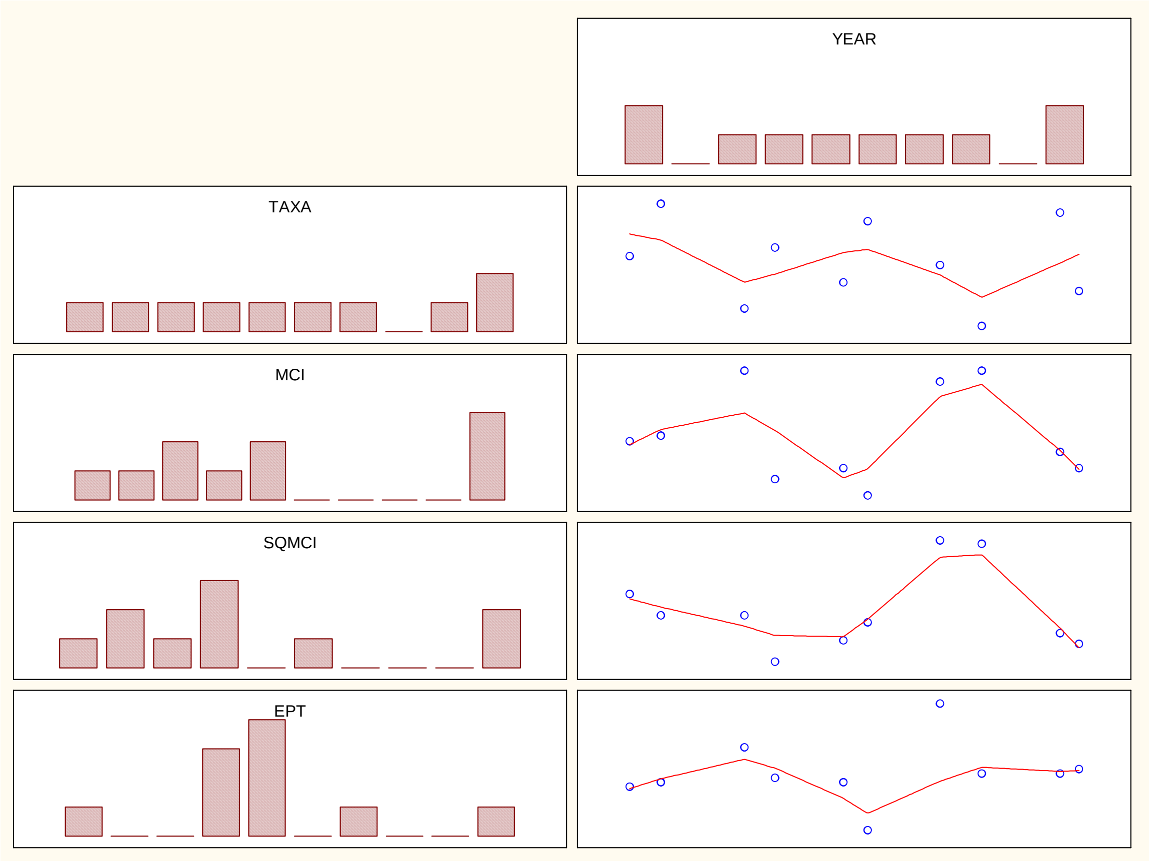
Kendall Tau Correlations MD pairwise deleted Marked correlations are significant at p <.05000					
	N	**Kendall**	**Z**	**p-level**	**p-exact**
TAXA & YEAR	20	0.093229	0.574705	0.565491	----
MCI & YEAR	20	-0.021393	-0.131876	0.895082	----
SQMCI & YEAR	14	0.011239	0.055989	0.955350	----
EPT & YEAR	20	-0.037747	-0.232688	0.816004	----

MKW000200



Kendall Tau Correlations	MD pairwise deleted Marked correlations are significant at p <.05000				
	N	Kendall	Z	p-level	p-exact
TAXA & YEAR	11	-0.074796	-0.320256	0.748774	----
MCI & YEAR	11	-0.055565	-0.237915	0.811947	----
SQMCI & YEAR	10	0.044947	0.180907	0.856441	.500
EPT & YEAR	11	-0.183494	-0.785674	0.432058	----

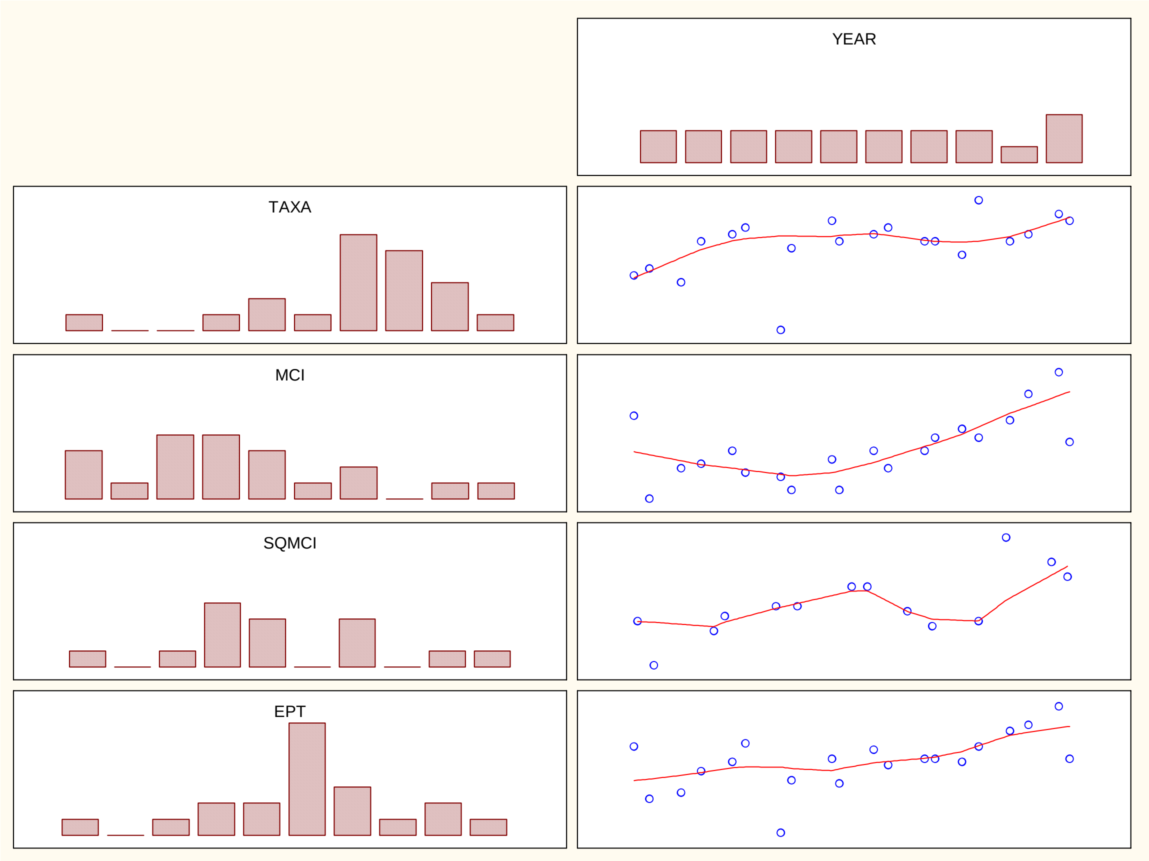
MKW000300



Kendall Tau Correlations MD pairwise deleted Marked correlations are significant at p <.05000

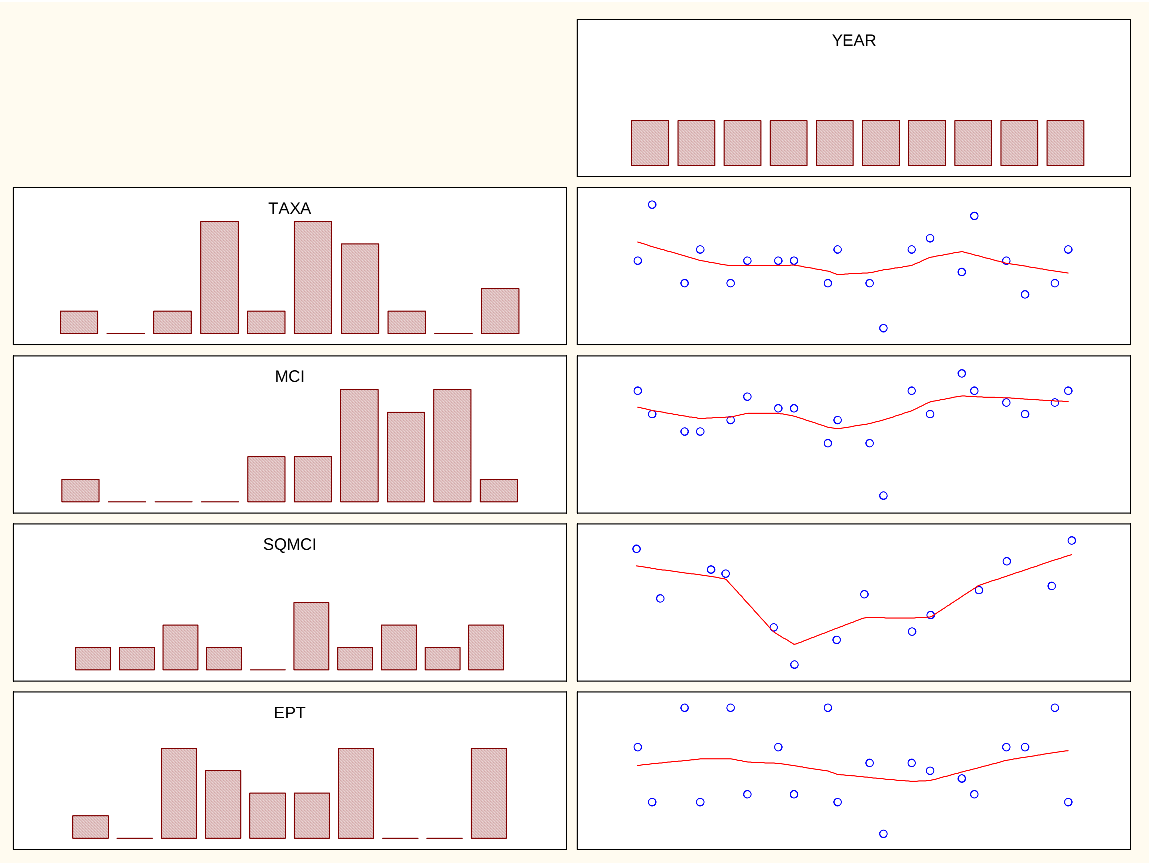
	N	Kendall	Z	p-level	p-exact
TAXA & YEAR	10	-0.155556	-0.626099	0.531250	.300
MCI & YEAR	10	-0.068199	-0.274497	0.783702	.431
SQMCI & YEAR	10	-0.179787	-0.723627	0.469295	.300
EPT & YEAR	10	0.340997	1.372487	0.169912	.108

MRK000420



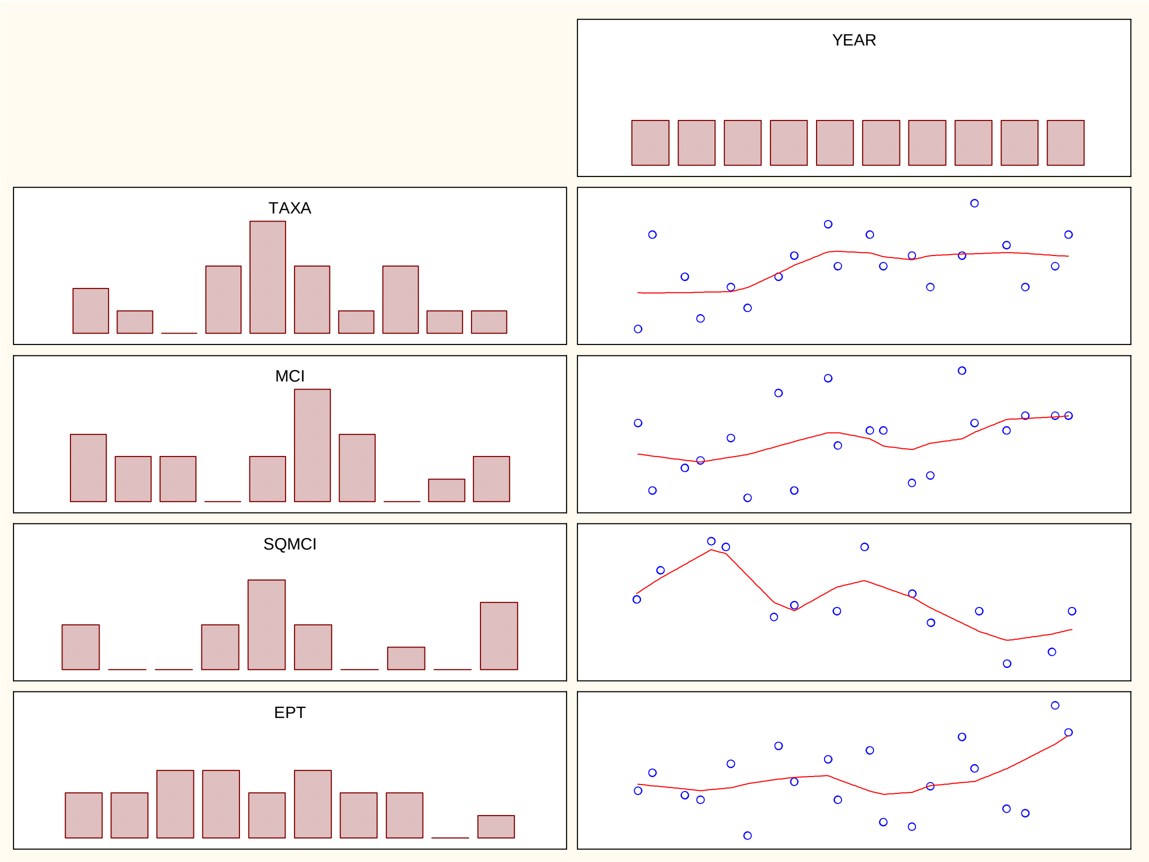
Kendall Tau Correlations MD pairwise deleted Marked correlations are significant at p <.05000					
	N	Kendall	Z	p-level	p-exact
TAXA & YEAR	20	0.411306	2.535463	0.011230	----
MCI & YEAR	20	0.481345	2.967212	0.003005	----
SQMCI & YEAR	14	0.491689	2.449490	0.014306	----
EPT & YEAR	20	0.451717	2.784573	0.005360	----

MWH000380



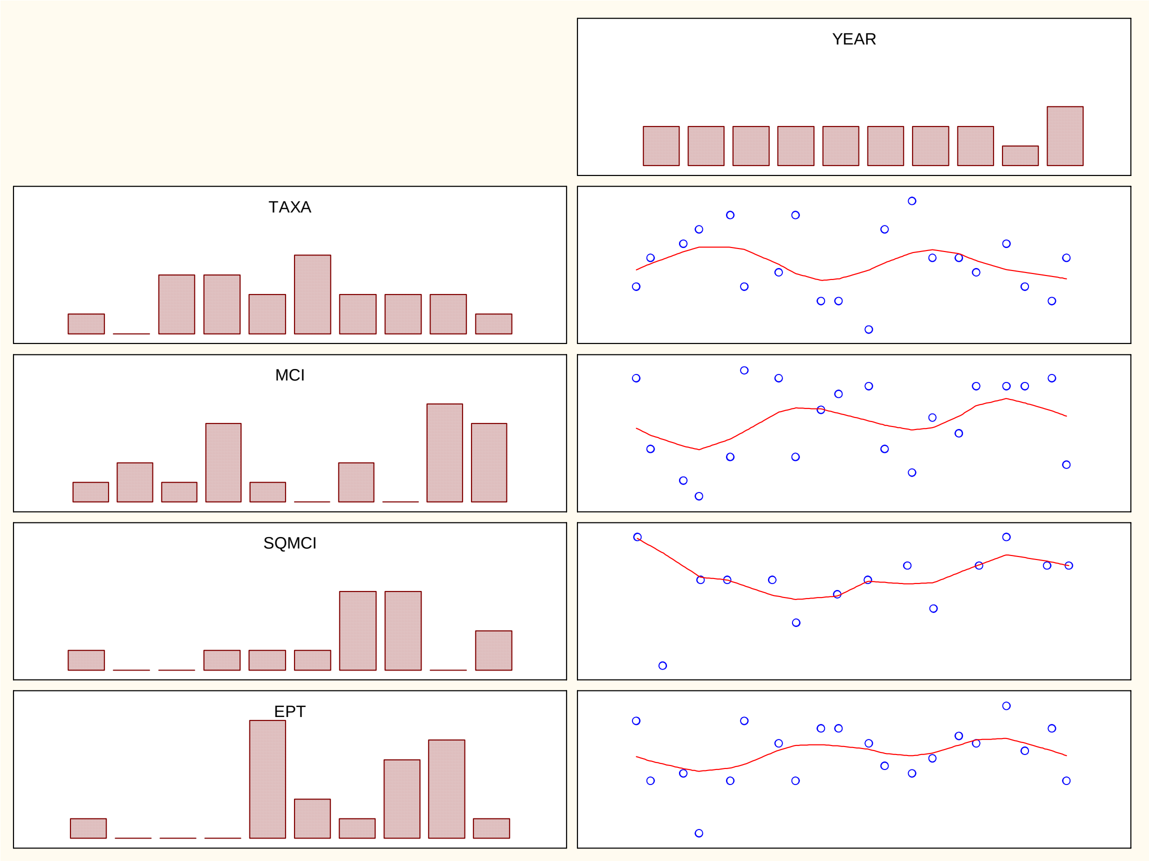
Kendall Tau Correlations	MD pairwise deleted Marked correlations are significant at p <.05000				
	N	**Kendall**	**Z**	**p-level**	**p-exact**
TAXA & YEAR	20	-0.045320	-0.279372	0.779959	----
MCI & YEAR	20	0.196865	1.213560	0.224916	----
SQMCI & YEAR	14	0.120879	0.602194	0.547045	----
EPT & YEAR	20	-0.011194	-0.069007	0.944984	----

MWH000490



Kendall Tau Correlations MD pairwise deleted Marked correlations are significant at p <.05000					
	N	**Kendall**	**Z**	**p-level**	**p-exact**
TAXA & YEAR	20	0.321728	1.98326	0.047338	----
MCI & YEAR	20	0.344166	2.12158	0.033873	----
SQMCI & YEAR	14	-0.460791	-2.29556	0.021701	----
EPT & YEAR	20	0.131927	0.81325	0.416075	----

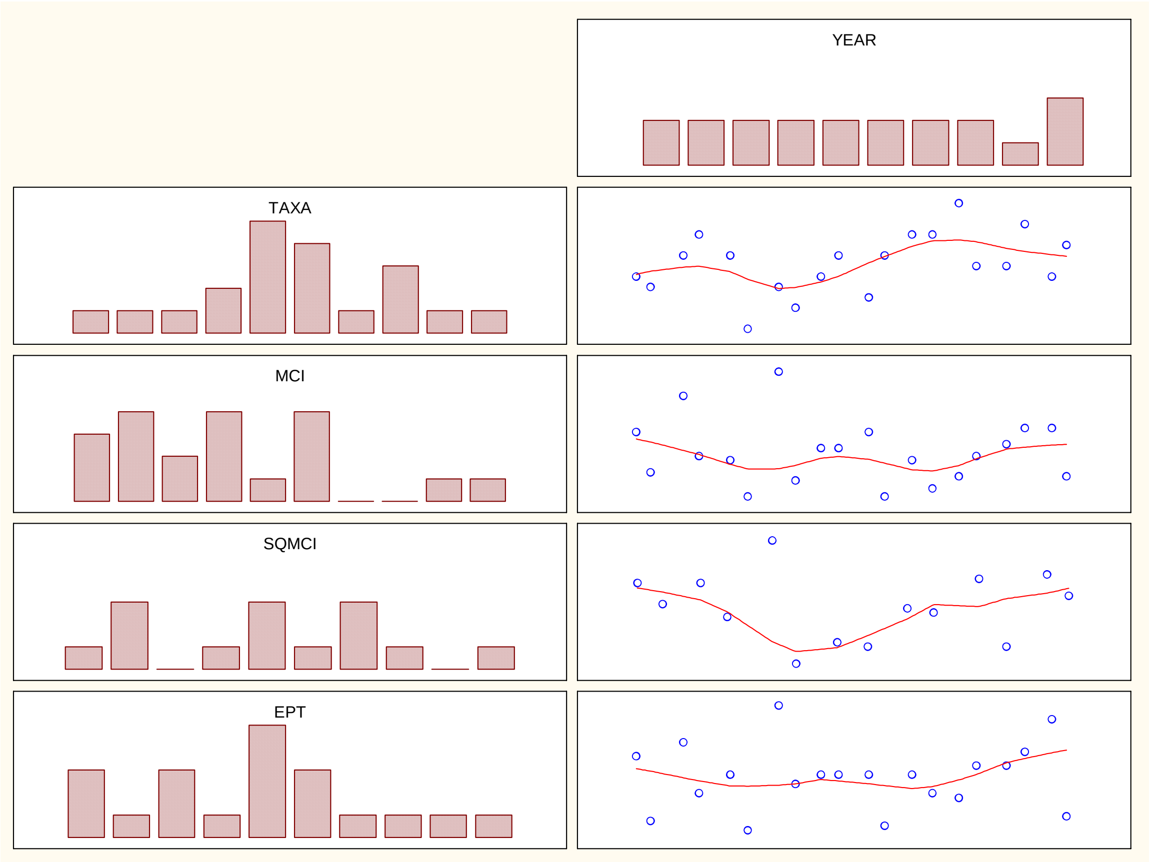
PAT000200



Kendall Tau Correlations MD pairwise deleted Marked correlations are significant at p <.05000

	N	Kendall	Z	p-level	p-exact
TAXA & YEAR	20	-0.098997	-0.610257	0.541691	----
MCI & YEAR	20	0.135562	0.835658	0.403347	----
SQMCI & YEAR	14	0.308607	1.537412	0.124192	----
EPT & YEAR	20	0.131244	0.809040	0.418492	----

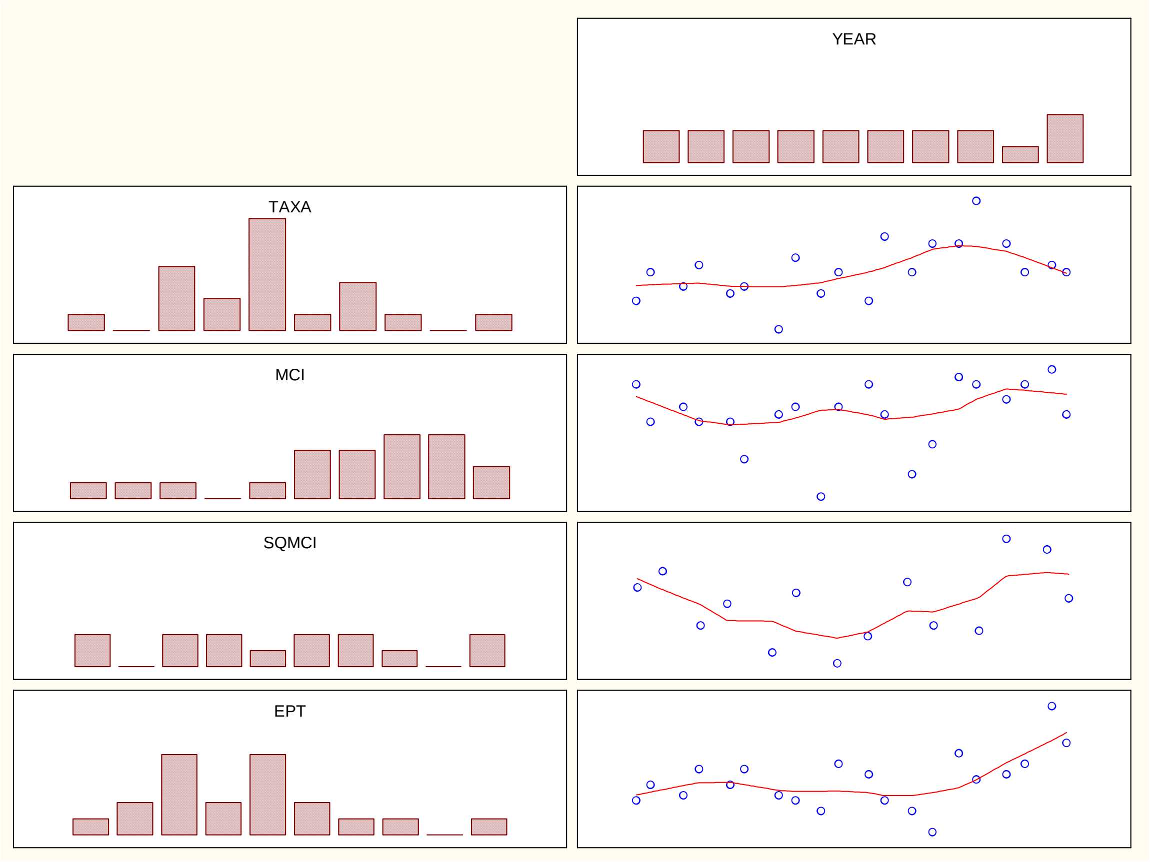
PAT000315



Kendall Tau Correlations MD pairwise deleted Marked correlations are significant at p <.05000

	N	Kendall	Z	p-level	p-exact
TAXA & YEAR	20	0.273424	1.685500	0.091892	----
MCI & YEAR	20	0.005363	0.033059	0.973627	----
SQMCI & YEAR	14	-0.011112	-0.055357	0.955854	----
EPT & YEAR	20	0.087003	0.536321	0.591737	----

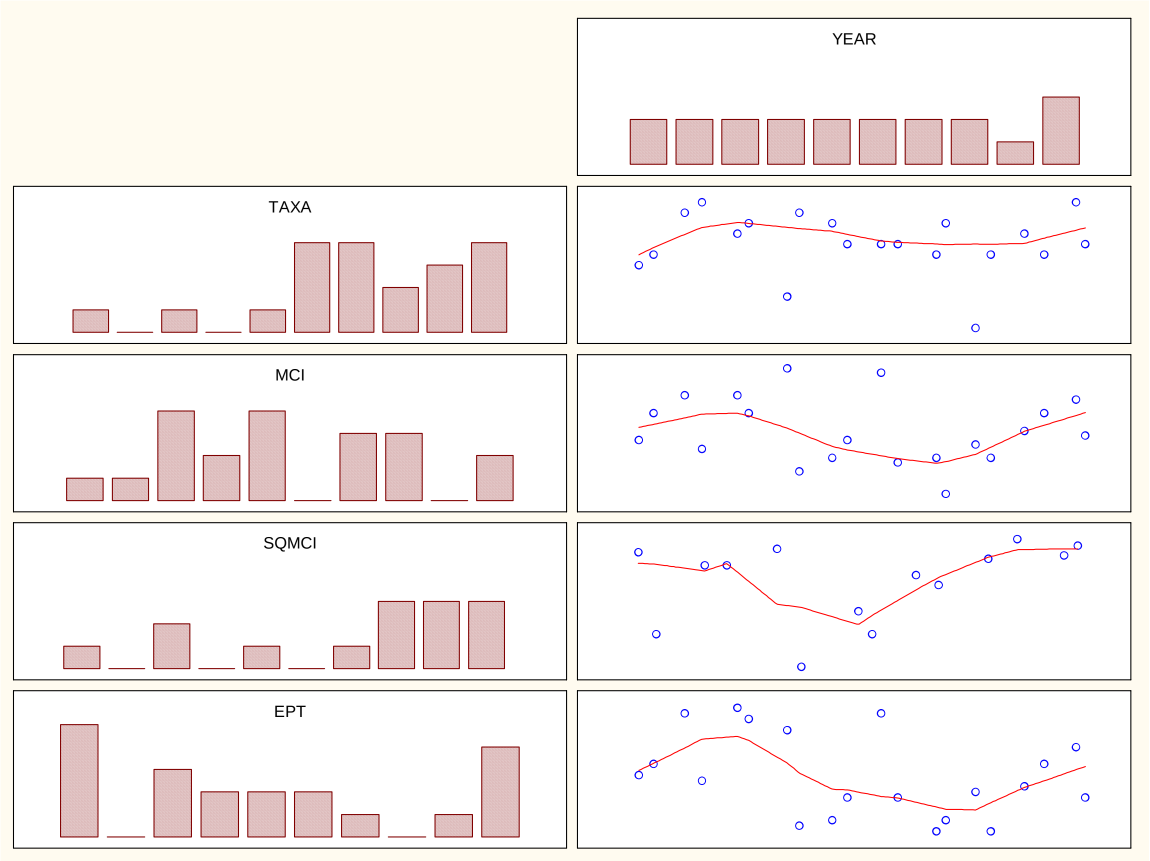
PAT000360



Kendall Tau Correlations MD pairwise deleted Marked correlations are significant at p <.05000

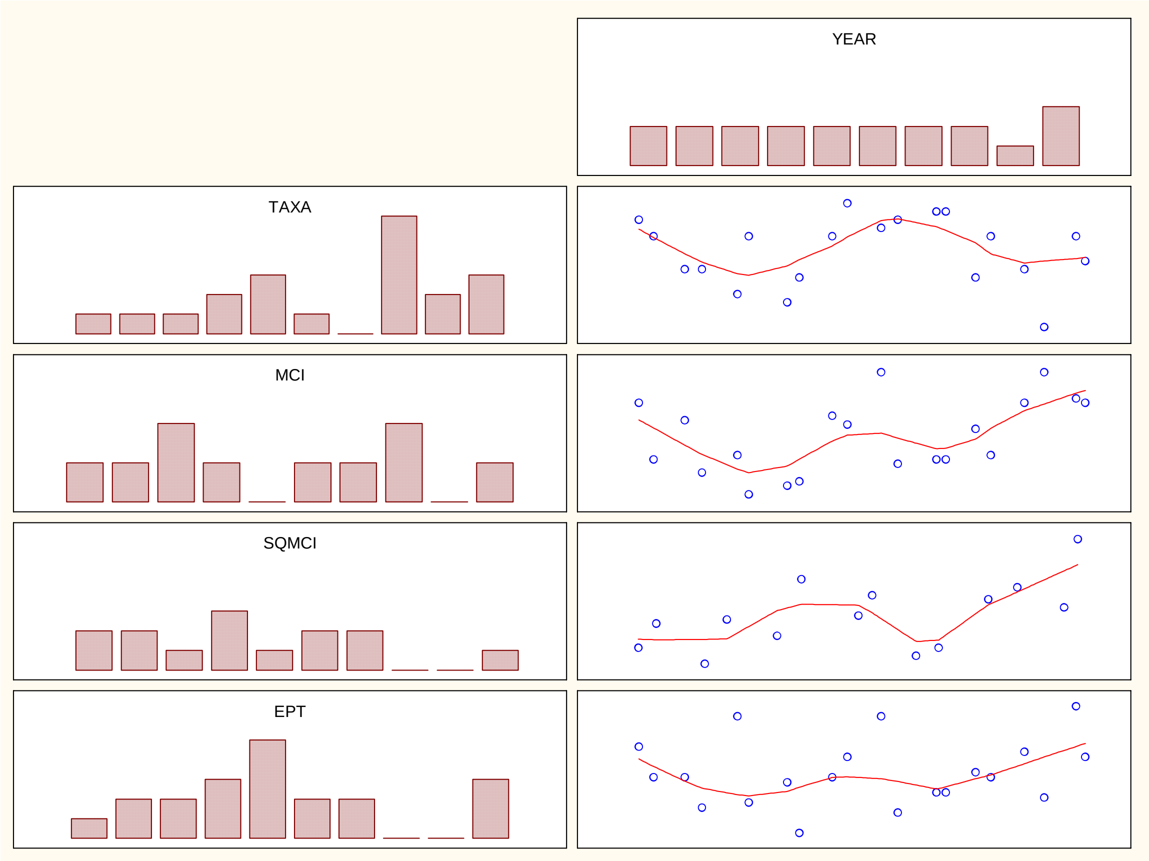
	N	Kendall	Z	p-level	p-exact
TAXA & YEAR	20	0.347489	2.142064	0.032188	----
MCI & YEAR	20	0.246784	1.521278	0.128190	----
SQMCI & YEAR	14	0.088399	0.440386	0.659658	----
EPT & YEAR	20	0.264229	1.628815	0.103352	----

PNH000200



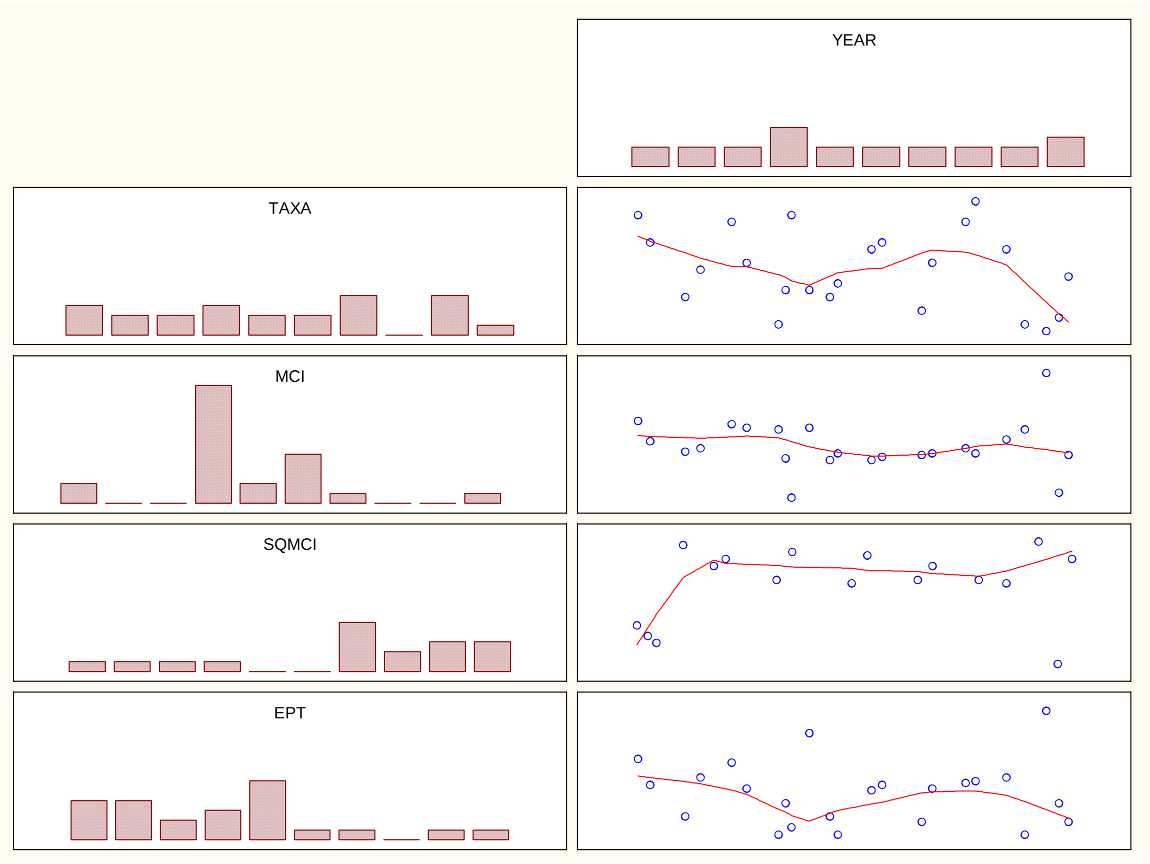
Kendall Tau Correlations MD pairwise deleted Marked correlations are significant at p <.05000					
	N	**Kendall**	**Z**	**p-level**	**p-exact**
TAXA & YEAR	20	-0.088507	-0.54560	0.585344	----
MCI & YEAR	20	-0.075286	-0.46410	0.642579	----
SQMCI & YEAR	14	0.277795	1.38391	0.166385	----
EPT & YEAR	20	-0.219878	-1.35542	0.175284	----

PNH000900



Kendall Tau Correlations MD pairwise deleted Marked correlations are significant at p <.05000					
	N	**Kendall**	**Z**	**p-level**	**p-exact**
TAXA & YEAR	20	-0.011000	-0.067806	0.945940	----
MCI & YEAR	20	0.311900	1.922681	0.054520	----
SQMCI & YEAR	14	0.419896	2.091831	0.036454	----
EPT & YEAR	20	0.113241	0.698064	0.485137	----

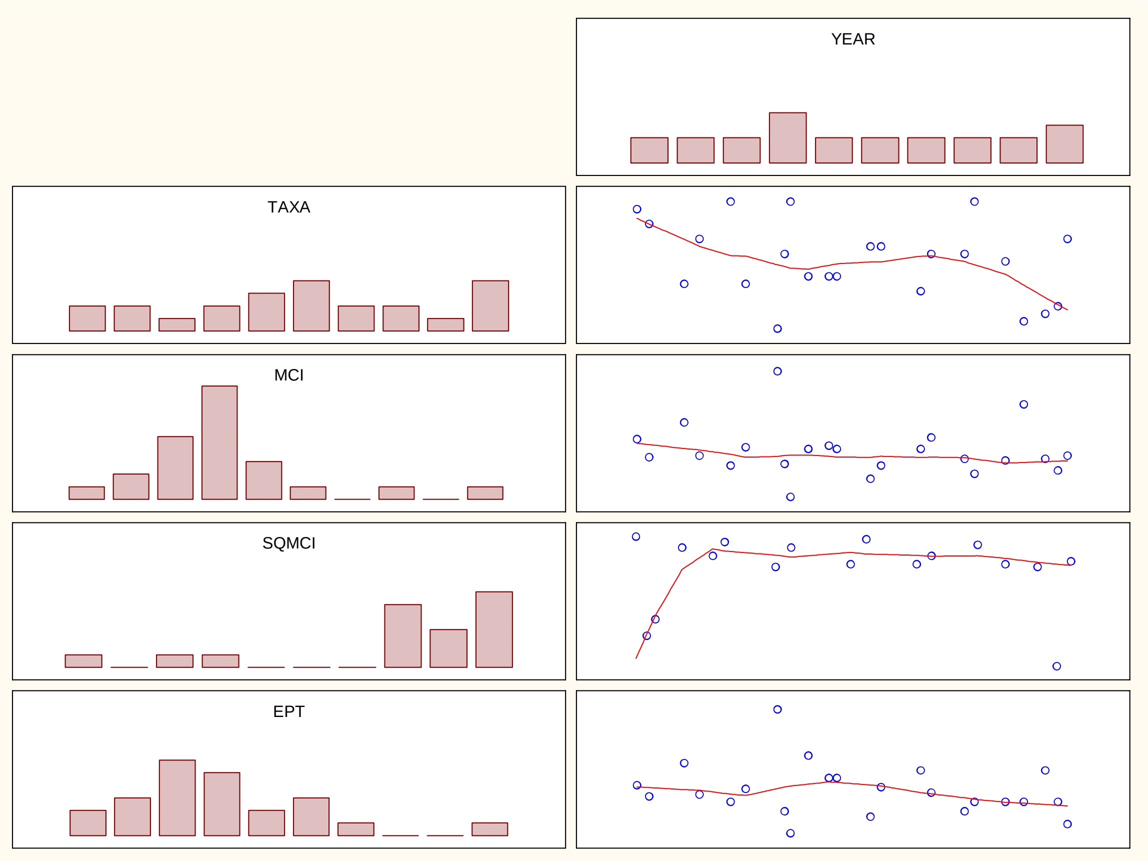
STY000300



Kendall Tau Correlations MD pairwise deleted Marked correlations are significant at p <.05000

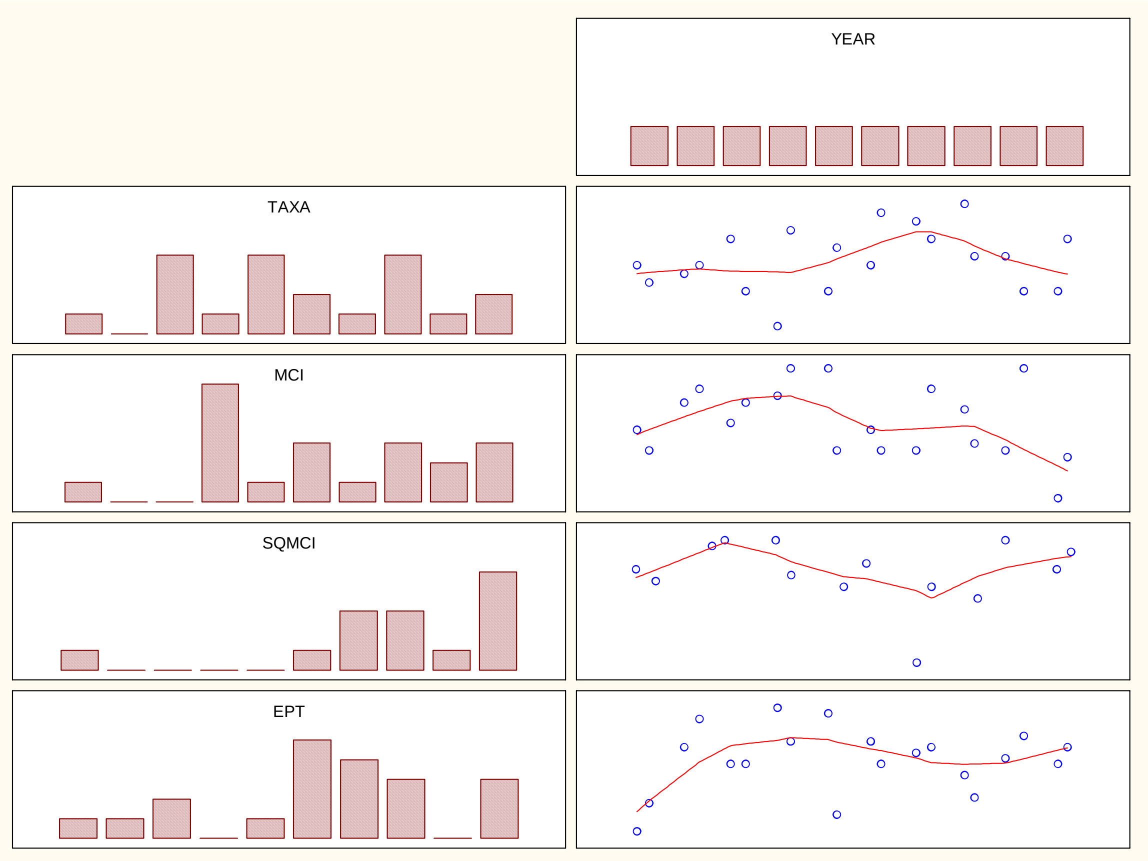
	N	Kendall	Z	p-level	p-exact
TAXA & YEAR	23	-0.164680	-1.10037	0.271173	----
MCI & YEAR	23	-0.140580	-0.93934	0.347558	----
SQMCI & YEAR	17	0.090249	0.50559	0.613144	----
EPT & YEAR	23	-0.056347	-0.37650	0.706542	----

STY000400



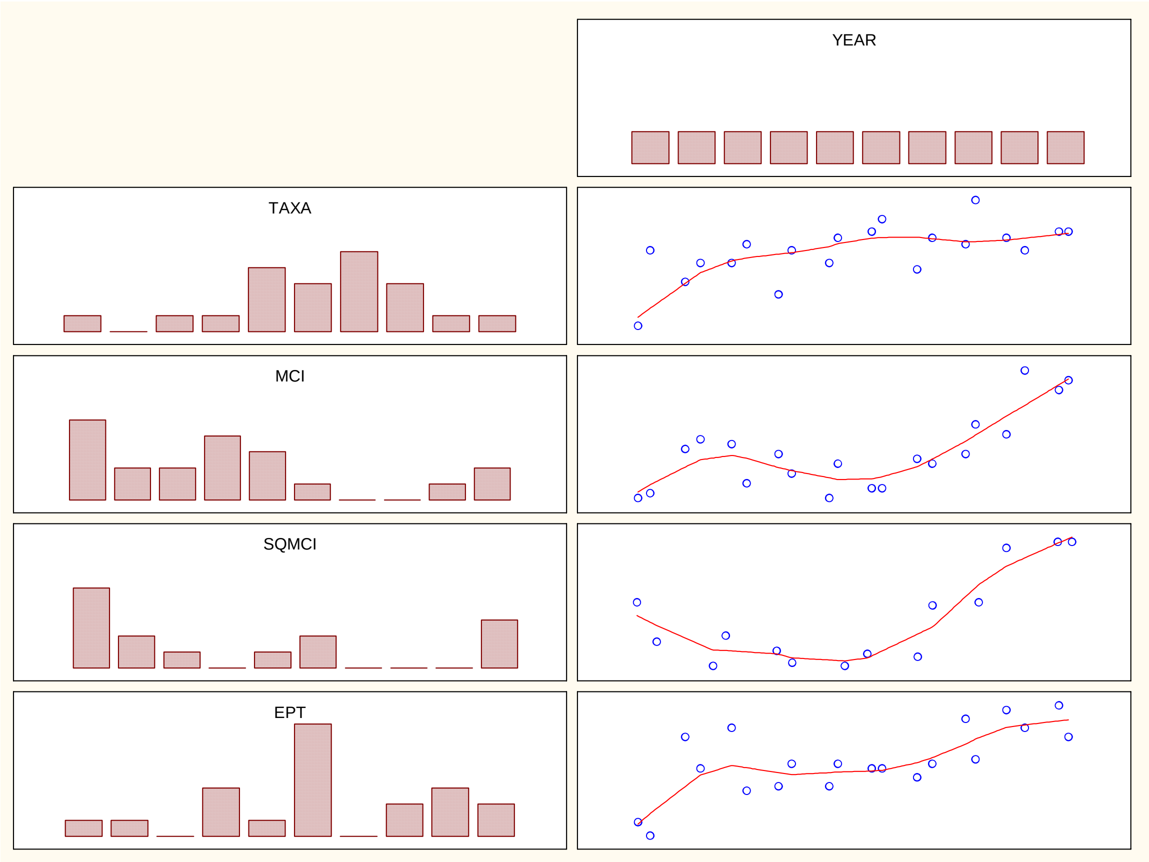
Kendall Tau Correlations MD pairwise deleted Marked correlations are significant at p <.05000					
	N	**Kendall**	**Z**	**p-level**	**p-exact**
TAXA & YEAR	23	-0.190340	-1.27182	0.203437	----
MCI & YEAR	23	-0.148011	-0.98898	0.322671	----
SQMCI & YEAR	17	-0.135373	-0.75839	0.448220	----
EPT & YEAR	23	-0.211027	-1.41005	0.158526	----

TMR000150



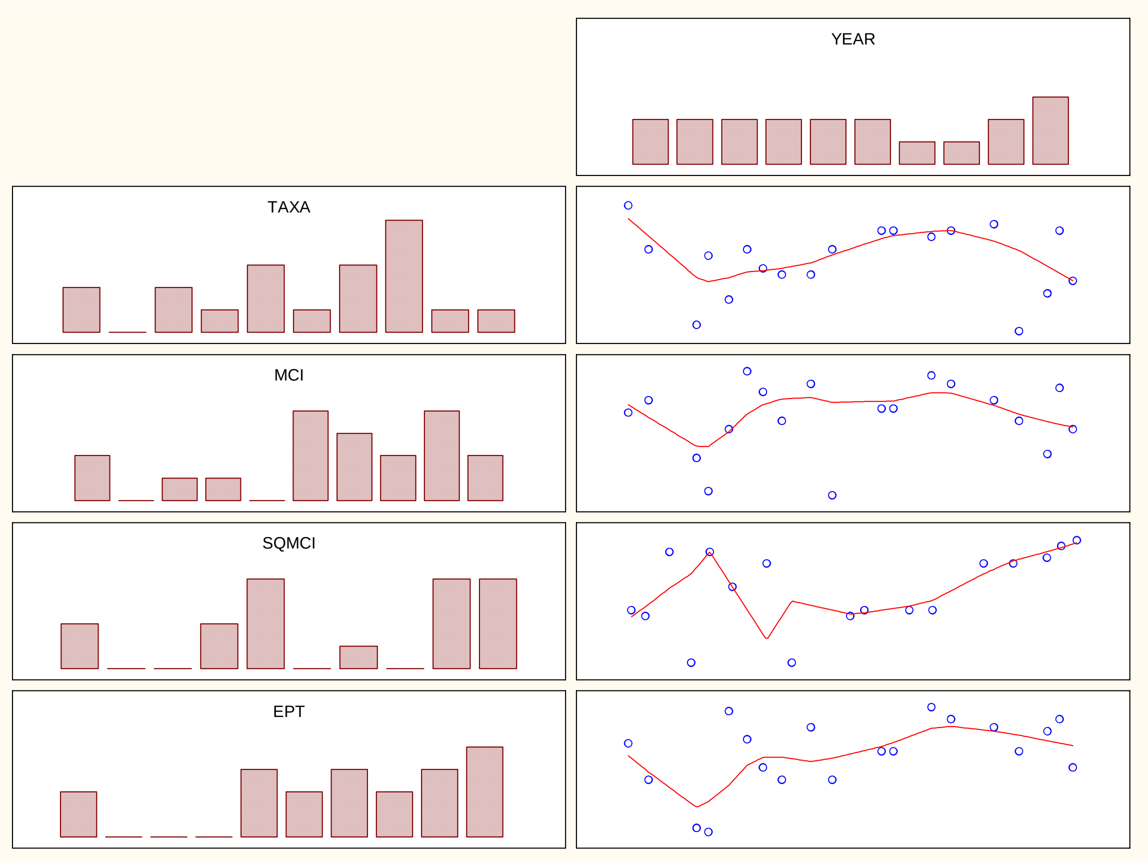
Kendall Tau Correlations MD pairwise deleted Marked correlations are significant at p <.05000					
	N	**Kendall**	**Z**	**p-level**	**p-exact**
TAXA & YEAR	20	0.158137	0.97482	0.329647	----
MCI & YEAR	20	-0.208993	-1.28832	0.197634	----
SQMCI & YEAR	14	-0.090432	-0.45051	0.652342	----
EPT & YEAR	20	0.054074	0.33333	0.738883	----

TMR000375



Kendall Tau Correlations MD pairwise deleted Marked correlations are significant at p <.05000					
	N	**Kendall**	**Z**	**p-level**	**p-exact**
TAXA & YEAR	20	0.507130	3.126161	0.001771	----
MCI & YEAR	20	0.446834	2.754469	0.005879	----
SQMCI & YEAR	14	0.402291	2.004128	0.045056	----
EPT & YEAR	20	0.512280	3.157907	0.001589	----

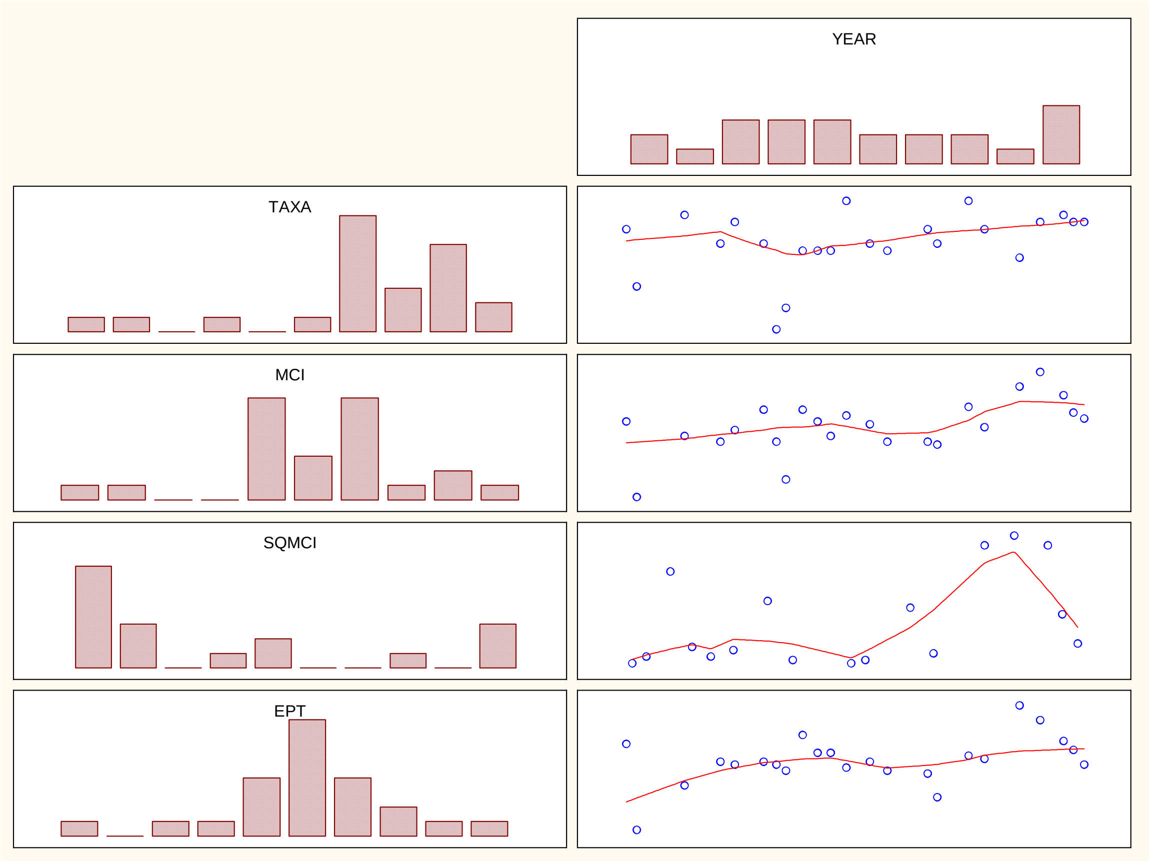
WAA000200



Kendall Tau Correlations MD pairwise deleted Marked correlations are significant at p <.05000

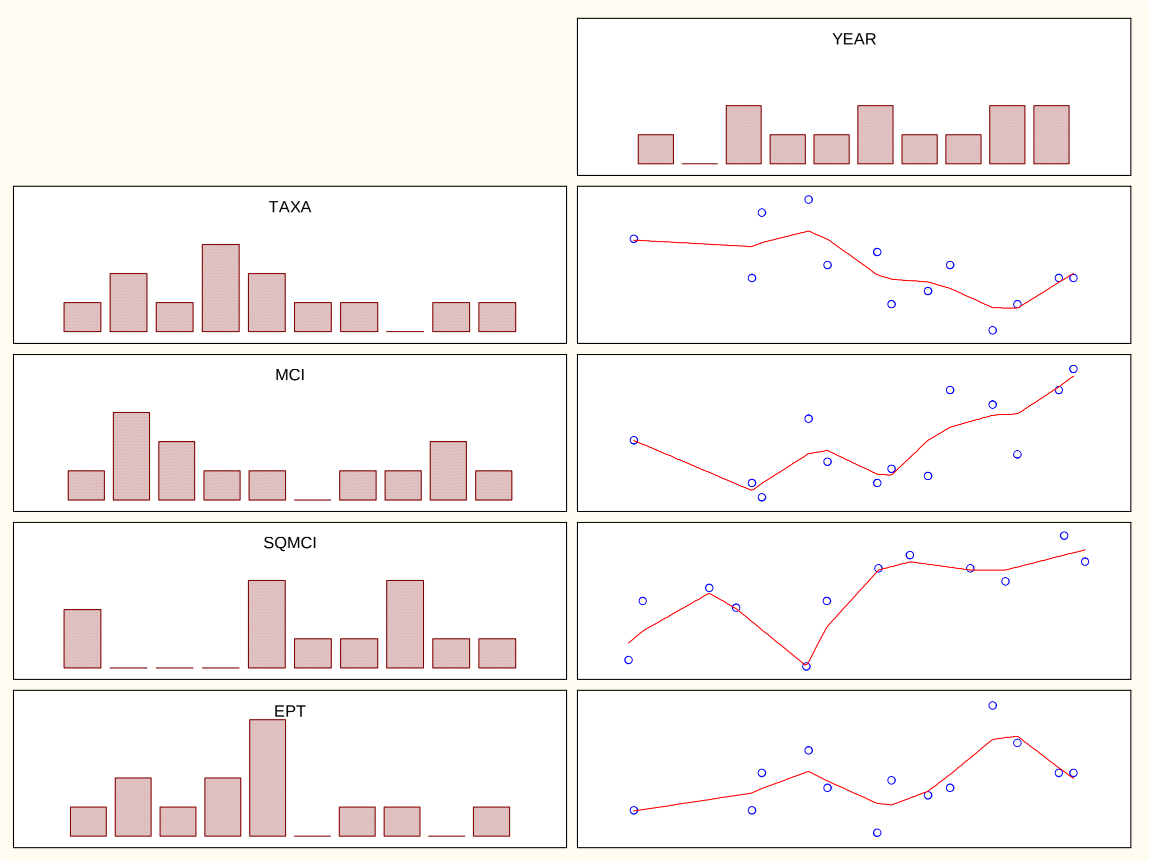
	N	Kendall	Z	p-level	p-exact
TAXA & YEAR	19	0.066295	0.396613	0.691653	----
MCI & YEAR	19	0.059354	0.355086	0.722526	----
SQMCI & YEAR	17	0.385026	2.156990	0.031006	----
EPT & YEAR	19	0.264361	1.581547	0.113753	----

WAA000447



Kendall Tau Correlations MD pairwise deleted Marked correlations are significant at p <.05000					
	N	**Kendall**	**Z**	**p-level**	**p-exact**
TAXA & YEAR	23	0.248729	1.661971	0.096519	----
MCI & YEAR	23	0.313935	2.097664	0.035935	----
SQMCI & YEAR	17	0.298541	1.672484	0.094429	----
EPT & YEAR	23	0.204846	1.368747	0.171078	----

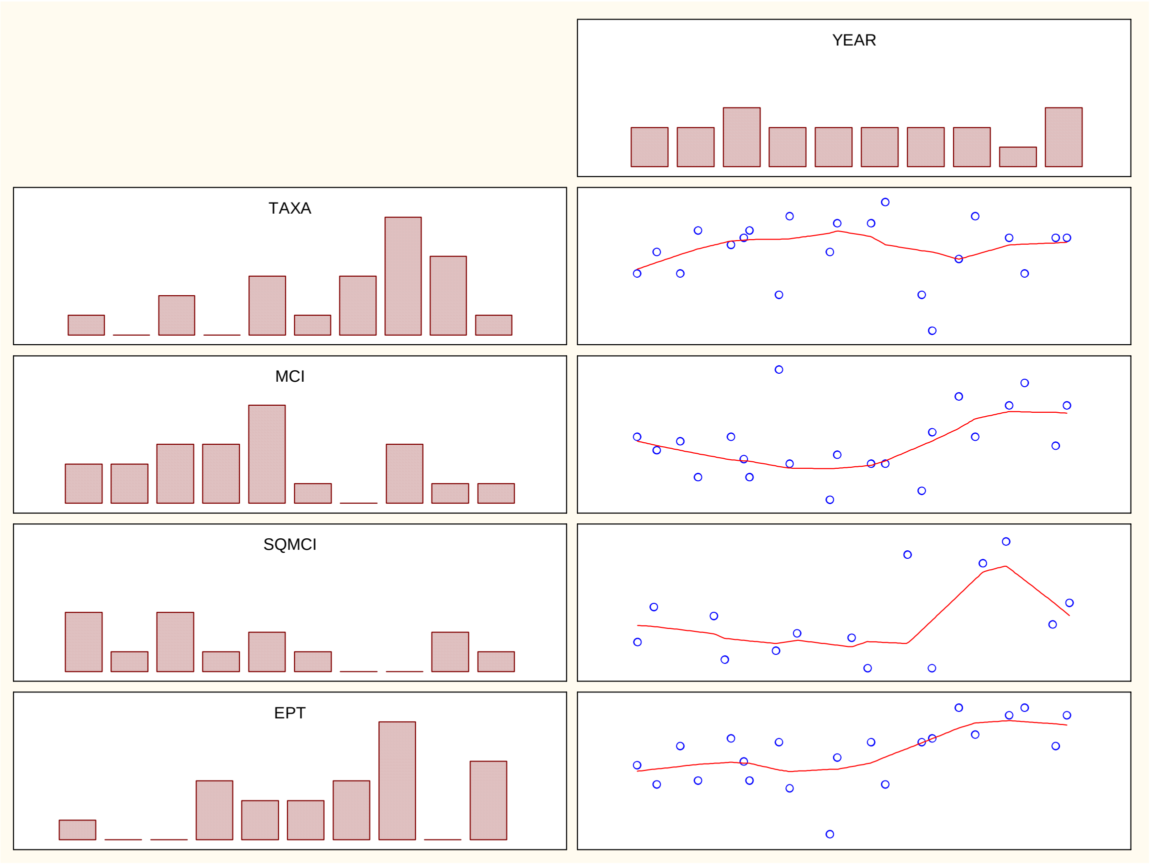
WAI000110



Kendall Tau Correlations MD pairwise deleted Marked correlations are significant at p <.05000

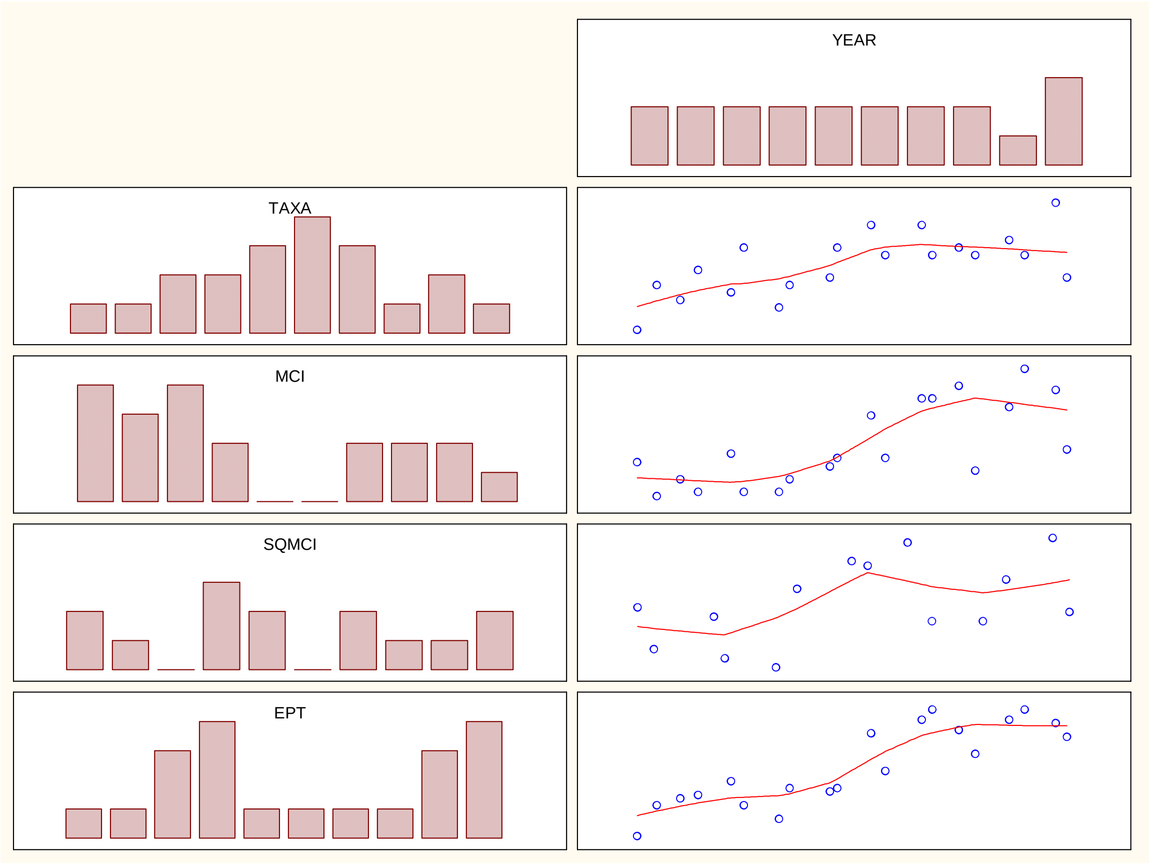
	N	Kendall	Z	p-level	p-exact
TAXA & YEAR	13	-0.410821	-1.95497	0.050586	----
MCI & YEAR	13	0.467572	2.22503	0.026079	----
SQMCI & YEAR	12	0.584685	2.64616	0.008141	----
EPT & YEAR	13	0.357812	1.70272	0.088621	----

WGA000260



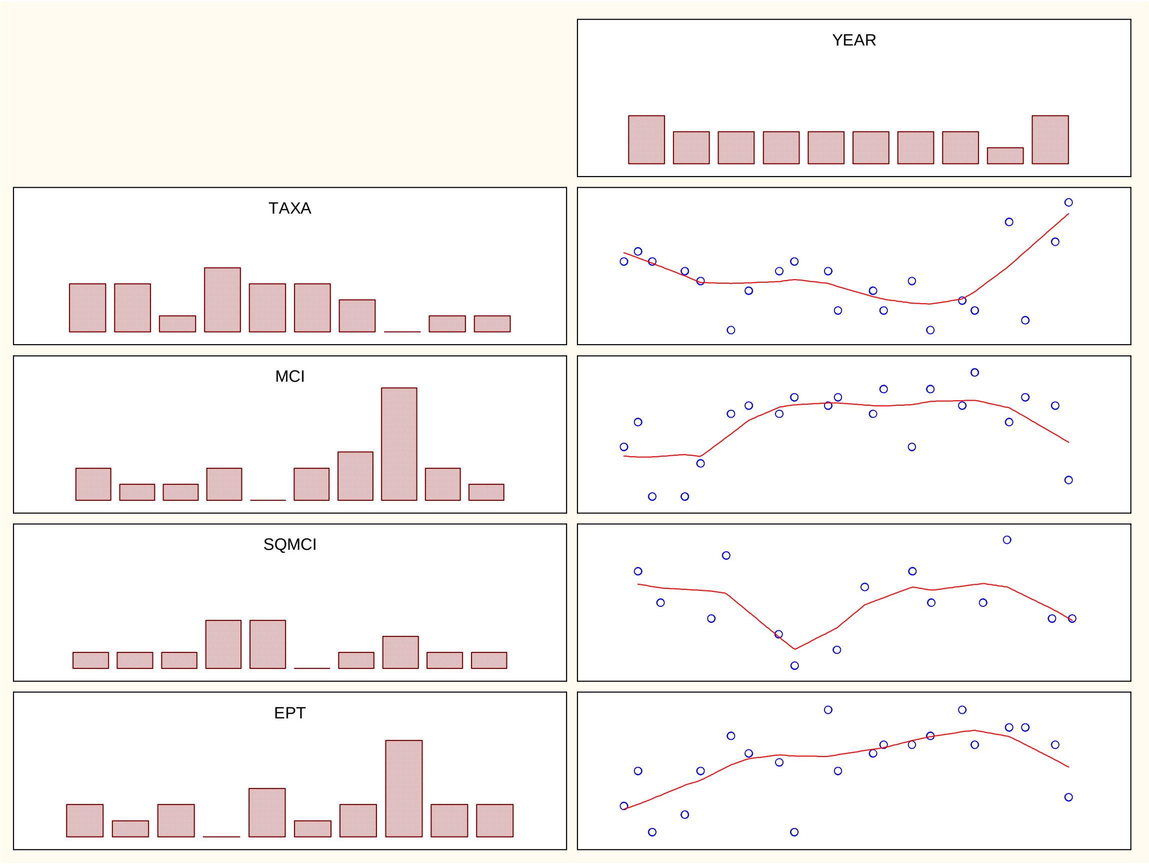
Kendall Tau Correlations MD pairwise deleted Marked correlations are significant at p <.05000					
	N	**Kendall**	**Z**	**p-level**	**p-exact**
TAXA & YEAR	21	0.098581	0.625136	0.531882	----
MCI & YEAR	21	0.203922	1.293141	0.195962	----
SQMCI & YEAR	14	0.198898	0.990867	0.321750	----
EPT & YEAR	21	0.452663	2.870498	0.004098	----

WGA000450



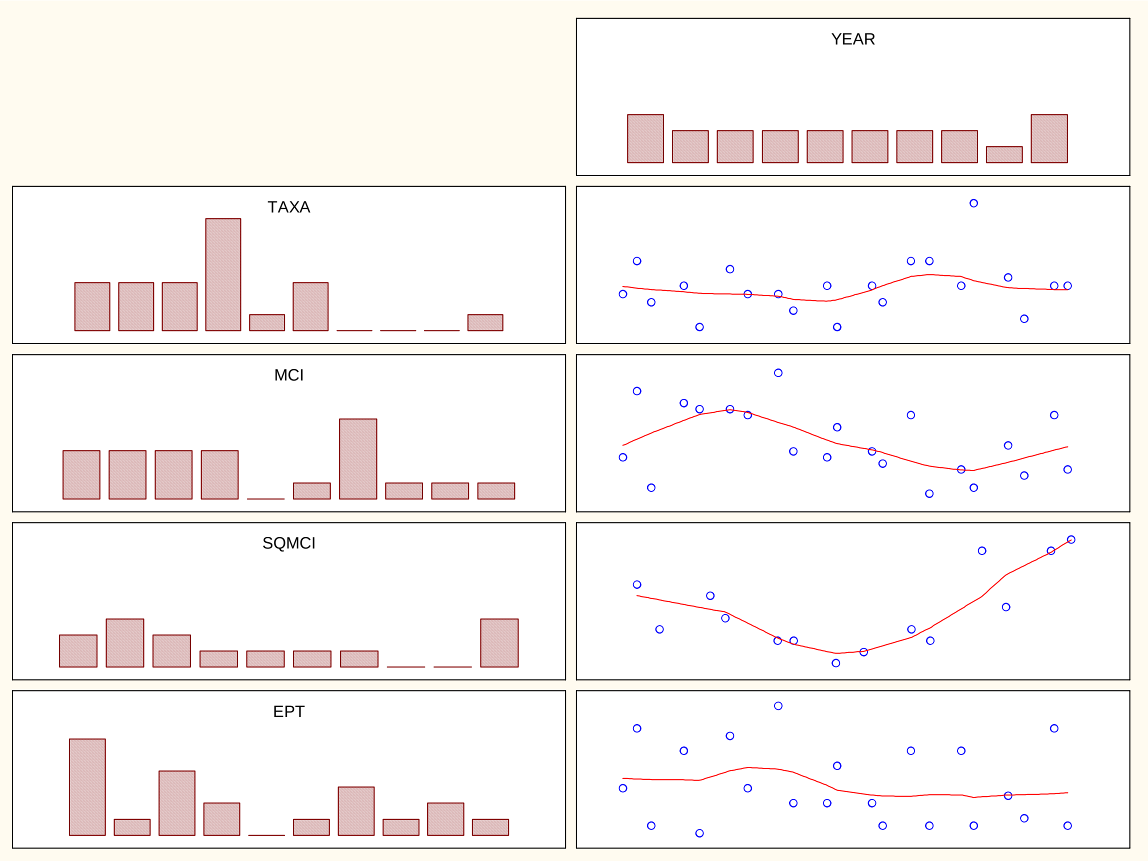
Kendall Tau Correlations MD pairwise deleted Marked correlations are significant at p <.05000					
	N	**Kendall**	**Z**	**p-level**	**p-exact**
TAXA & YEAR	20	0.467640	2.882727	0.003942	----
MCI & YEAR	20	0.577614	3.560655	0.000370	----
SQMCI & YEAR	14	0.287297	1.431253	0.152358	----
EPT & YEAR	20	0.659612	4.066120	0.000048	----

WGG000115



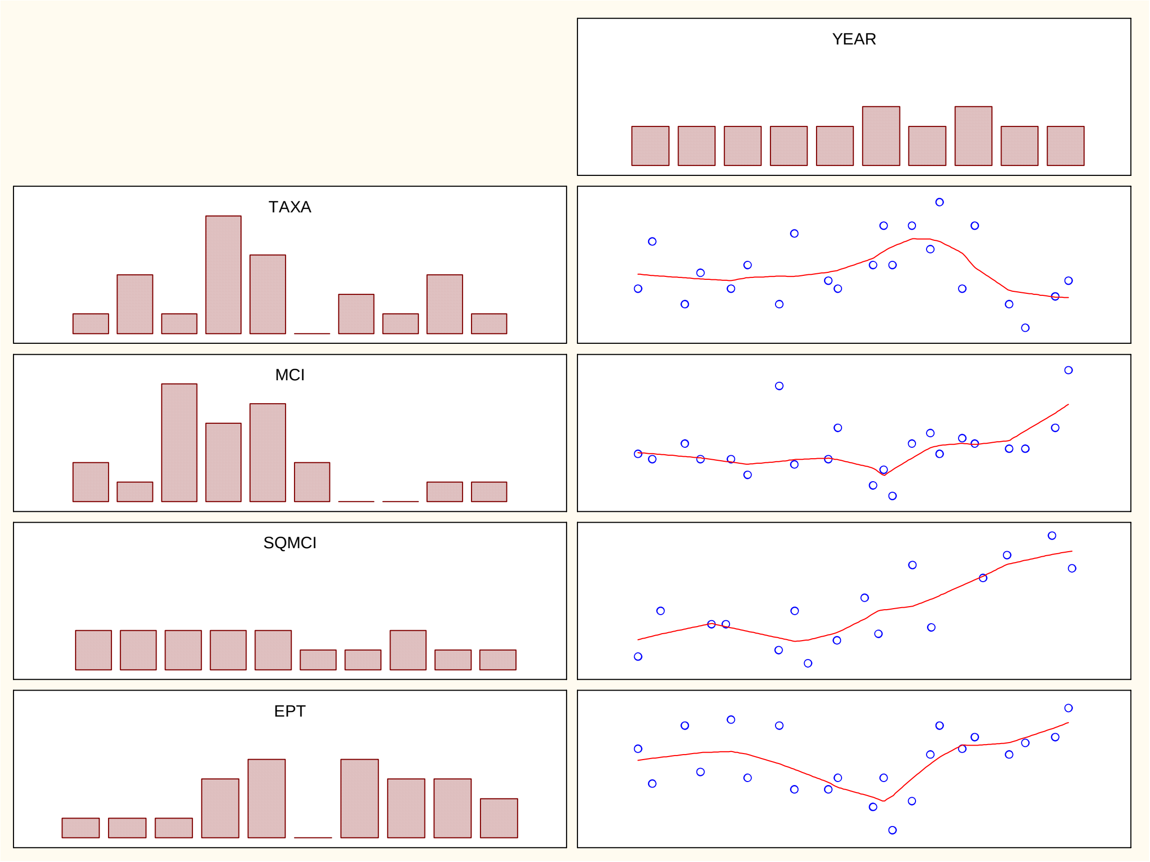
Kendall Tau Correlations MD pairwise deleted Marked correlations are significant at p <.05000					
	N	**Kendall**	**Z**	**p-level**	**p-exact**
TAXA & YEAR	21	-0.137314	-0.870759	0.383886	----
MCI & YEAR	21	0.336898	2.136389	0.032648	----
SQMCI & YEAR	14	-0.045751	-0.227921	0.819708	----
EPT & YEAR	21	0.423897	2.688083	0.007186	----

WGG000150



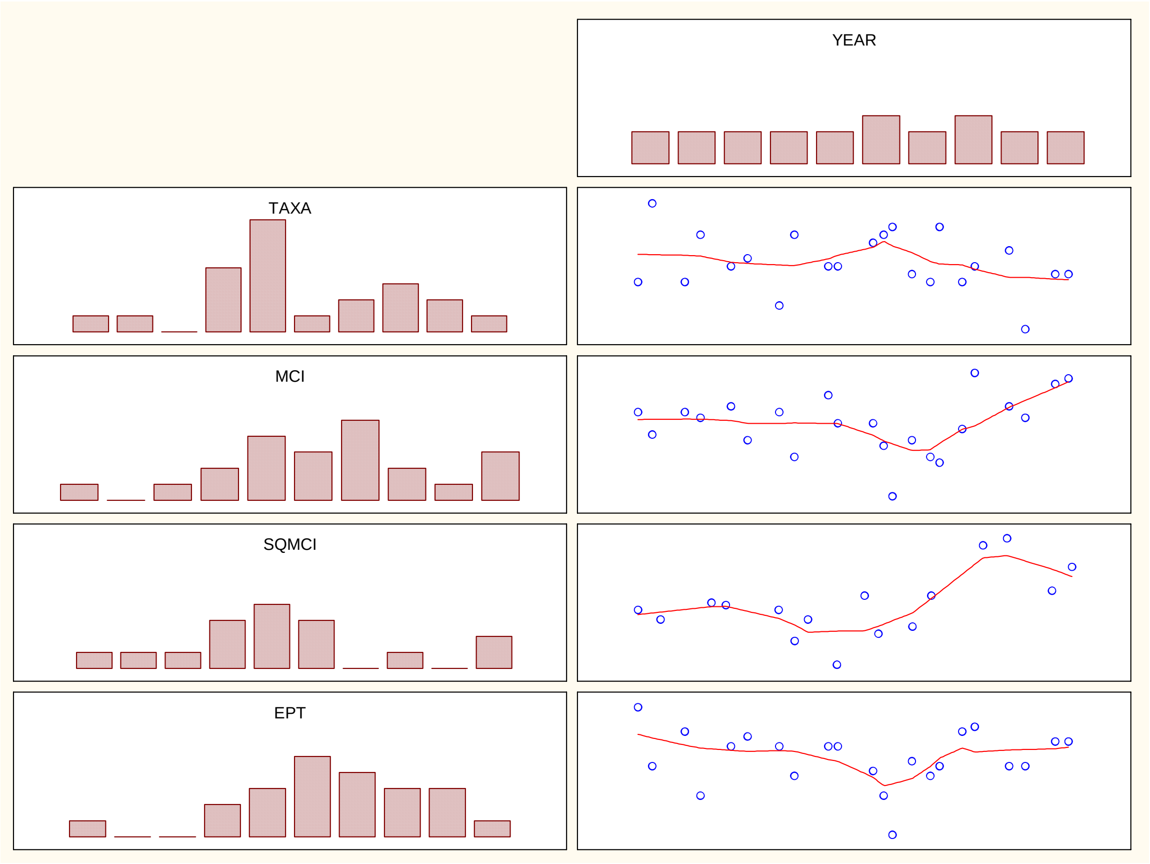
Kendall Tau Correlations MD pairwise deleted Marked correlations are significant at p <.05000	N	Kendall	Z	p-level	p-exact
TAXA & YEAR	21	0.146342	0.92800	0.353406	----
MCI & YEAR	21	-0.349580	-2.21681	0.026636	----
SQMCI & YEAR	14	0.180863	0.90102	0.367577	----
EPT & YEAR	21	-0.149404	-0.94742	0.343424	----

WGG000500



Kendall Tau Correlations MD pairwise deleted Marked correlations are significant at p <.05000					
	N	**Kendall**	**Z**	**p-level**	**p-exact**
TAXA & YEAR	22	0.040385	0.263055	0.792508	----
MCI & YEAR	22	0.244531	1.592810	0.111203	----
SQMCI & YEAR	16	0.487412	2.633343	0.008455	----
EPT & YEAR	22	0.146053	0.951352	0.341426	----

WGG000540



Kendall Tau Correlations MD pairwise deleted Marked correlations are significant at p <.05000					
	N	**Kendall**	**Z**	**p-level**	**p-exact**
TAXA & YEAR	22	-0.126527	-0.824163	0.409847	----
MCI & YEAR	22	0.066090	0.430489	0.666840	----
SQMCI & YEAR	16	0.346019	1.869439	0.061562	----
EPT & YEAR	22	-0.031410	-0.204598	0.837886	----

WGG000665

Error! Objects cannot be created from editing field codes.

Kendall Tau Correlations MD pairwise deleted Marked correlations are significant at p <.05000

	N	Kendall	Z	p-level	p-exact
TAXA & YEAR	20	-0.022389	-0.138013	0.890230	----
MCI & YEAR	20	0.186683	1.150793	0.249817	----
SQMCI & YEAR	14	0.522254	2.601759	0.009275	----
EPT & YEAR	20	0.080443	0.495885	0.619976	----

WGG000895

Error! Objects cannot be created from editing field codes.

Kendall Tau Correlations MD pairwise deleted Marked correlations are significant at p <.05000

	N	Kendall	Z	p-level	p-exact
TAXA & YEAR	21	0.224691	1.424847	0.154202	----
MCI & YEAR	21	0.144237	0.914661	0.360370	----
SQMCI & YEAR	14	0.303447	1.511710	0.130608	----
EPT & YEAR	21	0.093414	0.592370	0.553603	----

WGG000995

Error! Objects cannot be created from editing field codes.

Kendall Tau Correlations MD pairwise deleted Marked correlations are significant at p <.05000

	N	Kendall	Z	p-level	p-exact
TAXA & YEAR	20	0.488084	3.00875	0.002623	----
MCI & YEAR	20	0.080007	0.49320	0.621873	----
SQMCI & YEAR	14	0.331497	1.65145	0.098648	----
EPT & YEAR	20	-0.177950	-1.09696	0.272660	----

WKH000100

Error! Objects cannot be created from editing field codes.

Kendall Tau Correlations MD pairwise deleted Marked correlations are significant at p <.05000

	N	Kendall	Z	p-level	p-exact
TAXA & YEAR	6	-1.00000	-2.81801	0.004832	.001
MCI & YEAR	6	0.20000	0.56360	0.573025	.360
SQMCI & YEAR	6	0.41404	1.16677	0.243305	.235
EPT & YEAR	6	0.00000	0.00000	1.000000	----

WKH000500

Error! Objects cannot be created from editing field codes.

Kendall Tau Correlations MD pairwise deleted Marked correlations are significant at p <.05000

	N	Kendall	Z	p-level	p-exact
TAXA & YEAR	20	0.124717	0.768805	0.442009	----
MCI & YEAR	20	0.170222	1.049321	0.294030	----
SQMCI & YEAR	14	0.441996	2.201928	0.027670	----
EPT & YEAR	20	0.166249	1.024828	0.305444	----

WKH000920

Error! Objects cannot be created from editing field codes.

Kendall Tau Correlations MD pairwise deleted Marked correlations are significant at p <.05000

	N	Kendall	Z	p-level	p-exact
TAXA & YEAR	21	0.151644	0.961629	0.336236	----
MCI & YEAR	21	0.378907	2.402781	0.016271	----
SQMCI & YEAR	14	0.077783	0.387496	0.698389	----
EPT & YEAR	21	0.461560	2.926914	0.003423	----

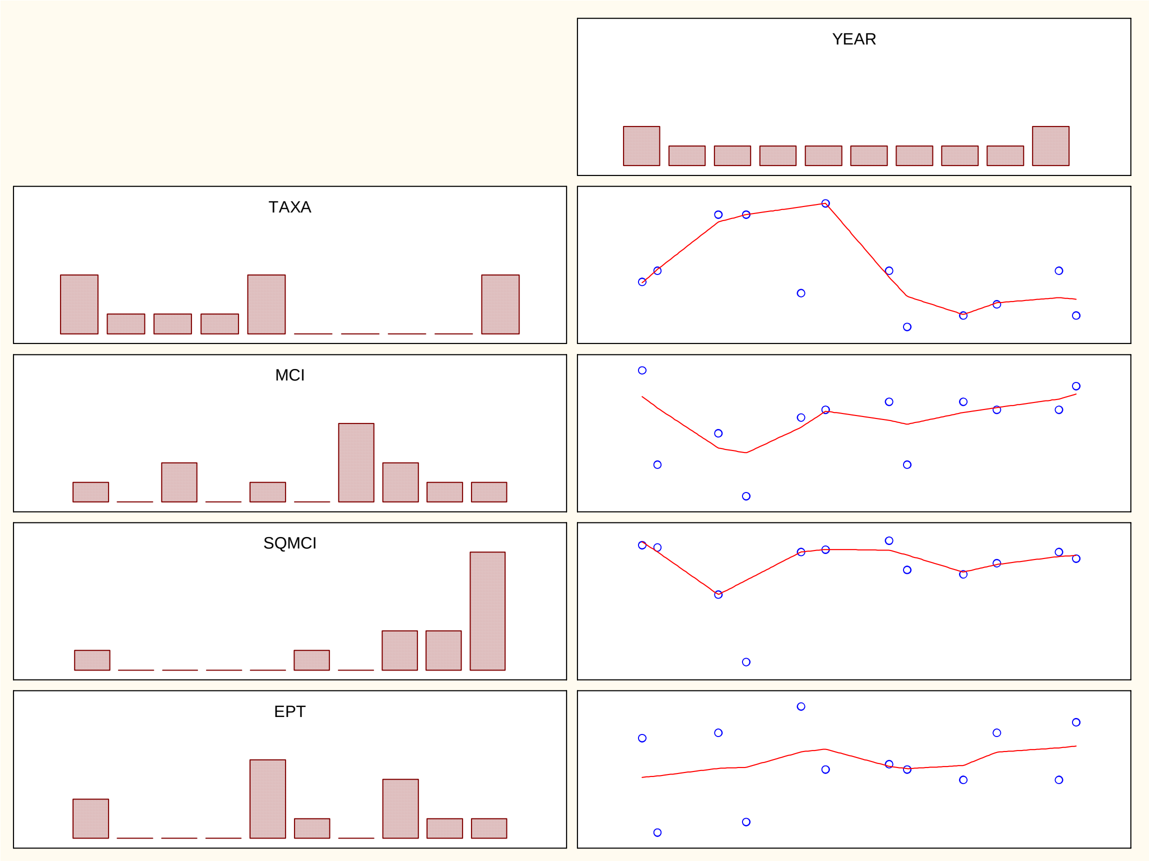
WKH000950

Error! Objects cannot be created from editing field codes.

Kendall Tau Correlations MD pairwise deleted Marked correlations are significant at p <.05000

	N	Kendall	Z	p-level	p-exact
TAXA & YEAR	19	0.292021	1.747027	0.080633	----
MCI & YEAR	19	0.370595	2.217095	0.026617	----
SQMCI & YEAR	14	0.056194	0.279946	0.779519	----
EPT & YEAR	19	0.492671	2.947419	0.003204	----

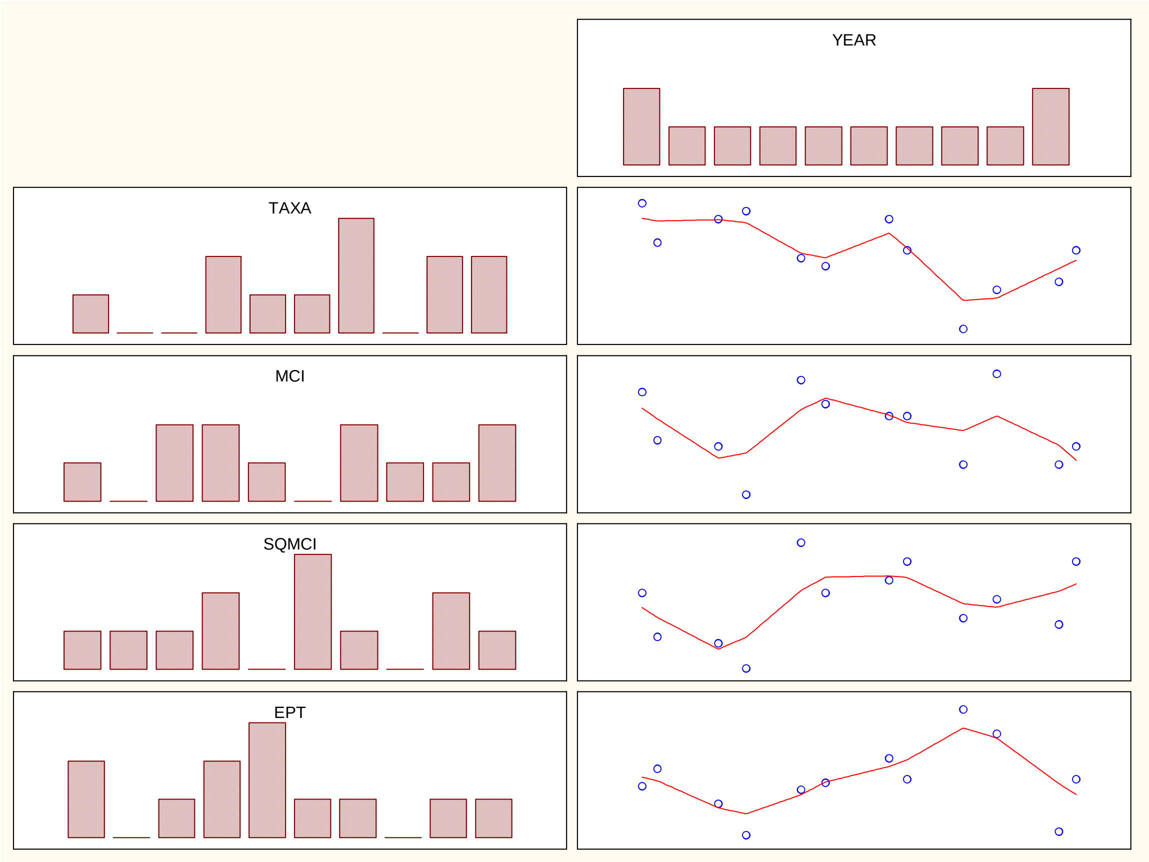
WMK000100



Kendall Tau Correlations MD pairwise deleted Marked correlations are significant at p <.05000

	N	Kendall	Z	p-level	p-exact
TAXA & YEAR	12	-0.299445	-1.35522	0.175346	----
MCI & YEAR	12	0.299445	1.35522	0.175346	----
SQMCI & YEAR	12	-0.106873	-0.48369	0.628609	----
EPT & YEAR	12	0.108556	0.49130	0.623212	----

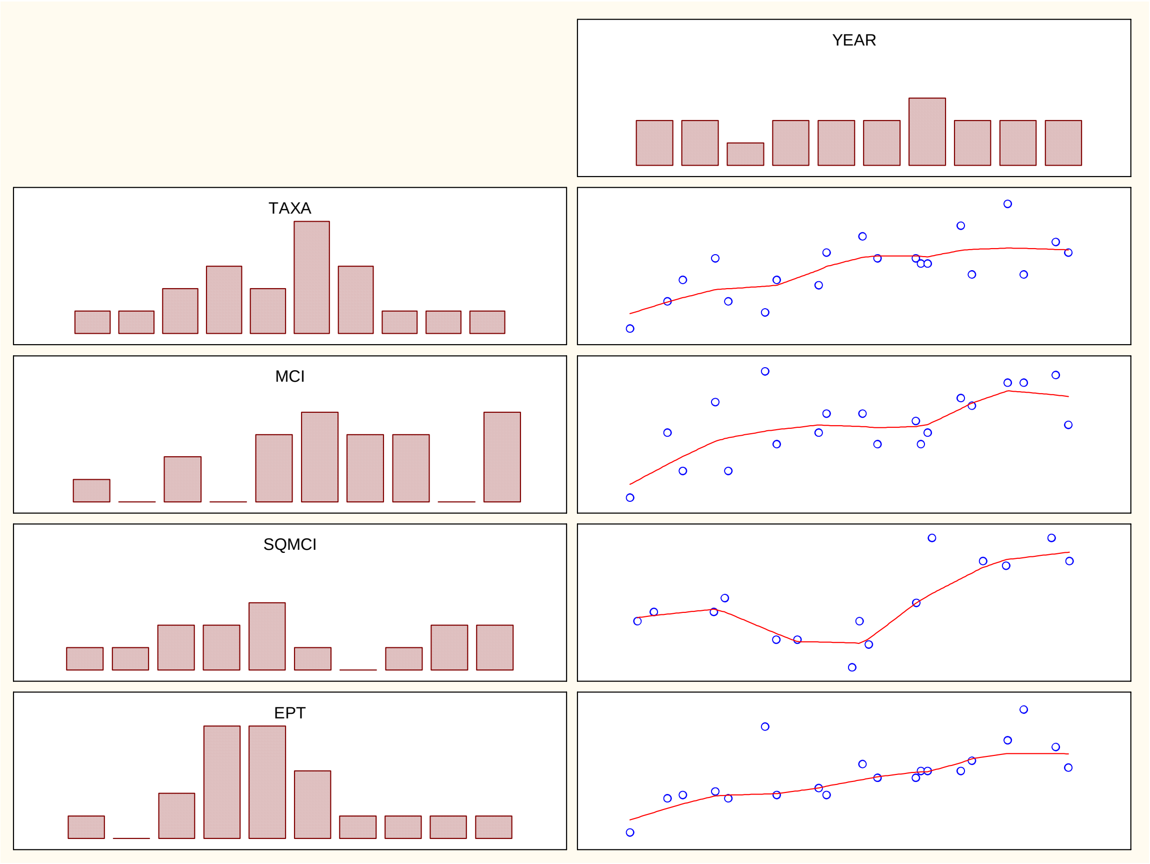
WMK000298



Kendall Tau Correlations MD pairwise deleted Marked correlations are significant at p <.05000

	N	Kendall	Z	p-level	p-exact
TAXA & YEAR	12	-0.492366	-2.22834	0.025858	----
MCI & YEAR	12	-0.139573	-0.63168	0.527598	----
SQMCI & YEAR	12	0.153864	0.69636	0.486205	----
EPT & YEAR	12	0.259550	1.17467	0.240128	----

WTR000850



Kendall Tau Correlations MD pairwise deleted Marked correlations are significant at p <.05000					
	N	**Kendall**	**Z**	**p-level**	**p-exact**
TAXA & YEAR	20	0.462473	2.850872	0.004360	----
MCI & YEAR	20	0.426001	2.626049	0.008638	----
SQMCI & YEAR	15	0.390360	2.028370	0.042522	----
EPT & YEAR	20	0.656066	4.044261	0.000052	----

Appendix 2 Spearman rank probability values for taxa richness, MCI, SQMCI, and %EPT$_{taxa}$ with probabilities deemed significant using the False Discovery Rate highlighted with ***. The critical probability cutoff level is 0.00654.

Site	Record	TAXA	MCI	SQMCI	EPT
HTK000350	1	0.092089	0.010554	0.01419	0.019393
HTK000425	2	0.029264	0.015012	0.000025***	0.125187
HTK000745	3	0.001275***	0.026926	0.535498	0.477096
INH000400	4	0.337077	0.049966	0.101033	0.424424
KPA000250	5	0.925517	0.314636	0.033717	0.508268
KPA000700	6	0.563588	0.014678	0.027196	0.000184***
KPA000950	7	0.353442	0.214549	0.819304	0.016549
KPK000250	8	0.912575	0.894445	0.841537	0.269971
KPK000500	9	0.638874	0.11425	0.113869	0.468996
KPK000660	10	0.346926	0.001025***	0.241948	0.020374
KPK000880	11	0.881173	0.032991	0.003642***	0.153096
KPK000990	12	0.147333	0.18695	0.001585***	0.038434
KPN000275	13	0.006453***	0.574465	0.000641***	0.355245
KPN000360	14	0.373189	0.26978	0.000413***	0.398949
KRP000300	15	0.054344	0.001811***	0.992774	0.120305
KRP000660	16	0.000060***	0.000157***	0.996402	0.000001***
KTK000150	17	0.296137	0.374677	0.934611	0.873907
KTK000248	18	0.617752	0.001662***	0.000055***	0.080667
MGE000970	19	0.038177	0.867934	0.867934	0.088347
MGH000950	20	0.002650***	0.000202***	0.012049	0.033407
MGN000195	21	0.481332	0.940748	0.762852	0.996995
MGN000427	22	0.623926	0.963284	0.542247	0.730166
MGO000050	23	0.002966***	0.698985	0.12565	0.647349
MGO000190	24	0.205371	0.843348	0.514731	0.678259
MGT000488	25	0.017495	0.349927	0.23155	0.188681
MGT000520	26	0.000074***	0.027394	0.073794	0.678794
MHW000060	27	0.397503	0.807225	0.921731	0.745897
MKW000200	28	0.881792	0.645465	0.786682	0.288802
MKW000300	29	0.551325	0.639175	0.585534	0.275023
MRK000420	30	0.007514	0.008535	0.01198	0.012824
MWH000380	31	0.860731	0.168058	0.645514	0.722036
MWH000490	32	0.090256	0.069119	0.01399	0.484168
PAT000200	33	0.55721	0.531337	0.231268	0.47895
PAT000315	34	0.095434	0.72727	0.725774	0.873501
PAT000360	35	0.010477	0.13446	0.544516	0.06884
PNH000200	36	0.772604	0.366842	0.280061	0.07542
PNH000900	37	0.960573	0.083218	0.02422	0.692231
STY000300	38	0.262184	0.246308	0.491287	0.557604
STY000400	39	0.22541	0.253366	0.436299	0.130336
TMR000150	40	0.307853	0.185989	0.736049	0.784614
TMR000375	41	0.000612***	0.006544***	0.03847	0.003004***
WAA000200	42	0.98991	0.791261	0.066908	0.146508
WAA000447	43	0.109161	0.022245	0.076093	0.133018
WAI000110	44	0.025692	0.024619	0.003281***	0.069258
WGA000260	45	0.572768	0.153131	0.245829	0.004159***
WGA000450	46	0.003293***	0.000329***	0.160488	0.000004***
WGG000115	47	0.643203	0.037418	0.924482	0.013533
WGG000150	48	0.444832	0.025347	0.343263	0.27833
WGG000500	49	0.904669	0.06637	0.006089***	0.386217
WGG000540	50	0.491936	0.603738	0.027479	0.788082
WGG000665	51	0.970527	0.351433	0.004663***	0.846712
WGG000895	52	0.170672	0.711332	0.214706	0.787416
WGG000995	53	0.011641	0.682456	0.190158	0.291934
WKH000100	54	0.001249***	1	0.428188	0.953117
WKH000500	55	0.436305	0.574039	0.023891	0.381302
WKH000920	56	0.495491	0.027078	0.818316	0.004210***
WKH000950	57	0.065794	0.048196	0.603407	0.003494***
WMK000100	58	0.102793	0.413965	0.530174	0.71721
WMK000298	59	0.013295	0.688481	0.353971	0.294223
WTR000850	60	0.002900***	0.034076	0.020011	0.000131***

The FDR.exe programme (written by Ian Jowett & Graham McBride of NIWA) that produced the above output requires two input files (sample size file & coefficients file). The shaded column of site codes is <u>not</u> part of the input file for the FDR analysis.

Sample size file

Site	TAXA	MCI	SQMCI	EPT
HTK000350	18	18	14	18
HTK000425	18	18	14	18
HTK000745	18	18	14	18
INH000400	33	33	21	33
KPA000250	12	12	11	12
KPA000700	12	12	11	12
KPA000950	12	12	11	12
KPK000250	13	13	12	13
KPK000500	16	16	13	16
KPK000660	20	20	14	20
KPK000880	20	20	14	20
KPK000990	12	12	12	12
KPN000275	38	38	25	38
KPN000360	28	28	25	28
KRP000300	21	21	13	21
KRP000660	21	21	13	21
KTK000150	12	12	12	12
KTK000248	11	11	11	11
MGE000970	6	6	6	6
MGH000950	20	20	14	20
MGN000195	22	22	16	22
MGN000427	20	20	14	20
MGO000050	20	20	14	20
MGO000190	20	20	14	20
MGT000488	21	21	14	21
MGT000520	21	21	14	21
MHW000060	20	20	14	20
MKW000200	11	11	10	11
MKW000300	10	10	10	10
MRK000420	20	20	14	20
MWH000380	20	20	14	20
MWH000490	20	20	14	20
PAT000200	20	20	14	20
PAT000315	20	20	14	20
PAT000360	20	20	14	20
PNH000200	20	20	14	20
PNH000900	20	20	14	20
STY000300	23	23	17	23
STY000400	23	23	17	23
TMR000150	20	20	14	20
TMR000375	20	20	14	20
WAA000200	19	19	17	19
WAA000447	23	23	17	23
WAI000110	13	13	12	13
WGA000260	21	21	14	21
WGA000450	20	20	14	20
WGG000115	21	21	14	21
WGG000150	21	21	14	21
WGG000500	22	22	16	22
WGG000540	22	22	16	22
WGG000665	20	20	14	20
WGG000895	21	21	14	21
WGG000995	20	20	14	20
WKH000100	6	6	6	6
WKH000500	20	20	14	20
WKH000920	21	21	14	21
WKH000950	19	19	14	19
WMK000100	12	12	12	12
WMK000298	12	12	12	12
WTR000850	20	20	15	20

Coefficients file (contains Spearman rank correlations between year and index value)

The shaded column of site codes is not part of the input file for the FDR analysis. The sign of the correlation indicates whether the trend is positive or negative.

Site	TAXA	MCI	SQMCI	EPT
HTK000350	0.408816	0.586279	0.637479	0.544792
HTK000425	-0.51356	0.56289	0.885949	0.375
HTK000745	0.69805	0.520108	0.181112	0.179073
INH000400	-0.1725	0.344005	0.367698	0.14387
KPA000250	0.030303	0.317473	0.64058	0.212019
KPA000700	0.1856	0.681426	0.659729	0.876347
KPA000950	0.294118	0.386527	0.078166	0.672598
KPK000250	-0.03385	-0.04091	-0.06475	-0.33056
KPK000500	-0.12715	0.410505	0.45985	0.195104
KPK000660	0.22197	0.67775	0.334817	0.514188
KPK000880	-0.03571	0.478098	0.720628	0.331693
KPK000990	-0.44484	0.408855	0.80497	0.601778
KPN000275	-0.43421	0.094024	0.635518	0.15422
KPN000360	0.174965	0.215927	0.652045	0.16586
KRP000300	0.42571	0.639216	0.002793	0.349607
KRP000660	0.762046	0.73292	0.001391	0.847312
KTK000150	-0.32916	-0.28192	0.026596	-0.05142
KTK000248	-0.16977	0.827641	0.921698	0.54841
MGE000970	-0.83591	0.088273	0.088273	0.746352
MGH000950	0.634607	0.738277	0.64889	0.477109
MGN000195	-0.15842	-0.01683	-0.08196	0.000853
MGN000427	0.116776	-0.011	-0.17817	0.082291
MGO000050	0.629041	-0.09222	-0.42921	-0.109
MGO000190	0.295833	0.047203	-0.19025	-0.0989
MGT000488	0.512621	-0.21473	-0.34187	-0.29853
MGT000520	0.755723	0.480706	0.492205	0.096037
MHW000060	0.200154	0.058267	0.028954	-0.07733
MKW000200	-0.05093	-0.15668	0.098466	-0.35173
MKW000300	-0.21474	-0.16976	-0.19693	0.382723
MRK000420	0.578679	0.571105	0.649282	0.545662
MWH000380	-0.04191	0.320671	0.134959	-0.08487
MWH000490	0.388764	0.414616	-0.63849	0.166038
PAT000200	-0.1396	0.148767	0.342065	0.167998
PAT000315	0.383128	-0.08321	-0.10311	0.038038
PAT000360	0.558535	0.346524	0.17719	0.414994
PNH000200	-0.06898	-0.21318	0.310424	-0.40636
PNH000900	0.011814	0.396819	0.596908	0.09439
STY000300	-0.24384	-0.25187	0.17922	-0.12895
STY000400	-0.26296	-0.24826	-0.20224	-0.32493
TMR000150	0.240123	-0.30832	-0.09911	0.065251
TMR000375	0.698561	0.586703	0.557162	0.628409
WAA000200	-0.00311	0.065074	0.454375	0.346206
WAA000447	0.342939	0.474222	0.441424	0.322798
WAI000110	-0.61367	0.617204	0.771692	0.518829
WGA000260	0.130535	0.323087	0.332231	0.59843
WGA000450	0.623769	0.721634	0.396462	0.835665
WGG000115	-0.10737	0.456696	0.027933	0.529665
WGG000150	0.17621	-0.48644	0.273942	-0.24803
WGG000500	0.027112	0.398291	0.653064	0.19431
WGG000540	-0.15466	0.117112	0.549455	-0.06081
WGG000665	-0.00883	0.219955	0.707324	0.046177
WGG000895	0.310531	0.085864	0.353727	0.062623
WGG000995	0.551886	0.097543	0.372099	-0.24792
WKH000100	-0.97101	0	0.40303	0.031265
WKH000500	0.184441	0.133737	0.598016	0.206962
WKH000920	0.157447	0.481568	0.067628	0.597782
WKH000950	0.430477	0.45876	0.152223	0.634927
WMK000100	-0.49376	0.26025	-0.20142	0.117022
WMK000298	-0.68851	-0.12943	0.293809	0.330397
WTR000850	0.630163	0.47554	0.592228	0.752085