NATIONAL INSTITUTE OF ECONOMIC AND SOCIAL RESEARCH

GROWTH AND SIZE OF FIRMS

Peter E. Hart and Nicholas Oulton

Discussion Paper no. 77

2 DEAN TRENCH STREET SMITH SQUARE LONDON SW1P 3HE TELEPHONE 071 222 7665 FACSIMILE 071 222 1435

REGISTERED CHARITY NUMBER 306083

NATIONAL INSTITUTE DISCUSSION PAPERS

Discussion Papers range over the whole field of macro and micro economics including studies in the field of European integration, productivity and industrial policy and macroeconomic simulations.

This publication is available at £3.00 per single copy (£4.00 non-EU), £20.00 for nos. 1-10 in the series, or £22.00 for 11-20, and subsequent sets of 10 (£30.00 non-EU). If you would like to order any of these papers, please contact Gill Clisham at the Institute who will then invoice you as appropriate.

GROWTH AND SIZE OF FIRMS

Peter E. Hart and Nicholas Oulton

University of Reading and National Institute of Economic and Social Research

NIESR Discussion Paper No. 77

January 1995

ABSTRACT

This paper uses a dataset of some 87,000 independent UK companies, the great majority small to medium size, to investigate the relationship between firm size and firm growth. Three measures of size are considered: employment, sales and net assets. Growth over 1, 2 and 4 years is examined, using the Galtonian model of regression towards the mean. For the sample as a whole we find strong support for regression towards the mean: that is, growth is negatively related to initial size and proportionately more new jobs are created by small firms than by large firms. However, when the sample is broken down by size group, we find that regression towards the mean only occurs for the smallest firms, e.g. those with less than 8 employees. For larger firms, there is essentially no relationship between growth and size. Even for the smallest firms, our results may be due to transitory factors.

Keywords *Growth, job creation, size distribution, small firms*
JEL codes *L11, J23*

1. INTRODUCTION[1]

Are more new jobs generated by small firms than by large firms? In the USA the usual answer is, "Yes". Davis, Haltiwanger and Schuh (1993) cite quotations in support of this view from President Clinton, the US Small Business Administration, and an impressive array of commercial opinion. This widespread belief, which influences government tax policy, is based on a series of academic studies of employment growth and size of firm including the influential work of Birch (1987). Davis *et al.* (1993) then argue that this belief is wrong: it is based on unsuitable Dun and Bradstreet data and in any case the data have been misinterpreted by not allowing for the dynamics of size distributions.

In the UK there is an equally widespread belief that small firms are the most important generators of new jobs and an equally impressive list of official and commercial support for this view could be compiled. The most recent examples are the Kleinwort Benson Securities study of CBI data since 1976 (reported in the Independent of August 9th 1994) and the Department of Employment study (reported in the Employment Gazette, August 1994) which show that small firms grow more quickly than large firms. There is also an extensive academic literature on this topic, some of which may be open to the criticisms advanced by Davis *et al.* (1993). For example, Gallagher and Doyle (1986) used Dun and Bradstreet data in this and in subsequent publications. Storey and Johnson (1987) compared two univariate size distributions of firms instead of performing a dynamic analysis of a bivariate size distribution in their study of job generation and size of firm. Presumably, Davis *et al.* (1993) would also claim that when Dunne and Hughes (1994) found that small firms grew more rapidly than large firms 1975-80 and 1980-85, using the EXSTAT database, they used initial net assets as a measure of size and thus failed to allow for transitory components.

It is possible, therefore, that the conclusions that small firms in the UK are growing more quickly than large firms and are generating more jobs are as erroneous as Davis *et al.* (1993) claim them to be in the USA. The present paper investigates this possibility using the hitherto unexploited, large database created by OneSource Information Services Ltd. The 50 to

[1] This research has been made possible by a grant from 3i plc. We would like to thank Ewen Macpherson, Chief Executive of 3i, for his support for this work. Any conclusions and views expressed are our own and should not be assumed to be necessarily the same as those of 3i.

80 thousand companies studied here include very small firms (nearly 22,000 companies with less than 17 employees) and overcome the problem of the lack of coverage of small firms in the EXSTAT database, which is clearly recognised by Dunne and Hughes (1994). Obviously, it is essential to have adequate data on the lower tail of the size distribution of companies if we wish to examine the relative growth of small firms and their role as job generators. Our paper attempts to fill this gap.

In explaining the importance of the dynamics of company size distributions in this context, we are led to the extensive literature on stochastic growth processes. The size distribution of firms is positively skew and there are many stochastic models of growth which account for this. Such processes are important first because the existence of large numbers of small firms makes it possible for a large number of new jobs to be generated if each small firm took on only one extra employee. For example, if the 10.6 per cent of companies below 17 employees in our OneSource sample holds for the total population of 2,346,614 firms in Bannock and Daly (1990), then an extra employee in each of these small firms would increase employment by nearly 250,000. Secondly, while many of the models incorporate Gibrat's (1931) law of proportionate effect, stating that the proportionate growth of a firm is independent of its initial size, it is also possible to use a first order Galton-Markov process which allows the proportionate growth of small firms to exceed that of large firms and still generate the kind of positively skew size distribution which is observed in practice. If small firms are in fact growing more quickly than large firms, there is another reason for emphasising their importance in the generation of new jobs.

Stochastic models are discussed in section 4. Needless to say, when emphasising stochastic factors we certainly do not wish to suggest that systematic forces have no effect on firms' growth. This is obviously not true, as has long been recognised. The stochastic model in Simon and Bonini (1958) provided for constant returns to scale above minimum efficient size. Lucas (1978) formulated a theory which links the skew size distribution to the distribution of entrepreneurial talent, while Jovanovic (1982) linked it to the distribution of efficiency. Even more progress in combining economic and statistical theory has been made in the context of income distributions, which are also positively skew and appear to be generated by multiplicative stochastic shocks. Creedy, Lye and Martin (1994) combine an economic model of the labour market with a Wiener process to generate a generalised gamma distribution

which corresponds to the observed distribution of wages. We concentrate on stochastic models simply because it is convenient to do so when investigating the relationship between the size and growth of firms, which lies at the root of the debate on the job generating propensities of small and large firms.

Before discussing these models we describe the data in section 2 and the measures of size in section 3. Section 5 presents the results of our statistical analysis of the relationship between the size and growth of firms in the UK over 1990-94. The conclusions are given in section 6.

2. THE DATA

Firms in the private sector may be companies, partnerships or sole proprietorships. A complete count of all the different types of firm, especially sole proprietorships, encounters formidable statistical problems created by the 62 per cent increase in the number of self-employed between March 1979 and March 1994. As a proportion of the total workforce, this increase was even more striking, from 7.6 per cent in 1979 to 13 per cent in 1994.[2] Many of the self-employed are unemployed workers trying to avoid the dole by using their redundancy payments and their skills, by themselves or in partnerships, to earn a living (Storey, 1987, pp. 175-183). In some cases companies have encouraged their employees to become self-employed, in order to avoid paying employers' superannuation and national insurance contributions. In return companies have provided them with short-term contracts as consultants etc. This casualisation of work has obvious advantages for any company with fluctuations in its demand for labour and is regarded as extending the practice of 'just in time' management of raw materials and components.

Because changes in the numbers of self-employed reflect the casualisation of work rather than increases in the numbers of entrepreneurs, the statistical analysis in this paper is confined to the company sector, or at least to that part of it in the OneSource database. The properties of this database are most easily seen by comparing it with the other major databases used to measure the behaviour of firms. It

[2] See Employment Gazette, November 1994, Table 1.1, page S6 and Historical Supplement, October 1994, Table 1.1, page 6.

must be admitted that restricting firms to companies may well confine the smallest new entrants to those partnerships and sole proprietorships which are growing so rapidly that they acquire corporate status. This might create an upward bias in the recorded growth of the smallest companies, but similar problems arise in other databases which are not confined to companies.

The databases used as comparators are the Annual Census of Production (ACOP), Datastream, Dun and Bradstreet, and EXSTAT.

The Annual Census of Production (ACOP)

The Annual Census of Production is confined to the manufacturing, construction and utility industries, whereas the OneSource Database covers companies in all industries. The Business Statistics Office (BSO) maintains a register of legal business units (companies, partnerships, public authorities) based on Value Added Tax data for the industries it covers. Within this register the BSO records the employment of each local unit (factory or site) and these local units are linked to the legal units which own them. The ACOP term 'business' refers to the unit which reports to it and this depends on the level of the disaggregation of the accounting data kept by the firm. If the lowest level of aggregation is the company then this is the 'business' which reports to ACOP. If the data are recorded for each local unit belonging to the company, then the 'business' is the local unit.

A higher level of aggregation is the 'enterprise' which comprises a group of companies controlled by the same parent company. This comes nearest to the economist's definition of the firm. The OneSource database does not provide information on enterprises as such, but it does divide its companies into subsidiaries and non-subsidiaries (independent companies). The latter are used here.

The key ACOP data on output, sales etc. are truncated. Businesses with less than 20 employees are excluded. Thus analyses of new entrants based on ACOP data, as in Geroski (1991), exclude most of the new businesses because we know from the data compiled from HM Inspector of Factories that between 1981 and 1984 over 81 per cent of the new independent businesses had less than 25 employees (Beesley and Hamilton, 1994, special tabulation). A similar truncation problem arises when VAT data are used to measure the number of firms, as in Ganguly (1985), because firms below the VAT threshold value of sales do not

have to register for VAT. This threshold was £37,600 in 1993/94 and has been increased in real terms over recent years. In contrast, the OneSource database is not truncated.

The number of businesses in the BSO register used by ACOP sometimes shows implausibly large changes. For example, between 1983 and 1984 the number increased from 107,000 to 138,000 due to the discovery of 30,000 extra small firms. The OneSource database has not experienced such sudden changes so far. While ACOP contains data on employment, output and sales, it excludes assets. The OneSource database has data on employment, sales and assets, and also value added, though the latter is only available for a relatively small number of firms. But the most important difference between ACOP and OneSource is that the data on individual firms in ACOP is confidential so no links between variables such as sales and employment, or between employment in two years, can be made at the level of the individual firm.[3]

Datastream

Datastream International is part of the Dun and Bradstreet Corporation and belongs to the same D&B group as Moody's Investors Service and Interactive Data. One of its many databases relates to company accounts. In 1989 this covered all UK quoted industrial companies, all major UK financial companies, all USM quoted companies, and 420 of the largest unquoted UK companies and foreign-owned subsidiaries. There were accounts for more than 80,000 unquoted companies. It is a very large database though it is not as large as the OneSource database with its grand total of more than 140,000 companies. Because the OneSource has more data on individual small companies, which are crucial to a study of the size and growth of firms, we use the OneSource data.

EXSTAT

EXSTAT is a commercial database of company accounts and includes the larger unquoted companies in addition to those quoted on the Stock Exchange. Dunne and Hughes (1994) found that of the 2,149 companies recorded in 1980 only 1,709 survived to 1985. They recognised that the vast majority of unquoted companies are excluded. This is clearly brought out by huge differences between the numbers of independent companies in the OneSource and the EXSTAT databases. Although 25

[3] Individual researchers can have regressions run for them by the CSO, but this is an expensive and time-consuming process. Also, researchers are still not allowed access to the individual data in the form, for example, of scatterplots.

per cent of the sample of EXSTAT companies used by Dunne and Hughes (1994) had less than 500 employees, the upper size limit of SME (small and medium enterprises) used by the European Community, the OneSource database is more appropriate for a comparison of the relative growths of small and large firms because of its superior coverage of small firms.

Dun and Bradstreet

Dun and Bradstreet provided credit ratings for some 360,000 firms in the UK in 1987. By 1989 they had records on some 450,000 UK businesses, according to Robson and Gallagher (1994). This database has been used by Gallagher and his colleagues at the University of Newcastle in a series of studies of the number of jobs generated by firms of different sizes. See, for example, Gallagher and Doyle (1986a), Gallagher and Stewart (1986), Stewart and Gallagher (1986), Doyle and Gallagher (1987), and Gallagher, Daly and Thomason (1991), Robson and Gallagher (1994). This database has been criticised by Storey and Johnson (1987) (1990). Because its primary purpose is to provide credit ratings, it is likely to be biased towards rapidly expanding small firms for which the data compiled have a market value. The less dynamic small firms will tend to be excluded. The under-sampling of small firms is accentuated by the fact that the database is biased towards companies and hence excludes many unincorporated small firms. Moreover, point estimates of a firm's employment are not compiled. Instead, each firm is allocated to an employment size class and that class is used to indicate size. Many clerical errors in the data have been discovered and over the years Gallagher and his colleagues have spent much time in cleaning and improving the data.

OneSource

The OneSource database is compiled by OneSource Information Services Ltd and comprises some 140,000 companies, of which about 87,000 are independent (non-subsidiaries). It is a subset of a much larger database of some 750,000 companies compiled by ICC Online Ltd from accounts held at Companies House. Companies House holds the records of some 1.2 million companies. The 750,000 companies are those which are considered by ICC to be actively trading. The data used in the present paper were extracted in February 1994 from a CD-ROM entitled "CD/Private+: UK". Since then, OneSource has changed its policy. It now prepares two CD-ROMs, namely one for 140,000 larger companies with more than £300,000 sales or net worth, entitled "UK Private+", and

the other for smaller companies below this size. The old CD-ROM we are using overlaps the two current CD-ROMs and is not the same as the first, in spite of a similar name and the same number of companies.

In addition to the latest available accounts data from each company (the profit and loss account and the balance sheet), the database also contains data relating to 1, 2 and 4 years earlier. This is essential to our analysis of the relationship between growth and size of company. Though our analysis covers broadly the period 1990-94, it should be emphasised that all our results relate to companies which were present in the database in February 1994.[4] Many thousands of companies have entered and left this database between 1990 and 1994 but the paper is confined to those which were present in both 1990 and 1994, called the "survivors". The others, called "births" and "deaths", merit detailed analysis in a separate paper.

Information on a total of 87,109 independent (non-subsidiary) companies and 53,084 subsidiaries was available. In the present paper, we concentrate on independent companies, in part because the financial results of subsidiaries depend on the accounting policies of their parent companies. The number of independent companies available for analysis depends on the size measure. Virtually all companies supply balance sheet data since this is a legal requirement, whatever the size of company. Sales need not be revealed by smaller companies, and no company is obliged to publish employment. Fortunately, however, many companies go beyond the minimum legal requirements. As can be seen from Table 1, information on employment, the key variable for any analysis of job-generation, was available for 50,441 independent companies. Table 2 shows that sales data were available for 57,812 independent companies and Table 3 gives data on net assets (total assets minus current liabilities) for 79,491 independents. The latter figure is less than the theoretical maximum of 87,109 and this is almost entirely due to the remaining companies having negative net assets; these had to be excluded since our analysis employs a logarithmic model.

[4] Though the data were extracted at a point in time, it should be noted that companies' accounting years differ and some companies are faster than others in supplying their accounts to Companies House. Hence the available data relate to slightly different time periods. The regression analysis to follow takes this into account by means of dummy variables. Dummy variables are also employed to control for different SIC Divisions to which the companies belong.

Clearly, the OneSource database is huge, with companies of all sizes from all sectors of the economy. It may be regarded as an enormous sample with sampling fractions increasing with size of company. For the larger companies the sampling fraction is probably nearly 100% so the sample approximates the population, in which case sampling errors are negligible and may be ignored. For the smaller companies the sampling fractions are lower but the size of sample is so large that sampling errors are very small and unimportant. As a consequence, we may expect all the standard statistical tests of the results to be significant. But this simply reflects the fact that we have enough data to discern differences which are significant in the statistical sense: it does not mean that the differences are important.

3. SIZES OF COMPANIES

The size of company may be measured in many ways. Employment, assets, sales, market value, and value added are some of the more common measures used. Each measure has its limitations. Employment is only one input and may be a misleading guide to the size of a large capital-intensive company. The value of a company's assets reported in its published accounts is usually in terms of historic costs and has obvious limitations in times of changing asset prices. Sales figures are also affected by price changes. Moreover, they do not correct for the value of inputs so that even a small retail company would tend to be large when sales are used to measure size. Market value can be used only for those companies quoted on the Stock Exchange and value added cannot always be computed from published company accounts. In practice the choice of measure is governed by the data available. This is the case with the OneSource database where the measures readily available include employment, sales and assets.

The number of employees is a discrete variable which creates problems when using it to measure the size of companies, particularly those in the smallest size classes. Fractions of employees are not recorded even though the number of part-time workers has increased substantially in recent years. In 1981 70 per cent of the workforce had full-time jobs. By 1986 this proportion was 65 per cent and by 1993 it had fallen to 61.8 per cent (Watson 1994). In principle, we should use Full Time Equivalents, but these figures are not reported. The continuous decline in the proportion of the workforce in full-time jobs qualifies the crude

counts of jobs created which, though very welcome, are so often part-time jobs.

More non-subsidiary companies in the OneSource database report sales than report employment, as can be seen in Tables 1 and 2. This gives sales an advantage over employment. The limitation of the sales measure resulting from the exclusion of inputs can be countered, to some extent, by using a dummy variable for industry group. This permits, for example, the average industry effect on sales growth to be distinguished from the sales growth of the average firm.

Although all companies in the OneSource database are required to provide balance sheet data, they are not required to give data on employment and sales, though many do. The net assets figure is defined as total company assets minus its current liabilities (i.e. those falling due within one year). Total assets are fixed plus intangible plus intermediate plus current assets. Current assets are stocks plus debtors plus other current assets. Table 3 shows that more companies report net assets than report the other measures, but against this there is the disadvantage that net assets can be negative whereas employment and sales are non-negative. Companies where current liabilities exceed total assets in any year had to be excluded because we use the logarithms of size.

The figures of sales and assets are interesting in their own right but for present purposes they may be regarded as controls to check the results obtained by using employment. Employment is the key variable because this paper is primarily concerned with the job generation propensities of companies of different sizes. Nevertheless, Tables 1-3 clearly show that the size distributions of companies have broadly similar shapes irrespective of the measure used and even though the number of companies varies from 50,441 to 79,491. The first moment distributions are also given so that the co-ordinates of the Lorenz curves are readily computed. The extreme positive skewness of each distribution is shown by the fact that the top one per cent of the companies has 69 per cent of total employment, some 86 per cent of total sales, and some 92 per cent of total net assets.

The size classes in Tables 1-3 equal unity in terms of logarithms to the base two and in effect take a logarithmic transformation of the data. The transformed size distributions are plotted in Figures 1-3 and are symmetrical. This is confirmed by Table 4 which summarises the distributions by the mean, standard deviation, skewness and kurtosis of the natural logarithms of size: the skewness is low. In fact the deviations of Figures 1-3 from a normal curve are fairly small with the possible exception of the kurtosis in Figure 2. Of course, with such large numbers of observations formal tests of normality, such as that by D'Agostino et al (1990) show statistically significant differences from strict normality. But as Berkson (1938) noted many years ago '...it is practically certain that any series of real observations does not follow a normal curve with *absolute exactitude* in all respects, and no matter how small the discrepancy between the normal curve and the true curve of observations, the Chi-square P will be small if the sample has a sufficiently large number of observations in it.'

Again, Savage (1954) argued that statistical tests should not be applied to extreme null hypotheses (those which could not readily be accepted by anyone as being exactly correct), but we should accept the hypotheses if they are approximately true and if they simplify our task. Ijiri and Simon (1977) in their classic work on skew distributions (particularly chapter 6) explain at length why tests such as Chi-square should not be used to test extreme hypotheses. In their words, 'This has been known to mathematical statisticians for many years, but their knowledge has been slow to percolate into textbooks, much less into practices of statistics, of journal editors or of referees' (p. 110). Similar critiques of the common use of significance tests in econometric studies may be found in the work of Leamer (1983) and McCloskey (1986). Thus a strong case can be made for assuming that the size distributions of companies by employment, sales or assets are lognormal, in spite of the measurements of skewness and kurtosis in Table 4. Such an assumption would be very convenient because the genesis of a lognormal distribution provides powerful insights into the relationship between a company's size and its proportionate growth.

So in the first instance, let us accept the lognormal hypothesis, at least as a working assumption, in the examination of the relationship between growth and company size, using stochastic models of firms' growth and paying special attention to any non-linearity of this relationship. This is done in sections 4 and 5.

4. STOCHASTIC MODELS OF FIRMS' GROWTH

The Gross Domestic Product of a nation increases when the output of its constituent firms increases. The growth of the firm is central to any explanation of the growth of an economy and it is not surprising that so many reasons have been suggested for such development. There is a long list of systematic forces which are thought to stimulate increases in output: investment in human and physical capital inputs, more efficient use of raw materials, better management and organisation of production, improved industrial relations, more intensive competition and the abolition of restrictive practices, lower taxation and the provision of stronger incentives, and many other causes of economic growth have been discussed in the extensive literature on the subject. Superimposed upon all these systematic forces is a large stochastic factor. A small minority of firms may implement all the correct policies to stimulate growth, but will still decline: at the other extreme, a small minority of firms will do all the wrong things, and will nevertheless grow. These perverse results will arise as the result of chance: storms and floods, earthquakes, wars, terrorism, change of government, Stock Exchange bubbles, health scares and a multitude of other random effects will influence a firm's growth. These stochastic shocks will outweigh the systematic forces in so many cases that the resulting skew size distribution of firms by output will appear to be generated by a multiplicative stochastic process.

The emphasis on the respective roles of small and large firms means that the important stochastic models generating a Pareto distribution, formulated for example by Steindl (1965), are not considered here because they are concerned solely with the upper tail of the size distribution. In any case the Pareto curve does not appear to be a good first approximation to the observed distribution of independent firms by employment.[5] Thus we turn to another popular theoretical distribution, the simple two-parameter lognormal, denoted by $\Lambda(\mu,\sigma^2)$. The two parameters μ and σ^2 denote the arithmetic mean and variance of the natural logarithms of employment.

If the size distribution of firms by employment is $\Lambda(\mu,\sigma^2)$, then the distribution by the natural logarithm of employment is normal, $Y \sim$

[5] This was confirmed by plotting $\ln[1-F(x)]$ against x, where $F(x)$ is the cumulative distribution function of x, the log of size. For a Pareto distribution, the plot should be a straight line, but visually this was far from the case.

$N(\mu,\sigma^2)$ where $Y = \ln X$. Let us denote the natural logarithm of the employment of the ith firm at time t by $Y_i(t)$, with $i = 1,...,n$, and $t= 0,...,T$. In deviations from the mean, $y_i(t) = Y_i(t) - \mu(t)$.

Let us represent all stochastic forces which shock firm i by the logarithmic stochastic disturbance term $\varepsilon_i(t)$, with $E[\varepsilon_i(t)]= 0$, and $E[\varepsilon_i^2(t)] = \sigma^2$ for all t. We also assume that $\varepsilon_i(t)$ is independent of $y_i(t\text{-}1)$.

In the first stochastic model which generates a lognormal distribution, the Wiener process, we assume that $\varepsilon \sim N(0,\sigma^2)$ and we write

(1) $y_i(t) = y_i(t\text{-}1) + \varepsilon_i(t)$, for all i.

Hence

(2) $$y_i(t) = y_i(0) + \sum_{\tau=1}^{t} \varepsilon_i(\tau).$$

For large t, $y_i(t) \to N(0,t\sigma^2)$ since $\Sigma\varepsilon_i$ is $N(0,t\sigma^2)$ and $y_i(0)$ can be neglected when t is large. Alternatively, we may assume that $y_i(0)$ is $N(0,\sigma^2)$ and so $y_i(t) \sim N(0,(t+1)\sigma^2)$. In either case it follows that $y(t)$ is normally distributed simply because it is the sum of independent normal distributions.

In the second stochastic model, the Lindeberg-Levy process, we drop the assumption that the ε_i are normal but still retain the assumption that the ε_i are identically and independently distributed. In particular, σ^2 is the same for all t. As shown by Fisz (1963), the standardised form of the sum in (2), $y_i(t)/\sigma\sqrt{t}$, tends to be normally distributed even if the ε_i are not.

A third model uses the Liapounoff central limit process with even less restrictive assumptions. For example, the distributions of ε_i are assumed to be independent but not identical. Fisz (1963) shows that even under these less restrictive assumptions the standardised variable

(3) $$Z_t = \Sigma\{Y(t) - \mu(t)\} / C_t$$

tends to be normally distributed, where $\mu(t) = E\{\varepsilon(t)\}$ and $C_t^2 = \Sigma\sigma_t^2$. It is assumed that $\lim_{t\to\infty} C_t^2 = \infty$ so each σ_t^2 is small relative to C_t^2.

Each of these three models may be used to interpret Gibrat's (1931) law of proportionate effect in (1), where the proportionate change in the size of a firm is independent of its initial size.

Equation (1) may be extended to allow $y_i(t\text{-}1)$ to influence $y_i(t)$ in a first order Galton-Markov process to provide a fourth stochastic model of firms' growth:

(4) $y_i(t) = \beta y_i(t\text{-}1) + \varepsilon_i(t)$.

Equation (4) is our basic regression model. For future reference we may note that we can define the reverse regression as

(4') $y_i(t\text{-}1) = \alpha y_i(t) + \eta_i(t\text{-}1)$.

The model of equation (4) is implicit in Galton (1889, 1892), was repeated by Kalecki (1945), and its slope coefficient was estimated by Hart and Prais (1956) by:

(5) $$b = \sum_{i=1}^{n} y_i(t)\, y_i(t\text{-}1) / \sum_{i=1}^{n} y_i(t\text{-}1)^2.$$

Note that b is a cross-section regression across i and not a time-series estimate over t. With this in mind, let us now drop subscript i.

If we assume that β is constant over time the revised form of equation (2) becomes

(6) $$y(t) = \beta^t y(0) + \sum_{\tau=1}^{t} \beta^{t-\tau}\varepsilon(\tau), \qquad t \geq \tau$$

Is $y(t)$ likely to be normal? Clearly, if $\beta = 1$ the Wiener, Lindeberg-Levy and Liapounoff models can be used to produce a normally distributed $y(t)$. If $\beta > 1$ then $y(0)$ dominates $y(t)$ and we would have to assume that $y(0)$ is normal, being determined long ago when β was unity. This, together with the Wiener assumption of normally distributed ε, would generate a normal $y(t)$. The Lindeberg-Levy assumption of constant variance does not hold because $V[\beta^{t-\tau}\varepsilon(\tau)]$ varies with t, even if $V[\varepsilon(\tau)] = \sigma^2$ for all t. But the Liapounoff theorem can be applied because

$$(7) \qquad C_t^2 = \sum_{\tau=1}^{t} \beta^{2(t-\tau)}\sigma^2(\tau) \qquad \text{and } \lim_{t\to\infty} C_t^2 = \infty.$$

If $\beta < 1$ neither the Lindeberg-Levy nor the Liapounoff theorems holds: the variances of $\beta^{t-\tau}\varepsilon(\tau)$ are not constant and the weights $\beta^{2(t-\tau)}$ diminish with increases in t so that in the limit C^2 does not equal infinity. To generate a normal $y(t)$, we would have to use the Wiener assumptions. But to assume that the ε are normally, independently and identically distributed is tantamount to assuming that the y are normal because a linear combination of independent normal distributions is normal. The conclusion is that we cannot use the central limit theorems of Lindeberg-Levy or of Liapounoff to justify a normal $y(t)$ when $\beta < 1$, which occurs when small firms grow more quickly than large firms.

Although the limiting distribution of $y(t)$ does not tend to normality when $\beta < 1$, the departure from strict normality is unlikely to be very large in practice for values of β which are not much below unity. With $\beta < 1$ we may ignore $\beta^t y(0)$ in equation (6) and consider the skewness and kurtosis of $y(t) = \sum\beta^{t-\tau}\varepsilon(\tau)$. Since the $\varepsilon(\tau)$ are independent, the rth cumulant is given by

$$(8) \qquad K_r\{\sum_{\tau=1}^{t} \beta^{t-\tau}\varepsilon(\tau)\} = K_r\{\beta^{t-1}\varepsilon_1\} + K_r\{\beta^{t-2}\varepsilon_2\} + \ldots$$

$$= \beta^{r(t-1)}K_r\{\varepsilon_1\} + \beta^{r(t-2)}K_r\{\varepsilon_2\} + \ldots + K_r\{\varepsilon_t\}$$

$$= K_r\{\varepsilon\}/(1 - \beta^r)$$

for large t, assuming the $\varepsilon(\tau)$ are identically distributed with zero means.

The following central moments of $y(t)$ may be obtained from the cumulants in (8):

$$(9) \qquad \mu_2 = K_2\{y(t)\} = \sigma^2/(1 - \beta^2) = C_t^2 = V[y(t)]$$

$$\mu_3 = K_3\{y(t)\} = K_3\{\varepsilon\}/(1 - \beta^3)$$

$$\mu_4 = K_4\{y(t)\} + 3\mu_2^2.$$

The skewness of $y(t)$ is given by

$$(10) \qquad \mu_3/\mu_2^{3/2} = K_3\{\varepsilon\}(1 - \beta^2)^{3/2}/\sigma^3(1 - \beta^3)$$

$$= K_3\{\varepsilon\}R_s/\sigma^3 \neq 0$$

where R_s (regression effect) is $(1 - \beta^2)^{3/2}/(1 - \beta^3)$.

The lowest estimate of β in Table 5 is 0.8299. Using this value, we may estimate R_s as 0.4054. Thus the skewness of $y(t)$ is only about 40 per cent of the skewness of $\varepsilon(t)$. The skewness of $\varepsilon(t)$ would have to be very large to make $y(t)$ depart very much from symmetry.

The kurtosis of $y(t)$ is given by

$$(11) \qquad \mu_4/\mu_2^2 - 3 = K_4\{y(t)\}/\mu_2^2$$

$$= K_4\{\varepsilon\}(1 - \beta^2)^2/\sigma^4(1 - \beta^4) \neq 0$$

$$= K_4\{\varepsilon\}R_k/\sigma^4$$

where $R_k = (1 - \beta^2)^2/(1 - \beta^4)$ measures the regression effect on kurtosis. Using the lowest estimate of β of 0.8299, we find that $R_k = 0.1844$. Thus the kurtosis of $y(t)$ is about 18 per cent of the kurtosis of $\varepsilon(t)$.

It can be seen in (10) and (11) that as $\beta\to 1$, $y(t)$ approaches normality. In practice with values of β just below unity, we should expect the size distribution of firms to be approximately, but not exactly, lognormal. This is confirmed by Figures 1-3 which show the size distribution of independent companies in the UK by their employment, sales and net assets. The skewness and the kurtosis measures in Table 4 show that the distributions are not exactly lognormal but the departures from strict lognormality are not very large.

The fact that the observed distributions are approximately lognormal means we are justified in assuming that a firm's size is heavily influenced by multiplicative stochastic shocks and the stochastic process of growth is summarised by the first-order Galton-Markov process in equation (4).

The use of the lognormal distribution certainly simplifies our task of measuring the rate of growth of the employment of small and large firms, but the fact remains that there are many other positively skew theoretical distributions which could be used. For example, Bartels (1977) proposed a short list of ten theoretical distributions, including the lognormal. In addition, there is the Pareto distribution favoured by Steindl (1965) and the Yule distribution used by Simon and Bonini (1958) and by Ijiri and Simon (1977).

The Yule distribution results from a stochastic process comprising Gibrat's law of proportionate effect and a constant birth rate of new firms into the smallest size class. It does not allow for the death of firms. Since the lognormal model in (1) or (4) relates to a constant set of firms and does not permit births, deaths or marriages (mergers), it might be regarded as less useful than the Yule distribution. Against this, the births permitted in the generation of a Yule distribution occur only in the smallest size class, whereas in practice they occur throughout the whole size distribution as a result of foreign multinational enterprises creating subsidiaries in the UK. Hence the assumed birth process of the Yule distribution might be misleading.

Ijiri and Simon (1977) also consider the serial correlation of firms' growth rates. Although such serial correlation of growth may be measured by correlating the cross-section residuals from estimates of (4) in consecutive periods, or by the multiple regression of $y(t)$ on $y(t-1)$ and $y(t-2)$, we do not yet have the OneSource data for enough years to investigate the important question of whether "success breeds success". But in principle it is possible to do this within the context of a lognormal model and it may well be important for the study of the persistence of the growth of small and large firms. Dunne and Hughes (1994), however, did not find that the growth of firms persisted from the period 1975-80 to the following quinquennium 1980-85 in their sample of 934 UK companies from the EXSTAT database.

Our OneSource data enable us to estimate (4) over only four years, which is insufficient to reveal serial correlation and barely sufficient time to allow Galtonian regression towards the mean to take place. A value of $\beta < 1$ would show that the employment of small firms grew more quickly than that of large firms during the period. As we shall see in the next section, it is possible that the ordinary least squares estimate of β is biased downwards as a result of transitory components in measured

employment, so that the faster growth of small firms might be exaggerated.

Equally important is the possibility that the assumption of linearity in (4) does not hold for small firms. For the period 1939-50 in the UK, Hart and Prais (1956) found that small companies grew more quickly than large companies and that there was a remarkable curvilinearity in the regression in the lower values, as shown by the plot of $E[y(1950) \mid y(1939)]$. It is possible that a similar curvilinearity exists for the period 1990-94, which is also considered in the next section.

5. SIZE AND GROWTH OF FIRMS

Table 5 gives the results of the Galtonian regressions specified in equation (4) estimated by OLS, for periods of one, two and four years and measuring size by employment, sales or net assets.[6] The number of independent companies available varies with the size measure and with the length of time period. But even the smallest number is over 29 thousand and is still an extremely large sample though it is only some 58 per cent of the total in Table 1. The decline in the number of companies in the sample panel as the time period increases reminds us that birth and death rates of companies, especially in the smallest size classes, are very large and we must emphasise once again that in this paper we are considering the size and growth of surviving companies only.

Table 5 shows that the cross section OLS estimate of β for each time period and for each size measure is significantly below unity. The large numbers of companies guarantee small standard errors, so little weight should be attached to the word 'significantly'. But the fact that each of the nine estimates of β is below unity suggests that Galtonian regression towards the mean was taking place in this period. That is, on the average, small firms grew more quickly than large firms. Since one size measure used is employment it follows that the smaller surviving firms

[6] Apart from the lagged size measure and a constant, these regressions also included nine dummy variables for the ten divisions of the 1980 SIC (SIC Division 0 was the omitted category) and three dummies for the accounting year (1993, 1992, and 1991, with pre-1991 the omitted category).

generated more jobs, in proportion to their size, than did the larger firms.[7]

This initial conclusion may be challenged on the grounds that even the four year period, which is the maximum allowed by the data, is very short and transitory rather than permanent components of firms' growths are likely to dominate the results. The standard argument is that the presence of such transitory components or errors of measurement make the OLS estimator biased downwards. If this argument is correct the values of b in Table 5 are misleading: the true values of β are higher and the initial conclusion that small firms grew more quickly than larger firms might not be justified.

One alternative estimate of β is provided by the geometric mean of b and $1/a$, where a denotes the OLS estimator of the reverse regression coefficient α in (4'), which is the slope of the locus of the conditional expectations $E[\, y(t\text{-}1) \mid y(t)]$. This estimator has a long and distinguished history as explained, for example, by Prais (1958), Leamer (1978) and Maddala (1992). It is shown in the column headed "compromise" in Table 6 and was 0.9963 for the whole sample of 29,230 companies by employment. Indeed, it can be seen that it was also very near unity, at 1.0252 and 0.9965 in Tables 7 and 8, for the even larger samples of companies by sales and by net assets. Hence those who believe that transitory components in the size measures of each firm are very important in the short run of only four years might claim that there is no evidence that small firms grew more quickly than larger firms: the true value of β is near unity and the simple Gibrat model is a good approximation to the actual relationship between the size and growth of firms.

The standard argument that such transitory components make the OLS estimates biased downwards is based on the errors-in-variables model in econometrics. But this in turn is based on a series of additional assumptions, including zero correlation between the errors in $y(t)$ and $y(t\text{-}1)$, which might not command universal assent. Hence we adopt another approach to the criticism of the OLS estimates of β by examining the assumed linearity of the direct regression.

[7] The results of Table 5 are not materially affected by industrial composition effects. See Appendix A where similar regressions for each Division of the SIC (1980) are presented.

The reason for suspecting that there is a non-linearity is revealed by Figure 4. This plots the employment of each company in years t and t - 4, after the effects of the dummy variables on industry and accounting year have been removed. It can be seen that the regression line rotates clockwise away from the centre of the scatter diagram, which might be explained by a non-linearity in the regression relationship for small companies. Such a non-linearity may be shown graphically, as in Figures 5-7, or by the sequential regression estimates in Tables 6-8.

For each size class of companies in year t-4, Figure 5 plots the geometric mean of the same companies in year t; the data underlying this graph are given in Table 9. The graph thus estimates the conditional expectations $E[\, y(t) \mid y(t\text{-}1)]$. A 45 degree line is drawn through the point of means and if the Gibrat model held exactly each point would lie on this line. Since this does not happen, the simple Gibrat model does not hold. In particular, we see that there is an increasing divergence away from the 45 degree line as we go down the distribution from size class 5 (9-16 employees) in year t-4. The smallest firms appear to have grown more quickly than the larger firms. It is possible that this appearance is deceptive because part-time employees have been counted as one. But this possibility may be checked by using sales and net assets as measures of size, which are continuous rather than discrete variables even at the bottom of the distribution. In fact, Figures 6 and 7 show that there is a non-linearity in the regression relationship when sales or net assets are used to measure size.

A more formal measurement of this tendency is provided by the sequential regressions in Tables 6-8. In Table 6 it can be seen that the OLS regression coefficient increases monotonically from 0.4372 to 0.8061 as larger companies are added to the regression. It declines from 1.1027 to 0.8629 as we go down the distribution and more and more smaller firms are included in the regression. In a time-series econometric analysis, with its typically few observations, a sequential Chow test based on the recursive least squares residuals could be used to test for structural breaks, as shown by Harvey (1976, 1981). But in the present cross-section analysis, with enormous numbers of observations we must expect to observe statistically significant structural breaks as more size classes are included. This simply means that we have collected enough data to reveal small differences as we go down the distribution to size class 5. But these small differences are not very important above size class 5. However, there does seem to be a tendency for β to rise steadily as the average size of companies in the sample increases. For the largest companies, β exceeds 1 suggesting a positive relationship between size and growth, even though these companies were on average shrinking in size over this period.

Below size class 5 the divergence from the 45 degree line in Figure 5 is systematic. The fact that the overall regression of 0.8372 is so much smaller than the regressions for the larger firms, which range from 1.1027 down to 0.8629 when only companies above 4 employees are included, indicates that it is a misleading guide to the general relationship between size and growth. The overall regression is too low because of the non-linearity at the bottom of the distribution and the great weight provided by the large numbers of small firms in the overall estimate.

Table 7 uses sales instead of employment as a measure of size. The results are similar even though more observations are available. For all 34,774 companies the estimate of β is 0.8349 instead of 0.8372 in Table 6. Once again, it increases monotonically as we go up the size distribution. A further check is provided by using net assets to measure size, for which even more observations are available, namely 55,098. Table 8 shows that the overall estimate of β is 0.8299 and the regression coefficient still increases as we go up the size distribution. The fact that the results based on net assets are virtually the same as those using employment or sales, in spite of the large differences in the numbers of observations, suggests that the precise size measure used does not really matter when comparing the proportionate growth and size of company.

The provisional conclusion is clear enough: within the OneSource databases the surviving companies in the smallest four employment size classes, i.e. those with no more than 8 employees, grew more quickly than the larger survivors over the period 1990-94. In terms of employment, these smallest companies generated proportionately more jobs than did the larger companies. In fact, in this time period employment rose in absolute terms amongst these companies while falling amongst larger ones (Table 9(a)).[8] Among companies with more than 8 employees there was little tendency for the proportionate growth of the firm to vary with its size, according to Figure 5. So why should the Galton-Markov model be inappropriate for the growth of the very smallest firms?

One possibility is that transitory components are more important for the smallest companies than for the rest. The values of R^2 decrease sharply as we go down the distributions in Tables 6-8 and introduce more and more smaller companies into the regression so that β cannot be estimated with any precision for the smallest companies. For example, in Table 6 the 5,562 companies with less than 8 employees show a substantial regression towards the mean, with $b = 0.5855$, but R^2 is only 0.1641, indicating tremendous turbulence at the bottom of the distribution. We may be observing noise rather than a systematic tendency for the faster growth of the smallest firms. The compromise regression estimate of 1.7614, which should counter this noise, indicates that the proportionate growth of the firm tended to decrease with increases in size among companies below 8 employees.

But it is also possible that it really is easier for a firm with one employee to double its size by employment than it is for a company with 10 employees. In economic terms, there is a minimum efficient scale of firm and until this size is reached, the firm experiences decreasing average costs and can enjoy rapid growth. After this point, its average cost curve flattens out and it enters the world of constant average and marginal costs experienced by firms operating at above minimum efficient scale.

Thus the above statistical and economic explanations of the observed faster growths of the smallest firms qualify the provisional conclusion and suggest that caution is required before relying on the job generation propensity of small firms as an employment policy.

6. CONCLUSIONS

In the Gibrat model the proportionate growth of a firm is independent of its size and there is no tendency for small firms to grow more quickly than large firms. In terms of employment, small firms do not create more jobs than large firms. Such a growth process generates a lognormal distribution of firms and the fact that the observed sized distribution by employment of the sampled companies is approximately lognormal

[8] Overall, employment fell by about 2 per cent over the 4 years considered here: see Table 9(a) and the slightly different sample of companies for whom data is available for 1, 2 and 4 years earlier which is reported in Appendix A.

might suggest that employment is not growing more quickly in small firms than in large firms.

A closer examination of skewness and kurtosis of the distribution, after logarithmic transformation, indicates that there are in fact slight deviations from lognormality. These deviations may be explained by a Galton-Markov growth process in which smaller firms grow more quickly than large firms and regress towards the mean of the distribution. This regression was estimated by Ordinary Least Squares using panel data of companies by employment, sales and net assets. It was found that this regression was less than unity, which would support the belief that small firms were generating proportionately more jobs over the period 1990-94.

But the OLS regressions might be biased downwards as a result of transitory components or errors-in-variables. Alternative estimates based on the geometric mean of the direct and reverse regressions do not reveal any tendency of Galtonian regression towards the mean.

These results hold for the whole size distribution. When sequential regressions are estimated, by introducing companies in ascending or descending size classes into the regression, it appears that there is a non-linearity in the basic loglinear regression and that the growth process of the very smallest firms (especially those below 8 employees) does differ from that of the larger firms. The loglinear growth process in equation (4), which yields satisfactory results for companies in the larger size classes, does not hold for the very smallest companies.

This non-linearity is confirmed by Figures 5-7 which show the locus of the logarithms geometric means of the distributions of company in year t conditional upon their sizes in year t-4. There are clear divergences of these conditional means away from the 45 degree line at the bottom of the joint distribution of company size in years t and t-4, whether size is measured by employment, sales or net assets.

The sequential regressions and the conditional means indicate that small firms were growing more quickly than large firms over the period 1990-94 and hence were generating proportionately more jobs. There may be economic reasons for this, including the relatively rapid growth of small firms towards their minimum efficient scale. There may also be statistical reasons. It is possible that the OneSource database has a

selection bias among the smallest companies: only the faster growing small firms are included because the slow growers are unlikely to seek company status.

In any case, the statistical result that on the average the very smallest companies have the fastest growth is recorded against a background of much turbulence among these small firms, as measured by the very low values of R^2 in the smallest size classes. It would be unwise to base any employment policy on the faster average growth of small firms until we know why the smallest companies enjoy such rapid growth. It is necessary to select a small sample from the companies in Table 1 and examine them in more detail to explain why they have generated more jobs. It is also necessary to investigate births and deaths of companies to see whether the net generation of jobs is greater for small firms. These are the next steps in our research.

REFERENCES

Bannock, G. and Daly, M.J. (1990). 'The size distribution of UK firms.' *Employment Gazette*, May, pp. 255-258.

Bartels, C.P.A. (1977). *Economic Aspects of Regional Welfare: Income Distribution and Unemployment.* Leiden: Martinus Nijhoff.

Beesley, M.E. and Hamilton, R.T. (1994). 'Entry propensity, the supply of entrants, and the spatial distribution of business units'. *Regional Economics*, 28 (3), pp. 233-239.

Berkson, J. (1938). 'Some difficulties of interpretation encountered in the application of the Chi-square test'. *Journal of the American Statistical Association,* 33, pp. 526-542.

Birch, D.L. (1979). *Job Creation in America: How Our Smallest Companies Put Most People to Work.* New York: The Free Press.

Creedy, J., Lye, J.N. and Martin, V.L. (1994). 'A labour market equilibrium model of the personal distribution of earnings.' University of Melbourne, Department of Economics Research Paper No. 414.

Davis, S.J., Haltiwanger, J., Schuh, S. (1993). 'Small business and job creation: dissecting the myth and reassessing the facts.' National Bureau of Economic Research, Inc. Working Paper No. 4492.

Doyle, J.R. and Gallagher, C.C. (1987). "The size distribution, growth potential and job generation contribution of UK firms.' *International Small Business Journal*, 6, pp. 31-56.

Dunne, P. and Hughes, A. (1994). 'Age, size, growth and survival: UK companies in the 1980s.' *Journal of Industrial Economics*, 42, pp. 115-140.

Fisz, M. (1963). *Probability Theory and Mathematical Statistics*. New York: Wiley.

Gallagher, C.C. and Doyle, J.R. (1986). 'Small firms - net job generators.' *British Business*. 17th October, pp. 36-8.

Gallagher, C.C. and Stewart, H. (1986). 'Jobs and the business life cycle in the UK.' *Applied Economics*, 18.8, pp. 875-900.

Gallagher, C.C., Daly, M.J. and Thomason, J.C. (1991). 'The growth of UK companies and their contribution to job generation, 1985-1987.' *Small Business Economics*, 3, 269-286.

Ganguly, P. (1985). *UK Small Business Statistics and International Comparisons.* London: Harper and Row.

Geroski, P. (1991). *Entry and Market Dynamics*. Oxford: Blackwell.

Gibrat, R. (1931). *Les Inegalités Economiques*. Paris: Sirey.

Hart, P.E. and Prais, S.J. (1956). 'The analysis of business concentration: a statistical approach.' *Journal of the Royal Statistical Society*, A, 119, pp. 150-191.

Ijiri, Y. and Simon, H.A. (1977). *Skew Distributions and the Sizes of Business Firms*. Amsterdam: North-Holland.

Jovanovic, B. (1982). 'Selection and the evolution of industry.' *Econometrica*, 50, pp. 649-670.

Leamer, E.E. (1978). *Specification Searches: Ad Hoc Inference with Nonexperimental Data*. New York: Wiley.

Leamer, E.E. (1983). 'Let's take the con out of econometrics'. *American Economic Review*, 73, pp. 31-43.

Lucas, R.E. (1978). 'On the size distribution of business firms.' *Bell Journal of Economics*, 9, pp. 508-523.

Maddala, G.S. (1992). *Introduction to Econometrics.* (2nd edition). New York: Macmillan.

McCloskey, D. (1986). *The Rhetoric of Economics*. Brighton: Wheatsheaf.

Prais, S.J. (1958). 'The statistical conditions for a change in business concentration.' *Review of Economics and Statistics*, 40, pp. 268-272.

Robson, G.B. and Gallagher, C.C. (1994). 'Changes in the size distribution of UK firms.' *Small Business Economics*, 6, pp. 299-312.

Savage, L.J. (1954). *The Foundations of Statistics*. New York: Wiley.

Simon, H.A. and Bonini, C.P. (1958). 'The size distribution of business firms.' *American Economic Review*, 48, pp. 607-617.

Steindl, J. (1965). *Random Processes and the Growth of Firms*. London: Griffin.

Stewart, H. and Gallagher, C.C. (1986). 'Business death and firm size in the UK.' *International Small Business Journal*, 4, pp. 42-57.

Storey, D.J and Johnson, S. (1987). *Job Generation and Labour Market Change.* Basingstoke: Macmillan.

Storey, D.J. and Johnson, S. (1990). 'A review of small business employment databases in the United Kingdom.' *Small Business Economics*, 2, pp. 279-299.

Watson, G. (1994). 'The flexible workforce and patterns of working hours in the UK.' *Employment Gazette*, July, pp. 239-247.

Table 1 Size distribution of employment: independent companies

	Range of employees	*No. of companies*	*% of companies*	*Mean employment*	*Total employment*	*Share of total employment (%)*	*Elements of Lorenz curve* Cumulative % of companies	Cumulative % of employment
1	1	536	1.1	1.0	536	0.0	1.1	0.0
2	2	2,013	4.0	2.0	4,026	0.0	5.1	0.0
3	3-4	4,075	8.1	3.5	14,274	0.1	13.1	0.2
4	5-8	6,645	13.2	6.4	42,586	0.4	26.3	0.6
5	9-16	8,666	17.2	12.2	105,486	1.0	43.5	1.6
6	17-32	8,882	17.6	23.4	207,776	2.0	61.1	3.5
7	33-64	7,972	15.8	46.5	371,073	3.5	76.9	7.0
8	65-128	5,769	11.4	89.8	517,995	4.9	88.3	11.9
9	129-256	2,903	5.8	177.4	514,973	4.8	94.1	16.7
10	257-512	1,399	2.8	354.9	496,547	4.7	96.9	21.4
11	513-1024	721	1.4	711.3	512,830	4.8	98.3	26.2
12	1025-2048	356	0.7	1,437	511,489	4.8	99.0	31.0
13	2049-4096	193	0.4	2,911	561,852	5.3	99.4	36.3
14	4097-8192	117	0.2	5,792	677,633	6.4	99.6	42.6
15	8193-16384	81	0.2	11,650	943,645	8.9	99.8	51.5
16	16385-32768	55	0.1	23,252	1,278,883	12.0	99.9	63.5
17	32769-65536	35	0.1	45,329	1,586,502	14.9	100.0	78.4
18	65537-131072	19	0.0	86,453	1,642,609	15.4	100.0	93.9
19	> 131072	4	0.0	163,359	653,436	6.1	100.0	100.0
	TOTAL	50,441	100.0	211.0	10,644,151	100.0		

Table 2 Size distribution of sales: independent companies

	Range of sales, £k	*No. of companies*	*% of companies*	*Mean sales, £k*	*Total sales, £k*	*Share of total sales (%)*	*Elements of Lorenz curve Cumulative % of companies*	*Cumulative % of sales*
1	0-100	2,923	5.1	42.7	124,814	0.0	5.1	0.0
2	101-200	1,943	3.4	149.1	289,773	0.0	8.4	0.0
3	301-400	3,292	5.7	298.1	981,488	0.1	14.1	0.1
4	401-800	13,621	23.6	611.5	8,328,575	0.5	37.7	0.5
5	801-1600	12,937	22.4	1,130	14,624,942	0.8	60.0	1.4
6	1601-3200	8,974	15.5	2,280	20,457,274	1.1	75.6	2.5
7	3201-6400	5,920	10.2	4,475	26,491,343	1.5	85.8	4.0
8	6401-12800	3,673	6.4	9,037	33,193,129	1.9	92.2	5.8
9	12801-25600	2,202	3.8	17,730	39,042,055	2.2	96.0	8.0
10	25601-51200	1,057	1.8	35,699	37,734,118	2.1	97.8	10.1
11	51201-102400	518	0.9	69,490	35,996,006	2.0	98.7	12.1
12	102401-204800	276	0.5	141,501	39,054,304	2.2	99.2	14.3
13	204801-409600	182	0.3	286,356	52,116,701	2.9	99.5	17.2
14	409601-819200	90	0.2	571,530	51,437,709	2.9	99.6	20.1
15	819201-1638400	73	0.1	1,185,472	86,539,456	4.8	99.8	24.9
16	1638401-3276800	61	0.1	2,411,387	147,094,607	8.2	99.9	33.2
17	3276801-6553600	43	0.1	4,645,804	199,769,572	11.2	100.0	44.3
18	6553601-13107200	17	0.0	9,221,816	156,770,872	8.8	100.0	53.1
19	> 13107200	10	0.0	84,000,000	840,000,000	46.9	100.0	100.0
	TOTAL	57,812	100.0	30,971	1,790,046,738	100.0		

Table 3 Size distribution of net assets: independent companies

	Range of net assets, £k	*No. of companies*	*% of companies*	*Mean net assets, £k*	*Total net assets, £k*	*Share of total net assets (%)*	*Elements of Lorenz curve Cumulative % of companies*	*Cumulative % of net assets*
1	0-100	23,725	29.8	43.5	1,033,220	0.1	29.8	0.1
2	101-200	12,104	15.2	145.5	1,761,632	0.1	45.1	0.2
3	301-400	12,794	16.1	288.9	3,696,300	0.3	61.2	0.5
4	401-800	11,204	14.1	569.5	6,380,831	0.5	75.3	1.0
5	801-1600	8,152	10.3	1,125	9,172,965	0.7	85.5	1.7
6	1601-3200	4,821	6.1	2,240	10,799,353	0.9	91.6	2.6
7	3201-6400	2,747	3.5	4,464	12,262,435	1.0	95.0	3.6
8	6401-12800	1,567	2.0	8,979	14,070,714	1.1	97.0	4.7
9	12801-25600	898	1.1	17,958	16,126,491	1.3	98.1	5.9
10	25601-51200	528	0.7	35,458	18,721,692	1.5	98.8	7.4
11	51201-102400	317	0.4	72,488	22,978,569	1.8	99.2	9.2
12	102401-204800	209	0.3	139,104	29,072,778	2.3	99.5	11.5
13	204801-409600	134	0.2	284,294	38,095,342	3.0	99.6	14.5
14	409601-819200	91	0.1	568,674	51,749,352	4.1	99.7	18.6
15	819201-1638400	69	0.1	1,151,082	79,424,658	6.3	99.8	24.9
16	1638401-3276800	59	0.1	2,235,475	131,893,025	10.4	99.9	35.3
17	3276801-6553600	31	0.0	4,470,308	138,579,548	10.9	99.9	46.2
18	6553601-13107200	23	0.0	9,098,067	209,255,541	16.5	100.0	62.8
19	> 13107200	18	0.0	26,200,000	471,600,000	37.2	100.0	100.0
	TOTAL	79,491	100.0	15,941	1,266,674,446	100.0		

Table 4 **Measures of size: summary statistics for independent companies**

	Size measure (in logs)		
	Employment	*Sales*	*Net assets*
Mean	3.1582	7.2015	5.5539
Standard deviation	1.5197	1.6628	1.9635
Skewness	0.7487	0.1932	0.4366
Kurtosis	1.5794	3.1876	1.8350
Number of companies	50,441	57,812	79,491

Note: Kurtosis defined so as to be zero for a normal distribution. 87,109 independent companies in all.

Table 5 **Galtonian regressions for 3 measures of size: size variable lagged 1, 2 or 4 periods**

		Number of lags in RHS size variable		
Statistic	*Size measure*	1	2	4
b	Employment	0.9645	0.9348	0.8372
	Sales	0.9362	0.9280	0.8349
	Net assets	0.9525	0.9307	0.8299
s.e.	Employment	0.0013	0.0021	0.0043
	Sales	0.0026	0.0033	0.0049
	Net assets	0.0016	0.0021	0.0033
N	Employment	48,238	37,098	29,230
	Sales	55,307	43,227	34,774
	Net assets	75,833	65,369	55,098
R^2	Employment	0.9531	0.9037	0.7489
	Sales	0.8846	0.8387	0.7122
	Net assets	0.9188	0.8675	0.7258

Note: Estimated equation is (4). Standard errors are corrected for heteroskedasticity. Constant, three accounting year dummies and nine SIC Division dummies included but not reported.

Table 6 Galtonian regressions on sub-samples: employment at *t* on employment at *t* - 4

					Independent companies		
Employment Range		b^D	*s.e.*	b^R	*Compro--mise*	*N*	R^2
<=4	1	0.4372	0.0531	12.4477	2.3329	2,534	0.1032
<=8	2	0.5855	0.0261	5.2986	1.7614	5,562	0.1641
<=16	3	0.6799	0.0147	2.8162	1.3837	10,084	0.2956
<=32	4	0.7378	0.0096	1.9549	1.2009	15,155	0.4311
<=64	5	0.7776	0.0070	1.5859	1.1105	20,146	0.5443
<=128	6	0.7872	0.0056	1.4081	1.0529	24,481	0.6168
<=256	7	0.7964	0.0050	1.3288	1.0287	26,794	0.6547
<=512	8	0.8026	0.0048	1.2860	1.0160	27,955	0.6773
<=1024	9	0.8061	0.0047	1.2590	1.0074	28,488	0.6918
TOTAL	10	0.8372	0.0043	1.1856	0.9963	29,230	0.7489
>4	11	0.8629	0.0046	1.2465	1.0371	26,696	0.7305
>8	12	0.8746	0.0050	1.3018	1.0671	23,668	0.7060
>16	13	0.8888	0.0060	1.4021	1.1163	19,146	0.6637
>32	14	0.9069	0.0074	1.5351	1.1799	14,075	0.6169
>64	15	0.9410	0.0089	1.6452	1.2443	9,084	0.5966
>128	16	0.9763	0.0114	1.7702	1.3146	4,749	0.5747
>256	17	1.0182	0.0148	1.8743	1.3814	2,436	0.5692
>512	18	1.0753	0.0199	1.8999	1.4293	1,275	0.5890
>1024	19	1.1027	0.0278	1.8878	1.4428	742	0.6021

Note b^D is OLS estimate of β from direct regression of equation (4); b^R is inverse of OLS estimate of α from reverse regression (4′). Compromise estimate is geometric mean of these. Standard errors are heteroskedasticity corrected. Constant, three accounting year dummies and nine SIC Division dummies included but not reported.

Table 7 Galtonian regressions on sub-samples: sales at *t* on sales at *t* - 4

					Independent companies		
Sales Range, £k		b^D	*s.e.*	b^R	*Compro--mise*	*N*	R^2
<=100	1	0.5146	0.0547	5.8475	1.7347	1,687	0.2089
<=200	2	0.6117	0.0364	3.9271	1.5499	3,165	0.2625
<=400	3	0.6911	0.0245	2.8367	1.4002	5,941	0.3348
<=800	4	0.7322	0.0171	2.3493	1.3115	12,668	0.4086
<=1600	5	0.7584	0.0128	2.0399	1.2438	19,386	0.4618
<=3200	6	0.7807	0.0097	1.7647	1.1738	24,909	0.5253
<=6400	7	0.7979	0.0078	1.5778	1.1220	28,925	0.5795
<=12800	8	0.8035	0.0066	1.4630	1.0842	31,527	0.6170
<=25600	9	0.8104	0.0058	1.3883	1.0607	33,071	0.6467
<=51200	10	0.8124	0.0055	1.3474	1.0463	33,817	0.6622
<=102400	11	0.8118	0.0054	1.3239	1.0367	34,208	0.6707
TOTAL	12	0.8349	0.0049	1.2588	1.0252	34,774	0.7122
>100	13	0.8561	0.0045	1.3096	1.0589	33,087	0.6941
>200	14	0.8597	0.0047	1.3266	1.0679	31,609	0.6823
>400	15	0.8670	0.0050	1.3513	1.0824	28,833	0.6682
>800	16	0.8818	0.0062	1.4321	1.1238	22,106	0.6390
>1600	17	0.8897	0.0079	1.5482	1.1736	15,388	0.6009
>3200	18	0.9112	0.0102	1.7037	1.2460	9,865	0.5637
>6400	19	0.9532	0.0131	1.8989	1.3454	5,849	0.5398
>12800	20	0.9922	0.0161	2.0514	1.4267	3,247	0.5266
>25600	21	1.0684	0.0205	2.2446	1.5486	1,703	0.5212
>51200	22	1.1701	0.0282	2.3614	1.6623	957	0.5355
>102400	23	1.1574	0.0361	2.3005	1.6317	566	0.5506

Note b^D is OLS estimate of β from direct regression of equation (4); b^R is inverse of OLS estimate of α from reverse regression (4′). Compromise estimate is geometric mean of these. Standard errors are heteroskedasticity corrected. Constant, three accounting year dummies and nine SIC Division dummies included but not reported.

Table 8 Galtonian regressions on sub-samples: net assets at t on net assets at t - 4

Independent companies

Net Assets Range, £k		b^D	*s.e.*	b^R	*Compro- -mise*	*N*	R^2
<=100	1	0.5810	0.0118	3.0583	1.3330	17,451	0.2045
<=200	2	0.6718	0.0084	2.2142	1.2196	26,219	0.3182
<=400	3	0.7315	0.0063	1.7837	1.1423	35,047	0.4269
<=800	4	0.7648	0.0051	1.5474	1.0879	42,505	0.5157
<=1600	5	0.7845	0.0043	1.4215	1.0560	47,578	0.5785
<=3200	6	0.7944	0.0039	1.3496	1.0354	50,679	0.6172
<=6400	7	0.8000	0.0037	1.3090	1.0233	52,419	0.6414
<=12800	8	0.8034	0.0036	1.2776	1.0131	53,493	0.6599
<=25600	9	0.8065	0.0035	1.2585	1.0075	54,060	0.6726
<=51200	10	0.8087	0.0034	1.2422	1.0023	54,428	0.6832
<=102400	11	0.8124	0.0034	1.2309	1.0000	54,650	0.6926
TOTAL	12	0.8299	0.0033	1.1965	0.9965	55,098	0.7258
>100	13	0.8867	0.0043	1.3441	1.0917	37,647	0.6912
>200	14	0.8931	0.0052	1.4040	1.1198	28,879	0.6679
>400	15	0.9050	0.0063	1.4749	1.1553	20,051	0.6447
>800	16	0.9211	0.0079	1.5744	1.2042	12,593	0.6193
>1600	17	0.9526	0.0097	1.7040	1.2740	7,520	0.5969
>3200	18	0.9913	0.0123	1.8127	1.3405	4,419	0.5872
>6400	19	1.0367	0.0152	1.8521	1.3856	2,679	0.5994
>12800	20	1.0699	0.0198	1.8650	1.4125	1,605	0.6156
>25600	21	1.0892	0.0241	1.8460	1.4180	1,038	0.6345
>51200	22	1.0744	0.0318	1.8291	1.4019	670	0.6361
>102400	23	1.0818	0.0410	1.8247	1.4050	448	0.6409

Note b^D is OLS estimate of β from direct regression of equation (4); b^R is inverse of OLS estimate of α from reverse regression (4'). Compromise estimate is geometric mean of these. Standard errors are heteroskedasticity corrected. Constant, three accounting year dummies and nine SIC Division dummies included but not reported.

Table 9 Geometric mean size in year t, conditional on size in year t-4

(a) Employment

Size range in year t-4 (no.)	*Year t*	*Year t-4*	*Number of companies*
<=1	2.9	1.0	190
>1 & <=2	3.3	2.0	637
>2 & <=4	4.8	3.4	1,707
>4 & <=8	7.4	6.2	3,028
>8 & <=16	12.6	11.7	4,522
>16 & <=32	22.3	22.6	5,071
>32 & <=64	42.1	45.3	4,991
>64 & <=128	72.8	87.9	4,335
>128 & <=256	132.5	173.6	2,313
>256 & <=512	244.7	349.5	1,161
>512 & <=1024	444.0	695.7	533
>1024 & <=2048	853.1	1,381.2	313
>2048 & <=4096	2,154.2	2,793.2	162
>4096 & <=8192	4,083.3	5,587.4	104
>8192	21,624.7	23,761.9	163
TOTAL	31.6	32.4	29,230

(b) Sales

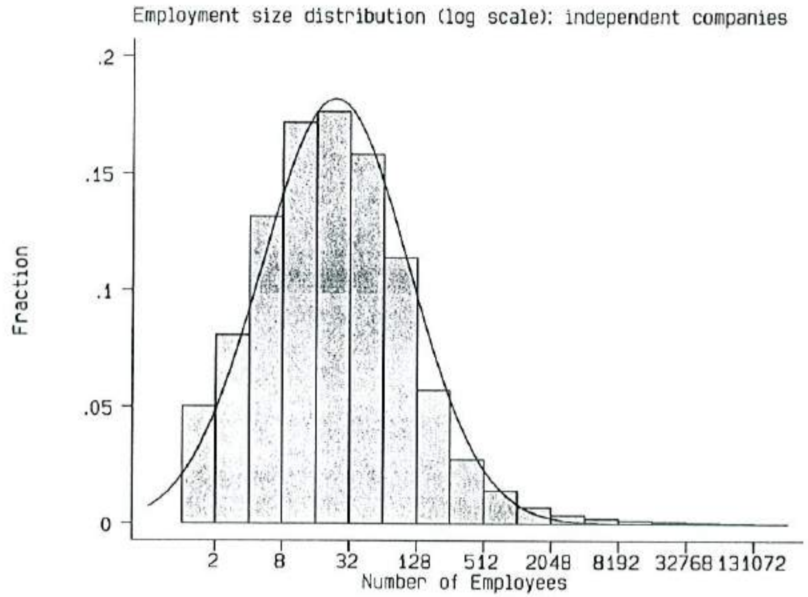
Size range in year t - 4 (£k)	*Year t*	*Year t-4*	*Number of companies*
<=12.5	27.2	5.3	270
>12.5 & <=25	37.6	18.4	226
>25 & <=50	68.7	36.5	405
>50 & <=100	105.1	75.6	786
>100 & <=200	201.6	147.6	1,478
>200 & <=400	390.1	298.7	2,776
>400 & <=800	728.6	583.3	6,727
>800 & <=1600	1,262.3	1,123.4	6,718
>1600 & <=3200	2,345.5	2,244.6	5,523
>3200 & <=6400	4,284.4	4,432.1	4,016
>6400 & <=12800	7,440.7	8,846.7	2,602
>12800 & <=25600	13,998.7	17,433.0	1,544
>25600 & <=51200	24,213.0	34,875.9	746
>51200 & <=102400	39,356.6	69,211.1	391
>102400 & <=204800	92,125.7	137,334.4	235
>204800 & <=409600	185,521.5	271,721.2	121
>409600	1,588,479.8	1,427,725.2	210
TOTAL	1,620.6	1,473.7	34,774

Table 9 (cont.) **Geometric mean size in year *t*, conditional on size in year *t*-4**

(c) Net assets

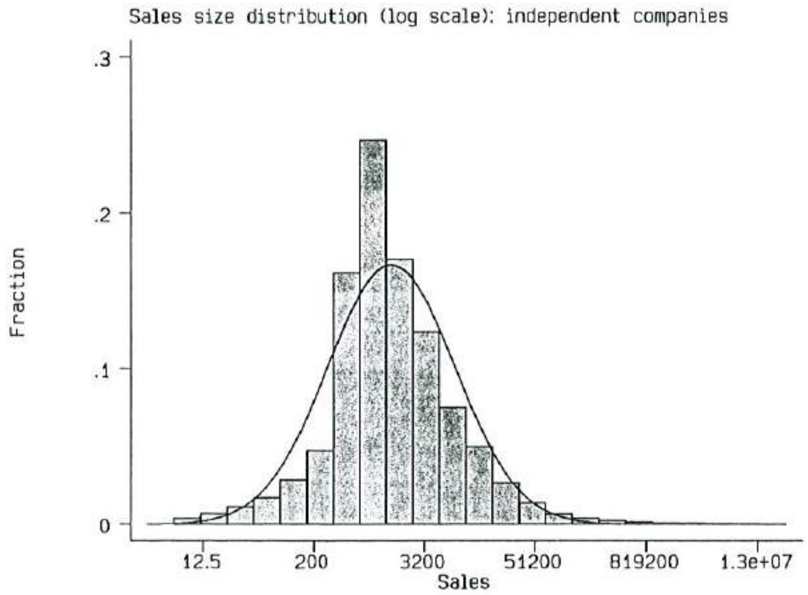
Size range in year t - 4 (£k)	*Year t*	*Year t-4*	*Number of companies*
<=12.5	21.4	5.1	2,551
>12.5 & <=25	38.6	18.6	2,829
>25 & <=50	61.9	37.0	4,806
>50 & <=100	106.7	73.0	7,265
>100 & <=200	198.8	143.3	8,768
>200 & <=400	366.2	282.5	8,828
>400 & <=800	668.5	559.3	7,458
>800 & <=1600	1,231.5	1,102.2	5,073
>1600 & <=3200	2,174.0	2,184.8	3,101
>3200 & <=6400	3,953.6	4,391.3	1,740
>6400 & <=12800	7,066.0	8,811.6	1,074
>12800 & <=25600	13,367.8	17,553.4	567
>25600 & <=51200	24,362.6	34,780.4	368
>51200 & <=102400	62,769.8	73,132.3	222
>102400 & <=204800	132,572.4	137,627.1	143
>204800 & <=409600	249,586.8	287,833.4	109
>409600	1,545,552.5	1,322,061.7	196
TOTAL	334.0	244.3	55,098

FIGURE 1



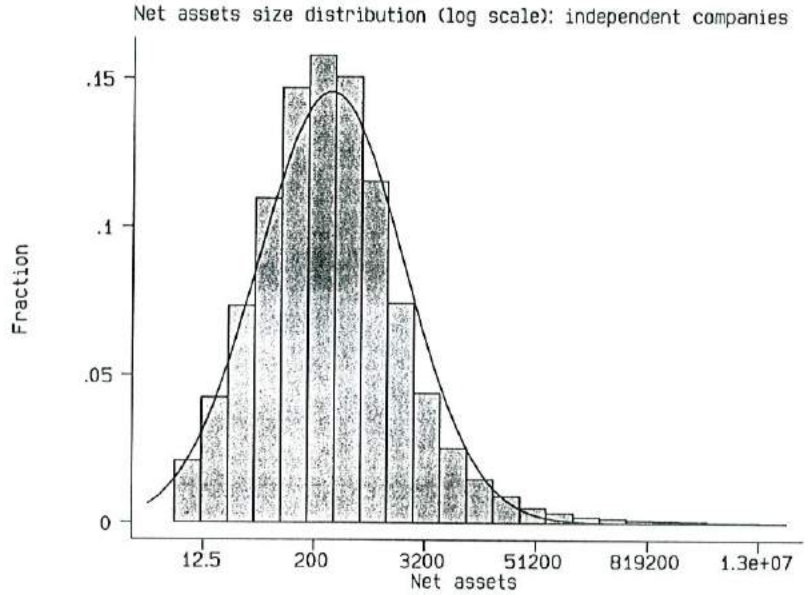
Source Table 1.

FIGURE 2



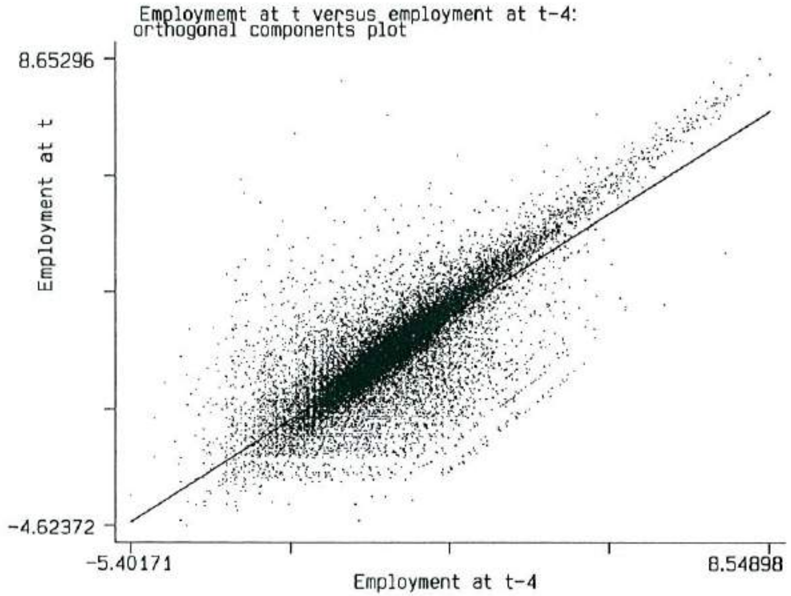
Source Table 2.

FIGURE 3



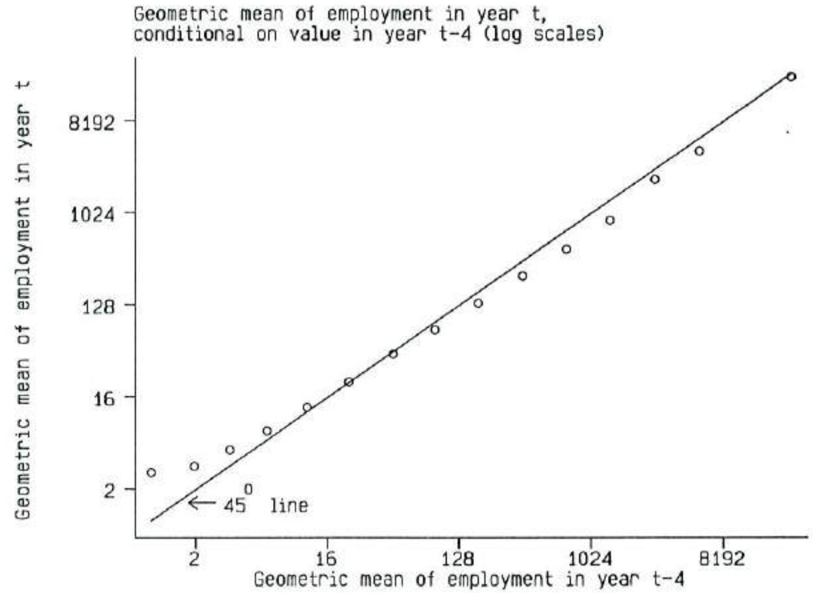
Source Table 3.

FIGURE 4



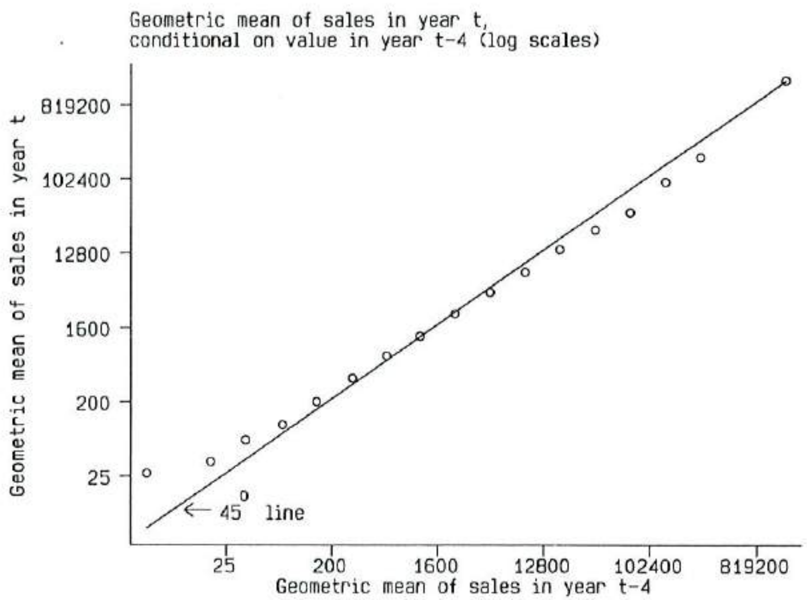
Source OLS estimate of equation (4) plus dummies; 29,230 independent companies.

FIGURE 5



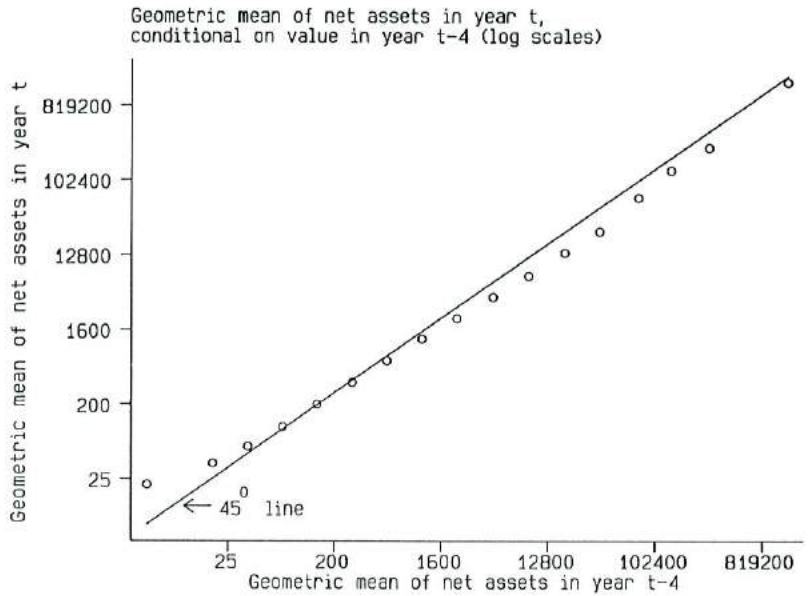
Source Table 9(a). The 45 degree line is drawn through the point of means.

FIGURE 6



Source Table 9(b). The 45 degree line is drawn through the point of means.

FIGURE 7



Source Table 9(c). The 45 degree line is drawn through the point of means.

APPENDIX A

In this appendix, we provide some further details on our sample. First, we look at whether our results are affected by disaggregating by SIC Division. We have repeated our basic regression (4) for each of the 10 Divisions of the 1980 SIC and for each of our three measures of size. The results are in Table A1. Though there is some variation in the estimate of β, particularly in Divisions with relatively small sample sizes, on the whole the estimates are remarkably similar. Thus there is no evidence that our conclusions about the effect of size on growth are influenced by industrial composition effects.

Secondly, we give some descriptive statistics for the sub-sample of companies for which data are available for 1, 2 and 4 years earlier, that is, those surviving for at least 4 years (Table A2). From this table we can see that compared with 4 years earlier, the geometric mean of employment fell by 1.8 per cent amongst these survivors. The growth rates of sales and net assets are less interesting since these are in nominal terms.

Table A1 **Galtonian regressions, by SIC(1980) Division: growth over 4 years (independent companies)**

SIC (1980) Division	*b*	*s.e.*	*N*	R^2
(a) Employment				
0 Agriculture	0.9138	0.0243	668	0.7933
1 Energy & water	0.7719	0.0547	106	0.8071
2 Metals & chemicals	0.9185	0.0166	817	0.8806
3 Engineering & vehicles	0.8654	0.0108	3,270	0.7968
4 Other manufacturing	0.8526	0.0109	3,877	0.7798
5 Construction	0.8496	0.0135	2,656	0.7502
6 Distribution & hotels	0.8439	0.0075	9,977	0.7381
7 Transport & comms.	0.8683	0.0178	1,509	0.7296
8 Banking & finance	0.7736	0.0122	4,493	0.6156
9 Other services	0.8004	0.0179	1,857	0.6933
All (SIC 0-9)	0.8372	0.0043	29,230	0.7489
(b) Sales				
0 Agriculture	0.8240	0.0348	757	0.6892
1 Energy & water	0.7050	0.0730	120	0.6756
2 Metals & chemicals	0.9284	0.0201	813	0.8696
3 Engineering & vehicles	0.9120	0.0114	3,404	0.8050
4 Other manufacturing	0.8819	0.0127	4,035	0.7854
5 Construction	0.8111	0.0166	3,439	0.6678
6 Distribution & hotels	0.8443	0.0084	11,986	0.7125
7 Transport & comms.	0.8396	0.0223	1,886	0.6916
8 Banking & finance	0.7905	0.0118	6,078	0.6319
9 Other services	0.7324	0.0235	2,256	0.6451
All (SIC 0-9)	0.8349	0.0049	34,774	0.7122
(c) Net assets				
0 Agriculture	0.8258	0.0258	1,167	0.7133
1 Energy & water	0.8781	0.0591	149	0.8307
2 Metals & chemicals	0.9029	0.0143	1,513	0.8355
3 Engineering & vehicles	0.8515	0.0086	6,702	0.7650
4 Other manufacturing	0.8603	0.0090	7,116	0.7587
5 Construction	0.7808	0.0109	5,615	0.6585
6 Distribution & hotels	0.8047	0.0060	18,844	0.6799
7 Transport & comms.	0.8479	0.0144	2,945	0.7005
8 Banking & finance	0.8390	0.0071	8,456	0.7278
9 Other services	0.8010	0.0152	2,591	0.6855
All (SIC 0-9)	0.8299	0.0033	55,098	0.7258

Note Equation fitted is (4), with either employment, sales or net assets as the dependent variable. Constant and accounting year dummies included, but not reported. Standard errors are corrected for heteroskedasticity.

Table A2 **Summary statistics for 3 measures of size: current versus 1, 2 and 4 years ago**

	Log Employment (N=28,539)		*Log Sales (N=34,022)*		*Log Net assets (N=54,039)*	
	Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*
Current	3.4740	1.5817	7.4089	1.7872	5.8557	1.8872
1 year earlier	3.5117	1.5658	7.4426	1.7343	5.8524	1.8474
2 years earlier	3.5384	1.5597	7.4602	1.7031	5.8042	1.8373
4 years earlier	3.4922	1.5616	7.3075	1.7210	5.5321	1.9015

Note: For each size measure, the sample includes only those companies for which values are available in all 4 years.

52 Economic performance and education: the nature of Britain's deficiencies, S.J. Prais
53 Smoothing the transition to skilled employment: How can school-based vocational guidance be improved?, Valerie Jarvis
54 Expectational errors and bankruptcy, Garry Young
55 Making decisions about education and training: a note on practice and procedures in careers guidance in France and Germany, Hilary Steedman
56 Industrial investment and economic policy, David G. Mayes and Garry Young
57 Did the retreat of UK trade unionism accelerate during the 1990-3 recession?, Paul Geroski, Paul Gregg and Thibaut Desjonqueres
58 When the time was right? The UK experience of the ERM, Ray Barrell, Andrew Britton and Nigel Pain
59 Mortgage equity withdrawal; causes and consequences, Peter Westaway
60 Cointegration and forecast evaluation: some lessons from National Institute forecasts, Nigel Pain
61 The role of soft law in the evolution of rules for a Single European Market: the case of retailing, Duncan Matthews and David G. Mayes
62 Increasing returns and externalities in UK manufacturing: myth or reality?, Nicholas Oulton
63 Small samples and structural change: a simulation study of consumption, Andrew P. Blake and Peter F. Westaway
64 The Phillips curve in empirical macro-models of the World Economy, James Sefton and Stephen Wright
65 Solvency and cycles (or Lord make us good, but not yet), Ray Barrell
66 The influence of foreign factor prices and international taxation on fixed investment in the UK, Garry Young
67 Fiscal solvency and fiscal policy, Ray Barrell, James Sefton and Jan in't Veld
68 Full employment in a market economy, Andrew Britton
69 How effective are state employment agencies? Jobcentre use and job matching in Britain, Paul Gregg and Jonathan Wadsworth
70 Targetting inflation with nominal interest rates, Andrew P. Blake and Peter F. Westaway
71 Investigating structural changes in UK export performance: the role of innovation and direct investment, Andrew P. Blake and Nigel Pain
72 More work in fewer households? Paul Gregg and Jonathan Wadsworth
73 The net national product and exhaustible resources: the effects of foreign trade, J.A. Sefton and M.R. Weale
74 Policy regimes and the persistence of wage inflation and unemployment, Robert Anderton
75 Schooling as preparation for life and work, Helvia Bierhoff and S.J. Prais
76 The persistence of UK employment and unemployment, Robert Anderton and Soteris Soteri