

Comparison of Parameter Estimators for the Dagum Distribution

Alina Jędrzejczak¹, Dorota Pekasiewicz², Wojciech Zieliński³

Abstract

The Dagum model is frequently applied to the analysis of income and wage distributions all over the world. It has many desirable statistical properties and turned out to be well fitted to empirical distributions in different divisions. The estimates of its parameters can be applied to the evaluation of numerous income distribution characteristics, including inequality, poverty and wealth indices, dispersion measures based on quantiles and concentration curves. They can also be used to compare income distributions in space and over time. The estimation of these characteristics needs reliable Dagum distribution estimates. The paper is devoted to the analysis of statistical properties of various estimators of the Dagum distribution parameters.

Keywords: size distributions, Dagum distribution, parameter estimation,

JEL Classification: C13, C15

DOI: 10.14659/SEMF.2018.01.18

1 Introduction

The Dagum distribution was originally derived by Camilo Dagum (1977), when he investigated the income elasticity of the cumulative distribution function (CDF) in several developed and developing countries. The elasticity turned out to be a monotonic decreasing and bounded function of CDF. The resulting distribution turned out to present many advantages over its competitors and thus was frequently applied to the analyses of income distributions in many countries and by different divisions (Jędrzejczak, 2014, 2015a, 2015b). Its CDF depends on three parameters: the scale parameter and the two shape parameters, and it can be considered as a special case of the generalized beta distribution of the second kind (GB2) and of the Burr type III (inverse Burr XII) distribution (see: Kleiber and Kotz, 2003). Unlike the gamma or the lognormal distribution the Dagum distribution has an explicit mathematical formula for its CDF and its p^{th} quantile. It generally presents very good consistency with empirical income and wage distributions (that are known to be unimodal and positively skewed) as well as with the non-modal wealth distributions, depending on the

¹ University of Łódź/Institute of Statistics and Demography, Department of Statistical Methods, 41/43 Rewolucji 1905 St., 90-214 Łódź, Poland, jedrzej@uni.lodz.pl.

² Corresponding author: University of Łódź/Institute of Statistics and Demography, Department of Statistical Methods, 41/43 Rewolucji 1905 St., 90-214 Łódź, Poland, pekaszewicz@uni.lodz.pl.

³ Warsaw University of Life Science /Department of Econometrics and Statistics, Nowoursynowska 159 St., 02-776 Warsaw, Poland, wojciech_zielinski@sggw.pl

product of its shape parameters. For some values of shape parameters the distribution is unimodal, otherwise it is non-modal. Also, the number of finite moments of the distribution (in practice 2 or 3) depends on the value of the shape parameter. Moreover, the model possesses the important property of weak Pareto law as its cumulative distribution function converges to the Pareto model, which is considered ideal for high income groups. Recently, the Dagum distribution has been studied from a reliability point of view and used to analyse survival data (e.g.: Domma, et al., 2011) due to its hazard function which can be monotonically decreasing, an upside-down bathtub, or bathtub and then upside-down bathtub shaped. Several estimation methods for the Dagum model parameters have been discussed in the literature. Amongst them, the maximum likelihood, pseudo maximum likelihood, the method of percentiles, the method of moments and a nonlinear least squares on the quantile function proposed by Dagum himself (Dagum, 1977) are the most frequently applied. The review of these methods can be found in: Dey et al. (2017), Kleiber and Kotz (2003). The method of maximum likelihood, the most popular one, presents many desirable properties including consistency, invariance, asymptotic efficiency and normality. A simulation study performed by Domański and Jędrzejczak (1998) revealed that the ML estimates of the Dagum model parameters are normally distributed and efficient for very large samples (greater than 7000). One needs also remember that the ML estimators from parametric distributions have robustness problems and are sensitive to extremes (see e.g. Victoria-Feser and Ronchetti, 1994; Victoria-Feser, 2000). Moreover, as it has been emphasized in Kleiber and Kotz (2003), there is a path in the Dagum distribution parameter space along which the likelihood becomes unbounded. A simulation study concerned with other estimation methods (except for the maximum likelihood) was reported in Dey et al. (2017), but it did not lead to practical conclusions due the choice of the parameters and sample sizes, unrealistic from the point of view of income distribution analyses. In the light of these finding it becomes desirable to examine the estimation methods for the Dagum model in order to assess their estimation errors for sample sizes used in practical applications.

The main objective of this paper was to recognise the methods that can be applied for moderate sample sizes which are often considered in practice (socio-economic groups, family types or NUTS2 regions in Poland). In the second section we briefly present the characteristics of the Dagum distribution. The next section comprises the review of selected estimation methods that can be applied to the case of three-parameter Dagum model. In section 4 we present the results of Monte Carlo experiments designed to assess the accuracy of estimators.

2 Dagum distribution

The probability density function of the Dagum distribution $D(a, v, \lambda)$ is given by:

$$f_{a,v,\lambda}(x) = \frac{av}{\lambda} \left(\frac{x}{\lambda} \right)^{av-1} \left(1 + \left(\frac{x}{\lambda} \right)^v \right)^{-a-1} \quad \text{for } x > 0 \quad (1)$$

and the cumulative distribution function takes on the form:

$$F_{a,v,\lambda}(x) = \left(1 + \left(\frac{x}{\lambda} \right)^v \right)^{-a} \quad \text{for } x > 0, \quad (2)$$

where $\lambda > 0$ is scale parameter, $v > 0$ and $a > 0$ are shape parameters determining the Lorenz curve and inequality measures.

The quantile function of the Dagum distribution is the following:

$$Q_{a,v,\lambda}(q) = \lambda \left(q^{\frac{1}{a}} - 1 \right)^{-\frac{1}{v}} \quad \text{for } 0 < q < 1, \quad (3)$$

while the moments about the origin of the random variable X are:

$$E_{a,v,\lambda} X^m = \lambda^m \frac{\Gamma\left(1 - \frac{m}{v}\right) \Gamma\left(a + \frac{m}{v}\right)}{\Gamma(a)} \quad \text{for } m < v, \quad (4)$$

where Γ is the gamma function.

From (4) it is seen that the number of finite moments of the distribution (1) does not exceed the value of the parameter v . Thus this parameter determines the tail of the Dagum distribution.

As the Dagum distribution is dedicated to the analysis of income and wages, it is convenient to have the explicit formulas for Gini and Zenga inequality measures based on the distribution parameters. The popular Gini index for the distribution (1) takes on the form

$$G = \frac{\Gamma(a) \Gamma\left(2a + \frac{1}{v}\right)}{\Gamma(2a) \Gamma\left(a + \frac{1}{v}\right)} - 1 \quad \text{for } v > 1, \quad (5)$$

while the Zenga (1984) index, defined on the basis of distribution and income quantiles, becomes:

$$Z = 1 - \exp \left[\frac{1}{v} \left[\Psi(a) + \Psi\left(1 - \frac{1}{v}\right) - \Psi\left(a + \frac{1}{v}\right) - \Psi(1) \right] \right], \quad (6)$$

where Ψ is the digamma function.

It is worth noting that both formulae (5) and (6) depend only on the shape parameters of the Dagum distribution: a and v , which can be considered as the distribution “inequality” and “equality” parameters, respectively.

3 Estimation methods

The section is devoted to the brief description of the estimation methods which can be especially useful in the case of the three-parameter Dagum distribution applied to income data. The comprehensive review of this methods can be found in Dey et al. (2017). After the theoretical consideration of their properties, we choose the following methods of parameter estimation: Method Maximum Likelihood Method (ML), Methods of L -Moments (LM), Method of Ordinary Least-Squares (OLS) and Method of Weighted Least-Squares (WLS)

Let X be a random variable, X_1, X_2, \dots, X_n be a sample and $X_{1:n} \leq X_{2:n} \leq \dots \leq X_{n:n}$ be order statistics.

One of the estimation methods frequently applied for the Dagum model parameters is the maximum likelihood. The likelihood function of the Dagum distribution (1) takes the form:

$$L(a, v, \lambda; X_1, \dots, X_n) = \left(\frac{av}{\lambda}\right)^n \prod_{i=1}^n \left(\frac{X_i}{\lambda}\right)^{av-1} \left(1 + \left(\frac{X_i}{\lambda}\right)^v\right)^{-a-1}. \quad (7)$$

To find maximum likelihood estimators we have to solve (numerically) the following system

of equations: $\frac{\partial L}{\partial a} = 0$, $\frac{\partial L}{\partial v} = 0$, $\frac{\partial L}{\partial \lambda} = 0$.

Another estimation method that can be applied in case of the three-parameter Dagum model, is the method of L -moments (LM). The L -moments are defined as (Hosking, 1990):

$$v_r = \frac{1}{r} \sum_{k=0}^{r-1} (-1)^k \binom{r-1}{k} EX_{r-k:r} \text{ for } r \geq 1. \quad (8)$$

For the Dagum distribution the first three L -moments take the following forms:

$$v_1 = \lambda \frac{\Gamma\left(a + \frac{1}{v}\right)}{\Gamma(a)} \Gamma\left(1 - \frac{1}{v}\right), \quad (9)$$

$$v_2 = \lambda \left(\frac{\Gamma\left(2a + \frac{1}{v}\right)}{\Gamma(2a)} - \frac{\Gamma\left(a + \frac{1}{v}\right)}{\Gamma(a)} \right) \Gamma\left(1 - \frac{1}{v}\right), \quad (10)$$

$$v_3 = \lambda \left(\frac{2\Gamma\left(3a + \frac{1}{v}\right)}{\Gamma(3a)} - \frac{3\Gamma\left(2a + \frac{1}{v}\right)}{\Gamma(2a)} + \frac{\Gamma\left(a + \frac{1}{v}\right)}{\Gamma(a)} \right) \Gamma\left(1 - \frac{1}{v}\right). \quad (11)$$

Unbiased estimators of those moments are equal to (Dey et al., 2017):

$$l_1 = \frac{1}{n} \sum_{i=1}^n X_{i:n}, \quad (12)$$

$$l_2 = \frac{2}{n} \sum_{i=1}^n \frac{(i-1)}{n-1} X_{i:n} - l_1, \quad (13)$$

$$l_3 = \frac{6}{n} \sum_{i=1}^n \frac{(i-1)(i-2)}{(n-1)(n-2)} X_{i:n} - \frac{6}{n} \sum_{i=1}^n \frac{(i-1)}{(n-1)} X_{i:n} + l_1. \quad (14)$$

Estimators obtained by the method of L -moments are the solution of the following system of equations: $l_m = v_m$ for $m = 1, 2, 3$.

The ordinary least square estimators (OLS) and weighted least square estimators (WLS) were proposed by Swain et al. (1988). It is well known that

$$E[F_{a,v,\lambda}(X_{i:n})] = \frac{i}{n+1}, \quad D^2[F_{a,v,\lambda}(X_{i:n})] = \frac{i(n-i+1)}{(n+1)^2(n+2)}. \quad (15)$$

The ordinary least square estimators are obtained by minimizing with respect to a, v, λ the function:

$$\min_{a,v,\lambda} \left(\sum_{i=1}^n \left[F_{a,v,\lambda}(X_{i:n}) - \frac{i}{n+1} \right]^2 \right). \quad (16)$$

The weighted least square estimators are obtained by minimizing with respect to a, v, λ the function:

$$\min_{a,v,\lambda} \left(\sum_{i=1}^n \frac{(n+1)^2(n+2)}{i(n-i+1)} \left[F_{a,v,\lambda}(X_{i:n}) - \frac{i}{n+1} \right]^2 \right). \quad (17)$$

4 Simulation study

Several Monte Carlo experiments have been carried out to examine basic statistical properties of the estimators $\hat{\lambda}, \hat{v}$ and \hat{a} of the Dagum model parameters λ, v and a . The experiments involved four estimation methods described in section 3, namely: ML, LM, OLS and WLS. For each estimation method and each sample size ($n=1000, 2000$) the following steps were performed:

- drawing $N=1000$ independent random samples from the Dagum distribution,

- estimation of the Dagum model parameters for each sample,
- the estimators $\hat{\lambda}$, $\hat{\nu}$, \hat{a} and their empirical relative bias and empirical relative root means squared error were calculated:

$$B(\hat{\theta}) = \frac{\frac{1}{N} \sum_{i=1}^N (\hat{\theta}_i - \theta)}{\theta} \cdot 100\%, \quad \text{RMSE}(\hat{\theta}) = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{\theta}_i - \theta)^2}}{\theta} \cdot 100\%.$$

Throughout the experiments we assumed a constant value of the scale parameter ($\lambda = 1$), while the values of the inequality parameter ν , responsible for the right tail of the Dagum distribution, changed from 2 to 4 in order to comprise the wide variety of distributions, presenting light as well as heavy tails. The light-tailed distributions ($\nu=3.5$ and $\nu=4.0$) had three finite moments; among the heavy-tailed ones we considered the distributions with two moments ($\nu=2.5$ and $\nu=3.0$) or with only first moment ($\nu=2$). The parameter a fluctuated between 0.2 and 1.2 what made it possible to control the distribution inequality. Selected this way the sets of the Dagum model parameters embraced the distributions with different levels of the Gini ratio including the values typical for income analyses (Table 1).

Table 1. Gini index.

ν	a					
	0.2	0.4	0.6	0.8	1.0	1.2
2.0	0.704	0.600	0.549	0.519	0.500	0.487
2.5	0.618	0.503	0.449	0.419	0.400	0.387
3.0	0.552	0.433	0.380	0.351	0.333	0.321
3.5	0.499	0.380	0.330	0.303	0.286	0.274
4.0	0.456	0.339	0.291	0.266	0.250	0.240

Fig. 1 shows the examples of Dagum densities for different sets of parameters. They confirm outstanding flexibility of this distribution discussed in section 1. Fig. 2-3 depict the values of the empirical bias and RMSE of \hat{a} obtained by the above-mentioned estimation methods, while the corresponding characteristics obtained for $\hat{\nu}$ are presented in Fig. 4-5.

In Table 2 we present the relative root means squared errors (RMSEs) for the estimators of the Dagum model parameters obtained by four estimation methods : ML, LM, OLS and WLS and for sample sizes $n=1000$ and $n=2000$.

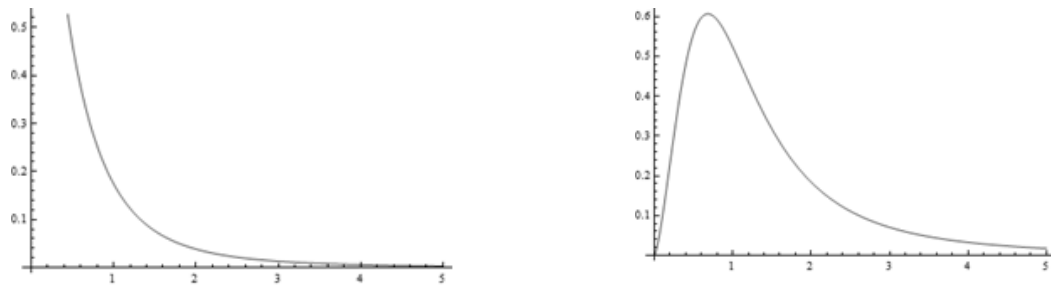


Fig. 1. Density functions of Dagum distribution: $D(0.2;2,1)$ -left; $D(1.2;2,1)$ -right.

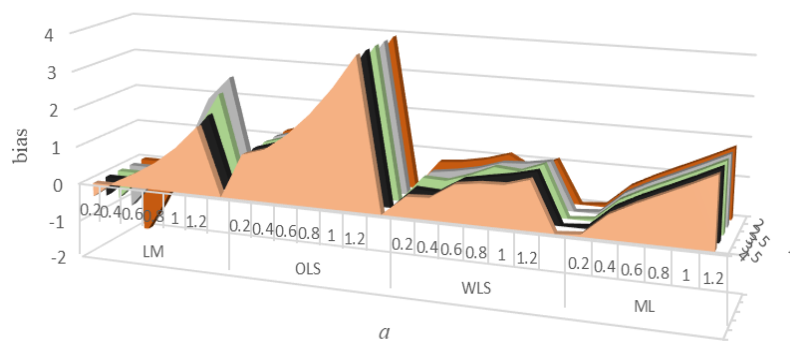


Fig. 2. Empirical relative bias of various estimators of a .

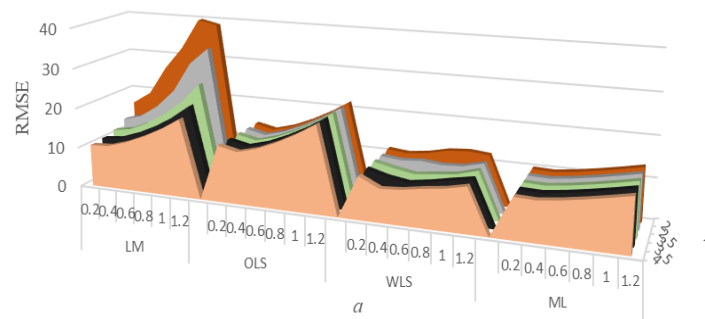


Fig. 3. Empirical relative RMSE of various estimators of a .

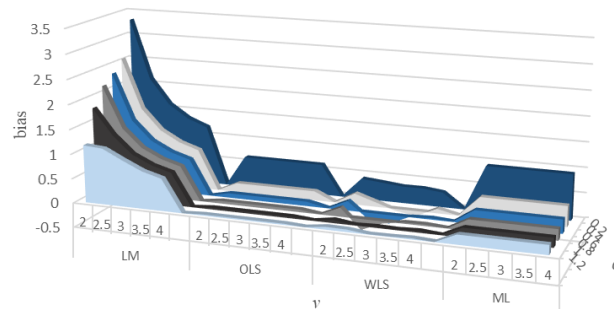


Fig. 4. Empirical relative bias of various estimators of v .

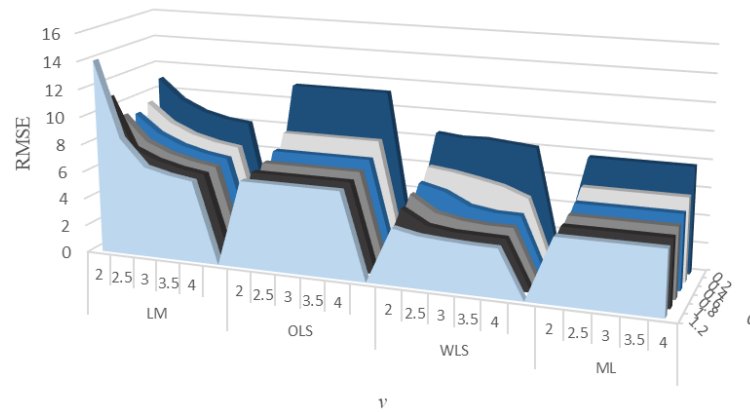


Fig. 5.. Empirical relative RMSE of various estimators of v .

Table 2. Characteristics of estimators of the Dagum model parameters.

Distribution	Estimation method	RMSE(\hat{a})		RMSE(\hat{v})		RMSE($\hat{\lambda}$)	
		$n=1000$	$n=2000$	$n=1000$	$n=2000$	$n=1000$	$n=2000$
$D(0.2,2,1)$	LM	16.510	12.574	10.944	8.097	11.971	9.037
	OLS	14.054	10.022	11.210	7.860	9.959	7.390
	WLS	11.088	7.883	8.522	5.976	8.848	6.544
	ML	9.993	6.982	7.703	5.240	8.354	6.065
$D(0.6,2.5,1)$	LM	17.572	12.359	7.659	5.572	9.727	7.204
	OLS	14.306	9.892	7.110	4.920	8.314	5.990
	WLS	10.684	7.279	5.312	3.804	6.462	4.541
	ML	10.438	7.243	5.412	3.743	6.403	4.590
$D(0.8,3,1)$	LM	17.261	11.820	6.742	4.825	8.053	5.840
	OLS	16.328	11.120	6.689	4.623	7.717	5.487
	WLS	9.601	6.400	3.942	2.800	4.635	3.239
	ML	11.288	7.800	5.066	3.517	5.643	4.013
$D(1.2,4,1)$	LM	19.513	12.884	5.195	4.180	6.552	4.659
	OLS	21.702	14.244	6.302	4.356	7.109	4.966
	WLS	11.131	7.433	3.475	2.497	3.797	2.680
	ML	13.313	9.119	4.691	3.277	4.786	3.373

The estimators of the parameter a obtained by the WLS or ML method are characterized by the smallest RMSE. Among considered estimators of v the smallest biases are observed for the WLS method. The smallest root mean squared errors of the estimator of v have been

noticed for the estimators obtained by ML or WLS method. LM estimators were found to present the smallest precision (highest RMSE of a and highest RMSE of $\nu < 2.5$) amongst all the statistics considered in the study.

It is worth noting that the ML method tends to overwhelm the remaining ones when the sample size is increasing and when the inequality level is high what can be observed for the first distribution $D(0.2, 2, 1)$, corresponding to $G=0.704$, shown in the Table 2. For the distributions presenting smaller inequality (e.g. $D(0.8, 3, 1)$ with $G=0.351$) the WLS method seems more appropriate. It can also be noticed (Fig. 2-5) that the precision of the estimators of the Dagum model parameters ν and a is getting better (smaller RMSE) together with the increasing values of ν and decreasing values of a . It is what could have been expected as the ν parameter is responsible for the right tail of the Dagum density, determining the number of finite moments of the distribution (see: eq. 4). Contrary to this, the a parameter is positively correlated with the distribution inequality (see: eq. 5 and 6) so its high values denote highly dispersed distributions. In general, the sample sizes $n=1000$ are still too small to confirm satisfactory level of RMSE, what is especially evident for a (Fig. 3). Just for $n=2000$ and for moderate inequality the RMSE values do not exceed 5% of the estimated parameters values.

5 Conclusions

The Dagum distribution is widely assumed as a theoretical model for income distributions in empirical analyses. Its parameters are to be estimated from sample data for whole countries and for different subpopulations. The main objective of this study was to recognise the estimation methods that can be applied for moderate sample sizes which are often considered in practice (socio-economic groups, family types or NUTS2 regions). The Monte Carlo experiments revealed that the smallest root mean squared errors of the estimators have been obtained when the ML or WLS method were applied. LM estimators were found to present the smallest precision (highest RMSE of a and highest RMSE of $\nu < 2.5$) amongst all the statistics considered in the study. The ML method tends to overwhelm the remaining ones when the sample size is increasing and when the inequality level is higher. The precision of the estimators of the Dagum model shape parameters ν and a is getting better (smaller RMSE) together with the increasing values of ν and decreasing values of a . To obtain satisfactory results the sample size must be at least $n=2000$.

References

- Dagum, C., (1977). A New Model of Personal Income Distribution: Specification and Estimation. *Economie Appliquee*, 30, 413-437.
- Dey, S., Al-Zahrani, B. & Basloom, S. (2017). Dagum Distribution: Properties and Different Methods of Estimation. *International Journal of Statistics and Probability*, 6, 74-92.
- Domma, F., Giordano., S. & Zenga., M. M. (2011). Maximum Likelihood Estimation in Dagum Distribution with Censored Sample. *Journal of Applied Statistics*, 38, 2971-2985.
- Domański, C. & Jędrzejczak, A. (1998). Maximum Likelihood Estimation of the Dagum Model Parameters. *International Advances in Economic Research*, 4, 243-252.
- Hosking, J. R. M. (1990). L-Moments: Analysis and Estimation of Distributions Using Linear Combinations of Order Statistics. *Journal of the Royal Statistical Society. Series B (Methodological)*, 52, 105-124.
- Jędrzejczak, A. (2014). Income Inequality and Income Stratification in Poland. *Statistics in Transition*, 15, 269-283.
- Jędrzejczak, A. (2015a). Regional Income Inequalities in Poland and Italy. *Comparative Economic Research*, 18, 27-45.
- Jędrzejczak, A. (2015b). Asymptotic Properties of Some Estimators for Gini and Zenga Inequality Measures: a Simulation Study. *Statistica & Applicazioni*, 13, 143-162.
- Kleiber, C. & Kotz, S. (2003). *Statistical Size Distributions in Economics and Actuarial Sciences*. Wiley, Hoboken.
- Swain, J., Venkatraman, S. & Wilson, J. (1988). Least squares estimation of distribution function in Johnson's translation system. *Journal of Statistical Computation and Simulation*, 29, 271-297.
- Victoria-Feser, M. P. (2000). Robust Methods for the Analysis of Income Distribution, Inequality and Poverty. *International Statistical Review*, 68, 277-293.
- Victoria-Feser, M. P. & Ronchetti, H. (1994). Robust Methods for Personal Income Distribution Models. *The Canadian Journal of Statistics*, 22, 247-258.
- Zenga, M. (1984). Proposta per un indice di concentrazione basato sui rapporti fra quantili di popolazione e quantili reddito. *Giornale Degli Economisti e Annali di Economia* 48, 301–326.