Opening the Leontief’s black box

A. Mustafin *, A. Kantarbayeva

Satbayev University, 22 Satbayev St., Almaty, 050013, Kazakhstan

* Corresponding author.
E-mail address: butsuri123@gmail.com (A. Mustafin).

Abstract

The paper attempts to reconstruct what goes on inside the black box of the fixed proportions production function underlying the matrix of technical coefficients in the famous Leontief’s input–output analysis. The approach is based on an idea that the conversion of inputs to outputs in industrial supply chains occurs in much the same manner as does the conversion of substrates into different molecules in enzyme-catalyzed biochemical reactions. A bottleneck principle is derived according to which at any moment the steady-state output of a multi-resource, one-product supply chain is completely controlled by a sole, most deficient, factor of production. This dramatic shrinkage of complexity is made possible owing to two key features of the input–output production units comprising the chain: saturated response of the output to the inventory level and weak buffer inventory leakage.

Keywords: Applied mathematics, Economics, Business

1. Introduction

The neoclassical production function, first introduced by Philip Wicksteed [1, p. 3], relates the rate of production of a certain commodity (output), \( Y \), to a set of the involved factors of production (inputs), \( \phi_1, \phi_2, \ldots, \phi_n \):

\[
Y = F(\phi_1, \phi_2, \ldots, \phi_n).
\]  

(1)

It is appropriate to say a few words on the terminology used herein. In the present paper we restrict ourselves to physical factors of production. In its broad meaning,
factor of production is any entity which can lead to increased output as its availability is increased. The factors of production may be of two types: consumable and nonconsumable. By convention, consumable factors are referred to as “inputs”. Interchangeably with “inputs” we will call consumable factors “resources” as it is generally accepted in ecology (e.g., [2, ch. 2]), in spite of economists quite often identify resources with factors of any kind.

The resources are material substances. Consuming the resource means tending to reduce its availability. As the production process takes place, the resource is certainly consumed, used up. By consumption we understand irreversible conversion, physical embodiment, of a resource into a material product.

The nonconsumable factors of production, or funds, such as land, capital and labor, are not resources by this definition. This is not to imply that nonconsumable factors are less important, but that they must be treated in a different way from resources. Funds are not materially transformed into an output they produce. They are transforming tools that turn the involved resources into a product, but are not themselves embodied physically in the product. Although funds are not used up, their amount can change and they are subject to wear-and-tear.

In terms of dimensional analysis, output \( Y \) in formula (1), being the quantity of the commodity produced in a unit of time, is a flow variable. Factors of production that represent resources, most commonly are flows, although in some cases they may be stocks. Fund factors always are stock variables.

Cybernetically, a manufacturing technology for the single-product case may be considered as a converter of the resources \( R_1, R_2, \ldots, R_n \) into a product \( P \) by means of the funds \( \varphi_1, \varphi_2, \ldots, \varphi_m \):

\[
R_1, R_2, \ldots, R_n \xrightarrow{\text{TECHNOLOGY}} P
\]

\[
\varphi_1, \varphi_2, \ldots, \varphi_m
\]

In terms of the systems approach, resources are fed to the converter from the outside, while funds act inside of the black box of the technology. Nonconsumable factors are not spent on the output, however they function to make the transformation of resources to product feasible and efficient, and to enable control of that transformation.

It is commonly supposed that the resources are not interactive: a change in the availability of one resource is assumed to have no effect on the rate of supply of any other resource.

A special case of production function implying zero substitutability between inputs is the Leontief technology [3, p. 38], first proposed for the mathematical apparatus of
the input–output (IO) analysis which arose to deal with the problem of interindustry demand. It has the form

\[ Y = Y_0 \min \left( \frac{x_1}{x_0}, \frac{x_2}{x_0}, \ldots, \frac{x_n}{x_0} \right), \]

where \( x_i \) are inputs, \( x_0 \) are the constant per unit input requirements, \( Y \) is output, and \( Y_0 \) is the scale factor having the dimension of \( Y \).

In Leontief’s approach, production activities of a specific geographic region (nation, province, etc.) are disaggregated into \( n \) sectors (industries) and the transaction of goods among the sectors is analyzed. Each of \( n \) industries both produces a single kind of commodity (output) and consumes products from other industries (inputs) in the process of producing its own output. The intersectoral flows are described in a tabular form of the so-called IO table that records the purchases and sales across the sectors of an economy over a given period of time. Traditionally, an IO table is divided into four quadrants. Focusing solely on quadrant II (“northwest”) that depicts interindustry transactions, we are not going to detail quadrants I, III and IV here. Important these organic parts of IO table may be, they are not essential to understanding of how the Leontief production function is related to the IO analysis. Quadrant II shows the ways that raw materials and intermediate products are combined to produce outputs for sale to other industries and to ultimate consumers. It is \( n \times n \) square matrix whose elements represent flows of commodities, which are both produced and consumed in the process of production of goods. These flows are called inter-industry flows or intermediate demand. (As opposed to final demand, that denotes consumption of goods by non-industry consumers like households, government or exports, who are not producers of other goods by themselves.) The rows of such a matrix reflect the distribution of a producer’s output throughout the economy; the columns show the composition of inputs required by a particular sector to produce its output. Namely, the element \( z_{ij} \) represents the amount of the \( i \)th sector’s output used by the \( j \)th sector (including the consumption of own product, when \( j = i \)) to produce its gross output \( Z_j \). In compiling an interindustry transactions table, the entries can be expressed in physical units (e.g., tons of grain or kWh of energy) or in terms of monetary value (e.g., dollars or euro). For reasons of convenience, the monetary form of IO table dominates.

To produce one unit of the \( j \)th good, \( a_{ij} = z_{ij}/Z_j \) units of the \( i \)th good are needed as inputs in sector \( j \). The quantities \( a_{ij} \) are referred to as technical coefficients (the terms IO coefficients and direct input coefficients are also often used) and are usually assumed to be constant. It is important to note that IO analysis assumes linear relations between inputs and outputs from different sectors as well as linear relations between outputs and final demand. This assumption means that there are no increasing or decreasing returns to scale in production or factor substitution.
Table 1. US technical coefficients 2006 [4].

| PRODUCERS AS CONSUMERS | 1   | 2   | 3   | 4   | 5   | 6   | 7   |
|-------------------------|-----|-----|-----|-----|-----|-----|-----|
| 1 Agriculture           | 0.2403 | 0.0000 | 0.0014 | 0.0345 | 0.0001 | 0.0018 | 0.0007 |
| 2 Mining                | 0.0028 | 0.1307 | 0.0079 | 0.0756 | 0.031 | 0.0004 | 0.0066 |
| 3 Construction          | 0.0035 | 0.0002 | 0.001 | 0.0019 | 0.0039 | 0.0072 | 0.0242 |
| 4 Manufacturing         | 0.1858 | 0.0959 | 0.2673 | 0.3311 | 0.0581 | 0.0558 | 0.1027 |
| 5 Trade, transport and utilities | 0.0774 | 0.0379 | 0.1063 | 0.1003 | 0.0698 | 0.0329 | 0.0439 |
| 6 Services              | 0.0875 | 0.1298 | 0.1262 | 0.1239 | 0.1846 | 0.2889 | 0.2029 |
| 7 Other                 | 0.0102 | 0.0096 | 0.0095 | 0.0233 | 0.0223 | 0.0192 | 0.0225 |

Moreover, all producers in a given industry are assumed to employ the same production technology.

Technical coefficients also form an $n \times n$ matrix. The columns of such a matrix describe the composition of inputs required by a particular sector to produce a unit of its output. A real example of the matrix of technical coefficients is shown in Table 1 borrowed from [4, p. 715]. It is derived from a highly aggregated, seven-sector version of the 2006 United States IO table by dividing each cell of the intermediate matrix by gross output ascribed to its column. Thus, for instance, in the first row–fourth column, the technical coefficient ($a_{14}$) is equal to 0.0345. This coefficient shows the rate at which inputs are transformed into outputs. Here, agricultural products valued at 0.0345 dollars are purchased by manufacturing sector in order to produce 1 dollar of manufactured output.

One may see that the very definition of the technical coefficients implies there is only one technique for producing output and requires combination of the inputs in a fixed ratio; the elasticity of substitution is zero. This corresponds to a production function of $j$th industry

$$Z_j = \min \left( \frac{z_{1j}}{a_{1j}}, \frac{z_{2j}}{a_{2j}}, \ldots, \frac{z_{nj}}{a_{nj}} \right),$$

which is the same as (2) up to notation. Referring to the technical coefficients given in Table 1, it is possible to write down, for example, the Leontief production function for the mining industry (ref. column 2):

$$Z_2 = \min \left( \frac{z_{22}}{0.1307}, \frac{z_{32}}{0.0002}, \frac{z_{42}}{0.0959}, \frac{z_{52}}{0.0379}, \frac{z_{62}}{0.1298}, \frac{z_{72}}{0.0096} \right).$$

(Note that the production of agriculture is not used in the production of mining sector, and hence $a_{12} = 0$ and $z_{12}/a_{12} \to \infty$. However this infinitely large ratio may be omitted from the “min” operator because it does not affect the process of searching for the smallest among the ratios.)

In the basic version of the Leontief production function intended for static analysis, inputs are meant to be flow variables—resource influxes. Besides, all quantities are measured in money values because IO models are typically implemented using...
databases in money values only. The more recent developments of the IO analysis, including the dynamic extensions of the basic IO model, do admit stock arguments and allow for physical units [5, 6].

In such sectors as agriculture, forestry and fishery the use of production function with stock arguments proved to be more appropriate because it is dictated by the very specifics of the relationship between yield (harvest, crop, catch, etc.) and nutrients. This is expressed by the Liebig production function [7]:

\[ Y = \min \left( f_1 \left( x_1 \right), f_2 \left( x_2 \right), \ldots, f_n \left( x_n \right) \right), \]  

where \( Y \) is the observed crop yield, \( x_i \) are the concentrations of nutrients, and \( f_i \) are the potential yield functions for \( i = 1, \ldots, n \). The Liebig production function was first proposed by Edgar Lanzer and Quirino Paris [8]. As a matter of fact, the equation (3) is a mathematical formulation of the famous “law of the minimum” commonly credited to a German chemist Justus von Liebig who widely popularized the conjecture originally proposed by his compatriot, a botanist Carl Sprengel [9]. This law states that the rate of growth of a plant or crop is affected not by the most abundant mineral resource, but by the most deficient one. Essentially, a plant will only yield as much as the least available nutrient allows.

As Paris noticed [7], in contrast to the Leontief production function, the statement known as Liebig’s law of the minimum does not necessarily imply the linearity of the relation between nutrients and yield; Liebig asserted only that the nutrients involved and the output stand in direct relation. Quite often, however, the term “Liebig production function” is being reserved for the special case of linear relations \( f_i \) in the equation (3), when the latter is formally identical with the Leontief production function; production functions with specific nonlinear terms \( f_i \) bear different names.

To get a certain idea of the empirical Liebig production function, the statistical estimates by means of maximum likelihood methods carried out by Paris [7] are worth citing. In his work the experimental data on response of corn to applied nitrogen (\( N \)) and phosphorus (\( P \)) on a calcareous Ida silt loam soil in western Iowa have been analyzed. Nitrogen was applied in the form of ammonium nitrate and phosphorus in the form of concentrated superphosphate. The estimates of a Liebig model corresponding to linear relations are as follows:

\[ Y = \min \left( 29.04 + 0.9601 N, 20.04 + 1.2219 P, 124.61 \right), \]  

where the crop yield, \( Y \), is in bushels per acre, and the amounts of nutrients, \( N \) and \( P \), are in pounds per acre. Let us remark here, that, strictly speaking, the quantities \( N \) and \( P \) in (4) are rather integral doses than actual concentrations.

In some instances agricultural economists tend to treat arguments of the Liebig production function as flows, thereby making it the Leontief production function.
For example, Nijland et al. [10] propose the following response function “plant production–nutrients” claiming it to be the Liebig production function:

\[
r = \min\left(20, \frac{S_N}{0.2}, \frac{S_P}{0.02}\right),
\]

where \(r\) is the rate of dry matter production in \(g/(m^2 \cdot d)\), and \(S_N\) and \(S_P\) are the respective supply rates of nitrogen and phosphorus in \(g/(m^2 \cdot d)\).

Neither the Leontief, nor the Liebig production functions are derived; they are merely postulated. Formally, the Leontief function can be inferred from the production function with constant elasticity of substitution (CES), as is done, e.g., in works [11, p. 20] and [12, 13]. But still the CES production function per se is nothing more than a formal mathematical construction.

In an industry sector, a path from primary resources to a finished product may run through a complex supply chain network of “elementary” production units—IO converters. In such a network, the product of one unit, or workstation, serves as the resource for the other. We restrict our consideration to the case of converging structure of supply chain, which allows for multiple initial resources and a single final product. The following commutative diagram shows an example of one-product supply chain with six initial suppliers:

Here circles stand for IO production units, while arrows indicate material flows.

As empirical studies evidence, in some occasions, at least in the short run, the production function of a whole sector can be effectively represented by the Leontief function [14]. The resulting lumped, or aggregated, production function depends not on the total number of the factors of production involved, but on the scarcest one (limiting factor). The behavior of the entire system turns out to be governed by only very few degrees of freedom. The questions arise:

- What are the reasons for such an enormous reduction of the description? Namely, what peculiar properties of the supply chain generate the production function of the Leontief type?
- Does the production system have the property of scale invariance?
- How are the formalisms of Leontief and Liebig interrelated?

These and other questions are addressed in the present paper.
2. Related work

At present, queueing theory, a major branch of operations research that deals with mathematical models of waiting lines using the apparatus of probability theory, is held to be the principal mathematical tool to describe production processes (e.g., [15]). A comprehensive review of the literature on applications of queueing theory in manufacturing is clearly beyond the scope of this paper. Instead, we briefly review the results of this theory most relevant to our study.

A queueing phenomenon in manufacturing is characterized by three main elements: (i) jobs (production lots or parts); (ii) the queue (buffer); and (iii) the server (machine), the purpose of which is to handle the jobs. The lots arrive from outside the system according to a statistical distribution and the distribution describes their interarrival times. Any lot joins a queue in the buffer and waits until the machine is available. At various times, lots are selected for processing by the machine. The basis on which the jobs are selected is determined by the queue discipline. Typically, jobs are served in order of arrival.

Important to us are implications of the most basic and tractable queueing model M/M/1. In standard Kendall’s notation, letters “M” and “M” in the descriptor designate exponential distribution of interarrival and processing times respectively (“M” comes from Markovian), and “1” indicates that the number of parallel servers is just one. Let $\lambda^{-1}$ be the mean interarrival time and $\mu^{-1}$ the mean processing time. Define the utilization, $u = \lambda / \mu$, as the fraction of time the machine is not idle. Utilization has no dimension and can never exceed 1, otherwise, the queue length will explode. Let $W$ be the current work in progress (WIP), i.e. the total number of jobs in the manufacturing system, or the total length of the queue. Then it can be shown [15, ch. 8] that in steady-state conditions, the utilization is related to the WIP by the formula $W = u / (1 - u)$. The throughput, $Y$, of any queueing system is the rate at which jobs successfully leave the system. For the M/M/1 infinite buffer case, $Y = \lambda$ if the system is stable. (Everything that arrives must eventually depart.) Therefore, one may obtain $Y = \mu W / (1 + W)$. As is seen, there is a trade-off between the throughput and the WIP in the M/M/1 model. If a high throughput is required, the machine should always be busy and the inventory level needs to be high, thereby lengthening the time a job is in the manufacturing system (from entering the buffer in front of the machine until leaving the machine). Conversely, if the WIP level is low, the machine is not processing for most of the time, yielding a small throughput.

The deterministic relation between the WIP and the throughput of a steady-state production process bears a name of “clearing function”. Dealing with stochastic model of real production process involves severe computational load. Instead, there is a general consensus to employ proper aggregate models intended to represent average behavior in some sense. Nowadays the formalism of nonlinear
clearing functions devised by Uday Karmarkar [16] shows considerable promise in production planning. The Karmarkar clearing function has the following form:

\[ Y = Y_m W / (K + W), \]  

where \( Y \) is the throughput and \( W \) is the WIP. Constant \( Y_m \) represents the maximum yield achieved by the production system at saturating WIP level; constant \( K \) is the WIP at which the throughput is half of the maximum. The alternative forms of clearing function sharing the common properties of monotonicity, concavity and saturation have been proposed as well [17].

Our research attempts to reconstruct what goes on inside the black box of the fixed proportions production function. The approach differs significantly from the conventional and is based on an idea that the conversion of inputs to outputs in an economy occurs in much the same manner as does the conversion of substrates into different molecules in enzyme-catalyzed biochemical reactions. Indeed, in a living cell, a substrate molecule binds to an enzyme to form a substrate-enzyme complex. The complex then breaks up into a product and the original enzyme, which can then catalyze a new reaction (e.g., [18]). In any production unit of a supply chain, funds play role of enzymes.

Presumably, a Vanderbilt University professor Nicholas Georgescu-Roegen, the founder of ecological economics and bioeconomics [19, 20], was the first to come up with an idea that a production factor such as labor is like a catalyst: “The assumption is that labor, while an indispensable factor in the production of any commodity, can be substituted by other factors beyond any limit. Or in other words, an output of any size can be obtained by any industry with as little labor as we may wish, provided that nonlabor inputs are available in unlimited amounts. Since this property presents an obvious analogy with that of a catalyst in a chemical reaction, we propose to say that labor is catalytic in industry \( G_k \) if it has the mentioned property in that industry” [21, p. 319] (ref. also [22]).

Virtually at the same time, a Novosibirsk-based cyberneticist Igor Poletaev proposed a general mathematical theory of systems with limiting factors, in which production yield is given by a switching function of the Leontief–Liebig type with such arguments as flows of input components and stocks of fund components [23, 24]. Here is his definition of a process with limiting factor: “The typical prototypes of the process of production-consumption type are chemical reaction in the presence of a catalyst, and process of industrial production. In both cases, three kinds of components are involved: inputs (substrate, raw material), funds (catalyst, equipment) and outputs (product of reaction, commodity). The process, in the course of which the quantities of all the participating components ought to be in strictly fixed proportions, we will call stoichiometric process, or process with a limiting factor, or L-process, for brevity” [23, p. 72, own translation].
Independently, a physicist Dmitrii Chernavskii from Moscow, the originator of the concept “protein–machine” [25, 26], deepened the analogy between biosynthesis and industrial production. In particular, he pointed out: “Both by spirit and by methods of research, modeling the economic and production processes is closely related to the subject we expounded above. There is nothing surprising in that biological systems with their basic variables—concentrations of substances—are similar to economic ones, where variables are the quantities of products or commodities, and the role of enzyme concentration is played by the number of machines in a shop or automatic line. In this regard, both the kinetic models of biophysics and biochemistry, and the economic models belong to the common branch of cybernetics, the so-called theory of complex systems” [27, p. 134, own translation].

Regrettably, the above mentioned inspiring insights remained barely noticed over the years. Though it is well known that in search of explanation for the lawlike regularities governing economic systems scholars would repeatedly turn to patterns of biological organization. Suffice it to recall the great names of Adam Smith, Thomas Malthus, Thorstein Veblen and Alfred Marshall. The similarity of laws of competition in the world of living organisms and among economic entities seemed obvious to them. Evolutionary terms like “natural selection”, “niche”, “survival of the fittest” firmly entrenched in the economic science as self-explanatory, almost intuitive concepts. However, later on, as the axiomatization of the economics progressively followed the model of theoretical mechanics, the paths of evolutionary biology and economic theory diverged for a long time.

The collapse of mechanical philosophy in natural science was not followed by a rapid and unconditional departure from it in economic theory. The situation started to change in mid 1980s with publication of the seminal work by Richard Nelson and Sidney Winter [28] and the newly emerged field of evolutionary economics. Nelson and Winter first pointed out the existence of two dialectically opposite processes in an economic development—variability and selection, similar to biological mutations and Darwinian selection. The former means the emergence of innovations as a result of heuristic search process that combines both dynamic and stochastic behavior of firms, while the latter corresponds to competitive survival and adaptation. The role of genome in their theory is assigned to “routines”—patterns of behavior. The routines include: (a) information of managerial and technological nature that is handed down within the organization and also can be transferred to another firm; (b) standard investment operations; and (c) the technique of innovative search.

By a number of key features, both economic and biological systems belong to the type of self-organizing:

(i) They are open, continuous-flow systems, exchanging energy, matter and information with the environment. At the microeconomic level of description,
the concept of “industrial metabolism” proposed by Robert Ayres [29] seems to be extremely helpful. The word ‘metabolism’ in its original biological meaning characterizes the totality of internal biochemical processes in living organism. An individual cell or the whole organism consumes energy-rich, low-entropy substances to maintain its basic functions, as well as for growth and reproduction. This process is necessarily accompanied by the release of high-entropy waste. Industrial metabolism is an integrated set of physical processes aimed at transforming raw materials, energy, labor and capital into goods and associated waste in a steady mode of operation. The analogy between biological and industrial metabolism is not only about the fact that in both types of systems takes place the conversion of material substances driven by a flow of free energy. For dissipative systems of such a type, relaxation to an equilibrium analogous to “heat death” is impossible: nonzero flows through the system persist even in a steady state;

(ii) Both systems are nonlinear, where autocatalytic processes may occur due to the presence of positive feedback loops. Self-reproduction (self-copying) serves to preserve the information previously created and stored in the system. Economic agents are able to store and hand down managerial and technological information physically recorded in books, databases, technical documentation, software, etc. In biology, at the level of organism, self-reproduction occurs owing to the replication of genetic information recorded in DNA as a sequence of nucleotides. At the same time, biological gene cannot be matched to one unique concept from the economy;

(iii) Both systems are path-dependent (historically-conditioned), irreversible, in the sense that each economic agent or each organism develops, qualitatively changing in time. Its current state is the result of both dynamic and statistical events. Variability in self-replication is the main source of new information. Hereditary technological information is subject to accidental changes as a result of the heuristic nature of innovation motivated by entrepreneurial activity. The mutational variability of the genetic material is due to irremovable thermal noise affecting enzymatic reactions. Another source of randomness—sexual recombination of genes—has no analog in economic evolution. But in economic development, it is possible to inherit new features acquired through learning, that means Lamarckian evolution;

(iv) Both systems are hierarchical systems, where each structural level has its own characteristic space-time scale;

(v) Finally, both types of systems share such an important common property of self-organizing systems, as effects of competition. Any emergence of an ordered structure is a result of competition between unstable growing modes: the “surviving” mode suppresses all the rest and imposes its specific structure on the system [30]. The selection is based on a breakdown of the established parity among the competing agents, caused by the arrival of better adapted mutants. For the selection to be
possible, a limitation should be imposed on the total amount of unorganized or organized matter, or on metabolic inflows and outflows. As a rule, competition occurs for the access to scarce resources (in terms of their stock or flow): between business entities in economy and between biological populations. Let us emphasize, however, that it is not the very fact of similarity of economic and biological patterns that is important, but an awareness of the universality of laws of self-organization that lie at their basis.

All the above considerations lie behind our motives to apply a biologically-inspired approach to production process, production function and supply networks. Our research interest to the problem arose shortly after the appearance of the concept of industrial metabolism and was marked by the publication of papers [31, 32] by one of the authors. Treating the act of resource-to-product conversion as a sort of enzyme-catalyzed reaction was shown to result in hyperbolic dependence of the output on the inventory, similar to the Michaelis–Menten saturation curve. The throughput of a cascade of such converters is shown to be determined by a single limiting production factor.

In recent years, sophisticated tools borrowed from the arsenal of queueing theory and supply chain management become more and more relevant in the studies of intracellular metabolic networks [33, 34, 35]. Research work in this direction is spoken of as “biologistics” [36]. In spite of the growing awareness that the biochemical activity of a living cell is similar to the operation of an industrial factory where products of one machine are used by other machines for manufacturing of their own products, biologistics, however, is lacking in attempts to recognize enzyme catalysis mechanism underlying a man-made production system, not the other way around. In other words, rather consider machine as an enzyme than protein as a molecular machine. To the best of our knowledge, the origins of the Leontief production function in supply chains has not been tackled so far to any noticeable extent. The present work is intended to fill that gap.

3. Theory/calculation

3.1. The model of a single resource, single product workstation

Consider a simple production unit—an IO processor, which converts a single resource into a product. The unit consists of a fixed number of identical machines. It is precisely the machinery that represents a nonconsumable factor of production, or fund, in the manufacturing unit under consideration. Each of the machines can process any of the arriving lots of the resource and we assume here that they do so one at a time. The resource arrives at the production unit from the outside, and if all machines are busy processing jobs, the arriving portion of the resource has to wait.
In our model the terms “inventory”, “work in progress” (WIP) and “buffer stock” are regarded as synonyms and mean the current stock of resource in the unit. Lots waiting for service pile up in a common buffer which feeds all machines. When a machine finishes the processing of its current job, it grabs another portion of the resource from the buffer.

We can write the scheme of this event in the form of pseudochemical equations:

\[
\begin{align*}
  &r \rightarrow x^a \\
  &\quad \quad \quad a + v^\beta \rightarrow b + y \\
  &\quad \quad \quad a \cdot u \cdot \beta
\end{align*}
\]

(6)

Here \(r\) is the supply rate of the resource, \(x\) is the inventory, \(q\) is the specific rate with which the resource is being lost (or dispatched to a storage, or exported elsewhere), \(u\) is the number of idle machines, \(v\) is the number of busy (operative) machines, and \(y\) is the quantity of successfully produced commodity. The constants \(a, b, \alpha\) and \(\beta\) depict the various rates with which these processes proceed.

A flow diagram such as the one given by equations (6) can be translated into a set of differential equations that describe rates of change of stock quantities of the participating material agents. The diagram (6) encodes both the sequence of steps and the rates with which these steps occur. To write corresponding equations, we naturally can choose to use what chemists call the “law of mass action”, which states that when two or more agents are involved in a conversion step, the rate of conversion is proportional to the product of their quantities. By convention, the mass-action rate constants are the proportionality constants. Unlike standard chemical reaction schemes, in (6) more than one rate constant—by the number of input and output agents—may correspond to one transformation step, because dimensions of quantities may not coincide.

Suppose units of measurement of the variables involved in (6) are as follows: [\(x\)] = tons of resource, [\(r\)] = h, [\(u\)] = [\(v\)] = machines, [\(y\)] = tons of product, and [\(r\)] = tons of resource/h. Then the units of measurement for the rate constants will be: [\(a\)] = 1/(machines · h), [\(b\)] = tons of product/(machines · h), [\(\alpha\)] = 1/(tons of resource · h), and [\(\beta\)] = [\(q\)] = 1/h.

As to the physical meaning of the rate constants, \(a\) is the capture rate of a unit of the resource by a machine, \(a\) stands for how many idle machines get involved in work in a unit of time per unit of resource, \(\beta^{-1}\) is the mean processing time of a machine (service time of a job at a machine), and \(b\) is the output per a machine, or the number of units of product that a single machine will deliver at one unit of time. They are indicated in the diagram as arrow labels.
Keeping track of each participant allows us to derive the following set of equations:

$$\frac{dx}{dt} = r - aux - qx, \quad (7a)$$
$$\frac{du}{dt} = \beta v - aux, \quad (7b)$$
$$\frac{dv}{dt} = aux - \beta v, \quad (7c)$$
$$\frac{dy}{dt} = bv, \quad (7d)$$

where $t$ is time. All parameters in the model are nonnegative.

Adding equations $(7b)$ and $(7c)$ reveals a conserved quantity $u_0$, the total number of machines, idle and busy:

$$u + v = u_0. \quad (8)$$

This is not at all surprising, since fund is neither formed nor destroyed in the process of manufacturing.

With the aid of $(8)$ the system $(7)$ can be simplified by eliminating either $u$, or $v$. We arbitrarily choose to eliminate $u$. Furthermore, we see that $(7d)$ is just a slave equation with respect to $(7a)$, $(7b)$ and $(7c)$; it can always be solved later on, once solutions for $x$, $u$ and $v$ are known. These steps lead to the following:

$$\frac{dx}{dt} = r - ax(u_0 - v) - qx, \quad (9a)$$
$$\frac{dv}{dt} = ax(u_0 - v) - \beta v. \quad (9b)$$

Introduce new dimensionless parameters: the loss rate constant $\gamma = q/(au_0)$ and the influx $\varrho = ar/(a\beta u_0)$, such that $\gamma \ll 1$ and $|\varrho - 1| \gg \gamma$. In terms of $\gamma$ and $\varrho$, to $O(\gamma)$ the steady-state solutions of $(9)$ are

$$\bar{x}_\pm = \begin{cases} \frac{\varrho}{a} \left( \frac{\varrho^2}{\gamma} + \frac{1}{\varrho - 1} \right) - \frac{\gamma \varrho}{(\varrho - 1)^2}, & \text{for } \pm(\varrho - 1) > 0; \\ \frac{\beta a}{\alpha} \left( \frac{1}{1 - \varrho} - \frac{\gamma}{(1 - \varrho)^2} \right), & \text{for } \pm(1 - \varrho) > 0; \end{cases} \quad (10a)$$
$$\bar{v}_\pm = \begin{cases} u_0 \left( 1 - \frac{\gamma}{\varrho - 1} \right), & \text{for } \pm(\varrho - 1) > 0; \\ \omega u_0 \left( 1 - \frac{\gamma}{1 - \varrho} \right), & \text{for } \pm(1 - \varrho) > 0. \end{cases} \quad (10b)$$

In $(10)$, fixed point $(\bar{x}_+, \bar{v}_+)$ is physically feasible because (i) it is always positive, and (ii) asymptotically stable: both eigenvalues of the Jacobian matrix of the system $(9)$ evaluated at $(\bar{x}_+, \bar{v}_+)$ have negative real parts. The solution $(\bar{x}_-, \bar{v}_-)$ is nonphysical because it yields $\bar{x}_- < 0$ for all legitimate values of $\varrho$. Moreover, it is always saddle-type unstable. From now on we drop subscript at the feasible fixed point. Inserting $\bar{v}$ into $(7d)$ gives the steady-state output of the IO unit under consideration, i.e. the production function. To zeroth order in $\gamma$,

$$\frac{dy}{dt} = b \min \left( ar/(a\beta), u_0 \right). \quad (11)$$
(Here we reverted to the dimensional parameters.) Quite apparently, the equation (11) is a Leontief–Liebig production function for two factors, \( r \) and \( u_0 \). The former is flow and the latter is stock. They enter the function on equal terms. In our approach, flows and stocks are “equalized in rights”. Thus, the steady-state output of an elementary resource-product converter is determined either by the resource supply rate or by the given installed capacity, whichever is in shortest availability. At subcritical arrival rates, when \( \rho < 1 \), the fraction of busy machines is of order \( \rho \), so the machinery is not a limiting factor. At supercritical arrival rates, however, when \( \rho > 1 \), all the machines are engaged barely coping with the huge WIP that becomes inversely related to \( \gamma \).

In steady state, one obtains from (9b) how \( \bar{x} \) and \( \bar{v} \) are related: \( \bar{v} = au_0\bar{x} / (\beta + a\bar{x}) \).

Substituting this in (7d) yields
\[
\frac{dy}{dt} = bu_0\bar{x} / (\beta / a + \bar{x}).
\] (12)

Being the dependency of the throughput (i.e. the number of lots per unit of time that leave the manufacturing system) on the current WIP, (12) is nothing but the clearing function of the production unit (6) under consideration. In more exact terms, we derived the clearing function of the Karmarkar type [16] (cf. (5)).

One can recognize (12) as another version of the Michaelis–Menten equation of enzyme kinetics [18]. The hyperbolic IO relationship of the type (12) is not uncommon in biology: it describes the sigmoidal oxygen-binding curve of hemoglobin and the fraction of a macromolecule saturated by ligand as a function of the ligand concentration (Hill equation [37]), growth rate of microorganisms in a nutrient solution (Monod equation [38]), numerical response of predator to prey population density (Holling type II response [39]), and the like.

A distinguishing characteristic of the equation (12) is saturated response of the output to the inventory. For low levels of \( \bar{x} \), the output is roughly proportional to \( \bar{x} \). At high \( \bar{x} \) levels, though, the rate of production approaches a constant value, \( bu_0 \).

By the way, the equation (12) may be regarded as a production function of the IO processor (6) in stock-flow representation. (The fact that the clearing function is a kind of production function was also mentioned earlier by Salah Elmaghraby [40, p. 132].)

Recall that in deriving (12) we considered the production unit in a steady-state mode of operation. Now we are going to show that under certain additional assumptions the relationship (12) remains valid even in nonsteady-state conditions.

To begin with, we nondimensionalize the equations (9) by introducing the following scaled variables and parameters (in addition to already defined \( \gamma \) and \( \phi \)): \( \xi = ax / \beta \), \( \eta = v / u_0 \), \( \tau = tau_0 \), and \( \epsilon = au_0 / \beta \).
The quantity \((au_0)^{-1}\) is chosen to be a new unit of time. It is a characteristic time a job spends waiting (in the buffer) before beginning service. In other words, it is the resource lifetime in the production unit. This time is seen to be inversely proportional to the total number of installed machines, \(u_0\).

The dimensionless equations then become

\[
d\xi/d\tau = \rho + (\eta - 1 - \gamma)\xi, \\
\varepsilon \, d\eta/d\tau = \xi - \eta(1 + \xi).
\]

(13)

According to the chosen scaling, \(\varepsilon\) is the ratio of the processing time to the characteristic waiting time. Hereinafter we assume this ratio to be small: \(\varepsilon \ll 1\). This is a necessary condition for most of the subsequent reasoning to be valid, although we are aware that the assumption made cannot be ensured for every existing production process. We will turn back to this topic in Section 4.

Inasmuch as \(\varepsilon \ll 1\), the system (13) is singularly perturbed. The slow variable is resource, \(\xi\), and the fast variable is the number of machines in service, \(\eta\). The standard practice of reducing such systems is the adiabatic elimination of the fast variables, when the left-hand side in the fast equation is replaced by zero, thus turning this differential equation into an algebraic equation. It is assumed that the fast variables quickly relax to their momentary equilibrium, quasi-steady-state, values obtained from the algebraic equations, in which the slow variables are treated as parameters. “Frozen” slow variables do not move substantially in this short adaptation time of the fast variables. Quasi-steady-state values of the fast variables can then be expressed by values of the slow variables. The fast variables hastily adapt to the motion of the slow variables. The former are entrained (enslaved) by the latter. The utility of quasi-steady-state approximation is that it allows us to reduce the dimension of the system by retaining only slow variables in the model. One has to establish the validity of the adiabatic elimination in each specific case by using the recommendations of the singular perturbation theory (e.g., [41, ch. 3]). In particular, Fenichel–Tikhonov theorem requires, among other things, (i) quasi-steady state of the fast equation to be an isolated root of the algebraic equation \(d\eta/d\tau = 0\) and to retain stability at all allowed values of the slow variable, and (ii) initial conditions of the fast equation to fall within the domain of influence of that quasi-steady state. It is worthy of note that from the chemists’ side Max Bodenstein pioneered the quasi-steady-state approximation as far back as in 1913. The influential works to clarify the applicability of the technique to enzymatic reactions have been carried out by Włodzimierz Klonowski [42] and Lee Segel [43].

To decompose system (13) into fast and slow parts, introduce fast time variable \(\theta = \tau/\varepsilon\). Now rescale (13) by replacing \(\tau\) with \(\theta\varepsilon\) and, after taking \(\varepsilon = 0\), it becomes
\[
\frac{d\xi}{d\theta} = 0, \quad \frac{d\eta}{d\theta} = \xi - \eta(1 + \xi). \tag{14}
\]

This is the fast subsystem, where \(\xi\) is replaced by its initial value and treated as parameter. It yields the inner solution, valid for \(\tau = \mathcal{O}(\epsilon)\).

Setting \(\epsilon = 0\) in (13) leads to the slow subsystem

\[
\begin{align*}
\frac{d\bar{\xi}}{d\tau} & = \rho + (\eta - 1 - \gamma)\bar{\xi}, \\
0 & = \bar{\xi} - \eta(1 + \bar{\xi}). \tag{15a}
\end{align*}
\]

which produces the outer solution, valid for \(\tau = \mathcal{O}(1)\). In this singular limit as \(\epsilon \to 0\), the subsystem (15) defines a slow flow along the curve (slow manifold) given by (15b). Outer solution is valid for those values of \(\xi\), for which the quasi-steady states of the fast subsystem (14) are stable.

The quasi-equilibrium for the fast subsystem (14) is given by

\[
\eta = \frac{\xi}{1 + \xi},
\]

which is asymptotically stable for any positive \(\xi\).

Hence, it follows from (15) that for time scales on the order of \(\tau = \mathcal{O}(1)\) the process of resource-to-product conversion is given by the equations

\[
\begin{align*}
\frac{d\bar{\xi}}{d\tau} & = \rho - \bar{\xi}/(1 + \bar{\xi}) - \gamma \bar{\xi}, \\
\frac{d\bar{\zeta}}{d\tau} & = \mu \bar{\xi}/(1 + \bar{\xi}). \tag{16b}
\end{align*}
\]

Here we have written down the slave equation (7d) in dimensionless form (16b) by having introduced a normalized product quantity \(\zeta = y/y_0\) and the combined parameter \(\mu = b/(ay_0)\), where \(y_0\) is a proper unit for \(y\).

The equation (16b) is the clearing function in nondimensional form. It looks formally identical with (12), however as opposed to (12), the argument \(\bar{\xi}\) standing for WIP does not have to be constant in time. In deriving (16b) we did not require the production unit to operate in a steady-state mode. And yet, the number of busy machines, \(\eta\), being the fast variable, after a short transient of order \(\mathcal{O}(\epsilon)\) keeps in a quasi-steady state with respect to the current inventory, \(\xi\). In (16a), the resource supply rate, \(\rho\), may not be necessarily constant, but if the timescale of its typical variations is much longer than the machine turnover time, then the equation for the clearing function (16b) will remain valid.

### 3.2. Linear supply chain

Now we pass on to two serially connected production units operating by the generic mechanism as discussed in the preceding subsection:
The first (upstream) unit converts the resource \( x_0 \) to the product \( x_1 \), which, in turn, serves as a resource to the second (downstream) unit. The second unit uptakes \( x_1 \) and converts it to the product \( x_2 \). The two units may represent a fragment of a sequential supply chain, or a linear production line.

Upon adiabatical exclusion of the fast (fund) variables \( u_1 \), \( v_1 \), \( u_2 \) and \( v_2 \) the corresponding balance equations for the slow variables \( x_0 \), \( x_1 \) and \( x_2 \) become:

\[
\frac{dx_0}{dt} = r_0 - \frac{a_1 b_1 u_{10} x_0}{\beta_1 + \alpha_1 x_0} - q_0 x_0, \quad (18a)
\]
\[
\frac{dx_1}{dt} = \frac{a_1 b_1 u_{10} x_0}{\beta_1 + \alpha_1 x_0} - \frac{a_2 b_2 u_{20} x_1}{\beta_2 + \alpha_2 x_1} - q_1 x_1, \quad (18b)
\]
\[
\frac{dx_2}{dt} = \frac{a_2 b_2 u_{20} x_1}{\beta_2 + \alpha_2 x_1}. \quad (18c)
\]

Here \( u_{10} = u_1 + v_1 \) and \( u_{20} = u_2 + v_2 \) are the respective installed machinery of units 1 and 2.

Defining the dimensionless quantities \( \xi_0 = x_0 a_1 / \beta_1 \), \( \xi_1 = x_1 a_2 / \beta_2 \), \( \xi_2 = x_2 / x_{20} \),

\( \tau = a_1 u_{10} t \), \( \gamma_0 = q_0 / (a_1 u_{10}) \), \( \gamma_1 = q_1 / (a_2 u_{20}) \), \( \rho_0 = r_0 a_1 / (a_1 b_1 u_{10}) \), \( \rho_1 = b_1 a_2 u_{20} / (a_2 b_2 u_{20}) \), \( \mu_1 = a_2 u_{20} / (a_1 u_{10}) \), and \( \mu_2 = b_2 u_{20} / (a_1 u_{10} x_{20}) \), where \( x_{20} \) is an appropriate unit for \( x_2 \), we rescale system (18) to

\[
\frac{d\xi_0}{d\tau} = \rho_0 - \xi_0 / (1 + \xi_0) - \gamma_0 \xi_0, \quad (19a)
\]
\[
\frac{d\xi_1}{d\tau} = \mu_1 \left[ \rho_1 \xi_0 / (1 + \xi_0) - \xi_1 / (1 + \xi_1) - \gamma_1 \xi_1 \right], \quad (19b)
\]
\[
\frac{d\xi_2}{d\tau} = \mu_2 \xi_1 / (1 + \xi_1). \quad (19c)
\]

As a matter of convenience, introduce an auxiliary quantity

\( w_1 = \rho_1 \xi_0 / (1 + \xi_0) \)

such that \( \mu_1 w_1 \) is the dimensionless output by the first unit. The values of \( \xi_0 \) and hence, \( w_1 \), do not depend on parameters of the second unit. The equation (19a) has the following steady-state solutions:

\[
\bar{\xi}_0 = \begin{cases} 
\rho_0 / (1 - \rho_0) + O(\gamma_0) , & \text{for } \rho_0 < 1; \\
(\rho_0 - 1) / \gamma_0 + 1 / (\rho_0 - 1) + O(\gamma_0), & \text{for } \rho_0 > 1.
\end{cases}
\]

Accordingly, the steady states of \( w_1 \) turn out to be
\[
\overline{w}_1 = \begin{cases} 
\phi_0 \phi_1 + \mathcal{O}(\gamma_0), & \text{for } \phi_0 < 1; \\
\phi_1 + \mathcal{O}(\gamma_0), & \text{for } \phi_0 > 1.
\end{cases}
\]

To zeroth order in \( \gamma_0 \), this is equivalent to
\[
\overline{w}_1 = \phi_1 \min(\phi_0, 1). 
\]

(20)

To within a constant factor, formula (20) is the production function of the first unit, as is found above (cf. (11)).

For the steady-state values of \( \xi_1 \) we obtain from (19b):
\[
\overline{\xi}_1 = \begin{cases} 
\frac{\overline{w}_1}{(1 - \overline{w}_1)} + \mathcal{O}(\gamma_1), & \text{for } \overline{w}_1 < 1; \\
(\overline{w}_1 - 1)/\gamma_1 + 1/(\overline{w}_1 - 1) + \mathcal{O}(\gamma_1), & \text{for } \overline{w}_1 > 1.
\end{cases}
\]

(21)

It is a matter of direct verification to prove that positive steady states of the system of equations (19a) and (19b) are stable.

Substituting the steady-state values of \( \xi_1 \) in equation (19c) yields, to \( \mathcal{O}(1) \) in \( \gamma_0 \) and \( \gamma_1 \),
\[
d\xi_2/d\tau = \mu_2 \min(\overline{w}_1, 1) = \mu_2 \min(\phi_0 \phi_1, \phi_1, 1),
\]

where we used the equation (20). This is the dimensionless production function of the unbranched two-link supply chain (17). Turning back in to the dimensional quantities, we get
\[
dx_2/dt = b_2 \min(a_1 a_2 b_1 r_0/(a_1 a_2 \beta_1 \beta_2), a_2 b_1 u_{10}/(a_2 \beta_2), u_{20}). 
\]

(22)

This is another Leontief–Liebig production function for the arguments \( r_0 \), \( u_{10} \), and \( u_{20} \). Again, the overall output is controlled either by the resource supply rate or by an installed capacity of one of the two production units, whichever is more deficient.

Suppose, of two IO units placed in series in (17), the second unit happens to control the overall output, while parameters of the first unit do not affect the operation of the chain. According to equation (22), this situation corresponds to \( u_{20} < a_2 b_1 u_{10}/(a_2 \beta_2) \) and \( u_{20} < a_1 a_2 b_1 r_0/(a_1 a_2 \beta_1 \beta_2) \). As this takes place, there is a substantial level of the inventory, \( x_1 \), in the buffer of the second production unit—to the extent that the smallness of \( q_1 \) warrants. Indeed, by formula (21),
\[
\overline{x}_1 \approx \begin{cases} 
\frac{a_1 a_2 b_1 r_0 - a_1 a_2 \beta_1 \beta_2 u_{20}}{a_1 a_2 \beta_1 q_1}, & \text{for } a_1 r_0 < a_1 \beta_1 u_{10}; \\
\frac{a_2 b_1 u_{10} - a_2 \beta_2 u_{20}}{a_2 \beta_1 q_1}, & \text{for } a_1 r_0 > a_1 \beta_1 u_{10}.
\end{cases}
\]

(23)
In other words, the WIP piles up in the second unit to such a level, that makes the rate of production of \( x_2 \) practically insensitive to the variations in \( x_1 \):

\[
dx_2 / dt = b_2 u_{20}
\]

—in conformity with the equation (18c).

Now consider the case of \( u_{10} < a_2 \beta_2 u_{20} / (a_1 \beta_1) \) and \( u_{10} < \alpha_1 r_0 / (a_1 \beta_1) \) for which the first unit operates relatively slow. Then, in view of (22), the output of the second unit is given by

\[
dx_2 / dt = a_2 b_1 b_2 u_{10} / (a_2 \beta_2).
\]

Clearly, the rate of production of \( x_2 \) is completely determined by the throughput of the first unit and does not depend on the machinery of the second unit. In this case the steady-state WIP in the second unit can be estimated using the equation (21):

\[
\bar{x}_1 \approx \begin{cases} 
\alpha_1 b_1 b_2 r_0 / a_1 a_2 \beta_1 \beta_2 u_{20} - \alpha_1 a_2 b_1 r_0, & \text{for } \alpha_1 r_0 < a_1 \beta_1 u_{10}; \\
\frac{b_1 \beta_2 u_{10}}{a_2 \beta_2 u_{20} - a_2 b_1 u_{10}}, & \text{for } \alpha_1 r_0 > a_1 \beta_1 u_{10}.
\end{cases}
\]

The intermediate product \( x_1 \) does not pile up in the buffer of the second unit and undergoes conversion into the product \( x_2 \) without delay.

### 3.3. The converging branch

Finally, consider a production node in a chain that has at most one successor, but is supposed to have two predecessors. In this converging branch, characterized by the vertex with in-degree 2 and out-degree 1, two independent suppliers provide components \( x_1 \) and \( x_2 \) to the downstream manufacturer which then yields product \( x_3 \):

\[
\begin{align*}
\overset{r_1}{x_1} & \overset{q_1}{\downarrow} \\
\overset{r_2}{x_2} & \overset{a_2}{\downarrow} \overset{u}{\Rightarrow} \overset{\beta}{\Rightarrow} \overset{b}{\Rightarrow} x_3 \\
\overset{q_2}{\downarrow} & \overset{a}{\Rightarrow} \overset{u}{\Rightarrow} \overset{\beta}{\Rightarrow}
\end{align*}
\]

This structure can represent, for example, the fragment of a modular assembly supply chain, which nowadays has found applications in many manufacturing industries. In modular supply chain, product modules are being apportioned to intermediate sub-producers. As a result, only a few assembled modules will be delivered to the final producer, which reduces the complexity of the final assembly process.
It is a straightforward matter to draw the balance equations for the scheme (24). According to our assumption, the rate of uptake of either of two resources under enzymatic facilitation of the machinery would be proportional to $ux_1x_2$, where $u$ is the number of idle machines. Thus the equations describing the process will be

$$\frac{dx_1}{dt} = r_1 - a_1ux_1x_2 - q_1x_1,$$

$$\frac{dx_2}{dt} = r_2 - a_2ux_1x_2 - q_2x_2,$$

$$\frac{du}{dt} = \beta v - aux_1x_2,$$

$$\frac{dv}{dt} = aux_1x_2 - \beta v,$$

$$\frac{dx_3}{dt} = bv.$$  

(25a)

(25b)

(25c)

(25d)

(25e)

Noting from (25c) and (25d) that $u + v = u_0 = \text{const}$ and introducing dimensionless variables and parameters $\xi_1 = x_1\sqrt{ar_2/(\beta r_1)}$, $\xi_2 = x_2\sqrt{ar_1/(\beta r_2)}$, $\eta = v/u_0$, $\xi_3 = x_3/x_{30}$, $\tau = ta_0u_0\sqrt{br_2/(ar_1)}$, $\phi_1 = r_1\alpha/(a_1\beta u_0)$, $\phi_2 = r_2\alpha/(a_2\beta u_0)$, $\gamma_1 = q_1\sqrt{ar_1/(\beta r_2)}/(a_1u_0)$, and $\gamma_2 = q_2\sqrt{ar_2/(\beta r_1)}/(a_2u_0)$, we rewrite the system (25) in a nondimensional form

$$\frac{d\xi_1}{d\tau} = \phi_1 - (1 - \eta)\xi_1\xi_2 - \gamma_1\xi_1,$$

$$\frac{d\xi_2}{d\tau} = \mu_2 \left[\phi_2 - (1 - \eta)\xi_1\xi_2 - \gamma_2\xi_2\right],$$

$$\epsilon \frac{d\eta}{d\tau} = (1 - \eta)\xi_1\xi_2 - \eta,$$

$$\frac{d\xi_3}{d\tau} = \mu_3 \eta,$$  

(26a)

(26b)

(26c)

(26d)

where $\epsilon = a_1u_0\sqrt{r_2/(a\beta r_1)}$, $\mu_2 = a_2r_1/(a_1r_2)$, $\mu_3 = b\sqrt{ar_1/(\beta r_2)}/(a_1x_{30})$, and $x_{30}$ is a proper unit for $x_3$. Note that (26d) is slave equation.

Just as in the cases considered above, parameters $\epsilon$ and $\mu_2^{-1}$ characterize by how much the dynamics of the respective variables $\eta$ and $\xi_2$ is faster than that of $\xi_1$. Taking $\epsilon$ to be small while $\mu_2$ to remain within $O(1)$, the variable $\eta$ can be replaced by its quasi-steady-state value

$$\eta = \frac{\xi_1\xi_2}{(1 + \xi_1\xi_2)}.$$  

(27)

Plugging this in the system (26) we obtain slow equations

$$\frac{d\xi_1}{d\tau} = \phi_1 - \xi_1\xi_2/(1 + \xi_1\xi_2) - \gamma_1\xi_1,$$

$$\frac{d\xi_2}{d\tau} = \phi_2 - \xi_1\xi_2/(1 + \xi_1\xi_2) - \gamma_1\xi_2,$$

$$\frac{d\xi_3}{d\tau} = \mu_3 \eta.$$  

(28a)

(28b)

(28c)

The validity of the reduction of (26) to (28) is ensured, in conformity with Fenichel–Tikhonov theorem, by stability of quasi-steady state (27) of the fast equation (26c) at all positive $\xi_1$ and $\xi_2$.

For small loss parameters $\gamma_1$ and $\gamma_2$, steady-state solutions of the pair of equations (28a) and (28b) are as follows:
Besides, it can be shown that the positive fixed points of the pair of equations (28a) and (28b) are stable.

Substituting the steady-state values of $\xi_1$ and $\xi_2$ from (29) into (27) gives $\bar{\eta}$. Inserting the latter into equation (28c) yields, to $O(1)$ in small $\gamma_1$ and $\gamma_2$, the dimensionless production function of the converging branch (24):

$$\frac{d\xi_3}{d\tau} = \mu_3 \min(\phi_1, \phi_2, 1).$$

In its dimensional form, this will look as

$$\frac{dx_3}{dt} = b \min(a_{r1}/(a_1 \beta), a_{r2}/(a_2 \beta), u_0).$$

Clearly, the result (30) is a Leontief–Liebig production function of three factors of production: $r_1$, $r_2$, and $u_0$. This can be extended to multiple inputs. We have focused so far on models with just a few inputs where the concept of modeling with low-order “chemical reactions” is perhaps most natural. However, it is important to recognize that we use the notation of chemical reactions simply to describe things that combine and the things that they produce, and that this framework can be used to model higher-order phenomena in a similar way.

It is easily comprehended from the above analysis, that kinetics of any one-product supply chain of arbitrary length with multiple resources would lead to the overall production function of the Leontief type, providing individual production nodes of the chain follow the generic mechanism similar to that of the enzyme catalysis.

### 4. Discussion and conclusions

Going over to comment our results, we would like to emphasize that emergence of the bottleneck effect in a supply chain is stipulated by two key features of the suggested construction of the Leontief’s black box: (1) two strongly varying timescales involved in the production process—longer, for the inventory level, and shorter, for the number of machines engaged in processing, and (2) weak outflux of the inventory from the buffer.
Presence of time hierarchy makes possible the saturated response of the output to the WIP in the form of the Karmarkar clearing function (in fact the Michaelis–Menten equation), whereas the collateral buffer leakage secures finiteness of the steady-state inventory level.

The formalism of clearing functions is widely used for production planning. But how good is such an approximation? When is it expected to hold, and under what conditions would it fail? These questions are seldom if ever addressed in the current literature on operations research. The rare exception seems to be the review written by Dieter Armbruster [44] recognizing the quasi-steady-state nature of the clearing function and relative slowness of the WIP dynamics. In his own words: “Even at this low level of approximation there is a basic inconsistency: the clearing function is supposed to describe the outflux in steady state as a function of wip level. However, the clearing function is used with a wip level that is a function of time and is updated constantly to determine the outflux as a function of time. Hence, by making the outflux follow instantaneously any change in the wip level, the fundamental assumption is that the wip level changes slowly relative to the damping time of the underlying stochastic process. Therefore the fundamental assumption that justifies the use of a clearing function is that by the time the wip-level has reached a new state, the stochastic process determining the outflux is back in steady state. As a result the outflux is never in transient and always characterized by its steady state behavior. This is known as the quasi-steady assumption or the adiabatic model. The quasi-steady assumption poses a major problem for the applicability of any type of clearing function approach. Since almost no research in production planning is concerned with the specific nature of the stochastic process, there are no good estimates to my knowledge about the damping time of the stochastic processes. In fact, even the concept is ill-defined without discussing the timescales and magnitudes of the stochastic disturbances” [44, p. 291]. However the fast variable, which is supposed to stay in quasi-steady state towards the WIP, remains unspecified in the mentioned work. In terms of our bio-inspired model, with the background given above—especially with the concept of two timescales, we are able to suggest a more sound justification of the clearing function: the momentary number of machines in the operating state would be in a quasi-steady state with respect to the WIP provided the processing time is much shorter than the characteristic waiting time (the typical time it takes to load a machine with resource). Consequently, the condition $\epsilon \ll 1$ is expected to be sufficient to assure the validity of the clearing function in nonsteady supply chains.

Thus it is shown that the Leontief production function naturally appears in supply chains where output of each individual production node is universally characterized by a saturated response to the WIP. To ensure this type of response it is sufficient to assume that conversion of inputs to outputs in material production occurs similarly
to the conversion of substrates into different substances in enzyme-catalyzed biochemical reactions. The part of enzyme is played by machinery. In the general case this may be any nonconsumable, or primary, factor of production (fund), such as capital, land, or labor.

The production line consisting of units of such a type, has the property of scale invariance: the production function of the whole chain is similar to the production function of any constituent unit. As we found out, a more correct form of the Leontief function is not its conventional flow–flow notation, but the Leontief–Liebig form, where resources and funds intermingle.

It turns out that the output of a one-product supply chain (possibly with multiple inputs and converging branches) is solely controlled by the minimum of its input supplies and funds. The dependence of the output only on the properties of the bottleneck allows the production system to effectively simplify the control, acting only on the bottleneck unit. The considered self-regulation principle is useful for understanding the functioning of complex production networks.

Just as the deterministic approach fails to capture the discrete and stochastic nature of chemical reactions at low concentrations, so does the continuous mass-action treatment of production process at small quantities of factors of production, whether resources or funds. As many manufacturing processes involve IO conversions at extremely small quantities, such discrete stochastic effects are well relevant for our bio-inspired model. For some supply chains, large fluctuations in the WIP may be dangerous. The evolution of the number of parts of a given type due to interactions with machines–catalysts can be described by Markov processes, which can be formalized, for example, in terms of the chemical master equation [45]. Exploring these possibilities will constitute a future direction for work on the model.

**Declarations**

**Author contribution statement**

Almaz Mustafin, Aliya Kantarbayeva: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

**Funding statement**

The results presented in this paper are part of the self-funded research projects 0116RK00250 and 0116RK00275 registered at the Ministry of Education and Science of the Republic of Kazakhstan.
Competing interest statement

The authors declare no conflict of interest.

Additional information

No additional information is available for this paper.

Acknowledgements

No funding sponsors are involved in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

References

[1] P. Wicksteed, An Essay on the Co-ordination of the Laws of Distribution. No. 12 in Reprints of Scarce Tracts in Economic and Political Science, reprint of 1894 ed. London School of Economics, London, 1932, https://socialsciences.mcmaster.ca/econ/ugcm/3ll3/wickstead/wickess.pdf.

[2] D. Tilman, Resource Competition and Community Structure, Monographs in Population Biology, vol. 17, Princeton University Press, Princeton, NJ, 1982.

[3] W. Leontief, The Structure of American Economy, 1919–1939, an Empirical Application of Equilibrium Analysis, 2nd ed., Oxford University Press, New York, NY, 1951.

[4] R.E. Miller, P.D. Blair, Input–Output Analysis: Foundations and Extensions, 2nd ed., Cambridge University Press, New York, NY, 2009.

[5] F. Duchin, Input–output economics and material flows, in: S. Suh (Ed.), Handbook of Input–Output Economics in Industrial Ecology, in: Eco-Efficiency in Industry and Science, vol. 23, Springer Netherlands, Dordrecht, 2009, pp. 23–41.

[6] H. Weisz, F. Duchin, Physical and monetary input–output analysis: what makes the difference?, Ecol. Econ. 57 (3) (2006) 534–541.

[7] Q. Paris, The von Liebig hypothesis, Am. J. Agric. Econ. 74 (4) (1992) 1019–1028.

[8] E.A. Lanzer, Q. Paris, A new analytical framework for the fertilization problem, Am. J. Agric. Econ. 63 (1) (1981) 93–103.
[9] S.A. Ebelhar, W. Chesworth, Q. Paris, Law of the minimum, in: W. Chesworth (Ed.), Encyclopedia of Soil Science, Springer Netherlands, Dordrecht, 2008, pp. 431–437.

[10] G.O. Nijland, J. Schouls, J. Goudriaan, Integrating the production functions of Liebig, Michaelis–Menten, Mitscherlich and Liebscher into one system dynamics model, NJAS Wageningen J. Life Sci. 55 (2) (2008) 199–224.

[11] H.R. Varian, Microeconomic Analysis, 3rd ed., Norton, New York, NY, 1992.

[12] L. Csontos, S. Ray, The Leontief production function as a limiting case of the CES, Indian Econ. Rev. 27 (2) (1992) 235–237, http://www.jstor.org/stable/29793581.

[13] T. Saito, How do we get Cobb–Douglas and Leontief functions from CES function: a lecture note on discrete and continuum differentiated object models, J. Ind. Organ. Educ. 6 (1) (2012) 1–14.

[14] A. Tipper, Capital-Labour Substitution Elasticities in New Zealand: One for All Industries?, Working Paper 12-01, Statistics New Zealand, Wellington, New Zealand, 2012, http://www.stats.govt.nz/~media/Statistics/surveys-and-methods/methods/research-papers/working-papers/capital-labour-substitution-elasticities-nz.pdf.

[15] W.J. Hopp, M.L. Spearman, Factory Physics: Foundations of Manufacturing Management, 3rd ed., McGraw-Hill, New York, NY, 2008.

[16] U.S. Karmarkar, Manufacturing lead times, order release and capacity loading, in: S.C. Graves, A.H.G. Rinnooy Kan, P.H. Zipkin (Eds.), Logistics of Production and Inventory, in: Handbook in Operations Research and Management Science, vol. 4, North Holland, Amsterdam, 1993, pp. 287–329, chap. 6.

[17] H. Missbauer, R. Uzsoy, Optimization models of production planning problems, in: K.G. Kempf, P. Keskinocak, R. Uzsoy (Eds.), Planning Production and Inventories in the Extended Enterprise: A State of the Art Handbook, Volume 1, in: International Series in Operations Research and Management Science, vol. 151, Springer, New York, NY, 2011, pp. 437–507, chap. 16.

[18] A. Cornish-Bowden, Fundamentals of Enzyme Kinetics, 4th ed., Wiley-Blackwell, Weinheim, 2012.

[19] H.E. Daly, J. Farley, Ecological Economics, 2nd ed., Island Press, Washington, DC, 2010.
[20] R. Bobulescu, From Lotka’s biophysics to Georgescu-Roegen’s bioeconomics, Ecol. Econ. 120 (2015) 194–202.

[21] N. Georgescu-Roegen, Analytical Economics: Issues and Problems, Harvard University Press, Cambridge, MA, 1966.

[22] N. Georgescu-Roegen, On the case of catalytic labor, Int. Econ. Rev. 11 (2) (1970) 315–317.

[23] G.I. Kolesova, I.A. Poletaev, Nekotorye voprosy issledovaniia sistem s limitiruiushchimi faktorami (Selected problems in research of the systems with limiting factors), in: Upravliaemye sistemy (Controllable Systems), vol. 3, Institute of Mathematics, Siberian Branch of the USSR Academy of Sciences, Novosibirsk, 1969, pp. 71–80, http://odasib.ru/openarchive/Portrait.cshtml?id=Xu_kray_634993802507080078_1257 (In Russian).

[24] Yu.I. Gilderman, K.N. Kudrina, I.A. Poletaev, Modeli L-sistem (sistemy s limitiruiushchimi faktorami) (Models of L-systems (systems with limiting factors)), in: A.A. Liapunov (Ed.), Issledovaniia po kibernetike: Sbornik statei (Studies in Cybernetics: Collected Papers), Sovetskoe Radio, Moscow, 1970, pp. 165–210, http://odasib.ru/openarchive/Portrait.cshtml?id=Xu_kray_634993802507080078_1470 (In Russian).

[25] D.S. Chernavskii, E.N. Frolov, V.I. Goldanskii, A.A. Kononenko, A.B. Rubin, Electron tunneling process and the segment mobility of macromolecules, Proc. Natl. Acad. Sci. USA 77 (12) (1980) 7218–7221.

[26] D.S. Chernavskii, Yu.I. Khurgin, S.E. Shnol’, Kontseptsiiia “belok-mashina” i ee sledstviia (The “protein-machine” concept and its consequences), Biofizika 32 (5) (1987) 775–781 (In Russian).

[27] Yu.M. Romanovskii, N.M. Stepanova, D.S. Chernavskii, Chto takoe matematicheskia biofizika: kineticheskie modeli v biofizike (What is Mathematical Biophysics: Kinetic Models in Biophysics), Prosveshchenie, Moscow, 1971 (In Russian).

[28] R.R. Nelson, S.G. Winter, An Evolutionary Theory of Economic Change, Belknap Press of Harvard University Press, Cambridge, MA, 1982.

[29] R.U. Ayres, U.E. Simonis (Eds.), Industrial Metabolism: Restructuring for Sustainable Development, United Nations University Press, Tokyo, 1994.

[30] H. Haken, Synergetics: An Introduction. Nonequilibrium Phase Transitions and Self-organization in Physics, Chemistry, and Biology, 3rd ed., Springer Series in Synergetics, vol. 1, Springer, New York, NY, 1983.
[31] A.T. Mustafin, A Bottleneck Principle for Techno-Metabolic Chains, Discussion paper 504, Institute of Economic Research, Kyoto University, Kyoto, Japan, 1999.

[32] A. Mustafin, K printsipam promyshlennogo metabolizma (A contribution to the principles of industrial metabolism), AlPari (4–5) (1999) 80–83, https://www.researchgate.net/publication/319109711 (In Russian).

[33] E. Levine, T. Hwa, Stochastic fluctuations in metabolic pathways, Proc. Natl. Acad. Sci. 104 (22) (2007) 9224–9229.

[34] C. Kandemir-Cavas, L. Cavas, M. Baki Yokes, M. Hlynka, R. Schell, K. Yurdakoc, A novel application of queueing theory on the Caulerpenyne secreted by invasive Caulerpa taxifolia (Vahl) C. Agardh (Ulvophyceae, Caulerpales): a preliminary study, Mediterr. Mar. Sci. 9 (1) (2008) 67–76.

[35] P. Hochendoner, C. Ogle, W.H. Mather, A queueing approach to multi-site enzyme kinetics, Interface Focus 4 (2014) 20130077.

[36] D. Helbing, D. Armbruster, A.S. Mikhailov, E. Lefeber, Information and material flows in complex networks, Physica A 363 (1) (2006) xi–xvi.

[37] R. Gesztelyi, J. Zsuga, A. Kemeny-Beke, B. Varga, B. Juhasz, A. Tosaki, The Hill equation and the origin of quantitative pharmacology, Arch. Hist. Exact Sci. 66 (4) (2012) 427–438.

[38] J. Monod, The growth of bacterial cultures, Annu. Rev. Microbiol. 3 (1) (1949) 371–394.

[39] C.S. Holling, The functional response of predators to prey density and its role in mimicry and population regulation, Mem. Entomol. Soc. Can. 97 (S45) (1965) 5–60.

[40] S.E. Elmaghraby, Production capacity: its bases, functions and measurement, in: K.G. Kempf, P. Keskinocak, R. Uzsoy (Eds.), Planning Production and Inventories in the Extended Enterprise: A State of the Art Handbook, Volume 1, in: International Series in Operations Research and Management Science, vol. 151, Springer, New York, NY, 2011, pp. 119–166, chap. 7.

[41] C. Kuehn, Multiple Time Scale Dynamics, Applied Mathematical Sciences, vol. 191, Springer, New York, NY, 2014.

[42] W. Klonowski, Simplifying principles for chemical and enzyme reaction kinetics, Biophys. Chem. 18 (2) (1983) 73–87.

[43] L.A. Segel, M. Slemrod, The quasi-steady-state assumption: a case study in perturbation, SIAM Rev. 31 (3) (1989) 446–477.
[44] D. Armbruster, The production planning problem: clearing functions, variable lead times, delay equations and partial differential equations, in: D. Armbruster, K.G. Kempf (Eds.), Decision Policies for Production Networks, Springer, London, 2012, pp. 289–302.

[45] D.J. Higham, Modeling and simulating chemical reactions, SIAM Rev. 50 (2) (2008) 347–368.