

FROM GIBRAT'S LEGACY TO GIBRAT'S FALLACY. A BAYESIAN APPROACH TO STUDY THE GROWTH OF FIRMS

Elena Cefis, Matteo Ciccarelli and Luigi Orsenigo*

WP-AD 2002-19

Corresponding author: M. Ciccarelli, Departamento de Fundamentos del Análisis Económico, Universidad de Alicante, Campus de San Vicente del Raspeig, 03071 Alicante, Spain, e-mail: mmatteo@merlin.fae.ua.es.

Editor: Instituto Valenciano de Investigaciones Económicas, S.A.

Primera Edición Octubre 2002

Depósito Legal: V-4034-2002

IVIE working papers offer in advance the results of economic research under way in order to encourage a discussion process before sending them to scientific journals for their final publication.

* E. Cefis: Università di Bergamo; M. Ciccarelli: Universidad de Alicante; Università di Brescia and CESPRI, Università Bocconi.

**FROM GIBRAT'S LEGACY TO GIBRAT'S FALLACY.
A BAYESIAN APPROACH TO STUDY THE GROWTH OF FIRMS**

Elena Cefis, Matteo Ciccarelli and Luigi Orsenigo

A B S T R A C T

We aim at testing Gibrat's Law, a building block of the corporate growth dynamics. Using a Bayesian statistical framework that nests previous approaches, we provide evidence against Gibrat's law on average, within or across industries. Notwithstanding, data show only weak evidence of mean reversion, i.e. initial larger firms do not grow relatively slower than smaller firms. Moreover, differences in growth rates and in size steady state are persistent and firm-specific. Previous results confirming Gibrat's argument are likely to be incorrect being based on models that do not exploit appropriately all information contained in a panel data set.

KEYWORDS: Gibrat's Law; Panel Data; Heterogeneity; Bayesian Estimation, Gibbs Sampling

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” (Gibrat’s Law), which assumes that firm’s size follows a random walk and hence that firm’s growth is erratic. As a consequence there is no convergence within or across industries, and no stable or predictable differences in growth exist either in the short or in the long run. Rather, growth is 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 benchmark for discussing the processes of firms’ growth. Nevertheless, it is in contrast 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, many 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) (Baily and Chakrabarty, 1985; Mueller, 1990; Geroski et al., 1993; Cefis and Orsenigo, 2001).

However, the hypothesis that firms’ growth rates are erratic is often taken almost for granted and even 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 and Griliches 2000, McCloughan 1995).

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 an equation having the following form:

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

¹For a recent discussion, see Sutton (1997) and Bottazzi et al. (2000).

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 null hypothesis $H_0 : \beta = 1$ could not be rejected versus the alternative $H_1 : \beta < 1$. Empirical results are far from being uncontroversial. Some studies confirm the view that firm's size indeed follows a random walk ($\beta = 1$), at least as large firms are concerned. Nevertheless, the considerable body of results that reject the null hypothesis has been interpreted as data showing a reversion to the mean, at least in the sense that small or young firms tend to grow faster on average - but with a higher variance - than large or old ones². In other words, according to this empirical finding, in the long run firms would tend to converge to an ideal size (steady state), common to all firms.

In this work we suggest that both the random walk and the mean reversion view might be unwarranted. Previous literature attempts to verify Gibrat's law and its consequences using cross-section regressions or short-panel econometric techniques with homogeneity in the parameters across units and over time. We consider both approaches as *problematic*. The former ignores the information contained in unit-specific time variation in growth rates. The latter forces the parameters to be the same across individuals, thus pooling possibly heterogeneous units as if their data were generated by the same process, even if it considers information available for all periods and all cross sectional units.³ A fixed effect bias may emerge as a consequence of this assumption, as it is well known in the panel data literature (e.g. Hsiao et al. 1999).

The general conclusion of our study is a skepticism towards Gibrat's law as an unconditional referent of corporate growth dynamics. At the same time we point out that the *mean-reversion* argument must be qualified and re-formulated in a broader context. These considerations are justified by the fact that most results previously obtained could be econometrically biased. Even though they convey attention on important aspects of the data, they are based on methodologies that force units to be homogeneous and hence that exploit only one side of the information contained in a panel data set. As we will discuss, these methodologies are inappropriate for studying Gibrat' law and its implications.

The present study shares part of the view expressed in Goddard et al. (2002). The authors compare the properties of the standard cross sectional test of Gibrat's law with those of three alternative panel unit root tests, by means of simulated and real data on Japanese manufacturing. They conclude that Gibrat's law should be rejected, based on the idea that *cross sectional procedure produces biased parameter estimates and the test suffers from a loss of power if there are heterogenous individual firm effects while suitably designed panel tests avoid these difficulties*.

²Differentials 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).

³Notice that there is not much difference between the cross section and the panel data approaches if in the latter we force the parameters to be exactly equal in all units, a case that should be formally tested instead.

We share the general view that a panel data approach is a better vehicle to investigate Gibrat’s argument than cross section, but we argue that it is not just a matter of testing procedure that makes the difference, if the panel data model is not well suited. Our work proposes a general framework whose main characteristic is *plain heterogeneity*, i.e. heterogeneity in the intercept and in the slope of the statistical model, which has the twofold useful feature of exploiting **all** information contained in a panel data set and nesting previous studies on the same topic. Given that there are too many parameters to estimate relative to the number of time series observations for each cross sectional unit, our point of view is Bayesian. This means that a flexible prior on the parameters must be combined with information contained in the data (likelihood) to obtain posterior estimates. As it will be discussed in section 2, the procedure solves the small sample problem encountered by estimating separately using only the observations on unit i , since Bayesian estimates are exact regardless of the sample size, and, at the same time, it does not require the stringent assumption of equality of the coefficients across units. The chosen prior shares features with those of Lindlay and Smith (1972), Chib and Greenberg (1995, 1996), and Hsiao et al. (1999), and it is specified to have a hierarchical structure, which allows for various degrees of ignorance in the researcher’s information about the parameters. Both the econometric argument and the Bayesian technique are also related to Canova and Marcet (1998), who studied the issue of convergence of per-capita income across economic regions.

Using two different data sets, we find that: (i) the main assertion of the Gibrat’s law that firm’s growth rates are erratic over time is not true on average; (ii) there is a strong possibility that previous results, based on cross sectional or pooled panel data models are econometrically biased because they do not exploit all information contained in the data and hence they misspecify the econometric model without considering heterogeneity, even among firms of the same industry; (iii) estimated steady states differ across units, and firm sizes and growth rates do not converge within the same industry to a common limiting distribution; (iv) initial conditions are important determinants of the estimated distribution of steady states, but there is weak evidence of mean reversion, i.e. initial larger firms do not grow relatively slower than smaller firms. In other words, differences in growth rates and in the size steady state are firm-specific, rather than size-specific; (v) differences in firm size and in growth rates are likely to reduce at a very slow rate but they do not seem to disappear over time, i.e., they persist; (vi) the specified model provides an adequate fit to the data and results do not change under plausible alternative models. In other words, they are robust to more general families of prior information.

The paper is structured as follows. Section 2 discusses the statistical model. Section 3 describes data and comments on the estimation results. In section 4 we check the robustness of the results. Section 5 concludes, while details of the estimation and testing techniques are given in appendix.

2. THE ECONOMETRICS

2.1. Model specification

Given that our observations are collected across units and time, the evolution of size for all units is determined by a doubled indexed stochastic process $\{S_{it}\}$, where $i \in I$ indexes firms, $t = 0, 1, \dots$ 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

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

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 distributed with mean β_0 and variance σ^2 , then $\ln S_{it}$ follows a random walk and the limiting distribution of S_{it} is lognormal.

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

$$\ln S_{it} = \beta_0 + \beta \ln S_{it-1} + u_{it} \tag{2.1}$$

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

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

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

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

$$\ln \frac{S_{it}}{S_{it-1}} = \beta_0 + \beta_1 \ln S_{it-1} + u_{it}$$

where clearly $\beta_1 = \beta - 1$. In this case Gibrat's law is confirmed if data do not reject the null $\beta_1 = 0$.

In this work we follow a similar specification. The main difference is that we study the behavior of the (log of) each unit's size relative to the average, i.e., of the variable $g_{it} = \ln(S_{it}/\bar{S}_t)$, 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.

Therefore, we specify the following statistical model

$$g_{it} = \alpha_i + \rho_i g_{it-1} + \eta_{it} \quad (2.2)$$

where 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.

Notice that this specification is more general than either a simple cross sectional analysis or a homogeneous dynamic panel data model. On the one hand, Eq. (2.2) allows for a more efficient use of the information contained in the time dimension of the panel since the parameters of the model are estimated by using the firm sizes for all t 's. On the other hand, we are not forcing the parameters to be the same across units, as it is usually assumed in the empirical literature on Gibrat's law. The reason for considering different intercepts for each unit is simply to avoid the well known fixed effect bias due to lack of consideration of the heterogeneity typically found in *micro* data. Moreover we think that, even with a fixed-effect specification, the assumption of common slope is too restrictive. If units are heterogeneous in the slopes but the statistical model does not take this feature into account, then bias and inconsistency problems arise.⁴ It is not difficult to show that the neglect of coefficient heterogeneity in dynamic models creates correlation between the regressors and the error term and causes serially correlated disturbances (Pesaran and Smith, 1996, Hsiao et al., 1999). Hence, any traditional estimator is biased and inconsistent, the degree of inconsistency being a function of the degree of coefficient heterogeneity and the extent of serial correlation in the regressors.

Our main point in this paper is that the traditional results on Gibrat's law may be econometrically biased for the lack of consideration of possible heterogeneity in the data. This argument motivates the choice of the model specification (2.2), which is flexible enough to formally test the restrictions that previous studies impose on the data, i.e., $\alpha_i = \alpha$, and $\rho_i = \rho$, $\forall i$.⁵

Provided we can estimate the short run parameters for each unit, we are also able to estimate the steady states directly. Therefore we can test the Gibrat's law both for each single firm and on average, and we can separately examine three further implications of the law. Precisely, we are able to estimate the speed of adjustment $(1 - \rho_i)$ of each unit to its own steady state $(\alpha_i / (1 - \rho_i))$, a question

⁴Notice that the opposite is not true, in that a well specified heterogeneous model nests a model without such heterogeneity.

⁵In fact, other studies (e.g. Bottazzi et. al., 2000) sometimes estimate $g_{it} = \rho g_{it-1} + u_{it}$, forgetting the specific effect α_i , as if the expected proportionate rate of growth were zero. Even with the kind of normalization used in our paper, this is not a correct approach. If $u_{it} = \alpha_i + \eta_{it}$, then ols estimates are inconsistent because $E(g_{it}\alpha_i) \neq 0$ for each t , and inefficient given that the u_{it} 's are serially correlated.

related to the mean reversion argument and the decrease in the variance of the firm size over time. Second, we can verify whether steady states are all equal across units. Finally, if steady states are not common, the model specification can easily be used to test whether these differences across firms are transitory or permanent, i.e., whether there is persistence in size differences.

Given that there are too many parameters relative to the number of time series observations for each cross sectional unit, we impose a Bayesian prior on the parameters to be combined with information contained in the data (likelihood) to obtain posterior estimates. The procedure solves the small sample problem encountered by estimating separately using only the observations on unit i , since Bayesian estimates are exact regardless of the sample size, and, at the same time, it does not require the stringent assumption of equality of the coefficients across units.⁶

Let $\theta_i = (\alpha_i, \rho_i)'$. Eq. (2.2) can then be written in a more compact form as

$$g_{it} = X'_{it}\theta_i + \varepsilon_{it} \quad (2.3)$$

where $X_{it} = (1, g_{it-1})'$. Though allowing heterogeneity, the imposed prior distribution assumes that the intercept and the slope of the model do not differ too much across units. Concretely, the population structure is modelled as:

$$\theta_i \sim N(\bar{\theta}, \Sigma_\theta) \quad (2.4)$$

where $\bar{\theta}$ and Σ_θ are common to all individuals. In other words we assume that the parameters of each cross sectional units come from a distribution which is common to all firms. The variance of this distribution then determines the degree of uncertainty that the researcher has about the mean. Notice that this assumption is more general than forcing the parameters to be the same for each unit. This limiting case can be obtained by imposing $\Sigma_\theta = 0$. Our opinion is that these restrictions might be formally tested instead of simply imposed.

Notice also that (2.4) is just a prior assumption and must then be combined with the data to obtain posterior estimates. If data are sufficiently informative, the posterior need not be the same as the prior, as it will be clear from the estimation results.

For the prior information to be complete, we assume a normal-wishart-gamma structure

$$\bar{\theta} \sim N(\mu, C) \quad (2.5)$$

$$\Sigma_\theta^{-1} \sim W(s_o, S_o^{-1}) \quad (2.6)$$

$$\sigma_i^2 \sim IG\left(\frac{v^2}{2}, \frac{v^2\delta^2}{2}\right) \quad (2.7)$$

⁶See Canova and Marcet (1998) and Hsiao et al. (1998) for a more detailed discussion on this points.

where the notation $\Sigma_\theta^{-1} \sim W(s_o, S_o^{-1})$ means that the matrix Σ_θ^{-1} is distributed as a Wishart with scale S_o^{-1} and degrees of freedom s_o , and $\sigma_i^2 \sim IG\left(\frac{\nu^2}{2}, \frac{\nu^2 \delta^2}{2}\right)$ denotes an inverse gamma distribution with shape ν and scale $\nu\delta$. The hyperparameters μ, C, s_o, S_o, ν and δ are assumed known. Independence is also assumed throughout.

The entire specification (2.3) through (2.7) is standard in Bayesian literature (e.g. Gelfand et al., 1990) and has the advantage of being sufficiently flexible to answer the kind of questions posed in this work. In particular, notice that this specification easily nests the pure cross section, the fixed effect and the pure time series models. In fact, setting $\Sigma_\theta = 0$ is equivalent to impose equality of coefficients across firms. If the prior variance-covariance matrix is zero, no cross sectional heterogeneity is present and the parameter vector θ_i is pooled towards the common cross sectional mean $\bar{\theta}$. This setting therefore would replicate the cross sectional and the homogeneous panel data analysis, while the fixed effect specification is obtained by forcing only the variance of ρ_i, σ_ρ^2 , to be zero. In discussing the empirical results, we will call these two specifications the *pool ols* and the *fixed effect* respectively. On the contrary, if $\Sigma_\theta \rightarrow \infty$, the prior information is diffuse. This means that the degree of uncertainty about the mean is infinite, and hence that estimated parameters for different firms are similar to those obtained applying OLS to (2.3) equation by equation. In other words, when $\Sigma_\theta \rightarrow \infty$ only the time series properties of each g_{it} are used and the estimation results will resemble the *mean group* estimator proposed by Pesaran and Smith (1995). Finally when Σ_θ is a finite, positive definite matrix, the coefficients are estimated using information contained both in the cross-section and in the time series dimensions of the panel. In the empirical section we will denote the estimation results relative to this setting as *Bayes*. In a recent work, Hsiao et al (1999) establish the asymptotic equivalence of the Bayes estimator and the mean group estimator, and show that the Bayes estimator is asymptotically normal for large n (the number of units) and large T (the number of time periods) as long as $\sqrt{n}/T \rightarrow 0$ as both n and $T \rightarrow \infty$. They also show that the Bayes estimator has better sampling properties than other consistent estimators for both small and moderate T samples. Concretely, the bias of the Bayes estimator never exceed 10% for $T = 5$, while for $T = 20$ it is always less than 2%.

Posterior inference can then be conducted using the posterior distributions of the parameters of interest. Specifically, in discussing the validity of the Gibrat's law, we will be interested in examining whether the mean coefficient $\bar{\rho}$ is equal to one, as well as, how large is the percentage of firms for which the null ($\rho_i = 1$) is not rejected.

In discussing the implications of the law, we also need to verify the null hypothesis $\alpha_i/(1 - \rho_i) = \alpha_j/(1 - \rho_j) \forall i, j$, i.e., the null that the steady states are the same across units versus the alternatives that they are different. The rejection of the null $\alpha_i/(1 - \rho_i) = \alpha_j/(1 - \rho_j) \forall i, j$ provides evidence in favor of

lack of unconditional convergence to a common steady state. The final question, then would be if the initial differences in size are going to persist as time goes by. This issue is examined by running a cross sectional regression of the form:

$$\widehat{SS}_i = c + bg_{i0} + \omega_i \quad (2.8)$$

where \widehat{SS}_i is the mean of the posterior distribution of the steady state for unit i , and g_{i0} is its initial (scaled) size. A positive b would indicate that the distribution of initial size matters for the cross sectional distributions of steady states, while the magnitude of this estimate will provide an indication of how persistent these differences are.

2.2. Estimation and testing

The posterior distributions of the parameters of interest are obtained, as already remarked and as detailed in the appendix, by combining the prior information with the likelihood. More formally, if $\psi = (\theta_i, \bar{\theta}, \Sigma_\theta, \sigma_i^2)'$ is the vector of unknown parameters, and y represent the data, the Bayes rule

$$p(\psi | y) \propto p(\psi) l(y | \psi)$$

can be applied to obtain the joint posterior distribution of $\psi = (\theta_i, \bar{\theta}, \Sigma_\theta, \sigma_i^2)$. The marginal distribution of each element of ψ can be derived by integrating out the others. Given the complexity of our specification, this integration is analytically intractable and must rely on a numerical method. We use the Gibbs sampling. The ergodic mean of the marginal posterior distributions obtained from the Gibbs sampler are taken as our point estimate.

The null hypotheses $\bar{\rho} = 1$, $\rho_i = 1$, $\rho_i = \rho_j$, and $SS_i = SS_j, \forall i, j$ are verified by calculating the (log of) Posterior Odds ratio (PO) as in Leamer (1979) and Sims (1988). The null is not rejected whenever the computed statistics is positive. We also compute the largest prior probability to attach on the alternative in order for the data not to reject the null. This statistics, that we will call ω^* , represents the degree of confidence the researcher should attach on the null so that the data do not overturn her prior beliefs. Small values of this measure are the signal that the researcher should put more weight on the null to not reject it, or, equivalently, that the null is unlikely.

In order to choose a preferred model (e.g., $\Sigma = 0$ vs. $\Sigma > 0$) we compare the posterior predictive power of the two models instead of relying on PO ratios. The reason is that the PO ratio may not provide good inference in the case of comparing the no pooling ($0 < \Sigma < \infty$) and the complete pooling ($\Sigma = 0$) models. The argument can be sketched in a simple way. Let $y_i = (g_{i1}, \dots, g_{iT})'$. Assuming known variances σ_i^2 , the two models can be summarized as:

$$0 < \Sigma < \infty : p(y | \theta_1, \dots, \theta_n) = \prod_{i=1}^n N(y_i | \theta_i, \sigma_i^2), \quad p(\theta_1, \dots, \theta_n) \propto 1$$

$$\Sigma = 0 : \quad p(y | \theta_1, \dots, \theta_n) = \prod_{i=1}^n N(y_i | \theta_i, \sigma_i^2), \quad \theta_1 = \dots = \theta_n, \quad p(\theta) \propto 1$$

If we use the PO ratio to choose or average among these models, the ratio is not defined because the prior distributions are improper, and the ratio of density functions is $0/0$. Therefore, we should assign either proper prior distributions or improper prior carefully constructed as limits of proper prior. In both cases results are unsatisfactory, as shown by Gelman et al. (1995, p. 176-177).

The details of estimation and testing techniques are in appendix.

3. EMPIRICAL RESULTS

In this section we describe the data sets used in the analysis and present the empirical results. The latter are shown in Figures 1-9 and Tables 1-3.

3.1. The Data

The issue analyzed in this paper seems particularly relevant in the specific case of the pharmaceutical industry. The latter can be considered an ideal case where the process of growth behaves in accordance with Gibrat's Law, due to the peculiar pattern of innovation of that industry. As a matter of fact, innovation in this industry has often been described and conceptualized as a pure "lottery model", where previous innovation (in a particular submarket) does not influence in any way current and future innovation in the same or in other submarkets (Sutton, 1999). Thus, to the extent that firms' growth is driven by erratic innovation, it should also be erratic.

Data come from the data set PHID (Pharmaceutical Industry Database) developed at University of Siena. It collects data on the top incumbents in the seven major western markets (France, Germany, Italy, Spain, UK, Canada, and USA) during the period 1987-1998. As we are interested in the process of internal growth of firms, we use sales as proxy for firm size, considering as unity of analysis the international firm.⁷ Therefore, sales for each firm stand for the sum of their sales in each of the national markets. Furthermore, in order to control for merger and acquisition processes during the period, 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 during the period of observation. Hence, if two firms merged during the period, we consider them merged from the start, summing up their sales from the beginning.

The data set is constituted of 210 firms that are the results of the intersection of the top 100 (in terms of sales) in each national market, at the beginning of the

⁷Other proxies can be used. We tried invoices and the number of employees as alternative proxies for size without experimenting different qualitative results.

period of observation. A balanced panel is obtained by censoring the entrants among the top ranks, using 199 observations out of 210.

Figures 1 report the histograms at each time period for these firms. The distributions do not show important departures from stability over time. The only relevant feature which is worthwhile observing is that firms whose size was relatively low at the beginning (left tale of the empirical distributions, years 87/88) have probably moved towards the center of the distribution by the end of the sample. We will come back to this point in section 4, where the robustness of the results is checked by specifying differently the distributional assumption both of the error term ε_i and of the population structure θ_i .

3.2. The speed of convergence

The first set of results which is worthwhile commenting is contained in Table 1 and in Figure 2.

For four different settings of Σ , Table 1 reports the following information. The first four lines summarize the posterior distribution of $\bar{\rho}$, showing the mean, the median and the 68% confidence bands. Line 5 reports the value of the posterior density computed at the posterior means of the parameters. Line 6 reports the percentage of firms for which we could not reject the null $\rho_i = 1$, while in line 7 and 8 the average $\log(PO)$ and the ω^* are computed to test for unit root across firms. Lines 9 and 10 reports the same statistics for the firms whose growth behaves erratically ($\rho = 1$) and for those this is clearly not the case ($\rho < 1$). Lines 11-12 show the statistics for testing the equality $\rho_i = \rho_j$. Finally lines 13-14 report the statistics for testing the equality $SS_i = SS_j$, with $i \neq j$. The entire posterior distribution of $\bar{\rho}$ under the four settings is also shown in Figure 2.b, while the histograms of mean individual ρ_i are reported in Figure 2.a. Figure 2.c plots the scatter relation between the mean convergence rates $(1 - \rho_i)$ and the initial firms' sizes, to examine the mean reversion argument.⁸

Several important facts can be discussed.

First of all, notice that by forcing the units to have the same coefficients α_i and ρ_i (pool ols) or just the same slope (fixed effect) we obtain the results generally obtained in the literature. Therefore, under this set of restrictions the model is indeed able to reproduce the standard cross-sectional/pooling-panel and the fixed effect regression results (e.g. Goddard et al (2002), p.417, table 1). In these two cases we cannot reject the null hypothesis that the average ρ is equal to 1, because the $\log(PO)$ ratio is positive, meaning that the posterior odds favor the null, and the largest prior probability we should assign on the alternative in

⁸The histograms and the scatter plots are based on the average of the estimated posterior distribution of ρ_i firm by firm. The information contained in the histogram is therefore different from the one contained in the posterior distribution of $\bar{\rho}$ which represents the common part of ρ_i across firms, or the central value to which ρ_i would collapse if there were no heterogeneity in the data.

Figure1. Histograms of observed data
Pharmaceutical industry

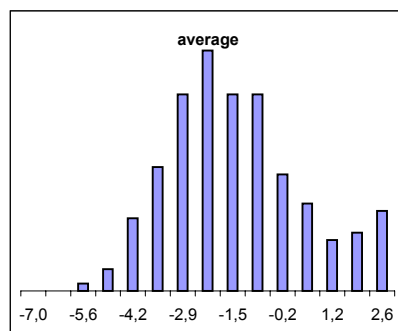
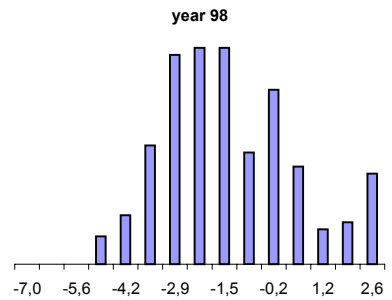
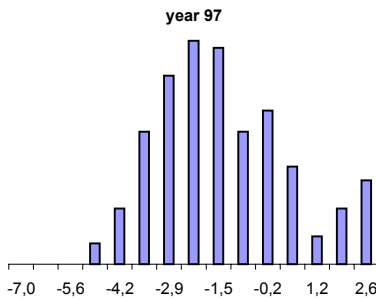
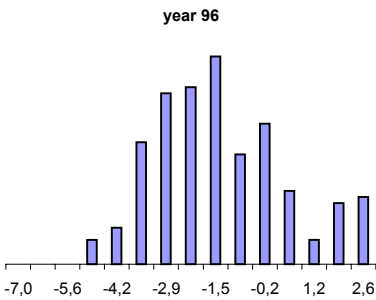
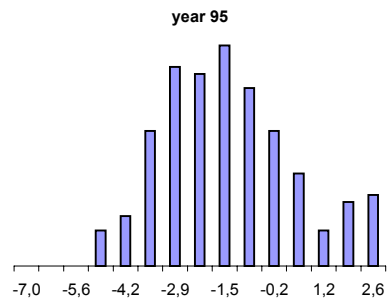
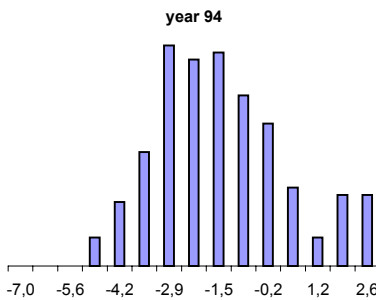
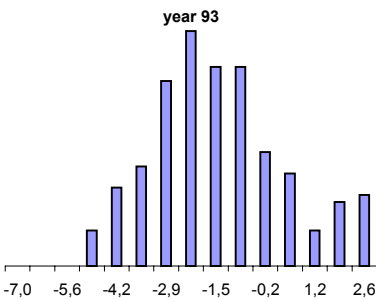
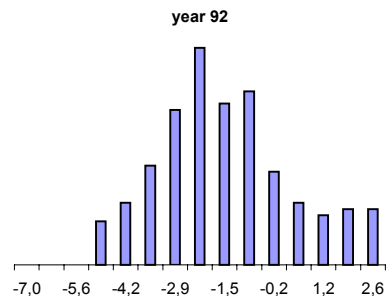
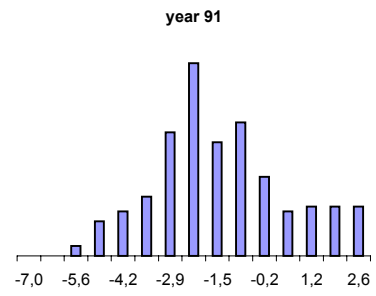
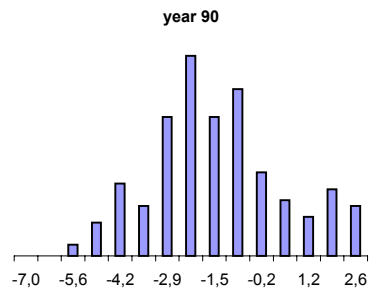
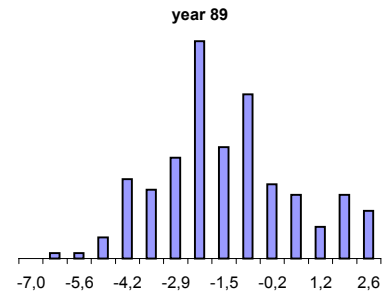
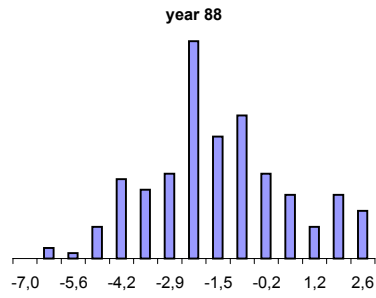
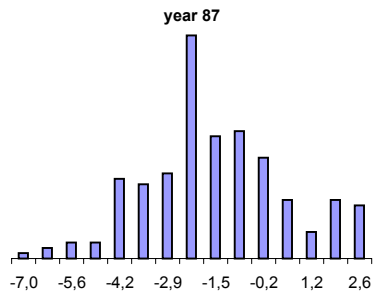
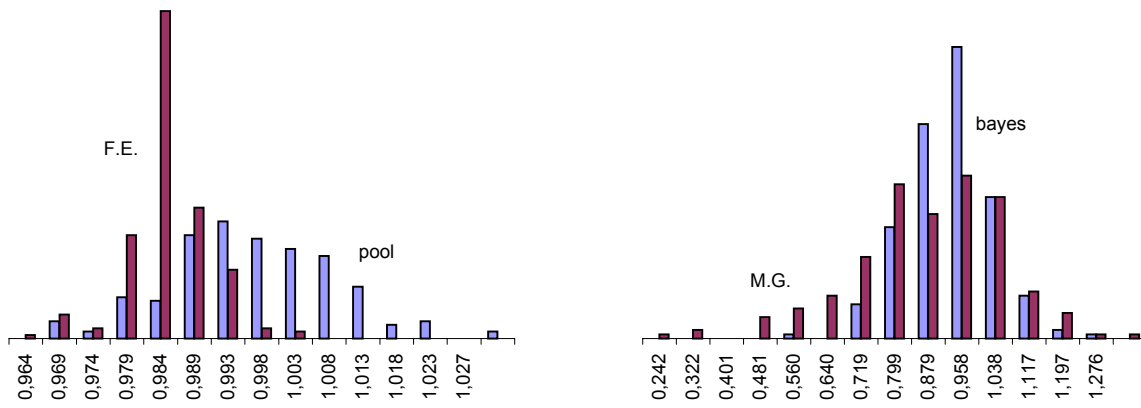
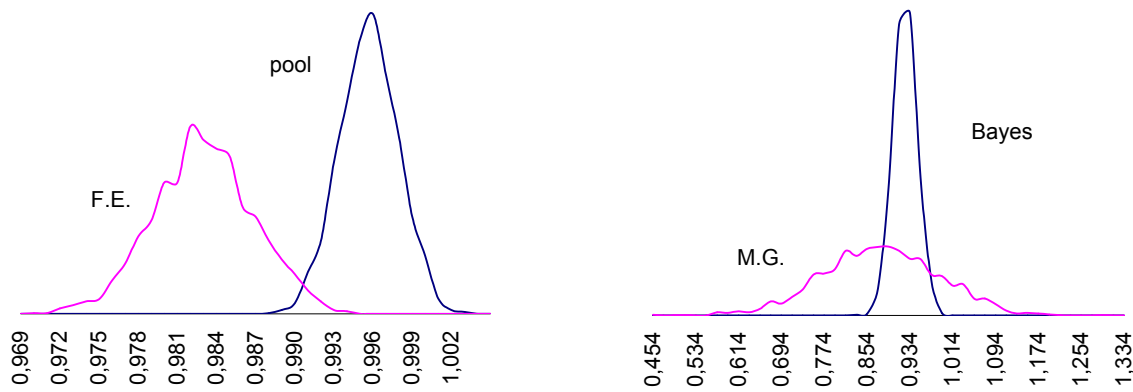


Figure 2. Convergence rates: pharmaceutical industry

(a) Histogram of individual $\rho(i)$



(b) Posterior distribution of average ρ



(c) Mean reversion

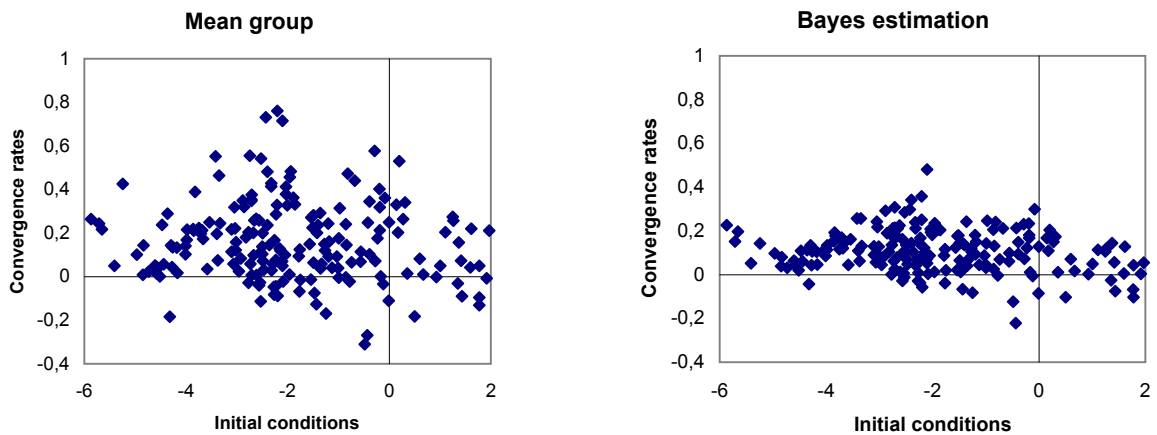


Table 1. Estimation and testing: Pharmaceutical industry

		pool ols		fixed effect		mean group		bayes	
		<i>Posterior distribution of ρ</i>							
1	16%	0,9924		0,9787		0,7612		0,8924	
2	mean	0,9948		0,9827		0,8686		0,9142	
3	median	0,9948		0,9827		0,8688		0,9141	
4	84%	0,9972		0,9868		0,9774		0,9361	
5	P.D.	4455,59		4505,75		4704,13		4677,45	
		<i>Testing $\rho = 1$</i>							
6	% $\rho = 1$	86,08		84,09		23,54		21,71	
7	ln(PO)	0,39		0,34		-0,99		-6,61	
8	ω^*	0,92		0,85		0,67		0,01	
		<i>$\rho = 1$ $\rho < 1$</i>		<i>$\rho = 1$ $\rho < 1$</i>		<i>$\rho = 1$ $\rho < 1$</i>		<i>$\rho = 1$ $\rho < 1$</i>	
9	ln(PO)	1,92	-1,35	1,05	-0,48	0,67	-2,06	0,58	-1,67
10	ω^*	0,97	0,61	0,94	0,77	0,90	0,50	0,91	0,60
		<i>Testing $\rho(i) = \rho(j)$</i>							
11	ln(PO)	-107,12		-15,94		-131,19		-67,87	
12	ω^*	0,00		0,00		0,00		0,00	
		<i>Testing $ss(i) = ss(j)$</i>							
13	ln(PO)	-2,37		-2,84		na		-703,16	
14	ω^*	0,09		0,06		na		0,00	

order for the data not to reject the null is very high ($\omega^* = 0.92$, and $\omega^* = 0.85$ respectively). The latter result means that we must have almost zero confidence in the null for the data not to overturn our prior beliefs. In other words when we force coefficients to be the same across units, on average the null is a posteriori highly likely. Results are confirmed if we test for unit root firm by firm, under the same set of restrictions. Concretely, for 84-86 percent of the firms in the sample we cannot reject the null of $\rho = 1$ (line 6). The same information is contained in Figure 2.a and 2.b, where the histograms of the posterior mean of the parameter ρ_i for each firm and the average posterior mean $\bar{\rho}$ are respectively plotted. The limiting distribution of the autoregressive parameter is not very dispersed around a mean *very close* to one.

When we allow for heterogeneous parameters across units (either $\Sigma \rightarrow \infty$ or $0 < \Sigma < \infty$), the average ρ ranges from 0,86 to 0,91 (line 2, *mean group* and *Bayes*). The average log(PO) is negative and favors the alternative versus the null of $\rho = 1$, while ω^* ranges from 0.01 (*Bayes*) to 0.61 (*mean group*), meaning that, especially for the Bayes estimator, a high prior probability should be attached on the null in order for the data not to reject it (lines 7-8). In other words, when the coefficients are estimated using the information contained both in the cross section and in the time series dimension of the panel, the size of the firms does not follow a random walk on average. The same statistics by firm (lines 6, 9-10) confirm that only for 21-23 percent of firms we cannot reject the null. A similar information is contained in Figures 2.a and 2.b where, as for the *pool ols* and the *fixed effect* cases, the histograms of the posterior mean of the parameter for each firm and the posterior distribution of $\bar{\rho}$ are plotted. The figures reveal that when we use the information contained both in the cross section and in the time series dimension, data show a considerable dispersion in the estimated distribution of ρ across firms, which renders the a posteriori probability of facing a random walk very unlikely on average. At the same time, the information contained in Fig. 2.a is also a way of testing the null hypothesis $\rho_i = \rho_j$. The substantial dispersion of ρ_i supports the view that the estimated autoregressive coefficients are far from collapsing toward the central value $\bar{\rho}$, and hence that the null hypothesis is likely to be rejected. The statistics shown in line 11-12 confirm this results under the four settings.

Line 5 shows that the posterior density is higher under these two settings than under the *pool* and *fixed effect* specifications. Therefore the models that allow for heterogeneity in the parameters are to be preferred, according to what discussed in the previous section, at least in the sense that they have a better predictive power. Notice that when only the time series dimension is used (*mean group*), given that we have just 12 time observations, a small sample downward bias in the estimation of the average ρ is present, as it is well known. In this case the distribution of average ρ is centered on a lower value than the Bayes estimator, though being more dispersed.

Finally, note that the simple observation of $\rho < 1$ is not sufficient to indicate the existence of “catch-up” or “mean reversion”. On the contrary, there seems to be a weak relation between the initial conditions and the speeds of adjustment ($1 - \rho_i$). As shown in Figure 2.c, there is a slight negative relationship between the two variables, but it seems not enough to claim that initial larger firms grow relatively slower than initial smaller firms. On the other hand, the chart and the results commented above indicate that it is also not true on average that big firms follow a random walk while small firms don’t, as it has been argued in recent works.⁹ Therefore it is not the levelling out in growth rates between large and small firms that bounds the overall rise in the variance of firm sizes, but rather the absence of a unit root on average. At this respect, notice however that $\rho < 1$, and hence, failure of Gibrat’s law, is not incompatible with a growing variance. To show this, assume in our model specification that $\alpha_i = \lambda g_{i0}$. Then the model becomes

$$g_{it} = \lambda g_{i0} + \rho_i g_{it-1} + \eta_{it} \quad (3.1)$$

or, going backward

$$g_{it} = \lambda g_{i0} \sum_j \rho_i^j + \rho_i^t g_{i0} + \sum_j \rho_i^j \eta_{it-j}$$

Therefore the variance of g_{it} is

$$\begin{aligned} Var(g_{it}) &= \left(\lambda \sum_j \rho_i^j \right)^2 Var(g_{i0}) + \rho_i^{2t} Var(g_{i0}) \\ &\quad + \sum_j \rho_i^{2j} \sigma_{i\eta}^2 + \left(\lambda \sum_j \rho_i^j \right) \rho_i^t Var(g_{i0}) \end{aligned}$$

Notice that when $|\rho_i| < 1$ this expression converges to

$$Var(g_{it}) = \left(\frac{\lambda}{1 - \rho_i} \right)^2 Var(g_{i0}) + \frac{\sigma_{i\eta}^2}{1 - \rho_i^2}$$

as t becomes sufficiently large. Therefore it may very well be the case that $Var(g_{i0}) < Var(g_{it})$, even in the case of a failure of the Gibrat’s law and without implying that predictions of g_{it+k} become increasingly uncertain as k gets larger.

Summarizing this first set of results, the above discussion suggests the view that only a small percentage of firm sizes drift unpredictably over time and clearly diverge within industry, while the size of the vast majority of firms in the sample converges to a stable steady state. An interesting question now is to see whether firms converge to the same steady states or not. A negative answer to the question does not necessarily mean that differences in firm sizes are permanent and not transitory, because firms can converge to different steady states and the “biblical prophecy” that small firms may have greater steady states than big ones can hold.

⁹See Lotti et al. (2000), for instance.

3.3. The steady state

Focusing on the cases with higher posterior predictive densities, which in our opinion are the most reasonable ones, the dispersion of steady states is substantial. Figure 3.a plots a histogram of firms estimated posterior steady states.¹⁰ The histogram is constructed so that firms are grouped in 10 classes of steady state size: up to 5%, 6-10%, 11-35%, 36-60%, 61-85%, 86-100%, 101-115%, 116-125%, 126-135%, above 136%, where 100 is the average size, i.e., the steady state level of g_{it} which we would obtain if all the units converged to the same steady state (the bold line in the figure). Clearly the estimated steady state distribution is far from collapsing toward the central value. Table 1 (lines 13-14) reports, as said, the statistics for the hypothesis that the steady states are the same across units. Under the four settings $\log(PO)$ is negative and ω^* is zero, meaning that the null is highly unlikely, or that unless we assume that the alternative is impossible ($\omega^* = 0$), the null hypothesis is always overturned by the data. Whenever the value of the statistics is not available, $\log(PO)$ must be regarded as minus infinity.

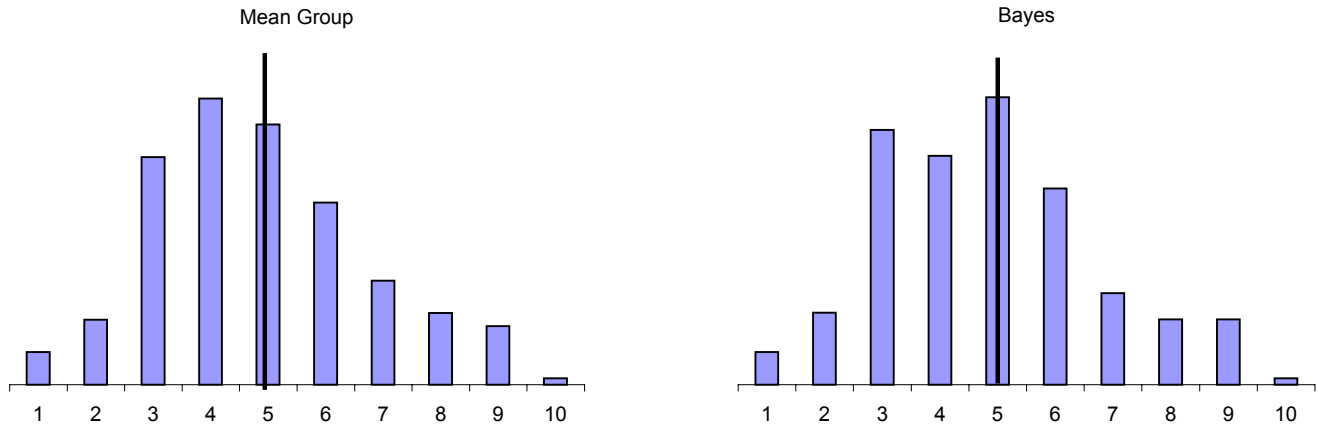
These results indicate that the estimated distribution of steady states is non-degenerate, i.e., firms converge to different steady states. The next question is to find the appropriate variables which may account for the cross-sectional dispersion in estimated steady states. This is not the purpose of the paper, though we can at least propose a natural candidate to explain the limiting distribution. Figure 3.b plots the estimated steady states against the initial (scaled) size levels. It is clear that there is a strong positive relationship, i.e., initially large firms have also the highest steady states and the initial ranking is largely maintained. Figure 3.c measures the strength of this relation running a cross sectional regression of estimated steady states on a constant and the initial condition (see Eq 2.8 above). Clearly the estimated b is positive, i.e., the distribution of the initial size of firms matters for the limiting distribution of the steady states. In other words, differences in firm size are persistent, and given the estimated value of 0.7, one would argue that inequalities are *strongly* persistent. The \bar{R}^2 can be interpreted as a measure of long run mobility.¹¹ A small value would suggest that individual units may move up and down in the ranking, whereas a high value indicates that the ordering in the initial distribution is the same as in the steady states. The latter seems to adjust better to our estimation results. A slope of 0.7 in the cross sectional regression suggests that, on average, the gap between the big and the small firms will be reduced in the limit only by 30%, while $\bar{R}^2 = 0.57$ indicates that the initial conditions alone explain almost 60% of the variation of the cross sectional distribution of steady states. We regard this results as strong evidence in favor of persistence of differences in firm sizes.

¹⁰Whenever $\rho > 1$, we compute steady states using the small sample formula $S(i) = \alpha_i \frac{1-\rho_i^{T+1}}{1-\rho_i} + \rho_i^T g_{i0}$ at each draw of the Monte Carlo in the Gibbs Sampler

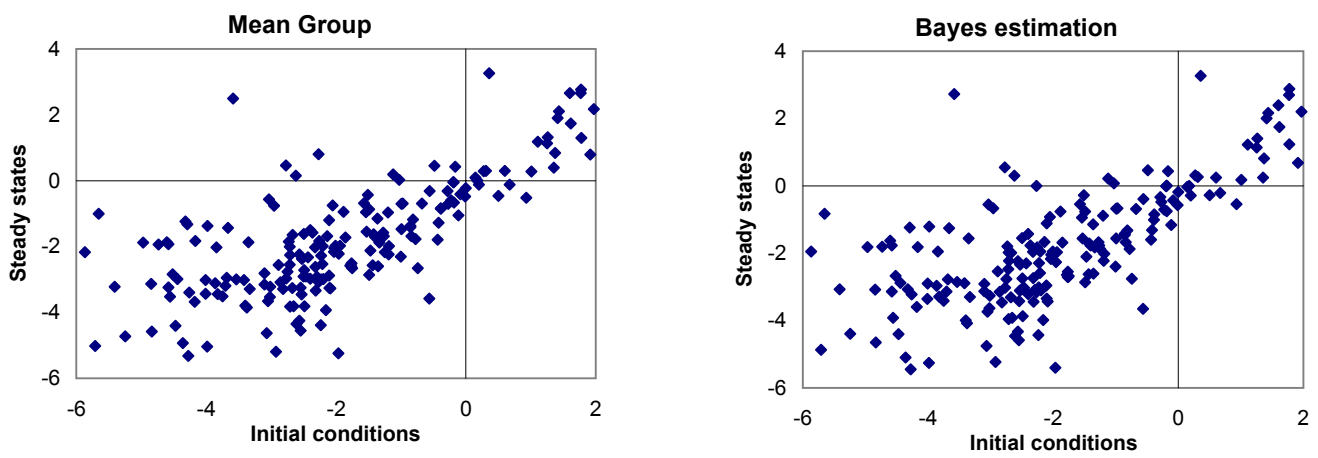
¹¹See Canova and Marcet, 1998, for instance.

Figure 3. The steady state: pharmaceutical industry

(a) Histogram. Posterior estimates



(b) Persistence of differences



(c) test of persistence in differences of firm size

constant	initial condition	R_bar**2	constant	initial condition	R_bar**2
-0,37	0,71	0,57	-0,4	0,695	0,54
(0.12)	(0.044)		(0.105)	(0.052)	

Summing up, the first set of results does not confirm the conventional wisdom that firm’s size drifts erratically over time. Even more interesting, data show that firms converge to different steady states, that their speed of convergence does not depend on initial size and that the steady state is strongly correlated to initial size. In other words, there is only weak evidence of reversion to the mean, and firms’ size differentials tend to persist over time.

4. ROBUST INFERENCE AND SENSITIVITY ANALYSIS

In this section we check the robustness of previous results by performing the analysis with a different data set and different prior assumptions.

4.1. A different data set: UK manufacturing firms

Compared to the Pharmaceutical database, the new data set has a shorter time series dimension and includes different industries belonging to one single country. The analysis is performed on a sample of 267 UK manufacturing firms. Data constitute a balanced panel of five years, from 1988 to 1992. The histograms for each year are plotted in Figure 4. The same remarks as for the previous data set can apply.

The aim of this subsection is twofold. On the one hand we can verify if our main results on Gibrat’s argument are robust to a different data set. On the other hand, given that the new data set contains different industries and hence another level of possible heterogeneity, one could cast light on some other features which explain firms growth across industries.

The model specification is the same as before, and so are the prior assumptions. The estimation results, shown in Table 2 and Figures 5-6, confirm the previous findings. In particular, when both types of information (cross-sectional and time series) are controlled for, we reject the null hypothesis of unit root both on average (table 2, line 7-8) and by firms (table 2, lines 6, 9-10). The average autoregressive parameter ρ ranges from 0,55 (*mean group*) to 0,81 (*Bayes*) in the settings where heterogeneity is controlled for, while it is 0,99 in the plain pool and in the fixed effect cases (Table 2, lines 1-4). The dispersion of the mean estimates of ρ_i by firms (Fig. 5.a) is again a way of rejecting the null hypothesis that $\rho_i = \rho_j$ in all settings (see also Table 1, lines 11-12), although the mean group and the Bayes specification provide more dispersed estimators. There seems to be even less relation between the convergence rate and the initial conditions, with respect to the previous data set (Fig. 5.c), confirming the view that no mean reversion is present in the sample data.

All together, these findings support again the *fact* that firm size *does not* drift unpredictably over time on average. Notice that when we pool the data forcing the parameters to be the same across units, for most firms we do not reject the null that Gibrat’s law holds (Table 2, line 6-8). Even for firms where the test

Figure 4. Histograms of observed data

UK manufacturing

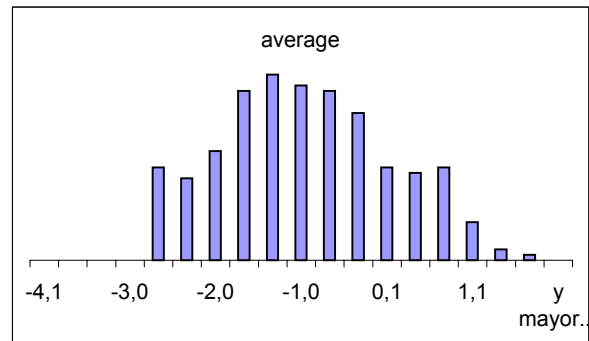
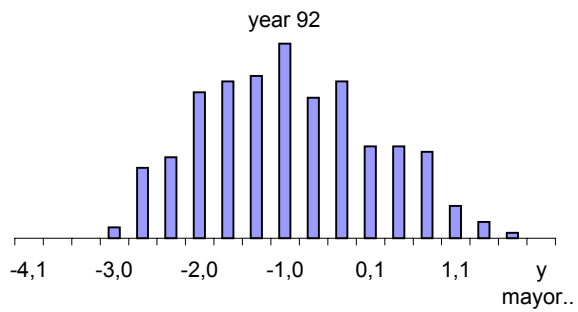
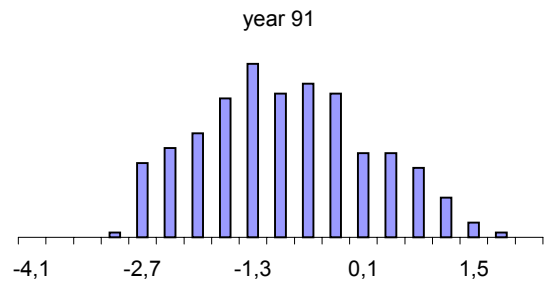
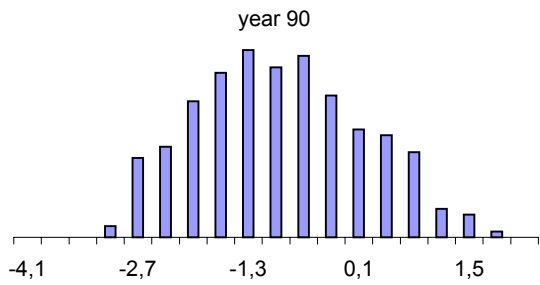
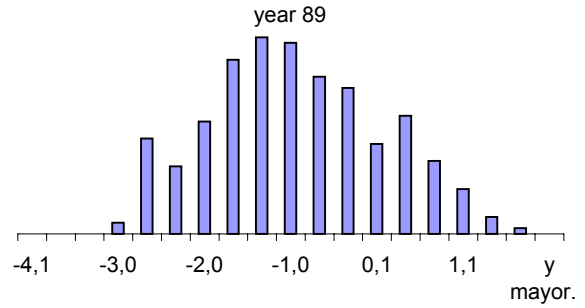
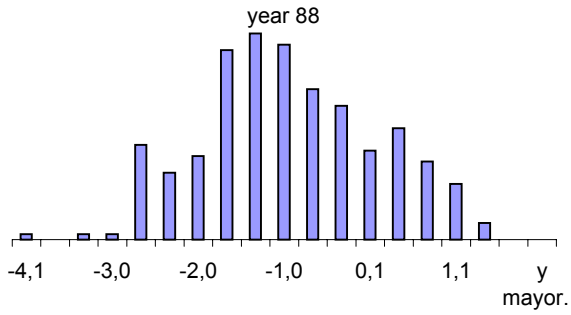
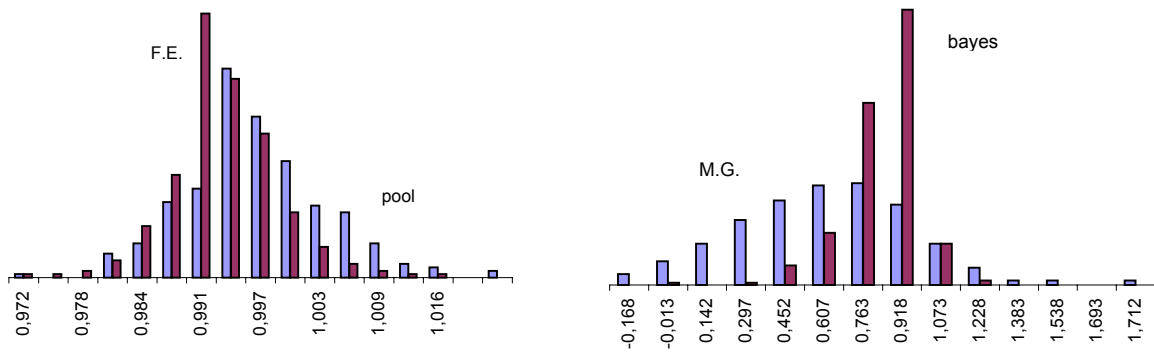
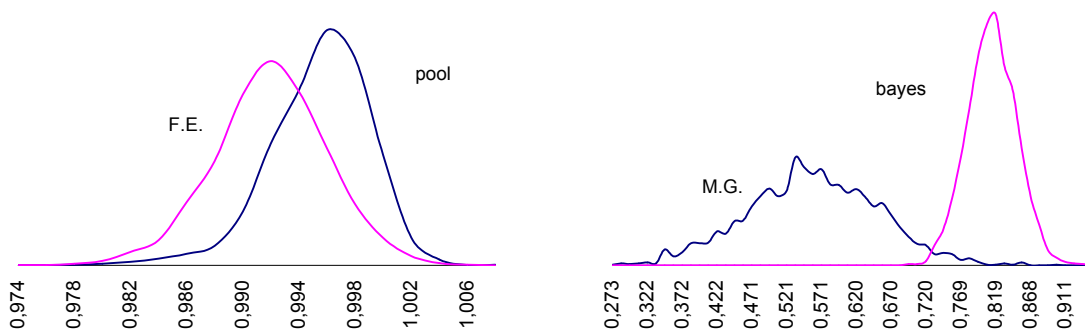


Figure 5. Convergence rates: UK manufacturing

(a) Histogram of individual $\rho(i)$



(b) Posterior distribution of average ρ



(c) Mean reversion

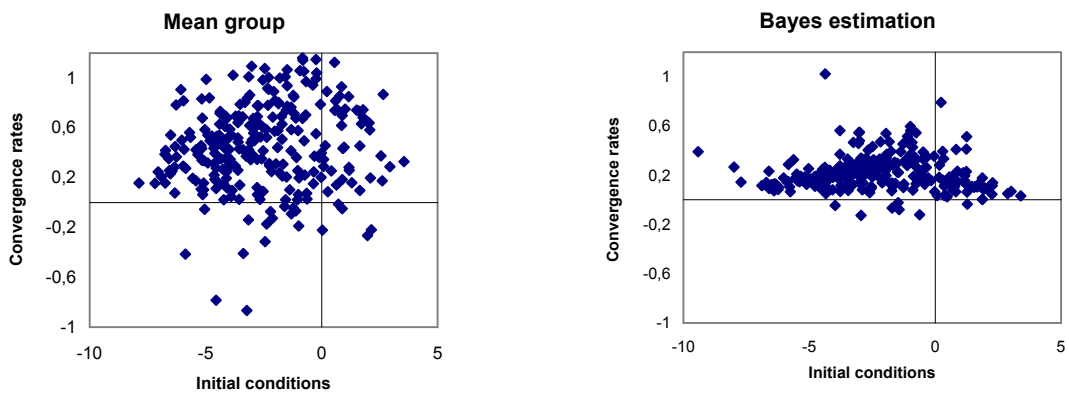


Table 2. Estimation and testing: UK manufacturing

		pool ols	fixed effect	mean group	bayes				
		<i>Posterior distribution of ρ</i>							
1	16%	0,9915	0,9874	0,4596	0,7726				
2	mean	0,9949	0,9913	0,5516	0,8072				
3	median	0,9952	0,9915	0,5508	0,8074				
4	84%	0,9983	0,9954	0,6456	0,8422				
5	P.D.	421,48	445,14	1449,06	845,83				
		<i>Testing $\rho = 1$</i>							
6	% $\rho \geq 1$	87	89,44	8,87	4,23				
7	ln(PO)	1,84	0,68	-11,49	-14,71				
8	ω^*	0,98	0,92	0,00	0,00				
		<u>$\rho = 1$</u>	<u>$\rho < 1$</u>	<u>$\rho = 1$</u>	<u>$\rho < 1$</u>	<u>$\rho = 1$</u>	<u>$\rho < 1$</u>		
9	ln(PO)	1,55	-0,35	1,27	-0,32	0,44	-6,99	0,25	-2,26
10	ω^*	0,97	0,94	0,96	0,93	0,78	0,34	0,83	0,54
		<i>Testing $\rho(i) = \rho(j)$</i>							
11	ln(PO)	-19,39	-11,70	-115,51	-93,67				
12	ω^*	0,00	0,00	0,00	0,00				
		<i>Testing $ss(i) = ss(j)$</i>							
13	ln(PO)	-1,64	-1,79	na	na				
14	ω^*	0,16	0,14	na	na				

rejects the unit root hypothesis, the largest prior probability to attach on the alternative in order for the data not to reject the null is very high (0,93-0,94), meaning that the specifications with little or no heterogeneity in practice would never reject the null of unit root. On the contrary, when we do not impose this restriction only 4 to 9 percent of the firms in the sample behave according to a random walk in the size. Among these firms there seems not to be a clear pattern, at least across industries. In table 3 we report the number of firms and the percentage of firms with $\rho = 1$ for each industry, as well as the scaled size of the average firm in the industry. It can be noticed that only in one industrial group (motor vehicles & parts) all firms have a ρ significantly less than 1, while in the others the percentage of firms whose size drifts unpredictably does not follow a clear pattern at a first glance (see also the histogram below the table). Thus, for instance, the simple argument that big firms may follow a random walk while small firms certainly not is contradicted by the evidence that in the first industry (metal manufacturing), where the average size is relatively high, 50 percent of firms grow erratically, while in industry 14 (footwear & clothing), whose firms are among the smallest, all firms follow Gibrat's law.

Conclusions on the limiting behavior of firm size can be appreciated from Fig. 6. As before, we reject the null hypothesis of equal steady states. The dispersion of estimated steady states is again substantial (Fig. 6.a) and, unless we assume a priori that the alternative is impossible, the null hypothesis will always be overturned by the data (in Table 2, lines 13-14 the $\log(PO)$ is negative and ω^* is approximately zero under the four settings). Moreover differences in firms size are extremely persistent. The evidence contained in Fig. 6.b and c is overwhelming. The position in the initial size distribution of a given unit strongly determines the position of the same unit in the steady state distribution. On average, the gap between the big and the small firms will be reduced in the limit only by 10%, while the initial conditions alone explain more than 80% of the variation of the cross sectional distribution of estimated steady states.

Although results may suffer from a small sample bias, they point out once again the importance of controlling for heterogeneity in the coefficients when a panel data set is used, as well as the importance of using efficiently all information contained in the sample.

4.2. Different prior assumptions: model checking

In this section we assess the fit of the model to the data and the plausibility of the specification for the purposes for which the model has been used. The basic questions are: how much does the posterior predictive distribution fit to data? and how much does previous posterior inference change when other reasonable probability models are used in place of the present model? It is important to distinguish between the two questions because, even if the present model provides an adequate fit to data, the posterior inferences can still differ under plausible

Table 3. UK manufacturing firms by industry

Sectors (SIC80 - digit)	INDUSTRY number	firms	% of $\rho = 1$	average scaled size
Metal manufacturing (22)	1	8	0,50	0,520
Non- metallic manufacturing (24)	2	9	0,11	0,180
Chemical (25 and 26)	3	28	0,07	0,246
Other metal goods (31)	4	21	0,19	0,034
Mechanical engineering (32)	5	59	0,24	0,052
Office and data machinery (33)	6	12	0,08	0,086
Electrical & electronic machinery (34)	7	45	0,16	0,080
Motor vehicles & parts (35)	8	8	0,00	0,195
Other transport (36)	9	6	0,33	0,369
Instrument engineering (37)	10	18	0,17	0,036
Food, drink & tobacco (41/42)	11	9	0,44	0,295
Textiles (43)	12	11	0,09	0,122
Leather goods (44)	13	1	1,00	2,718
Footwear & clothing (45)	14	3	1,00	0,002
Timber (46)	15	6	0,33	0,046
Paper & printing (47)	16	11	0,18	0,086
Rubber & plastics (48)	17	10	0,10	0,062
Other manufacturing (49)	18	2	0,50	0,059

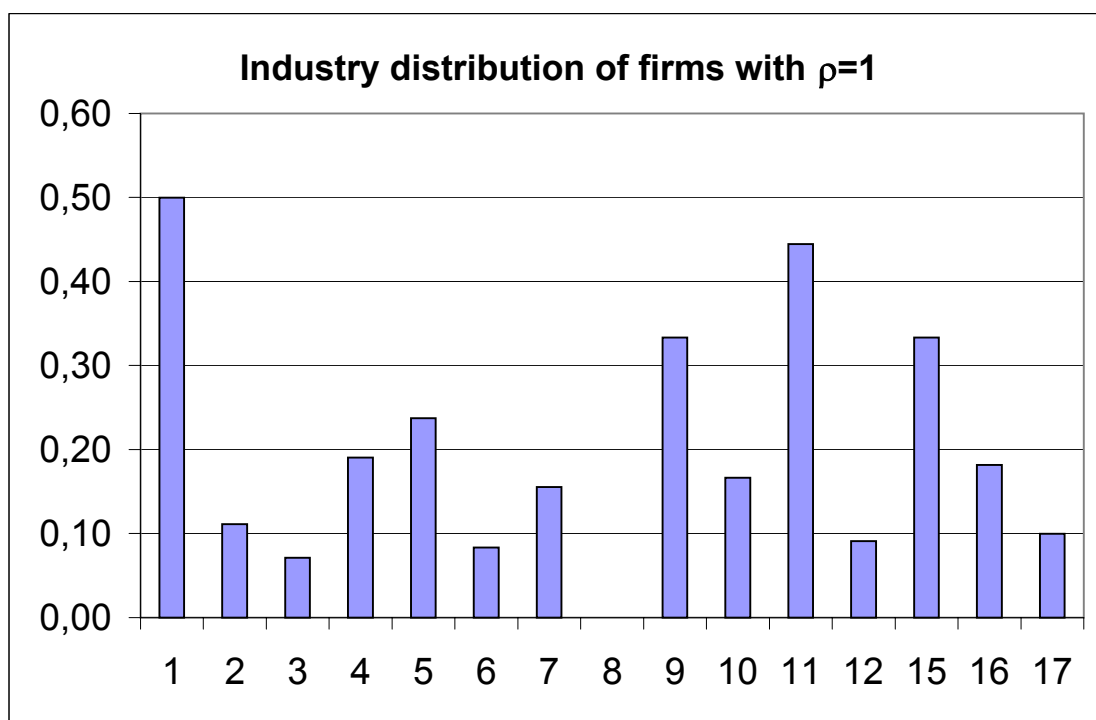
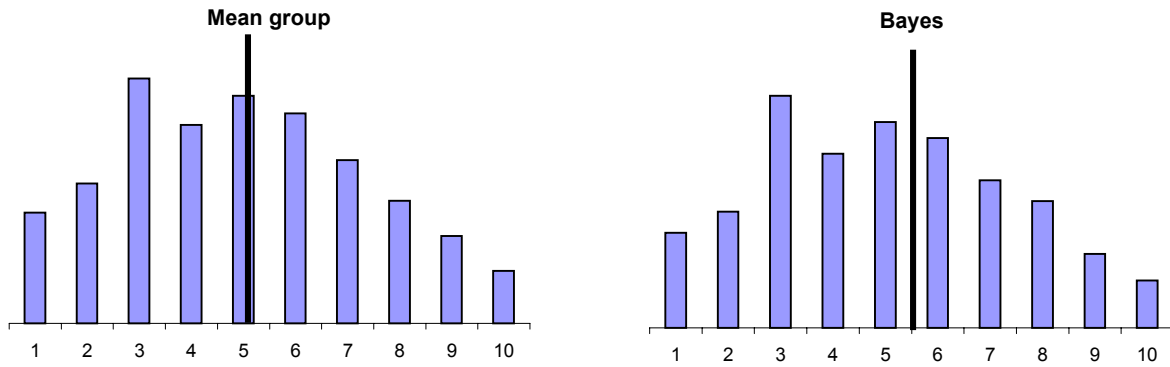
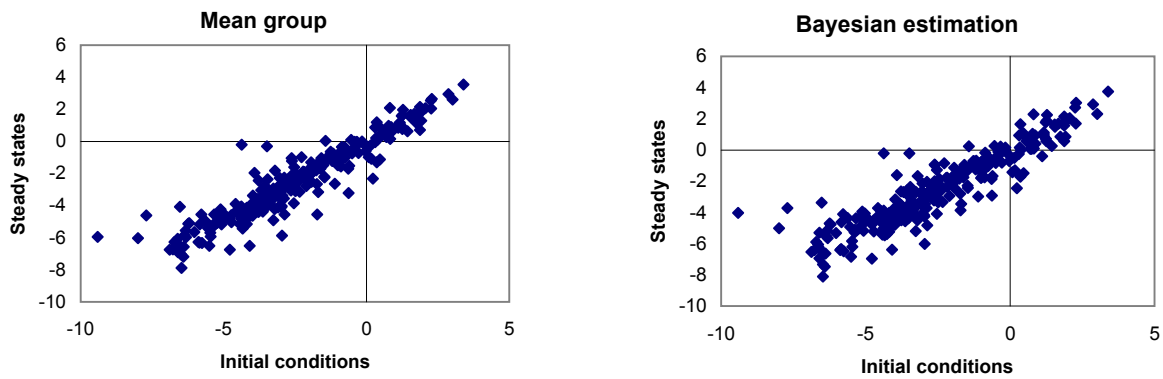


Figure 6. The steady state: pharmaceutical industry

(a) Histogram. Posterior estimates



(b) Persistence of differences



(c) test of persistence in differences of firm size

<u>constant</u>	<u>initial condition</u>	<u>R_bar**2</u>	<u>constant</u>	<u>initial condition</u>	<u>R_bar**2</u>
-0,21	0,928	0,88	-0,29	0,88	0,83
(0.063)	(0.024)		(0.08)	(0.032)	

alternative models.

The technique used for checking the fit of our model to data was to draw simulated values from the posterior predictive distribution of replicated data and to compare these samples to the observed data. This search is combined with a more general family of distributions for both the error term in the regression (2.3) and the population structure (2.4) to analyze the sensitivity of the results. Concretely, we assume the t_ν distribution in place of the normal. The t distribution has a longer tail than the normal and can be used for accommodating occasional unusual observation in the data distribution or occasional extreme parameters in the prior distributions or hierarchical model. Histograms for observed data (Figures 1 and 4) do not reveal extreme values in both data sets. Therefore we expect results based on the normal distribution for data fit the data better and use the t_ν distribution just for sensitivity analysis. In other words, as long as the assumptions on data distribution are concerned, the t_ν is chosen simply as a robust alternative to the normal, provided the degrees of freedom ν are fixed at values no smaller than prior understanding dictates in our example. On the other hand, given the small time series dimension of the panel, we would expect a better predictive performance by assuming a t_ν distribution with finite ν for the parameters of the hierarchical model.

The procedure for carrying out both posterior predictive model check and sensitivity analysis is the following. Let y be the observed data and ψ the vector of parameters (including all hyperparameters). Define y^{rep} as 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:

$$p(y^{rep}) = \int p(y^{rep} | \psi) p(\psi | y) d\psi$$

The discrepancy between the model and the data is measured as suggested by Gelman et al. (1995) 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 (*p-value*) of the quantity, and computed using posterior simulations of (ψ, y) . This p-value is defined as the probability that the replicated data could be more extreme than the observed data, as measured by the test quantity:

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

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 discrepancy measure chosen is the χ^2 discrepancy quantity, an omnibus measure for routine checks of fit, and is defined as

$$T(y, \psi) = \sum_i (y_i - E(y_i | \psi))' [var(y_i | \psi)]^{-1} (y_i - E(y_i | \psi)) \quad (4.1)$$

Other technical details are in appendix.

We compute these statistics, the p-values and the relevant results for inference of interest, first fitting a range of t_ν distributions with $\nu = 10, 50, 100$ for the errors in (2.3), maintaining the normality assumption on θ_i , and then fitting a range of t_ν distributions with $\nu = 5, 10, 15, 50$ for the vector θ_i in (2.4), maintaining the normality assumption on ε_i . Infinite degrees of freedom have already been fitted (the normal-normal data model). Results are reported in Figures 7-9 and Table 3. Most of them refer only to the pharmaceutical industry because those for the UK manufacturing are very similar.

Some comments are in order.

Figure 7 reports scatter plots showing prior and posterior simulation of the chosen test quantity (4.1) based on 2500 simulations from the posterior distribution of (ψ, y^{rep}) for different values of the degrees of freedom of the t . The p-value is computed as the proportion of points in the upper-left half of the plot with respect to an imaginary 45° line. As expected, more extreme values are encountered when a t distribution is fitted for the error terms than when a t replace the normal for the vector θ_i . The values of the posterior predictive density in table 3 confirm this finding. Among the specifications fitted, the one assuming normal errors ($\nu = \infty$) has the highest predictive power (P.D. = 4677,45). Under the assumption of normality of the error term ε_i , the t_{15} distribution for θ_i is also well performing (P.D. = 4342,99). Notice that the p-values when $\theta_i \sim t_\nu$ is always around 0.02, regardless of ν . Figures 8-9 report the histograms of replicated data for both data sets. They are computed as posterior averages over the simulation draws at each time across firms, with the benchmark specification: normal ε_i -normal θ_i . They look very similar to those already shown in Figures 1 and 4 for observed data, confirming that our basic assumptions are clearly able to capture the variation observed in the data. We regard these findings as evidence that no potential failings of our model are present, in the sense that it does not produce any systematic differences between the simulations and the data, and hence that it fits well to data.

Finally, for each value of ν , Table 3 reports also the posterior distribution of $\bar{\rho}$ and the relevant statistics to test for unit root, for the cases *Bayes* and *pool*. It is easy to check that under all tried values of ν , results are essentially the same as those obtained under the normal-normal model as displayed in Table 2, except for the fact that under the assumption $\theta_i \sim t_\nu$, the 68% confidence interval is larger than under the normality assumption, as expected. The *plain heterogeneous* model (Bayes) always provides estimates of ρ far from being equal to 1 a posteriori, while the pool model is not able to reject on average Gibrat's

Table 4. Robust inference and sensitivity analysis

Pharmaceutical industry

a. $y \sim t\text{-student}(\nu), \theta \sim \text{Normal}$

		ν	10	50	100	inf
Bayes	median ρ		0,913	0,912	0,913	0,914
	16% 84%		0,891 0,935	0,891 0,935	0,890 0,935	0,892 0,934
	P.D.		2629,33	3460,83	3736,86	4677,45
	ln(PO)		-6,961	-6,782	-6,320	-6,610
	ω^*		0,006	0,008	0,013	0,010
	p-value		0,997	0,388	0,185	0,120
Pool	median ρ		0,994	0,994	0,994	0,995
	16% 84%		0,991 0,997	0,991 0,998	0,991 0,997	0,992 0,997
	P.D.		1942,25	3327,66	3514,02	4455,59
	ln(PO)		0,840	0,849	0,772	0,393
	ω^*		0,937	0,937	0,945	0,921
	p-value		0,994	0,411	0,180	0,130

b. $\theta \sim t\text{-student}(\nu), y \sim \text{Normal}$

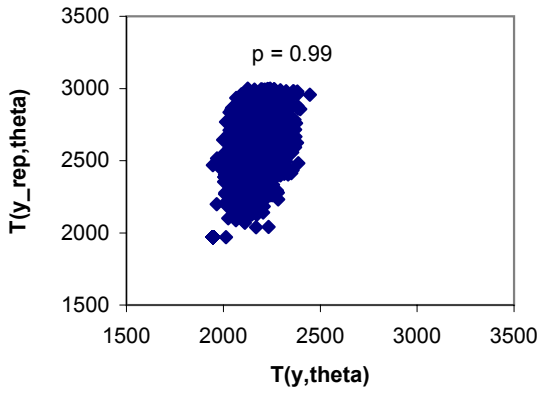
		ν	5	10	15	30
Bayes	median ρ		0,893	0,893	0,891	0,888
	16% 84%		0,834 0,950	0,857 0,931	0,860 0,924	0,835 0,918
	P.D.		2340,19	4459,74	4532,62	3784,17
	ln(PO)		-2,501	-3,917	-5,066	-3,361
	ω^*		0,371	0,136	0,041	0,201
	p-value		0,017	0,020	0,018	0,022
Pool	median ρ		0,980	0,987	0,990	0,987
	16% 84%		0,924 1,083	0,945 1,032	0,957 1,023	0,967 1,021
	P.D.		1906,27	4262,26	4341,99	3529,69
	ln(PO)		0,424	0,825	1,584	0,298
	ω^*		0,905	0,934	0,947	0,841
	p-value		0,018	0,016	0,018	0,019

Figure 7. Sensitivity analysis

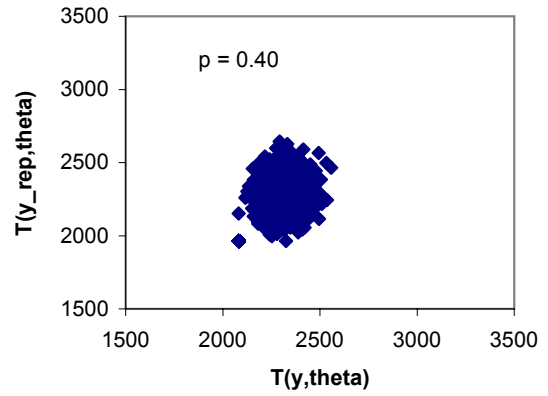
Pharmaceutical industry

$$y \sim t\text{-student}(\nu)$$

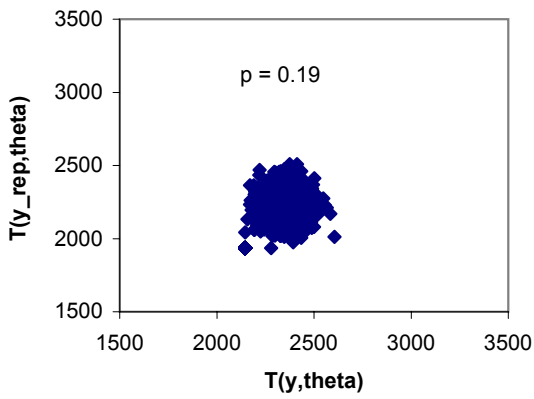
$\nu = 10$



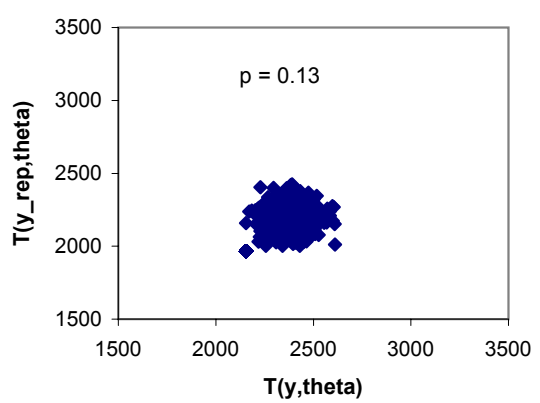
$\nu = 50$



$\nu = 100$

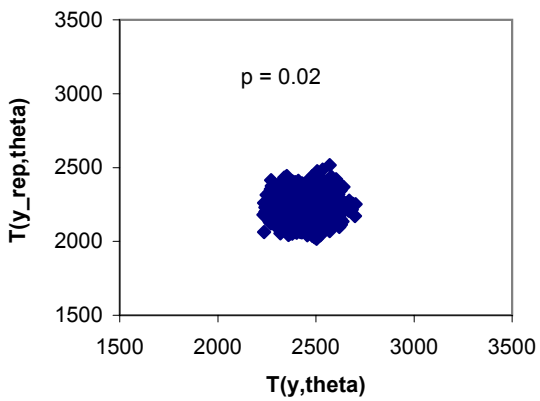


$\nu = \text{inf}$



$$\theta \sim t\text{-student}(\nu)$$

$\nu = 10$



$\nu = 15$

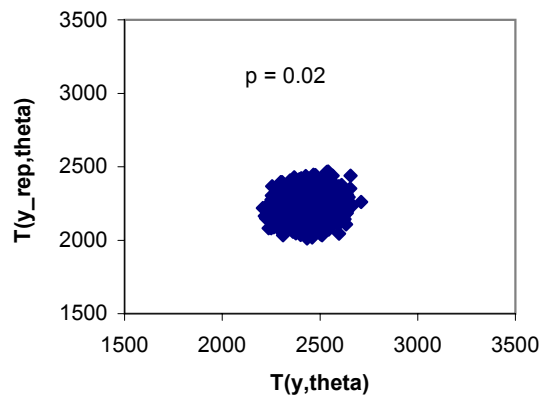
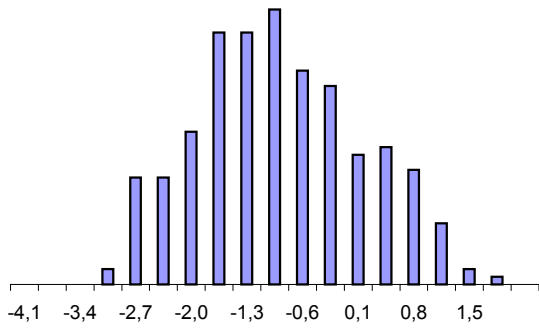


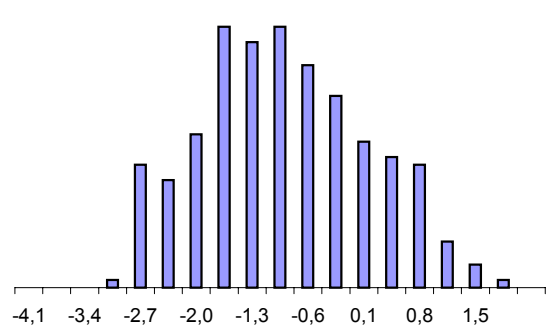
Figure 9. Histograms of replicated data

UK manufacturing

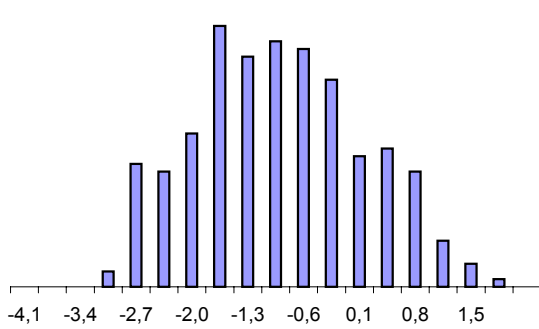
y_rep 89



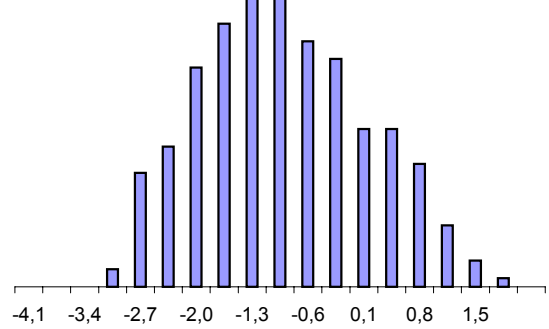
y_rep 90



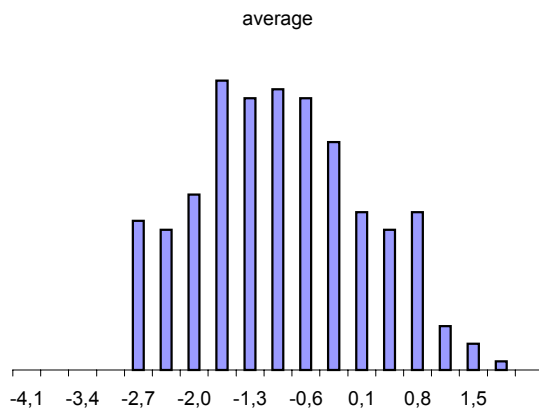
y_rep 91



y_rep 92



average



argument. The resulting inference therefore is unchanged, meaning that it is robust to a model expansion which uses the t_ν as a more general and robust alternative to the normal.

5. SUMMARY AND CONCLUDING REMARKS

Results discussed above can be summarized as follows:

(i) The estimated average speed of adjustment is far from being zero when the information contained both in the cross sectional and in the time series dimension is used. This implies that the main assertion of the Gibrat's law that growth rates are erratic is not true on average, within or across industries;

(ii) When we allow for heterogeneity both in the intercepts and in the slope coefficients, data show a considerable dispersion in the estimated distribution of ρ across firms, whereas when we force the parameters to be the same across units the distribution of ρ is centered around values very close to one. This confirms our initial suspect that previous results, based on cross sectional or pooled panel data models may be econometrically biased because they do not exploit all information contained in the data and hence they misspecify the econometric model without considering heterogeneity, even among firms of the same industry. The null hypothesis $\rho_i = \rho_j$ is a posteriori very unlikely, meaning that the distribution across firms of the autoregressive parameter is far from collapsing to the central value $\bar{\rho}$, as a priori imposed in the cross section or in the pool-panel data models;

(iii) There is only weak evidence of mean reversion. Even if on average $\rho < 1$, this does not necessarily mean that initial larger firms grow relatively slower than smaller firms. Therefore the overall rise in the variance of firm growth turns out to be bounded, but for different reasons than the conventional one linked to the levelling out in growth rates between large and small firms. In any case, as shown in section 3.2, the variance may increase, as time goes by, even if $\rho < 1$;

(iv) Estimated steady states differ across units, and firm sizes do not converge within or across industries to a common limiting distribution. 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;

(v) Initial conditions are important determinants of the estimated distribution of steady states. Differences are likely to reduce at a very slow rate but they do not seem to disappear over time. A firm with an initial size below the average is going to narrow the gap with respect to bigger 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 a robust alternative to the

priors chosen.

Our conclusion is that the simple empirical *fact* on the growth of firms is that growth is not erratic or that firm's size *does not* drift unpredictably over time, as in many previous studies was claimed, and hence that Gibrat's argument does not hold on average. Moreover, the result that small firms tend to grow faster than large ones is also not confirmed. Rather, initially smaller firms tend to remain smaller in steady state and their convergence rate is not higher as compared to larger firms. In other words, there are systematic differences in growth rates among firms. These differences are not size-specific and may depend on other firm-specific features that are not observable in our data. Given that these results are robust to different data sets either within or across industries, and to general models, they open rooms for investigating further the determinants of firms growth.

Most likely, size is not the only correct variable which growth should be conditioned on. Other sources of heterogeneity (age being a primary - but certainly not the only - candidate) may 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 cast serious doubts on the validity and the robustness of all those conventional econometric exercises that do not treat heterogeneity appropriately. How pervasive these problems might be is a fundamental question for future research.

A P P E N D I X

1. The Posterior Distributions

Given the prior information previously specified, we look for the posterior density of the parameter vector $\psi = \left(\theta_i, \bar{\theta}, \Sigma_\theta^{-1}, \{\sigma_i^2\}_{i=1}^N \right)$ which is given by

$$p(\psi | y, y_{i0}) \propto f\left(y | \theta_i, \bar{\theta}, \Sigma_\theta^{-1}, \{\sigma_i^2\}_{i=1}^N, y_{i0}\right) p(\psi | y_{i0})$$

Assuming a vague prior for σ_i^2 , i.e., taking $v_o = 0$, the joint density of all the parameters can be written as

$$\begin{aligned} p\left(\theta_i, \bar{\theta}, \Sigma_\theta^{-1}, \sigma_i^2 | y, y_{i0}\right) &\propto \prod_{i=1}^N \sigma_i^{-T} \exp\left[-\frac{1}{2} \sum_{i=1}^N \sigma_i^{-2} (y_i - X_i \theta_i)' (y_i - X_i \theta_i)\right] \\ &\quad \times |\Sigma_\theta|^{-\frac{N}{2}} \exp\left[-\frac{1}{2} \sum_{i=1}^N (\theta_i - \bar{\theta})' \Sigma_\theta^{-1} (\theta_i - \bar{\theta})\right] \\ &\quad \times |C|^{-\frac{1}{2}} \exp\left[-\frac{1}{2} (\bar{\theta} - \mu)' C^{-1} (\bar{\theta} - \mu)\right] \\ &\quad \times |\Sigma_\theta|^{-\frac{1}{2}(s_o - k - 1)} \exp\left[-\frac{1}{2} \text{tr}\left((s_o S_o) \Sigma_\theta^{-1}\right)\right] \\ &\quad \times \prod_{i=1}^N \sigma_i^{-2} \end{aligned}$$

where $y_i = (g_{i1}, \dots, g_{iT})'$ and $X_i = (X'_{i1}, \dots, X'_{iT})'$. The first line of the formula represents the standard likelihood conditional on the initial conditions and the others represent the prior information.

As said in the text, in order to obtain the marginal posterior distributions of each component of ψ , a numerical integration is needed. We use the Gibbs sampler, a sampling-based approach, firstly introduced by Geman and Geman (1984) and successively popularized by Gelfand and Smith (1990) among others. If we dispose of the full conditional distributions of the parameters, the idea is to construct a Markov chain on a general state space such that the limiting distribution of the chain is the joint posterior of interest. The relevant conditional distributions are obtained from the above formula. For example, the conditional distribution for θ_i is obtained by combining line one with line two, completing the square for θ_i . The conditional distribution of $\bar{\theta}$ is obtained by combining line two with line three, completing the square, and so on. Concretely, the conditional distributions needed to implement the Gibbs sampler are the following:

$$p(\theta_i | y, \psi_{-\theta}) = N\left[A_i \left(\sigma_i^{-2} X'_i y_i + \Sigma_\theta^{-1} \bar{\theta}\right), A_i\right] \quad i = 1, \dots, n$$

$$\begin{aligned}
p(\bar{\theta} | y, \psi_{-\bar{\theta}}) &= N \left[B \left(n \Sigma_{\theta}^{-1} \tilde{\theta} + C^{-1} \mu \right), B \right] \\
p(\Sigma_{\theta}^{-1} | y, \psi_{-\Sigma_{\theta}^{-1}}) &= W \left[\left(\sum_{i=1}^n (\theta_i - \bar{\theta}) (\theta_i - \bar{\theta})' + s_o S_o \right)^{-1}, s_o + n \right] \\
p(\sigma_i^2 | y, \psi_{-\sigma_i^2}) &= IG \left[T/2, \left((y_i - X_i \theta_i)' (y_i - X_i \theta_i) \right) / 2 \right] \quad i = 1, \dots, n
\end{aligned}$$

where $A_i = \left(\sigma_i^{-2} X_i' X_i + \Sigma_{\theta}^{-1} \right)^{-1}$, $B = \left(n \Sigma_{\theta}^{-1} + C^{-1} \right)^{-1}$, $\tilde{\theta} = (1/n) \sum_i \theta_i$, $W()$ denotes the wishart, $IG()$ the inverse gamma distribution and $\psi_{-\gamma}$ denoting ψ without γ .

After iterating, say, M times, the sample value $\psi^{(M)}$ can be regarded as a drawing from the true joint posterior density. Once this simulated sample has been obtained, any posterior moment of interest or any marginal density can be estimated using the ergodic theorem. Convergence to the desired distribution can be checked as suggested by Gelfand and Smith (1990).

In our exercise the number of Monte Carlo iterations is set equal to 5000, and the first 1000 are discarded. Convergence is achieved already with the first 3000 iterations. Moreover, the third stage of the hierarchy is assumed vague, or non-informative, i.e., we set $C^{-1} = 0$. This means that the only hyperparameters to be assumed known are s_o and S_o , i.e., the degrees of freedom and the scale matrix of the wishart prior for Σ_{θ} . These hyperparameters control the four settings under which we estimate the model. Concretely, the *pool ols* ($\Sigma_{\theta} = 0$) is approximated by choosing $S_o = \text{diag}(0.0001 \ 0.0001)$; for the *fixed effect* ($\sigma_{\rho} = 0$) we choose $S_o = \text{diag}(0.1 \ 0.0001)$. The setting *mean group* ($\Sigma_{\theta} \rightarrow \infty$) is approximated with $S_o = \text{diag}(100 \ 100)$. Finally for the *Bayes* setting we choose $S_o = \text{diag}(1. \ 1.)$. In the four cases prior degrees of freedom are chosen randomizing uniformly over the interval $(3, 10)$, i.e. $s_o \sim \text{uniform}(3, 10)$.

The conditional posterior distributions need to be modified when a t_{ν} distribution is used in place of the normal either for the error term or for the population structure. Recalling that the t_{ν} can be interpreted as a mixture of normal distributions with variances distributed as scaled inverse- χ^2 , the Gibbs sampler is easily extended to include one or more conditional distributions. For instance, in the case of $y_i \sim N(X_i \theta_i, \sigma_i^2 I_T)$ and $\theta_i \sim t_{\nu}(\bar{\theta}, \Sigma_{\theta})$, the latter is equivalent to

$$\begin{aligned}
\theta_i \mid v_i &\sim N(\bar{\theta}, v_i \Sigma_{\theta}) \\
v_i &\sim \text{Inv-}\chi^2(\nu, 1)
\end{aligned}$$

where $\text{Inv-}\chi^2(\nu, s^2)$ denotes a scaled inverse- χ^2 with scale factor s^2 .

Therefore, the change in the distributive assumption on the population structure is equivalent to an expansion of the model. This, in turn, implies that the posterior distributions of the parameters must change accordingly and also that a new variable, v_i , must enter the Gibbs Sampler. A similar hierarchy holds for y_i when $y_i \sim t_{\nu}(X_i \theta_i, \sigma_i^2 I_T)$.

The sensitivity analysis is performed by comparing different sets of results corresponding to different values for ν . To compute the test-quantities $T(y, \theta)$ and tail-area probabilities we proceed as follows. If we already have, say, L simulations from the posterior density of ψ , we draw one y^{rep} from the predictive distribution for each simulated ψ . We now have L draws from the joint posterior distribution $p(y^{rep}, \psi | y)$. The posterior predictive check is the comparison between the realized test quantities, $T(y, \psi^l)$, and the predictive test quantities $T(y^{rep,l}, \psi^l)$. The estimated p-value is just the proportion of these L simulations for which the test quantity equals or exceeds its realized value; that is, for which $T(y^{rep,l}, \psi^l) \geq T(y, \psi^l)$, $l = 1, \dots, L$. For further technical details on model checking and sensitivity analysis, see also Gelman et al. (1995, Ch. 6 and 12).

2. Testing

The statistics used to test the hypotheses discussed in the text is the logarithm of the posterior odds (PO) ratio suggested by Leamer (1978) and applied by Sims (1988). The statistics can be written as

$$\ln(PO) = \ln \left[\frac{(1 - \omega)}{\omega} \frac{\phi(q)}{\Phi(q)} |V|^{-\frac{1}{2}} \right] \quad (2.1)$$

where the c.d.f of the standard Normal distribution, $\Phi()$, and its p.d.f., $\phi()$, are evaluated at $q = \sqrt{(R\gamma - r)' V^{-1} (R\gamma - r)}$, V is the matrix RQR' , Q is the posterior variance covariance matrix of γ , R is a restriction matrix, r is the value to which we are restricting the vector γ , and ω is the prior probability we assign to the alternative, i.e., the probability we initially put uniformly on the interval $(0, 1)$, while $1 - \omega$ is the probability we put on the null $R\gamma = r$.

For instance, in the concrete case of testing for unit root on average, q becomes equal to the conventional t-statistics for $\rho = 1$, $(1 - \rho^*) / \sigma_\rho^*$, and $|V|^{-\frac{1}{2}} = 1 / \sigma_\rho^*$, where variables with a (*) denote estimated posterior means. On the other hand, in the case of testing that steady states are the same across units, the standard normal c.d.f. and p.d.f. are evaluated at $q = \sqrt{(RS)' (RQR')^{-1} (RS)}$ where S is the $n \times 1$ vector containing posterior estimates of the (linearized) steady states of each unit, R is the $(n - 1) \times n$ restriction matrix with ones on the main diagonal, -1 on the following upper diagonal and zero elsewhere, and Q is the variance covariance matrix of the posterior estimates of the (linearized) steady states.

Notice that by selecting $\omega < 1$ we are implicitly placing higher weight on the null hypothesis, since ω is spread over infinitely many possible alternative values. For this reason, in testing for unit root, Sims (1988, p. 471) suggests to choose a value in the interval $(0.5, 1)$, proposing as a reasonable value 0.8. In this paper, to verify the null $\rho = 1$, we randomize uniformly over this interval, while we assume $\omega = 0.5$ to test for equality of steady states and for $\rho_i = \rho_j$.

As said in the text, we can also compute the largest prior probability ω^* to attach to the alternative for the test to accept the null, given the data. Such a prior probability can be computed from (2.1) and it is equal to $\omega^* = 1/(1 + \exp(w))$ where

$$w = \ln(\Phi) + 0.5 \ln(|V|) - \ln(\phi)$$

For more details on this testing procedure, see Sims (1988) and Canova and Marcet (1998).

References

- [1] Baily M.N. and A.K. Chakrabarty (1985), ‘Innovation and Productivity an US Industry’, *Brookings Papers on Economic Activity*, 2, pp.609-632.
- [2] Bottazzi G., G. Dosi, M. Lippi, F. Pammolli, M. Riccaboni, (2000), ‘Process of Corporate Growth in the Evolution of an Innovation- driven industry. The case of pharmaceuticals’, LEM Working paper n.2000/5, Scuola Superiore Sant’Anna, Pisa.
- [3] Canova F. and A. Marcet. (1995), ‘The Poor Stay Poor: Non-Convergence across Countries and Regions’, CEPR, Discussion paper n.1265.
- [4] 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.
- [5] Chib S. and E. Greenberg (1995), ‘Hierarchical Analysis of SUR Models with Extensions to Correlated Serial Errors and Time-Varying Parameter Models’, *Journal of Econometrics*, 68, 409–431.
- [6] Chib, S. and E. Greenberg (1996), ‘Markov chain Monte Carlo simulation methods in econometrics’, *Econometric Theory*, 12, 409–431.
- [7] Dosi G., O. Marsili, L. Orsenigo and R. Salvatore (1995), ‘Technological Regimes, Selection and Market Structure’, *Small Business Economics*, 7, pp. 1-26.
- [8] 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.
- [9] Ericson R. and A. Pakes (1995), ‘Markov-Perfect Industry Dynamics: a Framework for Empirical Work’, *Review of Economic Studies*, 62(1), pp.53-82.
- [10] 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
- [11] Evans D.S. (1987b), ‘Tests of Alternative Theories of Firm Growth’, *Journal of Political Economy*, 95(4), pp. 657-74.
- [12] Farinas J.C. and L. Moreno (2000), ‘Firms’ Growth, Size and Age: A Non-parametric Approach’, *Review of Industrial Organization*, 17(3), pp. 249-65

- [13] 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.
- [14] 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.
- [15] Hart P.E. and N. Oulton (1999), ‘Gibrat, Galton and Job Generation’, *International Journal of the Economics of Business*, 106(3), pp.1242-1252.
- [16] 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.
- [17] Ijiri Y. and H. Simon (1977), *Skew Distribution and the Sizes of Business Firms*, Amsterdam, North-Holland Publishing.
- [18] 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.
- [19] 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.
- [20] Gelman A., J.B. Carlin, H.S. Stern and D.B. Rubin (1995), *Bayesian Data Analysis*, Chapman & Hall, London.
- [21] Geman S. and D. Geman (1984), ‘Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images’, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6):721-741.
- [22] Geroski P.A. (1998), ‘The growth of Firms in Theory and in Practice’, CEPR working paper 2092, London.
- [23] Geroski, P., S. Machin and Van Reenen, J. (1993), ‘The Profitability of Innovating Firms’, *RAND Journal of Economics*, 24(2), pp. 198-211.
- [24] Geroski P.A., S. Lazarova, G. Urga and C.F. Walters (2001), ‘Are Differences in Firm Size Transitory or Permanent?’, mimeo, London Business School.
- [25] Gibrat, R. (1931), *Les inegalites economiques*. Librairie du Recueil Sirey, Paris.

- [26] 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.
- [27] 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.
- [28] Jovanovic B. (1982), 'Selection and the Evolution of Industry', *Econometrica*, 50(3), pp.649-70.
- [29] Klette T.J., M. Jarle and Z. Griliches (2000), 'Do Subsidies to Commercial R&D Reduce Market Failures? Microeconomic Evaluation Studies', *Research Policy*, 29(4-5), pp. 471-95
- [30] Leamer E.E. (1979), *Specification Searches. Ad Hoc Inference with Non-Experimental data*, New York: Wiley.
- [31] 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.
- [32] Lotti F., E. Santarelli, M. Vivarelli (2000), 'Does Gibrat's Law Hold among Young, Small Firms?', LEM Working paper n.2000/5, Scuola Superiore S'Anna, Pisa.
- [33] Mata J. (1994), 'Firm Growth during Infancy', *Small Business Economics*, 6(1), pp. 27-39
- [34] McCloughan P.(1995), 'Simulation of Concentration Development from Modified Gibrat Growth-Entry-Exit Processes', *Journal of Industrial Economics*, 43(4), pp.405-433.
- [35] Mueller D.C. (ed.) (1990), *The Dynamics of Company Profits: an International Comparison*, Cambridge, MA, Cambridge University Press.
- [36] Pakes A. and R. Ericson (1998), 'Empirical Implications of Alternative Models of Firms Dynamics', *Journal of Economic Theory*, 79(1), pp. 1-45.
- [37] Pesaran M.H. and R. Smith (1996), 'Estimating Long Run Relationships from Dynamic Heterogeneous Panels', *Journal of Econometrics*, 68, pp.79-113.
- [38] Simon H. and C. Bonini (1958), 'The Size Distribution of Business Firms', *American Economic Review*, 58(4), pp.607-617.
- [39] Sims C. (1988), 'Bayesian Skepticism on Unit Root Econometrics', *Journal of Economic Dynamics and Control*, 12, 463-474.

- [40] Sutton J. (1997), 'Gibrat's Legacy', *Journal of Economic Literature*, 35, pp.40-59.
- [41] Sutton J. (2000), 'The Variance of Firm Growth Rates: the 'Scaling' Puzzle', mimeo for the Schumpeterian Society, Manchester.
- [42] 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