Centro di Ricerche
e Documentazione
"Piero Sraffa"

Effective Rank and Dimensionality Reduction: from Complex Disaggregation Back to a Simple World

Lefteris Tsoulfidis

Centro Sraffa Working Papers
n. 57

July 2022

ISSN: 2284 -2845
Centro Sraffa working papers
[online]

Effective Rank and Dimensionality Reduction: from Complex Disaggregation Back to a Simple World

Lefteris Tsoulfidis

University of Macedonia

Abstract

In recent years there is a revival of political economy, and discussions are about the near linearities of price rate of profit trajectories. In this article, we argue that economy's input-output data are of low effective dimensionality, meaning that there is overfitting in that it takes only a few eigenvalues and respected eigenvectors for an adequate representation of the movement of prices, and that some of the fundamental features of the economy may be tracked down with the use of a low dimensional system.

Keywords: Labor theory of value; Randomness hypothesis**;** Vertical integration; Effective rank; Eigendecomposition

JEL codes: B24; B51; C67; D46; D57; E11; E32

1. Introduction

In recent years, the research has repeatedly shown that the shape of the price rate of profit trajectories and the wage rate of profit curves are near linear. Curved trajectories do exist, but they are relatively few, and even fewer are the trajectories with a single extremum, and we do not exclude the possibility of two extrema in the relevant region. The explanations offered for these linearities were based on the characteristic distribution of the eigenvalues of the system matrices. More specifically, in the usual dimensions of input-output matrices, the dominant eigenvalue is significantly higher (by 40% to 60%) than the second, followed by the third and a few more, their exact number depending on the size of the matrices. The remainder eigenvalues form a long tail and paint an exponentially falling distribution.

Three hypotheses have been put forward to explain this distribution of eigenvalues and the associated with this linearities:

1. The randomly distributed input-output coefficients (Bródy 1997; Schefold 2020).

2. The closeness of vertically integrated compositions of capital (VICC) between sectors (Shaikh 2016).
3. The low effective-rank or effective dimensionality of the utilized matrices shapes the exponential fall in their eigenvalues, which in turn determines the near-linear features of PRP and WRP curves (Mariolis and Tsoulfidis 2018, Tsoulfidis 2021 and 2022).

The purpose of this study is to examine the extent to which these three hypotheses are consistent with the available evidence (for more see Ferrer-Hernández and Torres-González 2022; Torres-González 2022) and proceed with the less researched third hypothesis by operationalizing a new metric of effective rank based on Shannon entropy.

The remainder of the article is structured as follows: section 2 examines the realism of these competing explanations and introduces the concept of effective rank (and dimensionality) to identify the number of eigen- or singular-values that condition the behavior of the entire economic system. Section 3 illustrates the theoretical discussion by utilizing actual input-output data of the US economy of 15 sectors for (the most recent) year 2020, so the reader may have a better grasp of the usefulness and reliability of the approach. The fourth section concludes with the idea that there is overfitting of data and that fewer data and dimensions compressed in two or three sectors would be adequate to convey the essential behavioral features of the system.

2. Effective rank and dimensionality

Our research has shown that the first of the above hypotheses does not corroborate with the available evidence. The reason is that although a random or rather a near random matrix gives rise to an exponentially falling distribution of eigenvalues. However, it does follow that every skew distribution of eigen- or singular-values comes from a random matrix. Our empirical analysis in Tsoulfidis (2021 and 2022) has shown that the random matrix hypothesis does not pass the statistical tests. First, because the actual output vector, $\mathbf{x}$ of the input-output coefficient matrix is quite different from the standard or right-hand-side (r.h.s.) output vector, $\mathbf{s}$ derived either directly from matrix $\mathbf{A}$ or by its multiplication by the Leontief inverse, $\mathbf{H} = \mathbf{A}[\mathbf{I} - \mathbf{A}]^{-1}$. The idea is that if the two vectors are no different to each other, it follows that the price – rate of profit trajectories and the wage – rate of profit curves will be linear. Second, the employment coefficients vector $\mathbf{l}$ also differs significantly from the left-hand-side (l.h.s.) unique positive eigen vector, $\boldsymbol{\pi}$ of the matrix $\mathbf{A}$ or of $\mathbf{H}$. The idea is that if differences between these two vectors are minimal, it follows that the economy is described by the case of an equal composition of capital between sectors. From the above, it follows that in order for the randomness hypothesis to hold, the vectors $\mathbf{d}_1 = \mathbf{x} - \mathbf{s}$ and $\mathbf{d}_2 = \mathbf{l} - \boldsymbol{\pi}$ must have their correlation coefficient equal to zero and zero must be their respective covariance coefficient.

Our findings in testing the USA input-output tables of the years 2007 and 2014 of dimensions 54 industries (Timmer *et al*. 2015) and of the years 2012 of 70 industries (www.bea.gov) suggest that the correlation between $\mathbf{d}_1$ and $\mathbf{d}_2$ is statistically significant, and therefore the randomness hypothesis is not consistent with the data in spite of the fact that the covariance of the two vectors was found zero[1]. The zero covariance does not help, because of its dependence on the normalization condition. However, the same is not true with the correlation coefficient, and if the correlation coefficient is positive and statistically significant, the randomness hypothesis does not hold on purely statistical reasoning. Besides, there are other more intuitive and systematic reasons related to the nature of technological change and the associated input-output coefficients, whose value is declining over time (Carter 1970 and Tsoulfidis and Tsaliki 2019). The persistence of the ranking of industries according to backward, forward and their total linkages is another reason that renders the randomness hypothesis not coming to terms with the empirical evidence (Tsoulfidis and Athanasiadis 2022).

The exponentially decreasing distribution of eigenvalues is also consistent with the remaining two hypotheses from which the closeness of VICCs to the economy-wide average is quite appealing to researchers. The idea is that if the VICCs are too close to each other, except for just a few, it follows that the maximal eigenvalue (along with a few others) will be crucial for the behavior of the entire economy lending support to the conceptualization of one commodity world (OCW) economies. The remainder of eigenvalues will be flocking together at negligibly small values, whose effect will not be felt in the economy. The trouble with this hypothesis is that the estimation of VICCs depends on equilibrium prices for which we need the VICCs. In short, there is cyclicality, which can be hardly overcome unless the estimations are carried out in terms of labor values or market prices or simply by stipulating that all three kinds of prices end up in quite close estimates. However, the question becomes, how can one decide between too different or too similar VICCs? There is no such metric, and the notion of the VICC, although intuitively in the right direction, nonetheless requires further qualifications. Thus, it becomes imperative to invoke (if not contrive) a metric that is independent of prices.

Consequently, we are left with the third in line hypothesis which we need to introduce first and then discuss its explanatory content. Roy and Vetterli (2007) are from the first that proposed a metric for the estimation of the effective rank of a matrix[2]. In order to find the required number of terms to be included in the representation, they employ the Shannon (1948) entropy index or the spectral entropy defined as

$$S = -\sum_{i}^{n} \sigma_i \log \sigma_i \tag{1}$$

where $\sigma_i{}'$s stand for the normalized singular values of the matrix, whose effective rank we want to estimate, with i=1, 2, …, n. Thus, $\sigma_i = s_i / \sum_i^n s_i$ where $s_i = s_1 \geq s_2 \geq \cdots \geq s_n \geq 0$ are the singular values.

[1] All four vectors are normalized in the unit simplex, that is, the sum of their elements is equal to one.
[2] Their metric is inspired by the work of Campbell (1960).

By stipulating that $0\log(0) = 0$, the effective rank ($erank$) of our matrix can be written

$$erank(\mathbf{H}R) = e^S \quad (2)$$

It follows that the more similar the singular values, the higher the entropy, whose maximum is attained when $\sigma_i = n^{-1}$ for all $i = 1, 2, \ldots, n$. In the hypothetical case that all the $\sigma_i{}'$s are of the same value, the entropy will be $-\log n^{-1}$. The exponential of this term gives an effective rank equal to one whereas the maximal nominal rank might be n, that is, the number of linearly independent rows or columns. In the case of a random matrix its effective rank will be 1 and the nominal n.

However, the following statement by Roy and Vetterli (2007): "In the sequel, all logarithms are to the base *e* and we adopt the convention that $0\log 0 = 0$", unfortunately, made the present author utilize natural logarithms and the derived results were not of any help at all. But as they say "every cloud has a silver lining", which in this case led the research to indirect estimates of the effective rank through an eigendecomposition of the matrix $\mathbf{H}R = \mathbf{A}[\mathbf{I} - \mathbf{A}]^{-\mathbf{1}}R$, where R the reciprocal of the maximal eigenvalue of $\mathbf{H}$. The matrix $\mathbf{H}R$ can be restated by an eigen or spectral decomposition (Meyer 2001, pp. 243-4, Mariolis and Tsoulfidis 2018). That is, the matrix $\mathbf{H}R$ is cast in matrix terms formed from its eigenvalues and eigenvectors, such that the sum of these terms gives the original matrix. Thus, we may write

$$\mathbf{H}R = (\mathbf{y}_1\mathbf{x'}_1)^{-1}\mathbf{x'}_1\mathbf{y}_1 + \lambda_2(\mathbf{y}_2\mathbf{x'}_2)^{-1}\mathbf{x'}_2\mathbf{y}_2 + \cdots + \lambda_n(\mathbf{y}_n\mathbf{x'}_n)^{-1}\mathbf{x'}_n\mathbf{y}_n \quad (3)$$

where, λ_i $(i = 1, 2, \ldots, n)$ stands for the normalized eigenvalues of the matrix $\mathbf{H}$ with the dominant $\lambda_1 = 1$, and $\mathbf{y}$ and $\mathbf{x}$ are the l.h.s. and r.h.s. eigenvectors, respectively. The prime over the vector $\mathbf{x}$ indicates its transpose. The first or the maximal eigenvalue is denoted by $\lambda_1 = 1$ whereas the second eigenvalue by λ_2 and the remainder or subdominant eigenvalues by λ_n. Since each of the formed matrices is the result of multiplication by two vectors, it follows that their respective rank will be equal to one. In adding more terms, we merely increase the rank of the resulting matrices according to the number of their terms.

It is of great interest to test if the eigendecomposition of a matrix of input-output coefficients and the metric based on the Shannon index give the same effective rank. In this case, we argue that by combining these two measures, we arrive at more definitive (from a practical point of view) conclusions about the effective rank of the matrix. For this reason, in the section below, we introduce an illustration based on actual input-output data of the US economy of 2020, the most recent data as of this writing.

3. An illustrative example based on input-output data of the USA (2020)

We utilize the more recent input-output table of the US economy of the year 2020 starting with the 15x15 sectoral structure of total requirements, or what is the same as the Leontief inverse, $[\mathbf{I}-\mathbf{A}]^{-1}$.[3] The matrix of input-output coefficients $\mathbf{A}$ is obtained by inverting the Leontief inverse and subtracting it from the identity matrix, $\mathbf{I}$. Thus, we arrive at the matrix $\mathbf{H}=\mathbf{A}[\mathbf{I}-\mathbf{A}]^{-1}$, that is, the matrix of vertically integrated input-output coefficients. Furthermore, we estimate the vector of employment coefficients $\mathbf{l}$ by dividing the sectoral wages by the respective output both available in the commodity by industry table of the same source. We adjust these findings by dividing by the economy-wide average wage as this is given by the social security administration (https://www.ssa.gov). The workers consumption goods vector $\mathbf{b}$ is obtained by multiplying the so-obtained average money wage times the share of consumption goods in the total of each sector. With the help of these vectors and matrices (see Tsoulfidis 2021, 2022, and the literature cited there) we estimate the actual trajectories through the following formula

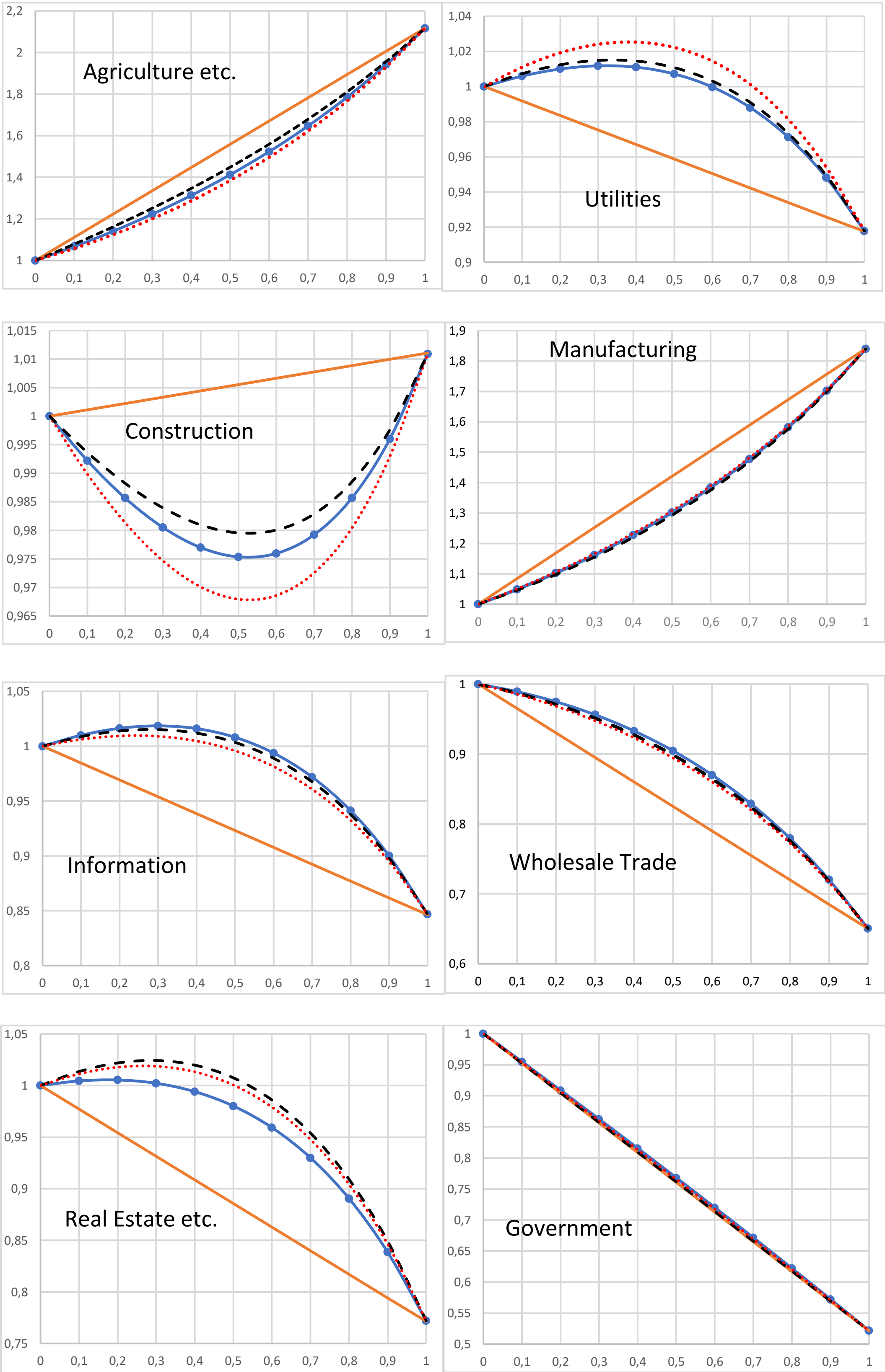
$$\mathbf{p}=(1-\rho)\mathbf{v}[\mathbf{I}-\mathbf{H}R\rho]^{-1} \tag{4}$$

where $\rho\equiv r/R$ is the relative rate of profit, that is, the ratio of the rate of profit, r corresponding to the reciprocal maximal eigenvalue of the matrix $\mathbf{A}[\mathbf{I}-\mathbf{A}-\mathbf{bl}]^{-1}$ and R, the maximal rate of profit corresponding to the reciprocal eigenvalue of matrix $\mathbf{H}$. Finally, $\mathbf{v}$ stands for the vector of labor values $\boldsymbol{\lambda}=\mathbf{l}[\mathbf{I}-\mathbf{A}]^{-1}$ which obtain their monetary expression (direct prices), when they are normalized according to $\mathbf{v}=\boldsymbol{\lambda}(\mathbf{ex})(\boldsymbol{\lambda}\mathbf{x})^{-1}$, where $\mathbf{e}$ is the row (1x15) summation vector.

For reasons of clarity of presentation and economy in space, in Figure 1 below we select to display eight out of our fifteen price curves for illustrative purposes. In each of the panel of eight graphs we also display the three approximations (linear, quadratic and cubic) according to relation (3). In the panels of graphs of Figure 1 we display just a few sectors (those with the most curved trajectories except for the last one, which is almost linear) for illustrative purposes. On the horizontal axis of each of the graphs, we display the relative rate of profit ρ, and on the vertical axis the ratio of estimated price $\mathbf{p}$ over the values $\mathbf{v}$, or p_i/v_i. The straight lines refer to the linear approximations, the dashed black lines stand for the square approximation, the red dotted lines represent the cubic approximations, finally, the blue line with the round markers stands for the actual estimated prices whose paths we want to approximate. The crossing of line of price-value equality indicates a change in the characterization of capital intensity (Sraffa 1960).

[3] The following are the fifteen sectors: 1. Agriculture etc., 2. Mining, 3. Utilities, 4. Construction, 5. Manufacturing, 6. Wholesale trade, 7. Retail trade, 8. Transportation and warehousing, 9. Information, 10. Finance, insurance, real estate, 11. Professional and business services, 12. Educational services, health care, and social assistance, 13. Arts, entertainment, recreation, accommodation, and food services, 14. Other services, 15. Government.

Figure 1. Linear, Quadratic and Cubic Approximations



We find that the linear approximation (as this is judged by the Mean Absolute Deviation) is quite satisfactory even in case that one would only accept a relatively minimal deviation. The quadratic approximation, in general, is an improvement over the linear, even for this small size of input-output description, but one cannot say the same with the cubic, which we find, in most cases, excessive and therefore redundant.

One would be wondering of whether the same answer we would derive through the exponential of the Shannon index of entropy. We apply the singular value decomposition (SVD) method in the matrix **H**. The idea is that there will be subdominant eigenvalues negative and complex numbers. They differ from the eigenvalues of the same matrix **H**, in that they are the positive square roots of the eigenvalues of the matrix $\mathbf{H}'\mathbf{H}$, which are no different from those of the matrix $\mathbf{H}\mathbf{H}'$. Our estimates are shown in Table 1 below:

Table 1. Singular values, Shannon's entropy and Effective rank.

Ranking of Singular Values	Singular Values (1)	Normalized Singular Values (2)	Common Logarithms of (2) (3)	The Product of (2)x(3) (4)	
1	1.266776	0.476572	-0.32187	-0.15339	
2	0.500926	0.188453	-0.7248	-0.13659	
3	0.25281	0.095109	-1.02178	-0.09718	
4	0.171247	0.064425	-1.19095	-0.07673	
5	0.138969	0.052281	-1.28165	-0.06701	
6	0.096752	0.036399	-1.43891	-0.05237	
7	0.073059	0.027486	-1.56089	-0.0429	
8	0.045746	0.01721	-1.76422	-0.03036	
9	0.03557	0.013382	-1.87348	-0.02507	
10	0.024936	0.009381	-2.02774	-0.01902	
11	0.017067	0.006421	-2.19242	-0.01408	
12	0.015397	0.005792	-2.23714	-0.01296	
13	0.008659	0.003258	-2.48711	-0.0081	
14	0.006393	0.002405	-2.61887	-0.0063	
15	0.00379	0.001426	-2.84591	-0.00406	
Sum:	2.658	1.000		Shannon (S)	-0.746
				erank=e^{-S}	2.109

From Shannon's index of entropy, S, whose exponential is equal to 2.109, the effective rank of the system matrix is equal to 2, the rank of a matrix is an integer. A result absolutely consistent with the approximations through the eigendecomposition, where we found only marginal improvements adding the quadratic term whereas the cubic term did not improve the approximation in our 15x15 input-output structure, an indication that we should not go beyond the quadratic term in dimensions of this size input-output structure. The same effective rank equal to two we got by using the absolute eigenvalues instead of the singular ones. However, having to choose between the two, the singular values are preferred because they contain all the required information. By contrast, in the case of eigenvalues, the presence of complex numbers prevents the use of the common logarithms, and by taking the absolute values of these numbers some information may be lost, so that in some marginal cases the derived threshold integer may misrepresent the effective rank. By contrast, the singular values are always positive, and we can take their common logarithms without any loss of information.

4. Results and their evaluation

We have also experimented with input-output data of various dimensions for the same country and years. The differences we found were in the decimals, which, however, do not play any role because at the end the rank must be a one-digit number. More specifically, our estimates for the US economy of the benchmark years 2007 and 2012 showed that the 15x15 dimensions gave that a quadratic approximation would be adequate. The higher dimensions (71x71 industries) input-output matrices, when tested for the same (2007 and 2012) years (www.bea.gov), gave an effective rank twice higher than that of the 15x15 dimensions. However, the spectral decomposition indicated that, for all practical purposes, a cubic term is a satisfactory enough approximation. The fourth or fifth terms did not improve the approximation (Tsoulfidis 2022). We have also tested the 405x405 dimensions input-output data, which gave an effective rank equal to eight. We did not, at present, try eigen approximations for these super high input-output tables. We have also estimated the old 65x65 tables of the BEA, which gave effective ranks or dimensions equal to four. Not surprisingly, the eigen or singular values distribution has remained the same over the years.

The results in the case of matrices of lower dimensions 54x54 of the USA, 2007 and 2014 (Timmer et al. 2015) were quite similar. In both matrices, we found that the quadratic approximation of the price trajectories is more than satisfactory (Tsoulfidis 2021). In contrast, the cubic and the quartic terms did not add much information, even in those trajectories characterized by the highest curvature. These particular trajectories are those of the minimal difference between prices and labor values, indicating the closeness of their VICCs to the economy-wide average or the standard ratio. The results for the other countries, to the extent tested, were no different from those of the US economy. The distribution of eigen and singular values displayed a repeated pattern described by the exact same parameters of an exponential equation whose fit in the distribution of the eigenvalues of all years and countries tested has been extremely good. These results lead to the idea that there are certain regularities embedded deeply in the available input-output data and they are manifested through the skew distribution of eigen or singular values, which in turn determine the effective rank and dimensions of the system matrices. From a mathematical point of view, the idea of the effective rank and dimensions and their estimation through the above based on the Shannon entropy index is quite reasonable. After all, the top few singular values are distinct and quite different from the bulk of singular values, and these top singular values are those that compress a lot more information than the rest of the singular values combined.

Finally, the matrix of fixed capital stock derived through the capital flows tables indicated much lower dimensions, and the quadratic term would be more than enough. After all, the second eigen or spectral value in these matrices is markedly lower than the maximal. Besides, in capital stock matrices, as expected, there are too many rows with zero elements. The idea is that neither the consumer goods industries nor services produce any capital goods, so their rows are filled either with zeros or with relatively small numbers. It is important to point out that the multiplication of the capital stock matrix by the Leontief inverse gives rise to a new matrix whose form takes on that of the capital stock matrix. In counting the number of zeros in our 65x65 capital stock matrix, we found 39 rows which, when added to the zeros scattered to the rest of cells, amounted to 61 percent of total figures of the capital stock matrix, without counting the near-zero negligibly small elements (Tsoulfidis 2021, pp.71-78 and 181).

From our discussion so far follows that both the spectral decomposition and the effective rank operate complementary to each other and help us approximate economic reality, as this is described in its input-output structure, with solid analytical tools capable of extracting its essential features. The hitherto analysis has shown that Samuelson's (1962) OCW description of the economy was an oversimplification, but so was Ricardo's corn model, Marx's schemes of simple reproduction based on the assumption of equal organic composition of capital between departments, and the currently in use economic growth models. Our findings of near-linear price trajectories by no means suggest that the neoclassical theory is adequate in dealing with real-world features. On the contrary, the problems of the marginal productivity theory of income distribution remain. The idea is that the possible equality of marginal productivity of a factor production with its payment is the result of an identity and not of a causal relationship running from the marginal product of capital to its payment, as expected in neoclassical theory (see Shaikh 2016, ch. 9). Furthermore, the assumption of given endowments with near-perfect substitutability and the subjective nature of preferences permeate the whole neoclassical analysis, regardless of whether it refers to the pure exchange economy or production, which is theorized as indirect exchange. There is a better, by far, alternative couched on the labor theory of value that was abandoned for mainly ideological reasons. Our analysis so far has shown that for the usual input-output structure of the economy the first two eigen or singular values are adequate for the construction of models that mimic the operation of the entire economy. In this respect, the principal components analytical method may be used and it has been used profitably to this direction (Tsoulfidis and Athanasiadis 2022).

5. Concluding remarks

In short, the applied factorization method revealed that the structure of the economies is simpler than is usually thought, and a lot of information is compressed in the maximal eigenvalue of the system matrices while the remaining eigen- or rather singular-values add little additional information. Thus, by limiting ourselves to the first few terms of the eigendecomposition, we obtain a satisfactory approximation of the price trajectories consequent upon changes in income distribution. In doing this, we end up with the view that the actual economies are not like an OCW. The latter would require equal capital intensities between industries, which is another way to say that the system's matrices would have nominal and effective rank equal to one. This does not mean that our multi-commodity world requires all commodities and dimensions to uncover its structural features. In a nutshell, we are dealing with overfitting data and over-dimensional representations of the actual economies. Our analysis has shown that the deep laws of motion of the system can be laid bare by de-noising our data and meaningfully compressing the dimensions of the system to just a few.

References

Bródy, A. (1997), The second eigenvalue of the Leontief matrix, *Economic Systems Research*, 9(3): 253-258.

Campbell, L. (1960), Minimum coefficient rate for stationary random processes, *Information and Control*, 3(4): 360-371.

Carter, A. (1970), *Structural Change in the American Economy*, Boston, Harvard University Press.

Ferrer-Hernández J. and Torres-González L. (2022), Some recent developments in the explanation of the empirical relationship between prices and distribution, *Contributions to Political Economy*, 41(1): 1-29.

Mariolis, T., and Tsoulfidis, L. (2018), Less is more: Capital theory and almost irregular uncontrollable actual economies, *Contributions to Political Economy,* 37(1): 65–88.

Meyer, C. (2001), *Matrix Analysis and Applied Linear Algebra*, New York, Society for Industrial and Applied Mathematics.

Roy, O., and Vetterli, M. (2007), The effective rank: A measure of effective dimensionality, *15th European Signal Processing Conference (EUSIPCO)*, Poznan, Poland.

Samuelson, P. (1962), Parable and realism in capital theory: The surrogate production function, *The Review of Economic Studies*, 29(3): 193-206.

Schefold, B. (2020), What remains of the Cambridge critique of capital theory, if reswitching and reverse capital deepening are empirically rare and theoretically unlikely?, *European Journal of Economics and Economic Policies: Intervention*, 17(2): 220-240.

Shannon, C. (1948), A mathematical theory of communication, *The Bell System Technical Journal*, 27(3): 379-423.

Shaikh, A. (2016), *Capitalism: Competition, Conflict, Crises*, Oxford, Oxford University Press.

Sraffa, P. (1960), *Production of Commodities by Means of Commodities. Prelude to a Critique of Economic Theory*, Cambridge, Cambridge University Press.

Timmer, M., Dietzenbacher, E., Los, B., Stehrer, R. and de Vries, G. (2015), An illustrated user guide to the World Input-Output Database: The case of global automotive production, *Review of International Economics*, 23(3): 575-605.

Torres-González, L. (2022), The Characteristics of the productive structure behind the empirical regularities in production prices curves, *Structural Change and Economic Dynamics*, doi: 10.1016/j.strueco.2022.04.007.

Tsoulfidis, L. (2021), *Capital Theory and Political Economy: Prices, Income Distribution and Stability*, London, Routledge.

Tsoulfidis, L. (2022), Capital Theory Debates: New Developments and Direction, *Investigación Económica,* 81(319): 3-36.

Tsoulfidis, L. and Athanasiadis, I. (2022), A New Method of Identifying Key Industries: A Principal Component Analysis, *Journal of Economic Structures*, 11(2): 1-23.

Tsoulfidis, L. and Tsaliki, P. (2019), *Classical Political Economics. Theories of Value, Competition, Trade and Long Cycles*, Berlin, Springer.

Author contact information:

Lefteris Tsoulfidis

University of Macedonia

Thessaloniki, Greece

Email: Lnt@uom.edu.gr