«ΣΠΟΥΔΑΙ», Τόμος 46, Τεύχος 1ο-2ο, Πανεπιστήμιο Πειραιώς / «SPOUDAI», Vol. 46, No 1-2, University of Piraeus

THE APPLICATION OF BETA AND GAMMA DISTRIBUTIONS TO UNDER-REPORTED INCOME VALUES

By
*George Vaphiades**
Agricultural Bank of Greece
Research and Planning Division

Abstract

The paper "The application of Beta and Gamma Distributions to Under-reported Income Values" considers two functional forms for the distribution of incomes —the Gamma and the Beta density functions. A model of under-reported income values is applied to each functional form assuming only one-sided errors. The parameters of the distribution governing the true values are identified on the basis of a sample on the observed values. So far as the dispersion of true incomes is concerned, some conclusions seem to be substantiated by empirical results obtained elsewhere. (JEL, D31)

1. Introduction

In this short paper I consider the case of under-reported income values on the basis of a model suggested by Krishnaji (1970). I assume that the true values follow either the gamma or the beta distribution and I proceed to identify the parameters of the distribution governing the true values on the basis of a sample on the observed values. So far as the dispersion of true incomes is concerned, some conclusions seem to be substantiated by empirical results obtained elsewhere, Salem and Mount (1974), Thurow (1970).

2. The Case of the Gamma Distribution

Krishnaji proposed an "errors-in-variables" model where he assumed that income values are under-reported as it may be in the case of data collected

* Helpful comments received from an anonymous referee of this journal are gratefully acknowledged. Any remaining errors are the sole responsibility of the author.

through sample surveys. If we let Y be a random variable representing the reported income values and X a r.v. corresponding to the true values, the two random variables are related to each other according to the equation

$$Y = RX, \text{ where R is a random variable} \qquad (1)$$

independent of X with range [0,1]

The above relation implies that Y is always an understatement of X; only one-sided errors are considered in the model and not observation ones. Assuming that R follows the power function distribution (a special case of the beta distribution) Krishnaji shows that the Pareto density possesses certain invariance properties.

While the limited applicability of the Paretian fits is acknowledged, the author correctly points out that an obvious implication of (1), namely that on the assumption of log-normality of X and R the r.v. Y will be log-normally distributed with different parameters, should be rejected as the variable R cannot have an infinite range of values. Another case which is not considered by Krishnaji is the assumption that the r.v. X follows the gamma distribution. Salem and Mount (1974) approximated the distribution of personal income in the U.S.A. for the years 1960 to 1969 by a two-parameter gamma density function, which was found to give a better fit than the log-normal one.

Let us assume that the true income values follow the p.d.f. $f(x;a,\lambda) = \frac{\lambda^a}{\Gamma(a)} x^{a-1} e^{-\lambda x}$ for $X > 0$ and $a, \lambda > 0$ where λ is an indicator of scale and a an indicator of inequality since skewness $Sk = \frac{1}{\sqrt{a}}$. We have no information about the error variable R but we know its range [0,1], therefore we may assume that it is a beta variable with the p.d.f.

$$h(r) = \frac{(p+q-1)!}{(p-1)!\,(q-1)!} r^{p-1} (1-r)^{q-1} \quad \text{for} \quad [0,1] \quad \text{and} \quad p, q > 0.$$

Upon the restriction $\rho = a - q$, it has been proved[1] that the product of the gamma variable X with parameters (a, λ) and the independent beta variable

1. See Weatherburn (1961).

($\rho = a - q$, q) will be a gamma variable with parameters ($a - q, \lambda$). In other words, the p.d.f. of the observed or reported income values will be

$$\varphi (y;a - q, \lambda) = \frac{\lambda^{a-q}}{\Gamma (a - q)} y^{a-q-1} e^{-\lambda y} \quad \text{for } Y > 0. \tag{2}$$

Comparing the skewness of the distributions of X and Y we have $Sk_x = \frac{1}{\sqrt{a}} < Sk_y = \frac{1}{\sqrt{a-q}}$. Therefore there is less inequality in the distribution of the true income values, the difference depending upon the value of the parameter q.

If under-reporting of incomes takes place there may be less unemployment than the reported one, but we can legitimately conclude that the official figures showing the real national product are, in fact, under-estimated.

The above result coincides with the one stated by Salem and Mount who argue that, in applying the gamma distribution, inequality is shown to decrease when unemployment decreases or when the real gross national product increases.

It is interesting to note that the same results apply when one considers two independent random variables, X1 and X2, corresponding to the observed and unreported parts of income respectively and following the Gamma distribution. More specifically, if X1 is a Gamma r.v. with parameters a_1, λ and X2 a Gamma r.v. with parameters a_2, λ then the true income values X= X1+X2 follow the Gamma distribution[2] with parameters $a = a_1 + a_2$, λ. The skewness of the X-distribution (true income values) is $Sk_x = \frac{1}{\sqrt{a_1+a_2}}$ which is less than the corresponding $Sk_{X_1} = \frac{1}{\sqrt{a_1}}$, of the X1 distribution (observed income values). Assuming an additive relationship between observed and unreported income values there is less inequality in the distribution of the true income values.

3. The case of the Beta distribution

This case arises if we assume that there is a maximum true income value Xmax, and that X/Xmax=W follows the beta density with parameters ρ and σ, that is

2. See Johnson and Kotz (1970).

$$f(\omega; \rho, \sigma) = \frac{1}{B(\rho, \sigma)} \omega^{\rho-1} (1 - \omega)^{\sigma-1} \quad \text{for } W \in [0,1]$$
$$\rho, \sigma > 0$$

The beta distribution was applied by Thurow (1970) to the U. S. Bureau of Census income distribution statistics for households from 1949 to 1966 and, as Thurow argues, it fits well the actual income distributions.

Assuming again that income values are under-reported we can follow the same procedure as in section 2 retaining all the assumptions with respect to the error variable. Consequently we have U = RW, where U = Y/Xmax and X, Y correspond to true and observed income values respectively.

As far as the variable U is concerned Jambunathan (1954) has proved that the product of two independent beta variables of the first kind with parameters (ρ, σ) and (p, q) is a b_1 $(\rho, \sigma + q)$ variable, provided that $p = \rho + \sigma$. Therefore the p.d.f. of U will be

$$\varphi(u; \rho, \sigma + q) = \frac{1}{B(\rho, \sigma+q)} u^{\rho-1} (1 - u)^{\sigma+q-1} \quad \text{for} \quad [0,1]$$

As $\sigma < \sigma + q$ there is less dispersion and higher median incomes in the distribution of observed income values. If under-reported income values are related to the existence of an under-ground economy which may lead to a higher GNP per capita and less unemployment the resulting distribution of true income values will have higher inequality and smaller median incomes. Direct comparison with Thurow's empirical results is rather difficult and it may be proved misleading. Thurow considers the effect of growth, employment and other macroeconomic variables on the parameters of the Beta distribution. According to Thurow growth, measured in terms of constant dollar per employee $\left(\frac{\quad}{E}\right)$, leads to higher real incomes but does not have any major impact on the dispersion of income. Rising employment, on the other hand, leads to more equality but a falling share for personal income leads to more inequality. Thurow points out that in booms employment rises but the proportion of income going to persons falls.

On the assumption that under-reporting of income is related to less unemployment it seems that, in the case of the Beta density and contrary to Thurow's evidence, rising employment may not lead to more equality in the income distribution.

When income values are described by the Beta density what is likely to play an important role in this rather surprising result of greater inequality appearing in the distribution of true income values is the imposition of a maximum income value. It is interesting to note that this point applies to the Gamma density too. The imposition of a maximum income value Xmax to the Gamma density produces a truncated Gamma density of the form

$$\frac{x^{\alpha-1} \quad e^{-\lambda x}}{\int_0^{Xmax} t^{-1} \ e^{-\lambda t} dt}$$

for [0, Xmax] with moments $\mu'_\rho(x) = \dfrac{\Gamma_{Xmax}(\alpha + \rho)}{\Gamma_{Xmax}(\alpha)}$.

The skewness of this distribution is $Sk = \dfrac{1}{\dfrac{\Gamma_{Xmax}(\alpha + 1)}{\Gamma_{Xmax}(\alpha)}}$

which means that as Xmax increases inequality decreases.

When Xmax→ ∞ then

$$Sk = \frac{1}{\sqrt{\dfrac{\Gamma(\alpha + 1)}{\Gamma(\alpha)}}} = \frac{1}{\sqrt{\dfrac{\alpha!}{(\alpha - 1)!}}} = \frac{1}{\sqrt{a}}.$$

4. Conclusions

In conclusion, if one believes in distributions as descriptions of behaviour the following results are of some importance.

i) Although one will often be able to find some distribution for the proportion of income revealed which will render parameters of the income distribution unidentifiable if true income is either Gamma or Beta distributed the parameters of the distributions are not identifiable from data on observable income.

Johnson N. and Kotz S. (1970), Distributions in Statistics: Continuous Univariate Distributions-1, John Wiley and Sons Inc.

Krishnaji N. (1970), "Characterization of the Pareto distribution through a model of under-reported incomes", Econometrica, Vol. 38, pp. 251-3.

Krishnaji N. (1970), "A characteristic property of the Yule distribution", Sankhya A., No 32, pp. 343-346.

Prichard J. and Eagle R. (1965), Modern Inventory Management, John Wiley, New York.

Salem A.B.Z. and Mount T.D (1974), "A convenient descriptive model of income distribution: the gamma density", Econometrica, Vol. 42, pp. 1115-27.

Thurow L. (1970), "Analyzing the American income distribution", The American Economic Review, Papers and Proceedings, Vol. 60, pp. 261-269 ("Distributions Issues: Trends and Policies").

Weatherburn C. E. (1961), A First Course in Mathematical Statistics, University Press, Cambridge, UK.

ii) In the existence of underground economy the parameter a of the true Gamma density increases so that there is less inequality in the distribution of the true income values. On the other hand the true Beta density will have higher inequality and smaller median incomes.

iii) The imposition of a maximum income value on true incomes affects income inequality when the Gamma and Beta densities are considered. More specifically, the imposition of a maximum income value to the Gamma density produces a truncated Gamma with decreasing inequality as the maximum value increases.

The stochastic multiplicative model Y = RX has also important applications in other fields. In inventory decision making, X represents the demand for an item within a unit time interval and Y item units in stock within the same time interval (Prichard and Eagle, 1965). In discounting cash flows, X represents a payment to be paid at some future time and Y the present value of the payment (Artikis et al., 1991 and 1992).

A stochastic multiplicative model as in (1) appropriately modified to account for discrete random variables X, Y is given by Y = [RX], where [RX] denotes the integral part of Y U X. Krishnaji (1970), though not referring to financial modes, used Y = [RX], with R uniformly distributed in [0,1], to establish a characterization of a zero-truncated Yule distribution. Furthermore, Artikis et al. (1994) have used a modified form of the above model in certain selecting and under-reporting processes. It seems that further study should be carried out on the properties and the financial applications of the two stochastic multiplicative models.

References

Artikis T. and Jerwood D. (1991), "On the unimodality of present value distributions", European Journal of Operational Research, No 52, pp. 107-112.

Artikis T., Jerwood D. and Voudouri A. (1992), "Analytical and simulation techniques for discounting binomial random sums", Mathematical and Computer Modelling, No 16 (3), pp. 56-58.

Artikis T., Voudouri A. and Malliaris M. (1994), "Certain selecting and underreporting processes", Mathematical and Computer Modelling, No 20 (1), pp. 103-106.

Jambunathan M. V. (1954), "Some properties of beta and gamma distributions", The Annals of Mathematical Statistics, Vol. 25, pp. 401-2.