Sławomir Dorosiewicz, Tadeusz Dorosiewicz*

SIMILARITY ANALYSIS OF GROWTH CYCLES

1. Introduction

This paper introduces the definition of similarity measure of two functions. This measure is a quantitative characteristic of similarity of functions and it can be helpful in comparison of the sets of qualitative and quantitative data on economic activity. The measure includes two special cases: functions defined on intervals and functions with discrete domains. Similarity measure admits many generalisations, e.g. it can be extended on the space of stochastic processes. This fact allows testing several hypotheses about mutual properties of time series. These measures are then applied to analyse the similarity between the actual dynamics of the aggregate economic activity in Poland and its representation by survey data.

It seems important to define the tools for quantifying the similarity of functions and the sets of data. These tools, called in this paper *similarity measures,* enable us to compare quantitative time series and survey data and quantify the similarity between them. In the suggested comparative method, similarity of economic dynamics is defined by the measure of the compatibility of the functions. The considered functions are mainly dynamics functions (see Dhrymes, 1971, and Judge et al., 1985), whose analytical form we consider as a mathematical model of the dynamics of compared processes. There are many possible similarity measures. Some of them seem to be useful in theoretical and empirical studies. This paper concentrates on definitions and properties of some of them.

2. Similarity measure

A tool for comparing similarity of functions is a similarity relation denoted by $\succeq$. Technically, $\succeq$ is a binary relation on $X \times X$, where X is given functional space. We can interpret $(f_1, g_1) \succeq (f_2, g_2)$ as "f_1 is at least as similar to g_1 as f_2 is similar to g_2." The axioms of similarity relations may be given by the following definition.

Definition 1. A relation $\succeq$ on $X \times X$ will be called *similarity relation* if the following conditions hold:

---

* Sławomir Dorosiewicz is Professor of Mathematics at the Department of Economic Analysis, Warsaw School of Economics; Tadeusz Dorosiewicz is Professor of Economics at the Department of Management and Finance, Warsaw School of Economics.

This is an extended and corrected version of the paper "Similarity Measures of Growth Trends and Cycles," presented at the 26th CIRET Conference in Taipei in 2002.

(a) $\succeq$ is reflexive, i.e. $(f,f)\succeq(f,f)$ for all $f\in X$;

(b) $\succeq$ is transitive:

$$(f_1,g_1)\succeq(f_2,g_2)\wedge(f_2,g_2)\succeq(f_3,g_3)\Rightarrow(f_1,g_1)\succeq(f_3,g_3);$$

(c) $(f_1,g_1)\succeq(f_2,g_2)\Rightarrow(g_1,f_1)\succeq(f_2,g_2)$ and

$(f_1,g_1)\succeq(f_2,g_2)\Rightarrow(f_1,g_1)\succeq(g_2,f_2)$.

(d) if $f\in X$, then (f,f) is the maximal element of $\succeq$.

Reflexivity is sensible property for similarity. Transitivity is a strong assumption. It goes to a concept of rationality. Axiom (c) follows from the symmetric nature of similarity: similarity of the first object to second one is "the same" as the similarity of the second object to the first. Axiom (d) is also natural. Any object is more similar to itself than to any other object.

From $\succeq$ we can derive strict similarity $\succ$ and indifference $\sim$ relations:

$$(f_1,g_1)\succ(f_2,g_2)\Leftrightarrow(f_1,g_1)\succeq(f_2,g_2)\text{ but not }(f_2,g_2)\succeq(f_1,g_1),$$

$$(f_1,g_1)\sim(f_2,g_2)\Leftrightarrow(f_1,g_1)\succeq(f_2,g_2)\text{ and }(f_2,g_2)\succeq(f_1,g_1).$$

Definition 2. Let $\succeq$ be a similarity relation on X. If there exists a function $\mu : X\times X\rightarrow\Re$ such that

$$(f_1,g_1)\succeq(f_2,g_2)\Leftrightarrow\mu(f_1,g_1)\geq\mu(f_2,g_2)\quad\forall f_1,f_2,g_1,g_2\in X, \tag{1}$$

then μ will be called the *similarity measure* (SM) or, equivalently, $\succeq$ is represented by the *similarity measure*.

Of course, SM is a weak order on $X\times X$. If $\succeq$ is represented by SM, then $\succeq$ is complete (i.e. for each (f_1,g_1) and (f_2,g_2) we have $(f_1,g_1)\succeq(f_2,g_2)$ or $(f_2,g_2)\succeq(f_1,g_1)$), therefore $\succeq$ is the preference relation on X. In this case SM is the utility function for $\succeq$. It easily follows from Debreu theorem[1] that continuity of $\succeq$ is sufficient for the existence of the similarity measure for $\succeq$.

[1] This is the well-known sufficient condition for a preference relation on a set Y to be

Theorem 1. *If $\succeq$ is continuous and complete similarity relation on $X \times X$, then $\succeq$ is represented by a SM μ. Moreover, μ can be chosen such that $|\mu| \leq 1$.*

In practice it is more natural to define first the measure μ and then a similarity relation according to formula (1).

There are of course many ways to define SM. We now introduce one of them and generalise it in several ways. We start from the following example of similarity measures of differentiable functions defined on the interval.

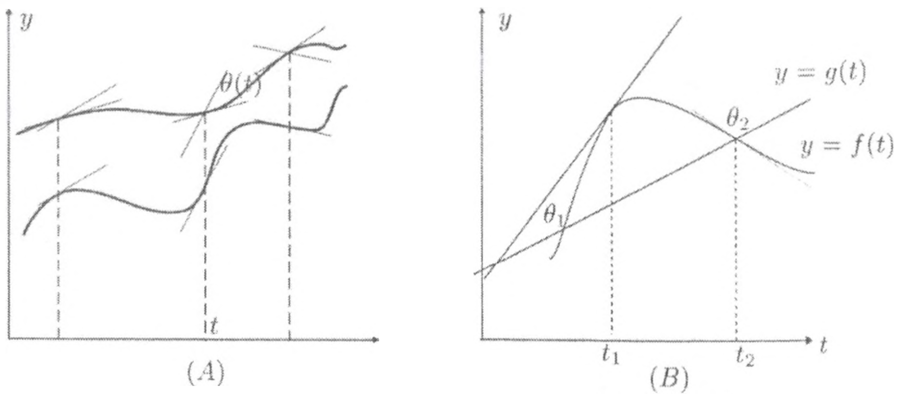
Example 1. (See Dorosiewicz et al.,1998). Let $I =]a,b[$, where $-\infty \leq a < b \leq \infty$, be given interval, $C^1(I, \Re)$ – the space of all continuously differentiable real-valued functions on I. We define the similarity measure in two steps:

(a) the formula of density of measure in a given point $t \in I$,

(b) definition of SM on I.

The density of the similarity measure of $f, g \in C^1(I, \Re)$ in the point $t \in I$ is the cosine of an angle $\theta(t)$ between the tangent lines to the graphs of f and g in the points $(t, f(t))$ and $(t, g(t))$ respectively (see Fig.1 (A)).

Figure 1. (A) Similarity measure of smooth functions. (B) The construction of the similarity measure: the difference between $\mu(f, g)(t_1)$ and $\mu(f, g)(t_2)$



The functions f, g are increasing in the neighbourhood of t_1, hence we have $\mu(f, g)$

---

represented by utility function: if a preference relation $\succeq$ is continuous (i.e. for each $y \in Y$ the sets $\{y \in Y : y \succeq x\}$ and $\{x \in Y : x \succeq y\}$ are closed), then $\succeq$ is represented by a utility function (i.e. there exists $u : Y \rightarrow \Re$ such that $x \succeq y \Leftrightarrow u(x) \geq u(y)$ for all $x, y \in Y$).

$(t_1) = \cos\theta_1$, where $\theta_1 \in [0, \frac{\pi}{2}]$. In the point t_2, where f decreases, g increases, $\mu(f, g)(t_2) = \cos\theta_2$, where $\theta_2 > \frac{\pi}{2}$, so $\mu(f, g)(t_2) < 0$.

Let us consider the following situation showed in the Fig. 1 (B). If $\theta_1 = \pi - \theta_2$, then obviously $\cos\theta_1 = \cos\theta_2$, but it seems that the function f is more similar to g at t_1 than in t_2 (in the neighbourhood of t_1 both f and g increase, in t_2 they do not). The measure of similarity expressed by the following formula will reflect this:

$$\mu(f,g)(t) = \mathrm{tg}\theta(t)\vartheta(f'(t)g'(t)) = \frac{|1 + f'(t)g'(t)|}{\sqrt{(1+(f'(t))^2)(1+(g'(t))^2)}}\vartheta\left(f'(t)g'(t)\right), \tag{2}$$

where $\mu(f,g)(t)$ denotes the value of measure in the point t, $\vartheta(a)$ which is equal 1 if $a \geq 0$ and $-1 + \exp(x)$ elsewhere.

Using the similarity measure in point t we can define the value of the measure on I, where $-\infty \leq a < b \leq \infty$:

$$\mu_I(f,g) = \liminf_{\substack{\epsilon \to b-0 \\ \delta \to a+0}} \frac{1}{\epsilon - \delta}\int_\delta^\epsilon \mu(f,g)(t)dt. \tag{3}$$

If $I = [a,b]$ $(-\infty < a < b < \infty)$, then $\mu(f,g) = \frac{1}{b-a}\int_a^b \mu(f,g)(t)dt$

For sufficiently smooth functions (i.e. with continuous derivatives of the order n for some $n \geq 1$), one can define the SM by formula (3), with $\mu(f,g)(t)$ replaced by:[2]

$$\mu^{(n)}(f,g)(t) = \sum_{p=0}^{n-1} \omega_p\, \mu(f^{(p)}, g^{(p)})(t) \tag{4}$$

with the positive weights ω_p $(p = 0,\ldots,n-1)$. These weights may be for example harmonic: $\omega_p = \frac{1}{p+1}\left(\sum_{j=1}^{n}\frac{1}{j}\right)^{-1}$.

Let $C^n(I,\Re)$ denote the space of all functions mapping a compact interval I into $\Re$ with continuous n-th derivative. The simple properties of (4) are collected in the following

Proposition 1.

(a) *$\mu^{(n)}$ is normalised, that is* $-1 \leq \mu^{(n)}(f, g) \leq 1$;

---

[2] If $n = 1$, then of course μ_1 is the same as the measure given by (2).

(b) $\mu^{(n)}(f,f) = 1$. *Conversely, if* $\mu^{(n)}(f,g) = 1$, *then* $f = g + c$ *for some constant function* c;

(c) $\mu^{(n)}(f,g) = \mu^{(n)}(g,f)$;

(d) *for any constant c:* $\mu^{(n)}(f,g) = \mu^{(n)}(f,g+c)$;

(e) $\mu^{(n)}$ *is continuous in the norm topology of* $C^n(I,\Re)$ *defined by the norm* $\|f\| = \Sigma_{p=0}^{n} \sup_{t\in I} |f^{(p)}(t)|$.

Example 2. Now let us consider the case with discrete domain

$$D = \{1, ..., m\}, \tag{5}$$

where $m \geq \infty$, and two real functions f, g defined on D. In this case, the measure of similarity of f and g is the average value of measures in the points of D:

$$\mu(f,g) = \begin{cases} \liminf\limits_{n\to\infty} \frac{1}{n-1} \sum\limits_{t=1}^{n-1} \mu(f,g)(t) & if \quad m = \infty, \\ \frac{1}{m-1} \sum\limits_{t=1}^{m-1} \mu(f,g)(t) & if \quad m < \infty, \end{cases} \tag{6}$$

where for $t \in D$:

$$\mu(f,g)(t) = \frac{|1 + \Delta f(t)\Delta g(t)|}{\sqrt{(1+(\Delta f(t))^2)(1+(\Delta g(t))^2)}} \vartheta(\Delta f(t)\Delta g(t)) \tag{7}$$

and for $\phi = f,\ g$ $\Delta\phi$ defines the (first-order) difference of ϕ:

$$\Delta\phi\,(t) = \phi(t+1) - \Delta\phi(t).$$

The equations (6) and (7) are very similar to (2) and (3). The difference $\Delta\phi$ corresponds in a natural way with the derivative ϕ'. The formula equivalent to (4) can be obtained in the natural way: high order derivatives should be replaced by appropriate high-order differences.

The construction from the last example can be generalised in several ways. Namely, let D be a nonempty set, $\mathcal{M}$ – a given σ-algebra of subsets of D, and $\nu : \mathcal{M} \to \Re$ a nonnegative and σ-finite measure such that $\nu(D) > 0$ and the map: $t \mapsto \mu(f,g)(t)$ is ν-measurable[3]. The measure of similarity of functions f, $g : D \to \Re$ is defined by:

[3] For example, μ may be given by (4) for some n and ν may be the Lebesgue measure on $\Re$

$$\mu_D(f,g) = \inf_{(D_n)} \liminf_{n \to \infty} \frac{1}{\nu\left(\bigcup_{k=1}^{n} D_k\right)} \int_{\bigcup_{k=1}^{n} D_k} \mu(f,g)(t) d\nu(t), \tag{8}$$

where inf covers all partitions (D_n) of D such that ν (D_n) $< \infty$. The σ-finiteness of ν implies, that this family is nonempty. If ν is finite (i.e. ν (D) $< \infty$) and sequentially continuous then (8) can be reduced to

$$\mu_D(f,g) = \frac{1}{\nu(D)} \int_D \mu(f,g)(t) d\nu(t) \tag{9}$$

For example, if $D =]a, b[$, f, g are smooth, and ν denotes the Lebesgue measure on D, then we obtain (3). If D has the form (5) and ν is the computing measure[4] on D, then we obtain formula (6).

3. Similarity of stochastic processes

The ideas mentioned above can be applied to construct the measure of similarity of stochastic processes. The case with continuous-time stochastic processes is much more complex. Formula (2) is not useful in many important cases because it requires the differentiability of the compared processes. Unfortunately, many of stochastic processes are not differentiable in any way[5]. For example, almost all (with probability one) trajectories of separable Wiener process are continuous but not differentiable. Moreover, the Wiener process is not differentiable in mean-square sense (see e.g. Gihman, 1979, and Karatzas, 1991). This implies the necessity of changes in the definition of the similarity measure.

For discrete-time stochastic processes the measure of similarity can be defined analogously to (6) – (7). More precisely, let $\xi = (\xi_t)_{t \in T}$ and $\eta = (\eta_t)_{t \in T}$ ($T = \{1, ..., m\}$) be stochastic processes. It is easy to see that for any $t \in T$:

$$\lambda_t(\xi, \eta) = \frac{|1 + \Delta\xi_t \Delta\eta_t|}{\sqrt{(1 + (\Delta\xi_t)^2)(1 + (\Delta\eta_t)^2)}} \vartheta\left(\Delta\xi_t \Delta\eta_t\right) \tag{10}$$

is random variable. The value of similarity measure of ξ, η can be obtained from the formula analogous to (6):

---

[4] For any $A \subset D$, $\nu(A)$ is the number of elements belonging to A.

[5] In this paper the generalized stochastic processes are not considered.

$$\mu(\xi,\eta) = \frac{1}{m}\sum_{t=1}^{m} \mathrm{E}\left(\lambda_t(\xi,\eta)\right) \quad \text{for } m < \infty, \tag{11}$$

and

$$\mu(\xi,\eta) = \liminf_{n\to\infty} \frac{1}{n}\sum_{t\le n} \mathrm{E}\left(\lambda_t(\xi,\eta)\right) \text{ for } m = \infty. \tag{12}$$

It follows from the Lebesgue majority theorem, that λ_t (ξ, η) has finite all moments (esp. mean), so $\mu\,(\xi, \eta)$ is well-defined.

Example 3. Let us compute the values of similarity measure of selected discrete- time stochastic processes: independent white noises (ε_t), (δ_t) and random walks

$$\xi_t = \xi_{t-1} + \varepsilon_t\,,\ \eta_t = \eta_{t-1} + \delta_t\,,\ P(\xi_0 = 0) = P(\eta_0 = 0) = 1\,,\ t = 0, ..., m. \tag{13}$$

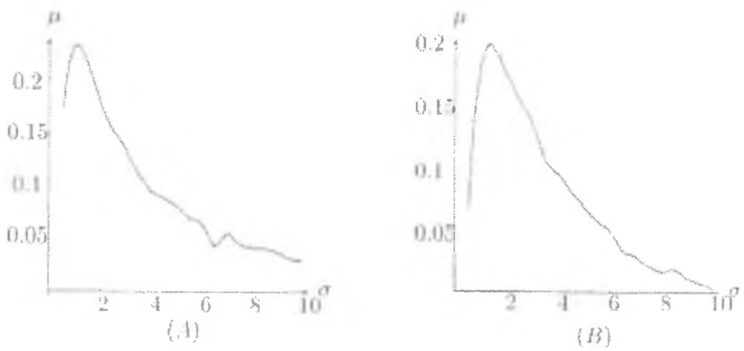
Because of the complexity of formula (11), the integrals may be computed only using numerical procedures like Monte Carlo (MC) simulations:[6]

(A) Similarity μ of two independent Gaussian white noises;

(B) Similarity of independent random walks, see Fig. 2 (B).

The results are shown in Fig. 2.

Figure 2. (A) Similarity of two independent Gaussian white noises with standard deviation σ. (B) Similarity of two independent Gaussian random walks. Dependence of the similarity on standard deviation of the differences: results of MC simulations.



Random variables $\lambda\ (\xi,\eta)\ =\ \frac{1}{m}\ \Sigma_{t=1}^{m}\ \lambda_t\ (\xi,\eta)$ can be used to construct the statistics verifying the hypothesis about (non-) stationarity of process generating given

[6] The computations were done with Mathematica v.3 system.

time series. These statistics may be used together with well-known stationarity tests, i.e. Dickey-Fuller tests (Dickey et al., 1979, 1981), Phillips tests (Phillips et al., 1987, 1988), etc. To show this idea, let us consider independent random walks processes (13), where $t = 1, ..., m$, (ε_t), (δ_t) are independent Gaussian white noises with mean zero and standard deviations σ_1 and σ_2 respectively.

Let B be the time-lag operator: $(B\eta)_t = \eta_{t-1}$, and let $[x]$ denote the integer part of number x. The Lebesgue majority theorem and Lindenberg-Levy theorem imply that the random variables

$$\Lambda = \lambda\left(\frac{\xi}{\sigma_1}, \frac{\eta}{\sigma_2}\right), \Lambda_1 = \lambda\left(\frac{\xi}{\sigma_1}, \frac{B^{[m/2]}\eta}{\sigma_2}\right), \Lambda_2 = \lambda\left(\frac{\xi}{\sigma_1}, 1\right), \tag{14}$$

have finite all the moments and their distributions are asymptotically ($m \to \infty$) normal. This observation gives the asymptotical tests to verify the hypothesis (13).

If the sample is small, the random variables (14) are not in general Gaussian. In this case the sample standard deviation

$$S_m(\xi) = \left(\frac{1}{m-1}\sum_{t=1}^{m}\left(\xi_t - \frac{1}{m}\sum_{i=1}^{m}\xi_i\right)^2\right)^{1/2}$$

may be significantly different from σ_1. The Monte Carlo simulations enable to estimate the quantiles of statistics:

$$\lambda\left(\frac{\xi}{S_m(\xi)}, 1\right). \tag{15}$$

The results depend on the sample size. They are presented in Table 1.

Table 1. Estimated quantiles of random variable

α	α-quantiles of (15) for sample size m:				
	10	20	50	100	150
0.05	0.69	0.77	0.86	0.91	0.93
0.1	0.72	0.80	0.88	0.92	0.94
0.2	0.75	0.82	0.90	0.94	0.96
0.3	0.78	0.85	0.91	0.95	0.96
0.4	0.80	0.86	0.93	0.96	0.97
0.5	0.82	0.88	0.94	0.96	0.98

Formula (15) provides a natural unit-root test for verifying the stationarity of time series. The null hypothesis H_0 and the alternative one H_a are the following:

H_0 : data are drawn by random walk process (13);

H_a : data are generated by stationary $AR(1)$ process

$$\xi_t = c_1 + a\xi_{t-1} + \varepsilon_t, \quad t = 1, \dots, \mathrm{m}.$$

The critical set contains small values of statistics: H_0 is rejected if the value of (15) is smaller than the appropriate quantile from Table 1. The graph of estimated power function for sample size $m = 10$ and $m = 50$ is presented on Fig. 3. The power function of this test obviously depends on the sample size m. For small $m \leq 30$ it is comparable with power function of Dickey-Fuller (DF) test (see Fig. 4). For larger samples, $m > 30$, DF test is uniformly stronger than (15), but the difference of power is not very significantly large.

It is worth to notice that the random variables (14) may also be applied to verify higher-order integration and cointegration (see Engle, 1987, and Hamilton, 1994) of time series. In this case the differences $\Delta\xi$, $\Delta\eta$ in formula 10) should be replaced by d-order ones $\Delta^d \xi$, $\Delta^d \eta$.

Figure 3. The power function M of (15). Probability level $\alpha = 0.05$. Sample size $m = 10$ (A), and $m = 50$ (B).

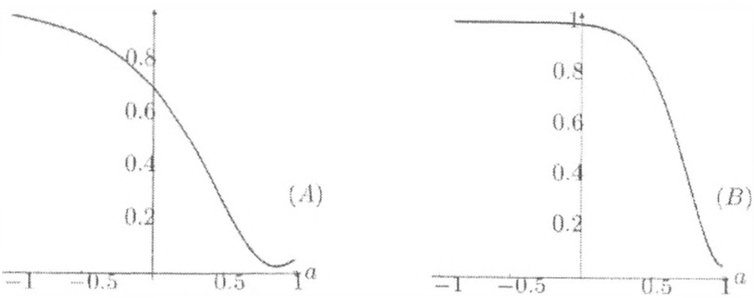
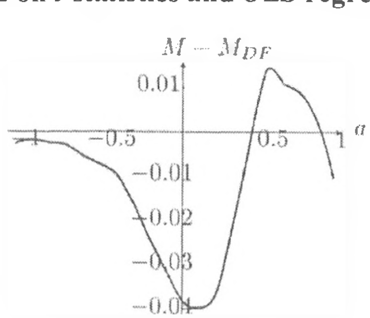


Figure 4. Estimated difference $M - M_{DF}$ of power functions M, M_{DF} of (15) and Dickey-Fuller test based on t-statistics and OLS regression. Sample size $m = 20$.



4. Similarity of general indicators of business activity based on qualitative and quantitative data

In this section we evaluate the similarity of two sets of indicators of general business activity for Poland. The first set consists of the indicators $ZHG1$, ... , $ZHG4$ and $ZGG1$, ... , $ZGG4$, based on qualitative data. These indicators are a weighted average of indices reflecting business tendency (sales, production etc.) in major sectors of economy.[7] The second set consists of statistical indicators of the aggregate output, including GDP, industrial production index IP, and GCI (general coincident indicator) – a weighted average of indices reflecting output or sales in major sectors of economy weighted by their yearly shares in GDP.[8] The latter are considered to be reference indexes to check the adequacy and forecasting properties of the qualitative indicators.

The analysis was made on time series calculated by Z. Matkowski for the period 1994-2001 at monthly intervals.

The values of similarity between time series (and some of their components) are shown in Table 2. The results of an extended analysis, including leads or lags, are shown in Table 3.

Table 2. Average values of similarity between the selected indexes

Index	*GDPI*	*IPI*	*GDPI_mcd*	*GDPI_tc*	*IPI_mcd*	*IPI_tc*
*ZHG*1	0.22	0.07	**0.39**	**0.36**	**0.30**	**0.31**
*ZHG*2	0.23	0.09	0.36	0.33	0.20	0.28
*ZHG*3	0.23	0.11	0.36	0.33	0.23	0.28
*ZHG*4	**0.23**	**0.11**	0.32	0.31	0.27	0.23
*ZGG*1	0.14	0.12	0.28	0.26	0.14	0.13
*ZGG*2	0.14	0.10	0.27	0.24	0.18	0.18
*ZGG*3	0.15	0.12	0.28	0.27	0.24	0.19
*ZGG*4	**0.28**	**0.13**	**0.34**	**0.33**	**0.31**	**0.25**
*ZHG*1_*mcd*	0.26	0.17	0.53		0.35	
*ZHG*1_*tc*	0.32	0.15		0.58		0.48
*ZHG*2_*mcd*	0.27	0.17	0.57		0.42	
*ZHG*2_*tc*	0.40	0.20		0.64		0.55
*ZHG*3_*mcd*	0.32	0.20	0.57		0.40	
*ZHG*3_*tc*	0.36	0.18		0.60		0.54
*ZHG*4_*mcd*	**0.37**	0.17	**0.59**		**0.52**	
*ZHG*4_*tc*	**0.42**	0.19		**0.61**		**0.64**
*ZGG*1_*mcd*	0.30	0.13	0.45		0.28	
*ZGG*1_*tc*	0.42	0.17		0.56		0.53
*ZGG*2_*mcd*	0.30	0.14	0.43		0.30	

[7] These indicators, developed by Z. Matkowski, differ in formulas and weights. Indexes *ZHG*1, ... , *ZHG*4 base on data from the Research Institute of Economic Development, Warsaw School of Economics, while indexes *ZGG*1, ... , *ZGG*4 use survey data from the Central Statistical Office. (See Matkowski, 2000, 2002).

[8] For simplicity, the GCI reference index has been dropped in our computations.

*ZGG*2_*tc*	0.37	0.15		0.47		0.49
*ZGG*3_*mcd*	0.43	0.18	0.54		0.46	
*ZGG*3_*tc*	0.44	0.16		0.52		0.48
*ZGG*4_*mcd*	**0.42**	**0.21**	**0.58**		**0.50**	
*ZGG*4_*tc*	**0.48**	**0.21**		**0.64**		**0.63**

mcd – seasonally adjusted index, smoothed by MCD-average, *tc* – trend + cycle (9 or 13 month moving average), suffix *I* denotes constant base index (1995 = 100).

Looking at these tables, we can make the following observations:

- The dynamics of *ZHG*1, ..., *ZGG*4 is significantly similar to the reference index, especially GDP. The values of similarity measure increasingly rise if the time series are smoothed. These indexes seem to be good to asses the current economic trend.
- The best fit to the actual development of economy (actual changes of reference index) is given by the indexes *ZHG*1, *ZHG*4 and *ZGG*1, *ZGG*4. Similarity analysis does not allow to choose one of them as the best indicator. Perhaps it can be done together with statistical and other methods (see Matkowski, 2002).
- The most similar dynamics against the reference indexes, either synchronically or with some lead (≤ 12 months), is shown by *ZHG*1, *ZHG*3, *ZHG*4, *ZGG*3 and *ZGG*4. These indexes seem to be good leading indicators. They also can be the basis to find another formula for good leading indicators.

Table 3. Maximum similarity between reference indexes (GDP or IP) and the current and delayed survey based indicators

Index	*GDPI*	*GDPG*	*GDPR*	*IPI*	*IPG*	*IPR*
*ZHG*1	0.33 (-10)	0.33 (-11)	0.25 (0)	0.21 (-10)	0.12 (-12)	0.12 (-6)
*ZHG*2	0.43 (-10)	0.27 (-1)	0.18 (0)	0.23 (-10)	0.16 (-7)	0.14 (-10)
*ZHG*3	**0.49 (-10)**	**0.40 (-12)**	0.20 (0)	0.26 (-10)	0.22 (-7)	0.16 (-1)
*ZHG*4	0.41 (-10)	0.28 (-1)	0.24 (-10)	**0.25 (-10)**	0.20 (-3)	0.15 (-11)
*ZGG*1	0.43 (-9)	0.29 (-6)	0.32 (-10)	0.21 (-5)	0.23 (-3)	0.18 (-12)
*ZGG*2	0.44 (-9)	0.33 (-6)	0.28 (-10)	0.23 (-9)	0.25 (-7)	0.17 (-10)
*ZGG*3	**0.48 (-8)**	**0.40 (-7)**	**0.35 (-11)**	0.25 (-5)	0.23 (-6)	0.17 (0)
*ZGG*4	**0.47 (-10)**	0.37 (-1)	0.31 (-12)	**0.29 (-10)**	**0.26 (-6)**	**0.22 (-11)**

The suffix *I* means constant base index (1995 = 100), *G* – growth rate against the same month in the preceding year, *R* – growth rate against the preceding month. The element in *i*-th row and *j*-th column is $\max\{\mu(B^{\tau}\, ind_i,\, ref_j) : \tau = 0, \dots, 12\}$. The number of leads is given in the parenthesis.

The most adequate formula of the general indicator of business activity can be obtained by comparing similarity between the actual reference index (GDP, IP, or another one) and actual or delayed values of the survey-based indicators. The similarity analysis provides a natural criterion for a good indicator: its dynamics should be similar to the dynamics of reference index. Moreover, its changes should lead the changes of reference index. Formally, the optimal indicator *ind** is a solution of the following optimisation problem:

$$\text{Maximise } \mu(ref, B^{\tau} ind), \tag{16}$$

subject to: *ind* belongs to the set of admissible indicators and $0 \le \tau \le \tau^*$, where *ref* denotes the reference index, B is the lag-operator, τ^* is an arbitrary chosen maximum number of lags (time difference) between *ind* and *ref*. It is worth to notice that this methodology does not require stationarity of time series.

It follows from Table 3 that the indexes $ZHG1,\dots, ZHG4$, $ZGG1,\dots, ZGG4$ (especially $ZHG3$, $ZHG4$ and $ZGG3$, $ZGG4$) have similar dynamics to GDP and IP.

But, slightly better results could be reached by taking an average of different indicator variants. Solving the problem (16) with $\tau^* = 12$, one can obtain the following results:

- The indicator

$$\begin{aligned} ind1 &= 0.07\,ZHG3 + 0.16\,ZHG4 + 0.11\,ZGG1 + 0.14\,ZGG2 \\ &\quad + 0.20\,ZGG3 + 0.31\,ZGG4, \\ &\quad \mu(B^{10} ind1, GDP) = 0.32 \end{aligned} \tag{17}$$

  is the most similar to GDP among all linear combinations of $ZHG1 - ZGG4$. The maximum of similarity is attained with $\tau = 10$. This means that (17) is a quite good leading indicator.

- The indicator

$$\begin{aligned} Ind2 &= 0.21\,ZHG1 + 0.25\,ZHG2 + 0.27\,ZHG3 + 0.26\,ZHG4, \\ &\quad \mu(B^{11} ind2, GDP) = 0.31 \end{aligned} \tag{18}$$

  is the most similar to GDP among all linear combinations of indexes $ZHG1-ZHG4$. The maximum of similarity is attained with $\tau = 11$.

- The indicator

$$\begin{aligned} Ind3 &= 0.34\,ZHG2 + 0.50\,ZHG3 + 0.16\,ZHG4, \\ &\quad \mu(B^{10} ind3, IP) = 0.17 \end{aligned} \tag{19}$$

  is the most similar to IP among all linear combinations of indexes $ZHG1-ZHG4$. The maximum of similarity is attained with $\tau = 10$.

5. Conclusion

Similarity measures may be useful tools in analysing data, particularly in comparing dynamic structure and building econometric models.

The definition of similarity measures can be extended on the space of stochastic processes on given probabilistic space. Depending upon the definition of the similarity measure, several properties of stochastic processes may be investigated and a number of hypothesis concerning data may be verified.

The definition of the similarity measure seems to be quite flexible, so it may lead to a large number of tests of probabilistic properties of data. It seems that in general the values of the power function of the test (15) are not very less than the standard Dickey-Fuller test. It may be very interesting to compare the power functions of these tests. It is very important to construct a number of sufficiently strong independent tests to examine the stochastic structure of data. Perhaps it is possible to find a statistic based on the similarity measure μ, with a power function greater than the mentioned tests of stationarity.

The similarity measures seem also to be a useful tool to analyse composite indicators of economic activity. The analysis of similarity, together with other methods, can be applied to choose the most appropriate formulas for leading indicators to assess current economic situation and to predict GDP growth rates.

6. References

Dhrymes P.J. 1971. *Distributed Lags. Problems of Estimation and Formulation.* San Francisco: Holden-Day Inc.

Dickey D.A.., W.A. Fuller. 1979. “Distribution of the Estimators for Autoregressive Time Series with a Unit Root.” *Journal of American Statistical Association* 74.

———. 1981. “Likelihood Ratio Statistics for Autoregressive Time Series with Unit Root.” *Econometrica* 49.

Dorosiewicz S., T. Michalski. 1998. “Podobieństwo funkcji w badaniach ekonomicznych (metody i przykłady)” [Similarity of functions in economic research, methods and examples]. *Przegląd Statystyczny* 2.

Engle R.F., C.W.J Granger. 1987. “Co-integration and Error Correction: Representation, Estimation and Testing.” *Econometrica* 55.

Gihman I.I., A.V. Skorochod. 1979. *The Theory of Stochastic Processes.* Springer Verlag.

Hamilton J.D. 1994. *Time Series Analysis.* Princeton, N.J.: Princeton University Press.

Judge G.G., W.E. Griffiths, H.R. Carter., H. Lütkepohl, T. Chao Lee. 1985. *The Theory and Practice of Econometrics.* John Wiley and Sons.

Karatzas I., S.E. Schreve. 1991. *Brownian Motion and Stochastic Calculus.* Springer Verlag.

Matkowski Z. 2000. "The Use of Survey Data in Economic Barometers for Poland." In: K.H. Oppenländer, G. Poser, B. Schips (eds.), *The Use of Survey Data for Industry, Research and Economic Policy,* Aldershot – Burlington – Singapore – Sydney: Ashgate.

——. 2002. "General Indicators of Business Activity for Poland Based on Survey Data." In: G. Poser, D. Bloesch (eds.), *Economic Surveys and Data Analysis,* Paris: OECD.

Phillips P.C.B. 1987. "Time Series Regressions with a Unit Root." *Econometrica* 55.

Phillips P.C.B., P. Perron 1988. "Testing for a Unit Root in Time Series Regression." *Biometrica* 75.