The Determinants of Income Dynamics

44

Gustavo De Santis and Giambattista Salinari

Abstract

Models of income distribution more or less succeed in linking the current *level* of household (or individual) income to household (or individual) characteristics. However, they are typically far less satisfactory in explaining income *dynamics*. Gibrat's model proves helpful in highlighting the predominant role of random-ness in the short run (here, 2–4 years), and this explains why other systematic influences are difficult to identify. One empirical regularity that does emerge, however, is that small incomes tend to increase more, and with more variability, than large ones. The traditional version of Gibrat's model does not incorporate this peculiarity, but this shortcoming can be overcome with a relatively minor modification of the original model.

Keywords

Income dynamics • Poverty and inequality • Statistical modelling

44.1 Household Income Dynamics and the Problem of Heteroscedasticity

The analysis of income dynamics is frequently carried out with regressions and OLS estimators, both of which assume homoscedasticity. But results can be biased if this condition is violated, and this, unfortunately, is virtually always the case with income dynamics.

G. De Santis (✉)
Department of Statistics, University of Florence, Italy
e-mail: desantis@disia.unifi.it

G. Salinari
University of Sassari, Piazza D'Armi, 17, 07100 Sassari, Province of Sassari, Italy
e-mail: gsalinari@uniss.it

At least two systematic violations of homoscedasticity have been observed. First, small incomes tend to increase more, and with more variability, than large ones [8]. Second, there are occupational categories whose incomes follow markedly different paths in terms of both average rates of increase and variability. Consider, for instance, two employed persons, a dependent and an independent worker. In any given period, the income growth rate of the independent worker tends to be not only higher but also more variable. Therefore, in order to determine, for instance, which of the two individuals is more likely to fall in poverty in any given time interval, both aspects (average and variance of growth rates) must be modelled.

And this is precisely what this paper sets out to do: present a model of income dynamics where heteroscedasticity is explicitly taken into account and can be (partly) explained by covariates. In so doing, we will also highlight the great variability of income trajectories—greater, indeed, than it is normally acknowledged in literature.

Already at the individual level, income trajectories are affected by very many variables: education, age, gender, industry, SES (Socio-Economic Status), changes in the labour market, economic cycles, etc. At the household level, which is what this paper refers to, the number of these influences increases exponentially, if one takes interactions into account, and soon becomes very difficult, if not impossible, to handle.

In order to simplify the problem, we break it in two: (A) we first model income trajectories in a theoretical, perfectly homogeneous population; and (B) later on, we complicate the model, by explicitly considering heterogeneity. This procedure permits us to disentangle the general features of income dynamics, common to all trajectories (point A), from those that are contingent on households characteristics (point B).

An empirical application to household incomes and their dynamics in the period 1998–2006 suggests that our model can keep both dimensions under control (the general process and the influence of specific situations, in terms of both average and variability of growth rates), and, both theoretically and empirically, compares favourably with other more conventional models of income dynamics.

44.2   Gibrat's Model: Traditional and Modified Version

Our model of income dynamics in a homogeneous population (see [8]), derives from [2] seminal paper, where incomes are assumed to evolve in a multiplicative form. In Gibrat's formulation, for every individual, income at time $t + k$ (y_{t+k}) can be obtained by multiplying income at time t (y_t) by a series of random shocks R_j occurring in between

$$y_{t+k} = y_t \prod_{j=1}^{k} R_j \tag{44.1}$$

which, in logarithms, becomes

$$\log(y_{t+k}) = \log(y_t) + \sum_{i=1}^{k} R_i \tag{44.2}$$

The multiplicative form implies that income is always strictly positive: depending on how income is defined and on the length of the period under consideration, this limitation may occasionally force researchers to drop a few rather extreme observations (but not in our application). In all cases, this does not constitute a major obstacle to the use of a model that has repeatedly proved very effective in the description of income dynamics.

Gibrat's model implies four main theoretical predictions, all of them relevant and empirically testable (see, e.g., [4–6, 8, 9]):

1. Log-growth is independent of log-income
2. The distribution of log-incomes approximates the normal
3. The variance of log-income increases linearly with k
4. The covariance of log-income increases with t, but decreases with k.

Empirical analyses, reported, for instance, by Hart [1, 3], or by Salinari and De Santis [7, 8], show that implications (44.2)–(44.4) hold, but implication (44.1) does not, because incomes that start from low levels tend to increase more and with more variability than the model predicts.

We contend that this divergence between theoretical expectations and empirical results depends on a sort of mechanical reason. In order to see this more clearly, it is convenient to transform Gibrat's model in its additive version:

$$y_{t+1} = y_t + y_t(R - 1) = y_t + y_t r \tag{44.3}$$

The growth rate $r(= R–1)$ can be thought of as the difference between what might be defined as the "gain rate" g and the "loss rate" l, so that Eq. (44.3) can be rewritten as:

$$y_{t+1} = y_t + G - L \tag{44.4}$$

where both $G(= g \cdot y_t)$ and $L(= l \cdot y_t)$ are random, non-negative variables. Losses L cannot be greater than $y_t + G$, otherwise income at time $t + 1$ would be negative. Therefore, L is a truncated random variable, the probability density function of which ranges between 0 and $y_t + G$. The impact of truncation is obviously greater for small than for large incomes, so that the growth rate of small incomes tends to be positively affected.

But why is it also more variable? Under our homogeneity assumption, gains and losses can be imagined to occur to individuals as in a Bernoulli experiment, where the probability of getting a "unit" of loss (or gain) is proportional to individual income. In this case, the gains and losses experienced by households distribute as a Poisson, the mean and variance of which increase linearly with income. As a consequence, the variance of the growth rate $V[r] = V[g–l]$ decreases as income increases.

A possible way of improving over Gibrat's model is to use a truncated normal distribution (T) to model income variations ($\Delta y = y_{t+1} - y_t$). Let these variations be normally distributed with mean and variance proportional to y_t. Since variations cannot be smaller than $-y_t$ (otherwise y_{t+1} would be negative), we restrict the domain of the probability distribution to the interval $(-y_t, +\infty)$. The probability density function $f(\cdot)$ of the relevant truncated random variable T is therefore:

$$f(\Delta y | y_t, \alpha, \beta) = \frac{\phi\left(\frac{\Delta y - a y_t}{b y_t}\right)}{1 - \Phi\left(\frac{-y_t - a y_t}{b y_t}\right)} \tag{44.5}$$

where a and b are two parameters defining the mean ($a y_t$) and the variance ($b y_t$) of the normal distribution, $\phi(\cdot)$ is the probability density function of the standard normal variable, and $\Phi(\cdot)$ its cumulative function.

We can now define Gibrat's "modified model" as:

$$y_{t+1} = y_t + T \tag{44.6}$$

where the two parameters a and b can be estimated numerically, for instance with maximum likelihood. Figure 44.1 shows a comparison between the actual and the expected (44.5) income variations registered for 5,195 Italian households in a series of 2-year intervals (1998–2000, 2000–2002,..., 2004–2006), conditional on the income class to which households belonged at the beginning of each period. Note that the variability of absolute income variations increases with income and that our model curves approximate reality rather well, especially if one considers that no covariate has been taken into consideration yet (homogeneity assumption).

The model of Eq. (44.6) can be extended to the case of heterogeneous populations, where households differ by their socio-economic characteristics. Let $\mathbf{x}_{t,i} = (1, x_{1,i}, \ldots, x_{p,i})$ be a time-dependent vector of covariates for household i, while $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \ldots, \alpha_p)$ and $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)$ are two vectors of parameters describing the effects produced by these covariates, respectively, on the mean and on the variance of the distribution of the variations. Model (44.5) can now be generalized as follows:

$$\begin{cases} a_i = \exp\left(\boldsymbol{\alpha}' \mathbf{x}_{t,i}\right) \\ b_i = \exp\left(\boldsymbol{\beta}' \mathbf{x}_{t,i}\right) \end{cases} \tag{44.7}$$

44.3 SHIW Data

For the empirical part of our study, we use micro data taken from the Survey on Household Income and Wealth (SHIW). The Bank of Italy carries out this survey every other year, on about 8,000 households (about 20,000 individuals): it
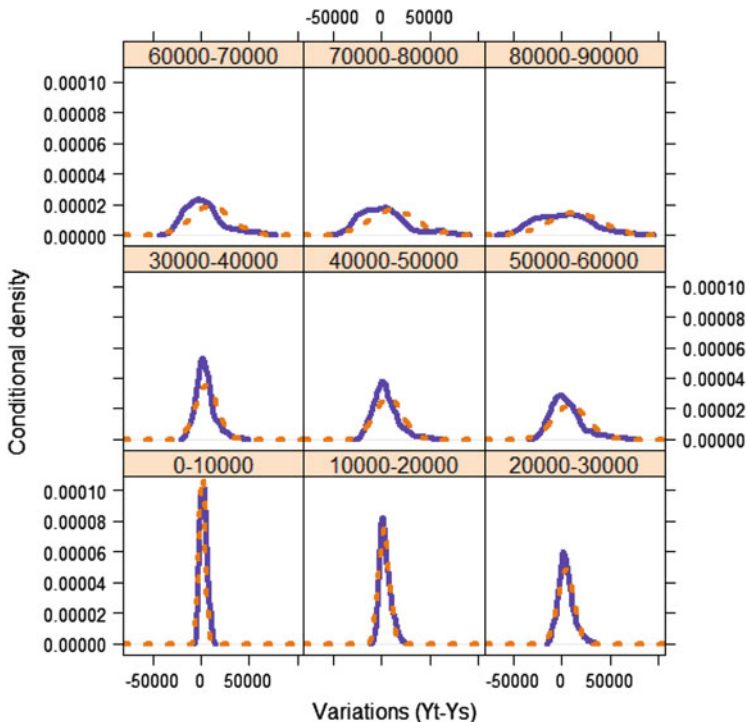
Fig. 44.1 Probability density distribution of conditional income variations (Italy 1998–2006). The variations (Δy) are observed in the periods 1998–2000,..., 2004–2006 for 5,195 Italian households initially belonging to different income classes (0–10,000,..., 80,000–90,000 Euros). The *solid line* indicates the empirical densities estimated with Gaussian kernel. The *dashed lines* indicate the theoretical distribution generated by Eq. (44.5) (with $a = 0.16$ and $b = 0.32$)

is described in detail on the website of the Bank,[1] from where elementary data can be downloaded, and, most importantly, it contains a panel part that we will exploit in this application.

We work on five rounds of the survey (years 1998, 2000, 2002, 2004 and 2006), and we study the change in the total net household income over a period of either 2 years (1998–2000; 2000–2002;...) or 4 years (1998–2002; 2000–2004;...). We work on a subset of the panels, selected on the basis of two criteria: the demographic structure of the household must not change in the period under examination (2 years or 4 years, depending on the application; see further in the text) and, at the beginning of the period, the reference person must not be older than 64 years and must be in the labour market. After dropping a few outliers (fewer than 1%, with abnormally high or low growth rates), we are left with about 5,100 biannual transitions and 2,400 quadrennial transitions, approximately equally distributed among the various time intervals considered.

---

[1]http://www.bancaditalia.it/statistiche/indcamp/bilfait.

44.4   A Regression Analysis of Income Dynamics

Before applying our modified version of Gibrat's model, let us take a step backward and consider the results produced by an ordinary regression analysis of household income and growth rates, which we will later compare to our own.

We are interested in modeling income dynamics, using "independent" explanatory variables. Actually, most of the variables that we use are not truly exogenous (as we will briefly discuss below) but our interest here is not really about how to "explain" income: merely to show that our approach (modified Gibrat's model) works better than other, more traditional ones, using the same set of explanatory variables.

In the traditional version, the dependent variable is the growth rate (Table 44.1, columns 2 and 3); we, instead, model transition probabilities (not reported here; but see, e.g. [8]) which result, among other things, in growth rates (Table 44.2). In both cases, we use a set of "independent" variables, of two broad types:

(a) *Demographic*: number of children $(0, 1, 2, 3+)$, region (North, Center, South), age, and educational attainment of the reference person $(1 = $ illiterate or primary school; $2 = $ secondary education; $3 = $ tertiary education; $4 = $ university degree or PhD).

(b) *Labour force*: occupation of the reference person (dependent, self-employed, unemployed, other), period.

As for the traditional approach, although we are basically interested in dynamics (growth rates), we also model income levels. In the latter case, we use:

$$\log(y_t) = \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_{t,i} + \varepsilon_{t,i} \tag{44.8}$$

while in the former (growth rates—both for 2- and 4-year periods), we use:

$$\log\left(\frac{y_{t+1,i}}{y_{t,i}}\right) = \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_{t,i} + \varepsilon_{t,i} \tag{44.9}$$

Our results are summarized in Table 44.1. In all the cases, the baseline is a household living in the north of Italy, with one child and whose reference person is dependent worker, aged 45, with an average level of education. The static regression of income (first column) confirms what is generally known: all the covariates we use affect income in the expected direction: incomes are higher for better educated and older (but younger than 65 years) reference persons, while living in the south, or being a dependent worker (or, worse still, unemployed), depresses income. Finally, income increases with the number of children. In all of these cases, the interpretation must be particularly cautious, because reverse causation, selection and unobserved heterogeneity surely play a major role, but, as mentioned before, we will not attempt to investigate these issues here; we are simply comparing models of income dynamics, and we are using this regression on income levels as an entry point, whose utility will appear shortly. This simple model "explains" about 43% of the variance in incomes.

Table 44.1 Regression analysis of income and income growth rates

Dependent variable	Income		Growth rate over 2 years		Growth rate over 4 years	
Intercept	10.223	***	0.087	***	0.167	***
Children 0	−0.089	***	−0.025		−0.056	
Children 2	0.077	***	0.010		0.034	
Children 3	0.069	**	0.040	*	0.051	
Age (Stand.)	0.140	***	−0.014	*	−0.011	
Education 1	−0.254	***	0.052	**	0.095	**
Education 3	0.240	***	0.009		0.008	
Education 4	0.562	***	0.023		0.006	
Self-Empl.	0.064	***	0.008		0.037	
Unempl.	−0.616	***	0.231	***	0.403	***
Other	−0.160	***	0.036	.	0.084	**
Period 2	0.067	***	−0.011		0.012	
Period 3	0.110	***	0.002		−0.010	
Period 4	0.168	***	−0.023		/	
Center It.	−0.124	***	0.017		0.030	
South It.	−0.476	***	−0.008		−0.028	
R^2	0.43		0.01		0.03	

Significance level: *** $= 1\%;$** $= 5\%;$* $= 10\%$. Periods: $1 =$ from 1998 (ref.); $2 =$ from 2000; $3 =$ from 2002; $4 =$ from 2004. Education: $1 =$ low; $2 =$ medium (ref.); $3 =$ high; $4 =$ very high. For the reasons of comparability, R^2 for the growth rates regressions have been calculated as var[exp(fitted.values)]/var(actual.growth.rates)

If we now turn to the analysis of dynamics, i.e. of 2-year growth rates (second column) we find that only a few of our covariates exert a measurable effect, and the explained variance is much lower—merely 1%. The only variables that seem to matter are those associated with a low income at the start (which favours a more rapid increase in the period): for instance being unemployed, or with a low education.

Basically the same happens over a longer time span (4 years, third column), with a very slight improvement in the overall goodness of fit (R^2, from 1% to 3%). In short, while our covariates are associated with income levels, they do not seem to be associated with growth rates. How is this possible?

In order to answer this question, we note that:

(a) Income growth is highly erratic
(b) Growth is a cumulative process: very small systematic differences, even if they go unnoticed in the short run, may result in large differences in the long run
(c) Income trajectories can (and indeed do seem to) depend on their starting point
(d) What we observe here are net incomes, which are also influenced by the effects of the fiscal policy (progressive taxation, subsidies, etc.)

Table 44.2 Analysis of income rates of growth through the modified Gibrat's model

| Coeff. | Variation after 2 years | | | | Variation after 4 years | | | |
	Estim. α		Estim. β		Estim. α		Estim. β	
Intercept	−1.498	*	−0.599	*	−0.934	*	−0.266	*
Children 0	0.114	*	−0.010		−0.260	*	−0.002	
Children 2	−0.105		0.028		−0.067		−0.150	*
Children 3	0.025		0.075		0.081		−0.058	
Age (Stand.)	−0.083	*	−0.009		0.059		0.013	
Education 1	−0.063		0.233	*	0.150		0.357	*
Education 3	−0.249	*	−0.121	*	−0.155	*	−0.089	
Education 4	−0.232	*	−0.057		−1.160	*	−0.597	*
Self-Empl.	0.203	*	0.295	*	0.201	*	0.289	*
Unempl.	0.518	*	0.347	*	0.568	*	0.486	*
Other	0.047		0.023		0.294	*	0.080	
Period 2	−0.071		−0.306	*	0.045		−0.216	*
Period 3	0.011		−0.111	*	0.004		−0.214	*
Period 4	0.150	*	−0.121		\		\	
Center It.	−0.011		−0.077		0.093		0.091	*
South It.	0.118		0.196	*	0.142	*	0.050	
R^2	0.02				0.05			

* Significant at 10%

R^2 calculated as var(fitted.growth-rates)/var(actual.growth-rates)

As for the first two points, consider, for example, "Self-employment" in Table 44.1 (2nd and 3rd column): the coefficients are positive (income increases more rapidly for this category), but not significant. However, the mean income of the self-employed is significantly higher than that of the dependent workers. Besides, over a 4-year period, the effect of self-employment is slightly more significant than it is over a 2-year period, and this is a pattern that holds for most of our variables.

Unfortunately, this explanation cannot be simply extended to all our covariates. Take "Education 1", for instance: this has a positive and significant effect on both the biannual and quadrennial growth rates, but the income of the poorly educated workers is significantly smaller than that of highly educated ones. Our interpretation is that this apparent contradiction depends on the different starting income of the two groups: highly educated workers are richer at the beginning (they typically come from richer families, and their entry wage is higher), but they tend to lose some of their initial advantage as time goes by.

44.5 Income Dynamics from Gibrat's Perspective

Let us now apply our modified Gibrat's model to the same data set. Our starting values are the estimates of the coefficients obtained in the previous regression (Table 44.1, columns 2 and 3), and the confidence intervals for the coefficients

have been estimated through bootstrapping (500 repetitions). The results of our analysis are summarized in Table 44.2, where the α coefficients represent the effect produced by the covariates on the average increment, and, apart from the intercept, are directly comparable with those estimated in the regression of growth rates. The β coefficients, instead, represent the effects of the covariates on the variance of these rates of growth, and for these there is no corresponding value in the preceding table.

The α parameters of Table 44.2 tell basically the same story as in Table 44.1: variations in income are, on average, more strongly positive for the unemployed and for the self-employed; the evolution is relatively worse for the well educated, and the effect of age is not so clear. Note, however, that the variance explained by the model, although still very low, improves slightly and, more importantly, the cases when the α parameters are significant (if only at the 10% level) are now considerably more numerous than in Table 44.1. For example, self-employment exerts a significant positive effect in all the cases of Table 44.2, while this effect is weak or insignificant in the estimation of Table 44.1. We advance two tentative explanations for this improvement: (1) our modified Gibrat's model corrects the bias produced by what we referred to as the "differential dynamics of income" (small incomes increase faster and with a greater variability than others); (2) an optimization process based on likelihood is less influenced by the presence of outliers.

A distinctive feature of our modified Gibrat's model is that it allows us to measure the effects produced by household's characteristics also on the variability of income, and not only on its average. Our results suggest that low education, unemployment and self-employment all contribute to an increase in the variability of income variations in the subsequent period (of either 2 or 4 years). Note that periods 2 and 3 of both analyses (2 years and 4 years increments) are characterized by a significant lower variability than the baseline period (period 1). This may depend on the macroeconomic situation: the period 2002–2005 witnessed a slowdown of economic growth in Italy, with an average annual growth rate of the GDP of only about 0.6%. And in times of economic recession, the variability of income variations typically shrinks.

44.6   Conclusions

Our modified Gibrat's model seems to describe income dynamics better than other models from at least three different points of view:

1. We can theoretically justify, and model, why smaller incomes increase more, and with more variability, than others.
2. In our estimation, we circumvent the problem of heteroscedasticity, which is explicitly modelled. In our case, the bias turns out to be relatively modest (α coefficients are similar in Tables 44.1 and 44.2), but this need not be always the case.
3. Finally, and perhaps most importantly, our modified Gibrat's model permits us to model the *variance* of income dynamics, and to measure the impact of covariates

on this variance. This has several advantages: for instance, it allows researchers to better identify the population subgroups who are more exposed to the risk of poverty in any given time interval.

References

1. Fields, G.S., Duval Hernández, R.D., Freije Rodríguez, S.F., Sánchez Puerta, M.L.: Income Mobility in Latin America, mimeo (2006). http://www.cid.harvard.edu/Economia/Mexico06 %20Files/fieldsetal%20102306.pdf
2. Gibrat, R.: Les inégalités économiques: applications aux inégalité des richesses, à la concentration des entreprises, aux populations des villes, aux statistiques de familles, etc., d'une loi nouvelle, la loi de l'effet proportionnel, Librerie du Recueil Sirey, Paris (1931)
3. Hart, P.E.: The comparative statics and dynamics of income distributions. J. R. Stat. Soc. A Gen. **139**(1), 108–125 (1976)
4. Kalecki, M.: On the Gibrat distribution. Econometrica **13**(2), 161–170 (1945)
5. Mitzenmacher, M.: A brief history of generative models for power law and lognormal distributions. Internet Math. **1**(2), 226–251 (2003)
6. Neal, D., Rosen, S.: Theories of the distribution of earnings. In: Atkinson, A.B., e Bourguignon, F. (eds.) Handbook of Income Distribution, vol. 1, Chapter 7, pp. 379–427. Elsevier, London (2000)
7. Salinari, G., De Santis, G,: On the Evolution of Household Income, Luxemb. Income Study Work. Pap. Ser., No. 488, July (2008)
8. Salinari, G., De Santis, G.: The evolution and distribution of income in a life-cycle perspective. Riv. Ital. di Econ. Demogr. e Stat. **63**(1–2), 257–274 (2009)
9. Shorrocks, A.F.: Income mobility and the markov assumption. Econ. J. **86**(343), 566–578 (1976)