How the Diffusion of Money may Direct the Distribution of Income

Kryssanov, V.V., Ogawa, H.
College of Information Science and Engineering, Ritsumeikan University, Kusatsu, Japan
Buryakov, I.
TOSEI Corp., Tokyo, Japan

Abstract: In the presented study, an attempt was made to explain the mechanism of income distribution, using the metaphor of the physical diffusion process. The corresponding generating mechanism is defined in terms of the real GDP, the average price level, and uniformity characteristics of the economy. The standard exponential distribution is derived as the model of income distribution in the simplest case, and it is shown how this model can be extended to accommodate to the realities of income analysis. A pilot study of the application of the proposed modeling framework to the JTA tax return data for 1970-2001 is described, and its results are briefly discussed. Finally, some conclusions are drawn at the paper's end.

Keywords: income distribution, exponential probability distribution, Laplace-Stieltjes transform

1. Introduction

Since the publication of Pareto's celebrated work on the distribution of income more than a century ago (Pareto, 1897), statistical properties of income remained in the focus of a number of theoretical and empirical studies reported in the specialized literature. Until the mid 70^{th} of the last century, the Pareto and the lognormal (Gibrat, 1931) models were seldom challenged and were consequently used as a foundation of various macro-economic models. The advent of digital data processing technologies and the following dramatic increase in the size of data samples available for statistical analysis has, however, revealed that: 1) the simple classic models seldom, if ever, satisfactory approximate empirical income distributions, and 2) no single simple model is capable of theoretically explaining various distributions of income in economies observed at different times and in different parts of the world. The research community has then been quick to respond with models and flexible approximating forms of ever increasing complexity offered to achieve adequate fits to actual data.

Miscellaneous extensions and combinations of the Pareto and lognormal models have been proposed (e.g., Reed, 2001; Souma, 2002) to accurately replicate the behavior of income data. At the same time, new models, such as beta (Thurow, 1970), Gamma (Salem and Mount, 1974), and Singh and Maddala (Burr) (1976) probability distributions, have been proposed and demonstrated a better precision in predicting the empirical statistics. As a step to finding a reasonably universal model, known income distributions have been suggested to generalize with appropriate distribution families, such as generalized beta and generalized beta of the second kind (e.g. McDonald, 1984), and the corresponding theoretical justifications have been proposed (Parker, 1999).

In spite of a significant progress made in reproducing the income data behavior with distribution forms having 3 to 5 free parameters, the problem of income distribution modeling can hardly be considered solved. While the early simple models have often been criticized for their ad-hoc nature and lack of micro-foundations, more advanced models typically have their parameters defined in terms of some behavioral or decision-making patterns that assumingly persist (among the individuals or organizations) in the economy (e.g. Parker, 1999). Such patterns are, as a rule, not directly observable, and their actual dynamics are unknown. This makes quite speculative attempts to ascribe an indicative meaning to parameters of the income distribution models that otherwise cannot be verified with methods other than simulation. In view of the latter, even models accurately mimicking the empirical statistics may appear ad hoc curve-interpolating forms, unless explicit links of the model parameters with some observable (macro-) characteristics of the economy are drawn and can be verified with data other than that of income distribution.

In this paper, we present a very general class of income distribution models, which results from the consideration of various directing processes responsible for the distribution of money over different sectors of the economy. In the next section, we first show that the standard exponential distribution may often be a good first-approximation model for the statistics of income. With the support of a simple experiment, it is then demonstrated in Section 3 that even the simple exponential model can provide meaningful results when analyzing income inequality in "equal societies," such as in the case of Japan. Section 4 discusses connections between the proposed income generating mechanism and some macroeconomic processes and gives possible generalizations of the exponential model. Section 5 concludes the paper.

2. The diffusion of money

In the following, we will assume that the income in an industry is proportional to the money circulating in this industry. Let $S_0(t)$ be the income in some "0-th" industry (or sector of the economy), and $M_0(t)$ be the total money in this industry; our assumption is then explicitly formulated as:

$$S_0(t) = k\,M_0(t)\,, \tag{1}$$

where coefficient $0 < k < 1$. For the sake of simplicity, we will consider $k = const$. In many cases, it is convenient to assume that the 0-th industry is the observed one, i.e. the industry for which the income is analyzed.

The dynamics of $S_0(t)$ in a society is determined by many factors, such as current economic situation, investment climate, legislation, and the like. Some of these factors can explicitly be accounted for by setting a specific value of k, while many other factors will implicitly shape up the competition among different industries in the economy for a share in the economy's total money supply M. The latter process for the 0-th industry can be specified as follows:

$$\frac{dM_0}{dt} = a_0 \frac{dM}{dt} + \eta_0\,, \tag{2}$$

where coefficient $a_0(t) > 0$, and $\eta_0(t)$ is the stochastic (Gaussian) noise with zero mean.

The total money supply M depends on the monetary base, the currency-deposit ratio, and the reserve-deposit ratio. In a situation when the government does not print money, the financial intermediation process has a constrained outcome and, therefore, there exists a theoretical maximum of money circulating in the economy. Under this maximum, the monetary base reaches its highest value, while the currency-deposit and reserve-deposit ratios fall to their lowest rates. Let us denote L the maximum quantity of money in the economy – the M2 monetary aggregates that include assets such as saving accounts in addition to coins, paper currency, and checkable deposits. The temporary dynamics of M can be estimated as a difference in the dynamics of L and $\sum_{j=0}^{N} M_j$, where M_j stands for the money circulating in the *j*-th industry, and N is the number of industries competing for the money. Equation (3) can then be generalized as follows:

$$\frac{dM_i}{dt} = a_i \left(vL - \sum_{j=0}^{N} M_j \right) + \eta_i\,, \tag{3}$$

where v is the velocity of money, $i = 0,1,...,N$, and a_i and η_i are analogs of a_0 and η_0, respectively, in Equation (2).

Equations (3) describe a diffusion process on a hyperplane $\sum_{j=0}^{N} M_j = vL$ formed by $N+1$ monetary subsystems M_j, each of which corresponds to an industry with different production factors. Due to the hyperplane condition, there can be only N mutually independent subsystems, say $M_1,...,M_N$; for a free market economy, it appears natural to assume that values of $M_1,...,M_N$ are all uniformly distributed on the interval $[0, vL]$. Taking into account condition (1), the hyperplane equation can be re-written as $\sum_{j=0}^{N} S_j = \sum_{j=0}^{N} k M_j = k v L$. The probability that $S_n \geq S_0$, S_n is the income in the *n*-th industry, $n = 1,...,N$, can be calculated as a product of the marginal probability density functions $f_n(S_n) = \frac{1}{k v L}$ that yields $\Pr[S_1 \geq S_0,...,S_N \geq S_0] = \left(1 - \frac{S_0}{k v L}\right)^N$. Probability theory defines the cumulative distribution function (CDF) for some x_0 taken from $X = \{x_0, x_1,..., x_N\}$, the set of all random variables that obey a given probabilistic law as $F(x_0) = \Pr[x_1 < x_0,..., x_N < x_0]$. In this context and assuming that $N >> 1$, we arrive at a probability distribution function of S_0 defined as follows:

$$F(S_0) \propto 1 - \left(1 - \frac{S_0}{k v L}\right)^N \approx 1 - \mathrm{e}^{-\frac{S_0}{\beta}}, \qquad (4)$$

where parameter $\beta > 0$ is, up to the factor $k v$, determined by the total available money L averaged over the $N+1$ industries: $\beta = \frac{k v L}{N+1}$. Using the quantity theory of money, the latter equation can be re-formulated as follows:

$$\beta = \frac{k p q}{N+1}, \qquad (5)$$

where p is the average price level, and q is the real GDP in the economy.

3. Experiment

The study of income distribution was always hampered by the quality of empirical data. Among the typical questions about the sample "nosiness" that need to be addressed when selecting data for validation of an income distribution model, we would point to the following few. Income of what economic unit – an individual or a family – should be considered and for what period of time, how to account for the demographic dynamics affecting the earnings, and what source of data should be used.

Coarsely tabulated personal income data for the period from 1970 to 2001 fiscal year obtained from the Japanese Tax Administration (JTA) was used in the presented experiment. The tabulation is as follows: income less than 0.7 million yen per year, 0.7-1.0 (more than 0.7 but less than 1.0 million yen), 1.0-1.5, 1.5-2.0, 2.0-2.5, 2.5-3.0, 3.0-4.0, 4.0-5.0, 5.0-6.0, 6.0-7.0, 7.0-8.0, 8.0-10.0, 10.0-12.0, 12.0-15.0, 15.0-20.0, 20.0-30.0, 30.0-50.0, and income grater than or equal to 50.0 million yen. The data characterizes individuals who filed tax returns and does not include all the individuals with income, as many of them might not have to submit returns according to the Japanese tax law (or did not do so, violating the law). As it was typical for a Japanese family in the specified period of time to have only one member employed full-time, the personal income distribution may not principally be different from the family income distribution in the considered case. Besides, as the demographic situation in Japan did not undergo major changes in the inspected period of time, it may be assumed that the level of the "demographic noise" in the data-sample is approximately the same for every year. It seems reasonable to also assume, on cultural grounds, that even though some of the returns do not reflect the actual situation with earned income, the percentage of the deception remained constant every year.

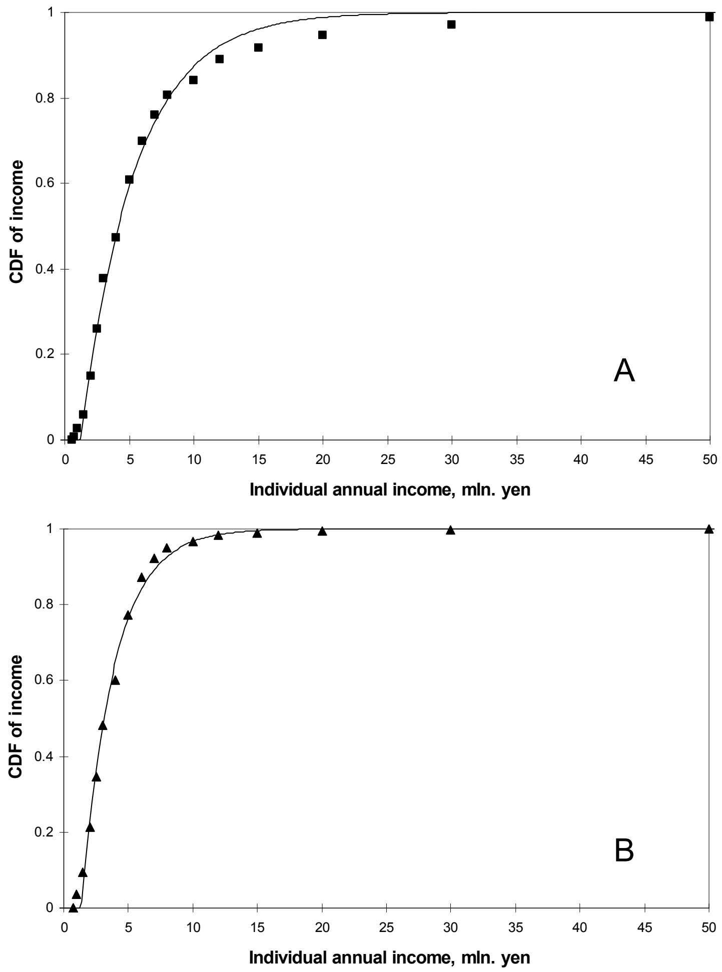

Figure 1. Modeling the distribution of income in Japan in 2000 with the exponential distribution model. Solid lines show the MSE model fit to the aggregated data from all sectors of the economy (A, empirical values represented with filled boxes), and to the data from one business-sector (B, represented with filled triangles)

Figure 1 (A) shows a typical, for the given data collection, result of fitting the simple exponential model (4) to the tax return data. The obtained fit ($\beta \approx 4.247$) is rather poor, especially in the right-tail area of the distribution: obviously, the observed income distribution has a heavier tail than the one of the standard exponential. This result should not, however, appear surprising because the considered data is not homogeneous in respect to the industries or economy sectors, but the model assumptions require the homogeneity condition to hold. Given the structure of the available data, it generally appears difficult to obtain a sample that would reflect income in a single industry. At the same time, it was observed from the data that the empirical distribution becomes much closer to the exponential, as we "narrow down" the categories of the earners – see Figure 1 (B) for a typical result ($\beta \approx 2.543$). For 1970-2001, JTA reported aggregated returns along with returns in four economy sectors: business, farming, "operating," and "other."

One possible way to improve the fit would be to extend the model by considering, for example, a finite mixture of exponentials with a justification that one exponential should stand for each homogeneous (in terms of the model parameters) sector of the economy. Another way would be to consider β as a random variable itself and fit the data to various infinite mixtures of exponentials. In both cases, however, more detailed knowledge about the tax return data and the Japanese economy in the inspected period (e.g. the "true" structure of the economy or the distribution function of β) would be required to validate the models.

By looking into the properties of model (4), one would notice that the exponential distribution assumes an income variability – "inequality" – growing as the square of the average income (for the exponential distribution, β determines the mean, and β^2 – the variance). Figure 2 displays the "empirical inequality," defined as a difference between the average income and the most probable (i.e. frequent) income, as a function of the average income squared calculated from the available data collection (note that the Gini coefficient is a constant 0.5 for the exponential distribution and may not thus be a good indicator of inequality in our case).

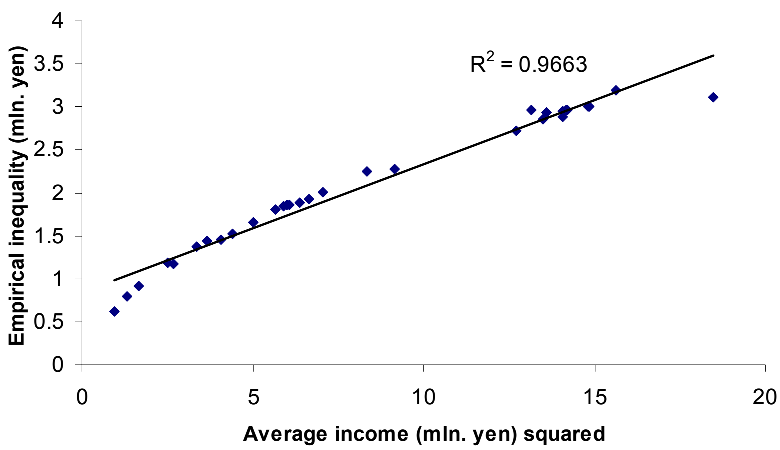

Figure 2. The income inequality as a function of the average income squared calculated from the JTA data (for the "aggregated income" category) for 1970-2001. The filled diamonds represent the empirical results, and the solid line shows a linear MSE fit

Leaving aside the discussion of whether the Japanese society is or was "equal" (see, for example, Fukawa, 2006), results presented in Figure 2 support our hypothesis inferred from equation (4) that an average income growth causes the quadratic increase in social disparity (the latter assertion, of course, needs further refinement and verification with different data). This, together with the results of Figure 1, A and B, supports our claim that even in its simplest form, the proposed model of income distribution can be used as a loose approximation model to analyze empirical data.

4. Discussion

Equation (5) derived in Section 2 suggests that under other similar conditions, a higher real GDP would result in a higher average income. On the other hand, such macroeconomic phenomena as deflation and industry diversification may lead to a decrease in the average income. It then appears interesting to explore how the corresponding parameters of the Japanese economy affected the average income in the considered period of time.

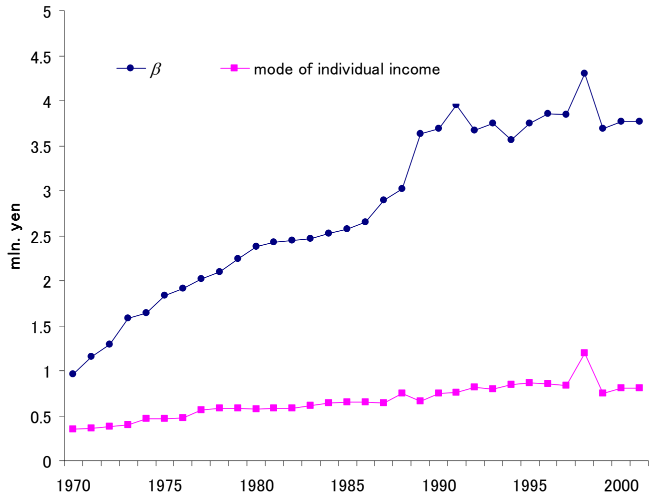

Figure 3. The dynamics of earnings in Japan calculated from the JTA personal income data. (The sharp rise of average income in 1989 is due to changes in the tabulation ranges used by the JTA)

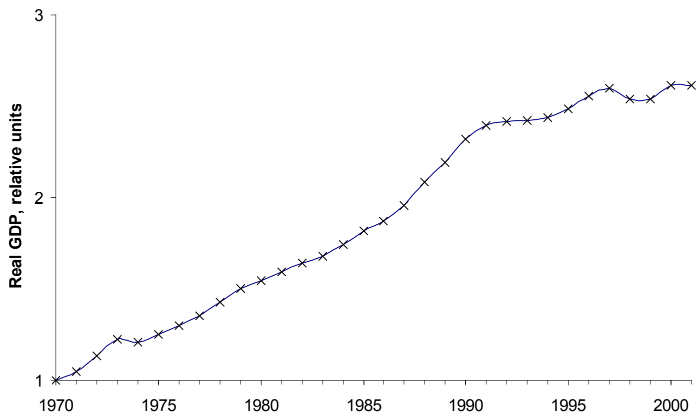

Figure 4. The dynamics of real GDP in Japan in 1970-2001 (USDA data points are indicated with crosses)

Figure 3 shows the dynamics of earnings in Japan calculated from the JTA dataset. Figure 4 then displays the real GDP growth in Japan for the same period of time (data from USDA, 2007). A simple visual inspection of the figures reveals that the real GDP and the average income (i.e. β) are strongly correlated: the corresponding linear correlation coefficient $r > 0.98$ (note also that the most probable income, although was growing, remained relatively independent from the dynamics of GDP). At the same time, until approximately 1991, the rate of the income growth was higher than that of GDP, while for 1991-2001 we observe the opposite development. In view of equation (5), a possible explanation for this fact would be that beginning from 1991, the income growth was slowed down by deflation. Qualitatively, the latter is confirmed by the data of Bank of Japan (see BoJ, 2000). We would also speculate that it is the beginning of 1990th, when intensifying competition among the existed industries and tough market conditions began forcing Japanese companies to diversify their production and look for new market niches by creating new industries. Although needing a thorough quantitative analysis and verification, the proposed parameterization of the average income thus appears a convenient tool to explore the interplay effects of the basic macroeconomic indicators.

Returning to the issue of goodness-of-fit, we would like to point to the following facts. Somewhat differently from previous works on the use of the exponential distribution for income distribution modeling (e.g. see Dragulescu and Yakovenko, 2001a and 2001b), the theoretical framework of Section 2 implies that in most practical situations, the empirical distribution should indeed deviate from the standard exponential, owing to various inhomogeneities peculiar to the data generation and measurement processes. When the deviations (does not matter how large) are observed for F, an empirical distribution with a completely monotone probability density function (PDF), one still can always obtain a sound statistical fit with a hyperexponential model of the following form:

$$F^{(n)}(S_0) = \sum_{i=1}^{n} c_i (1 - \mathrm{e}^{-\frac{S_0}{\beta_i}}), \tag{6}$$

where $c_1 + c_2 + ... + c_n = 1$, that is a generalization of equation (4) for the case of income in $n \geq 1$ industries. It can be shown that $F^{(n)}$ asymptotically converges to F (see Feldmann and Whitt, 1998, where details of a relevant parameter estimation algorithm are also given).

A further natural generalization of equation (4) can be obtained if one considers the scale parameter β as itself a random variable with a distribution function $F(\beta)$. In many situations, it would be natural to incorporate a directing process for β (e.g. to account for an uncertainty in determination of GDP or else for fluctuations in average prices over different economy sectors) into the model. The corresponding infinite mixture of exponentials can be written as

$$F^c(S_0) = 1 - F(S_0) = \int_0^{\infty} \mathrm{e}^{-\frac{S_0}{\beta}} d\,F(\beta), \tag{7}$$

where $F^c(S_0)$ is the tail, or survival, function. Equation (7) defines a Laplace-Stieltjes transform that can be used to replicate behavior of a very broad class of empirical distributions of income. It has been shown, for example, that if $F(\beta)$ is the standard beta distribution, equation (7) produces distributions with an exponentially damped power tail, and if $F(\beta)$ is the beta distribution of the second kind, one obtains distributions with a power tail (see Abate and Whitt, 1999, for details).

Thus, while the choice of a specific distribution of the form (4), (6), or (7) should be justified by a particular study's goal and available information about the analyzed economy, the proposed modeling framework can accommodate virtually all types (i.e. with exponential, semi-exponential, or heavy tail) of income empirical distributions.

5. Conclusions

A thorough study of the income distribution mechanism and the related social phenomena and processes would need to rely on a variety of micro- and macroeconomic indicators, rather than merely on income data. It is also desirable to have a convenient (i.e. simple or at least computable) and verifiable theoretical model not only connecting the data with the economic indicators, but also providing for accurate predictions of the empirical observations. Surprisingly, few of the known models of income distribution satisfy these criteria.

The modeling framework described in this paper, although does not pretend to deal with all the important indicators of an economy, offers a simple explanation of how the distribution of money among the industries forming an economy would shape up the distribution of income in that economy. The proposed model oversimplifies the reality but, at the same time, it has a potential to generate, through extensions, models of nearly arbitrary complexity.

The presented experiment is a rather superficial attempt to explore the mechanism of income distribution in Japan. Unfortunately, the aggregated character of the available data (especially, in regard to the distribution tails) prevented us from a more detailed analysis that would, possibly, allow for arriving at a closed functional form with a better predictive capability. This, together with an analysis of various social processes (e.g. as was previously done in Kryssanov *et al*., 2008) that possibly direct the distribution of income is left for future study.

References

Abate, J., and Whitt, W. (1999). Modelling service-time distributions with non-exponential tails: Beta mixtures of exponentials. *Stochastic Models* **15**, 517-546.

BoJ (2000). Price Developments in Japan: A Review Focusing on the 1990s. Bank of Japan, available from: http://www.boj.or.jp/en/type/ronbun/ron/research/data/ron0010a.pdf

Dragulescu, A.A., and Yakovenko, V.M. (2001a). Exponential and power-law probability distributions of wealth and income in the United Kingdom and the United States. *Physica A* **299**, 213–221.

Dragulescu, A.A., and Yakovenko, V.M. (2001b). Evidence for the exponential distribution of income in the USA. *European Physical Journal B* **20**, 585-589.

Feldmann, A., and Whitt, W. (1998). Fitting mixtures of exponentials to long-tail distributions to analyze network performance models. *Performance Evaluation* **31**, 245-279.

Fukawa, T. (2006). Income distribution in Japan based on IRS 1987-2002. *The Japanese Journal of Social Security Policy* **5** (1), 27-34.

Gibrat, R. (1931). *Les Inegalites Economiques*. Sirely, Paris.

Kryssanov, V.V., Rinaldo, F.J., Kuleshov, E.L., and Ogawa, H. (2008). A Hidden Variable Approach to Analyze “Hidden” Dynamics of Social Networks, in: Friemel, Thomas N. (Ed.), *Why Context Matters. Applications of Social Network Analysis*, pp. 17-35. VS Verlag, Wiesbaden (Germany).

McDonald, J.B. (1984). Some Generalized Functions for the Size Distribution of Income. *Econometrica* **52** (3), 647-663.

Pareto, V. (1897). *Course d’Economie Politique*. Macmillan, Paris.

Parker, S.C. (1999). The generalised beta as a model for the distribution of earnings. *Economics Letters* **62**, 197-200.

Reed, W. (2001). The Pareto, Zipf and other power laws. *Economics Letters* **74**, 15-19.

Salem, A.B.Z., and Mount, T.D. (1974). A Convenient Descriptive Model of Income Distribution: the Gamma Density. *Econometrica* **42** (6), 1115-1128.

Singh, S.K., and Maddala, G.S. (1976). A Function for Size Distribution of Incomes. *Econometrica* **44** (5), 963-970.

Souma, W. (2002). Physics of personal income, in: H. Takayasu (Ed.), *Empirical Science of Financial Fluctuations – The Advent of Econophysics*, pp. 343-352. Springer, Tokyo.

Thurow, L.C. (1970). Analyzing the American income distribution. *American Economic Review, Papers and Proceedings* **60**, 261-269.

USDA (2007). The International Macroeconomic Data Set. USDA, available from: http://www.ers.usda.gov/Data/Macroeconomics/Data/HistoricalRealGDPValues.xls