Qual Quant (2009) 43:913–940
DOI 10.1007/s11135-008-9168-9

ORIGINAL PAPER

A drunk and her dog: a spurious relation? Cointegration tests as instruments to detect spurious correlations between integrated time series

Esther Stroe-Kunold · Joachim Werner

Published online: 9 April 2008

Abstract A significant correlation between integrated time series does not necessarily imply a meaningful relation. The relation can also be meaningless, i.e. spurious. Cointegration is sometimes illustrated by the metaphor of 'a drunk and her dog'. The relation between integrated processes is meaningful, if they are cointegrated. To prevent spurious correlations, integrated series are usually transformed. This implies a loss of information. In case of cointegration, these transformations are no longer necessary. Moreover, it can be shown that cointegration tests are instruments to detect spurious correlations between integrated time series. This paper compares the Dickey–Fuller and the Johansen cointegration test. By means of Monte Carlo simulations, we found that these cointegration tests are a much more accurate alternative for the identification of spurious relations compared to the rather imprecise method of utilizing the R^2-and *DW*-statistics recommended by some authors. Furthermore, we demonstrate that cointegration techniques are precise methods of distinguishing between spurious and meaningful relations even if the dependency between the processes is very low. Using these tests, the researcher is not in danger of either neglecting a small but meaningful relation or regarding a relation as meaningful which is actually spurious.

Keywords Spurious correlation · Spurious regression · Cointegration · Multivariate time series analysis · Psychological process research · Longitudinal analysis · Stationarity · Monte Carlo experiments

The question whether a correlation is spurious or meaningful has a high relevance for social and behavioral sciences. In psychology, in particular, many of the postulated causal or feedback mechanisms reside within the organism. Internal processes such as the development of intellectual abilities, personality traits, and emotional skills are cases in point (Haig 2003). For decades, spurious correlations have been "fallacies of statistical method in psychological measurement" (Peatman 1937).

E. Stroe-Kunold · J. Werner (✉)
Department of Psychology, University of Heidelberg, Hauptstr. 47-51, 69117 Heidelberg, Germany
e-mail: Joachim.Werner@psychologie.uni-heidelberg.de

By analyzing only differenced time series, all information about potential (long-run) relationships between time variables is lost. This is an inappropriate solution to possible spurious correlations (Hendry 1986, p. 204). In case of *cointegration*—sometimes explained by the metaphor of 'a drunk and her dog' (Murray, 1994; see further explanations below)—differencing of the time series is no longer necessary. Some authors postulate that cointegration models can be regarded "as remedies to the problems of 'spurious regression' arising from non-stationary time series" (Lin and Brannigan 2003, Abstract). This paper investigates whether a test for cointegration can be thought of "as a pre-test to avoid 'spurious regression' situations" (Granger 1986, p. 226).

The aims of this paper are: (a) introduce to the problem of spurious regression and the concept of cointegration, (b) clarify the relationship between both concepts, (c) find out in which way cointegration tests can contribute to the indication of spurious correlations (by means of Monte Carlo simulations), and (d) point out that the application of cointegration methods allows interesting conclusions about the psychological processes under investigation.

1 Time series analysis as a tool of psychological process research

1.1 Psychological process research

Humans experience time through changes in their environment. Changes become obvious through a sequence of events separated by intervals (Rinne and Specht 2002, p. 4). The term 'process' stands for a development over time, gradually approaching a certain state. *Process research* has a long history in psychology. Lasswell, one of the early pioneers of empirical process research, investigated the development of conscious and unconscious affects of subjects participating in psychoanalytic discourses (Lasswell 1937).

In contrast to *episodic research*, process research focuses on the investigation of temporal development as a whole. A therapy process, for instance, can be regarded as a *time series*. Since the end of the 1970s, time series analysis has been increasingly utilized as a research procedure appropriate for evaluating psychotherapy (see Glass et al. 1975, *inter alia*). As many examples demonstrate, the psychological process perspective is not restricted to processes of psychotherapy (see Stroe-Kunold and Werner 2007).

Cointegration methodology enables modeling human short- and long-term dynamics (Stroe-Kunold and Werner, in press). Before investigating the benefits of this method as a cure for spurious correlations, a short introduction to time series analysis is given.

1.2 Introduction to time series analysis

A time series is a sequence of repeated measurements on the same object. This leads to dependent measurements. Conventional statistical algorithms are based on the assumption of independent measurements and are thus no longer appropriate.

In the time-domain, processes can be modeled through ARIMA (Autoregressive Integrated Moving Average) (p, d, q)-models (Box and Jenkins 1976). This article refers only to AR(p)-processes and their multivariate counterpart (vector autoregression).

Autoregression means that the time series falls back on itself in each point of time t. An AR(p)-process is written

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + \varepsilon_t$$

where ϕ quantifies the magnitude of dependence; p specifies the order of dependence; ε_t is a sequence of purely independent and identically distributed random variables (*innovations*).

Within the scope of ARIMA-modeling, *stationarity* is the most important and problematic assumption necessary for the estimation of process parameters. (Weak) stationarity is given if the first moments (mean, variance/covariance) of the generating process show stability over time. The most simple stationary process is known as *white noise*. *Differencing* is the procedure of eliminating *stochastic drifts* in order to make a process stationary. An instationary time series x_t without deterministic components is *integrated* of order d, i.e. $x_t \sim I(d)$, if it can be transformed into a stationary process through differencing d-times. For the first differencing ($d = 1$), we write: $\Delta x_t = x_t - x_{t-1}$. Thus, an $I(1)$-process stands for an instationary, an $I(0)$-process for a stationary series. Usually, the literature refers to 'integrated' as 'instationary'. The influence of *random shocks* on these *difference-stationary* (*DS*-) processes does *not* dissipate over time leading to their instability. They have d unit roots and their variance tends to infinity for $t \to \infty$. *Random walks* as ARIMA(0,1,0)-processes have an infinite memory and stand for a very common case of instationarity.

A variety of psychological processes is integrated: based on a large empirical study, Glass et al. (1975, pp. 115–117) found that 44 of 95 (i.e. approx. 46%) of psychological time series were difference-stationary (38 $I(1)$- and 6 $I(2)$-processes). In their recent studies, Fortes et al. (2004) and Ninot et al. (2005) analyzed 50 time series reflecting the dynamics of global self-esteem and physical self. All series were classified as integrated.

A special case of integration are *fractionally integrated* processes (so-called ARFIMA-models) where d is not restricted to discrete and positive values. In the last years, ARFIMA-models have been increasingly applied in psychological research for modeling long-term dependencies known as *pink noise* or $1/f$-noise (Gilden 1997, 2001). Pink noise can be observed in diverse temporal phenomena, especially in the field of cognition (Farrell et al. 2006; Gilden et al. 1995, *inter alia*), but also in the temporal dynamics of self-esteem (Delignières et al. 2004), of the tics in Gilles de la Tourette syndrome (Peterson and Leckman 1998), or in the fluctuations of mood of patients suffering from a bipolar disorder (Gottschalk et al. 1995).

For the analysis of several processes in the context of their common relational structure, the univariate approach is no longer sufficient. Usually, the development of one variable depends at least on the development of another variable. The prognosis of each variable connected in such a system can be improved if the researcher models and analyzes them together. *Multivariate* (vector) autoregressive *VAR*(p)-*models* are direct generalizations of univariate time series models: single stochastic processes are replaced by vector processes and scalar parameters by parameter matrices.

$$\boldsymbol{x}_t = \boldsymbol{\Phi}_1 \boldsymbol{x}_{t-1} + \boldsymbol{\Phi}_2 \boldsymbol{x}_{t-2} + \cdots + \boldsymbol{\Phi}_p \boldsymbol{x}_{t-p} + \boldsymbol{\varepsilon}_t$$

describes a (m-dimensional) vector autoregressive process of order p, where $\boldsymbol{\Phi}_i$ ($i = 1, 2, \ldots, p$) are ($m \times m$)-parameter matrices with $\boldsymbol{\Phi}_i \neq \boldsymbol{0}$ and $\boldsymbol{\varepsilon}_t$ a vector of white noise.

A bivariate VAR(1)-process, for instance, is a system of two component processes written as

$$\begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix} = \begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix} \begin{pmatrix} x_{1,t-1} \\ x_{2,t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}.$$

In a psychological study about atopic dermatitis, a researcher might wish to investigate the temporal interaction of moods (e.g. depression) and a certain psychoneuroimmunological parameter (see Brosig and Brähler 2001). For this purpose, it would be recommended to use a VAR(1)-model consisting of the depression-process $x_{1t} = \phi_{11}x_{1,t-1} + \phi_{12}x_{2,t-1} + \varepsilon_{1t}$

and the psychoneuroimmunological process $x_{2t} = \phi_{21}x_{1,t-1} + \phi_{22}x_{2,t-1} + \varepsilon_{2t}$. Many methodological approaches of the bivariate case can be transferred to cases of higher order (Banerjee et al. 1993, p. 138). Therefore, the results of this study are based on bivariate cointegration relations.

2 Spurious regression: defining the problem

Psychologists investigate the problem of spurious regression/correlation as regards content as well as from a statistical perspective. In social psychology and cognition research, for instance, spurious correlations are a characteristic of judgment biases in group impression formation (Meiser 2006). They may result from representational conflicts as a part of contingency learning (Fiedler et al. 2002) or play a role in stereotype formation as a research field of social cognition (Meiser and Hewstone 2006). As another example, spurious correlations occur in mathematical thinking, when a student perceives a correlation between an irrelevant feature in a problem and the algorithm used for solving that problem and then proceeds to execute the algorithm when detecting the feature in a different problem. This distortion of the learning process leads to ineffectual problem solving (Ben-Zeev and Star 2001). This paper investigates spurious regression from a statistical perspective. At the same time, the methodological perspective has crucial implications for the psychological contents.

One basic lesson psychologists learn in their undergraduate statistics education is that correlation does not necessarily imply causation (Haig 2003). The statistical practice in psychology and related sciences is governed by the so-called *regularity theory* of causation (Harré and Madden 1975; see also Simon 1985). According to this theory, a causal relation is an actual, or alternatively hypothetical, regularity between different events. This paper refers to the correlation between processes, i.e. events recorded over time. The term *correlation* describes a relation without a particular direction, while the term *regression* stands for a directed, respectively causal relationship between two processes.

In 1926, Yule pointed out that “we sometimes obtain between (…) time-variables quite high correlations to which we cannot attach any physical significance whatever, although under the ordinary test the correlation would be held to be certainly ‘significant”’. He concluded: “the occurrence of such ‘nonsense correlations’ makes one mistrust the serious arguments that are sometimes put forward on the basis of correlations between time-series” (Yule 1926, p. 2). Granger and Newbold (1974) called this *spurious regression*. In this case, the regression between nonstationary time series is significant even if there is no relationship between the series. Thus, the problem of spurious regression can be also described in terms of the so-called *Type I error*, i.e. the error made when H_0(the tested hypothesis) is wrongly rejected. The fact that the usual significance tests on the coefficients are invalid can be regarded as one major consequence of autocorrelated errors in regression analysis beside others (Granger and Newbold 1974, p. 111).

The term spurious correlation/regression is ambiguous in the methodological literature (Kendall and Buckland 1982; Prather 1988). It was introduced by Pearson (1897) at the end of the 19th century to describe the situation in which a correlation is identified between two ratios or indexes although the original values are random observations on uncorrelated variables. This first definition does not explicitly refer to time variables like the one of Yule mentioned above. Nevertheless, its terminological discussion in the scientific community can contribute also to the further understanding of spurious relations between time series. Haig (2003) distinguishes between the term ‘nonsense/illusory correlations’ and the term ‘spurious correlations’. The first refers to those accidental correlations for which no sensible, natural causal interpretations can be provided. An amusing example would be the high positive correlation between birth rate and number of storks for a period in Britain. The latter describes Pearson's original idea: spurious correlations are accidental correlations that are not brought about by their claimed natural causes. Being the product of accidents, they cannot be genuine correlations. They are rather "artifacts of methods that arise from factors such as sample selection bias, use of an inappropriate correlation coefficient, large sample size, or errors of sampling, measurement, and computation" (Haig 2003, p. 128). Misleadingly, some researchers describe genuine correlations produced by common or intervening causes with the term 'spurious correlation'. These *indirect* correlations are *not* spurious: if general intelligence is the common cause of correlated IQ scores on the vocabulary and arithmetic subtests of an intelligence test, then those subtest scores are indirectly and genuinely correlated. Those correlations are not spurious as general intelligence explains why the correlation obtains (example taken from Haig 2003, p. 129).

Social and behavioral scientists have to make sure that the relationship under investigation is non-spurious. Quertemont et al. (2004) demonstrate the pitfalls of many correlational studies. They found that most of the conclusions in the study of Carey et al. (2003) were based on spurious correlations leading them to conclude that the response of rats to a novel environment negatively predicts their subsequent response to cocaine. The authors discuss how spurious correlations can be avoided. Their paper refers to single points of time. What concerns time series, Roberts (2000) proved that the results reported by Hitiris (1997) referring to the determinants of aggregate health care expenditure were seriously affected by a spurious regression problem. Roberts found this by using cointegration techniques.

Granger and Newbold (1974) investigated Yule's findings and concluded from their simulation study (with 100 replications and 50 observations [T]) that the probability of accepting the hypothesis of 'no relationship' becomes very small when regressions involve independent random walks. Similarly, Banerjee et al. (1993, 74ff.) found in simulation studies for $T = 100$ (10,000 replications) that the correct null hypothesis ($\beta_1 = 0$) was rejected in 75.3% of all the cases for a significance level of 0.05 for two *independent* nonstationary processes. Nelson and Kang (1984) found a slightly lower rejection quota of 64% ($p < 0.05$) and a determination coefficient R^2 with an average of 0.501 with a Durbin–Watson statistic of 0.260 ($T = 100$; 1,000 replications) for two unrelated random walks.

2.1 Monte Carlo replication study

These results were confirmed by our replication study additionally distinguishing between three significance levels: 0.05 (significant [s]), 0.01 (*very* significant [ss]) as well as *extremely* significant [sss] for $p < 0.0001$. The Monte Carlo simulations were run over two independent random walks for 1,000 replications with T varying in steps of 25 (from $T = 25$ up to $T = 500$). The results are listed in Table 1 and illustrated in Fig. 1. For $T = 100$, we found a total significance quota of 74.6%. On the 5%-level, only 7% refer to significant regressions. 15.2% refer to *very* significant regressions (ss for $p < 0.01$). What is most amazing is that over 50% of the regressions are *extremely* significant (sss for $p < 0.0001$). The percentages for s and ss approximately remain the same over the variations of T while the non-significant percentage (n) is constantly decreasing. Figure 1 shows clearly that the percentage of extremely significant regressions (sss) gradually increases over the variations of T, with 65.9% for $T = 200$ and 77.1% for $T = 400$. Starting from $T = 400$ the percentage of sss remains somehow constant. This finding underlines that the problem of spurious regression can make a psychological researcher postulating an extremely significant relationship between two processes *that actually does not exist*.

Table 1 Percentage of significant regressions of two independent random walks with varied length (T) for 1,000 replications

T	25	50	75	100	125	150	175	200	225	250
% n	49.6	33.1	28.6	25.4	21.6	19.5	16.9	19.2	16.4	14.9
% s	13.9	9.9	8.7	7	7.2	5.5	4.5	6.1	6.4	4.3
% ss	19.9	20.8	15.2	15.2	13	12.2	13.6	8.8	9.1	11
% sss	16.6	36.2	47.5	52.4	58.2	62.8	65	65.9	68.1	69.8
T	275	300	325	350	375	400	425	450	475	500
% n	14.3	14.9	14.9	14.9	12.2	12.4	11.6	10.2	10.7	10
% s	4.7	4.4	3.8	3.5	4.5	3.8	3.4	3.9	3.6	3
% ss	8.8	8	8.7	8.1	7.5	6.7	7.9	8.2	7	7.5
% sss	72.2	72.7	72.6	73.5	75.8	77.1	77.1	77.7	78.7	79.5

n, *not* significant for $p > 0.05$; s, significant for $p < 0.05$; ss, *very* significant for $p < 0.01$;, sss, *extremely* significant for $p < 0.0001$

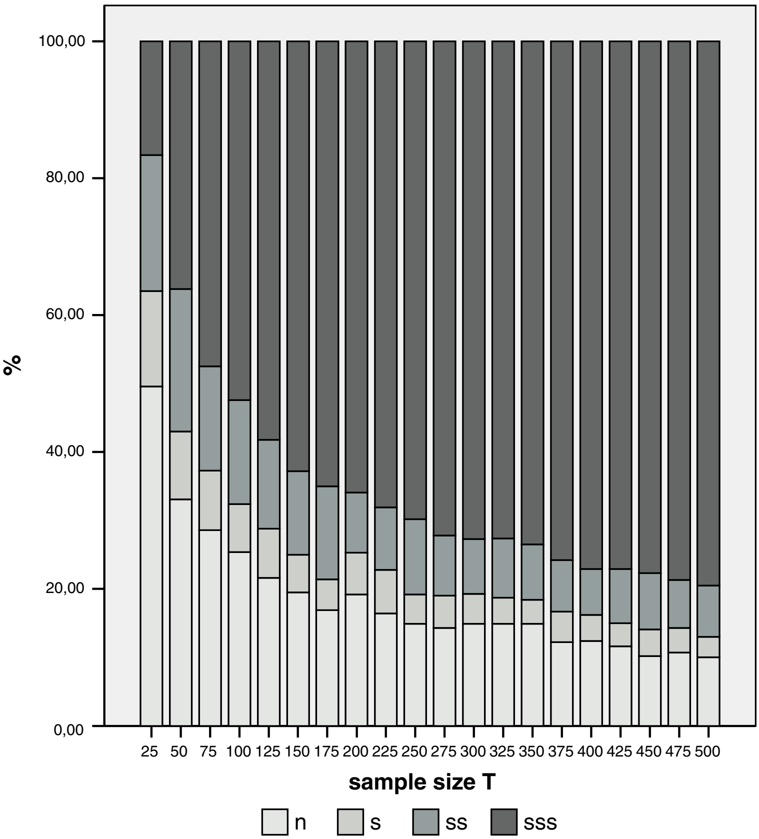

Fig. 1 Percentage of significant regressions of two independent random walks with varied length (T), received from 1,000 replications. Further explications in Table 1

Without doubt, the naive use of regression procedures can lead to dramatic mis-conclusions since the true nature of a process is never known a priori. In a personal interview, conducted by Phillips (1997, p. 262), C.W.J. Granger described that these findings were regarded as an error of programming when presented for the first time at the London School of Economics.

The problem of spurious regression between integrated processes can be avoided if the parameters of the regression model are estimated with the differenced series. However, differencing facilitates the mis-identification of apparent non-causalities. Moreover, differencing leads to a loss of crucial information. The analysis of differenced time series only reveals information about their *short-term* behavior. Possible *long-term* equilibrium relationships or adjustment dynamics between the integrated processes cannot be observed any longer. In other words, differencing of each variable in a multivariate system of time series "may indeed distort the relationship between the original variables" (Lütkepohl 1991, p. 351). This paper aims to show, in which way stochastic trends—as one possible reason for instationarity—can be appropriately handled by using cointegration methods. These methods have lead to a *change of the paradigm* in time series analysis.

3 Cointegration

Differencing of each component in a VAR(p)-process levels out long-term effects. In cointegrated systems, it is possible to model instationary processes and their information. Cointegrated processes describe a stable system. As the component processes are permanently influenced by developmental changes, the system may temporarily deviate from the long-run equilibrium, but always returns to it. The equilibrium connects the components with each other and can be written as a *linear combination* (Hamilton 1994, p. 572). Murray (1994) illustrates cointegration with the metaphor of a drunkard and her dog: both for themselves move like random walks. In spite of the instationarity of each individual's way, one would say: "if you find the drunkard, the dog is unlikely to be very far away."

Two instationary processes (integrated of order d [$I(d)$]) are cointegrated, if their linear combination is stationary [$I(0)$]. According to the conventional rules, one would say: if x_{1t} and x_{2t} are $I(1)$, each linear combination should be $I(1)$ as well. If cointegration is given, the linear combination $x_{1t} - \beta_2 x_{2t}$ is $I(0)$ (Granger 1988, 556f.). Thus, for the case of cointegration, "some of the previously stated rules do not hold" (Granger 1981, p. 127). $\boldsymbol{\beta} = (\beta_1, \beta_2)$ is the *cointegration vector* with β_1 usually normed to 1. The case described with components integrated of order $1 (\boldsymbol{x}_t \sim CI(1, 1))$ is most common in research practice (Rinne and Specht 2002, p. 540; Stier 2001).

The relationship between the processes involved can also be represented as (*vector*) *error-correction models* ((V)ECM). ECMs are the most important representations of cointegrated processes (Stier 2001, p. 318) where deviations ("errors") from the equilibrium are dynamically corrected. The adjustment coefficient ($\boldsymbol{\alpha}$) appearing only in the ECM and the cointegration vector ($\boldsymbol{\beta}$) are tools for modeling *long-run* equilibrium as well as *short-term* dynamic adjustment within the multivariate system (Darrat and Al-Yousif 1999; Harris and Sollis 2003; Stroe-Kunold and Werner, in press *inter alia*).

4 Cointegration tests

In the following Monte Carlo studies, two cointegration tests are used, the so-called *Dickey–Fuller cointegration test* and the *Johansen cointegration test*.

In bivariate models, only one possible cointegration relationship can be identified. For these models, testing on cointegration seems easy if the parameters of the equation $x_{1t} - \beta_2 x_{2t}$ are known: the rejection of the null hypothesis 'instationarity of the linear combination of two $I(1)$-series' indicates cointegration. For this purpose, the well-known unit-root-tests (Dickey and Fuller 1979, 1981) are appropriate. A time series which can be made stationary by differencing is said to have a *unit root*. The Dickey–Fuller test implemented in SAS is a method for testing whether a time series has a unit root. It tests the hypothesis H_0: 'The time series has a unit root' versus H_1: 'The time series is stationary' based on tables provided in Dickey (1976) and Dickey et al. (1984). For this reason, the procedure can be called *Dickey–Fuller cointegration test*.

In practice, the cointegration vector $\boldsymbol{\beta}$ is *not* known. For this case, Engle and Granger (1987) propose a two-step procedure. In case of the following Monte Carlo simulations, however, $\boldsymbol{\beta}$ is known. Therefore, the procedure first described is sufficient.

For models consisting of more than two series, it is important to find out how many cointegration relationships exist. The *Johansen cointegration test* (Johansen 1988, 1991; Johansen and Juselius 1990) is an appropriate procedure. It can also be used in the case of bivariate models. This *Full-Information Maximum-Likelihood-Approach* (*FIML*) (Hamilton 1994, 571ff.), based on the concept of *canonical correlations* (Stier 2001, 329ff.), consists of three steps (Hamilton 1994, 635ff.): (1) calculate auxiliary regressions; (2) calculate canonical correlations, and (3) calculate maximum likelihood estimates of parameters. Two significance tests result from this three-step FIML-approach: the *trace-test* and the *λ-max-test*. Both tests are called *cointegration rank tests*. The λ-max-test based on the same test statistic was not used here and is not described further in this paper.

The *trace-criterion* is the difference between the log-likelihood-function under H_1 (LLF_{H1}) and under H_0(LLF_{H0}). H_1 postulates the existence of m cointegration relations while H_0 assumes $h < m$ cointegration relations:

$$\mathrm{LLF}_{H1} - \mathrm{LLF}_{H0} = -\frac{T}{2} \sum_{i=h+1}^{m} \log(1 - \hat{\lambda}_i)$$

T is the number of observations, m is the amount of variables, $\hat{\lambda}_i$ stands for the eigenvalues. The doubling of this likelihood-quotient is equivalent to the trace of a multidimensional *Wiener process* (for the characterization of this process see Hamilton 1994, 477ff.). The statistic for testing the hypothesis that not more than h cointegration vectors with $m - h$ unit roots exist ($H_0 : r \leq h$ against $H_1 : r > h$) can be written as:

$$\lambda^{trace} = -T \sum_{i=h+1}^{m} \log(1 - \hat{\lambda}_i), \quad h = 0, 1, 2, \ldots, m - 1$$

The trace-statistic is calculated sequentially for each h (from $h = 0$ to $h = m - 1$). The VARMAX-procedure implemented in SAS uses the test statistic just described and tests H_0: Rank $= r$ against H_1: Rank $> r$. The amount of cointegration relations determines the rank r. In practice, this means that the first row (of the SAS-output) tests $r = 0$ against $r > 0$ and that the second row tests $r = 1$ against $r > 1$ etc. By default, the critical values at the 5%-significance level are used (SAS/ETS Software 2001, p. 54). This test could be regarded as a test with the null hypothesis 'spurious regression' and the alternative hypothesis 'cointegration'.

In this paper, the simulations are restricted to bivariate models. That's why only the first row of the output is relevant as only one cointegration relation is possible. The comparison of

```
                    The VARMAX Procedure

             Cointegration Rank Test Using Trace

                                                          5%
   H0:        H1:                                      Critical
 Rank=r     Rank>r      Eigenvalue          Trace        Value

      0          0          0.3697        71.7561        12.21
      1          1          0.0199         2.9917         4.14
```

Fig. 2 Exemplary excerpt of a SAS-output of the Johansen cointegration test for a given bivariate model and $T = 150$. Model characteristics in Table 5 (Model C_I_2)

the test statistics and critical values in this row of Fig. 2 shows that there is one cointegration relation.

5 Cointegration and spurious regression

De Jong (2003 Abstract) defines spurious regressions as "regressions in which an integrated process is regressed on another integrated process while there is no cointegration". Spurious correlations lead to the popularity of cointegration (Kumar 1995). As stated previously, spurious correlations can only be avoided without bias, if the instationarity of the series is *appropriately* modeled. Cointegration is a strategy of modeling multivariate time series by excluding spurious correlations (see also Banerjee et al. 1993, p. 158). The logic seems trivial: spurious regression can only be observed for *independent* integrated processes. Otherwise, the regression is not spurious. If cointegration is given, the regression between two integrated time series is real (Maddala and Kim 1998). In other words, spurious regression and cointegration are *opposite concepts* as cointegration implies a *meaningful* relation between integrated time series.

5.1 Monte Carlo study I: can spurious correlations be indicated by analyzing R^2- and Durbin–Watson-values?

Some authors recommend the analysis of R^2- and Durbin–Watson- (*DW*-) statistics in order to identify spurious regression between time series. Granger and Newbold (1974, p. 111) realized that the regression is spurious if the degree of fit measured by the coefficient of multiple correlation R^2 is apparently high and the *DW*-statistic is extremely low (see also Rinne and Specht 2002, p. 539). Phillips (1986, 318f.) postulates that low values of *DW* and moderate values of R^2 indicate spurious regression. A high value of R^2 should not, on the ground of traditional tests, be regarded as evidence of a significant relationship between autocorrelated series. Banerjee et al. (1993, p. 139) explain that "any regression for which $R^2 > DW$ is one that is likely to be spurious". This could be interpreted as a lack of any equilibrium relationship among the variables in the regression, which in turn implies a non-stationary error-term and so very strong autocorrelation in the regression residuals.

The following Monte Carlo experiment shows that the operationalization of spurious regression by these statistics can only partly contribute to the indication of spurious regression in multivariate time series.

5.1.1 Method

The performance of R^2- and *DW*-statistics for indicating spurious relations is assessed on 17 VAR(1)-models consisting of two random walks. As a second step, this performance is

Table 2 Different types of temporal relations in VAR(1)-models

	Model matrix	Component series	Pattern of relation
A	$\mathbf{\Phi}_1 = \begin{pmatrix} \phi_{11} & 0 \\ 0 & \phi_{22} \end{pmatrix}$ Diagonal matrix	$x_{1t} = \phi_{11}x_{1,t-1} + \varepsilon_{1t}$ $x_{2t} = \phi_{22}x_{2,t-1} + \varepsilon_{2t}$	*Independence* (for uncorrelated ε_{1t} and ε_{2t})
B	$\mathbf{\Phi}_1 = \begin{pmatrix} \phi_{11} & \phi_{12} \\ 0 & \phi_{22} \end{pmatrix}$ Upper triangular	$x_{1t} = \phi_{11}x_{1,t-1} + \phi_{12}x_{2,t-1} + \varepsilon_{1t}$ $x_{2t} = \phi_{22}x_{2,t-1} + \varepsilon_{2t}$	*Causality* x_{2t} (input) $\rightarrow$ x_{1t} (output)
C	$\mathbf{\Phi}_1 = \begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix}$ ϕ_{12} *and* $\phi_{21} \neq 0$	$x_{1t} = \phi_{11}x_{1,t-1} + \phi_{12}x_{2,t-1} + \varepsilon_{1t}$ $x_{2t} = \phi_{21}x_{1,t-1} + \phi_{22}x_{2,t-1} + \varepsilon_{2t}$	*Feedback* $x_{1t} \longleftrightarrow x_{2t}$

Table adapted from Stroe-Kunold and Werner (2007)

compared with the one of the cointegration tests described above (for the same models → *Monte Carlo Study II*). All results are received from 1,000 replications.

Social scientists are especially interested in the question, which sample size (T) is necessary for applying cointegration techniques. Therefore, we consider variations of T in the range of smaller values (=50, 100, 150, 200) with a prospect for a larger $T (= 500)$.

In VAR(1)-models, different types of relation (no relationship (A), causality (B), feedback (C)) between the participating series are possible. These are listed in Table 2.

To keep the relationship symmetric, the impacts of one series on the other and vice versa are identical for feedback models, i.e. $\phi_{12} = \phi_{21}$ for this case.

Pattern A is the case of independent time series. For this case, the regression can only be spurious. Cointegrated models can show pattern B or C. We checked R^2- and *DW*-statistics for all patterns. As in Granger and Newbold (1974), we calculated the corrected coefficient of R^2 additionally. The corrected formula proposed by Wherry (1931) is commonly used (Werner 1997, pp. 89–95)

$$R_c^2 = 1 - \frac{n-1}{n-p}\frac{SS_{\text{error}}}{SS_{\text{total}}} = 1 - \frac{n-1}{n-p}(1 - R^2)$$

where n is the number of experimental subjects and p is the number of predictors. The multiple coefficient of determination is adjusted to the number of degrees of freedom. In SAS for Windows this formula is called *adjusted R-square*.

Within the categories of causal (B; Table 4) and feedback models (C; Table 5) we distinguish additionally between models with components graphically *not* drifting apart (B_I_x or C_I_x, respectively) or drifting apart (B_II_x or C_II_x, respectively). If $\beta_2 = -1$ the component series of a multivariate process do *not* drift apart graphically while they do so for $\beta_2 \neq -1$ (here $\beta_2 = -2.5$; for details see Stroe-Kunold and Werner, in press).

Figure 3 exemplarily shows VAR(1)-processes with components *not* drifting apart (top line) and components drifting apart (bottom line).

For the purpose of interpretation, the degree of dependency of the two component processes (*dep*) is operationalized by $|\phi_{12}| + |\phi_{21}|$. *dep* is systematically varied in each of the four model categories ($dep = 0$ [model A1]; $dep = 0.25$; $dep = 0.5$; $dep = 0.75$; $dep = 1.0$).

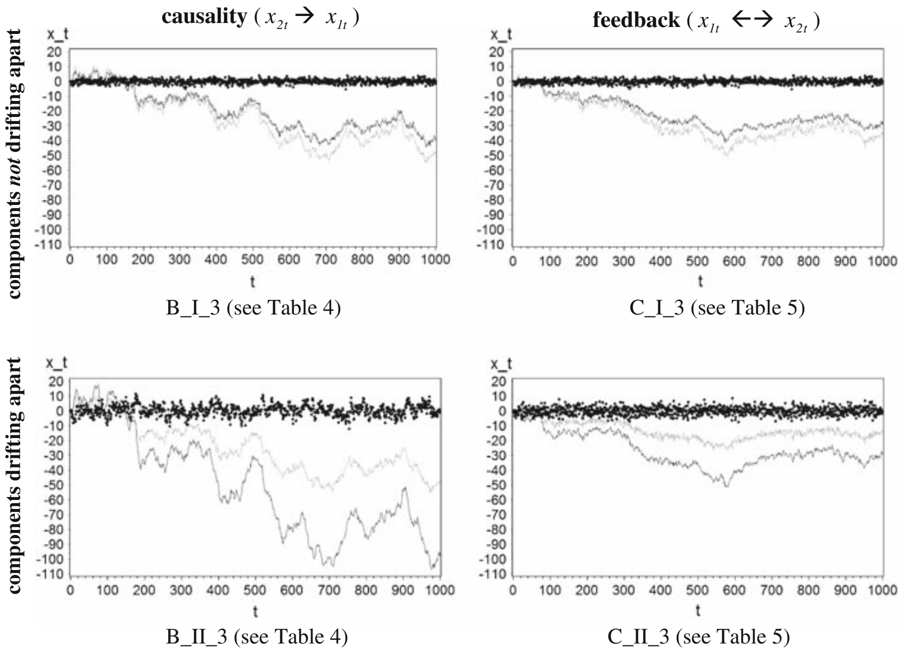

Fig. 3 Exemplary cointegrated processes with components not drifting apart (top line) and components drifting apart (bottom line) with same degree of dependency of 0.75 for all models. (For the purpose of illustration, x_{2t} was multiplied with a factor of 1.25 in order to distinguish the component series more clearly. The upper series is the linear combination of the darker series x_{1t} and the lighter series x_{2t}. $T = 1{,}000$ was chosen in order to illustrate the long-term development of the VAR(1)-processes. In reality, much less points of time are necessary)

Table 3 R^2-, R_c^2- and DW-statistics for a bivariate model consisting of two independent random walks (Type A; see Table 2); mean values over 1,000 replications; rounded to four decimals after the comma

Type		Model matrix	EV[a]	dep[b]	β_2	T	R^2	R_c^2	DW
A independence	A1	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$	1	0	–[c]	50	0.6045	0.5964	0.1913
			1			100	0.5234	0.5186	0.0995
						150	0.4842	0.4807	0.0687
						200	0.4588	0.4561	0.0516
						500	0.4143	0.4132	0.0222

[a] As soon as the model matrix shows one eigenvalue $EV = 1$, the components are random walks, i.e. $I(1)$-processes

[b] For the purpose of interpretation, the dependency of the two component processes (*dep*) is operationalized by $|\phi_{12}| + |\phi_{21}|$

[c] Two independent component series cannot be cointegrated. Thus, no cointegration vector $\boldsymbol{\beta}$ is existent for this type of model

5.1.2 Results

Tables 3–5 list the R^2-, R_c^2- and DW-statistics for all types of models.

Figure 4 illustrates these results.

Table 4 R^2-, R_c^2- and DW-statistics for a bivariate model consisting of two random walks with a causal relationship (Type B; see Table 2); mean values over 1,000 replications; rounded to four decimals after the comma

Type		Model matrix	EV	dep	β_2	T	R^2	R_c^2	DW
B causality	B_I_1	$\begin{pmatrix} 0.75 & 0.25 \\ 0 & 1 \end{pmatrix}$	1	0.25	−1	50	0.8657	0.8629	0.6068
			0.75			100	0.8982	0.8972	0.5483
						150	0.9178	0.9172	0.5316
						200	0.9316	0.9313	0.5205
						500	0.9643	0.9642	0.5076
	B_I_2	$\begin{pmatrix} 0.5 & 0.5 \\ 0 & 1 \end{pmatrix}$	1	0.5	−1	50	0.9102	0.9084	1.0676
			0.5			100	0.9358	0.9351	1.0295
						150	0.9494	0.9491	1.0177
						200	0.9585	0.9583	1.0104
						500	0.9788	0.9787	1.0016
	B_I_3	$\begin{pmatrix} 0.25 & 0.75 \\ 0 & 1 \end{pmatrix}$	1	0.75	−1	50	0.9254	0.9239	1.5319
			0.25			100	0.9475	0.947	1.5128
						150	0.959	0.9587	1.5055
						200	0.9666	0.9664	1.5027
						500	0.9829	0.9829	1.4967
	B_I_4	$\begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}$	1	1	−1	50	0.9297	0.9283	1.9971
			0			100	0.9507	0.9502	1.9973
						150	0.9615	0.9613	1.995
						200	0.9686	0.9685	1.9961
						500	0.984	0.984	1.9931
	B_II_1	$\begin{pmatrix} 0.9 & 0.25 \\ 0 & 1 \end{pmatrix}$	1	0.25	−2.5	50	0.8416	0.8384	0.3162
			0.9			100	0.8738	0.8726	0.2479
						150	0.8959	0.8953	0.2293
						200	0.9126	0.9122	0.2206
						500	0.9534	0.9534	0.2068
	B_II_2	$\begin{pmatrix} 0.8 & 0.5 \\ 0 & 1 \end{pmatrix}$	1	0.5	−2.5	50	0.8969	0.8948	0.4973
			0.8			100	0.9242	0.9234	0.4397
						150	0.9398	0.9394	0.4245
						200	0.9502	0.9499	0.4158
						500	0.9744	0.9744	0.4047
	B_II_3	$\begin{pmatrix} 0.7 & 0.75 \\ 0 & 1 \end{pmatrix}$	1	0.75	−2.5	50	0.9209	0.9193	0.6806
			0.7			100	0.9438	0.9432	0.6323
						150	0.956	0.9557	0.6195
						200	0.9639	0.9637	0.6113
						500	0.9817	0.9817	0.6021
	B_II_4	$\begin{pmatrix} 0.6 & 1 \\ 0 & 1 \end{pmatrix}$	1	1	−2.5	50	0.9343	0.933	0.8655
			0.6			100	0.9541	0.9536	0.8251
						150	0.9643	0.9641	0.8143
						200	0.9708	0.9707	0.8071
						500	0.9853	0.9853	0.7996

The statistics for model A1 with independent series ($dep = 0$) are illustrated as a point of reference for all four model categories in Fig. 4. They are depicted as the three columns on the right in all categories. The *DW*-values are extremely low (close to zero) and definitely smaller then the R^2-values. The case of spurious regression is clearly indicated here.

For causal models (B-models) with components *not* drifting apart (left side top of Fig. 4) and $dep = 0.25$, the R^2-values are still higher than the *DW*-values but the proportion starts to change. Clearly, R^2- and R_c^2-statistics behave approximately identical. Starting from

Table 5 R^2-, R_c^2- and DW-statistics for a bivariate model consisting of two random walks with a feedback relationship (Type C; see Table 2); mean values over 1,000 replications; rounded to four decimals after the comma

Type		Model matrix	EV	dep	β_2	T	R^2	R_c^2	DW
C feedback	C_I_1	$\begin{pmatrix} 0.875 & 0.125 \\ 0.125 & 0.875 \end{pmatrix}$	1 0.75	0.25	−1	50	0.7683	0.7635	0.5989
						100	0.8162	0.8143	0.5385
						150	0.8534	0.8524	0.5236
						200	0.8744	0.8738	0.5134
						500	0.9323	0.9322	0.5035
	C_I_2	$\begin{pmatrix} 0.75 & 0.25 \\ 0.25 & 0.75 \end{pmatrix}$	1 0.5	0.5	−1	50	0.8301	0.8266	1.045
						100	0.876	0.8747	1.0107
						150	0.9049	0.9042	1.0031
						200	0.9203	0.9199	0.998
						500	0.959	0.9589	0.9952
	C_I_3	$\begin{pmatrix} 0.625 & 0.375 \\ 0.375 & 0.625 \end{pmatrix}$	1 0.25	0.75	−1	50	0.8528	0.8498	1.4928
						100	0.8958	0.8948	1.4836
						150	0.9212	0.9206	1.4833
						200	0.9344	0.9341	1.4843
						500	0.9668	0.9667	1.4875
	C_I_4	$\begin{pmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{pmatrix}$	1 0	1	−1	50	0.8583	0.8554	1.9375
						100	0.9007	0.8997	1.9547
						150	0.9251	0.9246	1.963
						200	0.9379	0.9376	1.9696
						500	0.9688	0.9687	1.98
	C_II_1	$\begin{pmatrix} 0.95 & 0.125 \\ 0.125 & 0.6875 \end{pmatrix}$	1 0.6375	0.25	−2.5	50	0.7741	0.7695	1.2619
						100	0.8262	0.8244	1.2909
						150	0.8558	0.8548	1.311
						200	0.8747	0.8741	1.3252
						500	0.9309	0.9308	1.3755
	C_II_2	$\begin{pmatrix} 0.9 & 0.25 \\ 0.25 & 0.375 \end{pmatrix}$	1 0.275	0.5	−2.5	50	0.7741	0.7695	1.2619
						100	0.8262	0.8244	1.2909
						150	0.8558	0.8548	1.311
						200	0.8747	0.8741	1.3252
						500	0.9309	0.9308	1.3755
	C_II_3	$\begin{pmatrix} 0.85 & 0.375 \\ 0.375 & 0.0625 \end{pmatrix}$	1 −0.0875	0.75	−2.5	50	0.7821	0.7777	1.8267
						100	0.8339	0.8322	1.9101
						150	0.8628	0.8619	1.953
						200	0.8811	0.8805	1.9809
						500	0.9352	0.9351	2.0644
	C_II _4	$\begin{pmatrix} 0.8 & 0.5 \\ 0.5 & -0.25 \end{pmatrix}$	1 −0.45	1	−2.5	50	0.7741	0.7695	1.2619
						100	0.8262	0.8244	1.2909
						150	0.8558	0.8548	1.311
						200	0.8747	0.8741	1.3252
						500	0.9309	0.9308	1.3755

$dep = 0.5$, the R^2-values are lower than the DW-values. We show later that cointegration tests indicate (Table 7 and Fig. 5) that the model with $dep = 0.25$ is cointegrated. The causal relationship cannot be spurious, even if it is not so strong. For this model category the statistics do not seem to be precise enough.

For causal models with components drifting apart (left side down of Fig. 4) all statistics are definitely lower with a maximum value of ≈1.0. For all degrees of dependency R^2-values are higher than DW-values. Of course, DW-values grow with increasing dep, nevertheless the

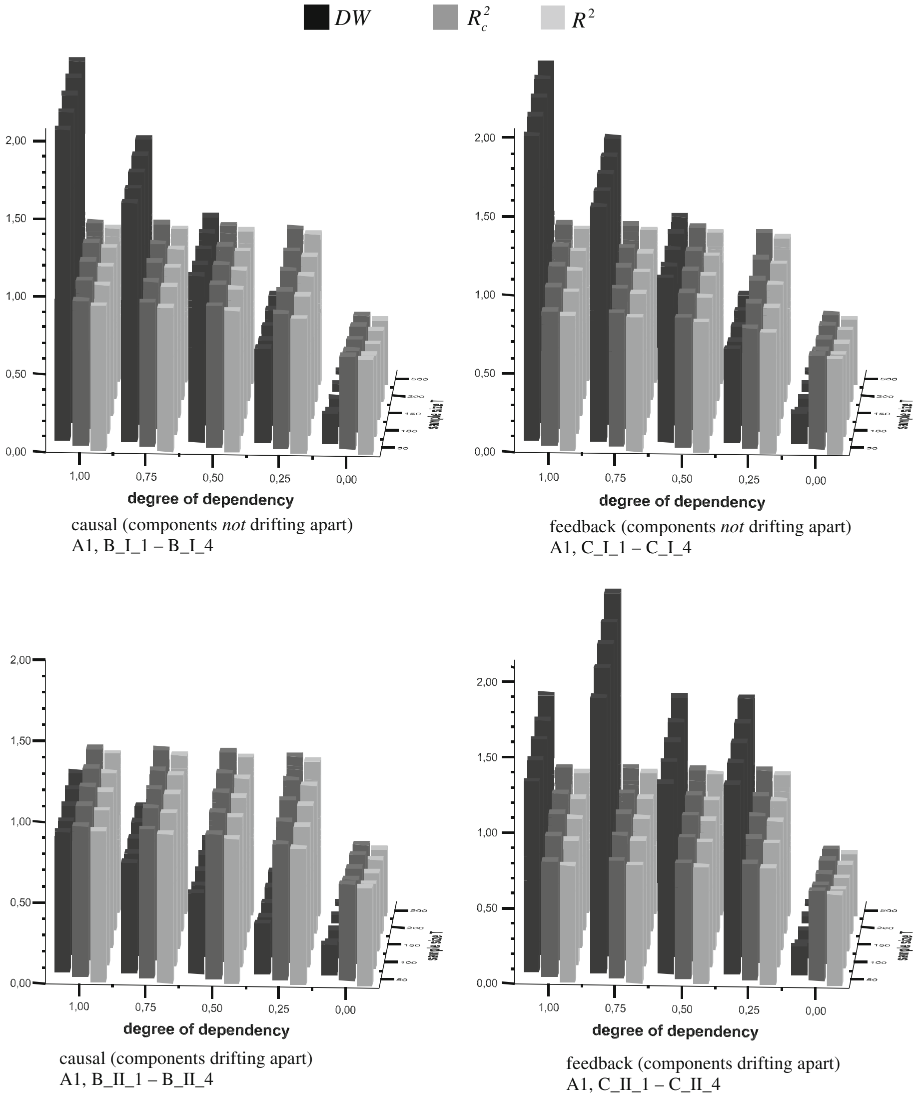

Fig. 4 R^2-, R_c^2- and DW-statistics for bivariate models consisting of two independent random walks, mean values received from 1,000 replications, illustrating the results listed in Tables 3–5. The 3^{rd} dimension illustrates the variations of the 'sample size T'

difference is not so clear as described above. For this category, the use of these statistics is not at all satisfactory for the purpose of identifying spurious regression.

For feedback models (C-models) with components *not* drifting apart (right side top of Fig. 4), the indication of spurious correlations by means of these statistics is as restricted as for the case of causal models with components *not* drifting apart (described above).

The category of feedback models with components drifting apart is the only model cluster for which the indication of spurious correlations by means of R^2-and DW-statistics is somehow satisfactory. For models with a $dep \geq 0.25$, DW-values are higher than R^2-values. Thus, they correctly indicate that the correlation is real and not spurious.

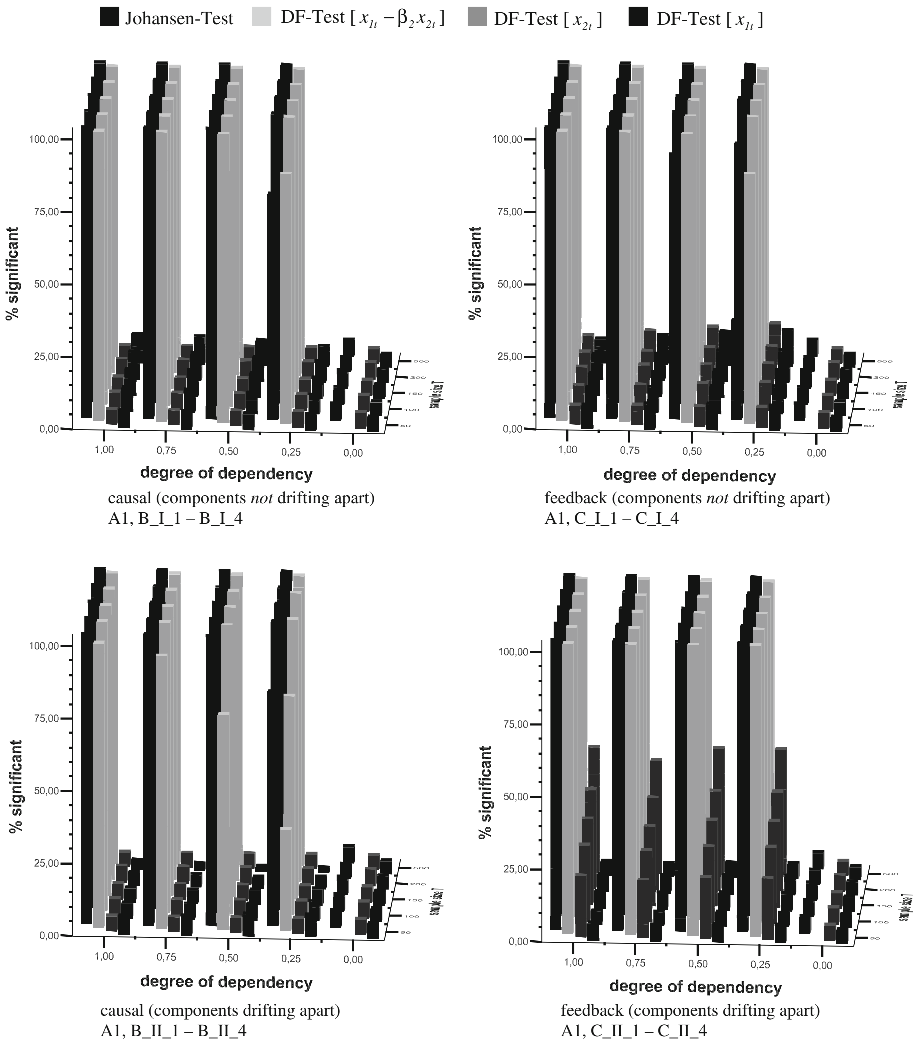

Fig. 5 Percentage of significant Dickey–Fuller- and Johansen cointegration tests, illustrating the results listed in Tables 6–8 (1,000 replications)

5.2 Monte Carlo study II: comparison with the results of cointegration tests

We compare these results with the results of the described cointegration tests gained from a Monte Carlo study run over the same models for the same variations of T.

5.2.1 Results

Tables 6–8 show the results illustrated by Fig. 5.

Again, we insert the results for model A1 (independent random walks) as a reference point for each model category.

Table 6 Percentage of non-significant (% ns)/significant (% s) Dickey–Fuller- and Johansen* cointegration tests for the model specified in Table 3 (1,000 replications; T varied)

Model	T	DF-test ($p < 0.05$)			Johansen-test ($p < 0.05$)
		% ns[x_{1t}]	% ns[x_{2t}]	% s[$x_{1t} - \beta_2 x_{2t}$]	% [$\boldsymbol{x}_t \sim CI(1,1)$]
A1	50	95.2	94.8	–**	5.7
	100	95.4	94.4	–	5.9
	150	94.9	94.7	–	6.1
	200	95.2	94.2	–	4.9
	500	95.6	95	–	5.4

* If the Johansen cointegration test is significant, the series—both integrated of order 1—are cointegrated, i.e $\boldsymbol{x}_t \sim CI(1,1)$

** Two independent component series cannot be cointegrated. Thus, no cointegration vector $\boldsymbol{\beta}$ weighting a linear combination is existent for this type of model. Therefore, the DF-test for this model ($dep = 0$) indicates no results for the linear combination in Figs. 5 and 6

Over all four model categories, the Johansen test clearly indicates cointegration for models with a $dep \geq 0.25$. For these models, the DF-test indicates stationarity of the linear combination in the majority of the cases over all model categories. Only for the model with $dep = 0.25$ in the category of causal models with components drifting apart the linear combination is stationary in merely 34.6% of the cases for $T = 50$. For $T \geq 50$, the linear combination is stationary for the majority of the cases.

Except for the last category of feedback models with components drifting apart, the Dickey–Fuller test indicates instationarity for all component series. Surprisingly, in this last category x_{2t} is stationary for a growing number of cases for increasing T. Obviously, the DF-test has problems in distinguishing between stochastic and deterministic trends in models with components drifting apart (see Hamilton 1994, p. 501). Further studies comparing the performance of the DF-test with the Augmented DF-test (ADF-test) might clarify this problem. Nevertheless, it is noteworthy that this does not necessarily contradict the idea of cointegration. Some authors explicitly include the definition of cointegration as a stationary linear combination of series that do not have the same degree of integration (Johansen 1995, 35ff.; Lütkepohl 1991, 352ff.). Engle and Granger (1991, p. 14) solve the discussion by stating that the inclusion of a stationary variable should neither affect the remaining coefficients nor the asymptotic critical values of the test statistics.

To sum up, cointegration tests seem to be a much more precise alternative for the identification of spurious relations compared to the rather imprecise method of utilizing the R^2- and DW-statistics for this purpose.

6 Cointegration tests as instruments to detect spurious correlations between integrated time series?

"A test for cointegration can be thought of as a pre-test to avoid 'spurious regression' situations" (Granger 1986, p. 226). This would be only true, if *all* meaningful relations between two integrated time series could be modeled as cointegrated relationships. Cointegration tests could only be used as instruments to detect spurious correlations between integrated time series if one found either spurious correlation/regression (independent series) or cointegration (dependent series) between integrated time series.

Table 7 Percentage of non-significant (% ns)/significant (% s) Dickey–Fuller- and Johansen cointegration tests for the models specified in Table 4 (1,000 replications; T varied)

Model	T	DF-test: non-significant ($p < 0.05$)			Johansen-test ($p < 0.05$)
		% ns[x_{1t}]	% ns[x_{2t}]	% s[$x_{1t} - \beta_2 x_{2t}$]	% [$\boldsymbol{x}_t \sim CI(1, 1)$]
B_I_1	50	94.7	94.8	86.1	77.8
	100	93.3	94.4	100	99.9
	150	92.9	94.7	100	100
	200	93	94.2	100	100
	500	93.2	95	100	100
B_I_2	50	94.6	94.8	99.6	99.9
	100	92.9	94.4	100	100
	150	91.9	94.7	100	100
	200	91.6	94.2	100	100
	500	91.5	95	100	100
B_I_3	50	94.2	94.8	100	100
	100	92.3	94.4	100	100
	150	91.1	94.7	100	100
	200	90.7	94.2	100	100
	500	90.4	95	100	100
B_I_4	50	93.8	94.8	100	100
	100	91.8	94.4	100	100
	150	90.8	94.7	100	100
	200	90.5	94.2	100	100
	500	89.9	95	100	100
B_II_1	50	92.2	94.8	34.6	81.1
	100	93.3	94.4	75.2	99.6
	150	94.5	94.7	96.1	100
	200	95.9	94.2	99.6	100
	500	98.3	95	100	100
B_II_2	50	92.4	94.8	74	99.7
	100	94.6	94.4	99.3	100
	150	95.3	94.7	100	100
	200	96.2	94.2	100	100
	500	98.1	95	100	100
B_II_3	50	93.3	94.8	94.1	100
	100	94.6	94.4	100	100
	150	95.4	94.7	100	100
	200	95.7	94.2	100	100
	500	97.4	95	100	100
B_II_4	50	93.5	94.8	98	100
	100	94.8	94.4	100	100
	150	95.1	94.7	100	100
	200	95.4	94.2	100	100
	500	96.6	95	100	100

6.1 Monte Carlo study III: do these tests indicate cointegration for a very low degree of dependency?

To investigate this issue further, we ran a Monte Carlo experiment with the two cointegration tests in order to find out whether these tests indicate cointegratedness already for series with a very low degree of dependency. If cointegration would be an ubiquitous und not just exotic characteristic of multivariate models consisting of integrated time series, cointegration tests could be regarded as refined instruments to avoid the danger of not identifying spurious correlations between integrated time series. By definition, two series are *not* independent if the

Table 8 Percentage of non-significant (% ns)/significant (% s) Dickey–Fuller- and Johansen cointegration tests for the models specified in Table 5 (1,000 replications; T varied)

Model	T	DF-test: non-significant ($p < 0.05$)			Johansen-test ($p < 0.05$)
		% ns[x_{1t}]	% ns[x_{2t}]	% s[$x_{1t} - \beta_2 x_{2t}$]	% [$\boldsymbol{x}_t \sim CI(1, 1)$]
C_I_1	50	91.6	93.2	86.1	34.1
	100	90.6	90.3	100	89.7
	150	90.4	89.3	100	99.6
	200	89.3	88.9	100	100
	500	87.7	86.8	100	100
C_I_2	50	91.3	93.3	99.6	91.1
	100	90.9	90.9	100	100
	150	90.8	89.9	100	100
	200	89.7	89.6	100	100
	500	88.5	88.1	100	100
C_I_3	50	91.4	93.1	100	100
	100	91.3	91.7	100	100
	150	91.6	90.3	100	100
	200	90.8	91.2	100	100
	500	90.1	89.7	100	100
C_I_4	50	92.1	93.4	100	100
	100	91.8	92.1	100	100
	150	92.4	91.3	100	100
	200	91.4	91.7	100	100
	500	91.6	91.6	100	100
C_II_1	50	93.9	78.9	100	99.8
	100	95.3	73.8	100	100
	150	94.2	70.3	100	100
	200	93.8	65.9	100	100
	500	94.9	56.8	100	100
C_II_2	50	93.9	78.9	100	99.8
	100	95.3	73.8	100	100
	150	94.2	70.3	100	100
	200	93.8	65.9	100	100
	500	94.9	56.8	100	100
C_II_3	50	94.3	80.2	100	100
	100	95.4	75.5	100	100
	150	94.5	72.7	100	100
	200	94	68.4	100	100
	500	95.4	61.2	100	100
C_II_4	50	93.9	78.9	100	99.8
	100	95.3	73.8	100	100
	150	94.2	70.3	100	100
	200	93.8	65.9	100	100
	500	94.9	56.8	100	100

degree of dependency is $\neq 0$. If cointegration tests were elaborate instruments, they would be able to show that the meaningfulness of the relation (i.e. cointegration) is low but existent.

6.1.1 Method

This second Monte Carlo study is an extension of study 1. We generated VAR(1)-models consisting of two random walks varied in the range of a very low degree of dependency. *dep* is varied from $dep = 0.05$ over $dep = 0.1$ and $dep = 0.15$ to $dep = 0.2$. The variations of T and the model categories are identical to those of study 1. All results are received from 1,000 replications.

6.1.2 Results

The results are listed in Tables 9 and 10 and illustrated in Fig. 6.

As expected, the pattern of the tests is not as clear as in the case of models with a higher degree of dependency. Thus, cointegration tests indicate that the dependency is low. For all model categories, the DF-test indicates instationarity for the component series, even if the instationarity of x_{2t} is slightly decreasing for increasing T in some of the cases. As discussed earlier, this does not need to contradict the idea of cointegration. For causal models with components *not* drifting apart and an extremely low degree of dependency ($dep = 0.05$), the DF-test indicates stationarity of the linear combination in more than 50% of the cases if $T \geq 150$. For causal models with components drifting apart, this is only the case for $T = 500$. For the feedback case, we find this for $T \geq 150$ (components *not* drifting) or $T \geq 50$ (components drifting). For higher degrees of dependency, the linear combination is indicated to be stationary much 'earlier'. As can be concluded from Fig. 6, the Johansen-test shows a quite similar pattern. For the case of minimal dependency ($dep = 0.05$), cointegration is indicated in more than 50% of the cases for causal models with components *not* drifting apart for $T \geq 200$, for $T \geq 150$ (causal models with components drifting apart), for $T \geq 500$ (feedback models with components *not* drifting apart) and for $T \geq 50$ (same models with components drifting apart), respectively. It becomes obvious that both tests indicate some meaningful relation between the integrated series. At the same time, their pattern of behavior clearly shows that this relation is not too strong. Furthermore, these results show that cointegration is an ubiquitous characteristic of VAR(1)-models consisting of two integrated time series. The comparison of the results for zero dependency ($dep = 0$) versus minimal dependency ($dep = 0.05$) underlines that these cointegration tests clearly distinguish between the 'spurious regression-case' and the case of a slight dependency.

It seems interesting to have a look at the R^2-and DW-statistics. We illustrate their behavior for the models simulated here in Fig. 7.

Obviously, the behavior of these statistics shows no regularity. Except for the last model category (feedback models with components drifting apart), they indicate a spurious relationship for the case of minimal dependency ($dep = 0.05$). Moreover, they do so still for higher degrees of dependency.

The comparison with these statistics underlines the advantages of the cointegration techniques. Their accurate usefulness for dealing with the problem of spurious relations is obvious. Using these tests, the researcher is not in danger of either neglecting a small but meaningful relation or regarding a relation as meaningful which is actually spurious.

7 Conclusions

The fact that many psychological processes are integrated means that considerable care has to be taken in what concerns the problem of spurious regression. The results presented here show that cointegration tests can be regarded as elaborate instruments to detect spurious correlations between integrated time series.

In the 1980s, Granger (1986, p. 227) described the "usefulness" of the concept of cointegration as one reason for gaining acceptance in econometrics. The aim of this paper was to exemplify one aspect of its usefulness for psychological research.

Without doubt, cointegration methodology shows a high potential of application for psychological process research (Stroe-Kunold and Werner, in press). Nevertheless, it seems unfamiliar to most psychologists. Although time series analysis has been introduced to social

Table 9 Percentage of non-significant (% ns)/significant (% s) Dickey–Fuller- and Johansen cointegration tests for causal models (Type B) with a minimal degree of dependency (1,000 replications; T varied)

Type		Model	EV	dep	β_2	T	DF-test ($p < 0.05$)			Johansen-Test ($p < 0.05$)
							% ns $[x_{1t}]$	% ns $[x_{2t}]$	% s $[x_{1t} - \beta_2 x_{2t}]$	% $[\boldsymbol{x}_t \sim CI(1,1)]$
B causality	B_I_5	$\begin{pmatrix} 0.95 & 0.05 \\ 0 & 1 \end{pmatrix}$	1	0.05	−1	50	94.7	94.8	17.2	12.5
			0.95			100	95.3	94.4	35	27.1
						150	95.6	94.7	58.6	44.4
						200	95.8	94.2	78.2	63.1
						500	94.8	95	100	99.9
	B_I_6	$\begin{pmatrix} 0.9 & 0.1 \\ 0 & 1 \end{pmatrix}$	1	0.1	−1	50	93.9	78.9	100	99.8
			0.9			100	95.3	73.8	100	100
						150	94.2	70.3	100	100
						200	93.8	65.9	100	100
						500	94.9	56.8	100	100
	B_I_7	$\begin{pmatrix} 0.85 & 0.15 \\ 0 & 1 \end{pmatrix}$	1	0.15	−1	50	94.1	94.8	55.7	43.2
			0.85			100	93.3	94.4	95.2	89.4
						150	94.1	94.7	99.8	99.2
						200	94	94.2	100	100
						500	94	95	100	100
	B_I_8	$\begin{pmatrix} 0.8 & 0.2 \\ 0 & 1 \end{pmatrix}$	1	0.2	−1	50	93.9	78.9	100	99.8
			0.8			100	95.3	73.8	100	100
						150	94.2	70.3	100	100
						200	93.8	65.9	100	100
						500	94.9	56.8	100	100

Table 9 continued

Type	Model	EV	dep	β_2	T	DF-test ($p < 0.05$)			Johansen-Test ($p < 0.05$)
						% ns $[x_{1t}]$	% ns $[x_{2t}]$	% s $[x_{1t} - \beta_2 x_{2t}]$	% $[\boldsymbol{x}_t \sim CI(1,1)]$
B_II_5	$\begin{pmatrix} 0.98 & 0.05 \\ 0 & 1 \end{pmatrix}$	1	0.05	−2.5	50	95.4	94.8	9.6	20.4
		0.98			100	96.9	94.4	13.6	38.2
					150	96.6	94.7	18.8	60.5
					200	97.7	94.2	29.3	75.4
					500	99.3	95	79.1	99.9
B_II_6	$\begin{pmatrix} 0.96 & 0.1 \\ 0 & 1 \end{pmatrix}$	1	0.1	−2.5	50	93.9	78.9	100	99.9
		0.96			100	95.3	73.8	100	100
					150	94.2	70.3	100	100
					200	93.8	65.9	100	100
					500	94.9	56.8	100	100
B_II_7	$\begin{pmatrix} 0.94 & 0.15 \\ 0 & 1 \end{pmatrix}$	1	0.15	−2.5	50	92.2	94.8	18.5	56.1
		0.94			100	93.7	94.4	44.8	91.1
					150	94.4	94.7	70.8	99.3
					200	96.3	94.2	88.9	99.9
					500	98.6	95	100	100
B_II_8	$\begin{pmatrix} 0.92 & 0.2 \\ 0 & 1 \end{pmatrix}$	1	0.2	−2.5	50	92.1	94.8	25.8	71
		0.92			100	93.1	94.4	62	98.2
					150	94.7	94.7	88	99.9
					200	96.2	94.2	97.5	100
					500	98.5	95	100	100

Table 10 Percentage of non-significant (% ns)/significant (% s) Dickey–Fuller- and Johansen cointegration tests for feedback models (Type C) with a minimal degree if dependency (1,000 replications; T varied)

Type		Model	*EV*	*dep*	β_2	T	DF-test ($p < 0.05$)			Johansen-Test ($p < 0.05$)
							% ns $[x_{1t}]$	% ns $[x_{2t}]$	% s $[x_{1t} - \beta_2 x_{2t}]$	% $[\boldsymbol{x}_t \sim CI(1,1)]$
C feedback	C_I_5	$\begin{pmatrix} 0.975 & 0.025 \\ 0.025 & 0.975 \end{pmatrix}$	1	0.05	−1	50	93.4	93.9	17.2	7.1
			0.95			100	92.8	92	35	10.6
						150	93	91.1	58.6	14.9
						200	91.2	90.5	78.2	23.5
						500	88.9	86.7	100	85.9
	C_I_6	$\begin{pmatrix} 0.95 & 0.05 \\ 0.05 & 0.95 \end{pmatrix}$	1	0.1	−1	50	92.5	93.9	35.4	10.2
			0.9			100	91.6	90.5	75.5	23.3
						150	91.6	89.8	95.7	45.8
						200	90.1	89.1	99.8	69.1
						500	87.9	86.7	100	100
	C_I_7	$\begin{pmatrix} 0.925 & 0.075 \\ 0.075 & 0.925 \end{pmatrix}$	1	0.15	−1	50	92.1	94	55.7	16.6
			0.85			100	90.9	90.1	95.2	45.7
						150	91	89.7	99.8	80.8
						200	89.9	88.8	100	95.7
						500	87.5	86.6	100	100
	C_I_8	$\begin{pmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{pmatrix}$	1	0.2	−1	50	93.9	78.9	100	99.8
			0.8			100	95.3	73.8	100	100
						150	94.2	70.3	100	100
						200	93.8	65.9	100	100
						500	94.9	56.8	100	100
	C_II_5	$\begin{pmatrix} 0.99 & 0.025 \\ 0.025 & 0.9375 \end{pmatrix}$	1	0.05	−2.5	50	93.9	78.9	100	99.8
			0.9275			100	95.3	73.8	100	100
						150	94.2	70.3	100	100

Table 10 continued

Type	Model	*EV*	*dep*	β_2	*T*	DF-test ($p < 0.05$)			Johansen-Test ($p < 0.05$)
						% ns $[x_{1t}]$	% ns $[x_{2t}]$	% s $[x_{1t} - \beta_2 x_{2t}]$	% $[\boldsymbol{x}_t \sim CI(1,1)]$
					200	93.8	65.9	100	100
					500	94.9	56.8	100	100
C_II_6	$\begin{pmatrix} 0.98 & 0.05 \\ 0.05 & 0.875 \end{pmatrix}$	1	0.1	−2.5	50	94.5	85.4	54.7	13.8
		0.855			100	95.1	77.9	94.3	43.4
					150	94.1	70	99.8	78
					200	93.2	65.4	100	95.8
					500	94.5	52.6	100	100
C_II_7	$\begin{pmatrix} 0.97 & 0.075 \\ 0.075 & 0.8125 \end{pmatrix}$	1	0.15	−2.5	50	94.2	82.6	79.1	27.2
		0.7825			100	95.1	74.3	99.9	77.6
					150	94	68.3	100	98.9
					200	93.4	63.8	100	100
					500	94.5	52.8	100	100
C_II_8	$\begin{pmatrix} 0.96 & 0.1 \\ 0.1 & 0.75 \end{pmatrix}$	1	0.2	−2.5	50	94.4	81.4	92.5	43.1
		0.71			100	95.1	73.3	100	96.3
					150	93.9	68	100	100
					200	93.4	63.4	100	100
					500	94.4	53.1	100	100

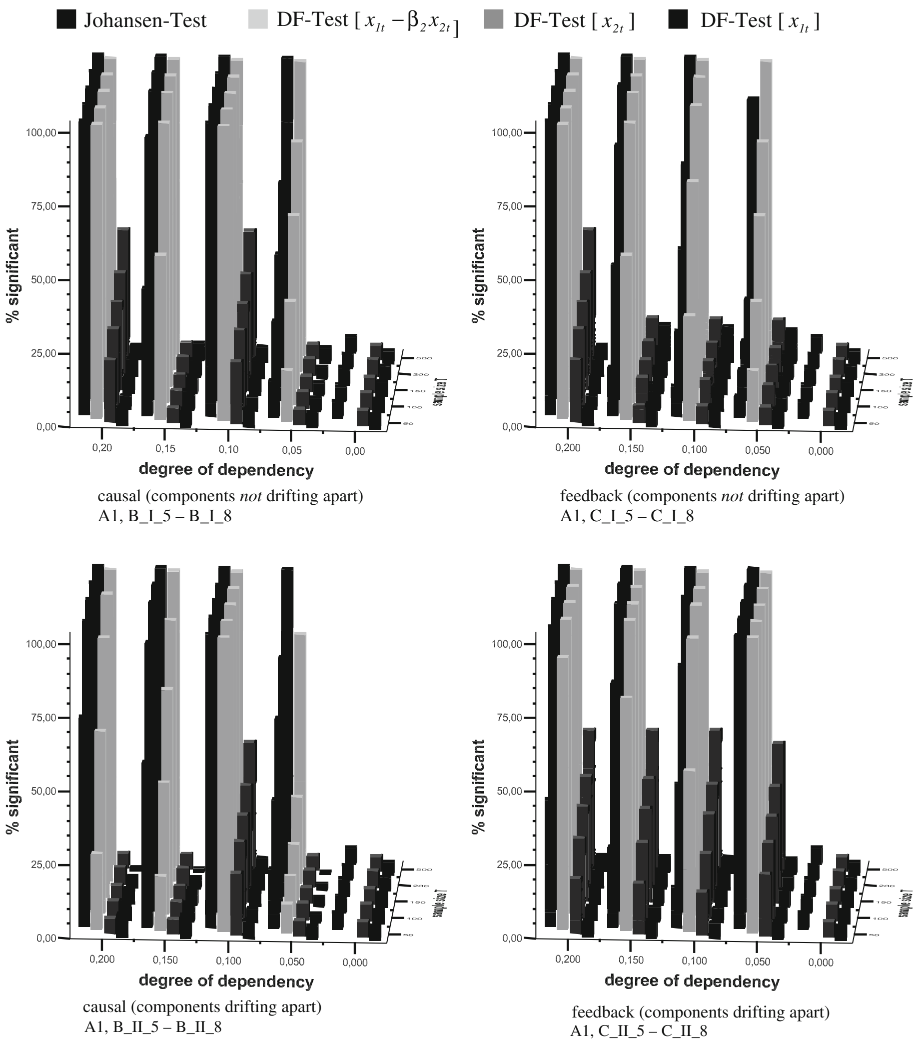

Fig. 6 Percentage of significant Dickey–Fuller- and Johansen cointegration tests, illustrating the results listed in Tables 9 and 10 (1,000 replications)

and behavioral sciences three decades ago by Glass et al. (1975) and further exemplified by Gottman (1981), it has not gained appropriate attention in psychological research for many years. In the last few years, however, an increasing interest in psychological processes can be observed (e.g. Ninot et al. 2001; Van Orden et al. 2003; Velicer and Fava 2003; Werner 2005)—though the studies are mainly restricted to univariate models.

At the same time, new methods of capturing data allow social scientists recording data samples of a sufficient size: the term *Ambulatory Assessment* stands for the use of ambulatory, mainly computer-based, measuring instruments. They can be used in everyday-conditions in order to record larger pools of observations of various psychological and physiological phenomena (Fahrenberg et al. 2007).

An increasing number of publications referring to cointegration can recently be found in sociology (e.g. Bremmer and Kesselring 2004; Lin and Brannigan 2003; Luiz 2001; McNown

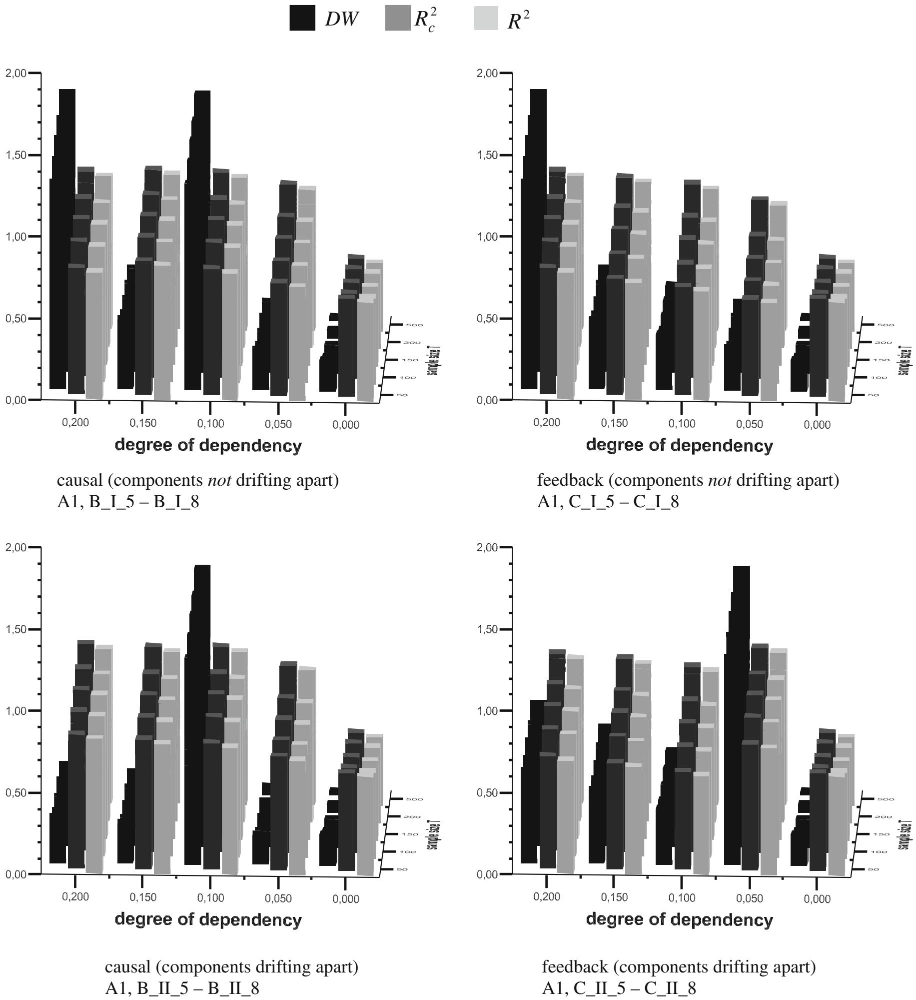

Fig. 7 R^2-, R_c^2- and DW-statistics for bivariate models consisting of two independent random walks (1,000 replications), illustrating the results listed in Tables 9 and 10

2003; Witt and Witte 2000). This can be regarded as a first step to introduce the cointegration approach into research on human behavior.

At the same time, it should not be overlooked that the concept of cointegration has many expansions. *Fractionally cointegrated* models, as one example, consist of fractionally integrated component series. In front of the background of the frequent application of ARFIMA-models described above, fractional cointegration is a promising concept to be investigated from a psychological perspective. Marmol and Velasco (2004) describe procedures to test the spurious regression hypothesis against the alternative of fractional cointegration. Another interesting expansion is the concept of *multi-cointegration* as the combination of various cointegration relations (see Granger and Lee 1990).

Haldrup (1994) extended the existing literature on spurious regression for the $I(1)$ case to models with $I(2)$ series. Future studies have to show whether cointegration tests can be regarded as remedies to the spurious regression problem also for this case.

Cointegration tests fulfill a useful function in what concerns a precise indication of spurious relations between integrated time series. Beyond this usefulness, social scientists have the chance to gain an elaborate insight into the dynamics of human processes by means of cointegration methodology.

Acknowledgments The research for this paper has been supported by a scholarship granted to the first author by the *German National Academic Foundation*.

References

Banerjee, A., Dolado, J.J., Galbraith, J.W., Hendry, D.F.: Co-integration, Error-Correction, and the Econometric Analysis of Non-stationary Data. Oxford University Press, Oxford (1993)

Ben-Zeev, T., Star, J.R.: Spurious correlations in mathematical thinking. Cogn. Instruct. **19**, 253–275 (2001)

Box, G.E.P., Jenkins, G.M.: Time-Series Analysis: Forecasting and Control. Holden-Day, San Francisco (1976)

Bremmer, D., Kesselring, R.: Divorce and female labor force participation: evidence from time-series data and cointegration. Atl. Econ. J. **32**, 175–190 (2004)

Brosig, B., Brähler, E.: Skin-mind-immune system: a vector-time series model. Z. Klin. Psychol. Psychiatr. Psychother. **49**, 279–296 (2001)

Carey, R.J., DePalma, G., Damianopoulos, E.: Response to novelty as a predictor of cocaine sensitization and conditioning in rats: a correlational analysis. Psychopharmacology **168**, 245–252 (2003)

Darrat, A.F., Al-Yousif, Y.K.: On the long-run relationship between population and economic growth: some time series evidence for developing countries. East. Econ. J. **25**, 301–306 (1999)

De Jong, R.M.: Logarithmic spurious regressions. Econ. Lett. **81**, 13–21 (2003)

Delignières, D., Fortes, M., Ninot, G.: The fractal dynamics of self-esteem and physical self. Nonlinear Dynamics Psychol. Life Sci. **8**, 479–510 (2004)

Dickey, D.A.: Estimation and testing of nonstationary time series. Unpublished Ph.D. Thesis, Iowa State University, Ames (1976)

Dickey, D.A., Fuller, W.A.: Distribution of estimators for autoregressive time series with a unit root. J. Am. Stat. Assoc. **74**, 427–431 (1979)

Dickey, D.A., Fuller, W.A.: Likelihood ratio statistics for autoregressive time series with a unit root. Econometrica **49**, 1057–1072 (1981)

Dickey, D.A., Hasza, D.P., Fuller, W.A.: Testing for unit roots in seasonal time series. J. Am. Stat. Assoc. **79**, 355–367 (1984)

Engle, R.F., Granger, C.W.J.: Co-integration and error correction: representation, estimation and testing. Econometrica **55**, 251–276 (1987)

Engle, R.F., Granger, C.W.J. (eds.): Long-Run Economic Relationships. Readings in Cointegration. Oxford University Press, New York (1991)

Fahrenberg, J., Myrtek, M., Pawlik, K., Perrez, M.: Ambulatory assessment—capturing behavior in daily life. A behavioral science approach to psychology. Psychol. Rundsch. **58**, 12–24 (2007)

Farrell, S., Wagenmakers, E.-J., Ratcliff, R.: 1/f noise in human cognition: is it ubiquitous, and what does it mean?. Psychon. Bull. Rev. **13**, 737–741 (2006)

Fiedler, K., Walther, E., Freytag, P., Stryczek, E.: Playing mating games in foreign cultures: a conceptual framework and an experimental paradigm for inductive trivariate inference. J. Exp. Soc. Psychol. **38**, 14–30 (2002)

Fortes, M., Ninot, G., Delignières, D.: The dynamics of self-esteem and physical self: between preservation and adaptation. Qual. Quant. **38**, 735–751 (2004)

Gilden, D.L.: Fluctuations in the time required for elementary decisions. Psychol. Sci. **8**, 296–301 (1997)

Gilden, D.L.: Cognitive emissions of 1/f noises. Psychol. Rev. **108**, 33–56 (2001)

Gilden, D.L., Thornton, T., Mallon, M.W.: 1/f noise in human cognition. Science **267**, 1837–1839 (1995)

Glass, G.V., Willson, V.L., Gottman, J.M.: Design and Analysis of Time-Series Experiments. Colorado Associated University Press, Boulder (1975)

Gottman, J.M.: Time-Series Analysis. A Comprehensive Introduction for Social Scientists. Cambridge University Press, Cambridge (1981)

Gottschalk, A., Bauer, M.S., Whybrow, P.C.: Evidence of chaotic mood variation in bipolar disorder. Arch. Gen. Psychiatry **52**, 947–959 (1995)

Granger, C.W.J.: Some properties of time series data and their use in econometric model specification. J. Econom. **16**, 121–130 (1981)

Granger, C.W.J.: Developments in the study of cointegrated economic variables. Oxf. Bull. Econ. Stat. **48**, 213–228 (1986)

Granger, C.W.J.: Causality, cointegration and control. J. Econ. Dyn. Control **12**, 551–559 (1988)

Granger, C.W.J., Lee, T.W.: Multicointegration. In: Fomby, Th.B., Rhodes, G.F. (eds.) Co-integration, Spurious Regressions, and Unit-Roots: Advances in Econometrics, vol. 8, pp. 71–84. JAI Press, Greenwich (1990)

Granger, C.W.J., Newbold, P.: Spurious regression in econometrics. J. Econom. **2**, 111–120 (1974)

Haig, B.D.: What is a spurious correlation? Underst. Stat. **2**, 125–132 (2003)

Haldrup, N.: The asymptotics of single-equation cointegration regressions with $I(1)$ and $I(2)$ variables. J. Econom. **63**, 153–181 (1994)

Hamilton, J.D.: Time Series Analysis. Princeton University Press, Princeton (1994)

Harré, R., Madden, E.H.: Causal Powers. Basil Blackwell, Oxford (1975)

Harris, R., Sollis, R.: Applied Time Series Modelling and Forecasting. Wiley, Chichester (2003)

Hendry, D.F.: Econometric modelling with cointegrated variables: an overview. Oxf. Bull. Econ. Stat. **48**, 201–212 (1986)

Hitiris, T.: Health care expenditure and integration in the countries of the European Union. Appl. Econ. Lett. **29**, 1–6 (1997)

Johansen, S.: Statistical analysis of cointegration vectors. J. Econ. Dyn. Control **12**, 231–254 (1988)

Johansen, S.: Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. Econometrica **59**, 1551–1580 (1991)

Johansen, S.: Likelihood-Based Inference in Cointegrated Vector Autoregressive Models. Oxford University Press, New York (1995)

Johansen, S., Juselius, K.: Maximum likelihood estimation and inference on cointegration—with applications to the demand for money. Oxf. Bull. Econ. Stat. **52**, 169–210 (1990)

Kendall, M.G., Buckland, W.R.: A Dictionary of Statistical Terms. Longman, London (1982)

Kumar, T.K.: Cointegration and error correction models: a historical and methodological perspective. J. Quant. Econ. **11**, 143–154 (1995)

Lasswell, H.D.: Changes in an experimental subject during a short series of psychoanalytical interviews. Imago **23**, 375–380 (1937)

Lin, Z., Brannigan, A.: Advances in the analysis of non-stationary time series: an illustration of cointegration and error correction methods in research on crime and immigration. Qual. Quant.: Int. J. Methodol. **37**, 151–168 (2003)

Luiz, J.-M.: Temporal association, the dynamics of crime, and their economic determinants: a time series econometric model of South Africa. Soc. Indic. Res. **53**, 33–61 (2001)

Lütkepohl, H.: Introduction to Multiple Time Series Analysis. Springer, New York (1991)

Maddala, G., Kim, I.-M.: Unit Roots, Cointegration and Structural Change. Cambridge University Press, Cambridge (1998)

Marmol, F., Velasco, C.: Consistent testing of cointegrating relationships. Econometrica **72**, 1809–1844 (2004)

McNown, R.: A cointegration model of age-specific fertility and female labor supply in the United States. South. Econ. J. **70**, 344–358 (2003)

Meiser, Th: Contingency learning and biased group impressions. In: Fiedler, K., Juslin, P. (eds.) Information Sampling and Adaptive Cognition, pp. 183–209. Cambridge University Press, New York (2006)

Meiser, Th., Hewstone, M.: Illusory and spurious correlations: distinct phenomena or joint outcomes of exemplar-based category learning? Eur. J. Soc. Psychol. **36**, 315–336 (2006)

Murray, M.P.: A drunk and her dog: an illustration of cointegration and error correction. Am. Stat. **48**, 37–39 (1994)

Nelson, C.R., Kang, H.: Pitfalls in the use of time as an explanatory variable in regression. J. Bus. Econ. Stat. **2**, 73–82 (1984)

Ninot, G., Fortes, M., Delignières, D.: A psychometric tool for the assessment of the dynamics of the physical self. Eur. Rev. Appl. Psychol. **51**, 205–216 (2001)

Ninot, G., Fortes, M., Delignières, D.: The dynamics of self-esteem in adults over a six-month period: an exploratory study. J. Psychol. **139**, 315–330 (2005)

Pearson, K.: Mathematical contributions to the theory of evolution: on a form of spurious correlation which may arise when indices are used in the measurement of organs. Proc. R. Soc. Lond. **60**, 489–498 (1897)

Peatman, J.G.: Hazards and fallacies of statistical method in psychological measurement. Psychol. Rec. **22**, 365–390 (1937)

Peterson, B.S., Leckman, J.F.: The temporal dynamics of the tics in Gilles de la Tourette syndrome. Biol. Psychiatry **44**, 1337–1348 (1998)

Phillips, P.C.B.: Understanding spurious regression in econometrics. J. Econom. **33**, 311–340 (1986)

Phillips, P.C.B.: The ET interview: Professor Clive Granger. Econom. Theory **13**, 253–303 (1997)

Prather, J.E.: Spurious correlation. In: Encyclopedia of Statistical Science, vol. 8, pp. 613-614. Wiley, New York (1988)

Quertemont, E., Brabant, C., Tirelli, E.: Response to novelty as a predictor for drug effects: the pitfalls of some correlational studies. Psychopharmacology **173**, 221–224 (2004)

Rinne, H., Specht, K.: Zeitreihen—Statistische Modellierung, Schätzung und Prognose. Vahlen, München (2002)

Roberts, J.: Spurious regression problems in the determinants of health care expenditure: a comment on Hitiris (1997). Appl. Econ. Lett. **7**, 279–283 (2000)

SAS/ETS, Software: Changes and Enhancements, Release 8.2. SAS Institute Inc, Cary (2001)

Simon, H: Spurious correlation: a causal interpretation. In: Blalock, H.M. (ed.) Causal Models in the Social Sciences, pp. 7–21. Aldine, New York (1985)

Stier, W.: Methoden der Zeitreihenanalyse. Springer, Berlin (2001)

Stroe-Kunold, E., Werner, J.: Are psychological processes cointegrated? Present role und future perspectives of cointegration methodology in psychological research Psychool. Rundsch **58**, 225–237 (2007)

Stroe-Kunold, E., Werner, J.: Modeling Human Dynamics by Means of Cointegration Methodology (in press)

Van Orden, G.C., Holden, J.G., Turvey, M.T.: Self-organization of cognitive performance. J. Exp. Psychol. Gen. **132**, 331–350 (2003)

Velicer, W.F., Fava, J.L.: Time series analysis. In: Schinka, J., Velicer, W.F. (eds.) Research Methods in Psychology—Handbook of Psychology, vol. 2, pp. 581–606. Wiley, New York (2003)

Werner, J.: Lineare Statistik—Allgemeines Lineares Modell. Weinheim, BELTZ (1997)

Werner, J. (ed.): Zeitreihenanalysen mit Beispielen aus der Psychologie. Logos, Berlin (2005)

Wherry, R.: A new shrinkage formula for predicting the shrinkage of the coefficient of multiple correlation. Ann. Math. Stat. **2**, 440–457 (1931)

Witt, R., Witte, A.: Crime, prison, and female labor supply. J. Quant. Crimonol. **16**, 69–85 (2000)

Yule, G.U.: Why do we sometimes get nonsense-correlations between time-series? J. R. Stat. Soc. **89**, 1–64 (1926)