Advances in Economics, Business and Management Research, volume 147
Proceedings of the International Conference on Policies and Economics Measures for Agricultural Development (AgroDevEco 2020)

Development of Information Model of National Economy Turnover Using Big Data Technology

Pshenichnikov V.V.
Department of Finance and Credit
Voronezh State Agrarian University named after Emperor Peter the Great
Voronezh, Russia
e-mail: wladwp@yandex.ru

***Abstract* — The article substantiates the concept of an information model of the payment turnover of the country's economy using big data technology. The information model is designed to reflect the economic situation of payment entities, industries, regions, and the country as a whole on the basis of cash transactions in cash, non-cash and electronic forms. The basis for developing an information model of payment turnover is a single classifier of payment destination. It is proposed to adapt the unified classifier of payment assignment used in the Republic of Kazakhstan to the system of indicators of financial statistics of the Russian Federation.**

Keywords* — *big data, money, information model, payment turnover.

I. Introduction

Modern payment turnover is a continuous movement of the three main forms of money – cash, non-cash and electronic. At the present the system of statistical accounting of cash and non-cash money flows is functioning quite well. The emergence and spread of electronic money in the conditions of the formation of the digital economy complicated the process of official statistical accounting of the volume of emission and circulation of money outside the banking sector.

The purpose of the study is to substantiate the concept of an information model that characterizes the economic situation of the subjects of payments, industries, regions and the country as a whole on the basis of ongoing cash transactions in cash, non-cash and electronic forms.

II. Research Methodology

In the Russian practice an attempt has already been made to create a similar model in the form of a regional matrix balance of payment turnover grouped by the economic content of payments. It was compiled in 1974 in the Estonian SSR by the Laboratory of Financial Planning Models of the Central Economics and Mathematics Institute of the USSR Academy of Sciences and the Main Computing Center of the USSR State Bank. The matrix balance of the payment turnover encompassed an array of information from all 18 institutions of the State Bank of the Estonian SSR for the period from November 18 to 22, 1974. It reflected the state of foreign economic, inter-industry and intra-industry settlement-payment and credit relations for the studied period of time. The review of payment transactions carried out in this way showed the advantages of the matrix method of accounting for money transactions in the economy [4]. With the transition from a planned economy to a market-based one the method of matrix balance constructing has not lost its relevance and can be successfully used in modern conditions based on the technology of big data processing.

Big data is a collection of different approaches, methods and tools, processing not only structured, but also unstructured data to achieve specific goals and objectives. Big data allows the financial and banking sector to provide its services through any Internet access point, providing customers with individual targeted offers. Due to the massive personalization of the customer base, which is analyzed with the help of big data, external systems of banking and financial institutions monitor the location of the client in relation to nearby retailers, and internal – analyze earlier requests and purchases made by the client to find and show him available in his location proposals that best fit his needs and preferences. Analysis of big data allows the computer to see certain hidden patterns that a person can not observe. This creates opportunities to optimize such spheres of public life as public administration, telecommunications, production, distribution and consumption of various goods. The key sources of big data are the Internet (blogs, media, social networks, forums, and other sites); document archives of various enterprises and organizations; reading of various kinds of tools.

The development of an information model of the country economy turnover at the macro level should be based, in our opinion, on a transactional approach of measuring the amount of money serving as a medium of circulation and a means of payment. In the framework of this approach it is believed that the availability of money is an essential condition for the exchange of goods by subjects of economic relations. In this regard the definition of active money supply stands out – cash in circulation used for real maintenance of economic relations. An active money supply is understood to be the aggregate of all legal means of payment within a given jurisdiction, presented in cash, non-cash and electronic form.

III. The Research Results

The proposed prototype of the information model of the country economy turnover at the macro level (Table 1) illustrates the consolidated matrix balance of the payment turnover of the country, which includes regional matrix balances of payment turnover, for example, for each subject of the Russian Federation, compiled in terms of cash, non-cash and electronic money . Separate data sets for the formation of such balances are collected and processed by the territorial institutions of the Bank of Russia.

TABLE I. THE PROPOSED PROTOTYPE OF THE INFORMATION MODEL OF THE COUNTRY ECONOMY TURNOVER AT THE MACRO LEVEL

Money Inflow	**Money Outflow**							
	Outside world	***Central Bank***	***Government bodies***	***Financial sector***	***Sectors of the economy of the non-financial sector***	***Households***	***Total***	***The total amount of excess of the outflow of money over their inflow***
Outside world								
Central Bank								
Government bodies								
Financial sector								
Sectors of the economy of the non-financial sector								
Households								
Total								
The total amount of excess of the inflow of money over their outflow								

The prototype of the information model in its most general form describes the settlement and payment relations between the following sectors of the payment turnover of the economy of the country: the outside world, the Central Bank, government bodies, the financial sector, sectors of the economy of the non-financial sector, households. The inflow of money into the indicated sectors of the payment turnover is reflected in the rows of the table, and the outflow of money is reflected in the columns of the table. The total amount of excess of the inflow of money over their outflow, or the outflow of money over their inflow in each sector of the payment turnover is reflected respectively in the bottom row or in the right-hand column of the table.

The basis for the development of the information model of the payment transaction is a unified classifier for the purpose of payments. In Russia such a unified classifier is not yet available, but local value classifiers are in effect, for example, “Classifier of types of taxes, fees and other fees charged by customs authorities” adopted by the Decision of the Customs Union Commission No. 378 dated September 20, 2010 (ed. from 30.10.2018) “On classifiers used to fill in customs documents” [8]. However, one of the Russian partners in the Customs Union, the Republic of Khazakhstan, has already gained some experience in using such a classifier approved by the Board of the National Bank of the Republic of Khazakhstan No. 388 dated November 15, 1999 “Rules for the application of the State Classifier of the Republic of Khazakhstan – a unified classifier for making payments” with amendments and additions dated August 31, 2016 No. 203, enacted from 01.01.2017 [10].

The unified classifier for the purpose of payments is a system of digital and alphabetic characters that allows identifying the payer, the payee and the purpose of the payment. This system should be included as a part of the details of settlement and payment documents in order to ensure the transparency of payments, systematize information about them, harmonize and simplify the procedure for paper and electronic payment documents. This system is formed from additional details of the payment document and the payment purpose code. The structure of a unified classifier for the purpose of payments is as follows: I II III IV VI VII VII IX X, where: I is a sign of the residence of the sender of money; II – the economy sector of the sender of money; III – a sign of residence of the beneficiary; IV – the economy sector of the beneficiary; V VI VII – the code of currency and precious metals; VIII – type of operation; IX – the nature of the payment; X – payment details.

Symbols I and II are formed in payment documents in the code field of the money sender. The sign of residence is determined in accordance with the currency legislation of the Republic of Khazakhstan and is affixed in the following order: “1” – resident and “2” – non-resident. Codes of sectors of the economy are reflected respectively in the code field of the sender of money, and in the code of the beneficiary. In the Republic of Khazakhstan, the following sector codes are used: A International organizations; 1 Central Government; 2 Regional and local governments; 3 Central (national) banks; 4 Other depository organizations; 5 Other financial institutions; 6 State non-financial organizations; 7 Non-governmental non-financial organizations; 8 Non-profit organizations serving households; 9 Households.

Symbols V VI VII are designated in accordance with the state classifier of the Republic of Khazakhstan – "Codes for the designation of currencies and funds", which was created in the framework of ISO 4217.

Symbols VIII, IX, X form the payment purpose code. The payment purpose code is affixed in accordance with the detailed table of payment purpose codes, which provides for 10 categories of payment purpose codes that determine the type of transaction: 0-Pension payments and benefits; 1 – Specific translations; 2 – Operations with foreign currency and precious metals; 3 – Deposits; 4 – Loans; 5 – Securities, promissory notes and deposit certificates issued by non-residents of the Republic of Khazakhstan and investments in foreign capital; 6 – Securities and promissory notes issued by residents of the Republic of Khazakhstan and investments in Khazakhstan's capital; 7 – Goods and intangible assets; 8 – Services; 9 – Payments to the budget and payments from the budget.

Within each category there are groups of payment purpose codes that establish the nature of the payment, in each of which different variants of codes are presented. When determining the payment purpose code, we must first determine which category the payment belongs to, then select the group of payment purpose codes and determine which of the options in the selected group the purpose of the payment is. When performing a transaction that has several payment settings, the purpose of the base payment is entered in the payment document. Putting more than one payment purpose code in payment documents is not allowed. It is not allowed affixing the payment purpose code, which is the title of the type of operation or the nature of the payment, if there is its further detail.

The unified classifier for the purpose of payments used in the Republic of Khazakhstan can be adapted to the system of indicators of financial statistics of the Russian Federation without special problems, since both countries, firstly, historically had a single system of statistical accounting within the same state and, secondly, at present, they are guided by a common methodological framework “The Guide to Monetary and Financial Statistics”, published by the International Monetary Fund in 2000 [2]. Moreover, the Russian Federation and the Republic of Khazakhstan could combine efforts to introduce a unified classifier for all participants of the Eurasian Economic Union in the future, also in order to apply the concept of an information model of payment turnover on a scale of the entire Eurasian Economic Union.

IV. TECHNOLOGICAL ASPECTS

The technology of using electronic money based on the transfer of information, as shown earlier requires encoding [9]. Encoding of information messages about the amount of electronic money to be transferred from the payer to the payee can be carried out either by numbers or by letters of the alphabet. The second option seems to be more preferable to us, at least because a mistake in at least one digit, when encoding the amount of payment, can dramatically distort this sum when decoding information, whereas an error, in at least one letter of the encoded numeral can lead only to a grammatical error in the designation of the amount of payment when decoding information about it. In addition, the encoded information message may contain not only the payment amount, but also its other requisites, in particular, the reason and purpose of the payment. Thus, mathematical parameters derived by B. Mandelbrot [5] at the intersection of information theory of C. Shannon [11] and psycholinguistics G. Zipf [12], as the quantitative characteristics of the turnover of electronic money, can be used.

Applying linguistic terms, it will be necessary to distinguish between the concepts: “language” as a system of rules; “discourse” as a sequence of written or spoken signs. Among the technical problems connected with discourse, the most pressing problems are initially cryptography (“cryptos” – “secret”, “hidden”), stenography (“stenos” – “compressed”) and telegraphy (“telos” – “far”) problems. However, these problems are solved within the framework of information theory. The cryptographer searches for a code, possibly devoid of structure, since any structure can help the enemy unravel its secret. The stenographer and the telegraph operator, each in their own way, search for the shortest possible code.

In general, knowledge of the statistics of discourse indicates how to replace frequently encountered speech clichés with single shorter signs. As a result, shorter transcripts are obtained as clichés help to unravel the secret, more "stable" cryptograms. The frequency of letters, as well as the frequency of *n*-grammes (sequences of *n* letters), was used by cryptographers and telegraphists from the very beginning. A few decades before C. Shannon, Samuel Morse realized that it would be nice to compare the shortest combinations of dots and dashes to the most frequently encountered letters, while a cryptographer for centuries rely on rough estimates of the frequencies mentioned above. However, in this sense, an insurmountable limit is reached very quickly, since individual letters appear to be too small to cause any own interest; knowledge of the frequency of their use says almost nothing about the frequency of using their large-sized associations. As for the *n*-gram, they are more useful, but too artificial from the linguistic point of view.

Words are more promising, because they are an integral part of the discourse, and also because the combined sequence of letters in words, separated by gaps, leads to a grandiose and global simplification. If we take a long sample consisting of k words taken from different information messages of one payer, then we can arrange all the words appearing there in descending order of their frequencies. The word that occurs most often gets rank 1, the word of rank 2 is the next most frequently used. We denote by $W(r)$ the word that has rank r in the given classification. We continue this procedure until the text is finished, and the word $W(r)$ will occur to us $i(r,k)$ times. For frequently recurring words such a classification will be unambiguous; there is an increasing ambiguity for words that occur 3, 2 or 1 times, but from our point of view it is not important. Thus, the frequency of the use of the word $W(r)$ will be determined as follows:

$$W\ (r) = \frac{i\ (r,\ k)}{k}\ . \qquad (1)$$

In order to enlarge the characteristics of the modeled process two approximations were made by B. Mandelbrot [7]:

The first approximation is Zipf's law;

The second approximation is the Zipf-Mandelbrot law.

The first approximation made it possible to establish the following fact: the definition of rank assumes that the frequency $i(r,k)/k$ varies in the direction opposite to the increase in r, but this definition does not in any way imply that it will vary inversely proportional to r, it is purely empirical discovery. A common method for establishing and confirming these relationships is to go to the logarithmic coordinates, where horizontal is the logarithm of the rank r, and the vertical is the logarithm of the frequency $i(r,k)/k$. The first approximation allows assuming that the diagram obtained this way will be a straight line with a slope of $n - 1$, parallel to the second bisector of the coordinate axes.

The second approximation made the following conclusions.

First, some of the most frequent words (small values of r) apparently do not obey any precise law, but in general they get lower than it would be by the law of first approximation. Secondly, if we omit the most frequent words, then the “rank-frequency” graph remains straight, but in general it is not true that this straight line is parallel to the second bisectrix of the coordinate axes, most often its slope is greater than 1; and the closer to 1 the “richer” a vocabulary is.

Combining these two remarks, B. Mandelbrot [6] came to the following formula:

$$i\,(r,k) = Pk(r+V)^{-B}, \quad (2)$$

where *P*, *V*, and *B* denote parameters that in the first approximation are reduced to *P* = 1/10, *V* = 0 and *B* = 1. The number 10 in this case has nothing to do with the decimal number system and is purely empirical in nature. *B* is the slope of the "rank-frequency" curve.

In the theory of information, it is a classic fact that if we equip a cryptographer and telegraph operator with one alphabet and make them encode word by word (separating words with gaps), then the optimal solutions will be the same for both, or rather each will receive a set of equivalent optimal solutions, and these two sets will be identical.

It is also known that for a system that is arbitrary in word frequency, the requirement to separate the codes with gaps in general case becomes a source of either structures that can help the enemy decrypt, or inefficiency in the use of letters. As it turned out, this system satisfies the second approximation. In other words, the frequency of the use of words is the only thing that needs to be adapted to the separation of words by means of gaps.

The use of extreme criteria, when it is required that a particular quantity be arbitrarily large or arbitrarily small, is widely known in physics: for example, the "least action principle", the "maximum entropy principle", and the "minimum entropy production principle". These criteria originate from the XVIII century and are based on the conscious transfer of various introspective "principles" of the least effort that were supposed to be inherent in human behavior.

"Entropy is the minimum of information that must be obtained in order to eliminate the uncertainty of the alphabet used by the source of information" [3]. The concept of entropy was introduced into science by the German physicist and mathematician R. Clausius in 1865 as a logical development of the thermodynamics of S. Carnot (the French physicist and mathematician), formulated, in turn, even before the discovery of the law of conservation of energy and based on the thermal hypothesis [1]. Entropy was defined as the ratio of the amount of heat in the body to its temperature and was postulated on the basis of experience that entropy never decreases, but, as a rule, increases. The main physical confirmation of this is the irrefutable fact that heat never passes from a cold body to a hot one. Further development of physics led to the appearance of statistical thermodynamics, which was based on the formula that relates entropy to the state of the system (the Boltzmann-Planck entropy):

$$S = k \ln P, \quad (3)$$

where: *S* – entropy of the system; *k* is the Boltzmann constant; *P* is the thermodynamic probability of a state. Thus, the second law of thermodynamics has become one of the fundamental laws of modern physics.

The introduction of the concept of entropy proved to be very fruitful for the development of the theory of information transfer, thanks to the striking similarity in the formal expression between the entropy of the probability of states (Boltzmann-Planck) and the information entropy of C. Shannon.

The discrete probability distribution is a set composed of positive numbers p_n, where $\sum p_n = 1$. Here the index *n* ranges from 1 to N, and N can be infinitely large. Then the amount of information on C. Shannon is determined as the sum:

$$H = -\sum p_n \log_x p_n, \quad (4)$$

where: *H* is the lower limit of the number of characters required to transmit a message; p_n is the probability of a lexical word; *x* is the number of characters used for encoding (for example, if a binary coding system consisting of two digits ("0" and "1") is used, then *x* = 2, if an alphabet consisting of 32 letters is used for text messages, then *x* = 32).

Thus, we have the opportunity to scientifically justify the magnitude of the flow of electronic money in a single transaction, which opens up opportunities for us to further search for quantitative parameters of the functioning of electronic money.

V. Conclusions

The information model of the country economy turnover at the macro level is able to illustrate the volume of money emission and turnover of all applied forms of money (cash, non-cash and electronic) servicing relations with the outside world, the Central Bank, government bodies, the financial sector, the non-financial sector, households. The prototype of the information model of the payment turnover of the economy of the country at the macro level is presented by us in a consolidated version. It can be deployed in the directions of information differentiation about individual sectors of the economy in the non-financial sector, expanding the list of operations in the context of certain types of organizations in the financial sector and the Central Bank, detailing payment relations with government bodies and the outside world.

A detailed information model will provide more extensive amounts of information about the payment turnover in the economy of the country. This is especially important for the operational (for example, monthly) identification of disproportions arising in relations between the subjects of payments. Moreover, this model can be useful for compiling forecasting options for the development of payment transactions in the framework of the implementation of national projects and federal targeted programs. In such cases all those indicators of the model that will be affected by the cash flows that ensure the implementation of a particular national project or of a particular federal targeted program will be subject to recalculation.

We see directions for further research in adapting the unified classifier of payment purposes used in the Republic of Khazakhstan to the system of indicators of financial statistics of the Russian Federation. Adapted to the conditions of the Russian Federation, a unified classifier for the purpose of payments can already be used to compile a detailed information model of the payment turnover of the economy of the country at the macro level.

References

[1] W.R. Ashby, An Introduction to Cybernetics. London: Chapman & Hall, 1956.

[2] V. Bartosova, V.V. Pshenichnikov, G. Ievtushenko, L. Petrova, P. Ovchar, "Accounting of transactions in electronic money: International trends", Acad. of Account. and Financial Studies J., vol. 23, spec. iss. 2, 2019.

[3] A.I. Dyomin, Information theory of economics: Macromodel, 3rd ed. Moscow: KomKniga, 2010.

[4] A.M. Kosoy, Payment Transactions: a Study and Recommendations. Moscow: Finansy i statistika, 2005.

[5] B.B. Mandelbrot, The Fractal Geometry of Nature. New York: Freeman, 1982.

[6] B.B. Mandelbrot, Fractal and Scaling in Finance: Discontinuity, Concentration, Risk. New York: Springer-Verlag, 1997.

[7] B.B. Mandelbrot, Fractal & Self-affinity: R/S, 1/f, Global Dependence, Relief & Rivers. New York: Springer-Verlag, 1997.

[8] On classifiers used to fill in customs documents 2018, Decision of the customs Union Commission of 20.09.2010 no. 378 (ed. from 30.10.2018). Retrieved from: http://www.consultant.ru/document/cons_doc_LAW_105110/.

[9] V.V. Pshenichnikov, A.V. Babkin, "Digital money as a product of the development of information and telecommunication technologies", pp. 266–272, 2017 [Proc. of the 2017 Int. Conf. Quality Management, Transport and Information Security, Information Technologies, IT and QM and IS 2017]. St. Petersburg.

[10] Rules for the application of the State Classifier of the Republic of Khazakhstan – a unified classifier for making payments 2016 Approved by resolution of the Board of the National Bank of the Republic of Kazakhstan dated 15.11.1999 no. 388 with amendments dated 31.08.2016 no. 203 Retrieved from: http://adilet.zan.kz/rus/docs/V000001319_

[11] C.E. Shannon, "A Mathematical Theory of Communication", Bell. Syst. Techn. J., vol. 27. pp. 379–423, 623–656, 1948.

[12] G.K. Zipf, Human Behavior and the Principle of Least Effort. Addison-Wesley Press, 1949.