Tjalling C. Koopmans Research Institute

Discussion Paper Series nr: 05-02

Testing Gibrat's Legacy: A Bayesian Approach to Study the Growth of Firms

Elena Cefis
Matteo Ciccarelli
Luigi Orsenigo

Universiteit Utrecht
Utrecht School of Economics

Tjalling C. Koopmans Research Institute
Utrecht School of Economics
Utrecht University

Vredenburg 138
3511 BG Utrecht
The Netherlands
telephone +31 30 253 9800
fax +31 30 253 7373
website www.koopmansinstitute.uu.nl

The Tjalling C. Koopmans Institute is the research institute and research school of Utrecht School of Economics.
It was founded in 2003, and named after Professor Tjalling C. Koopmans, Dutch-born Nobel Prize laureate in economics of 1975.

In the discussion papers series the Koopmans Institute publishes results of ongoing research for early dissemination of research results, and to enhance discussion with colleagues.

Please send any comments and suggestions on the Koopmans institute, or this series to M.Damhuis@econ.uu.nl

ontwerp voorblad: WRIK Utrecht

How to reach the authors

Please direct all correspondence to the first author.

Elena Cefis
Utrecht University
Utrecht School of Economics
Vredenburg 138
3511 BG Utrecht
The Netherlands
Email: e.cefis@econ.uu.nl

Matteo Ciccarelli
European Central Bank

Luigi Orsenigo
Università di Brescia and CESPRI
Università Bocconi

Utrecht School of Economics
Tjalling C. Koopmans Research Institute
Discussion Paper Series 05-02

Testing Gibrat's Legacy: A Bayesian Approach to Study the Growth of Firms

Elena Cefis[a]
Matteo Ciccarelli[b]
Luigi Orsenigo[c]

[a]Utrecht School of Economics
Utrecht University
University of Bergamo

[b]European Central Bank

[c]Università di Brescia and CESPRI
Università Bocconi

November 2004

Abstract
Gibrat's law is a referent model of corporate growth dynamics. This paper employs Bayesian panel data methods to test for Gibrat's law and its implications. Using a Pharmaceutical Industry Database (1987-1998), we find evidence against Gibrat's law on average, within or across industries. Estimated steady states differ across firms, and firm sizes and growth rates don't converge within the same industry to a common limiting distribution. There is only weak evidence of mean reversion: initial larger firms do not grow relatively slower than smaller firms. Differences in growth rates and in size steady state are persistent and firm-specific, rather than size-specific.

Keywords: Gibrat's Law, Firm Growth, Pharmaceutical Industry, Heterogeneity, Bayesian Estimation

JEL classification: C11, C23, D21, L11, L25

Acknowledgements
The financial supports of the European Union (ESSY Project, 'Sectoral Systems in Europe: Innovation Competitiveness and Growth'), of EPRIS project (European Pharmaceutical Regulation and Innovation System), and the Italian Ministry of University and Research (grant n. 9913443984, and ex 60%, grant n.CEFI6004, Dept. of Economics, University of Bergamo) are gratefully acknowledged. Part of Ciccarelli's research was undertaken when he was at the University of Alicante and benefitted from funding by DGIMCYT (BEC2002-03097) and Generalitat Valenciana (CTIDIB/2002/175). The views expressed in this paper are exclusively those of the authors and not those of the European Central Bank. The usual disclaimer applies

1 Introduction

After several years of neglect, considerable attention has recently been devoted by industrial economists to the study of the processes of corporate growth. The large majority of empirical studies in this field is based on testing the "Law of Proportionate Effects", also known as Gibrat's Law (Gibrat, 1931), which assumes that firm's size follows a random walk and hence that firm's growth is erratic. As a consequence there would be no convergence within or across industries, and no stable or predictable differences in growth would exist either in the short or in the long run. Rather, growth would be driven by small idiosyncratic shocks.[1]

Gibrat's Law was originally used as an explanation of the highly skewed distribution of firms' size. Even if the growth rate of each firm in an industry is unrelated to its current size, the variance of the firm size distribution and the level of concentration increase over time (Simon and Bonini, 1958, Ijiri and Simon 1974 and 1977). Subsequently, the Law of Proportionate Effects has become, both empirically and theoretically, a referent model for discussing the processes of firms' growth. Nevertheless, Gibrat's Law contrasts with most fundamental theories of firms' growth, ranging from standard models of convergence to an optimal size, to models where heterogeneous firms, facing idiosyncratic sources of uncertainty and discrete events, are subject to market selection, so that the most efficient firms grow while the others shrink and eventually leave the market (e.g. Geroski, 1998, for a discussion). Indeed, most recent theoretical models of firm's growth and industry evolution imply several violations of standard Gibrat-type processes (e.g. Jovanovic, 1982; Ericson and Pakes, 1995; Dosi et al., 1995; Pakes and Ericson, 1998; Winter, Kaniovski and Dosi, 2000). Moreover, Gibrat's Law is at odds with other observed empirical phenomena like the persistence of heterogeneity in some firms' characteristics and measures of performance, e.g. profits, productivity and - more controversially - innovation (see Baily and Chakrabarty, 1985; Mueller, 1990; Geroski et al., 1993; Cefis and Orsenigo, 2001; Cefis, 2003). However, the hypothesis that firms' growth rates are erratic is often taken almost for granted and considered as a *stylized fact* (Geroski, 1998). More generally, Gibrat's Law enters in the models and in the empirical discussion as a fundamental way of conceptualizing firm's growth (Klette et al., 2000; McCloughan, 1995) and, if anything, models are devised that capable to yield random growth as a result (Sutton, 1997, Geroski et al., 1997)

A large empirical literature has explored this issue in different data sets and with different statistical methodologies. Typically, the starting point of the analysis is a simple econometric model having the following form:

---

[1] For a recent discussion, see Sutton (1997) and Bottazzi et al. (2001).

$$\ln S_{it} = \beta_0 + \beta \ln S_{it-1} + u_{it} \tag{1}$$

where S_{it} is the size of firm i at time t, and u_{it} is an i.i.d. shock.

Gibrat's Law would be confirmed if the model $M_o : \beta = 1$ could not be rejected versus the alternative $M_1 : \beta < 1$. Empirical results are controversial. Some early studies (Hart and Prais, 1956; Simon and Bonini, 1958; Hymer and Pashigian, 1962) confirm the view that firm's size does indeed follow a random walk ($\beta = 1$), at least as large firms are concerned (Hall, 1987; Lotti, et al., 2003). Nevertheless, a considerable body of results rejects M_o, suggesting instead that firms size is mean reverting (Baldwin 1995, ch.5; Baily et al., 2000), in the sense that, conditional on firm survival, average firms' growth (and its variance) decline with firm size, holding firm age constant (Dunne et al., 1989; Evans, 1987a, 1987b; Hall, 1987, Caves, 1998).[2] In some versions, Gibrat's Law is considered to hold for large firms, whereas smaller companies grow faster, but with a higher variance.

In this work we reconsider these issues by addressing four interrelated questions. The first two questions are quite conventional. First, we ask if Gibrat's law holds, by testing the random walk assumption. Second, we verify the existence and the extent of mean reversion, i.e. the phenomenon that the rate of convergence to the steady state is lower as firms get larger. Then, we address some simple, but less conventional issues. Specifically, third, we check whether firms converge to a common steady state (as implied by the mean reversion argument) or to a firm-specific steady state size. Fourth, in the latter case, we investigate whether initial size differences persist, or there is a non-negligible number of initial smaller firms that are eventually able to catch-up or even to forge ahead.

These exercises are prompted by the consideration of the high degree of heterogeneity which is commonly observed among firms, even in very narrowly defined industries and lines of business. Such heterogeneity might be a significant factor influencing the the basic results obtained in the Gibrat's Law literature.

In fact, to answer the questions we address in this paper, we adapt a hierarchical Bayesian normal linear model (Lindley and Smith, 1972) to autoregressive time series panel data, i.e., data consisting of many time series generated by the same type of autoregressive model. The motivation for using this statistical framework can be articulated in several points.

First, the autoregressive model is chosen for sake of homogeneity with previous empirical studies.

Second, the results on Gibrat's law and its implications crucially depend on the total variation, i.e., on

[2] Differentials in growth rates have been explained with firm's age (Mata, 1994; Dunne, Roberts and Samuelson, 1989)), firm's size (Harhoff, Stahl and Woywode, 1998; Hart and Oulton, 1996; Hall, 1987; Evans 1987a and 1987b) or both (Farinas and Moreno, 2000).

the information contained both in the cross-sectional and in the time-series dimensions. Previous studies attempts to verify Gibrat's law using either classical modelling of time series and cross sectional data, or short-panel econometric techniques with homogeneity in the parameters across units and over time. We consider these approaches as *problematic*. Estimating equations like (1) firm by firm is possible only when a large time span is available, which is not always the case for micro data. Alternatively, cross section analysis ignores important information contained in unit-specific time variation in growth rates. Also, forcing the parameters to be the same across units, thus pooling possibly heterogeneous units as if their data were generated by the same process, is too restrictive. This is true even if one disposes of the information available for all periods and all cross sectional units, because by forcing units to be homogeneous we only exploits one side of the information contained in a panel data set. Recent studies (e.g. Pesaran and Smith, 1996) have shown that imposing the slope parameter to be homogeneous across units in an autoregressive panel data model distorts the estimation value of the parameter β towards the unit, irrespective of its true value, thus rendering less powerful a test of the Gibrat's law (see also Goddard et al., 2002). The hierarchical model approach reduces the estimation variability typically encountered in firm by firm regressions, and, at the same time, exploits coefficient similarities across firms without imposing the same population structure. Concretely, the hierarchical model allows for an exchangeable scheme where the parameter vectors vary across different firms, subject to a common distribution with unknown means and variance. In this sense, the model represents a satisfactory compromise between the regression model with the same coefficient for all firms and the time series regressions with different coefficients for each single firm.

Third, given the hierarchical model specification, the verification of Gibrat's law is based on the autoregressive parameter of the common underlying distributions, more than on the identification of firm-specific coefficients, which are also available. In other words, to verify Gibrat's law one can simply implement the appropriate test on the corresponding elements of the common mean vector, rather than on the corresponding individual firm-specific regression coefficients (Li, 1999).

Finally, in situations where several short time series are simultaneously modelled, the Bayesian paradigm is attractive because it offers a natural scheme for combining and weighting data from several similar sources (Nandram and Petruccelli, 1997). The Bayesian estimation of this hierarchical model is also computationally straightforward due to the recent advances in Bayesian statistics and Markov Chain Monte Carlo Methods (see Gelfand and Smith, 1990 for general applications and Hsiao et al., 1999, for an application to panel data model, among others).

All these considerations justify the use of the Bayesian hierarchical model whose main characteristic, plain heterogeneity, has the useful feature of exploiting in a more powerful way all the information contained in the panel data set.

Using 210 firms from a Pharmaceutical Industry Database on the sample 1987-1998, we do not find strong support for the law. Moreover, data show only weak evidence of mean reversion, i.e. initially larger firms do not systematically grow more slowly than smaller firms. Finally, differences in growth rates and in size steady state are firm-specific and persistent. The specified model provides an adequate fit to the data and results do not change under plausible alternative prior and model assumptions. In other words, they are robust to more general families of prior information.

The paper is structured as follows. Section 1 discusses the statistical model. Section 2 describes data and comments on the estimation results. In Section 3 we check the robustness of the results. Section 4 concludes.

2 Model specification

The evolution of size for all units is determined by a doubled indexed stochastic process $\{S_{it}\}$, where $i \in I$ indexes firms, $t = 0, 1, ..$ indexes time and I is the set of the first n integers. Following Sutton (1997), if ε_{it} is a random variable denoting the proportionate rate of growth between period $t-1$ and t for firm i, then

$$S_{it} - S_{it-1} = \varepsilon_{it} S_{it-1}$$

and

$$S_{it} = (1 + \varepsilon_{it}) S_{it-1} = S_{i0} (1 + \varepsilon_{i1}) (1 + \varepsilon_{i2}) ... (1 + \varepsilon_{it})$$

In a short period of time, ε_{it} can be regarded as small and the approximation $\ln(1 + \varepsilon_{it}) = \varepsilon_{it}$ can be justified. Hence, taking logs, we have

$$\ln S_{it} \simeq \ln S_{i0} + \sum_{t=1}^{T} \varepsilon_{it}$$

If the increments ε_{it} are independently and normally distributed, then $\ln S_{it}$ follows a random walk and the limiting distribution of S_{it} is lognormal. Therefore, the growth of the firm is unrelated to its current size and only depends on the sum of idiosyncratic shocks.

Hence, to test Gibrat's law, the vast majority of previous literature have used the following general logarithmic specification

$$\ln S_{it} = \beta_{i0} + \beta \ln S_{it-1} + u_{it} \tag{2}$$

where S_{it} is the size of firm i at time t, and u_{it} is a random variable that satisfies

$$\begin{aligned} E\left(u_{it} \mid S_{it-s}, \quad s>0\right) &= 0 \\ E\left(u_{it}u_{j\tau} \mid S_{it-s}, \quad s>0\right) &= \begin{cases} \sigma^2 & i=j,\, t=\tau \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

Gibrat's law is confirmed if the hypothesis $\beta = 1$ is not rejected by the data against $\beta < 1$.

An equivalent specification used by the literature and based directly on corporate growth rates is

$$\ln \frac{S_{it}}{S_{it-1}} = \beta_{i0} + \beta_1 \ln S_{it-1} + u_{it}$$

where clearly $\beta_1 = \beta - 1$. In this case Gibrat's law is confirmed if data do not reject $\beta_1 = 0$, against $\beta_1 < 0$.[3]

In this work we follow a similar autoregressive specification, introducing three main innovations with respect to the traditional empirical approaches. First, we study the behavior of the (log of) each unit's size relative to the average, i.e., of the variable $g_{it} = \ln\left(S_{it}/\bar{S}_t\right)$, where $\bar{S}_t$ represents the average size over all units at each time t. The use of the proportion of size g_{it} as our basic variable, instead of (the log of) plain size S_{it}, alleviates problems of serial and residual correlation, in that possible common shocks are removed by the normalization. Moreover, the variable g_{it} can be interpreted as the firm's market share. Second, we assume that even firms belonging to the same industries can differ substantially from each other. This (possibly intrinsic) heterogeneity is modelled in a general way by allowing all unknown parameters to be unit-specific. Finally, the latter feature is modelled in the context of a hierarchical linear model (Lindley and Smith, 1972) estimated with Bayesian techniques. As already shown in several studies, the panel-data hierarchical-model approach uses in a powerful way the variability contained both in the cross sectional and in the time series dimensions, allows the implementation of the relevant tests in a natural way, and is easy to estimate, given the recent advances in Bayesian statistics and MCMC techniques.

For our purposes, we assume that time series realizations $\{g_{it}\}_{t=t_i}^{\mathcal{T}_i}$ for n firms ($i = 1, ..., n$) are available, possibly of different lengths. Each series starts at time t_i and is generated by an autoregressive model of order 1 ($AR(1)$). Without lack of generality, the minimum t_i equals 1. The last observation occurs at time $\mathcal{T}_i$, for each firm. Assuming that there are no missing observation between t_i and $\mathcal{T}_i$ for each i, we let $\mathsf{T}_i = \mathcal{T}_i - t_i + 1$ denote the number of observations in the series for the ith firm. The initial conditions, g_{i0}, are observed and the subsequent estimation results are conditional on them. The following statistical

[3] In both cases the test is a one-tail test, the same that we use in our empirical analysis for consistency with the literature on testing for unit root. Note also that $\beta > 1$ ($\beta_1 > 0$) implies explosive growth paths, i.e., firms grow faster as they get larger. This situation is conceivable for a short period, but not indefinitely. Moreover, from a qualitative perspective its implications for market structure are similar to those of $\beta = 1$: concentration *would* increase over time, *al*though at a faster rate.

model is specified for each firm i:

$$g_{it} = \alpha_i + \rho_i g_{it-1} + \eta_{it} \qquad t \geq t_i \tag{3}$$

The random variables η_{it} are assumed normally and identically distributed, with mean zero and variance σ_i^2, and are uncorrelated across units and over time, i.e.,

$$\eta_{it} \mid \sigma_i^2 \sim N\left(0, \sigma_i^2\right) \qquad E\left(\eta_{it}\eta_{js}\right) = 0, \ \ \forall i \neq j, t \neq s \tag{4}$$

Two additional features are worth mentioning. First, our model specification allows for an intercept, α_i. Other studies (e.g. Bottazzi et. al., 2001) sometimes estimate a specification equivalent to $g_{it} = \rho g_{it-1} + u_{it}$, which, beside considering a common slope, also avoids the inclusion of the specific effect α_i, as if the expected proportionate rates of growth were zero. Second, another advantage of the hierarchical model adopted here is the possibility of handling unbalanced panels in a natural way. Therefore, firms with different observations can be included in the sample, which in turn allows us to both maximize the size of cross-sectional observations and minimize the survival bias.

For the sake of simplicity, let $\theta_i = (\alpha_i, \rho_i)'$ and $x_{it} = (1, g_{it-1})'$. Eq. (3) can then be written in a more compact form as

$$g_{it} = x_{it}'\theta_i + \eta_{it}, \qquad i = 1, 2, ..., n, \qquad t \geq t_i \tag{5}$$

The hierarchical structure is introduced into Eq. (5) with an exchangeable assumption on the population structure

$$\theta_i \sim N\left(\theta_c, \Sigma_c\right), \qquad i = 1, 2, ..., n \tag{6}$$

where $\theta_c \left(= (\alpha_c, \rho_c)'\right)$ and Σ_c are the unknown common mean and variance-covariance matrix, respectively. The chosen prior distribution assumes that intercept and slope of the model *do not differ too much* across units, so the firm-specific parameter vector θ_i is an independent random draw from the underlying common distribution (6). The matrix Σ_c controls the variability of the firm-specific regression parameter vector θ_i. The standard linear regression model with homogenous coefficients ($\theta_i = \theta_c$) for each firm is obtained by letting Σ_c be a null matrix.

Under this set up, the parameters of interest are $\left(\left\{\theta_i, \sigma_i^2\right\}_{i=1}^n, \theta_c, \Sigma_c\right)$. Gibrat's law can be tested by comparing the model $M_o : \rho_c = 1$ against $M_1 : \rho_c < 1$. A finding that ρ_c is not statistically different from 1 would confirm the law as holding over time and across firms. A third model specification, $M_2 : \rho_i = 1$ $(i = 1, ..., n)$ can also be examined. A finding that ρ_i $(i = 1, ..., n)$ is not statistically different from 1 would then be considered as the Gibrat's law holding uniformly over individual firms.

Further implications of the law can then separately be examined. Concretely, we can compare the speed of adjustment $(1 - \rho_i)$ of each unit to its own steady state, with the respective initial conditions, g_{i0}, a question related to the mean reversion argument and the decrease in the variance of the firm size over time. Also, we can verify whether steady states are all equal across firms, by comparing $H_o : \mathsf{SS}_i = \mathsf{SS}_j$ against $H_1 : \mathsf{SS}_i \neq \mathsf{SS}_j$, $\forall i \neq j$, where SS_i is the steady state of firm i. Finally, if steady states are not common, the model specification can easily be used to verify whether the long-run differences across firms are transitory or permanent, i.e., whether there is persistence in size differences. The latter can be done by comparing the posterior distribution of the steady states to the initial conditions.

3 Bayesian analysis of the model

A full implementation of the Bayesian approach is achieved here using the Gibbs sampler (e.g. Gelfand et al, 1990, for illustration of general models, and Nandram and Petruccelli, 1997, for an application to autoregressive time series panel data), a recursive Monte Carlo method which requires only knowledge of the full conditional posterior distribution of the parameters.

The analysis requires the specification of a prior for θ_c, Σ_c and σ_i^2. Assuming independence, as is customary in the literature, we take

$$\begin{aligned} & p\left(\theta_c, \Sigma_c^{-1}, \{\sigma_i^2\}_{i=1}^n\right) \\ \propto \quad & p\left(\theta_c\right) \cdot p\left(\Sigma_c^{-1}\right) \cdot \prod_{i=1}^n p\left(\sigma_i^2\right) \end{aligned} \tag{7}$$

to have a Normal-Wishart-Inverse Gamma structure:

$$\theta_c \sim N\left(\mu, C\right) \tag{8}$$

$$\Sigma_c^{-1} \sim W\left(s_o, S_o^{-1}\right) \tag{9}$$

$$\sigma_i^2 \sim IG\left(\frac{v}{2}, \frac{\delta}{2}\right) \tag{10}$$

The notation $\Sigma_c^{-1} \sim W\left(s_o, S_o^{-1}\right)$ means that the matrix Σ_c^{-1} is distributed as a Wishart with scale S_o^{-1} and degrees of freedom s_o, while $\sigma_i^2 \sim IG\left(\frac{v}{2}, \frac{\delta}{2}\right)$ denotes an inverse gamma distribution with shape $\nu/2$ and scale $\delta/2$. The hyperparameters μ, C, s_o, S_o, v and δ have to be specified by the researcher. Concretely, μ is the prior mean of the common mean vector θ_c; C controls the dispersion of our prior belief around θ_c: the larger the C matrix, the weaker the prior information on θ_c; s_o, the degrees-of-freedom parameter of the Wishart, controls the dispersion of Σ_c^{-1}, and S_o^{-1} the corresponding location: the bigger is s_o relative to the size of the cross-section, n, and the smaller is S_o, the smaller is the prior mean of Σ_c^{-1} making the prior on θ_i more informative and shrinking θ_i more towards the common mean θ_c; finally v and δ control the shape and the scale of the prior distribution for σ_i^2: a less informative prior is obtained by letting v and δ become smaller. In the following section, results are reported under

4 different prior specifications, by varying these hyperparameters in a reasonable range. In Appendix B, we perform a sensitivity analysis in order to investigate how much does our results change when we use other reasonable probability models.

Let $\psi = \left(\left\{\theta_i, \sigma_i^2\right\}_{i=1}^n, \theta_c, \Sigma_c\right)$ denote the parameters of interest , $Y = (G_1', ..., G_n')'$ denote the vector of data, where $G_i = (g_{it_i}, ..., g_{i\mathsf{T}_i})'$, and $X_i = \left(x_{i1}', ..., x_{i\mathsf{T}_i}'\right)'$. Given the likelihood from the normality assumption (4) and the prior information previously specified through (6) to (10), and conditioning on the initial observations, the joint posterior distribution of ψ is given by

$$
\begin{aligned}
p(\psi \mid Y) &\propto f(Y \mid \psi)\, p(\psi) \\
&\propto \prod_{i=1}^{n}\left\{\left(\sigma_i^2\right)^{-\frac{\mathsf{T}_i}{2}} \exp\left[-\frac{1}{2}\sigma_i^{-2}\left(G_i - X_i\theta_i\right)'\left(G_i - X_i\theta_i\right)\right]\right\} \\
&\quad \times \left|\Sigma_c\right|^{-\frac{n}{2}} \exp\left[-\frac{1}{2}\sum_{i=1}^{n}\left(\theta_i - \theta_c\right)'\Sigma_c^{-1}\left(\theta_i - \theta_c\right)\right] \\
&\quad \times |C|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}\left(\theta_c - \mu\right)' C^{-1}\left(\theta_c - \mu\right)\right] \\
&\quad \times \left|\Sigma_c\right|^{-\frac{1}{2}(s_o - k - 1)} \exp\left[-\frac{1}{2} tr\left(S_o \Sigma_c^{-1}\right)\right] \\
&\quad \times \prod_{i=1}^{N}\left(\sigma_i^2\right)^{-\left(\frac{v}{2}+1\right)} \exp\left[-\frac{\delta}{2\sigma_i^2}\right]
\end{aligned}
$$

where $k = 2$ denotes the dimension of the vector θ_i. The first line of the formula represents the standard likelihood conditional on the initial conditions and the others represent the different levels of the prior information.

The multiple integration needed to obtain the marginal posterior distributions of each component of ψ is not feasible analytically and must be performed numerically. As anticipated, we use the Gibbs sampler. The relevant conditional distributions are obtained from the above formula. Concretely, the steps through which the Gibbs sampler must cycle are easily shown to be the following:

(i) $p\left(\theta_i \mid Y, \psi_{-\theta_i}\right) = N\left[A_i\left(\sigma_i^{-2}X_i'G_i + \Sigma_c^{-1}\theta_c\right), A_i\right] \qquad i = 1, ..., n$

(ii) $p\left(\theta_c \mid Y, \psi_{-\theta_c}\right) = N\left[B\left(n\Sigma_c^{-1}\tilde{\theta} + C^{-1}\mu\right), B\right]$

(iii) $p\left(\Sigma_c^{-1} \mid Y, \psi_{-\Sigma_c^{-1}}\right) = W\left[\left(\sum_{i=1}^{n}\left(\theta_i - \theta_c\right)\left(\theta_i - \theta_c\right)' + S_o\right)^{-1}, s_o + n\right]$

(iv) $p\left(\sigma_i^2 \mid Y, \psi_{-\sigma_i^2}\right) = IG\left[\mathsf{T}_i/2, \left(G_i - X_i\theta_i\right)'\left(G_i - X_i\theta_i\right)/2\right] \qquad i = 1, ..., n$

where $A_i = \left(\sigma_i^{-2}X_i'X_i + \Sigma_c^{-1}\right)^{-1}$, $B = \left(n\Sigma_c^{-1} + C^{-1}\right)^{-1}$, $\tilde{\theta} = (1/n)\sum_i \theta_i$, and $\psi_{-\gamma}$ denotes ψ excluding γ.

The Bayesian point estimates and other quantities of interest are then obtained by taking the appropriate averages over the *useful* Gibbs draws, i.e., those draws for which convergence to the marginal posterior distributions has been achieved. Results are shown in terms of posterior point estimate of ρ_c, ρ_i, and SS_i. The comparison of the models is made bilaterally, i.e., $M_o : \rho_c = 1$ against $M_1 : \rho_c < 1$ and $M_2 : \rho_i = 1$ $(i = 1, ..., n)$, and $H_o : \mathsf{SS}_i = \mathsf{SS}_j$ against $H_1 : \mathsf{SS}_i \neq \mathsf{SS}_j$, $\forall i \neq j$. We base the comparison on the Bayes factor (BF), that is the ratio of marginal data densities under alternative models. The latter quantities are computed using the Gibbs output as suggested by Chib (1995).

In order to verify the mean reversion argument, we can compare the average firm-specific speed of convergence $1 - \rho_i$ with the initial conditions. If the useful draws of the Gibbs sampling goes from $\bar{L}$ to L, the average ρ_i is computed as $\bar{\rho}_i = \left(\bar{L} - L + 1\right)^{-1} \sum_{l=\bar{L}}^{L} \rho_i^{(l)}$, where $\rho_i^{(l)}$ is the lth draw for the ith firm. Given that uncertainty on ρ_i resulting form the model is ignored in this comparison, a better alternative will be using the draws for the individual ρ_i and computing $\Pr\{\rho_i < 1 \mid Y\}$, i.e., the posterior probability that the autoregressive coefficient for the ith firm is lower than the unit. The mean reversion is checked by comparing this quantity with the initial conditions. A negative relationship will provide evidence in favor of the mean reversion.

Finally, to check whether differences in size across firms are persistent or not, we compute the following probabilities: $p_1 = (1/n_1) \sum_i \mathcal{P}_i$, and $p_2 = (1/n_2) \sum_i \mathcal{Q}_i$, where $\mathcal{P}_i = \Pr\{\mathsf{SS}_i < 0 \mid g_{i0} < 0, Y\}$ is the probability that the posterior size steady state is lower than the average, given that the initial size is lower than the average, and $\mathcal{Q}_i = \Pr\{\mathsf{SS}_i > 0 \mid g_{i0} > 0, Y\}$ is the probability that the posterior size steady state is greater than the average, given that the initial size is greater than the average. n_1 and n_2 are the number of firms which started below and above the average respectively. The higher p_1 and p_2, the more *attractive* are the initial conditions and the more persistent are the initial differences in size. The complementary probabilities, $1 - p_1$ and $1 - p_2$ will then provide the *transition* probabilities of going from low to high and from high to low size respectively. A visual inspection of the size persistence argument is also easily obtained by comparing both the average posterior steady state for each firm ($\overline{\mathsf{SS}}_i = \left(\bar{L} - L + 1\right)^{-1} \sum_{l=\bar{L}}^{L} \mathsf{SS}_i^{(l)}$) and the unconditional probability $q = (1/n) \sum_i \mathcal{R}_i$, where $\mathcal{R}_i = \Pr\{\mathsf{SS}_i < 0 \mid Y\}$ with the initial conditions.

In order to check that the model provides an adequate fit to the data we follow Gelman et al. (1995, Ch. 6 and 12) and compare simulated values from the posterior predictive distribution of replicated data to the observed data. The procedure is the following. Let Y be the observed data and ψ the vector of parameters (including now all hyperparameters). Define Y^{rep} the *replicated* data that *could have been* observed, or, the data we *would* see tomorrow if the experiment that produced Y today were replicated with the same model and the same value of ψ that produced the observed data. The distribution of Y^{rep}

given the current state of knowledge, i.e. the posterior predictive distribution, is:

$$p(Y^{rep}) = \int p(Y^{rep} \mid \psi)\, p(\psi \mid Y)\, d\psi$$

The discrepancy between the model and the data is measured by defining a *discrepancy* measure $T(Y, \psi)$, which is a scalar summary of parameters and data. Lack of fit of the data with respect to the posterior predictive distribution is then measured by the tail-area probability (Bayes *p-value,* as defined by Gelman et al.) of the quantity, and computed using posterior simulations of (ψ, Y). This value is defined as the probability that the replicated data could be more extreme than the observed data, as measured by the test quantity:

$$\text{B-}p = \Pr(T(Y^{rep}, \psi)) \geq \Pr(T(Y, \psi))$$

where the probability is taken over the joint posterior distribution of (ψ, Y^{rep}). Major failures of the model typically correspond to extreme tail-area probabilities (less than 0.01 or more than 0.99). The general discrepancy measure chosen is the square root of the average of the n sums of squared residuals. The test-quantities and the tail-area probabilities are easily calculated as a by-product of the Gibbs sampler.

4 The Data

The pharmaceutical industry constitutes a particularly interesting testbed for Gibrat's Law. In fact, pharmaceuticals might be considered an ideal case where the process of firms' growth should behave in accordance with the Law of Proportionate Effects, as a consequence of the peculiar role and nature of innovation in this industry. As it is well known, pharmaceuticals is a highly innovation-intensive industry. Moreover, the innovative process in this sector has often been described and conceptualized as a pure "lottery model", whereby previous innovations (even in a particular submarket) do not influence in any way current and future innovation in the same or in other submarkets (Sutton, 2002).

Data come from the PHID (Pharmaceutical Industry Database) dataset, developed at CERM/EPRIS. The database provides longitudinal data for the sales of 210 firms in the seven largest western markets (France, Germany, Italy, Spain, UK, Canada, and USA) during the period 1987-1998. Values are in Thousands of Pound Sterling at constant 1998 exchange rate. The companies included in the dataset result from the intersection of the top 100 companies (in terms of sales) in each national market, obtaining a total of 210 companies. The PHID database had been constructed by aggregating the values of the sales of these firms in the different national markets: therefore, sales for each firm stand for the sum of their sales in each of the national markets. It is important to emphasize that the panel is unbalanced since processes of entry and exit are explicitly considered.

A few comments are in order here.

We use sales as proxy for firm size because it is the only measure available from the dataset and because sales are usually considered the best available proxy of firm's size in pharmaceuticals and in some recent studies on firms' growth (Hart and Oulton, 1996; Geroski et al., 1996, Higson et al., 2002). Alternative measures are more difficult to obtain on a longitudinal basis and suffer of a number of drawbacks. In particular, the use of employment - even if available - would unduly increase the lumpiness of the growth process, especially as it concerns divestitures, opening (or closure) of plants, R&D labs, etc., particularly as they occur at the international level.[4]

The use of a broad geographical definition for the relevant market and the consideration of the international firm - i.e. the sum of the sales in each national market - as the unit of analysis is justified in our view by the global nature of competition in the pharmaceutical industry. Many firms operate at the same time in different countries, as it concerns R&D, production and marketing. More importantly, successful drugs are sold worldwide and firms' growth depends crucially on the ability to be present at the same time in different countries. Hence, the world market - as approximated by the seven large countries for which data were available - seems to us an appropriate level of aggregation for capturing the locus of competition.

In terms of products, the market is defined here at the level of one single class at the 4 digit level of the Standard Industrial Classification, i.e. pharmaceutical products. In this respect, the definition of the market is quite narrow. It could be argued that the pharmaceutical industry is actually constituted by a collection of several (independent) submarkets or therapeutic categories, definable at extremely fine levels of disaggregation. Yet, firms' growth in this industry is fundamentally dependent on the process of diversification on a variety of submarkets (Sutton 2002, Bottazzi et al. 2001, Henderson et al., 1999). Thus, focusing the analysis on a single (or few) submarket(s) would imply missing an essential driver of firms' growth.

In this paper we exclusively focus on the process of *internal* growth of firms. For this reason, in order to control for mergers and acquisitions during the period of observation, we constructed "virtual-firms". These are firms actually existing at the end of the period for which we constructed backward the series of their data in the case they merged or made an acquisition. Hence, if two firms merged during the period, we consider them merged from the start, summing up their sales from the beginning. This procedure might introduce a bias in the intertemporal comparison of firms' size distributions along time, but it has the advantage of emphasizing the changes in the distributions that derive strictly from intra-market

[4]Employment, assets, sales, market value, and value added are some of the most common measures of company size. For a discussion on their advantages and their limits, see Hart and Oulton (1995).

competition (Bottazzi et al. 2001, p.1168). Furthermore, by constructing "virtual firms", we avoid to mix the phenomenon of pure (greenfield) entry as opposed to entry due to mergers. In fact, there is evidence that greenfield entrants are smaller than average firms (Baldwin et al., 1995; Acs, 1996)) whereas entrants which are the result of a merger are usually larger than average firms, like in the case of the entry of Novartis in 1996 following the exit of Ciba-Geigy and Sandoz in 1995. Our database confirms this finding: the greenfield entrants have smaller size than the average and they generally are the minimum values of the size distributions.

The methodology used to construct the dataset supports the use of the international market as the relevant locus of competition. Given this procedure, the smallest firms might seem to be under-represented. However, our dataset is not constituted only or even mainly by large companies. Descriptive statistics reported in the Appendix show that the sample includes several small and medium sized companies. In fact, the firm's size distribution is skewed towards the smallest firms of the sample, since the skewness is always significantly positive and the median is always much smaller than the mean.

As expected, the variable we construct, $g_{it} = \ln\left(S_{it}/\bar{S}_t\right)$ (in the Table, Ln_dev87...98) washes away the increasing trend of the total sales over time and removes the possible shocks common to all the industry. As a matter of fact the ratio of the standard deviation to the mean as well as the skewness and the kurtosis are nearly constant over time.

Finally notice that the minimum number of observations in the series for each firm is $\mathsf{T}_i = 2$ (for just one firm).

5 Estimation results

In this section we present the empirical results. They are shown in Figures 1-9 and Tables 1-3.

The estimation results are reported under four prior specifications. Table 1 describes the chosen hyperparameters.

Table 1. Prior information

prior	v	δ	s_o	S_o	μ	C
A	10	1	$k+2$	$20 \cdot I$	$\check{\mu}$	$\check{C}$
B	$\mathsf{T}_i + 10$	$\hat{\delta}_o$	$k+50$	$10 \cdot I$	$\tilde{\mu}$	$\tilde{C}$
C	5	1	$k+2$	$20 \cdot I$	$\tilde{\mu}$	$100 \cdot I$
D	5	1	$k+50$	$10 \cdot I$	$\tilde{\mu}$	$\tilde{C}$

The notation of the table is as follows:

1. $\check{\mu} = 0.5\left(\bar{\mu} + \mu_{pool}\right)$, $\check{C} = 0.5\left(\bar{C} + C_{pool}\right)$

2. $\bar{\mu} = (1/N)\sum_i \hat{\theta}_i$, with $\hat{\theta}_i = \left(X_i'X_i\right)^{-1} X_i'G_i$, $\mu_{pool} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'Y$, where $\mathbf{X} = diag\left(X_1, ..., X_n\right)$;

3. $\bar{C} = (1/N) \sum_i \left(\hat{\theta}_i - (1/N) \sum_i \hat{\theta}_i\right) \left(\hat{\theta}_i - (1/N) \sum_i \hat{\theta}_i\right)'$, and $C_{pool} = (\mathbf{X}'\mathbf{X})^{-1}$;

4. $\tilde{\mu} = (0 \;\; 0.96)'$ and $\tilde{C} = diag\,(10, 0.006)$;

5. $\hat{\delta}_o = (1/N) \sum_i (G_i - X_i'\breve{\mu})' (G_i - X_i'\breve{\mu})$.

The hyperparameters have been chosen following simple criteria and mixing sample information and previous empirical studies both on Gibrat's law and on hierarchical models. Prior A can be regarded as moderately non informative at all levels of the hierarchical structure. Prior B is more informative than prior A in all respects. First, it has a bigger s_o relative to the size of the cross-section, and a smaller is S_o: both features imply a smaller prior mean for Σ_c^{-1} making the prior on θ_i more informative and shrinking θ_i more towards the common mean θ_c. Second, the mean and the variance of μ_2, i.e. the hyperprior mean of ρ_c, have been taken from the previous empirical estimates of ρ_c in the literature as reported in Goddard et al. (2002, pp.417, table 1). We fitted an empirical distribution on this estimates and then computed its mean and variance. Figure 1 reports the histogram of these estimates. Finally, the shape ν_o and the scale δ_o of the prior on the variance have been chosen to match on average the sample variance.

Prior C is less informative than priors A and B at all levels of the hierarchy. Prior D is a mixture of priors B and C.

The Gibbs sampling is run in 4 cycles of 5000 iterations. Results are based on the last 5000 iterations, therefore discarding the first 15000 draws. Convergence has been checked following the method proposed by Brooks and Gelman (1998). It has been achieved already after 10000 iterations, using different seeds of the random number generator and different initial values of the unknown parameters.

Table 2 reports the mean and the 90% central part of the posterior distribution of ρ_c, the model testing as explained above (M_o vs. M_1, M_o vs. M_2, M_1 vs. M_2, and H_o vs. H_1), the marginal likelihood, i.e., the posterior density of the data, $(\ln(\hat{m}(y)))$, computed as in Chib (1995) and the Bayes p-value (B-p) to check the model fit of the data under the four priors.

As a general comment, notice that Prior A both fits the data well –according to the particular test quantity chosen– and produce the highest posterior density. Therefore most of the results discussed below, especially those concerning mean reversion and persistence of size differences, are based on the output generated under Prior A. Results based on the other priors are anyway qualitatively identical and therefore omitted to avoid replications.

Table 2. Estimation and testing: Benchmark model [5]

prior	$\rho_c\,(5\% - 95\%)$	$\ln(BF_{01})$	$\ln(BF_{02})$	$\ln(BF_{12})$	$\ln(BF_{ss})$	$\ln(\hat{m}(y))$	B-p
A	0.92 (0.85 − 0.99)	−6.19	420.94	427.13	−823.66	2201.03	0.59
B	0.94 (0.88 − 1.01)	−6.32	135.30	142.21	−835.45	2163.02	0.03
C	0.93 (0.85 − 1.02)	−4.72	453.53	458.25	−818.20	1977.23	0.97
D	0.96 (0.91 − 1.01)	−4.94	187.87	192.81	−725.42	2110.99	0.99

5.1 Does Gibrat's law hold?

Table 1 shows that, under all priors, the numerical estimate of ρ_c is different from 1. This can be seen also in the plots of the posterior densities of ρ_c (Figure 2). It is worth noting that in almost all cases the posterior distribution contains 1. However, the values of the Bayes factor BF_{01}, which compare $M_o : \rho_c = 1$ against $M_1 : \rho_c < 1$, are always lower than unity. To interpret these numbers, one can compute the highest prior probability to assign to model M_1 in order for the researcher to obtain posterior odds in favour of M_o. If π is the prior probability of model M_1, the Posterior Odds ratio is defined as the product of the prior odds ratio and the Bayes Factor,

$$PO = \frac{1-\pi}{\pi} BF.$$

Therefore, the highest prior probability to assign to model M_1 in order for the researcher to obtain posterior odds (just) in favour of M_o is $\pi^* = 1/\left(1 + \exp\left(1 - \ln\left(BF_{01}\right)\right)\right)$. Hence, for instance, under prior C one should assign at most $\pi^* = 0.00327$ to M_1 for the data to revert the conclusion and make the posterior inference favorable to M_o. Such a small probability would imply an implausible prior odds ratio of 304.8 in favour of M_o. We consider these numbers as a clear evidence against model M_o as compared to M_1. The former is however strongly favored when compared to model M_2, meaning that the posterior density of the sample data is much higher when we impose an average random walk across firms than when we impose the same assumption to all individual firms. Finally, M_2 is *a fortiori* not favored when compared to model M_1.

Overall, these findings do not confirm that Gibrat's law holds on average, over time and across firms. The histogram of $\bar{\rho}_i$, the posterior mean of ρ_i $(i = 1, ..., n)$ averaged across firms (notation above), provides a first visual inspection of the finding that several firms are far from following Gibrat's Law (Figure 3). An interesting issue is to check which are the firms that follow the law. A further straightforward analysis shows clearly that large firms have a posterior distribution of ρ_i centered on unity. Figure 4 (chart a) plots the posterior distribution of ρ_i for top 10% firms, i.e., firms whose initial size is in the top decile of

[5] BF_{01} compares $M_o : \rho_c = 1$ against $M_1 : \rho_c < 1$; BF_{02} compares $M_o : \rho_c = 1$ against $M_1 : \rho_i = 1$, $i = 1, .., n$; BF_{12} compares $M_1 : \rho_c < 1$ against $M_1 : \rho_i = 1$, $i = 1, .., n$; BF_{ss} compares $H_o : SS_i = SS_j$ against $H_1 : SS_i \neq SS_j$, $\forall i \neq j$. $\ln(\hat{m}(y))$ is the log marginal posterior density. B-p is the Bayes p-value (Gelman et al., 1995).

the initial distribution of sizes. The picture confirms the intuition that large firms do follow Gibrat's law, in line with previous finding of the literature (e.g. Hall 1987, Lotti et al.2003). The marginal likelihood ($\ln \hat{m}(y)$) of the model under prior A and the restriction that $\rho_i = 1$ for the top 10% firms is equal to 2146.58. The same figure is only 1849.38 when we restrict all the other firms to have $\rho_i = 1$. The positive Bayes factor resulting form the difference between the former and the latter provides odds in favor of the hypothesis that the size of large firms follows a random walk, even though the model without any restrictions (and under the same prior) has a higher posterior density (2201.03), as shown in Table 1. This is just saying that the sample data prefer a model where no random walk restriction is imposed on whichever firms.

On the contrary, if we look at the firms whose initial size belongs to the first decile of the distribution (bottom 10%, Figure 4b) we see that the posterior distributions are not centered on the unity, confirming the intuition that initially small firms do not follow Gibrat's law and have a higher speed of convergence to the steady state than initially large firms, which may have already reached their own steady states. These claims are also summarized in Figure 5 (charts "a" and "b"), which show the scatter plots of the average posterior speed of convergence versus the initial sizes. In chart "b" the top 10% and the bottom 10% have been excluded from the sample. The figures confirm that for very small firms the average speed of convergence is far away from zero, while for very large firms it is around zero (chart "a"). The evidence shows no clear pattern when the remaining firms are considered (chart "b").

In sum, this results would seem to confirm previous findings that Gibrat's Law holds only for very large firms but not for the others. The Law appears to hold only for 15% of the firms in our sample. Moreover, the failure of Gibrat's Law and the observation of higher rates of convergence for small firms do not necessarily imply mean reversion.

5.2 Do data show mean reversion?

From the same plots an initial assessment of the mean reversion argument can be drawn. Do smaller firms have a higher speed of convergence than larger firms? As argued above, this seems to be true only when the first and the last decile of the initial distribution of sizes are compared. If we take the "extreme-size" firms out of the sample, the evidence suggests that there is not a negative relation between the speed of convergence and the initial size.

The argument against a strong evidence of mean reversion remains the same also when we take into account the entire posterior distribution of ρ_i and not just its posterior mean. Figure 6 shows the relation between the initial condition and the posterior probability of ρ_i being lower than unity, i.e., $\Pr(\rho_i < 1 \mid Y)$. In chart "a" all firms are considered; in chart "b" again we have excluded the top and the bottom 10%. As from the previous scatter plots, it can be argued that it is indeed true for that very small firms $\Pr(\rho_i < 1 \mid Y)$ is quite high while for very large firms the same probability is low. However, when we consider only the central 80 percent firms of the initial distribution of sizes, this relation is very weak. We therefore cannot express a posterior confidence in favor of a mean-reversion claim. Finally, to the extent that mean reversion actually operates, yet the process is very slow indeed, given the observed values of the speed of convergence (Figures 5a and 5b)

5.3 Are the steady states equal?

But do firms' size converge to the same steady state, as it would be implied by a strict interpretation of the mean reversion argument? Table 2 (column 6: $ln(BF_{ss})$) shows that under all priors the Bayes factor overwhelmingly favors the model where steady states are not restricted to be equal. The evidence is reinforced by Figure 7, where the posterior distributions of the steady states of 10% randomly chosen firms are plotted. Both the Bayes factor and the chart support the evidence that the firms in the sample have very different steady states, confirming once more not only that Gibrat's Law does not hold on average, but also that firms do not converge to the same size. These results suggest that firm-specific characteristics are very important in the determination of firms' growth.

5.4 Do initial differences in size persist?

Gibrat's Law would imply that initial size differences would not tend to persist. The mean reversion argument would suggest that persistence of size differentials should be quite low, as small firms grow faster than large ones, even if the steady state sizes are different. Here, we check the persistence of differences in size. Figure 8 plots the posterior mean of the steady states versus the initial conditions. The positive relation favours the conclusion that the differences across firms are persistent, in that they depend strongly on the initial size.

In order to take into account also the uncertainty on the steady state resulting from the model and not just the posterior mean, we have further investigated the relation between the entire posterior distributions of steady states and the initial condition. First, dividing the initial sample into firms below and above the average, we plot the scatter points of the unconditional probability $q = (1/n)\sum_i \mathcal{R}_i$, where $\mathcal{R}_i = \Pr\{\mathsf{SS}_i < 0 \mid Y\}$, against the initial conditions. Figures 9 reinforces our first preliminary conclusion on the persistence, namely that most of the smallest firms have a high posterior probability of remaining below the average, while for almost all the largest firms the same probability is negligible. Second, we have computed the posterior probabilities $p_1 = (1/n_1)\sum_i \mathcal{P}_i$, and $p_2 = (1/n_2)\sum_i \mathcal{Q}_i$, where $\mathcal{P}_i = \Pr\{\mathsf{SS}_i < 0 \mid g_{i0} < 0, Y\}$ is the probability that the posterior size steady state is lower than the average, given that the initial size is lower than the average, and $\mathcal{Q}_i = \Pr\{\mathsf{SS}_i > 0 \mid g_{i0} > 0, Y\}$ is the probability that the posterior size steady state is greater than the average, given that the initial size is greater than the average. Results give $p_1 = 0.77$ and $p_2 = 0.80$, meaning that the probability of remaining in the same initial position is almost 80 per cent for both states (below and above the average) or, that there is only a 20 percent probability for a firm which starts below (above) the average to reach a steady state above (below) the average.

The same analysis can be refined by considering more quantiles of the initial distribution of sizes and not just the mean. Table 3 reports the average probabilities that a firm in a certain quartile of the initial conditions distribution has to end up in a quartile of the steady states distribution. For instance, the cell (1,1) reports the average probability of a firm with an initial size in the first quartile that its steady state is again in the first quartile; the cell (2,1) reports the average probability of a firm with an initial size in the first quartile that its steady state is in the second quartile; and so on. Therefore, the average probability of remaining in the same initial position is reported on the main diagonal, whereas the off diagonal elements represent the *transition* probabilities of moving from one state to another.

Probabilities in Table 3 reinforce previous results: initial conditions in firm's size matter for the firm's position in the steady state distribution. The probabilities of remaining in the same quartile of the distribution are always larger than those of moving in other quartiles, especially in the upper part of the distribution. It is worth noting that the state that shows the higher persistence is the last quartile: firms that start with a large size with respect to the average size are very likely to end up in steady states much larger than the average steady state.

In sum, these results indicate not only that Gibrat's Law is violated in our sample, but also that mean reversion is very weak: smaller firms tend to remain small and larger firms tend to remain large. Moreover, the speed of convergence to such steady state is very slow.

Table 3. Persistence of differences

	Initial	$\leq 25\%$	$25\% - 50\%$	$50\% - 75\%$	$> 75\%$
Steady state					
$\leq 10\%$		0.5219	0.3184	0.1251	0.0372
$25\% - 50\%$		0.2652	0.4582	0.2577	0.0240
$50\% - 75\%$		0.1296	0.1506	0.5263	0.1840
$> 90\%$		0.0833	0.0728	0.0909	0.7548

6 Summary and concluding remarks

The results of this paper can be summarized as follows:

(i) The main assertion of Gibrat's law that growth rates are erratic is not true on average, across firms and over time. The estimated average speed of adjustment is far from being zero on average when the information contained both in the cross sectional and in the time series dimension is used. However, the finding that the growth of initially very large firms follows a random walk is confirmed here.

(ii) Data show only a very weak evidence of mean reversion. Even if on average $\rho_c < 1$, this does not necessarily mean that initially larger firms grow more slowly than smaller firms. Our analysis shows that the relative speed of convergence of smaller firms is not necessarily higher than the one of larger firms, except in the extreme tails of the distribution.

(iii) Even more important, firm sizes do not converge to a common limiting distribution but to firm-specific steady state size: estimated steady states differ across units . This fact does not imply *per se* that firm size drifts unpredictably over time, as argued by some authors (see Geroski, 2001, p. 6). It is true that a unit root in the process of firm size implies divergence, but the reverse causality does not necessarily hold, as shown in this paper.

(iv) Initial conditions are important determinants of the estimated distribution of steady states. Initial differences in size do not seem to disappear over time and to the extent they do, the process occurs at a very slow rate. Thus, a firm with an initial size below the average is going to narrow the gap somewhat with respect to larger firms, but it does not seem to increase its relative size in the cross sectional distribution. In other words, differences in firm size persist.

(vi) The model we used to perform the analysis does not show failings in fitting to data. Moreover, results are unchanged with alternative models.

In sum, results obtained in this sample contradict two basic implications of both Gibrat's Law and the "generalised" mean reversion argument: almost no correlation is observed between initial size and speed of adjustment, while a strong correlation is found between initial size and the steady state. Thus, there seem to be systematic differences in growth rates among firms that are not size-specific and may depend on other firm-specific features that are not observable in our data. Given that these results are sufficiently robust to different prior specifications, they open rooms for investigating further the determinants of firms growth. Most likely, size is not the only variable which growth should be conditioned on. Other sources of heterogeneity (age being a primary - but certainly not the only - candidate) may more plausibly be responsible for differential growth rates of firms over time. In particular it would be interesting to explore some common features across clearly divergent/convergent firms as well as the role of other variables in the explanation of the cross sectional dispersion in estimated steady states. Finally, the mechanisms through which market selection operates in promoting the growth and the decline of firms should also be explicitly modelled and tested.

At a more general level, the results of this paper strengthens once more the argument that extreme attention has to be given to treating heterogeneity appropriately in econometric models.

References

[1] Acs Z.J. (ed.) (1996), *Small Firms and Economic Growth,* Elgar Reference Collection, International Library of Critical Writings in Economics, vol. 61; Cheltenham, U.K.

[2] Baily M.N. and A.K. Chakrabarty (1985), 'Innovation and Productivity an US Industry', *Brookings Papers on Economic Activity*, 2, pp.609-632.

[3] Baily M.N., C. Hulten, D. Campbell (2000), 'Productivity Dynamics in Manufacturing Plants', in D.B. Audretsch and S. Klepper,(eds.) *Innovation, Evolution of Industry and Economic Growth.* Volume 2. Elgar Reference Collection. International Library of Critical Writings in Economics, vol. 118. Cheltenham, U.K. and Northampton, Mass., Elgar.

[4] Baldwin, J.R, P.K. Gorecki, R. Caves, T. Dunne, J.Haltiwanger (1995), *The Dynamics of Industrial Competition: A North American Perspective*, Cambridge, Cambridge University Press

[5] Bottazzi G., G. Dosi, M. Lippi, F. Pammolli, M. Riccaboni, (2001), 'Innovation and Corporate Growth in the Evolution of the Drug Industry', *International Journal of Industrial Organization,* 19*,* pp.1161-1187.

[6] Brooks, S. P. and A. Gelman (1998), General methods for monitoring convergence of iterative simulations. Journal of Computational and Graphical Statistics*,* 7, 434-455.

[7] Caves, R. E. (1998), Industrial Organization and New Findings on the Turnover and Mobility of Firms, *Journal of Economic Literature*, 36 December), pp. 1947-1982

[8] Cefis E. (2003), 'Is There Persistence in Innovative Activities?', *International Journal of Industrial Organization,* 21(4)*,* pp.489-515.

[9] Cefis E. and L. Orsenigo (2001), 'The Persistence of Innovative Activities. A Cross-Countries and Cross-Sectors Comparative Analysis', *Research Policy*, 30(7), pp.1139-1158.

[10] Chib S. (1995), 'Marginal Likelihood from the Gibbs output', *Journal of the American Statistical Association*, 90, 1313–1321.

[11] Dosi G., O. Marsili, L. Orsenigo and R. Salvatore (1995), 'Technological Regimes, Selection and Market Structure', *Small Business Economics*, 7, pp. 1-26.

[12] Dunne T., M.J. Roberts and L. Samuelson (1989), 'The Growth and Failure of US Manufacturing Plants', *Quarterly Journal of Economics,*104(4), pp.671-698.

[13] Ericson R. and A. Pakes (1995), 'Markov-Perfect Industry Dynamics: a Framework for Empirical Work', *Review of Economic Studies*, 62(1), pp.53-82.

[14] Evans D.S. (1987a), 'The Relationship between Firm Growth, Size, and Age: Estimates for 100 Manufacturing Industries', *Journal of Industrial Economics*, 35(4), pp. 567-81

[15] Evans D.S. (1987b), 'Tests of Alternative Theories of Firm Growth', *Journal of Political Economy*, 95(4), pp. 657-74.

[16] Farinas J.C. and L. Moreno (2000), 'Firms' Growth, Size and Age: A Nonparametric Approach', *Review of Industrial Organization*, 17(3), pp. 249-65

[17] Gelfand A.E., S.E. Hills, A.Racine–Poon and A.F.M. Smith (1990), 'Illustration of Bayesian inference in normal data models using Gibbs sampling', *Journal of the American Statistical Association,* 85(412), pp.972–985.

[18] Gelfand A.E. and A.F.M. Smith (1990), 'Sampling–based approaches to calculating marginal densities', *Journal of the American Statistical Association,* 85(410):398–409.

[19] Gelman A., J.B. Carlin, H.S. Stern and D.B. Rubin (1995), *Bayesian Data Analysis*, Chapman & Hall, London.

[20] Geroski P.A. (1999), 'The growth of Firms in Theory and in Practice', in N. Foss and V. Mahnke (eds.), *Competence, Governance, and Entrepreneurship*, Oxford, Oxford University Press.

[21] Geroski, P., Machin, S.J. and Van Reenen, J. (1993), 'The Profitability of Innovating Firms', *RAND Journal of Economics*, 24(2), pp. 198-211.

[22] P. Geroski P.S, Machin S. J. and Walters C. F., Corporate growth and profitability, The Journal of Industrial Economics, 45, 2 (June 1997), 171-189

[23] Geroski P.A., S. Lazarova, G. Urga and C.F. Walters (2003), 'Are Differences in Firm Size Transitory or Permanent?', *Journal-of-Applied-Econometrics,* 18(1), pp.47-59.

[24] Gibrat, R. (1931), *Les inequalites economiques*. Librairie du Recueil Sirey, Paris.

[25] Goddard, J., J. Wilson and P. Blandon (2002), "Panel tests of Gibrat's law for Japanese manufacturing", *International Journal of Industrial Organization,* 20, 415-433.

[26] Hall B. (1987), 'The Relationship between Firm Size and Firm Growth in the U.S. Manufacturing Sector', *Journal of Industrial Economics*, 35(4), pp. 583-606.

[27] Harhoff D., K. Stahl and M. Woywode (1998), 'Legal Form, Growth and Exit of West German Firms-Empirical Results for Manufacturing, Construction, Trade and Service Industries', *Journal of Industrial Economics*, 46(4), pp. 453-488.

[28] Hart P.E. and N. Oulton (1996), 'Growth and Size of Firms', *The Economic Journal*, 106(Sept.), pp1242-1252.

[29] Hart P.E. and N. Oulton (1995), 'Growth and Size of Firms', National Institute of Economic and Social Research Discussion Paper No.77.

[30] Hart P.E. and S.J. Prais (1956), 'The Analysis of Business Concentration: A Statistical Approach', *Journal of the Royal Statistical Society*, 119, series A, pp.150-191.

[31] Henderson, R., L. Orsenigo, G.P. Pisano (1999), "The Pharmaceutical Industry and the Revolution in Molecular Biology: Interactions among Scientific, Institutional, and Organizational Change", in Mowery, D.C., Nelson, R.R. (eds.) *Sources of industrial leadership: Studies of seven industries.* Cambridge; New York and Melbourne: Cambridge University Press, pp. 267-311.

[32] Higson C„ S.Holly and P.Kattuman (2002), 'The Cross-Sectional Dynamics of the US Business Cycle: 1950-1999', *Journal of Economic Dynamics and Control*, 26, pp.1539-1555.

[33] Hsiao C., M.H. Pesaran and A.K. Tahmiscioglu (1999), 'Bayes estimation of short run coefficients in dynamic panel data models', in Hsiao et al. (eds.) *Analysis of panels and limited dependent variable models: in honor of G.S. Maddala*, Cambridge, Cambridge University Press.

[34] Hymer S. and P. Pashigian (1962), 'Firm Size and the Rate of Growth', *Journal of Political Economy,* 70(4), pp.556-569.

[35] Ijiri Y. and H. Simon (1974), 'Interpretations of Departures from the Pareto Curve Firm-Size Distribution', *Journal of Political Economy*, 82(2), part 1, pp.315-331.

[36] Ijiri Y. and H. Simon (1977), *Skew Distribution and the Sizes of Business Firms*, Amsterdam, North-Holland Publishing.

[37] Jovanovic B. (1982), 'Selection and the Evolution of Industry', *Econometrica*, 50(3), pp.649-70.

[38] Li K. (1989), 'Testing Symmetry and Proportionality in PPP: A Panel-Data Approach', *Journal of Business and Economic Statistics*, 17, pp. 409-418.

[39] Klette T.J., M. Jarle and Z. Griliches (2000), 'Do Subsidies to Commercial R&D Reduce Market Failures? Microeconometric Evaluation Studies', *Research Policy*, 29(4-5), pp. 471-95.

[40] Lindley D.V. and A.F.M. Smith (1972), 'Bayes estimates for the linear model' (with discussion), *Journal of the Royal Statistical Society*, B, 34, pp. 1–41.

[41] Lotti F., E. Santarelli, M. Vivarelli (2003), 'Does Gibrat's Law Hold among Young, Small Firms?', *Journal-of-Evolutionary-Economics*, 13(3), pp. 213-35

[42] Mata J. (1994), 'Firm Growth during Infancy', *Small Business Economics*, 6(1), pp. 27-39

[43] McCloughan P.(1995), 'Simulation of Concentration Development from Modified Gibrat Growth-Entry-Exit Processes', *Journal of Industrial Economics*, 43(4), pp.405-433.

[44] Mueller D.C. (ed.) (1990), *The Dynamics of Company Profits: an International Comparison,* Cambridge, MA, Cambridge University Press.

[45] Nandram B. and J.D. Petruccelli (1997), 'A Bayesian Analysis of Autoregressive Time Series Panel', *Journal of Business and Economic Statistics*, 15, pp. 328-334.

[46] Pakes A. and R. Ericson (1998), 'Empirical Implications of Alternative Models of Firms Dynamics', *Journal of Economic Theory*, 79(1), pp. 1-45.

[47] Pesaran M.H. and R. Smith (1996), 'Estimating Long Run Relationships from Dynamic Heterogeneous Panels', *Journal of Econometrics*, 68, pp.79-113.

[48] Simon H. and C. Bonini (1958), 'The Size Distribution of Business Firms', *American Economic Review*, 58(4), pp.607-617.

[49] Sutton J. (1997), 'Gibrat's Legacy', *Journal of Economic Literature*, 35, pp.40-59.

[50] Sutton J. (2002), 'The Variance of Firm Growth Rates: the 'Scaling' Puzzle', *Physica A*, 312, pp.577-590.

[51] Winter, S G, Y.M. Kaniovski and G. Dosi (2000), 'Modeling Industrial Dynamics with Innovative Entrants', *Structural Change and Economic Dynamics*, 11(3), pp. 255-93

Appendix A:Descriptive Statistics

Variable	Mean	Median	St. Dev.	Skewn.	St.Er.	Kurtosis	St.Er.	Min	Max
Sales_87	241078, 89	29298	553872, 55	3, 21	0, 17	10, 31	0, 34	229	3101318
Sales_88	270161, 54	34672, 00	619103, 71	3, 22	0, 17	10, 28	0, 34	325	3381416
Sales_89	300946, 61	39205	695208, 69	3, 27	0, 17	10, 73	0, 34	507	3890687
Sales_90	334374, 03	40150	774581, 63	3, 25	0, 17	10, 52	0, 34	102	4203022
Sales_91	380943, 28	51561	874492, 88	3, 28	0, 17	10, 94	0, 34	818	5150610
Sales_92	441702, 08	55082, 50	1026942, 76	3, 29	0, 17	11, 03	0, 34	14	6108043
Sales_93	459848, 30	60263, 50	1065013, 49	3, 27	0, 17	11, 01	0, 34	95	6611241
Sales_94	476994, 19	64368	1107718, 39	3, 28	0, 17	11, 16	0, 34	1885	7018260
Sales_95	518844, 59	67792	1193773, 04	3, 21	0, 17	10, 42	0, 34	597	7367569
Sales_96	560422, 12	76659	1295802, 52	3, 19	0, 17	10, 09	0, 34	2914	7637134
Sales_97	603968, 36	86271, 50	1390207, 25	3, 14	0, 17	9, 46	0, 33	3442	7585652
Sales_98	864057, 23	118964, 0	1982203, 50	3, 09	0, 17	8, 94	0, 33	4730	10038996
Ln_Dev87	−1, 85	−2, 11	2, 00	0, 15	0, 17	−0, 13	0, 34	−6, 96	2, 55
Ln_Dev88	−1, 80	−2, 05	1, 95	0, 22	0, 17	−0, 23	0, 34	−6, 72	2, 53
Ln_Dev89	−1, 77	−2, 04	1, 90	0, 29	0, 17	−0, 26	0, 34	−6, 39	2, 56
Ln_Dev90	−1, 80	−2, 12	1, 94	0, 17	0, 17	0, 06	0, 34	−8, 10	2, 53
Ln_Dev91	−1, 77	−2, 00	1, 90	0, 31	0, 17	0, 29	0, 34	−6, 14	2, 60
Ln_Dev92	−1, 83	−2, 08	1, 98	0, 00	0, 17	0, 98	0, 34	−10, 36	2, 67
Ln_Dev93	−1, 80	−2, 03	1, 91	0, 25	0, 17	0, 14	0, 34	−8, 48	2, 67
Ln_Dev94	−1, 79	−2, 00	1, 88	0, 43	0, 17	−0, 36	0, 34	−5, 53	2, 69
Ln_Dev95	−1, 77	−2, 04	1, 87	0, 41	0, 17	−0, 26	0, 34	−6, 77	2, 65
Ln_Dev96	−1, 75	−1, 99	1, 82	0, 51	0, 17	−0, 33	0, 34	−5, 26	2, 61
Ln_Dev97	−1, 71	−1, 95	1, 80	0, 51	0, 17	−0, 30	0, 33	−5, 17	2, 52
Ln_Dev98	−1, 72	−1, 99	1, 81	0, 50	0, 17	−0, 32	0, 33	−5, 21	2, 45

Appendix B: Some sensitivity analysis

In this appendix we briefly analyze how much does previous posterior analysis change if we use other reasonable probability models in place of the one used before. We have already pointed out that our main conclusions on Gibrat's law and its implication are unchanged under different prior of the same hierarchical model. Table 1 says that prior A is to be preferred to the others on the basis of the posterior density and of the model fit, given the test quantity chosen. Nonetheless, results are robust under the four prior assumptions. Here we specify three alternative probability models and verify how do they fit the same data in comparison to the benchmark and how do they affect the previous inference. Bayes factors that compare these models to the benchmark are also computed.

The first alternative model (MOD1) is a "robust" version of the benchmark and uses the Student-t distribution for the sample data in place of the normal (e.g. Gelman et al. 1995). The model is hierarchical and exchangeable in the population θ_i as before, but now we replace (4) with

$$\eta_{it} \sim t_\nu \left(0, \sigma^2\right) \tag{11}$$

where $t_\nu \left(0, \sigma^2\right)$ denotes a zero-mean Student-t distribution with ν degrees of freedom, and scale σ^2. The modification of the Gibbs sampler is straightforward in that the $t_\nu \left(0, \sigma^2\right)$ distribution is just a mixture of normal distributions with common mean and variances distributed as scaled inverse-χ^2. In our case $\eta_{it} \sim t_\nu \left(0, \sigma^2\right)$ is equivalent to

$$\begin{aligned} \eta_{it} \quad | \quad & h_i \sim N\left(0, h_i \sigma^2\right) \\ & h_i \sim \text{Inv-}\chi^2 (\nu, 1) \end{aligned}$$

where Inv-$\chi^2 (\nu, 1)$ denotes a scaled inverse-χ^2 with ν degrees of freedom and scale 1. Therefore an additional step to sample h_i must be added to (i)-(iv) above (Section 1.2):

(v) $\qquad p\left(h_i \mid Y, \psi_{-h_i}\right) = \text{Inv-}\chi^2 \left(\nu_n, s_n^2\right) \qquad i = 1, ..., n$

where $\nu_n = \nu + T_i$, and $s_n^2 = \left\{\left[(G_i - X_i\theta_i)' (G_i - X_i\theta_i) / \sigma_i^2\right] + \nu\right\} / \nu_n$. The other steps are slightly modified to account for the presence of h_i, while the assumption on σ^2 is, as before, $\sigma^2 \sim IG\left(\frac{v}{2}, \frac{\delta}{2}\right)$.

The second alternative (MOD2) is a hierarchical non-exchangeable model where there is no independence between θ_i and σ_i^2. Concretely we replace the second level of the hierarchy (6) with the following

$$\theta_i \sim N\left(\theta_c, \sigma_i^2 \Sigma_c\right), \qquad i = 1, 2, ..., n$$

where the prior variance of the population structure is tied to the sampling variance of the observation g_{it}. In this way, the prior belief about θ_i is calibrated by the scale of measurement of g_{it} and a high-variance prior distribution is induced on θ_i if σ_i^2 is large (e.g. Kadyiala and Karlsson, 1997 for similar priors in the context of VARs). The modification of the Gibbs sampler does not requires additional steps but only an adjustment on the conditional posterior distributions of all parameters due to the presence of σ_i^2 in the prior for θ_i.

The last model (MOD3) is a non-hierarchical, non-heterogeneous model where all firms have the same population coefficients. This model is chosen for the sake of comparison with that part of the empirical literature on Gibrat 's law based on cross-section or homogeneous panel-data setups. In this case the model is

$$\begin{aligned} & g_{it} = x_{it}'\theta_c + \eta_{it}, \\ \eta_{it} \quad | \quad & \sigma^2 \sim N\left(0, \sigma^2\right) \qquad i = 1, 2, ..., n, \qquad t \geq t_i \end{aligned}$$

where it is assumed that $p\left(\theta_c, \sigma^2\right) = p\left(\sigma^2\right) p\left(\theta_c \mid \sigma^2\right)$ with $p\left(\sigma^2\right) = IG\left(v/2, \delta/2\right)$ as before and $p\left(\theta_c \mid \sigma^2\right) = N\left(\mu, \sigma^2 C\right)$. It is easy to show that the conditional posterior distributions of θ_c and σ^2, given the hyperparameters v, δ, μ and C are

$$\begin{aligned} p\left(\theta_c \mid \sigma^2, Y\right) &= N\left[B_1\left(\sum_i X_i' G_i + C^{-1}\mu\right), \sigma^2 B_1\right] \\ p\left(\sigma^2 \mid \theta_c, Y\right) &= IG\left(v_n/2, \delta_n/2\right) \end{aligned}$$

where $B_1 = \left(\sum_i X_i' X_i + C^{-1}\right)^{-1}$, $v_n = v + k + \sum_i T_i$, $\delta_n = \delta + \sum_i \left(G_i - X_i\theta_c\right)' \left(G_i - X_i\theta_c\right) + \left(\theta_c - \mu\right)' C^{-1} \left(\theta_c - \mu\right)$, and $k = 2$ is the dimension of θ_c.

Each model is estimated under two prior assumptions on the hyperparameters, as described in Table 4.

Table 4. Prior information: Alternative models

model	prior	v	δ	ν	s_o	S_o	μ	C
MOD1	E	$\left(\sigma^2 = \check{\sigma}^2\right)$		4	$k+2$	$20 * I$	$\check{\mu}$	$\check{C}$
	F	5	1	4	$k+2$	$20 * I$	$\check{\mu}$	$\check{C}$
MOD2	G	5	1	—	$(\Sigma_c = I)$		$\check{\mu}$	$\check{C}$
	H	10	1	—	$k+2$	$20 * I$	$\check{\mu}$	$\check{C}$
MOD3	L	5	1	—	—	—	$\tilde{\mu}$	$100 \cdot I$
	N	10	1	—	—	—	$\check{\mu}$	$\check{C}$

Here $\check{\sigma}^2 = 0.5\left(\bar{\sigma}^2 + \sigma^2_{pool}\right)$, $\bar{\sigma}^2 = (1/NT)\sum_i \left(G_i - X_i'\hat{\theta}_i\right)' \left(G_i - X_i'\hat{\theta}_i\right)$, with $\hat{\theta}_i = \left(X_i' X_i\right)^{-1} X_i' G_i$, $\sigma^2_{pool} = (1/NT)\left(Y - \mathbf{X}\mu_{pool}\right)' \left(Y - \mathbf{X}\mu_{pool}\right)$, and the remaining notation has been defined in section 2. All prior hyperparameters have been chosen as before, reflecting both sample data information and previous studies. Notice, in particular, two features. First, σ^2 and Σ_c are assumed constant in the priors MOD1-E and MOD2-G respectively: In these cases the corresponding steps in the Gibbs sampler (iii and iv, respectively) are simply not activated. Second, overall the prior assumptions are such that, for each model specification, the first prior is relatively less informative than the second one.

Estimation results are reported in Table 5. As for MOD1 results are reported only for $\nu = 4$, we have also fitted a range of Student-t distributions with 1,2,5,10, 30 and infinite degrees of freedom (the latter being just the normal model already fitted in the previous section).

The main conclusion on the Gibrat's law seems to be robust across different model specifications. Table 5 shows that in all cases the values of the Bayes factor BF_{01} are always lower than unity, while Figure 11 provides evidence in favour of no apparent sensitivity of inferences to the hyperparameter ν: if anything, as ν decreases the posterior mean of ρ_c becomes lower. The remaining issues on the mean reversion and the persistence of differences in steady states are also supported with the same qualitative evidence reported previously.

Table 5. Estimation and testing: Alternative models

prior	ρ_c (5% – 95%)	$\ln(BF_{01})$	$\ln(BF_{02})$	$\ln(BF_{12})$	$\ln(BF_{ss})$	$\ln(\hat{m}(y))$	B-p
E	0.92 (0.85 – 1.00)	−5.51	553.15	558.66	−815.21	2258.23	0.97
F	0.91 (0.85 – 0.97)	−8.07	724.86	732.93	−871.71	2755.67	0.80
G	0.95 (0.87 – 1.03)	−1.05	424.73	425.78	−873.79	981.38	0.98
H	0.88 (0.70 – 1.05)	−48.01	621.96	668.97	−882.53	70.42	0.02
L	0.97 (0.92 – 1.01)	−1.97	—	—	—	899.82	0.25
N	0.97 (0.92 – 1.02)	−1.07	—	—	—	900.39	0.15

Finally two features deserve some attention. First, the "robust" alternative to the benchmark model, which uses a Student-t in place of the normal, provides the highest values of the marginal likelihood under a given model. In fact, under both MOD1-E and MOD1-F, the marginal likelihood is also higher than under any of the priors of the benchmark specification. This means that, in comparing a discrete set of models, as those described here, using Bayes factors, one would choose MOD1-F over all the others, or, alternatively, in averaging over all models, one should weight MOD1-F more. Second, when no heterogeneity is allowed for, the posterior distribution of ρ_c closely resembles the previous results of the literature. Incidentally, notice that under the latter prior assumption, the mean of ρ_c is a priori set equal to 0.90, and therefore that the posterior result is a genuine update of the prior. The fact that the posterior mean of ρ_c is higher than in specifications where heterogeneity is accounted for is a well known result in the classical analysis of dynamic panel data models when units are pooled as if they were homogeneous. Pesaran and Smith (1996), for instance, show how the neglect of coefficient heterogeneity in dynamic panel data models distorts the estimation value of the parameter ρ_c toward the unit, irrespective of its true value.