



E pluribus, quaedam. Gross Domestic Product out of a Dashboard of Indicators

Mattia Guerini^{1,2,3,4} · Fabio Vanni^{3,4,5} · Mauro Napoletano^{3,4,6}

Received: 17 April 2023 / Accepted: 12 February 2024
© The Author(s) 2024

Abstract

Is aggregate income enough to summarize well-being? We address this long-standing question by exploiting a quantitative approach that studies the relationship between gross domestic product (GDP) and a set of economic, social and environmental indicators for nine developed economies. We introduce a mathematical approach to the analysis of economic indicators. By employing dimensionality reduction and time series reconstruction techniques, we quantify the share of variability stemming from a large set of different indicators that can be compressed into a univariate index. We also evaluate how well this variability can be explained if the univariate index is assumed to be respectively the gross domestic product, national income, household income, or household spending. Our results indicate that all the four univariate measures are doomed to fail in accounting for the variability of all the domains. Even if GDP emerges as the best option among the four economic variables, its quality in synthesizing the variability of indicators belonging to other domains is poor (about 35%). Our approach provides additional support for policy makers interested in measuring the trade offs between income and other relevant social, health and ecological dimensions. Finally, our work adds new quantitative evidence to the vast literature criticizing the usage of GDP as a measure of well-being.

✉ Mattia Guerini
mattia.guerini@unibs.it

Fabio Vanni
fabio.vanni@uninsubria.it

Mauro Napoletano
mauro.napoletano@univ-cotedazur.fr

- 1 University of Brescia, Brescia, Italy
- 2 Fondazione ENI Enrico Mattei, Milano, Italy
- 3 Université Côte d'Azur, CNRS, GREDEG, Nice-Sophia Antipolis, France
- 4 Institute of Economics, Sant'Anna School of Advanced Studies, Pisa, Italy
- 5 Università degli Studi dell'Insubria, Varese, Italy
- 6 SciencesPo, OFCE, Paris, France

Keywords Gross domestic product · Well-being indicators · Data reduction techniques · Principal component analysis · Random matrix

JEL Classification C43 · I30 · I31

1 Introduction

In this paper we quantify the ability of Gross Domestic Product (GDP henceforth) and other aggregate income and expenditure variables to summarise well-being, defined by a broad set of economic, social, environmental sustainability and demographic indicators.¹ One of the main characteristics that allowed GDP to gain a central role among all the possible economic measures of well-being, is its quantitative, monetary and synthetic nature. Although other variables share these three desirable properties – e.g., Gross National Income (GNI), household income, and household spending – no other variable plays as predominant a role in policy decisions as the GDP. Together with the system of national accounting (SNA), GDP can be calculated across different countries on the same scale, thus allowing comparisons among them (Hoekstra 2019). But the monetary nature of GDP is also its main limitation. It is indeed difficult to assign a monetary value to all those human activities affecting living standards and well-being but lacking a regular market. And even where markets do exist, they might be highly imperfect, with prices that do not incorporate all the possible effects that the production and consumption of such goods and services generate – this is especially valid for activities generating significant externalities. For all these reasons a number of scholars and policy making institutions have challenged the idea of relying uniquely upon GDP as a tool to quantify well-being and to evaluate policies (Michalos 1982; Stiglitz et al. 2009). It is worth noticing, that also the system of national accounting explicitly indicates that “[...] GDP is often taken as a measure of human development index, but the SNA makes no claim that this is so and indeed there are several conventions in the SNA that argue against the welfare interpretation of the accounts” (United Nations 2009). For all these reasons, in the past decades a literature has developed around the creation of new well-being indicators. The reference to well-being (also when evaluating policies based solely upon GDP) allows policymakers to account for a “fully rounded humanity” rather than only for the mere production of goods and services. On the one side, the shift in political and policy focus towards the positive aspects of a “rounded humanity” can be considered as a positive development of the economic discipline because it encourages the contemplation of of being well – rather than only doing well. On the other side, however, also some perils arise from an extreme focus on well-being. Because many broad measures are seldomly based on individual questionnaires asking the interviewed some form of psychological introspection (e.g., question on life satisfaction, happiness, ...), they might distort the objectives of the policy agenda from scopes that satisfy the society at large. In particular they might

¹ In this paper we refer to well-being rather than welfare. This choice resides in the conception of welfare as a narrower measure of well-being. In particular, in economics, welfare relies upon the derivation of a utility function while well-being, instead, has a more lax definition and is considered to be somehow related to a broad set of variables and to the degree of satisfaction they provide to individuals (Maximo 1987).

focus on needs that tend to satisfy the private sphere of the individuals rather than the public one Taylor (2011).

As a consequence, *dashboards of indicators* have been created to better evaluate well-being in a society, by taking into account dimensions such as environmental sustainability, social equity, cultural development, economic vulnerability as well as the demographic one (see OECD 2011; Pinar et al. 2014; Roser 2014; Fitoussi and Durand 2018a,b; Ferran et al. 2018; Bacchini et al. 2020; Kalimeris et al. 2020, among the others). All these works focus on the multifaceted and complex nature of well being and, at the normative level, do not imply the complete replacement of GDP with the new metrics. They rather suggest that GDP shall be accompanied by alternative statistics.

We contribute to this literature by quantitatively evaluating the ability of GDP to capture the information embedded in a large set of social, economic and ecological indicators, which are constitutive of well-being (they are included in the construction of the most important well-being indicators. See OECD 2011; UNDP 2022). We make use of a widely known dimensionality reduction technique, the generalized Principal Component Analysis (PCA). Moreover, we combine PCA with the Random Matrix Theory (RMT) approach, which allow us to single out only the components that are statistically significant (see e.g., Onatski 2010). More precisely, this procedure involves comparing the empirically estimated eigenvalues with the distribution of eigenvalues that is generated by a Gaussian random model with only spurious correlations.² Using the leading component or the GDP as univariate measures of well-being, we reconstruct two alternative synthetic series for all the indicators. By comparing the synthetic series with the original counterparts, we quantitatively measure the ability of GDP to summarize the variability of all the indicators. We apply this strategy to nine advanced OECD economies and our findings suggest that univariate measures, and GDP among them, are only imperfect proxies of well-being. With respect to the ability of GDP to approximate single indicators, substantial heterogeneity is found at the country level with the possibility of poor performance, especially over the demographic and social equity dimensions. Overall, our results, confirm that one shall rely upon multivariate composite indices of well-being, which are more apt at capturing the interactions between different indicators also pertaining to heterogeneous domains.

2 Methodology

We start from N time series of well-being indicators observed for T periods and all sampled at the same frequency Δt . We denote the matrix of time-series by $\tilde{X}(t)$ and their complex Hilbert transformation by $X(t)$.³

² This technique has also been recently adopted in the business cycle and financial economics literature. (see respectively Guerini et al. 2023; Barbieri et al. 2021). The main advantage of RMT is that it provides more precise and accurate information about a panel of time-series compared to basic PCA analysis, which does not allow one to distinguish between factors reflecting spurious correlations obtainable with a finite number of observations and those that instead contain relevant information about the similarity of the series.

³ The Hilbert transformation on the series is useful because it will allow us to also capture the correlation between similar time series displaying time shifts in their co-movements.

2.1 Dimensionality Reduction

According to the generalized PCA (Ng et al. 2001) the time series of the indicators can be expressed as:

$$\mathbf{X}(t) = \mathbf{A}(t)\mathbf{V} \quad (1)$$

where $\mathbf{A}(t)$ is a $T \times N$ loading matrix and \mathbf{V} is a $N \times N$ matrix of eigenvectors.⁴ These eigenvectors are associated to the N -dimensional vector of eigenvalues $\boldsymbol{\lambda}$, computed from the spectral decomposition:

$$\mathbf{C}\mathbf{V} = \boldsymbol{\lambda}\mathbf{V} \quad (2)$$

where the correlation matrix \mathbf{C} can be estimated by $\widehat{\mathbf{C}} = \frac{1}{N}\mathbf{X}(t)\mathbf{X}(t)'$.⁵ The correlation matrix $\widehat{\mathbf{C}}$ is positive semi-definite and bears N non-negative and distinct eigenvalues $\boldsymbol{\lambda}$ with their associated eigenvectors \mathbf{V} . According to Principal Component Analysis (PCA, Vidal et al. 2016), each eigenvalue can be expressed as a linear combination of the original series and corresponds to a principal component, also explaining a portion of the total variance of the data proportional to its magnitude. Thus, the empirical density function of the eigenvalues can be expressed as:

$$\widehat{\rho}(\boldsymbol{\lambda}) = \frac{dn(\boldsymbol{\lambda})}{d\boldsymbol{\lambda}} \quad (3)$$

where $n(\boldsymbol{\lambda})$ indicates the number of eigenvalues larger than $\boldsymbol{\lambda}$.

To focus solely on principal components which are statistically significant one can compare the empirical density function $\widehat{\rho}(\boldsymbol{\lambda})$ with a theoretical benchmark distribution of eigenvalues that would have been generated under a known null-hypothesis. The random matrix theory (RMT) provides a well-specified theoretical null-hypothesis for such a statistical significance test (Onatski 2010). But for the RMT to hold, it is required that no autocorrelation exists in the series and that the series are infinitely dimensional – in the sense that both $N, T \rightarrow \infty$, with $Q = \frac{T}{N}$ finite. To be free from these two tight restrictions, one can alternatively rely upon the less demanding rotational random shuffling (RRS) simulations which, in the limit, converge to the same theoretical distribution of RMT (Iyetomi et al. 2011; Aoyama et al. 2017; Kichikawa et al. 2020), represented by the Marchenko-Pastur distribution:

$$\rho(\boldsymbol{\lambda}) = \begin{cases} \frac{1}{2\pi\sigma^2} \frac{\sqrt{(\lambda_M - \boldsymbol{\lambda})(\boldsymbol{\lambda} - \lambda_m)}}{\boldsymbol{\lambda}} & \text{if } \lambda_m \leq \boldsymbol{\lambda} \leq \lambda_M \\ 0 & \text{else} \end{cases} \quad (4)$$

where $\lambda_m = \sigma^2 \frac{(1 - \sqrt{Q})^2}{Q}$ and $\lambda_M = \sigma^2 \frac{(1 + \sqrt{Q})^2}{Q}$ represent the lower and upper bounds. Deviations between the empirical distribution $\widehat{\rho}(\boldsymbol{\lambda})$ and the theoretical one $\rho(\boldsymbol{\lambda})$, indicate the presence of some statistically significant components which can summarize the co-movements between the empirical indicators. This is exactly the reason why

⁴ The t symbol in parenthesis is made explicit for matrices representing time series.

⁵ Without loss of generality we assume all series to be stationary and standardized. This implies $\tilde{\mathbf{X}}(t) \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$.

PCA is considered a dimensionality reduction technique. In particular, the number of significant components is equivalent to the number of eigenvalues which exceed the theoretical (or simulated with the RRS) upper bound $\lambda_{.M}$ (Laloux et al. 2000).

2.2 Construction of Synthetic Indicators

Once the significant principal components have been selected according to the above-described procedure, one can construct a synthetic indicator of the original time series (“synthetic PC” henceforth) as follows:

$$\widehat{X}_J(t) = A_J(t)V_J \quad (5)$$

where the index J is an integer indicating the number of significant eigenvalues, $\widehat{X}_J(t)$ is a $T \times N$ matrix with the synthetic series, as generated using the J leading principal components (thus the index J), V_J is a $J \times N$ matrix of the estimated complex eigenvectors associated to the J significant eigenvalues, and $A_J(t)$ is a $T \times J$ matrix with the associated loadings. The synthetic series will be different from the original ones, and they represent the indicators that one would observe if the noisy component of each of the original indicators would be ignored.⁶

Furthermore, one can evaluate the quality of GDP at summarizing information about well-being (as provided by the large set of the N original series) by assuming that the GDP is the leading component that summarizes well-being. Formally, this corresponds to assuming that GDP replaces the component loadings $A_J(t)$, under the condition $J = 1$. With this assumption, one can obtain an alternative synthetic indicator (“synthetic GDP” henceforth) of the original time series as follows:

$$\widehat{X}_G(t) = \alpha A_G(t)V_1 \quad (6)$$

where $\widehat{X}_G(t)$ is the $T \times N$ matrix with the GDP-based synthetic series (thus the index G), $A_G(t)$ is a $T \times 1$ matrix with the Hilbert transform of the GDP series, V_1 is the $1 \times N$ eigenvector associated to the largest eigenvalue λ_1 and α is a rescaling factor measuring the deviation scalar from the dominant eigenvector.⁷

Once the synthetic indicators $\widehat{X}_J(t)$ and $\widehat{X}_G(t)$ are available, one can evaluate the quality of the matching with the original indicators $X(t)$. For that, we use the root mean squared error (RMSE) and the Hilbert correlation coefficient between the complexified series.⁸ Other more complicated alternatives are possible, for example the dynamic time-warp (DTW). However, it is not the aim of this paper to evaluate the quality of different similarity measures.

⁶ The noisy components are here intended to be all the non-significant ones according to the procedure developed in Section 2.1.

⁷ According to the small perturbation theory $\alpha = \frac{A_1(t)A_G(t)^*}{A_G(t)A_G(t)^*}$ and $A_G(t)^*$ is the complex conjugate of the GDP series (Stewart and Sun 1990; Ng et al. 2001).

⁸ We also report the Spearman and the simple Pearson correlation coefficients. However, since they are computed over the real part of the Hilbert transformed numbers, they cannot account for the dynamic cross-correlations of the variables and they are lower than the Hilbert correlation by construction.

3 Empirical Application

We employ 39 different time-series indicators capturing economic, environmental, social equity and demographic dimensions.⁹ Some of the indicators available in our analysis are also the loading components of the OECD better life index OECD (2011) and of the UNDP (2022) report. But one of the main limitations of the Better Life Index (recognized also by the authors) is that it is a cross-sectional index, incomparable over time. Therefore, in our analysis we only include those variables that can be consistently observed over a sufficient amount of periods (e.g., life expectancy at birth, infant mortality rates, ...), together with other variables excluded from the above mentioned well-being metrics, but that possibly relate to production of goods and services (e.g., energy consumption, housing prices, gender pay gap, ...). This allows us to compute the correlation of the principal component with the main univariate indexes often taken as indicators of national development and well-being also accounting for time delays.

All the indicators are sampled at annual frequency and cover the 1995–2015 period forming a balanced panel dataset for each country. Over the cross-section dimension, our analysis is performed on nine different advanced OECD economies.¹⁰ The panel dataset format and the variable measurement, together with the normalization, ensures a perfect comparability between the sampled countries.

After having transformed the indicators into stationary series by means of the first difference transformation, we apply the generalized PCA and the RMT procedures to test for the statistical significance of the estimated principal components. For most countries (all but Great Britain) we find that only the largest eigenvalue exceeds the RMT upper bound. This implies that one can significantly summarize a certain fraction of the data variance by means of a single variable. In particular, the fraction of variance captured by the first principal component is called the *absorption rate* and is reported in Table 1 (second column for all the four sub-tables). The leading principal component explains between 30% and 43% of the total variance provided by the original 39 series for the nine considered economies. Spain (ESP) displays the highest absorption rate (42.58%) whereas Germany (DEU) the lowest one (30.25%). In addition, Great Britain is the unique country for which the leading component is not statistically significant.¹¹

Overall, measuring well-being only by means of the best linear univariate predictor, represented by the leading component, one would lose about 57% to 70% of the overall information embedded in the series of the original indicators. This is a relevant amount of the total variation and points toward the failure of any univariate measure to account for the complex and multifaceted relationships between all the different indicators pertaining to different domains.

The last three columns in each sub-table (see Table 1) report the correlation between the leading principal component estimated over the 39 indicators reported in Appendix A and a possible synthetic univariate indicator, taken among the most

⁹ The complete list is presented in Appendix A and all series are publicly available at the OECD i-Library.

¹⁰ The nine countries and their short labels are Germany (DEU), Spain (ESP), France (FRA), Great Britain (GBR), Italy (ITA), Japan (JPN), Netherlands (NLD), Sweden (SWE) and United States (USA).

¹¹ The latter result points to the impossibility to summarize the information stemming from the indicators in a low dimensional space and warns against the usage of a univariate measure of well-being for this country.

Table 1 Fraction of variance explained by the dominant eigenvalue of the generalized PCA (absorption rate) and correlation between the four univariate indicators (reported in the table titles) with the principal component

Country	PCA Absorption	Correlation		
		Hilbert	Spearman	Pearson
Gross Domestic Product				
DEU	30.25%	87.62%	48.18%	79.64%
ESP	42.58%	88.85%	55.19%	78.80%
FRA	31.77%	90.33%	15.58%	25.58%
GBR*	30.76%	74.80%	64.29%	71.52%
ITA	38.40%	95.90%	50.78%	63.16%
JPN	33.91%	81.91%	28.57%	31.10%
NLD	29.99%	88.97%	84.94%	86.54%
SWE	30.65%	92.63%	68.31%	73.13%
USA	39.70%	91.22%	81.17%	89.35%
Gross National Income				
DEU	30.25%	84.44%	51.95%	68.15%
ESP	42.58%	77.76%	71.04%	77.47%
FRA	31.77%	73.89%	17.01%	6.94%
GBR	30.76%	64.06%	33.25%	58.82%
ITA	38.4%	82.19%	2.34%	10.04%
JPN	33.91%	84.92%	32.21%	27.21%
NLD	29.99%	65.82%	66.49%	65.50%
SWE	30.65%	82.63%	19.22%	37.31%
USA	39.7%	88.40%	68.96%	85.80%
Household Income				
DEU	30.25%	54.57%	18.70%	42.3%
ESP	42.58%	67.79%	60.00%	67.76%
FRA	31.77%	48.75%	24.81%	10.45%
GBR	30.76%	50.30%	35.58%	42.56%
ITA	38.40%	76.97%	3.90%	16.58%

important economic variables also taken as “targets” by policymakers.¹² An interesting result that we obtain from the comparison of these correlations between different indicators, is that for almost all countries the GDP is the one that correlates the most with the leading principal component. Also Figure 1 presents similar information by visually comparing the evolution of GDP, GNI, household income and expenditure in the US, with the evolution of the leading principal component (leading PC henceforth)

¹² There are three different columns because we report the correlation as computed by (i) the Hilbert transform correlation, (ii) the Spearman rank correlation, and (iii) the simple Pearson correlation. The Hilbert has to be preferred because of its ability to account also for time delays. Furthermore, except for few outliers (especially with national income, household income and household spending in Italy) the ranks of the three correlation coefficients are coherent.

Table 1 continued

Country	PCA Absorption	Correlation		
		Hilbert	Spearman	Pearson
JPN	33.91%	56.06%	50.65%	37.44%
NLD	29.99%	75.11%	69.87%	66.09%
SWE	30.65%	68.72%	26.62%	31.57%
USA	39.70%	65.02%	34.03%	61.52%
Household Expenditure				
DEU	30.25%	73.10%	32.34%	44.94%
ESP	42.58%	80.63%	83.90%	79.82%
FRA	31.77%	64.58%	33.64%	20.83%
GBR	30.76%	39.47%	4.55%	38.9%
ITA	38.40%	79.08%	0.65%	6.39%
JPN	33.91%	60.3%	40.65%	33.44%
NLD	29.99%	79.10%	71.95%	78.74%
SWE	30.65%	51.61%	6.10%	9.71%
USA	39.70%	88.05%	66.88%	87.79%

for the same economy. The correlation is outstandingly high for all indicators in the USA, with GDP recording the highest value (91%). Overall this suggests that, shall one be confined to the usage of a single indicator to capture well being, the idea of employing GDP might not be completely far fetched, as a first order approximation. In fact, it is fair to affirm that GDP growth mimics the dynamic of the best univariate linear approximation, specifically designed to maximize the absorption rate. However, as already mentioned, even the leading PC accounts only for about one third of the total variance and therefore, whenever possible, one should untie him/her-self from the constraint of the adoption of a single univariate indicator as a measure of well-being. We will also see in what follows that the fraction of variance not captured by these univariate indicators is particularly relevant for some domains (e.g., social justice and demography) making therefore any univariate indicator inappropriate for summarizing well-being.

The reminder of the paper aims at evaluating the ability of GDP and the leading principal component to reconstruct the original indicators. In what follows we focus our analysis on the USA and on GDP. Extensive results for the other eight countries and for the other three variables (i.e., GNI, household income and expenditure) are presented in the on-line supplementary material.¹³

To further quantify the loss of information when using a univariate measure of well-being we proceed with the construction of the synthetic indicators as described in Equations (5) and (6) (respectively “Synthetic PC” and “Synthetic GDP”) for all the 39 original variables of our sample. Results about four of these indicators, pertaining to different domains (economic, environmental, social and demographic) are presented

¹³ The on-line supplementary material is available at the authors’ github pages.

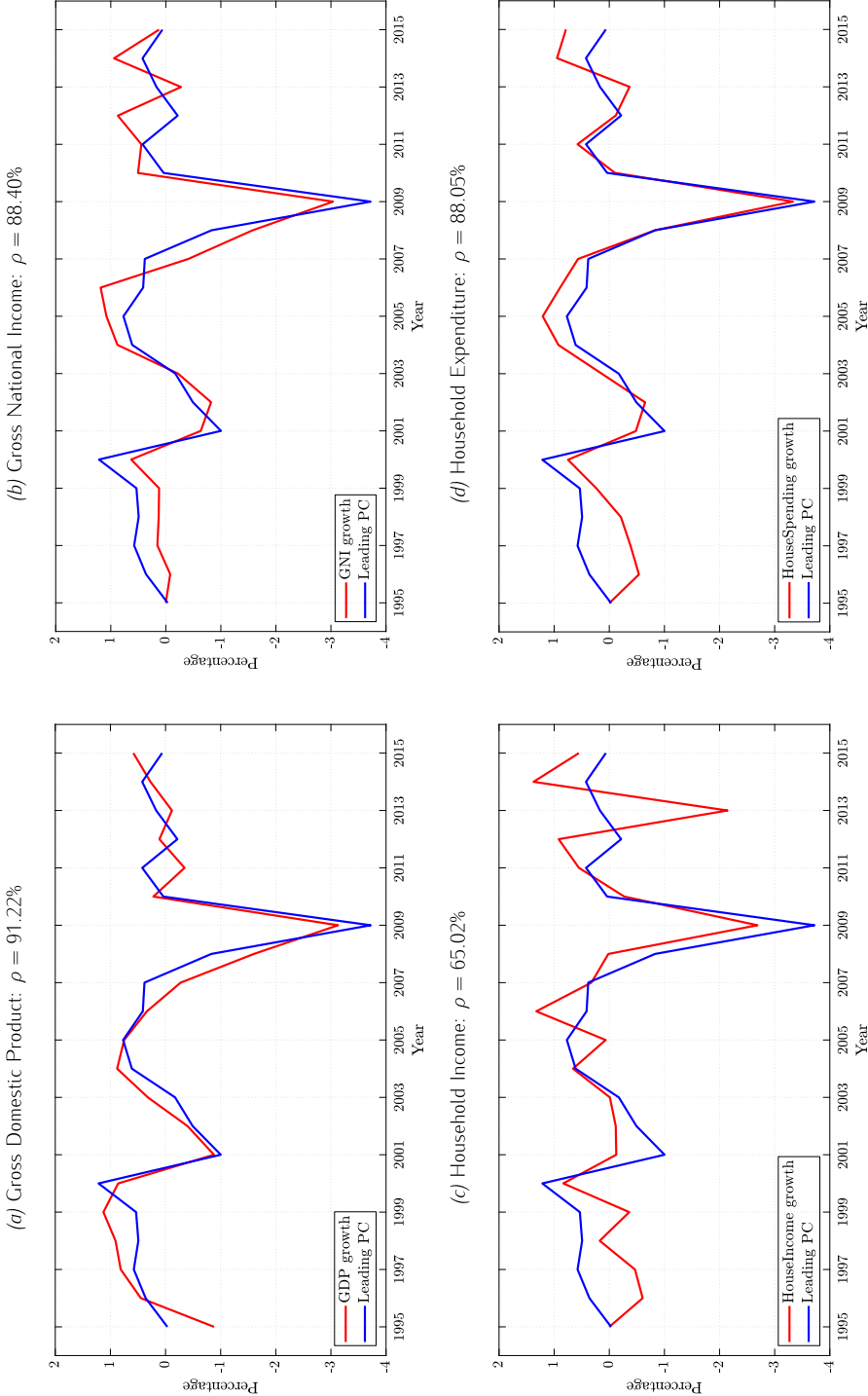


Fig. 1 Time series for the leading principal component in the United States (blue) together with the time series of the four different univariate indicators (red) reported in each subtitle (colour figure online)

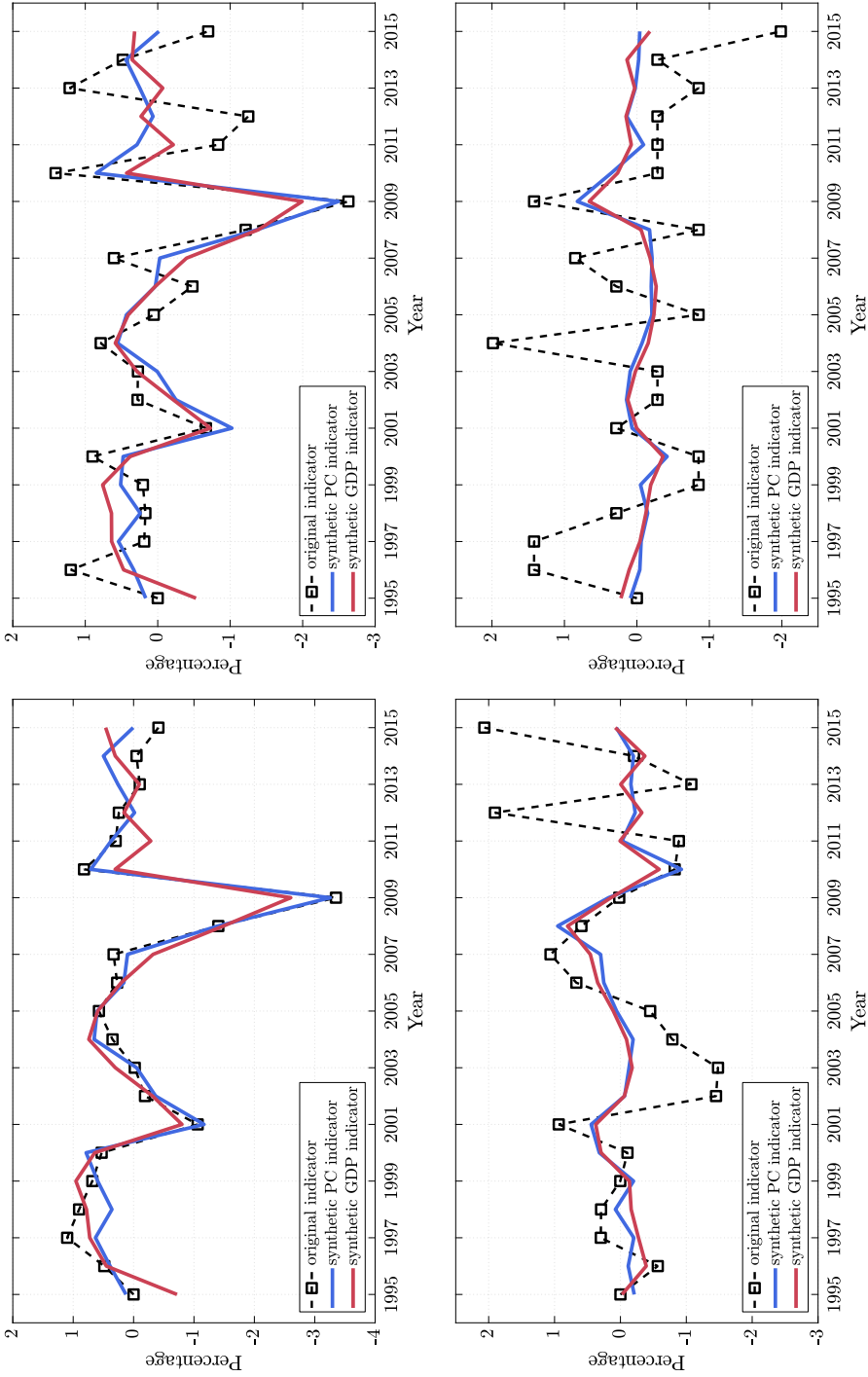


Fig. 2 Selected original indicators (dashed black lines with squares) and synthetic indicators constructed starting from the leading PC (blue lines) and the GDP (red lines) for the USA

Table 2 Comparison of original and synthetic indicators. For each indicator constructed with the leading PC or with the GDP growth, we report the Root Mean Squared Error in percentage terms with respect to the original indicator, the correlation between the series and the p-value of the correlation coefficient

Indicator	RMSE (in %)		Correlation		Correlation p-value	
	PC	GDP	PC	GDP	PC	GDP
Industrial Production	6.16	9.49	0.95926	0.90082	< 0.001	< 0.001
Greenhouse Gas Emissions	13.49	15.95	0.78587	0.68303	< 0.001	< 0.001
Gender Wage Gap	20.29	20.63	0.36802	0.33877	0.10071	0.14917
Life Expectancy	21.1	20.92	0.25489	0.29519	0.26481	0.2046

in Fig. 2. In addition, in Table 2 we also report the Percentage Root Mean Squared Error (RMSE%) for each combination of original and synthetic indicators, the correlation coefficient between them, and the correspondent p-values. Figure 2 reveals that the two synthetic indicators (the red and the blue lines) display similar patterns in all the domains considered. This is coherent with the high degree of correlation previously reported in Table 1. However, their performance in tracking representative varies wildly across the domains. For instance, the original series in the economic domain, as represented by the percentage change of industrial production in the manufacturing sector, is replicated with high precision (an error of about 6 to 9 percentage points). A similar qualitative result (but quantitatively worsened, cf. Table 2) is observed in the pollution domain, represented by the percent change of greenhouse gas emissions.¹⁴ In contrast, the synthetic reconstructed indicators are not capable of tracking the patterns for the variables selected as representative of social equity (the percent variation in the gender wage gap, with a RMSE of about 20.5%) and demography (measured by the percent variation in life expectancy). This is straightforwardly visible from the two panels in the bottom of Fig. 2. This can be explained by the fact that variables pertaining to these two domains are only *slow moving* and therefore they contribute less to the total variance of the dataset. Accordingly, the PCA assigns them lower weights in the loading matrix $A(t)$. As a consequence, neither GDP nor the leading PC are effective in their prediction. As a consequence, also the correlation coefficients between the original series and the synthetic ones are low and not significant (see Table 2) indicating the poor ability of the univariate indicators to capture variables in all the domains at the same time. Once again, these results point toward the failure of GDP and any univariate indicator to precisely map the well-being of a country.

Finally, Table 3 reports the percentage RMSE of the same four indicators for all countries in our sample. We find similar heterogeneity across indicators, with substantially larger RMSE for the social equity and demographic domains. Overall these results confirm that univariate measures of well-being are doomed to fail in their attempt of summarizing the information stemming from a large set of different indicators and pertaining to heterogeneous domains.

¹⁴ A similar performance holds true also for the CO_2 emissions indicator. This is also coherent with the empirical literature reporting a cointegration relationship (possibly time-varying) between GDP and emission of global pollutants (Mikayilov et al. 2018).

Table 3 Cross country comparison of original and synthetic indicators. We report the Root Mean Squared Error in percentage terms with respect to the original indicator

Country	PC RMSE (in %)				GDP RMSE (in %)			
	Ind.	Prod.	Greenh.	Gas	Gender	Wage	Gap	Life Exp.
DEU	8.45	17.77	21.62	21.30	8.17	20.45	21.73	21.08
ESP	10.29	10.78	21.82	21.20	11.08	14.51	21.79	21.76
FRA	7.83	20.05	20.85	20.51	7.34	20.70	21.64	21.03
GBR	19.20	12.24	13.34	20.06	16.71	21.15	21.98	22.64
ITA	12.09	10.50	14.22	21.55	11.20	12.88	15.20	21.34
JPN	14.81	13.99	19.20	21.18	11.12	15.47	22.86	22.42
NLD	16.25	20.57	21.23	21.68	14.09	20.83	21.82	21.97
SWE	8.23	19.69	20.66	21.81	7.31	18.83	21.28	21.82
USA	6.16	13.48	20.29	21.10	9.46	15.95	20.63	20.92

4 Conclusions

We can draw two main conclusions from our work. First, our results suggest that if one has the aim of measuring well-being, he/she cannot simply rely upon univariate measures. This is because even the best linear estimator, aimed at maximizing the explained variance of a large set of variables (i.e., the PCA), is only able to capture a small portion of the whole indicators' variance (about 30% to 42%). Overall, this suggests that univariate measures of well-being are doomed to fail and one shall rely also upon multivariate composite indices of well-being (Bacchini et al. 2020; Kalimeris et al. 2020) and sustainability (Pinar et al. 2014; Luzzati and Gucciardi 2015). These measures are more apt at capturing the complex interactions between different indicators also pertaining to very heterogeneous domains. Second, we also find that among the univariate alternatives aimed at summarizing a multitude of dimensions related to well-being, GDP fares quite well with respect to other alternatives. In particular, it delivers a performance which is similar to the one provided by the leading principal component of the series (i.e. the best linear indicator). Furthermore it has a lower prediction error vis-à-vis other univariate variables, such as the gross national income, the household income or the household expenditure. This militates in support of the predominant use of GDP as the leading measure of economic performance of a country Malay (2019).

This work could be extended in several ways. First, one might enlarge the number of series in the sample, especially with respect to the social, equity and environmental dimensions. This might lead to different estimates for the loadings $A(t)$, assigning different weights to the single indicators when constructing the synthetic indicators based on principal component analysis. The effect of the inclusion of new variables is *a priori* unclear. Clearly, the variance explained by the first factor might increase or decrease, depending upon the degree of correlation between the new indicators and the ones already in our sample. Furthermore, when new indicators are included, the number of significant factors might also vary, forcing one to account also for the higher

order principal components. Second, one might enrich the analysis by considering economies at a different stages of the country development process. In particular, this type of research could be useful to detect whether the usage of GDP, interpreted as a univariate indicator of well-being, is more or less appropriate in developed vs. developing economies. For both these extensions, however, the main difficulty lies in data availability, especially in relation to domains different from the economic one.

Acknowledgements We thank the Editor and two anonymous reviewers for useful comments and suggestions. All usual disclaimers apply. All authors received financial support by the European Union Horizon 2020 research and innovation program with grant agreement No. 822781 - GROWINPRO.

Funding Open access funding provided by Università degli Studi di Brescia within the CRUI-CARE Agreement.

Data availability The on-line supplementary material is available at the authors' github pages. Furthermore, footnote 13 is also linked to the following webpage that contains the data and codes allowing one replicate the paper's results: https://github.com/mattiaguerini/indicators_online_material.

Statements and Declarations.

competing financial interests The authors declare that they have no competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendix

A List of indicators

Table 4 List of indicators' names and labels, as sourced from the OECD i-Library database

Name	Label
Agricultural land (Total), Thousand hectares	Land
Average wages (Total), US dollars	Wages
Carbon dioxide (CO ₂), Tonnes per capita	CO ₂
Crude oil production (Total), Thousand of toe	Oilprod
Electricity generation (Total), Gigawatt-hours	ElectrGen
Employment rate (Total), Thousand persons	EmployRate
Fertility rates (Total), Children per woman	Fertility
Gender wage gap, Percentage	WageGap
Greenhouse gas (GHG), Tonnes per capita	GHG
Gross insurance premiums (Total), Million US dollars	Insurance
Health spending (Total), US dollars per capita	Healthspend
Hospital beds (Total), Per 1 000 inhabitants	HospBeds
Hours worked (Total), Hours per worker	HoursWorked
Housing prices, Real house prices	HousePrices
Housing prices, Rent price	RentPrices
Industrial production (Manufacturing)	IndustrialProd
Infant mortality rates (Total), Deaths live births	Mortality
Inflation of consumer price index (Total), Annual growth rate	CPI
Labour force (Total), Thousand persons	LabourF
Investment (Total), Million	Investment
Life expectancy at birth (Total), Years	LifeExpect
Municipal waste (Total), Thousand tonnes	Waste
Nutrient balance (Nitrogen), Kilograms per hectare	Nitrogen
Nutrient balance (Phosphorus), Kilograms per hectare	Phospho
Pharmaceutical spending (Total), US dollars per capita	PharmaSpend
Population (Total), Million persons	POP
Primary energy supply (Total), Million toe	EnergySupply
Producer price indices (Manufacturing), domestic market, Annual growth rate	PPI
Renewable energy (Total), Thousand toe	RenEnergy
Researchers (Total)	Research
Share prices (Total)	SharePrices
Social spending (Public), US dollars per capita	SocialSpend
Suicide rates (Total), Per 100,000 persons	Suicide
Tax revenue (Total), Million US dollars	Taxrev

Table 4 continued

Name	Label
Trade in goods and services (Exports), Million US dollars	TradeExp
Trade in goods and services (Imports), Million US dollars	TradeImp
Triadic patent families (Total)	Patents
Unemployment rate (Total), of labour force	Unemploy
Unit labour costs, by persons employed	UnitLabforce

toe = tonnes of oil equivalent

References

- Aoyama H, Fujiwara Y, Ikeda Y, Souma W (2017) Macro-econophysics: new studies on economic networks and synchronization. Cambridge University Press, Cambridge
- Bacchini F, Baldazzi B, Di Biagio L (2020) The evolution of composite indices of well-being: an application to Italy. *Ecol Indic* 117:106603
- Barbieri C, Guerini M, Napoletano M (2021) The anatomy of government bond yields synchronization in the eurozone. *LEM Working Paper Series* 2021/07
- Ferran PH, Heijungs R, Vogtländer JG (2018) Critical analysis of methods for integrating economic and environmental indicators. *Ecol Econ* 146:549–559
- Fitoussi J-P, Durand M et al (2018) Beyond GDP measuring what counts for economic and social performance: measuring what counts for economic and social performance. OECD Publishing, Berlin
- Fitoussi J-P, Durand M et al (2018) For good measure advancing research on well-being metrics beyond GDP: advancing research on well-being metrics beyond GDP. OECD Publishing, Berlin
- Guerini M, Luu DT, Napoletano M (2023) Synchronization patterns in the European Union. *Appl Econ* 55(18):2038–2059
- Hoekstra R (2019) Replacing GDP by 2030. Cambridge Books, Cambridge
- Iyetomi H, Nakayama Y, Yoshikawa H, Aoyama H, Fujiwara Y, Ikeda Y, Souma W (2011) What causes business cycles? Analysis of the Japanese industrial production data. *J Japanese Int Econ* 25(3):246–272
- Kalimeris P, Bithas K, Richardson C, Nijkamp P (2020) Hidden linkages between resources and economy: a “beyond-gdp” approach using alternative welfare indicators. *Ecol Econ* 169:106508
- Kichikawa Y, Iyetomi H, Aoyama H, Fujiwara Y, Yoshikawa H (2020) Interindustry linkages of prices—Analysis of Japan’s deflation. *PLoS One* 15(2)
- Laloux L, Cizeau P, Potters M, Bouchaud J-P (2000) Random matrix theory and financial correlations. *Int J Theo Appl Finance* 03(03):391–397
- Luzzati T, Gucciardi G (2015) A non-simplistic approach to composite indicators and rankings: an illustration by comparing the sustainability of the EU countries. *Ecol Econ* 113:25–38
- Malay OE (2019) Do beyond GDP indicators initiated by powerful stakeholders have a transformative potential? *Ecol Econ* 162:100–107
- Maximo M (1987) The difference between welfare and wellbeing and how objective the concept of a good life can be
- Michalos AC (1982) North American social report: a comparative study of the quality of life in Canada and the USA from 1964 to 1974, Volume 2. Springer Science & Business Media
- Mikayilov JI, Hasanov FJ, Galeotti M (2018) Decoupling of CO₂ emissions and GDP: a time-varying cointegration approach. *Ecol Indic* 95:615–628
- Ng AY, Zheng AX, Jordan MI (2001) Link analysis, eigenvectors and stability. In *International Joint Conference on Artificial Intelligence*, Volume 17, pp. 903–910. Citeseer
- OECD (2011). Measuring well-being and progress. How’s Life? 2020: Measuring Well-being <https://doi.org/10.1787/9870c393-en>
- Onatski A (2010) Determining the number of factors from empirical distribution of eigenvalues. *Rev Econ Stat* 92(4):1004–1016

- Pinar M, Cruciani C, Giove S, Sostero M (2014) Constructing the feem sustainability index: a choquet integral application. *Ecol Indic.* 39:189–202
- Roser M (2014) Human development index (HDI). *Our World in Data*
- Stewart GW, Sun J-G (1990) *Matrix perturbation theory*. Academic Press, Boston
- Stiglitz JE, Sen A, Fitoussi J-P et al. (2009) Report by the commission on the measurement of economic performance and social progress
- Taylor D (2011) Wellbeing and Welfare: A Psychosocial Analysis of Being Well and Doing Well Enough. *J Social Policy* 40(4):777–794
- UNDP U N D P (2022) Human development report 2021-22. UNDP (United Nations Development Programme)
- United Nations (2009) *System of national accounts 2008*. New York: United Nations. OCLC: ocn526091359
- Vidal R, Ma Y, Sastry SS (2016) *Generalized principal component analysis*, vol 5. Springer

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.