The growth of firms: from Gibrat's legacy to Gibrat's fallacy*

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

[a] University of Bergamo, Italy
[b] University of Alicante and Bank of Spain, Spain
[c] University of Brescia and CESPRI, Bocconi University, Milan, Italy

EPRIS working paper
August 2001

Abstract

In this paper we investigate some properties of the patterns of firms' growth in the pharmaceutical industry. Several recent studies are based on some version of the so-called Gibrat's Law, which assumes that firms' growth basically follows a random walk. We simply aimed at testing Gibrat's Law, using somewhat different statistical techniques than the usual ones (a Bayesian approach that takes into consideration heterogeneity at the firm level), as a first step towards a more systematic investigation of the patterns of firms' growth. The results suggest that: i) there seems to be strong evidence against the Gibrat's law (on average); ii) previous results are probably incorrect because they are based on models that do not control for potential heterogeneity in the slope coefficients; iii) differences in growth rates do not seems to disappear over time.

*The financial supports of the European Union (ESSY Project, TSER, contract n. SOE1-CT 98-1116 DGXII-SOLS), of EPRIS project (European Pharmaceutical Regulation and Innovation System), and the Italian Ministry of University and Research (grant n. 9913443984) are gratefully acknowledged. Corresponding author: Matteo Ciccarelli, matteo@merlin.fae.ua.es

1 Introduction

In recent years, the study of the processes of corporate growth has become again - after several years of neglect - a main focus of the attention of industrial economists. The large majority of these recent studies is based on some version of the so-called Gibrat's Law, which assumes that firms' growth follows a random walk. Despite some conflicting evidence (that shows important violations of the "Law of Proportionate Effects"), it would seem that the hypothesis that firms' growth rates are essentially erratic is now largely taken for granted and embodied without much further discussion in more general models of industrial dynamics and evolution (Geroski, 1998, Jovanovic, 1982). Even if Gibrat's Law is not embodied in theoretical models as such, it is considered at least a benchmark. More generally, Gibrat's Law enters in the models as a way of thinking to firm's growth, in the sense that growth is considered and modeled as erratic.

A large empirical literature has explored this issue in different data sets and with different statistical methodologies. Roughly speaking, the results support the views that: (i) firm's growth rates follow a random walk, (ii) they do not converge within or across industries, and (iii) no stable or predictable differences in growth exist either in the short or in the long run.[1]

Typically, this literature attempts to verify Gibrat's law and its consequences using a cross-section regression or short-panel econometric techniques with homogeneity in the parameters across units and over time. The first approach is problematic in that it ignores the information contained in unit-specific time variation in growth rates. The second approach (homogeneous panel data) is problematic because, even if it considers information available for all periods and all cross sectional units, it forces the parameters to be the same across individuals, thus pooling possibly heterogeneous units as if their data were generated by the same process.[2] 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. 1998).

In this paper we argue that more care should be put in the consideration of the Gibrat's law as a *stylized fact* or even only as a benchmark of the corporate growth dynamics. The point of this paper is that results previously obtained could be econometrically biased because even though they convey attention on important aspects of the data, they are based on methodologies which force units to be homogeneous, i.e., methodologies that could be inappropriate for studying the Gibrat' law and its implications, as we try to discuss in this work. We believe that it is not by chance that departures from Gibrat's law are encountered in studies where the analysis is either based on groups of firms homogeneously chosen (see. Lotti et al., 2000, Bottazzi et al.,2000), or constructed by modelling somehow the heterogeneity (Gerosky et al., 2001).

Using two different data sets, we find that: (i) the main assertion of the Gibrat's law that firm's growth rates drift unpredictably 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;

---

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

[2] In other words, there would not be much difference between the cross section and the panel data approaches if in the latter we forced the parameters to be exactly equal in all units, a case that should be formally tested instead.

(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; (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. In other words they persist.

The paper is structured as follows. Section 2 discusses the statistical model.[3] Section 3 describes data and comments on the estimation results. Section 4 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, $t = 0, 1, ..$ 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

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

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

$$\ln S_{it} \simeq \ln S_{i0} + \sum_{j=0}^{t-1} \varepsilon_{it}$$

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

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

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

$$\begin{aligned} E (u_{it} \mid S_{it-s}, \quad s > 0) &= 0 \\ E \left(u_{it}^2 \mid S_{it-s}, \quad s > 0\right) &= \sigma^2 \end{aligned}$$

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

[3] This section is based on the same line of reasoning as in Canova and Marcet (1998), who studied the issue of convergence of per-capita income across economic areas, a similar topic applied to a different context.

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 the same specifications, but instead of considering the log of S_{it}, we believe that it is more convenient to study the behavior of the (log of) each unit's size relative to the average, i.e., to study 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

$$g_{it} = \alpha_i + \rho_i g_{it-1} + \eta_{it} \tag{2}$$

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

Notice that our specification is more general than either a simple cross sectional analysis or a homogeneous dynamic panel data model. On the one hand, Eq. (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 Gibrait'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. 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., 1997, among others). 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 Gibrait'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), which is flexible enough to formally test the restrictions that previous studies impose on the data, i.e., $\alpha_i = \alpha$, and $\rho_i = \rho$, $\forall i$.[4]

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 more implications of the law. Precisely, we are able

[4] In fact, other studies (see 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 u_{it} is serially correlated.

to estimate the speed of adjustment $(1-\rho_i)$ of each unit to its own steady state $(\alpha_i/(1-\rho_i))$, a question 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 be used easily 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.[5]

Let $\theta_i = (\alpha_i, \rho_i)'$. Eq (2) can then be written in a more compact form as

$$g_{it} = X_{it}'\theta_i + \varepsilon_{it} \tag{3}$$

where $X_{it} = (1, g_{it-1})'$. The prior distribution we use assumes that the intercept and the slope of the model do not differ too much across units. Concretely, we assume

$$\theta_i \sim N\left(\bar{\theta}, \Sigma_\theta\right) \tag{4}$$

where $\bar{\theta}$ and Σ_θ are common to all individuals. In other words we are assuming 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 this restrictions might be formally tested instead of simply imposed.

Notice also that (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

$$\bar{\theta} \sim N(\mu, C) \tag{5}$$

$$\Sigma_\theta^{-1} \sim W\left(s_o, S_o^{-1}\right) \tag{6}$$

$$\sigma_i^2 \sim IG\left(\frac{v^2}{2}, \frac{v^2\delta^2}{2}\right) \tag{7}$$

where the notation $Z \sim W(q_o, Q_o)$ means that the matrix Z is distributed as a Wishart with scale Q_o and degrees of freedom q_o, and $\omega \sim IG(\zeta/2, \pi/2)$ denotes an inverse gamma distribution with shape ζ and scale π. The hyperparameters μ, C, s_o, S_o, v and δ are assumed known. Independence is also assumed throughout.

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

The entire specification (3) through (7) is standard in the Bayesian literature and has the advantage of being sufficiently flexible to answer the kind of questions posed in this work.[6] In particular, notice that setting $\Sigma_\theta = 0$, is equivalent to impose equality of coefficients across firms, i.e., 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. On the contrary, if $\Sigma_\theta \to \infty$, the prior information is diffuse, i.e., the degree of uncertainty about the mean is infinite, so that estimated parameters of different firms are similar to those obtained applying OLS to (3) equation by equation. In other words, when $\Sigma_\theta \to \infty$ only the time series properties of each g_{it} are used. 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.

Posterior inference can 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 clear divergence is present $(\rho_i \geq 1)$. 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:

$$\widehat{SS}_i = c + b g_{i0} + \omega_i \tag{8}$$

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 are these differences.

2.2 Estimation and testing

The posterior distributions of the parameters of interest are obtained, as already remarked, 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 D represent the data, the Bayes rule

$$p(\psi \mid D) \propto p(\psi)\ l(D \mid \psi)$$

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.

[6]For further details on the specification and the estimation see Hsiao et al. (1998) and Ciccarelli (2001), among others.

The null hypotheses $\bar{\rho} = 1$, $\rho_i = 1$, and $S_i = S_j$, $\forall i,\ j$, with $S_i = \alpha_i/(1-\rho_i)$ are verified by calculating the 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 π^*, 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.

The details of estimation and testing techniques are in appendix.

3 Empirical results

In this section we briefly describe the data sets used in the analysis and then present the empirical results. These are shown in Figures 1-5 and Tables 1-4.

3.1 The Data

The issue analyzed in this paper seems particularly relevant in the specific case of the pharmaceutical industry, which can be considered an ideal case where the process of growth behaves less dissimilar from the erratic model of Gibrat's Law due to the pattern of innovation peculiar of that industry. 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). 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. Therefore, sales for each firm stand for the sum of their sales in each of the national market. Furthermore, in order to control for merger and acquisition processes during the period, we constructed "virtual-firms". They 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. Therefore, if two firms, for example, merged during the period, we consider them merged from the start, summing up their sales from the beginning.The disaggregation goes up to the 4-digit-level of the Anatomical Therapeutic Classification. 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 period of observation. As we need a balanced panel to apply the Bayesian methodology, we censor the entrants among the top ranks, using 199 observations out of 210.

3.2 The speed of convergence

The first set of results which is worthwhile commenting is contained in table 1, where we report for three different settings of Σ the average estimate of ρ together with the first and the third quartile of its posterior distribution, and in Figure 1.B, which shows the entire posterior distributions under the three assumptions on Σ. Table 2 contains the test of unit root both for each single firm (2.a) and on average across firms (2.b)

Three important facts can be discussed.

First, by forcing the units to have the same coefficients ($\Sigma = 0$), we obtain the results generally obtained in the literature. Therefore, under this set of restrictions the model is able to reproduce the standard cross-sectional and pooling-panel regression results. In this case we cannot reject the null hypothesis that the average ρ is equal to 1 as shown in Table 2.b. In fact, when $\Sigma = 0$, the log of the PO ratio is positive, meaning that the posterior odds favor the null, and the largest prior probability we should assign on the alternative in order for the data not to reject the null is almost 1 ($\pi^* = 0.97$). The latter 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. Testing for unit roots firm by firm, under the same set of restrictions $\Sigma = 0$, provides the same results. Concretely, for 80 percent of the firms in the sample we cannot reject the null of $\rho = 1$ (Table 2.a). The same information is contained in Figure 1.A and 1.B, where the histogram of the posterior mean of the parameter ρ_i for each firm and the average posterior mean $\bar{\rho}$ are respectively plotted. From both figures it is clear that the limiting distribution of the autoregressive parameter is not very dispersed around a mean equal to one.

Second, when we allow for heterogeneous parameters across units ($0 < \Sigma < \infty$), the average ρ is equal to 0,88 (Table 1). The log(PO) is highly negative and favors the alternative versus the null of $\rho = 1$, while π^* is close to zero, meaning that a prior probability of almost 1 should be attached on the null in order for the data not to reject it (Table 2.b). 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. Testing for unit root by firm (Table 2.a) confirms that only for 22 percent of firms we cannot reject the null. A similar information is contained in Figures 1.A and 1.B where, as for the case $\Sigma = 0$, the histogram 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 rends the a posteriori probability of facing a random walk very unlikely on average. At the same time, the information contained in Fig. 1.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.[7] For the sake of completeness we considered also the situation where only the information in the time series dimension is used, i.e., the case $\Sigma \to \infty$. Given that we have just 12 time observations, a small sample downward bias in the estimation of the average ρ appears, as we would expect. In this case the distribution of average ρ is more shifted to the left with respect to the previous cases (Figure 1.B), and the null $\rho = 1$ is clearly rejected both on average and by firms (Table 2.b and 2.a respectively). The histogram (not plotted) reveals a dispersion very similar to the one obtained with the setting $0 < \Sigma < \infty$.

Third, note that, the simple observation of $\rho < 1$ is not sufficient to indicate the existence of

[7] Remember the histogram is based on the mean of the estimated posterior distribution of ρ_i firm by firm. The information contained in the histogram is therefore different form 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 where no heterogeneity in the data. We have also formally tested the null $\rho_i = \rho_j$. The $\log(PO)$ was equal to -32,74, and $\pi^* = 0$. These values are strong evidence in favor of the alternative.

"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 1.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 as been argued in recent works.[8] 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

$$g_{it} = \lambda g_{i0} + \rho_i g_{it-1} + \eta_{it} \tag{9}$$

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

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

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

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

as t becomes sufficiently large. Therefore it may very well be the case that $Var\left(g_{i0}\right) < Var\left(g_{it}\right)$, 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.

The above discussion confirms 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. The 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.

[8] See Lotti et al., 2000 for instance.

3.3 The steady state

Focusing on the case $0 < \Sigma < \infty$, which in our opinion is the most reasonable one, as already remarked, the dispersion of steady states is substantial. Figure 2.A provides a histogram of estimated posterior steady states for each firm.[9] The histogram is constructed so that firms are grouped in 9 classes of steady state size: up to 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 2.c reports the statistics for the hypothesis that the steady states are the same. Both the PO ratio and the π^* are zero, meaning that the null is highly unlikely, or that unless we assume that the alternative is impossible ($\pi^* = 0$), the null hypothesis is always overturned by the data.

The results indicate that the estimated distribution of steady states is non-degenerate, i.e., that 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 this paper, but we can at least propose a natural candidate to explain the limiting distribution. Figure 2.B plots the estimated steady states against the initial (scaled) size levels. It is clear that there is a strong positive relationship, i.e., initial big firm are those with the highest steady states and the initial ranking is largely maintained. Figure 2.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 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 0.71, one would argue that inequalities are *strongly* persistent. The $\bar{R}^2$ can be interpreted as a measure of long run mobility.[10] 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 0.71 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 29%, 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.

3.4 Robustness

In this section we briefly discuss the results obtained by using a different data set. The analysis is preformed on a sample of 267 UK manufacturing firms. Data constitute a balanced panel of five years, from 1988 to 1992.

The aim of this section is twofold. On the one hand we can verify if our results are robust to different data sets and on the other hand, given that in this new data set we have different industries,

---

[9] Whenever $\rho > 1$, 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.

[10] See Canova and Marcet, 1998, for instance.

and hence another level of possible heterogeneity, some common features across industries can also be captured to better explain firms growth.

The model specification is the same as before, and so are the prior assumptions. The estimation results, shown in Tables 3-4 and Figures 3-5, 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 4.b) and by firms (table 4.a). The average autoregressive parameter ρ is equal to 0,72 (Tab. 3 and Fig.3.B) in the setting $\Sigma > 0$, while it is 0,995 when $\Sigma = 0$. The dispersion of the mean estimates of ρ_i by firms (Fig. 3.A) is again a way of rejecting the null hypothesis that $\rho_i = \rho_j$. There seems to be even less relation between the convergence rate and the initial conditions, with respect to the previous data set (Fig. 3.C), confirming the view that no mean reversion is present in the sample data. 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 all firms we do not reject the null that Gibrat's law holds (Tab. 4.a, first column). On the contrary, when we do not impose this restriction only 12,4 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. Only in two industrial groups all firms have a ρ significantly less than 1, while in the others the percentage of firms whose size drifts unpredictably is almost uniformly distributed (Fig. 5).

Conclusions on the limiting behavior of firm size can be appreciated from Fig. 4. As before, we reject the null hypothesis of equal steady states. The dispersion of estimated steady states is again substantial (Fig. 4.A) and, unless we assume a priori that the alternative is impossible, the null hypothesis will always be overturned by the data (in Tab. 4.c $\pi^* = 0$). Moreover differences in firms size are extremely persistent. The evidence contained in Fig. 4.B and 4.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 17%, while the initial conditions alone explain 52% of the variation of the cross sectional distribution of estimated steady states.

4 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 follow a random walk process 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 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 not because of 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 compatibly with $\rho < 1$;

(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 the bigger firms, but she does not seem to increase its relative size in the cross sectional distribution. In other words, differences in firm size persist.

Our conclusion is that the simple empirical *fact* on the growth of firms is that growth is not erratic. In other words, growth rates *do not* drift unpredictably over time, as in many previous studies was claimed, and hence that Gibrat's argument does not hold on average. Given that these results are robust to different data sets either within or across industries, they open rooms to investigate further the determinants of firms growth. 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.

References

[1] Bottazzi, G., Dosi, G., Lippi, M., Pammolli, F., Riccaboni, M. (2000), Process of Corporate Growth in the Evolution of an Innovation- driven industry. The case o f pharmaceuticals, LEM Working paper n.2000/5, Scuola Superiore S'Anna, Pisa.

[2] Canova F., Marcet A. (1995), *The Poor Stay Poor: Non-Convergence across Countries and Regions*, CEPR, Discussion paper n.1265.

[3] Ciccarelli M. (2001), *Bayesian Inference in Heterogeneous Dynamic Panel Data Models*, Ph.D. Thesis, Universitat Pompeu Fabra, Barcelona.

[4] 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):972–985.

[5] 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.

[6] Geman S. and D. Geman (1984), Stochastic relaxation, Gibbs ditributions and the Bayesian restoration of images, *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 6(6):721–741.

[7] Geroski, P.A. (1998) The growth of Firms in Theory and in Practice, CEPR working paper 2092, London.

[8] Geroski, P.A., Lazarova S., Urga G., and Walters C.F. (2001), *Are Differences in Firm Size Transitory or Permanent?*, mimeo, London Business School.

[9] Hsiao C., Pesaran M.H., Tahmisciouglu A.K. (1997), "Bayes Estimation of Short Run Coefficients in Dynamic Panel Data Models", *mimeo*, Cambridge University.

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

[11] Leamer E.E. (1979), *Specification Searches. Ad Hoc Inference with Non-Experimental data*, New York: Wiley.

[12] Lotti F., Santarelli E., Vivarelli M. (2000), "Does Gibrat's Law Hold among Young, Small Firms?", LEM Working paper n.2000/5, Scuola Superiore S'Anna, Pisa.

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

[14] Sims C. (1998), Using a likelihood perspective to sharpen the econometric discourse: three examples, *mimeo.*

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

[16] Sutton, J. (2000), "The Variance of Firm Growth Rates: the "Scaling" Puzzle", mimeo for the Schumpeterian Society, Manchester.

A The posterior distribution of our parameters

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

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

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

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

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.

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. 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 three settings under which we estimate the model. Concretely, the setting $\Sigma_\theta = 0$ is approximated by choosing $S_o = diag\,(0.1 \quad 0.02)$ and $s_o \sim uniform\,(3,5)$. For $\Sigma_\theta > 0$ we choose $S_o = diag\,(10 \quad 1.0)$ and $s_o \sim uniform\,(3,50)$. Finally for $\Sigma_\theta \to \infty$ we set $S_o = diag\,(100 \quad 100)$ and $s_o \sim uniform\,(10,100)$.

B 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) to test for unit root. The statistics can be written as

$$PO = 2 \cdot \ln \left[\frac{(1-\pi)}{\pi} \frac{\phi(\tau)}{\Phi(\tau)} \frac{1}{\sigma_\rho} \right] \tag{10}$$

where $\Phi()$ is the c.d.f for the standard Normal distribution, $\phi()$ its p.d.f., and $\tau = (1-\rho)/\sigma_\rho$ stands for the conventional t-statistics for $\rho = 1$ where ρ and σ_ρ are the estimated posterior means of the respective distributions. π is the prior probability we assign to the alternative, i.e., the probability we initially put uniformly on the interval $(0,1)$, while $1-\pi$ is the probability we put on $\rho = 1$. Notice that by selecting $\pi < 1$ we are implicitly placing higher weight on the null hypothesis, since π 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) with annual data, proposing as a reasonable value 0.8. In this paper we randomize uniformly over this interval.

In order to test that steady states are the same across units, the (10) is modified as

$$PO = 2 \cdot \ln \left[\frac{(1-\pi)}{\pi} \frac{\phi(q)}{\Phi(q)} |\Sigma_\theta|^{-\frac{1}{2}} \right] \tag{11}$$

where 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 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 steady states. In this case we assume $\pi = 0.5$.

As said in the text we can also compute the largest prior probability π^* to attach to the alternative for the test to accept the null, given the data. Such a prior probability can be computed from (10) and(11) and it is equal to $\pi^* = 1/(1+\exp(w))$ where

$$w = \ln \Phi + \ln \sigma - \ln \phi$$

and

$$\ln|Q| + 2\ln(\Phi) + (n-1)\ln(2 \cdot 3, 14) + q^2$$

respectively.

For more details see Leamer (1978), Sims (1988) and Canova and Marcet (1998).

Figure 1. Convergence rates

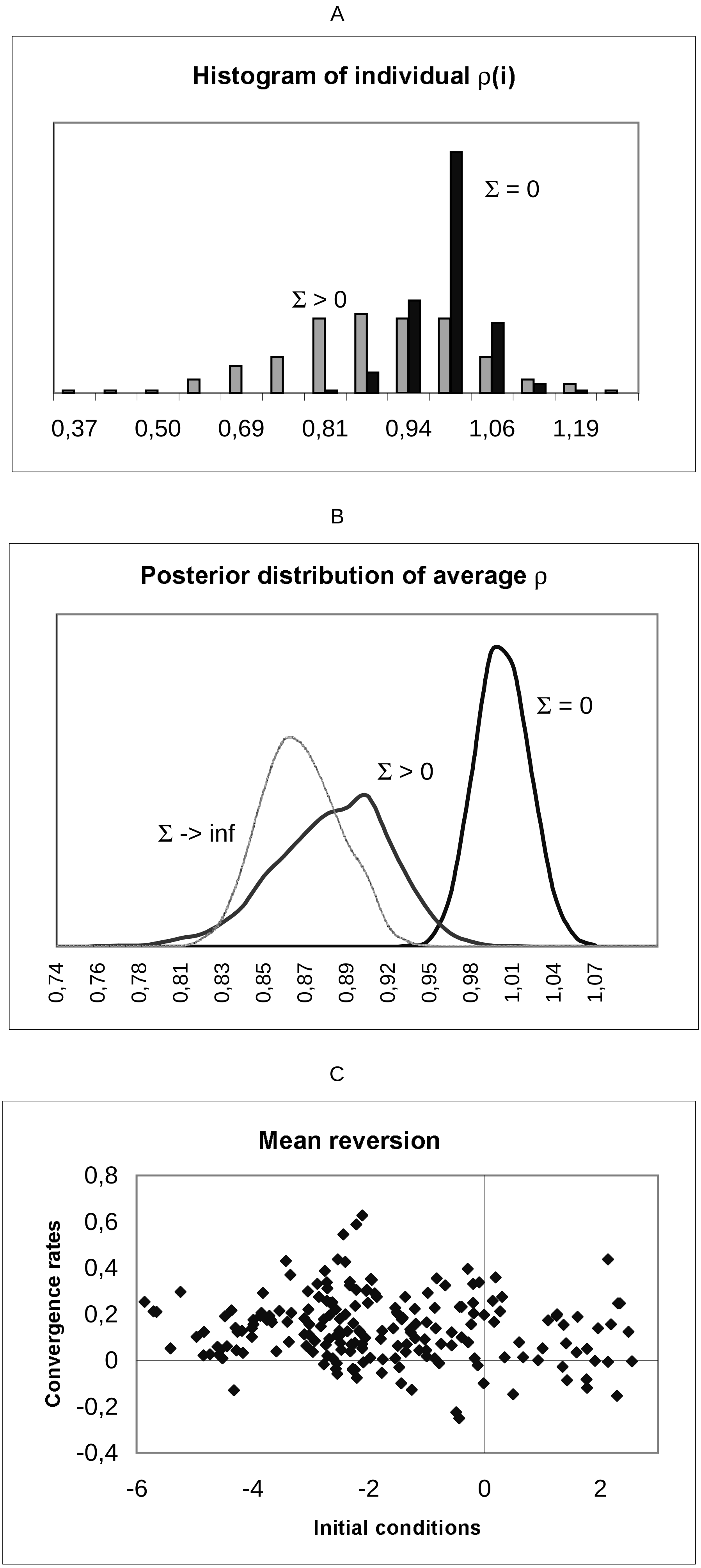
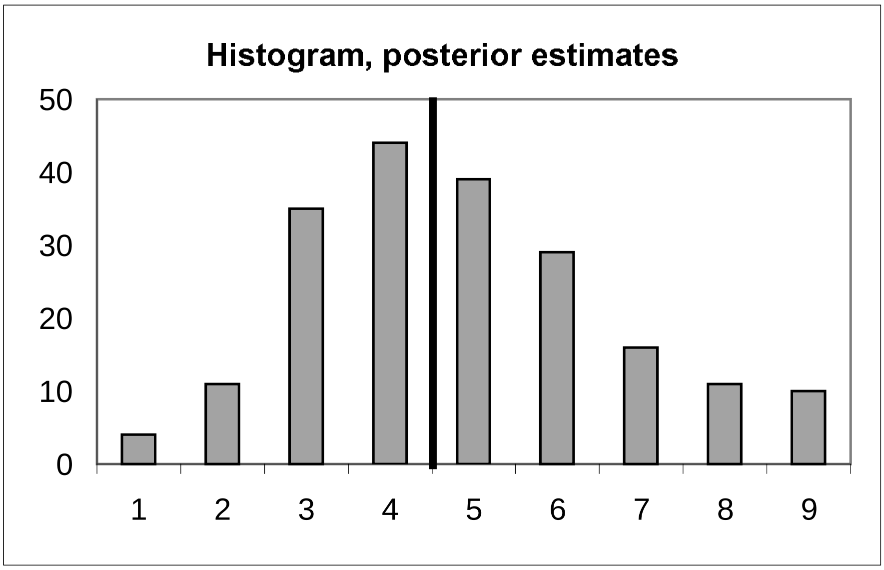
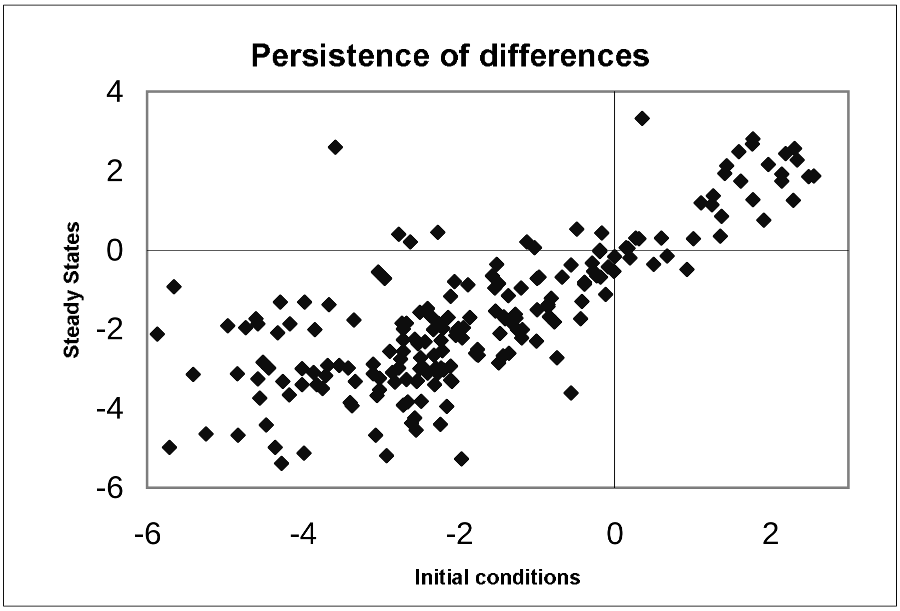


Figure 2. The steady state

A



B



C

test of persistence in differences of firm size		
constant	**initial condition**	**R_bar**2**
-0,37	0,71	0,57
(0.12)	(0.044)	

Table 1. Average estimates of ρ

	1st quart	mean	median	3d quart
$\Sigma = 0$	0,982	0,994	0,994	1,030
Σ -> inf	0,848	0,863	0,862	0,877
$0 < \Sigma <$ inf	0,860	0,883	0,882	0,924

Table 2. Testing

2.a unit root: by firm

	$\Sigma = 0$		**Σ -> inf**		**$0 < \Sigma <$ inf**	
	$\rho(i) = 1$	$\rho(i)<1$	$\rho(i) = 1$	$\rho(i)<1$	$\rho(i) = 1$	$\rho(i)<1$
%	80,00	20,00	20,10	79,90	21,60	78,40
ln(PO)	0,80	-1,26	0,69	-1,97	0,71	-1,95
π*	0,89	0,59	0,91	0,50	0,87	0,53

2.b unit root: across firms (ρ=1)

ln(PO)	2,09	-18,0	-3,6
π*	0,97	0,00	0,03

2.c equal steady states

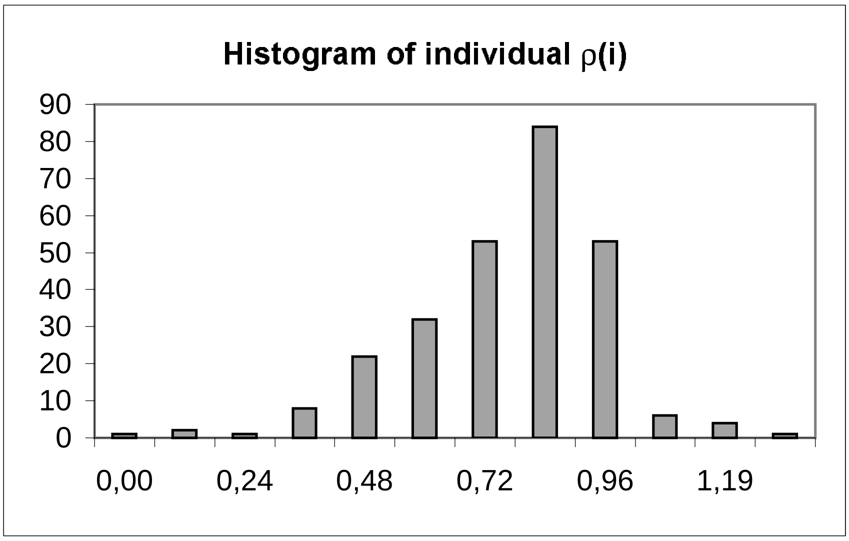
PO	0,000	0,001	0,000
π*	0,000	0,000	0,000

Notes

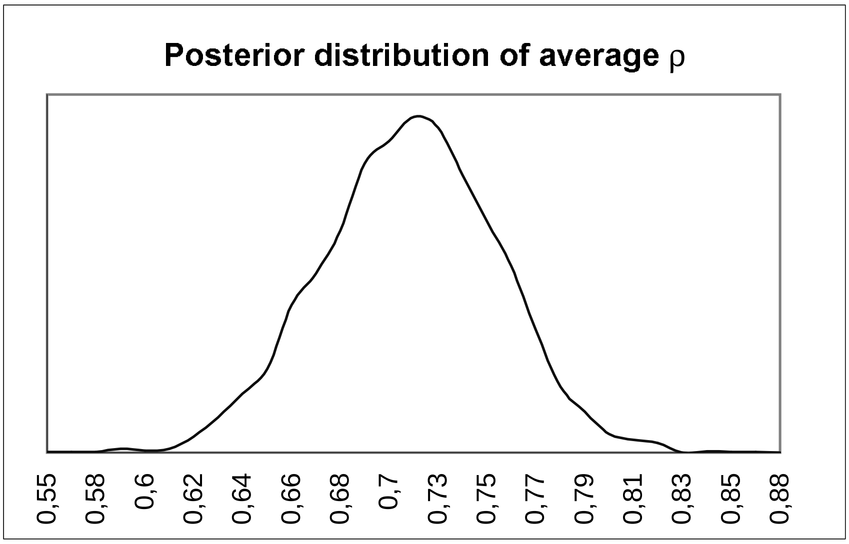
1. The prior probability π in 2.a and 2.b is chosen randomized over the interval (0.7, 1)
2. The prior probability π in 2.c is chosen equal to 0.5

Figure 3. Convergence rates (UK manufacturing)

A



B



C

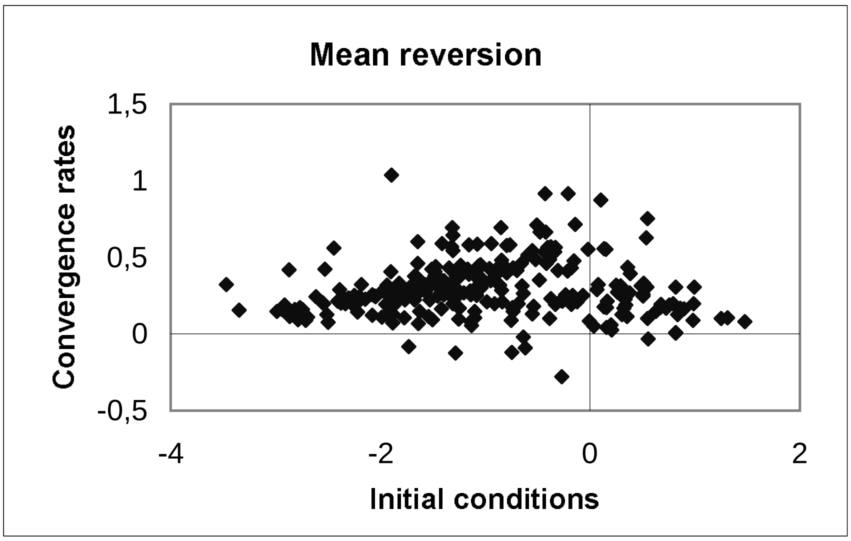
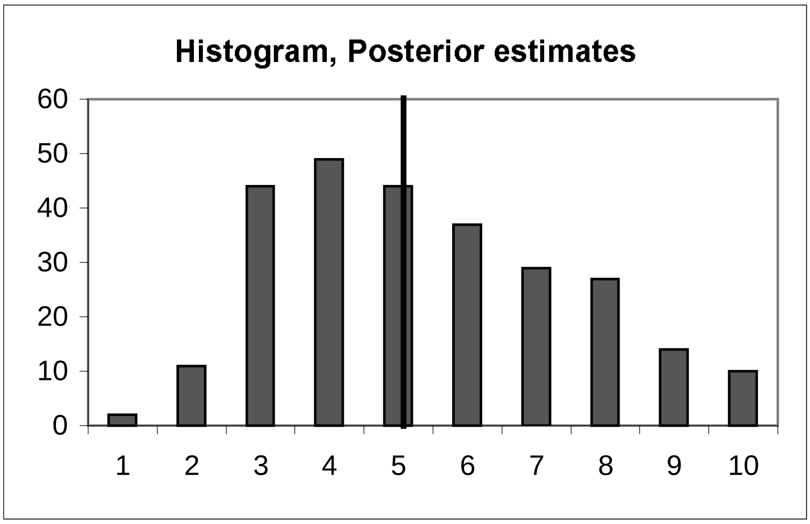
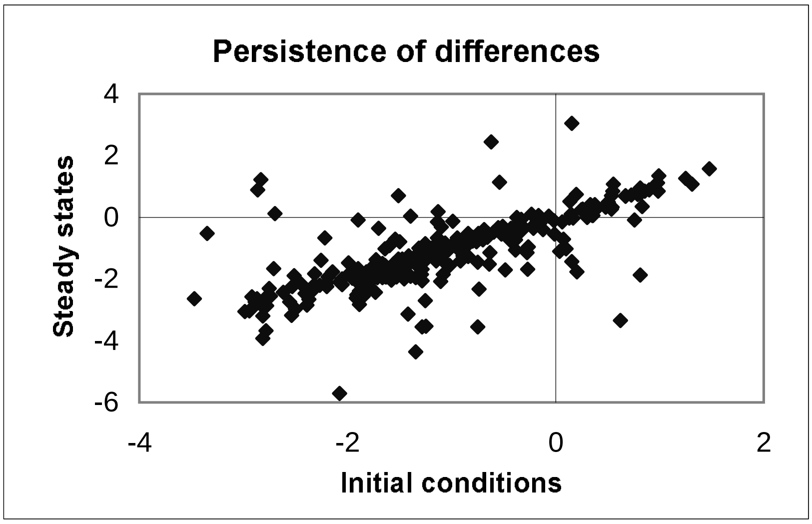


Figure 4. The steady state (UK manufacturing)

A



B



C

test of persistence in differences of firm size		
constant	initial condition	R_bar**2
-0,23	0,83	0,52
(0.07)	(0.049)	

Figure 5. Distribution of firms with ρ = 1, across industries

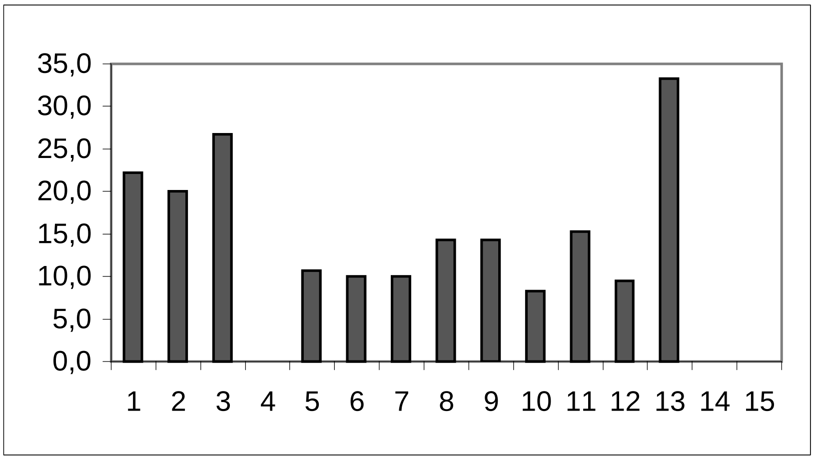


Table 3. Average estimates of ρ (UK manufacturing)

	1st quart	mean	median	3d quart
Σ = 0	0,992	0,994	0,995	1,000
0 < Σ < inf	0,600	0,718	0,734	0,832

Table 4. Testing (UK manufacturing)

4.a unit root: by firm

	Σ = 0		**0 < Σ < inf**	
	ρ(i) = 1	ρ(i)<1	ρ(i) = 1	ρ(i)<1
%	100,00	0,00	12,40	87,60
ln(PO)	3,15	na	0,24	-3,65
π*	0,96	na	0,75	0,43

4.b unit root: across firms (ρ=1)

ln(PO)	3,14	-26,6
π*	0,98	0,00

4.c equal steady states

PO	0,000	0,000
π*	0,000	0,000

Notes

1. The prior probability π in 4.a and 4.b is chosen randomized over the interval (0.5, 1)
2. The prior probability π in 4.c is chosen equal to 0.5