Analysis of inequality in fertility curves fitted by gamma distributions

Héctor M. Ramos, Antonio Peinado, Jorge Ollero and María G. Ramos[*]

Abstract

The aim of this paper is to analyse fertility curves from a novel viewpoint, that of inequality. Through sufficient conditions that can be easily verified, we compare inequality, in the Lorenz and Generalized Lorenz sense, in fertility curves fitted by gamma distributions, thus achieving a useful complementary instrument for demographic analysis. As a practical application, we examine inequality behaviour in the distributions of specific fertility curves in Spain from 1975 to 2009.

1. Introduction

The basic concept of inequality arises in many and diverse fields, and so it is difficult to provide a brief definition that will command universal acceptance. More specific contexts give rise to different versions of the concept that can be defined indirectly if we assume certain comparative criteria. Roughly speaking, inequality is a particular aspect of variability when the variables considered are nonnegative and represent quantities that can be transferred from one unit to another. Champernowne and Cowell (1998) provide a convenient reference on this topic. Several studies have approached the problem of ranking distributions by seeking a dominance relationship between concentration curves. In this context, the Lorenz curve and the Generalized Lorenz curve have been used to compare two income distributions in terms of inequality.

The purpose of this paper is to show that the partial orderings of distributions induced by such curves can provide a useful instrument for the demographic analysis of fertility curves. Therefore, our particular interest lies in the concepts of inequality

underlying the comparison of probability distributions using the Lorenz curve and the Generalized Lorenz curve. The first of these, strictly speaking, is an order of inequality (concentration), whereas the second, when the variables being compared represent incomes, is considered a welfare ordering (see Arnold et al. (1987) and Lambert (2001)). Furthermore, Ramos and Sordo (2002) showed that the Generalized Lorenz order is equivalent to the increasing concave order (Stoyan, 1983). The main results for stochastic orders can be found in Shaked and Shanthikumar (2007).

The aim of this study is to consider and analyze fertility curves from a new viewpoint, that of inequality. Through sufficient conditions that can be easily verified, we compare inequality, in the Lorenz and Generalized Lorenz sense, in fertility curves fitted by gamma distributions. The age-specific fertility rates f_x^t for each maternal age (X) and each year (t) are conventionally defined as the ratio of the number of births to women (x) years of age and the population of women of the same age at the midpoint of year (t). For each year, the observed series of age-specific fertility rates can be fitted (Duchêne and Gillet de Stefano, 1974) using the following curve:

$$g(x) = \frac{a\beta^{-\alpha}(x-\theta)^{\alpha-1}\exp[-(x-\theta)/\beta]}{\Gamma(\alpha)} \tag{1}$$

where $\Gamma(\cdot)$ denotes the complete gamma function and where $a = SFI^{(t)} = \sum_{x=15}^{49} f_x^t$ is the Synthetic Fertility Index (SFI), in which the summation extends from 15 to 49 years, the bounds being the woman's fertile period, with the value of 15 assigned when the mother is aged 15 years or younger and 49 when aged 49 years or older. Then, $(x-\theta)$ is the class mark of the age interval considered less the minimum fertile age; that is, $x - \theta = x + 0.5 - 15 = x - 14.5$. Expression (1) enables us to compare series corresponding to different years and to analyze behaviour over a broader time span. Abad et al. (2006) used this approach to fit fertility curves in the Andalusia region in southern Spain.

In the present study, we are not interested in the values of age-specific fertility rates, rather in analyzing the inequality present in the corresponding vectors. Neither are we interested in numerically quantifying inequality, which could be done using various standard measures, such as those associated with the Lorenz curve. On the contrary, our interest lies in comparing in absolute terms (whatever the specific measure applied) the inequality corresponding to two different years within a given population and the inequality corresponding to two different populations in a single period of time. To fulfill this aim, we do not consider age-specific fertility rates, rather the quotients $g_x^t = f_x^t(SFI)^{-1}$. In this way, the corresponding fitting curve $g(x)$ is the density function of a Gamma distribution (α, β, θ), with $\theta = 14.5$:

$$g(x) = \frac{\beta^{-\alpha}(x-\theta)^{\alpha-1}\exp[-(x-\theta)/\beta]}{\Gamma(\alpha)}, \quad x > \theta, \ \alpha > 0, \ \beta > 0, \ \theta = 14.5. \tag{2}$$

2. Results

The study of the distribution of fertility according to mothers' age is a scenario that can be easily transferred to the context of income distributions. Thus, it is only necessary to consider specific rates of fertility as the "income" contribution of a given age group of women to the "wealth" of the community in terms of the birth of new population members. This similarity enables us to approach the study of fertility curves from a new perspective. For the analysis and comparison of inequality, let us first employ the Lorenz order.

The Lorenz curve of any income distribution is the graph of the fraction of the total income owned by the lowest p–th fraction $(0 \leq p \leq 1)$ of the population as a function of p. If a nonnegative random variable X represents the income of a community, with distribution function $F_X(x)$ and finite expectation μ_X, then the Lorenz curve $L_X(p)$ is given by (Gastwirth, 1971):

$$L_X(p) = \mu^{-1} \int_0^p F_X^{-1}(t)dt, \quad 0 \leq p \leq 1,$$

where F_X^{-1} denotes the inverse of F_X:

$$F_X^{-1}(a) = \inf\{x : F_X(x) \geq a\}, \, a \in [0,1].$$

The Lorenz curve can be used to define a partial ordering as:

$$X \leq_L Y \Leftrightarrow L_X(p) \geq L_Y(p) \text{ for every } 0 \leq p \leq 1.$$

In this case, we can say that X does not show more inequality than Y (in the Lorenz sense). While for any finite population there is no problem in evaluating Lorenz curves, for a continuous distribution, a simple closed form for these curves is rarely available. In our case, the analytical difficulties involved in comparing two gamma distributions by means of the Lorenz order are overcome by taking into account that this order is invariant to scale transformation (i.e., it does not depend on β) and by applying the following sufficient condition (Arnold et al., 1987):

$$\text{Let } X_1 \sim \text{gamma}(\theta, \alpha_1) \text{ and } X_1 \sim \text{gamma}(\theta, \alpha_2) \, (\theta \text{ fixed}). \text{ Then,}$$

$$\alpha_1 \leq \alpha_2 \implies X_2 \leq_L X_1. \tag{3}$$

The definition of the Generalized Lorenz curve $GL_X(p)$ corresponding to the nonnegative random variable X with distribution function F_X defined by Shorrocks (1983) is:

$$GL_X(p) = \int_0^p F_X^{-1}(t)dt, \quad p \in [0,1]. \tag{4}$$

Consequently, scaling up the Lorenz curves to form the Generalized Lorenz curves will often reveal a dominance relationship that is not apparent from an examination of the means and Lorenz curves on their own. The Generalized Lorenz curve can be used to define a partial ordering on the class of nonnegative random variables as:

$$X \leq_{GL} Y \Leftrightarrow GL_X(p) \geq GL_Y(p) \text{ for all } p \in [0,1]. \tag{5}$$

We then say that X exhibits less inequality than Y in the Shorrocks (or Generalized Lorenz) sense. Generalized Lorenz ordering reflects a desire for both greater equality and higher mean values. Kleiber and Krämer (2003) made a detailed study of the decomposition of the Generalized Lorenz order for both components. Some results on this ordering can be found in Ramos et al. (2000). Once again, the analytical difficulties arising from comparing two gamma distributions by means of the Generalized Lorenz order are overcome by using the following result (Ramos et al., 2000):

$$\text{Let } X_i \sim \text{gamma}\,(\alpha_i, \beta_i) \;\; (i : 1, 2).\text{ If } \alpha_1 \leq \alpha_2 \text{ and } \alpha_1\beta_1 \leq \alpha_2\beta_2, \text{ then } X_2 \leq_{GL} X_1. \tag{6}$$

These sufficient conditions enable us to readily compare, from the standpoint of inequality in the Generalized Lorenz sense, distributions of age-specific fertility rates, normalized and fitted by gamma distributions.
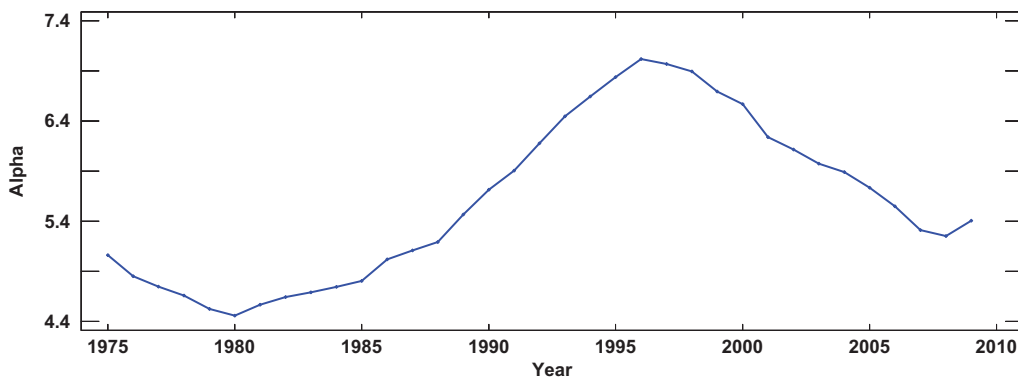
   In the following section as a practical application, we study the behaviour of inequality in the distributions of normalized specific fertility rates in Spain from 1975 to 2009. The analysis of fertility curves from the standpoint of inequality provides a tool that usefully complements demographic analysis based solely on the behaviour of the SFI, as the latter sometimes fails to detect certain situations of interest, as shown in the Conclusions section.

3. Application to Spanish data

Using official data (INE, 2010) for age-specific fertility rates in Spain for each maternal age and year from 1975 until 2009, we fitted the corresponding normalized rates by gamma (α, β, θ) distributions, with $\theta = 14.5$ using the maximum likelihood method to estimate the parameters. As shown in Table 1 and Figure 1, the α parameter decreases up until 1980, increases from 1980 to 1996 and then decreases again after 1996. According to the sufficient condition (3) (Arnold et al., 1987), we show that inequality (in the Lorenz sense) correspondingly increases in the first and third periods and decreases in the second period.

Table 1: *SFI and estimated* α, β *parameters, 1975–2009.*

Year	SFI	α	β	$\alpha \cdot \beta$	Year	SFI	α	β	$\alpha \cdot \beta$
1975	2.799	5.059	2.728	13.800	1993	1.266	6.448	2.242	14.459
1976	2.799	4.850	2.786	13.512	1994	1.202	6.646	2.215	14.722
1977	2.671	4.745	2.830	13.427	1995	1.173	6.838	2.188	14.963
1978	2.550	4.658	2.865	13.349	1996	1.160	7.020	2.163	15.186
1979	2.370	4.524	2.926	13.238	1997	1.173	6.969	2.206	15.375
1980	2.213	4.457	2.962	13.202	1998	1.153	6.895	2.254	15.541
1981	2.035	4.565	2.897	13.225	1999	1.191	6.695	2.339	15.658
1982	1.940	4.643	2.870	13.323	2000	1.231	6.570	2.394	15.725
1983	1.797	4.690	2.851	13.370	2001	1.241	6.239	2.526	15.758
1984	1.726	4.742	2.831	13.424	2002	1.259	6.115	2.582	15.792
1985	1.640	4.803	2.801	13.454	2003	1.306	5.975	2.652	15.844
1986	1.556	5.019	2.695	13.528	2004	1.325	5.890	2.694	15.871
1987	1.495	5.107	2.656	13.562	2005	1.341	5.735	2.774	15.912
1988	1.449	5.191	2.615	13.574	2006	1.377	5.549	2.865	15.896
1989	1.398	5.468	2.509	13.718	2007	1.392	5.312	2.982	15.838
1990	1.361	5.715	2.425	13.857	2008	1.459	5.252	3.015	15.833
1991	1.328	5.905	2.377	14.036	2009	1.394	5.404	2.969	16.046
1992	1.316	6.178	2.306	14.247					

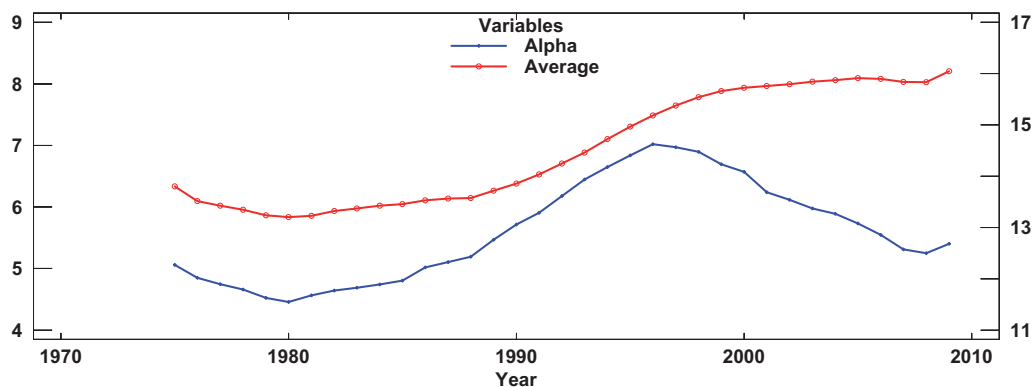**Figure 1:** *Estimated* α *parameters, 1975–2009.*

We can account for the decrease in inequality from 1980 as a consequence of greater birth control and the increased entry of women into the workplace. This meant that the maternal age increased and therefore became less concentrated. This trend was interrupted after 1996. We could explain this with the impact of major immigration into Spain during the previous decade, a population movement that contributed a substantial number of young women, most of whom arrived from countries with cultural backgrounds tending to favour maternity at a younger age. This means that a greater contribution to

births was concentrated in younger age groups (in the case of migrants), and in older groups within the native population, thus increasing inequality in accordance with the concept implicit in the Lorenz order.

From 2002 onward, the Spanish Institute of Statistics (INE, 2010) provides data broken down into the Spanish population and the foreign population. Thus, we can obtain the corresponding values for the estimated parameters (Table 2). Note that, for each year, the value of the α parameter is clearly higher for the population of Spanish origin. According to the sufficient condition of Arnold et al. (1987), the inequality in the distribution of fertility by age groups can then be said to be higher among foreign women. Once again, we can account for this by the higher concentration of fertility among younger age groups within the foreign (mainly immigrant) population. However, in this population and by taking into account the SFI values, the fertility rate appears to be decreasing. We could interpret this as the gradual assimilation of the cultural references of the host country with respect to birth rates.

Table 2: *SFI and estimated values of* α, β *parameters for Spanish and foreign populations.*

	Spanish				Foreign		
Year	SFI	α	β	Year	SFI	α	β
2002	1.209	6.852	2.350	2002	2.047	3.028	4.229
2003	1.256	6.883	2.356	2003	1.901	2.949	4.363
2004	1.275	6.862	2.376	2004	1.792	3.095	4.178
2005	1.296	6.824	2.402	2005	1.703	2.997	4.338
2006	1.331	6.656	2.472	2006	1.696	3.070	4.208
2007	1.328	6.501	2.540	2007	1.750	3.103	4.160
2008	1.382	6.441	2.571	2008	1.813	3.233	4.045
2009	1.331	6.528	2.563	2009	1.671	3.422	3.930

Figure 2: *Estimated* α *parameters and* $\alpha\beta$ *mean values, 1975–2009.*

Otherwise, our analysis of fertility curves could be approached from the standpoint of inequality in the sense of the Generalized Lorenz order. To do this, we employ sufficient condition (6) (Ramos et al., 2000). In this case, we must take into account not only the α values but also the $\alpha\beta$ mean values of the distributions being compared (Table 1). We can see that during the periods 1975–80 and 1980–96, the inequality behaviour was the same as the Lorenz sense. The abovementioned sufficient condition, however, was not attained during the period 1996–2005, during which the $\alpha\beta$ value increased while α decreased (Fig. 2).

In this case, we could resort to numerical procedures to determine whether any dominance exists between the corresponding Generalized Lorenz curves. As the expectations $\alpha\beta$ increased during this period, if there were a monotonic pattern of inequality in the Generalized Lorenz sense, then the inequality would necessarily be decreasing. This is shown immediately from (4) and (5) by taking account of the fact that when $p = 1$, expression (4) corresponds to the expression of the mean of the random variable X.

4. Conclusions

The analysis of fertility curves from the standpoint of inequality provides a tool that usefully complements demographic analysis based solely on the behaviour of the SFI, as the latter sometimes fails to detect certain situations of interest, as described below.

It appears reasonable to believe that a sustained rise in birth rates would arise naturally from higher fertility rates among younger women. In such a situation, the concentration of birth rates and, therefore, the degree of inequality must increase. Conversely, a sustained decline in birth rates over a given period would be associated with a decrease in inequality. However, it can be seen in Table 1 that a particular situation occurred in the period 1975–80 that could remain unnoticed if only the SFI values were considered. Although the latter index decreased during 1975–1996, inequality, according to both the Lorenz and the generalized Lorenz curves, did not fall during the period 1975–80, as would have been expected; on the contrary, it increased during this period. Thus, a detailed analysis of the age-specific fertility rates for the period 1975–80 (Table 3) reveals an anomalous behaviour pattern of the birth rate among younger women, in relation to the overall fertility rate. Indeed, SFI values indicate that the fertility rate decreased during this period while the age-specific fertility rates increased. This unexpected birth rate pattern among young women gave rise to a higher concentration of the birth rate and thus greater inequality in the distribution of fertility among age groups, despite the decrease in the SFI.

The sociological reasons for the above lie in the specificity of this particular historical period in Spain. It was a time of great social change, of transition from a dictatorship to a democracy. Society evolved from a situation of severe restrictions on individual freedoms affecting, among other aspects, sexual customs and behaviour, to a democratic context in which these freedoms were guaranteed. This, together with the fact that

birth control and family planning were less well established at the outset, provides a plausible explanation for the specificity of the period 1975–80 that was detected by simple inequality analysis.

Table 3: *Age-specific fertility rates, 1975–2009.*

Age	Year					
	1975	1976	1977	1978	1979	1980
15	3.323	4.059	4.205	4.644	5.182	5.151
16	8.168	9.633	9.643	10.21	10.912	11.259
17	17.574	20.783	21.23	21.64	22.29	22.175
18	33.238	37.494	37.955	38.887	38.182	36.832
19	50.679	56.712	59.153	59.226	59.573	55.133
20	76.447	85.061	85.313	83.838	80.556	78.252

References

Abad-Montes, F., Huete–Morales, M. D. and Vargas–Jiménez, M. (2006). Using the Gamma distribution to fit fecundity curves for application in Andalusia (Spain). Ch. 9 in *Distribution Modeling Theory*. World Scientific Publishing. Singapore.

Arnold, B. C., Robertson, C. A., Brockett, P. L. and Shu, B. Y. (1987). Generating ordered families of Lorenz curves by strongly unimodal distributions. *Journal of Business and Economic Statistics*, 5, 305–308.

Champernowne, D. G. and Cowell, F. A. (1998). *Economic Inequality and Income Distribution*. Cambridge University Press.

Duchêne, J. and Gillet de Stefano, S. (1974). Ajustement analytique des courbes de fécondité générale. *Population et Famille*, 32 (2), 53–94.

Gastwirth, J. L. (1971). A general definition of the Lorenz curve. *Econometrica*, 39, 1037–1039.

INE (2010). Indicadores Demográficos Básicos. Metodología. www.ine.es.

Kleiber, C. and Krämer, W. (2003). Efficiency, equity, and Generalized Lorenz dominance. *Estadística (Journal of the Inter–American Statistical Institute)*, 55 (Special Issue on Income Distribution, Inequality and Poverty), 173–186.

Lambert, P. (2001). *The Distribution and Redistribution of Income, 3rd ed.* Manchester University Press.

Ramos, H. M., Ollero, J. and Sordo, M. A. (2000). A sufficient condition for Generalized Lorenz order. *Journal of Economic Theory, 90, 286–292.*

Ramos, H. M. and Sordo, M. A. (2002). Characterizations of inequality orderings by means of dispersive orderings. *Qüestiió*, 26 (1–2), 15–28.

Shaked, M. and Shanthikumar, J. G. (2007). *Stochastic Orders.* Springer, New York.

Shorrocks A. F. (1983). Ranking income distributions. *Economica*, 50, 3–17.

Stoyan, D. (1983). *Comparison Methods for Queues and Other Stochastic Models*. Wiley, New York.