

Coherence: a new approach for analyzing interrelated serial data

Rebekka Mumm¹  · Detlef Groth²  · Michael Hermanussen³ 

¹University of Potsdam, IBB, Human Biology, 14469 Potsdam, Germany

²University of Potsdam, IBB, Bioinformatics, 14476 Potsdam-Golm, Germany

³Aschauhof 3, 24340 Eckernförde – Altenhof, Germany.

Citation:

Mumm, R./Groth, D./Hermanussen, M. (2024). Coherence: a new approach for analyzing interrelated serial data, Human Biology and Public Health 2. <https://doi.org/10.52905/hbph2024.2.95>.

Received: 2024-11-11

Accepted: 2024-11-21

Published: 2024-12-20

Copyright:

Unless otherwise indicated, this work is licensed under a [Creative Commons License Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/). This does not apply to quoted content and works based on other permissions.

Conflict of Interest:

There are no conflicts of interest.

Correspondence to:

Michael Hermanussen
email: michael.hermanussen@gmail.com

Keywords

secular trend, serial data, locally structured correlation, coherence, breakpoint analysis

Abstract

Background Serial public health data may or may not reflect living conditions, political background and/or certain targeted health interventions. Yet, the effect of political events and/or health interventions that only last for a few years or a single legislative period may be difficult to analyze as restricting the range of the variables, i.e. the time interval within which the data were obtained, restricts the power of correlation analyzes.

Objectives To provide a method to eliminate linear trends from serially obtained data and to visualize agreement between these data.

Method We combine information of both the X- and the Y-axis and assess the agreement (coherence) between variables by clockwise (positive correlation) or anticlockwise (negative correlation) rotation of the coordinates.

We provide an illustrative historic example of the coherence of infant mortality as an indicator of public health and body height in Germany between 1885 and 1995.

Results Calculating coherence between correlating variables eliminates linear trends and leaves residuals that correspond to local correlation coefficients.

The substantial changes in the coherence pattern of infant mortality and height exemplify the changes in the interaction between these variables during the transition from late feudalism to modern democracy.

Conclusion Assessing coherence patterns enables a sensitive assessment of inhomogeneity and temporal trend changes in serially obtained correlated variables.

Take home message for students

Coherence patterns combine the information of both the X- and the Y-axis of correlating variables by rotating the coordinates. Assessing coherence patterns enables a sensitive assessment of inhomogeneity and temporal trend changes in serially obtained correlated variables.

Introduction

Living conditions in Europe have rapidly changed in the last 150 years. The national health systems have improved, infant mortality has decreased, life expectancy has risen, the percentage of students has risen, European economies and their gross domestic products have grown, and body height of military conscripts has reached levels that have never been observed in Europe for the last 10.000 years (Rosenstock et al. 2019). In terms of statistics, these indicators of living conditions are time series, and they are strongly correlated. Neither of them however, when compared with each other, can provide an unequivocally correct measurement of, or can be considered causal to, the social and economic transition the European nations experienced during the last centuries.

Living conditions are “non-stationary” time series, i.e., their statistical properties change over time. Non-stationary time series cannot immediately be used in regression models as they create spurious correlations due to common trends in otherwise unrelated or weakly related variables. A solution to the problem is to convert non-stationary time series data into stationary time series. Several techniques have been proposed. Among these are differencing ($Y'_t = Y_t - Y_{t-1}$); logarithmic transformation ($Y'_t = \ln(Y_t)$); decomposition to break down the series into trend components, seasonal components, and residual components; seasonal differencing ($Y'_t = Y_t - Y_{t-12}$); smoothing by moving averages, detrending using regression models and subtracting the trend from the series Y_t , and others. All of them have in common that they focus on the series Y_t .

We suggest a different approach. Since it is not always obvious which of the variables of interest is the dependent and which the independent variable, we refrain from fo-

cus on the dependent variable Y_t , and instead suggest an approach that essentially corresponds to a stretched Bland-Altman plot (Bland and Altman 2011). Bland and Altman related the differences between two measurements, e.g., obtained by different measuring techniques, to the mean values of both measurements. Thus, they combined information of both variables, which essentially corresponds to a clockwise rotation of the X- and the Y-axis similar to the calculation of locally structured correlations (LSC) (Mumm et al. 2022). Both approaches eliminate linear trends and assess agreement between the variables. We use this approach both for positively and negatively correlated variables.

As we cannot say that the increase in body height of conscripts “agrees” with the drop in infant mortality or any other indicator of living conditions – each of these variables denote very different aspects of life – we will use the term “coherence” instead, and speak about the degree of coherence between variables and about changes of coherence over time. E.g., in the case of infant mortality in Germany and height of German conscripts which is highly correlated, we ask: has the coherence between infant mortality in Germany and height of German conscripts changed since the early 20th century? Does coherence differ between the years when Germans still lived in a rigid feudal society, and the years after the Second World War when the people recovered from the terror of the Nazi regime, from war, from foreign occupation, and poverty (Baghdady and Würz 2016), and experienced the “economic miracle (Wirtschaftswunder)” (Bökenkamp 2010)? We first demonstrate the calculation of coherence in simulated data and then provide an example of coherence between mean body height and infant mortality in different periods of German history.

Method

Variables can differ in range and/or dimension (such as body height (cm) and infant mortality (deaths per thousand)). Combining information of such variables requires standardization, e.g. transformation like $x_t = \frac{x_i - \bar{x}}{s}$ with x_i the individual value, \bar{x} the mean and s the standard deviation (z-transformation). If standardized variables are normally distributed, the linear regression line diagonally crosses the origin of their coordinates (by 45° , positive correlation, or anticlockwise, by -45° , negative correlation). The angle of the linear regression line may differ from 45° if variables are not standardized or not normally distributed.

In the case of a positive correlation between the variables, the difference ($y-x$) is plotted on the ordinate and the sum ($y+x$) on the abscissa, which leads to a clockwise rotation of the coordinate system by the angle of the linear regression. In the case of a negative correlation, the sum ($x+y$) is plotted on the ordinate and the difference ($x-y$) on the abscissa, which results in an anti-clockwise rotation of the coordinate system. In both cases, the approach eliminates the trend and leaves residuals that correspond to the correlation coefficient (Mumm et al. 2022).

The simulation

We simulate difference stationary time series (random walk series). The time series values Y_t are caused by a white noise disturbance variable U_t and a superimposed stochastic trend T , of the form

$$Y_t = Y_{t-1} + U_t + T$$

We correlate two time series. $T_1 = T_2 = 0$ may result in spurious correlations. $T_1 > 0$ and $T_2 > 0$ or $T_1 < 0$ and $T_2 < 0$ results in significant positive correlation, $T_1 < 0$ for variable1 and $T_2 > 0$ for variable2, or $T_1 > 0$

for variable1, and $T_2 < 0$ for variable2 results in significant negative correlations. In addition, we allow changes of T over time so that the slope of the correlation intermittently changes.

The example

Height of military conscripts and recruits of the armed forces of the German Empire and the Federal Republic of Germany was available for all birth cohorts from 1865 to 1975, except the years 1866–1874, 1877, 1878, 1895–1899, 1922–1937 (Jaeger et al. 2001; Rass 2023; Rass and Rohrkamp 2009; Institut für Wehrmedizinische Statistik und Berichtswesen, Remagen 1995; Nowak et al. 2020). We added data of 19-year old school boys from Burg/Fehmarn, North Germany (birth cohort 1927) (Träbert 1948) who represented a population of 17,000 city dwellers in an overall more rural community; and data of a 1954 school survey from Munich, Bavaria (birth cohort 1947) (Bach 1965). Missing values were added by flexible imputation (van Buuren 2019).

As we consider adult height to result from the cumulative effect of all economic, nutritional, social, educational, political and other factors that are critical to the growth of infants, children and adolescents, we refer adult height to the historic year at age 20, i.e., the birth cohorts 1865–1975 are referred to the historic years 1885–1995.

Based on political and economic events, we considered four major disruptions in the German history (World War I, the seizure of control by the Nazis, the founding of the Federal Republic of Germany, and later the increasing Europeanization) and accordingly divided the time from 1885 to 1995 into five periods.

(1) 1885–1916 (born 1865–1896): Late feudal period with three-class suffrage. Due to major hardship imposed on the civilian population during the First World War, we terminate this period already in 1916;

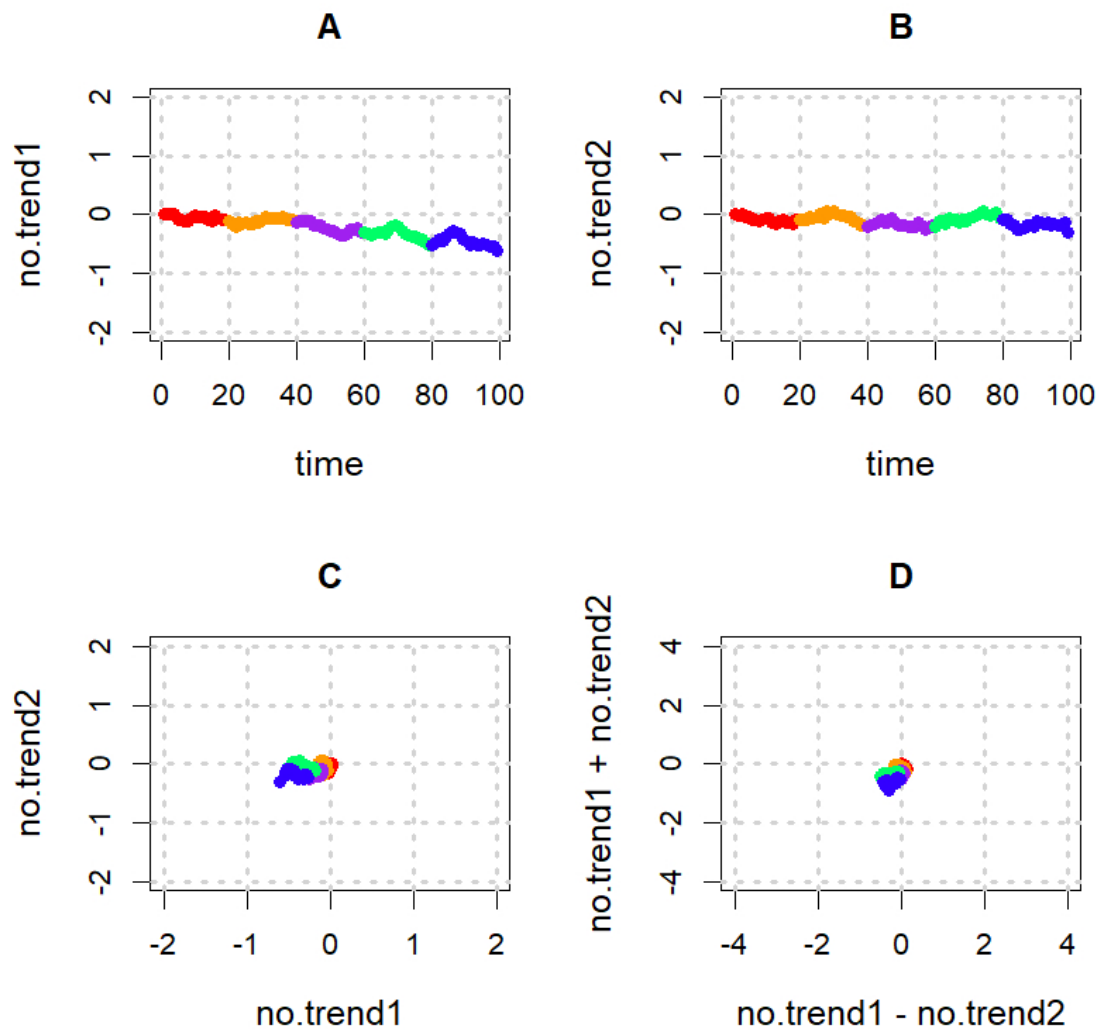


Figure 1 Two difference stationary time series created by a white noise disturbance variable (random walk). Panel A and panel B: the time series; panel C: spurious correlation between the time series; panel D: the coherence pattern. Colors indicate time segments of 20 units.

(2) **1918–1933 (born 1898–1913)**: Period of significant political instability (Glatzer and Glatzer 1983), famine (1916–1919), hyperinflation (1923) and economic failure, and the Great Depression in the late 1920s, followed by political radicalization until 1933;

(3) **1933–1947 (born 1913–1927)**: Nazi regime and Second World War with 2.25 million homes being destroyed and 2.5 million damaged, foreign occupation, million-fold displacement, and poverty (Baghdady and Würz 2016). Due to lack of data in the years 1945 and 1946, this period was extended to 1947 including the years of post-war famine (Grossmann 2011);

(4) **1947–1973 (born 1927–1953)**: foreign occupation and formation of the economically fast growing Federal Republic of Germany (FRG). The era included the economic miracle (Wirtschaftswunder) in the FRG and ended with the oil crisis in 1973 (Bökenkamp 2010);

(5) **after 1973 (born after 1953)**: “European period”. Germany increasingly integrated into the European community, with ever closer economic ties to its neighbor countries, rising mobility and mass tourism, and its increasingly unfair distribution of wealth (Blanchet and Martínez-Toledano 2023).

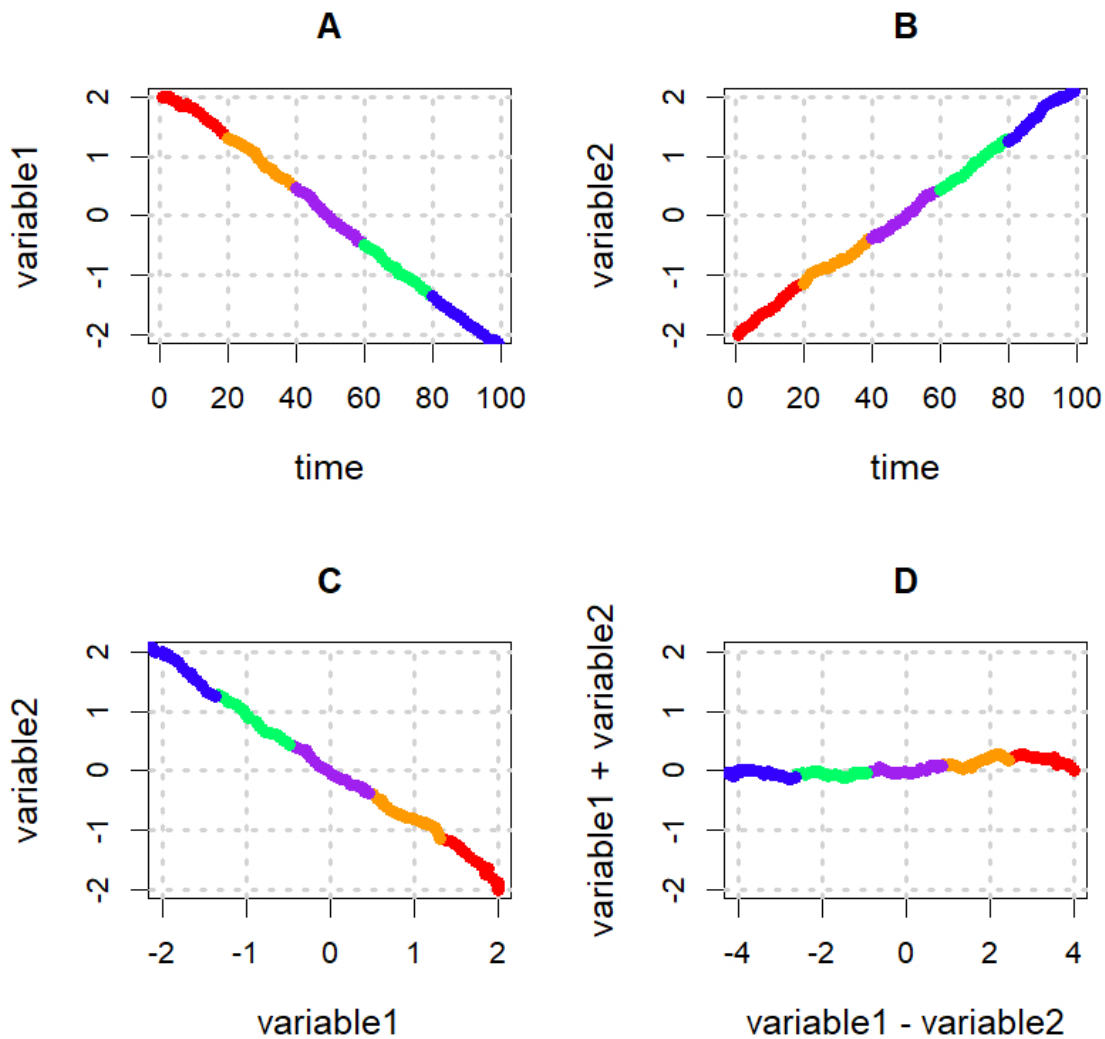


Figure 2 Two difference stationary time series created by a white noise disturbance variable and a superimposed stochastic trend. Panel A (negative trend) and panel B (positive trend): the time series; panel C: correlation between the time series; panel D: the coherence pattern. Colors indicate time segments of 20 units.

We exemplify the coherence between trends in body height and infant mortality (O'Neill 2024).

Results

First, we created two difference stationary time series by a white noise disturbance variable (random walk, Figure 1). There is some spurious correlation between the two variables (panel C), and the coherence pattern is nondescript (panel D).

Thereafter, we created difference stationary time series by a white noise disturbance variable and a superimposed stochastic trend. Figure 2 shows a representative example of a negative correlation. The coherence pattern (panel D) of random walk series does not coincide with the colored time segments.

Confounders affect the slope of a correlation. When confounders change over time, trends change. Coherence plots amplify trend changes. Figure 3 exemplifies inversely directed coherence patterns of two negatively correlated difference stationary time series. The patterns augment

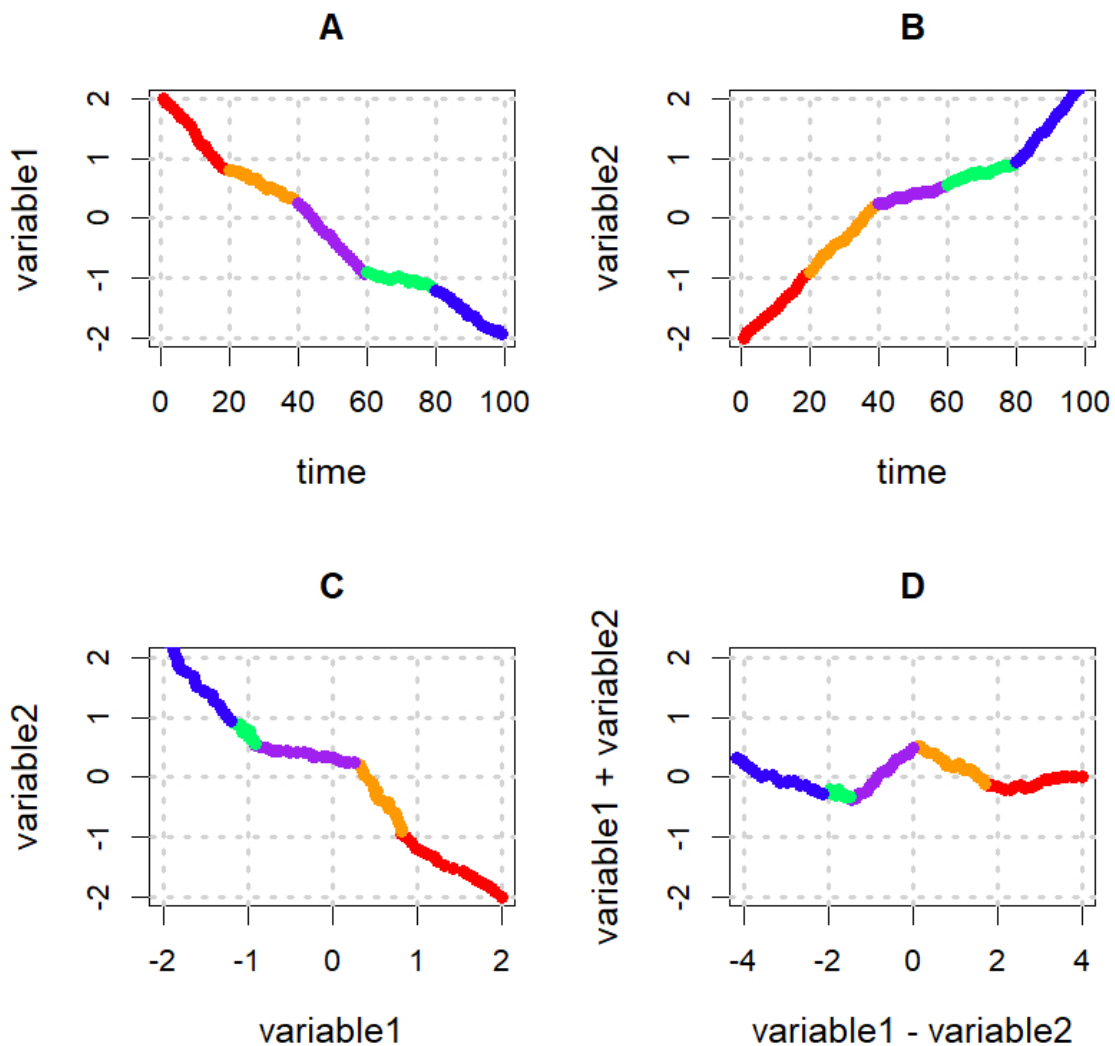


Figure 3 Two difference stationary time series created by a white noise disturbance variable and superimposed stochastic trends (confounders) that vary at intervals of 20 units. Panel A and panel B: the time series; panel C: correlation between the time series; panel D: the coherence pattern. Colors indicate time segments of 20 units at which the confounders change.

and highlight the different superimposed trends. The Figures 4 and 5 exemplify coherence patterns of difference stationary time series that are partially negatively, and partially positively correlated. We plot anticlockwise (4) and clockwise (5) rotated coherence patterns. Analog patterns occur when simulating random walk series with positive correlations.

Similar patterns occur when analyzing local correlations (Mumm et al. 2022) between infant mortality and conscript body height in the five major periods of German history (1885 to 1995) (Figure 6). The changes in the local correlation are empha-

sized by the inversely directed coherence patterns (panel D) and suggest differential confounding mechanisms for the red (late feudal) and the yellow (Weimar Republic), for the green and blue (democratic), and for the interposed purple period of the Nazi regime and the Second World War.

Discussion

Correlating serial data with other serial data obtained within the same time interval, often reveals trends that are neither

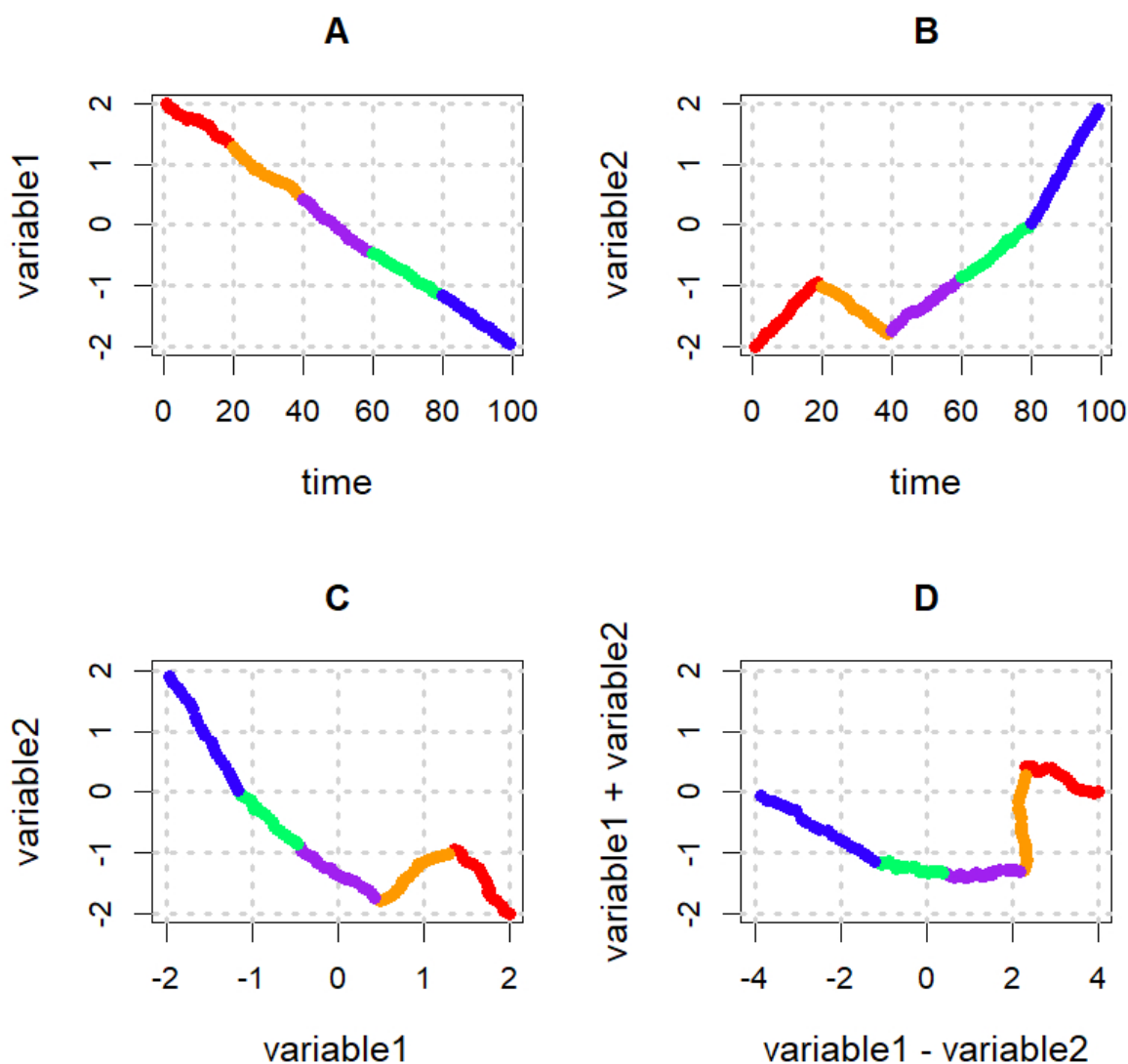


Figure 4 Two difference stationary time series created by a white noise disturbance variable and superimposed stochastic trends (confounders) that vary at intervals of 20 units. The time series are negatively correlated; the coordinate system is anticlockwise rotated; the figure corresponds to the previous figures.

causal nor homogenous. The strength of the association may vary, the slope of the regression lines may temporarily change, and partitioning and search for particular segments and “breakpoints” may be required (Jones and Molitoris 1984). Serially obtained public health data are particularly heterogenous and may or may not reflect living conditions, political background and/or certain targeted health interventions.

Yet, the effect of political events and/or health interventions that only last for a

few years or a single legislative period may be difficult to analyze as restricting the range of the variables, i.e. the time interval within which the data were obtained, restricts the power of correlation analyzes (Thorndike 1949; Bland and Altman 2011; Lakes 2013). Colored Bland-Altman plots have been used to mark points that „belong together“ through a third variable (e.g. (Lehnert 2015) or (MedCalc 2024)), but this or similar approaches have not been routinely used for the analysis of serial public health data.

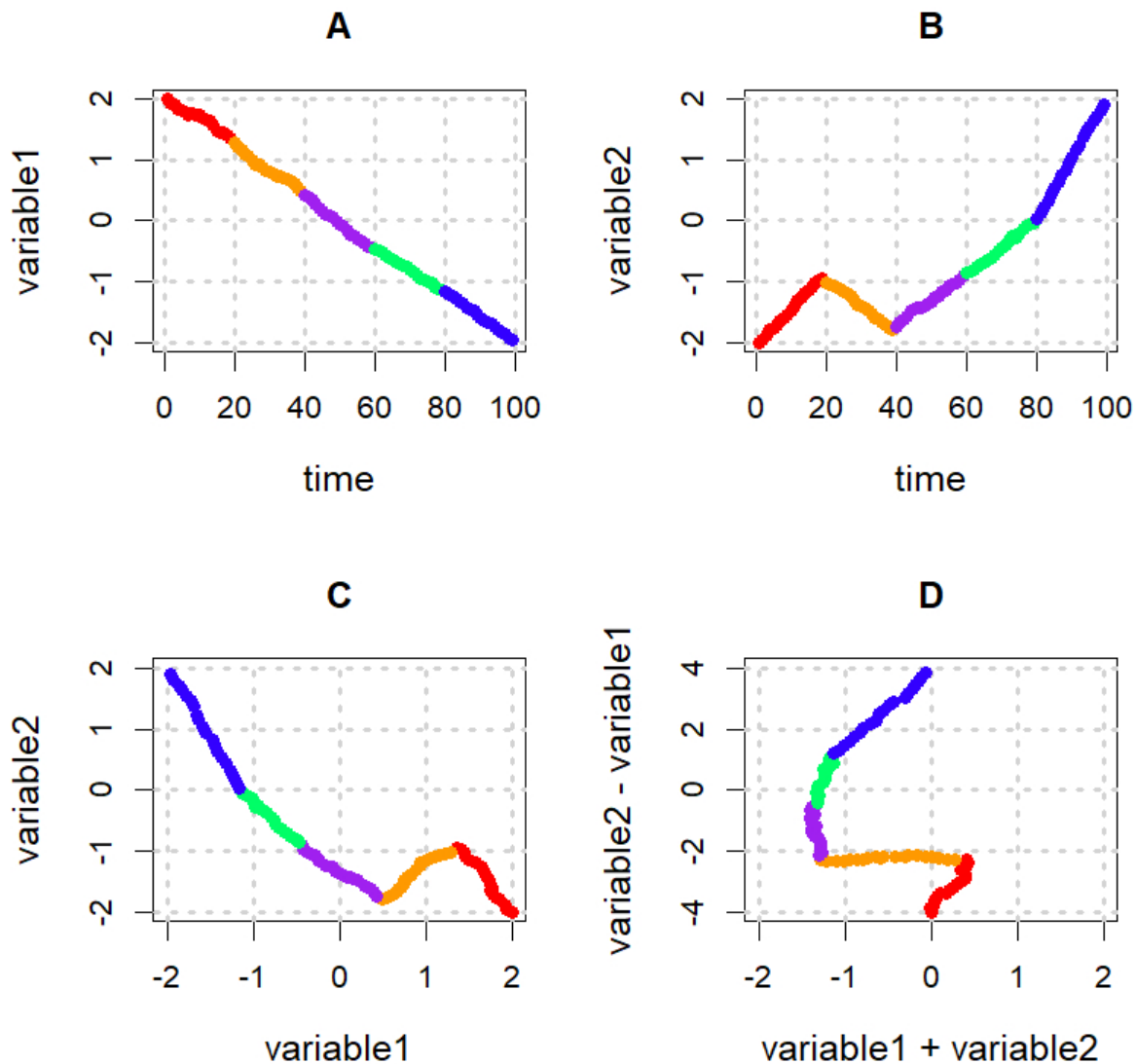


Figure 5 The same two time series as shown in Figure 4, but the coordinate system is clockwise rotated. The coherence pattern mirrors the pattern in Figure 4.

We recently proposed “locally structured correlations” (LSC) to depict inhomogeneity in the association of variables (Mumm et al. 2022). LSC combine information of both the X- and the Y-axis by rotating the coordinates and essentially present the coherence between serial data after eliminating prevalent trends. Here, we further developed this approach for the use in serially obtained both positively and negatively related data to facilitate the detection of disrupting events or intermittently occurring confounders.

Coherence patterns are not correlations. They do not relate the dependent and the

independent variables, but they relate the sums of both and the differences between them. Coherence patterns are sensitive to inhomogeneity and to temporal trend changes. Coherence patterns are similar to principal component analysis in that they reduce the dimensionality of large datasets by separating prevalent trends (Jolliffe and Cadima 2016). The technique enables visualizing local characteristics of the residuals along the time axis, and thus, may also be considered related to breakpoint analysis (Jones and Molitoris 1984).

The irregular coherence patterns of infant mortality and body height in Germany

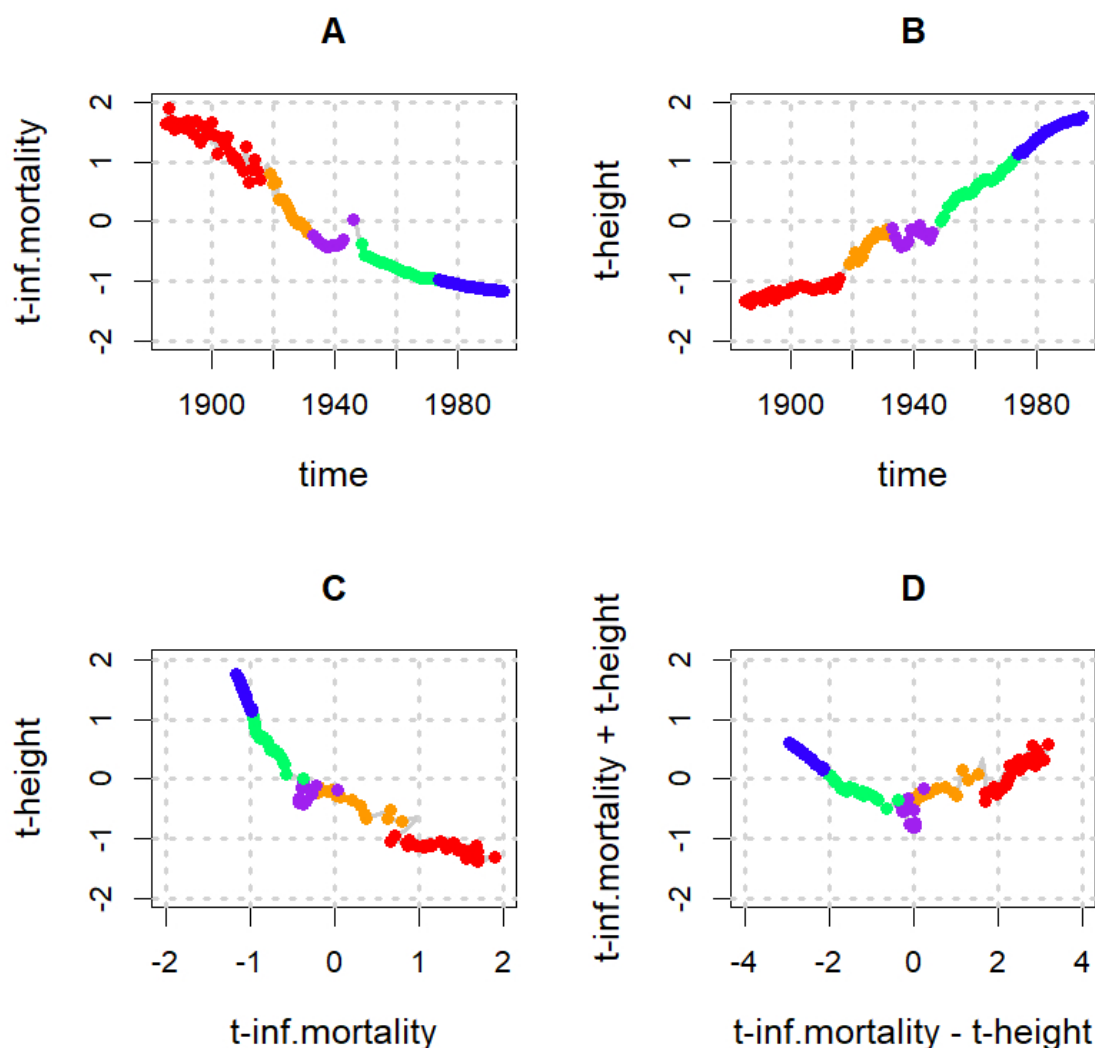


Figure 6 Coherence between z-transformed infant mortality (t-inf.mortality) and z-transformed body height (t-height). Panel A and panel B: the time series; panel C: correlation between the time series; panel D: the coherence pattern.

between 1885 and 1995 exemplify the substantial temporal changes in the interaction between the two historic variables and suggest similarly substantial changes in the mechanisms that have confounded these interactions.

Acknowledgements

Thanks are due to Christian Aßmann, Bamberg, for critical comments and suggestions.

References

- Bach, F. (1965). Untersuchungen über Körpergröße und Körpergewicht bei Kindern verschiedener sozialer Berufsgruppen. *Anthropologischer Anzeiger* 27, 289–296.
- Baghdady, Anne/Würz, Markus (2016). *Leben in Trümmern*. Available online at <https://www.hdg.de/lemo/kapitel/nachkriegsjahre/alltag/leben-in-truemmern.html>.
- Blanchet, T./Martínez-Toledano, C. (2023). Wealth inequality dynamics in Europe and the United States: understanding the determinants. *Journal of Monetary Economics* 133, 25–43. Available online at <https://www.sciencedirect.com/journal/journal-of-monetary-economics/vol/133/suppl/C>.
- Bland, J. M./Altman, D. G. (2011). Correlation in restricted ranges of data. *British Medical Journal Clinical Research* 342, d556. <https://doi.org/10.1136/bmj.d556>.

- Bökenkamp, G. (2010). *Das Ende des Wirtschaftswunders. Geschichte der Sozial-, Wirtschafts- und Finanzpolitik in der Bundesrepublik 1969 – 1998*. Berlin, Boston, De Gruyter, Oldenbourg.
- Glatzer, D./Glatzer, R. (1983). *Berliner Leben 1914 – 1918. Eine historische Reportage aus Erinnerungen und Berichten*. Berlin, Rütten & Löning.
- Grossmann, A. (2011). Grams, calories, and food: languages of victimization, entitlement, and human rights in occupied Germany, 1945–1949. *Central European History* 44, 118–148. <https://doi.org/10.1017/S0008938910001202>.
- Institut für Wehrmedizin und Berichterstattung, Remagen (1995). *Körperhöhen westdeutscher Wehrpflichtiger der Geburtsjahrgänge 1938 bis 1975*. Remagen, Wehrmedizin und Berichterstattung.
- Jaeger, U./Zellner, K./Kromeyer-Hauschild, K./Lüdde, R./Eisele, R./Hebebrand, J. (2001). Körperhöhe, Körpergewicht und Body Mass Index bei deutschen Wehrpflichtigen. *Historischer Rückblick und aktueller Stand. Anthropologischer Anzeiger* 59 (3), 251–273. <https://doi.org/10.1127/anthranz/59/2001/251>.
- Jolliffe, I. T./Cadima, J. (2016). Principal component analysis: a review and recent developments. *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences* 374 (2065), 20150202. <https://doi.org/10.1098/rsta.2015.0202>.
- Jones, R. H./Molitoris, B. A. (1984). A statistical method for determining the breakpoint of two lines. *Analytical Biochemistry* 141 (1), 287–290. [https://doi.org/10.1016/0003-2697\(84\)90458-5](https://doi.org/10.1016/0003-2697(84)90458-5).
- Lakes, K. D. (2013). Restricted sample variance reduces generalizability. *Psychological Assessment* 25 (2), 643–650. <https://doi.org/10.1037/a0030912>.
- Lehnert, B. (2015). *Bland-Altman-Plots*. University Medicine Greifswald, Germany. Available online at <https://cran.r-project.org/web/packages/BlandAltmanLeh/vignettes/Intro.html>.
- MedCalc (2024). *Bland-Altman plot with multiple measurements per subject*. MedCalc Software Ltd. Available online at <https://www.medcalc.org/manual/blandaltmanmultiple.php>.
- Mumm, R./Scheffler, C./Hermanussen, M. (2022). Locally structured correlation (LSC) plots describe inhomogeneity in normally distributed correlated bivariate variables. *Archives of Public Health* 80 (1), 30. <https://doi.org/10.1186/s13690-021-00748-4>.
- Nowak, O./Liczbinska, G./Piontek J. (Eds.) (2020). *Were inequalities in trends of body size observed in 19th/20th centuries? Economic and nutritional determinants of the height, weight and BMI in populations from the Prussian sector*. Abstract for a conference. Poznan, Poland.
- O'Neill, Aaron (2024). *Infant mortality rate (under one year old) in Germany from 1840 to 2020*. Available online at <https://www.statista.com/statistics/1042395/germany-all-time-infant-mortality-rate/>.
- Rass, Christoph (2023). *Anonymisierter biographischer Datensatz zu Mannschaften und Unteroffizieren von Heer, Luftwaffe und Waffen-SS, 1939–1945*. Online Anonymisierter biographischer Datensatz zu Mannschaften und Unteroffizieren von Heer, Luftwaffe und Waffen-SS, 1939–1945. *GESIS Studiennummer 8410. Datenbank und Handbuch Personenbezogene Quellen zu den bewaffneten Formationen des Dritten Reiches*. GESIS. Available online at <https://www.deutsche-digitale-bibliothek.de>.
- Rass, Christoph/Rohrkamp, René (2009). *Überregionale Erschließung personenbezogener Quellen zu Angehörigen der bewaffneten Formation des 'Dritten Reiches' (Deutsche Soldaten 1939 bis 1945)*. GESIS. Available online at Online verfügbar unter Anonymisierter biographischer Datensatz zu Mannschaften und Unteroffizieren von Heer, Luftwaffe und Waffen-SS, 1939–1945, GESIS Studiennummer 8410.
- Rosenstock, E./Ebert, J./Martin, R./Hicketier, A./Walter, P./Groß, M. (2019). Human stature in the Near East and Europe ca. 10,000–1000 BC: its spatiotemporal development in a Bayesian errors-in-variables model. *Archaeological and Anthropological Sciences* 11 (10), 5657–5690. <https://doi.org/10.1007/s12520-019-00850-3>.
- Thorndike, R. L. (Ed.) (1949). *Personnel selection: test and measurement techniques*. New York, Wiley.
- Träbert, H. (1948). *Ergebnisse einer Reihenuntersuchung von 17000 Personen im Frühjahr 1947*. *Zeitschrift für die gesamte Innere Medizin und ihre Grenzgebiete* 3.
- van Buuren, S. (2019). *Flexible imputation of missing data*. 2nd ed. Boca Raton, Florida, Chapman & Hall/CRC.