Output, Prices and the Distribution of Consumption in Rural India

BY

CLIVE BELL AND STEFAN KLONNER*

Südasien Institut der Universität Heidelberg and Economic Growth Center, Yale University

Südasien Institut, Im Neuenheimer Feld 330, 69120 Heidelberg, Germany
and
Economic Growth Center, P.O. Box 208269, New Haven, CT 06520-8269, USA

ABSTRACT

This paper analyzes the relation among agricultural output, inflation and the distribution of consumption in rural India, using the Singh-Maddala family to model the entire distribution parametrically. Employing a benchmark case in which growth is distributionally neutral and idiosyncratic shocks are completely smoothed, and using a GMM-estimator to deal with potential simultaneity between output and consumption, we conclude that: (i) growth was not distributionally neutral; (ii) good harvests (relative to trend) yielded improvements according to first-order stochastic dominance; (iii) slow growth before 1980 went with decreasing inequality; (iv) accelerated growth thereafter tended to increase inequality, though yielding improvements according to first-order stochastic dominance; (v) consumption-smoothing was incomplete.

Keywords: distribution, growth, Singh-Maddala distribution, India

JEL categories: D31, O40, O53

Acknowledgements

Bell gratefully acknowledges research support from the Deutsche Forschungsgemeinschaft (German Research Foundation). We are indebted to the participants in seminars at the Delhi School of Economics, Heidelberg University, the Indira Gandhi Development Research Institute, the National Institute of Public Finance and Policy, the L.S.E., Vanderbilt and Yale for useful comments on earlier drafts of this paper. Discussions with Timothy Besley, Robin Burgess and Martin Ravallion are also gratefully acknowledged. Invaluable assistance was rendered by Ansgar Wohlschlegel, who assembled and checked many of the data, and then arranged them in usable form. They are all fully absolved from any blame for the errors that remain.

---

* Corresponding author. Address: Economic Growth Center, P.O. Box 208269, New Haven, CT 06520-8269, USA. Phone: +203+4364827 Fax: +203+4323898, E-mail: stefan@klonner.de

Introduction

This paper analyzes the relation among inflation, output, and the size distribution of consumption in rural India over a period of almost four decades. The emphasis here is on explaining the movements of the entire distribution, as opposed to summary (mean-free) measures of inequality, or indices of poverty, on which there is an extensive literature (see, for example, Ahluwalia, 1978; Saith, 1981; Gaiha, 1989; Bell and Rich, 1994; Ravallion and Datt, 1996; Fan, Hazell and Thorat, 2000). The primary motivation is not so much the need for descriptive completeness, important though that is, but rather to arrive at conclusions about how social welfare has been influenced by growth, secular economic factors and transitory shocks.

The starting point is to characterize the observed size distributions of expenditures. We adopt a parametric approach, as pioneered by Thurow (1970), but since largely neglected. The advantage of such an approach is exploited as follows. First, if the distributions belong to a particular family, the changes in their scale and shape will stem from changes in the family's parameters, the values of which can be estimated from the data. The second step is to analyze changes in the parameters in the light of the equivalence between the ordering of distributions yielded by a general class of social welfare functions and those yielded by stochastic and Lorenz-dominance. The aim here is to find conditions among the parameters that yield an unambiguous ordering of the set of distributions with respect to stochastic dominance of some order and with respect to Lorenz-dominance, so that conclusions can be drawn about the effect of a change in a particular parameter on social welfare and inequality.

With this much accomplished, the third step, namely, identifying and analyzing the factors influencing the shape of the distribution, boils down to finding a set of variables that explain the movements of the parameters and estimating the magnitude and direction of their effects thereon. Particular attention will be paid to the role of movements in output. In the longer run, output and the distribution of rural incomes are influenced by, inter alia, population growth, technical change in agriculture, the provision of public infrastructure and the growth of non-farm employment opportunities. The available degrees of freedom are so limited, however, that only aggregate representations of such secular factors are possible in the econometric analysis. In the short run, transient shocks can also have important effects. Two kinds of such shocks will be considered. First, there are deviations of output from forecast values. Second, there are unanticipated price shocks, again interpreted as deviations of the price level from forecast values.

We must now address the fact that the Indian survey data pertain not to income, but to consumption. For, in principle at least, a variety of mechanisms may have been at work in smoothing consumption both within and across periods in the face of growing and fluctuating output per head. There are systemic shocks, such as droughts, floods, pestilence, improvements in technology and changes in the terms of trade, and idiosyncratic ones, such as illness in the family or the death of a work animal unrelated to epidemics. If there were ample opportunities to smooth consumption in the face of idiosyncratic shocks to current income, but there were limited possibilities to smooth income across periods, then one would expect the movements in the distribution of consumption to reflect changes in aggregate output as well as any forces making for permanent changes in the distribution of income, such as non-neutral technical progress.

Whether India's rural households have actually enjoyed such opportunities to smooth their consumption is, of course, another question. The extent to which the above factors make themselves felt as variability in the distribution of consumption at the aggregate level depends, in part at least, on public policy, especially in the realm of international trade and the management of food-grain reserves. Here, it can be argued that there were some serious mistakes, particularly in the stringent regulation of trade and in the failure to release stocks early (Joshi and Little, 1994). Since food-grains have a rather limited storage-life under

Indian conditions, mistakes in policy in such a regime will inevitably lead to fluctuations not only in the aggregate consumption of food, but also in aggregate rural expenditure as a whole.[1] For the buyers of land and animals, which are rural households' principal non-labor assets, are almost exclusively other rural households; and it is improbable that rural borrowing is financed largely by cuts in urban and public consumption. In this case, therefore, one would not expect aggregate expenditure to be especially smooth.

The plan of the paper is as follows. We begin by setting out a simple structure for the determination of output, inflation, income and consumption. This yields a benchmark case, which is useful in deriving structural forms for the econometric estimation. There follows a description of the data and a characterization of the distributions in the form of the hypothesis that all belong to the three-parameter family proposed by Singh and Maddala (1976), the parameter values for each year being estimated by a pseudo maximum likelihood procedure. We then proceed to estimate an econometric model of production, inflation and the size distribution of expenditures. Some concluding remarks follow.

The Framework

Consider a competitive economy in which individuals are endowed with labor, land and other factors of production, all of which they supply completely inelastically. In the absence of transfers, an individual's income is then simply the sum of the earnings yielded by renting out her entire vector of endowments at the going rates. Suppose the population is so numerous and the economy's aggregate endowment so distributed among them that the vector of individual incomes $\mathbf{y} = (y_1, ..., y_n)$ can be closely approximated by the continuous, parametric cumulative distribution function (c.d.f.) $G[(y/\alpha_y), \boldsymbol{\beta}_y]$, where α_y is a scale parameter and $\boldsymbol{\beta}_y$ is a vector of shape parameters.

Let aggregate output be produced under constant returns to scale, where the level of output is also influenced by the state of nature and technological shocks, which are represented by θ. The benchmark case is that where changes in output are distributionally neutral, that is, where each individual receives a fixed fraction of aggregate income. This can come about if, and only if, all factor-price ratios are independent of θ. In other words, a necessary and sufficient condition for the benchmark case to hold is that the effects of changes in θ on the technology be Hicks-neutral. It is then clear that after an appropriate rescaling of incomes, the c.d.f. $G[(y/\alpha_y), \beta_y]$ will be independent of θ: α_y will vary, but the shape parameters, and hence the Lorenz curve, will not.

Matters are much less tidy if the shocks represented by θ are not Hicks-neutral. Departures from 'neutrality' will then depend not only on the change in θ from one period to the next, but also on the elasticity of substitution in production and the extent to which factor endowment ratios vary across the population. If these departures are not too large, then given that $\mathbf{y}(\theta_t)$ can be closely approximated by the continuous c.d.f. $G[y/\alpha_y(\theta_t), \beta_y(\theta_t)]$, it is plausible that another c.d.f. from the same family can do the same for $\mathbf{y}(\theta_{t+1})$, where the shape parameters, and hence the Lorenz curve, now depend on the realization of θ in the period in question.

In applying this framework, there remains the difficulty that the National Sample Survey (NSS) data pertain not to income, but to consumption. Where idiosyncratic shocks to income are concerned, Townsend (1994) has established that in ICRISAT's well-studied villages at least, these are greatly, but not wholly, smoothed by a variety of mechanisms, whereas the effects of systemic shocks on households' consumption levels are substantial. Consider, therefore, the case where there is a complete set of state-contingent markets and all individuals have the same rate of time-preference. The ratio of the marginal utilities of consumption of any pair of individuals will be the ratio of the Lagrange multipliers associated with their respective life-time budget constraints, a ratio which is constant over time. It follows that if the felicity functions are isoelastic, then the ratio of their consumption levels will also be invariant over time: each individual will consume a fixed fraction of aggregate consumption (Deaton, 1992, pp. 35-36). This result corresponds precisely to the benchmark case with respect to income, and though the relation between aggregate consumption and aggregate income remains to be determined, the shape parameters for the distribution of consumption would be invariant to shocks in such a setting.

Turning to the distribution function of consumption, let individual *i*'s consumption in period *t* and state *s* (s = 1, ..., S) be denoted by $x_i(s_t)$, and let average consumption in period *t* be denoted by μ_{xt}. Define $\phi_i(s_t) \equiv x_i(s_t)/\mu_{xt}$. As noted above, if there is a complete set of markets in contingent claims, all individuals have the same rate of time-preference, and all felicity functions are isoelastic, then $\phi_i(s_t)$ will be constant over all periods for each and every *i*. For a continuum of individuals, the distribution of the ϕ_i's can be represented by a c.d.f. $F_\phi(.)$, say. In any period,

$$P(x_t < x) = P(\phi < x/\mu_{xt}) = F_\phi(x/\mu_{xt}).$$

In other words, if the distribution of consumption in period *t* can be represented by a continuous parametric distribution function $F_x^t(x) = F_x(x;\alpha_t, \beta_t)$, with scale parameter α_t and a vector of shape parameters β_t, then except for α_t, $F_x^t(\cdot)$ will be the same for all *t*. In the absence of any smoothing across periods, individual *i*'s consumption will also be a fixed fraction of aggregate income. The scale parameter α_t will be proportional to μ_{xt} and, in the absence of any smoothing, also to aggregate output. To summarize, with full consumption insurance within each period, but in the absence of storage or borrowing possibilities, we have

(1) $\alpha_t = ay_t$

and

$\beta_t = \mathbf{b}$.

We now relax this benchmark case by allowing for storage and borrowing possibilities as well as for inflationary shocks. Even if the assumption that there is a complete absence of storage and borrowing possibilities is relaxed, α_t cannot behave independently of y_t in the long run because long-run consumption can hardly be independent of long-run income. In the short run, however, one would still expect α_t to behave somewhat more smoothly than aggregate income. In reality, θ possesses stationary and trending components, whose combined effects are manifested in variations in y_t. It is therefore useful to write y_t as a function of a vector of trend terms of some sort, $\boldsymbol{trend}_t$, and a transitory component, u_{yt}:

(2) $y_t = y(\boldsymbol{trend}_t, u_{yt})$,

where $Eu_{yt} = 0$.

Turning to inflationary shocks, we assume that only unanticipated inflation has an effect on the distribution of incomes and thus on the distribution of expenditures. If inflation is a stationary process and the economic agents' relevant information set at time t consists of the vector $\boldsymbol{H}_t$, we can write the inflation rate as

(3) $\pi_t = \pi(\boldsymbol{H}_t, u_{\pi t})$,

where $u_{\pi t}$ represents unanticipated inflation, with $Eu_{\pi t} = 0$.

If some consumption smoothing occurs, it is also appropriate to add a lagged term to (1). A more general specification for the scale parameter of the distribution of expenditures is thus

(4) $\alpha_t = \alpha(\alpha_{t-1}, \boldsymbol{trend}_t, u_{y_t}, u_{\pi t})$.

If we relax the assumption of full insurance but keep the assumption that there is neither storage nor borrowing, non-neutral aggregate shocks to income will, in general, find their expression in a change of the shape of the consumption distribution. In the absence of such insurance, any shape parameter in the vector $\boldsymbol{\beta}_y$, and thus also in $\boldsymbol{\beta}$, depends on $\boldsymbol{\theta}$ in a complicated way. Consequently, an appropriate representation for $\boldsymbol{\beta}_t$ in this case is

(5) $\boldsymbol{\beta}_t = \boldsymbol{\beta}(\boldsymbol{trend}_t, u_{yt}, u_{\pi t})$.

Lagged values of $\boldsymbol{\beta}_t$, if added to the RHS of (5), should not be significant upon estimation; for as argued above, one would expect distributionally non-neutral, long-run changes in the environment to manifest themselves in a significant trend term, with transitory effects being explained by u_{yt} and $u_{\pi t}$.

The Data

The NSS publishes the data on per capita household consumption in grouped form, reporting both the mean nominal expenditure and the share in total population of each group.[2] Since the concern here is with movements in the distribution of real expenditures, the ideal would be a suitable price index for each expenditure group. The only available index for the entire period, however, is the consumer price index for agricultural laborers (CPIAL), who form a substantial part of the rural population and whose living standards are not much inferior to those of marginal farmers and service workers. It can be argued, therefore, that the CPIAL is a satisfactory deflator for the purposes of producing a series of real mean per capita household consumption, Ex_t.

Rural households derived their incomes overwhelmingly from agriculture or activities that depended on agricultural production, so that the relevant measure of aggregate economic performance is real agricultural value added per head of the rural population, y_t. In keeping with the deflation of the expenditure series, y_t is defined to be the level of nominal value added per head deflated by the CPIAL. As incomes and output are jointly determined in a common process, and current consumption is likely to depend on current income, any attempt to explain the movements of output and consumption runs into a potential problem of simultaneity. One plausible instrument for y_t would reflect the state of nature, especially the weather. In this connection, Lahiri and Roy (1985) have estimated an all-India index of rainfall quality with respect to rice for the period 1954-55 to 1978-79. While any index of this sort is bound to be somewhat coarse, it should be said that rice is India's most important crop. To complete the series for this index, R_t, it was necessary to experiment with various interpolations of the meteorological data in the *Statistical Abstracts* from 1979-80 onwards.

[Figure 1 about here]

The movements of Ex_t, y_t and R_t are depicted in Figure 1. The series for Ex_t fluctuated significantly, but with no apparent trend, during the period up to 1973-74; thereafter, the series is so broken until the resumption of annual surveying in 1986-87 that any attempt to draw further conclusions must be made very cautiously. In the latter year, the value of mean per capita household consumption attained a level 15 percent higher than any previously, a qualitative improvement that was maintained in subsequent years. Real per capita value added appears to possess a slight upward trend until 1973-74, the year-to-year fluctuations closely mirroring those in the rainfall index. A decade and two severe droughts later, its level was scarcely higher, but noticeable growth set in thereafter. What is striking about this series is that it is not obviously more volatile than that of mean per capita expenditure.

Characterizing the Expenditure Distributions

The maintained hypothesis is that the set of empirical distributions belongs to the three-parameter family proposed by Singh and Maddala (1976), the distribution function for which is

(6) $F_x(x) = 1 - [1 + (x/\alpha)^{\beta}]^{-\gamma}$.

Various studies have shown that within the family of three-parameter distribution functions, this functional form performs best in fitting empirical income distribution data. When applied to US income data, for example, it even outperformed such four-parameter distributions as the generalized beta of the first kind (McDonald, 1984).

It is clear from (6) that α is a scale parameter. When x is small relative to α, $F_x(x)$ is much more sensitive to changes in the shape parameter β than those in γ, and in the limit where x goes to zero, $F_x(x)$ depends only on β. For large values of x/α, however, $F_x(x) \cong 1 - (x/\alpha)^{-\beta\gamma}$, from which it is seen that the distribution has a Pareto right-hand tail, with Pareto-index $\beta\gamma$. Thus, whereas the lower tail is largely governed by β, the upper tail is determined by the product of the two shape parameters. Given α and γ, an increase in β reduces the size of both tails.

The next step is to rank such distributions according to an index of social welfare. The latter is assumed to be representable as an additively separable function of the form

(7) $W(F) = \int u(x)dF(x)$,

where $u(.)$ is an evaluation function of individual consumption. It turns out that this task is intimately related to establishing whether changes in each of the parameters α, β and γ induce shifts in $F_x(x)$ that satisfy first- or second-order stochastic dominance. The following equivalences hold: (i) F_1 is said to first-order dominate F_2 if, and only if, $W(F_1) \geq W(F_2)$ for all increasing functions u; and (ii) F_1 second-order dominates F_2 if, and only if, $W(F_1) \geq W(F_2)$ for all increasing concave functions u. Whereas $u' > 0$ implies that an increase in any individual's consumption is preferred to the status quo ante, $u'' < 0$ implies a preference for equality, in the sense that any mean-preserving, progressive transfer raises social welfare (Sen, 1997). It should be recalled that first-order stochastic dominance implies the second-order kind.

A careful distinction must be drawn between welfare, as defined by (7), and inequality. It is clear that increases in α alone yield first-order stochastic dominance, but leave the Lorenz curve, and hence inequality, unaltered. Where the shape parameters β and γ are concerned, Klonner (2000) has shown that $F(x, \beta_1, \gamma_1; \alpha)$ first-order stochastically dominates $F(x, \beta_2, \gamma_2; \alpha)$ if, and only if, $\beta_1 \geq \beta_2$ and $\beta_1\gamma_1 \leq \beta_2\gamma_2$. Where Lorenz dominance[3] is concerned, however, Wilfling and Krämer (1993) have shown that $F(x, \beta_1, \gamma_1; \alpha)$ Lorenz-dominates $F(x, \beta_2, \gamma_2; \alpha)$ if, and only if, $\beta_1 \geq \beta_2$ and $\beta_1\gamma_1 \geq \beta_2\gamma_2$.[4] Thus, where movements in the shape parameters alone are concerned, the rankings according to first-order stochastic dominance will be the reverse of those according to Lorenz-dominance, except in the case where the product $\beta\gamma$ remains constant. We were unable to find both necessary and sufficient conditions to rank distributions according to second-order stochastic dominance. A rise in β is necessary to obtain a preferred F according to this criterion; it can also be established that when γ is constant, a rise in β yields a preferred F if, and only if, $\gamma > 1 + 2/\beta$.

While increases in β and γ both yield improvements according to the Lorenz criterion, a rise in γ alone unambiguously reduces social welfare as defined by (7). A rise in β alone, however, improves social welfare so defined, provided aversion to inequality is sufficiently high.[5] So constrained, such an index of social welfare can live in peace with Lorenz dominance if an increase in β is accompanied by a sufficiently small increase in γ, and it will certainly do so if γ declines slightly. These requirements will be relaxed somewhat if α increases, since taken alone, an increase in α will yield an improvement according to first-order stochastic dominance, while leaving the Lorenz curve unchanged. It is clear that the use of summary measures of inequality such as the Gini coefficient suppresses much information when α, β and γ respond in different ways to the same underlying factors.

For each year in which the NSS conducted a survey, the values of α, β and γ were estimated by a pseudo maximum likelihood procedure (Klonner, 1999),[6] the results of which are graphed in Figure 2. It is seen that the scale parameter α exhibited some volatility at the beginning and the end of the entire period, but otherwise moved relatively smoothly and in a narrow range, which suggests a certain underlying stability in the distribution of consumption. Over the period as a whole, there is a hint of an upward drift in the series, and of a positive correlation with output per head. The behavior of β is similar, except that there is a more of a jump rather than an increase in volatility at the end of the period. In contrast, γ appears to decline somewhat over the period as a whole, with fluctuations that run counter to those in β. The product of β and γ rose fairly steadily at first, reaching an overall peak in 1967-68. There followed an erratic decline, and the series closed with fluctuations around a level a little above that ruling in the early years.

[Figure 2 about here]

The Econometric Results

We now turn to the estimation of the system of equations (2), (3), (4) and (5). Using a linear specification of (2), we found that an appropriate model to generate a stationary series of residuals $\hat{u}_{y_t}$ is to regress y_t not only on a uniform trend, but also on an additional trend term representing an acceleration of productivity growth from a certain date onwards. Using the standard method of determining a structural break endogenously,[7] we determined that accelerated growth started in 1982-83. For the estimation, we therefore specify (2) as

(8) $y_t = f_0 + f_1 \cdot trend_t + f_2 \cdot trend1_t + u_{yt},$

where *trend* is a term that serves as a catch-all for various secular factors, and *trend*1 reflects the fact of accelerated productivity growth since the early 1980's. To be precise, *trend*1 = 0 before 1982-83 and $trend1_t = trend_t - trend_{1981\text{-}82}$ afterwards.

Where the inflation equation (3) is concerned, the two most important variables in the agents' information set appear to be lagged values of the weather conditions and transitory shocks to income (Bell and Rich, 1994). We thus set $\boldsymbol{H}_t = \{R_{t-1}, u_{y_{t-1}}\}$. Solving (8) for u_{y_t} and substituting this into a linearized version of (3), we obtain

(9) $\pi_t = \delta_0 + \delta_1 R_{t-1} + \delta_2 \hat{u}_{y_{t-1}} + u_{\pi t}$,

where $\hat{u}_{y_t} \equiv y_t - (f_0 + f_1 \cdot trend_t + f_2 \cdot trend1_t)$.

The long gaps between 1973-74 and 1986-87 make estimation of the equations for the parameters α_t, β_t and γ_t especially awkward. In order to deal with the missing observations in this interregnum, we discarded the entire sub-period and spliced 1987-88 onto 1973-74, while retaining the values for 1986-87 as the one-period lagged values for 1987-88. There are also two missing observations in the earlier period, namely, 1962-63 and 1971-72, the values of which were estimated by linear interpolation in order to preserve degrees of freedom for all stages of the estimation.[8]

In the benchmark case (see eq. (1)), the scale parameter α is proportional to y. Hence, we chose a specification for α in levels, while taking the natural logarithm of the shape parameters β and γ (see also Thurow, 1970). Observe that the influence of current output on α is represented not directly, but rather through the trend terms and the shock $\hat{u}_{y_t}$, the constant f_0 being absorbed into a_0. Note that the vector of shape parameters $\boldsymbol{\beta}$ in (5) consists of the parameters β and γ in (6). The econometric specification of (4) and (5) is thus

(10) $\alpha_t = a_0 + a_1 \alpha_{t-1} + a_2\, trend_t + a_3\, trend1_t + a_4 \hat{u}_{y_t} + a_5 \hat{u}_{\pi_t} + u_{\alpha t}$,

(11) $\log\beta_t = b_0 + b_1 \log\beta_{t-1} + b_2\, trend_t + b_3\, trend1_t + b_4 \hat{u}_{y_t} + b_5 \hat{u}_{\pi_t} + u_{\beta t}$,

(12) $\log\gamma_t = c_0 + c_1 \log\gamma_{t-1} + c_2\, trend_t + c_3\, trend1_t + c_4 \hat{u}_{y_t} + c_5 \hat{u}_{\pi_t} + u_{\gamma t}$,

where $\hat{u}_{\pi_t} \equiv \pi_t - (\delta_0 + \delta_1 R_{t-1} + \delta_2 \hat{u}_{y_{t-1}})$ is interpreted as the unanticipated component of inflation. Lagged terms of the shape parameters have been added to equations (11) and (12) to test whether non-neutral shifts in the distribution of consumption are persistent. As argued above, the associated coefficients b_1 and c_1 should be insignificantly different from zero upon estimation. Note that the benchmark model, in which there is complete insurance within each period, but neither borrowing nor storage across periods, corresponds to the hypothesis that

$$H_0 : a_1 = a_5 = b_1 = b_2 = b_3 = b_4 = b_5 = c_1 = c_2 = c_3 = c_4 = c_5 = 0;$$

$$f_0/ f_1 = a_0/ a_2, f_0/ f_2 = a_0/ a_3 \text{ and } f_0 = a_0/a_4.$$

Because output and the distribution of income are determined in a simultaneous process, one must beware of a potential endogeneity bias arising from the use of $\hat{u}_{y_t}$ as a regressor when the system (8) - (12) is estimated by least squares. We therefore use Hansen's (1982) Generalized Method of Moments (GMM). This procedure not only offers a useful solution to the problem of simultaneity bias, but also yields consistent estimates of both the parameters and their covariances even if the disturbance terms are conditionally heteroskedastic. It also permits the specification to be tested through the use of over-identifying restrictions.

Hansen's (1982) asymptotic theory for stationary variables has been extended to trend-stationary ones by Andrews and McDermott (1995),[9] whose asymptotic theory will be applied here. Further, although the expenditure data are missing for a number of years, we want to retain the information for those years where equations (8) and (9) are concerned. Hence, we use a slight modification of the GMM procedure suggested by Klonner (2001), which accounts for partially missing data. Finally, based on some recent results (Donald and Newey, 2000; Hansen, Heaton and Yaron, 1996), we use the so-called continuous updating GMM-estimator.

The selection of the orthogonality conditions for each of the system's equations, Z_{vt}, where $v = y, \pi, \alpha, \beta, \gamma$, was based on the consideration that the instruments should be correlated with the regressors but uncorrelated with the equation's disturbance term.[10] Note that by including $\hat{u}_{\pi_t}$ in $Z_{\alpha t}$, $Z_{\beta t}$ and $Z_{\gamma t}$, we implicitly assume that inflation affects the distribution of expenditures but not vice versa. Here, the number of parameters to be estimated is large compared to the number of observations. We therefore limit the number of over-identifying restrictions to one per equation. This decision balances two considerations: first, at least one orthogonality condition per parameter is required for the sake of identification; and second, the number of observations should be substantially larger than the number of orthogonality conditions to justify the use of asymptotic theory.

[Table 1 about here]

We begin by noting that where the over-identifying restrictions of the model are concerned, the specification is satisfactory, in that the values of the chi-square statistic fall far short of conventional critical levels. At the same time, the benchmark model cannot be rejected (p-value = 0.33), the drastic reduction in the number of estimated parameters (15 fewer under the null hypothesis) being enough to offset the reduction in the test statistic.[11] Whether this rather striking finding should be accepted is an open question, for the sample is small and the power of this test, which is based on the likelihood-ratio principle, may be low. In the light of these considerations, we incline to the unrestricted model.

Beginning with the decomposition of y into time trends and a stationary component, the estimates of f_1 and f_2 are sharp. The inflation equation is rather less successful: lagged rain is only borderline significant, while shocks to lagged output appear to play no role in predicting inflation when rainfall is taken into account. That the coefficient on R_{t-1} is negative indicates that good weather in the previous year has a dampening effect on inflationary expectations in the present. This is a plausible finding, in as much as lagged rainfall is correlated with the lagged value of realized output, and so speaks for the importance of supply effects.

Turning to the scale parameter α, whereas there is clear evidence of persistence, transitory shocks to output do not appear to affect the distribution in a neutral way.

Unanticipated inflation, however, has a significant downscaling effect, which suggests that instead of full smoothing, households took some precautionary steps to reduce consumption, at least temporarily. The secular term, *trend*, has a positive (albeit borderline) coefficient, as expected. The term *trend*1, however, is not significant.

Where the shape parameters β and γ are concerned, there is no evidence of persistence, the coefficients b_1 and c_1 being insignificant at conventional levels, a result that is in accordance with the theoretical argument above. Interestingly, the effects of output shocks and unanticipated inflation on the distribution of expenditures run in the same direction, namely, of increasing β and decreasing γ. The coefficients b_5, c_4 and c_5 are significant at the five percent level, and b_4 is just borderline.

Combining these findings, we conclude that unexpectedly good harvests, i.e., positive output shocks, yield distributions which are preferred according to the criterion of first-order stochastic dominance; for β increases and the algebraic sum of b_4 and c_4 is negative, so that the product $\beta\gamma$ decreases. The positive, though statistically insignificant, estimate of a_4 works in the same direction through the scale parameter. Where the Lorenz-curve is concerned, no ranking is possible because an increase in β together with a decrease in $\beta\gamma$ implies that the corresponding Lorenz-curves intersect, the left section of the Lorenz-curve moving up and the right down.

Where the effects of unanticipated inflation are concerned, the picture is not immediately clear. While the response of the shape parameters to such shocks points to first-order stochastic dominance, the effect on the scale parameter is now statistically significant and pulls in the opposite direction. Evaluated at the sample means, we find that the mean of the expenditure distribution decreases by about 0.12% in response to an unanticipated inflationary shock of 1%, which outweighs the latter's effect on β and $\beta\gamma$. It must be said, however, that, for very small values of expenditure, the distribution function F shifts down in response to $\hat{u}_{\pi_t}$ by virtue of the positive impact $\hat{u}_{\pi_t}$ has on β. Thus, for some (possibly extreme) values of 'inequality-aversion' the distribution thus generated is preferred. Where the Lorenz-criterion is concerned, there is crossing, as before.

Turning to the trend terms, the evidence points to the presence throughout the period of secular forces that operated in favor of greater equality; for the coefficient on the term *trend* in the β-equation is positive and significant, whereas the corresponding coefficient in the γ-equation is statistically insignificant, which implies Lorenz-dominance. On the other hand, the term *trend*1 appears to work exclusively through γ, with a negative coefficient, which indicates that the accelerated growth of productivity since the early 1980s made for increasing inequality. No clear conclusions can be drawn in respect of first-order stochastic dominance, since a rise in β generates intersecting distribution functions. Nevertheless, the resulting changes in the distribution tend to be preferred under that criterion. The main effect is yielded by the positive (though only borderline significant) coefficient a_2, and when we account for both a_2 and b_2, we find that mean consumption rises about 0.18% per year in response to *trend*.

To summarize, first, the benchmark case, in which growth is distributionally neutral, cannot be rejected on the basis of the likelihood ratio test. In the unrestricted model, however, the prediction that transitory changes in output have an impact on the scale of the distribution of consumption through α does not find empirical support in these data. Unanticipated inflation, on the other hand, does have a statistically significant influence on this parameter. Second, while aggregate consumption appears to be far from smooth, there is strong evidence of persistence, so that the hypothesis of a complete absence of storage or borrowing facilities is hard to maintain. Third, it is interesting to note that while there appears to have been one set of equalizing secular forces at work over the whole sample period, which manifest themselves in the form of an upward trend in β, there is also another set operating through γ

that has ‘pulled’ against this equalizing trend from the early 1980s onwards. Focusing on the permanent forces expressed by *trend* and *trend*1 only, the resulting distributions cannot be Lorenz-ranked.[12] Employing the Gini coefficient as a summary measure of inequality, the effect of the decrease in γ outweighs the rise in β: the predicted value of the Gini coefficient rises at about 0.4 % per year from 1981-82 onwards, compared to a decrease of about 0.2 % per annum before that.[13] The brighter side of this coin is that the corresponding distributions so induced are preferred according to first-order stochastic dominance.[14] This latter development could be a consequence of accelerated technological progress, as well as of revised government policies, since the mid-eighties.

Conclusions

The parametric approach to modeling the distribution of expenditures has served the purposes of this paper well in two ways. First, purely descriptively, the Singh-Maddala three-parameter family appears to describe the data very satisfactorily. Second, changes in each parameter, when taken separately, admit of precise interpretations in terms both of first- and second-order stochastic dominance and of Lorenz-dominance.

The task of explaining the observed movements in these parameters is hindered by the shortness of the series, and any specification is bound to be parsimonious. A simple model of production combined with the assumption of complete insurance against idiosyncratic shocks provides a useful benchmark case, which cannot be rejected at standard levels. Given the smallness of the sample, however, there are grounds to extend the model to allow for the effects of unanticipated shocks to output and inflation. In this variant, the first finding is that the scale of the distribution exhibits significant persistence, which suggests that some inter-temporal smoothing was taking place. The second finding is that the shape of the distribution exhibits no persistence; but is influenced by unanticipated shocks as well as secular factors in the form of trend terms. Taken together, these findings are at odds with the result from the benchmark case that agricultural growth was distributionally neutral. Long-term departures from neutrality, such as they were, appeared to be the outcome of a tug of war between some secular factors that pulled one shape parameter in the direction of greater equality throughout and other forces that were making for more inequality through the second parameter from the early 1980s onwards. In contrast, short-term shocks were unambiguously non-neutral, good harvests (adjusted for trend) inducing Lorenz-crossing, though they are clearly preferred under first-order stochastic dominance.

How does the record appear in the light of these findings? Given that departures from neutrality were fairly modest throughout, the slow growth in productivity up to about 1980 must be judged as unsatisfactory, while the subsequent acceleration is to be applauded. It can be argued also that the mechanisms for the smoothing of consumption within and across periods are in need of improvement. Yet it can hardly be claimed that the record as a whole is unsatisfactory, let alone a dismal failure. In particular, strong claims that agricultural growth took an inherently unequalizing form over the long run is, on the above evidence, unwarranted: at most, the Scottish verdict of not proven should be returned.

Endnotes

[1] This is not to claim, of course, that substantial fluctuations would not have occurred under a policy of laissez-faire.

[2] Most of the data used in this paper are taken from Özeler et al. (1996). A table of the data and sources actually used is available from the authors upon request.

[3] $F_1(x)$ is said to Lorenz-dominate $F_2(x)$ if the Lorenz-curve corresponding to F_1 lies uniformly not below that corresponding to F_2 (Cowell, 2000). In this paper, F_1 is said to exhibit less inequality than F_2 if, and only if, F_1 Lorenz-dominates F_2. That is to say, the term inequality is always used in the relative sense.

[4] In contrast to Cowell's (2000) definition, which is adopted here, Wilfling and Krämer (1993) follow a common convention in mathematical statistics and define F_1 to Lorenz-dominate F_2 if the Lorenz-curve corresponding to F_2 lies uniformly not below that corresponding to F_1. We have stated Wilfling and Krämer's theorem based on Cowell's (2000) definition (see Footnote 3).

[5] To illustrate, consider the isoelastic specification $u(x) = x^{1-\delta}/(1-\delta)$. Then $\partial W(F;\delta)/\partial\beta > 0$ if, and only if, $\gamma > 1 + 2(1-\delta)/\beta$. In the Ramsey case ($\delta = 2$), this condition holds for most of the years in the present data.

[6] The estimation of parameters from grouped data commonly uses only group frequencies. Since the NSS also reports the means within each group, Klonner (1999) has exploited this additional information by deriving the asymptotic joint density of said means and the frequencies. The estimates are obtained by maximizing the logarithm of this density with respect to the parameters. Klonner (1999) also shows that the resulting estimates are always more efficient than the ones based on group frequencies alone.

[7] We tried different starting dates for the term $trend1_t$ in (8) and took the one that minimized the resulting sum of squared residuals.

[8] We also made an attempt to interpolate the latter two gaps using Kalman filter methods (see Harvey and Pierse, 1984). Employing only observations of the period from 1954-55 to 1973-74 we could not find a state space model that fitted the data sufficiently well to obtain significant estimates for the smoothing algorithm. On the other hand, we did not use the observations from 1986-87 onwards for this purpose, since we did not want to rule out the possibility of a structural break during the period 1974-75 to 1985-86.

[9] Applying so-called bounded trend asymptotics, their main result is that, given some regularity conditions hold, Hansen's (1982) asymptotic theory for the stationary case can also be applied to trend-stationary data.

[10] The orthogonality conditions are specified at the foot of Table 1.

[11] We test H_0 by estimating the unrestricted and restricted model separately. The difference between the values of the respective objective functions is asymptotically distributed as chi-square with 15 degrees of freedom (see Hansen, Heaton and Yaron, 1996).

[12] See the result on Lorenz-dominance above.

[13] This computation is based on the estimated coefficients b_2 and c_3 of Table 1.

[14] See the result on first-order stochastic dominance above.

References

Ahluwalia, M. S. (1978). "Rural Poverty and Agricultural Performance in India", Journal of Development Studies, 14: 298-323.

Andrews, D. W. K. and McDermott, C. J. (1995). “Nonlinear Econometric Models with Deterministically Trending Variables”, Review of Economic Studies, 62:343-360.

Bell, C. and Rich, R. (1994). "Rural Poverty and Aggregate Economic Performance in Post-Independence India", Oxford Bulletin of Economics and Statistics, 56: 111-33.

Cowell, F. (2000). "Measurement of Inequality", in Atkinson, A.B. and F. Bourguignon eds. Handbook of Income Distribution, vol. 1, Amsterdam: North Holland, pp. 87-166.

Deaton, A. (1992). Understanding Consumption, Oxford: The Clarendon Press.

Donald, G. D. and Newey, W. K. (2000). “A Jackknife Interpretation of the Continuous Updating Estimator”, Economics Letters, 67: 239-243.

Fan, S., P. Hazell and S. Thorat (2000). “Government Spending, Growth and Poverty in Rural India”, American Journal of Agricultural Economics, 82(4): 1038-1051.

Gaiha, R. (1989). "Poverty, Agricultural Production and Price Fluctuations in Rural India: a Reformulation", Cambridge Journal of Economics, 13: 333-352.

Hansen, L.-P. (1982). “Large Sample Properties of Generalized Method of Moments Estimators”, Econometrica, 50: 1029-1054.

Hansen, L.-P., Heaton, J., Yaron, A. (1996). “Finite-sample properties of some alternative GMM estimators”, Journal of Business and Economics Statistics, 14: 262-280.

Harvey, A.C. and Pierse, R.G. (1984). "Estimating Missing Observations in Economic Time Series", Journal of the American Statistical Association, 79: 125-131.

Joshi, V., and Little, I. M. D. (1994). India: Macroeconomics and Political Economy 1964-1991, Oxford: The Clarendon Press.

Klonner, S. (1999). "An Efficient Estimator for Grouped and Averaged Data", manuscript, Universität Heidelberg.

Klonner, S. (2000). ”The First-order Stochastic Dominance Ordering of Singh-Maddala Distributions”, Economics Letters, 69, 2: 123-128.

Klonner, S. (2001). “Generalized Method of Moments Estimation with Partially Missing Data”, manuscript, Universität Heidelberg.

Lahiri, A. and Roy, P. (1985). "Rainfall and Supply Response: A Study of Rice in India", Journal of Development Economics, 18: 315-34.

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

Özeler, B., Datt, G. and Ravallion, M. (1996). A Database on Poverty and Growth in India. Poverty and Human Resources Division, Policy Research Department, The World Bank.

Ravallion, M., and Datt, G. (1996). "How Important to India's Poor is the Sectoral Composition of Economic Growth?", World Bank Economic Review, 10: 1-26.

Saith, A. (1981). "Production, Prices and Poverty in Rural India", Journal of Development Studies, 17: 196-214.

Sen, A. (1997). On Economic Inequality, 2nd edition. Oxford: The Clarendon Press.

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

Townsend, R. M. (1994). "Risk and Insurance in Village India", Econometrica, 62: 539-591.

Thurow, L. C. (1970). ”Analyzing the American Income Distribution”, Papers and Proceedings, American Economics Association, 60, 261-269.

Wilfling, B. and W. Krämer (1993). ”The Lorenz-ordering of Singh-Maddala Income Distributions”, Economics Letters, 43: 53-57.

Table 1. GMM estimation results for the system of equations (8) - (12)

y	Parameter	Estimate	s.e.	t-Ratio
Constant	f_0	161.6194	2.9793	54.2483
trend	f_1	1.1293	0.1490	7.5776
*trend*1	f_2	4.7717	0.5374	8.8795
π				
Constant	δ_0	0.4410	0.1912	2.3063
lagged R	δ_1	-0.0038	0.0020	-1.9165
lagged $\hat{u}_y$	δ_2	0.0012	0.0015	0.8062
α				
Constant	a_0	8.0379	1.6907	4.7543
lagged α	a_1	0.3777	0.1291	2.9252
trend	a_2	0.0586	0.0326	1.7974
*trend*1	a_3	0.0271	0.0827	0.3278
$\hat{u}_y$	a_4	0.0198	0.0150	1.3188
$\hat{u}_\pi$	a_5	-5.6360	2.4824	-2.2704
$\log\beta$				
Constant	b_0	1.0382	0.1817	5.7129
lagged $\log\beta$	b_1	0.1714	0.1477	1.1609
trend	b_2	0.0043	0.0020	2.1608
*trend*1	b_3	0.0071	0.0051	1.3781
$\hat{u}_y$	b_4	0.0026	0.0014	1.9014
$\hat{u}_\pi$	b_5	0.3831	0.1571	2.4385
$\log\gamma$				
Constant	c_0	-0.3209	0.0595	-5.3943
lagged $\log\gamma$	c_1	0.0295	0.1346	0.2190
trend	c_2	0.0032	0.0025	1.2504
*trend*1	c_3	-0.0339	0.0096	-3.5398
$\hat{u}_y$	c_4	-0.0051	0.0020	-2.5053
$\hat{u}_\pi$	c_5	-0.5074	0.2561	-1.9814
$\chi^2_{(5)}$	3.21	p-value	0.77	

$Z_{yt} = \{1, trend_t, trend1_t, R_{t-2}\}$

$Z_{\pi t} = \{1, R_{t-1}, R_{t-2}, \hat{u}_{y_{t-1}}\}$

$Z_{\alpha t} = \{1, \alpha_{t-1}, trend_t, trend1_t, R_t, R_{t-2}, \hat{u}_{\pi_t}\}$

$Z_{\beta t} = \{1, \log\beta_{t-1}, trend_t, trend1_t, R_t, R_{t-2}, \hat{u}_{\pi_t}\}$

$Z_{\gamma t} = \{1, \log\gamma_{t-1}, trend_t, trend1_t, R_t, R_{t-2}, \hat{u}_{\pi_t}\}$

Figure 1. Annual mean per capita consumption, Ex_t, per capita value added in agriculture, y_t, rainfall, R_t, and inflation of the CPIAL in percent, π_t

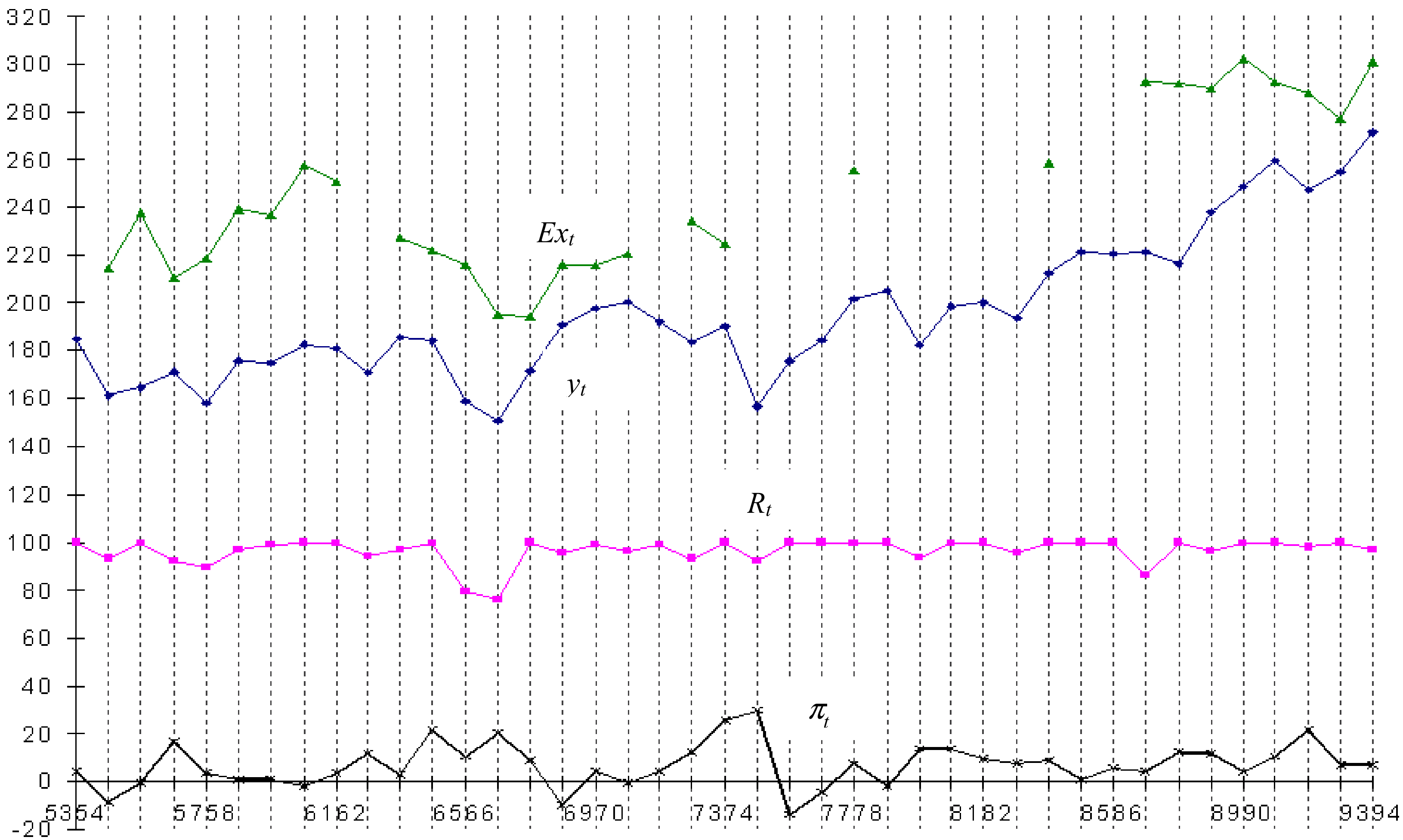
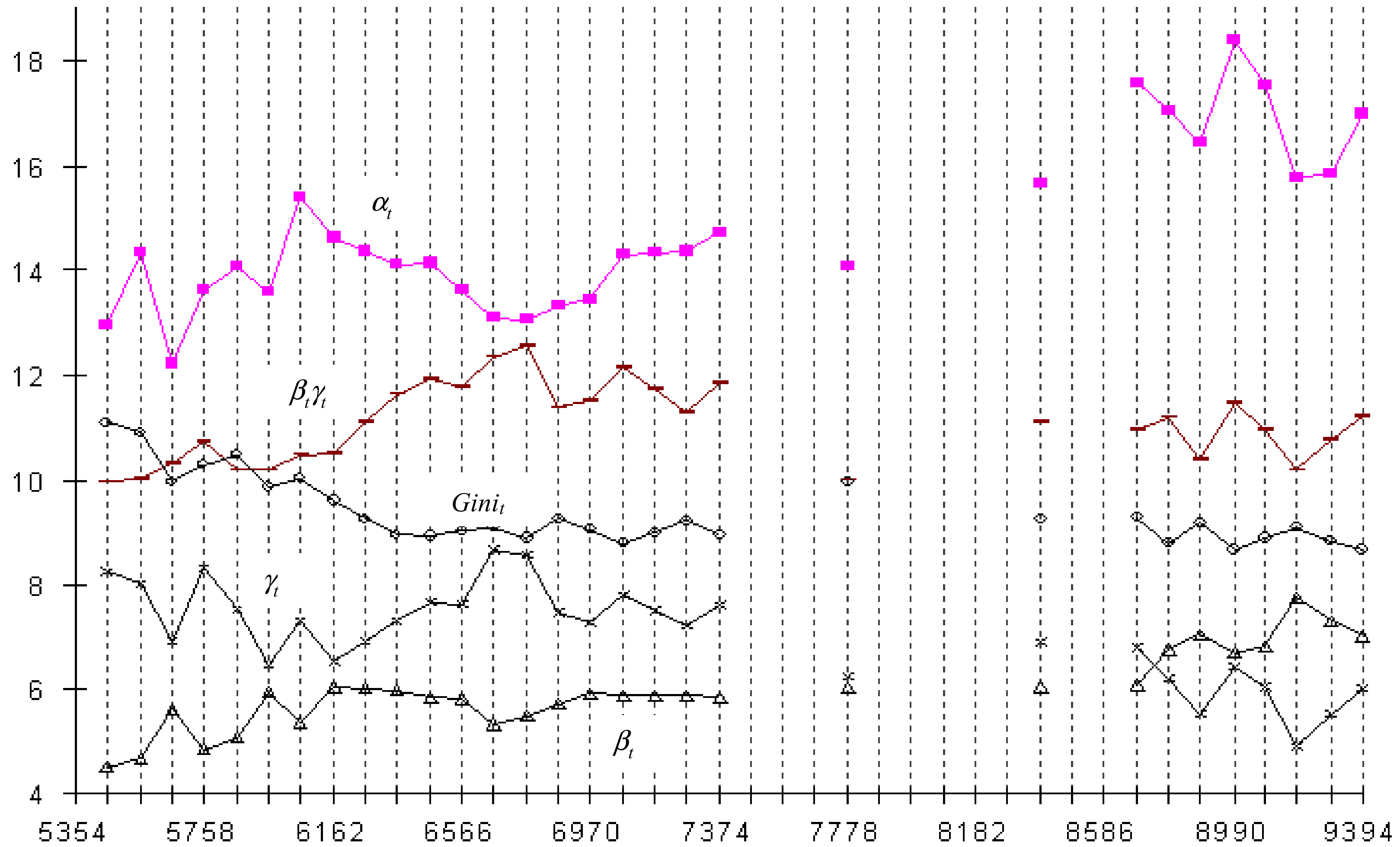


Figure 2. Estimated parameters of the Singh-Maddala distribution, α, β and γ, estimated Pareto index, $\beta\gamma$, and estimated Gini coefficient*



* to make the graph more accessible, the series of β, γ, the Pareto index and the Gini coefficient have been multiplied by 1.5, 10, 4 and 30, respectively.