

NBER WORKING PAPER SERIES

THE RELATIONSHIP BETWEEN FIRM
SIZE AND FIRM GROWTH IN THE
U.S. MANUFACTURING CENTER

Bronwyn H. Hall

Working Paper No. 1965

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 1986

The research reported here is part of the NBER's research program in Productivity. Any opinions expressed are those of the author and not those of the National Bureau of Economic Research.

The Relationship Between Firm Size and Firm Growth
in the U.S. Manufacturing Sector

ABSTRACT

This paper investigates the dynamics of firm growth in the U. S. manufacturing sector in the recent past. I use panel data on the publicly traded firms in the U. S. manufacturing sector: from a universe of approximately 1800 firms in 1976, I am able to follow most of them for at least three years, and over half of them from 1972 until 1983. I consider several problems, both econometric and substantive, which exist in analyzing this kind of data: the choice of size measure, the role of measurement error, and the effect of selection (attrition) on estimates obtained from this sample.

Using time series methods, suitably modified for panel data (where the number of time periods per observational unit is small), I analyze the behavior of employment over time and find that most of the change in employment in any given year is permanent in the sense that there is no tendency to return to the previous level. Year-to-year growth rates are largely uncorrelated and there is almost no role for measurement error. I find that Gibrat's Law is weakly rejected for the smaller firms in my sample and accepted for the larger firms. Other measures of size produce essentially the same results.

Correction for attrition from the sample changes the results somewhat: I use a simple model in which firms leave the sample because they are small and/or undervalued (since many exits are acquisitions) and find that Tobin's Q, the ratio of market valuation to the value of the underlying assets of the firm, is a much better predictor of exit probability than size alone (firms with low Q are more likely to exit the sample). When I use this estimate of the probability of exit to control for selection bias, Gibrat's Law is weakly rejected for firms of all sizes and there are significant positive effects on firm growth from both investment in physical capital and R&D expenditures, with R&D having a somewhat higher net effect.

Bronwyn H. Hall
National Bureau of
Economic Research
204 Junipero Serra Blvd.
Stanford, CA 94305

April 1986

THE RELATIONSHIP BETWEEN FIRM SIZE AND FIRM GROWTH
IN THE U.S. MANUFACTURING SECTOR

Bronwyn H. Hall¹

1. Introduction

The present paper is a first step in an investigation of the dynamics of firm growth in the U.S. manufacturing sector during the recent past. It updates work by earlier researchers on the relationship between firm size and growth using a more comprehensive dataset and modern econometric techniques to attempt to correct for some of the problems in estimating such a relationship. The ultimate conclusion is that the previously observed negative relationship between size and growth for smaller firms is robust to corrections for selection bias and heteroskedasticity, although this conclusion is clouded by the difficulties of separating nonlinearity from selection bias in the presence of size-related heteroskedasticity.

1. Stanford University and the National Bureau of Economic Research. I am indebted to Zvi Griliches for numerous helpful discussions and to Tim Bresnahan, Peter Reiss, Sherwin Rosen, David Evans, Tom MaCurdy, and members of the Stanford University Industrial Organization seminar for comments. Joy Mundy provided extremely able research assistance and Clint Cummins programmed the sample selection models in TSP. Parts of this research were supported by a National Science Foundation Grant (PRA 81-08635) and by the National Bureau of Economic Research.

I focus on the manufacturing sector in this study because my ultimate interest is in exploring the role of innovative activity in the growth of firms and my main indicators of such activity, Research and Development expenditures and patent applications, are primarily available for that sector (and, in fact, primarily done in that sector). However, in the modern U.S. economy, with the increasing importance of the computer-related service sector in innovation, this is a limitation of the study and should be kept in mind.

The dynamics of firm growth is an interesting and important topic for two reasons: 1) the growth of firms is a main ingredient in economic growth and job creation, and 2) the actual dynamics has an impact on the consequences of industrial concentration. The extent and speed at which smaller firms enter the market and grow successfully is an important check on the development of monopoly power by the large firms in the economy. The role of innovation in this process was stressed by Schumpeter and the Schumpeterian hypothesis remains to this day a controversial interpretation of the growth process in modern industrial economies.

In this paper I try to answer a version of the question "Do small firms grow faster?" and to reconcile my results with those of previous researchers. My version of the question might be more properly phrased as "Do small to medium-sized publicly traded manufacturing firms grow faster than large ones?" If they do, is it because of the way they are selected into our sample, or because of a difference in the rate and direction of innovative activity, or simply because the economy is finite and diminishing returns sets in eventually. I do not claim to be able to distinguish among all these alternatives completely, or even that only one must be true, but I will explore the implications of each

for the observed data.

Stochastic models of firm growth have been subjected to two kinds of empirical tests: the first posits a growth model which is stationary over time and then looks at the implications of this model for the equilibrium size distribution of firms. Various authors, beginning with Gibrat, have shown that the simplest version of a diffusion model, in which growth rates are independent of size, generates a log normal size distribution, albeit with an increasing variance over time. Mandelbrot (1963) provides a survey of this and other models in which he shows the conditions under which the equilibrium size distribution is a stable Pareto distribution. Boundary conditions on exit and entry are required in order to achieve a stable distribution in most cases. This has been investigated empirically by Simon and coworkers (e.g., Simon and Bonini 1958), Quandt 1966, and Hart and Prais 1956. Typically the size distribution conforms fairly well to log normal, with possibly some skewness to the right. The power of this kind of test is low, since the relationship of growth rates to size is not explicitly investigated. However, several of the existing theories, such as those of Lucas 1978 and the stochastic theory of Simon and Bonini have as their main implication these static distributions.

Other theoretical models of firm growth, such as work by Lucas (1978), Nelson and Winter (1982), Jovanovic (1982), and others, have more specific implications for the actual dynamics of individual firm growth. These models emphasize the role of the U-shaped cost curve and learning in the evolution of the firm size distribution. The empirical work in this area investigates the relationship of growth rates and size in a panel of firms. This work is exemplified by

Pashigian and Hymer and Mansfield in the sixties, and the more recent work by Birch, Armington and Odle, and Evans using the Dun and Bradstreet files (as cleaned by Brookings Institution for the Small Business Administration) and by Evans using Fortune 500 firms. Except for Evans, none of these researchers attempted to correct econometrically for biases induced by the selection into the sample and measurement error. One of the purposes of this paper is to investigate whether such biases have an appreciable effect on the results.

The first problem is the regression to the mean phenomenon: if the dependent variable in question is the growth rate measured as size in a final period less size in an initial period, and the independent variable is size in the initial period, measured with error, then firms which have transitorily low size due to measurement error will on average seem to grow faster than those with transitorily high size, assuming that all firms have the same growth rate. Using yearly observations from a panel of firms it is possible to control for this kind of random measurement error with instrumental variable techniques (see Griliches and Hausman 1985).

The second problem is probably somewhat more serious: measuring a growth rate from a panel requires that data on size be available for every firm in both the beginning and the end period. But small firms which have slow or negative growth are more likely to disappear from the sample than large firms, leading to another example of the well-known problem of sample selection bias. In addition, some of the most rapid-growing and successful small firms may not be present at the beginning of the period, which will produce biases in the other direction. In section 4 of the paper we present estimates of a simple model which attempts to control for this bias.

The plan of the paper is as follows: first I describe the data and present preliminary results on the role of measurement error in the size-growth relationship. This is followed by exploration of the time series behavior of employment growth, setting aside the issue of selection bias temporarily. Then I develop an econometric model of sample attrition and discuss the problems which arise in estimating such a model in the presence of heteroskedasticity and the absence of adequate instruments to separately identify the probability of firm survival. Finally I investigate the relationship of investment, both in physical capital and research and development, to firm growth, using the sample selection model I have developed to control for attrition. The paper concludes with some suggestions for further work.

2. Description of the Data

In this paper I confine my analysis to that part of the manufacturing sector which consists of publicly traded firms, since our sample is drawn from the Compustat files. This covers approximately ninety percent of the employment in the manufacturing sector in 1976, although only about one percent of the firms.² Thus the study is really about the relationship of growth and size across firms which have already reached a certain minimum size, large enough to require outside capitalization. We would argue that these are the firms of interest,

2. The total number of employees in our 1976 cross section is 16.7 million, reported on an enterprise basis. This figure may include some foreign employees. The total manufacturing employment reported by the Bureau of Labor Statistics for the same year is 19 million, collected on an establishment basis, and does not include foreign employees. The number of enterprises in the Census of Manufacturers in 1977 is approximately 300,000.

since the impact of employment growth by the smallest firms on the whole economy is likely to be negligible until they reach a certain size. It should be emphasized that this argument applies mainly to the manufacturing sector, where there are almost no privately held firms of any size.

The universe from which I draw my sample consists of 1778 firms in the manufacturing sector in 1976³ (see Bound et al 1984 for further description of this data). I considered two different panels selected from this universe: all the firms with employment data from 1972 through 1979, and all the firms with employment data from 1976 to 1983. The first set maximizes the data availability, since the basic universe of firms is as of 1976, while the second has the advantage that it begins in the year in which the sample was chosen and hence suffers from selection in only one direction. There are 1349 firms in the first sample and 1098 in the second; 962 of these firms are in both samples. The remainder of the firms either enter the sample during the period, or exit from the sample. A few merge with another firm so large that they become in effect a new firm; these observations are treated as exits

3. An earlier draft of this paper was based on a sample of 2577 firms, which included in addition all the firms on the Compustat full coverage file which were in the manufacturing sector. Further investigation of the full coverage sample has revealed that it is unsuitable for a study of growth and sample selection due to exit: many of the firms on the file are not really publicly traded and most do not have valuation information. In addition, many of them file for a year or two after a public offering under rule 15(d) of the 1933 Securities and Exchange Act, as amended in 1964, and then suspend filing because there are less than 300 shareholders of record. This is not really an exit, but we no longer can obtain data on the firm.

(and the new entity as an entry).⁴ Table 1 shows the industrial breakdown and average growth rates for the two samples. The overall growth rate of employment in this sample was about 2.9 percent in the first period and 0.8 percent in the second. There are substantial differences across the industries, with the so-called "high tech" industries (drugs, computing, equipment, communication equipment, and scientific instruments) typically growing more rapidly throughout both periods.

In the same table I show the average R&D to sales ratio and gross investment to sales ratio in 1976, in order to indicate the variability in the key "engines of growth" across industries. As expected, drugs, computers, communication equipment, and scientific instruments have much higher R&D to sales ratios than the other industries. The industries with high investment rates are chemicals, petroleum refining, computers, and the lumber and paper industry.

We first consider the possible role of measurement error in biasing a regression of changes on levels. A simple model of Markov growth with errors in variables would look like

$$E_t = X_t + w_t$$

$$X_t = X_{t-1} + u_t$$

with w and u uncorrelated white noise errors, X unobserved ("true"

4. This is obviously an inadequate treatment of an interesting aspect of growth in the manufacturing sector, but it involves a relatively small number of firms, and it is beyond to scope of the present paper to model major merger activity. We hope to explore the extent to which this kind of growth impacts on our estimates in future work.

employment), and E observed. Under this model, the true relationship between the change in employment and its level is

$$E (dX_t | X_{t-1}) = 0$$

but the estimated relationship will be

$$E (dE_t | E_{t-1}) = -(s_w / (s_x + s_w)) E_{t-1}$$

where s_w and s_x are variances of the error term and the unobservable respectively. For these data, the within to total variance ratio for log employment is approximately six percent. Under the simple Markov model presented above, this variance is the sum of twice the measurement error plus the variance of the disturbance u , so the largest negative value we would expect for this coefficient is -0.03 (divided by the number of years over which the employment change is computed).

In Table 2, I present the results of a simple regression of growth on size for our two different subperiods, with and without individual industry effects. The coefficient of $\log E_{72}$ in a regression of the annual growth rate from 72 to 79 is -1.14% , while that of $\log E_{76}$ in the growth rate is -1.06% . That is, doubling a firm's size decreases its annual growth rate by about eight tenths of one percent. In the remainder of the table I try to correct for possible measurement error bias in this relationship, but it remains remarkably stable. First, the pure random walk with measurement error model would predict an estimated coefficient of zero in the regression of the growth rate on size in the period preceding that from which the growth rate is measured, whereas in column 2 we obtain an estimate slightly smaller in absolute value as in the previous regression. Second, in the last column we regress the growth rate on size at the beginning of the period using size one year prior to the beginning of the period as an instrument, since by

assumption it is uncorrelated with the measurement error, but this regression yields almost the same result as the ordinary least squares estimate in column 3. In fact, if there is any bias at all, the correction has the wrong sign. This impression is confirmed by comparing regressions of one through seven year growth rates on size at the beginning of the period (not shown). Under the simple measurement error model presented above, the coefficient of size in such a regression is attenuated by the number of years over which the growth rate is computed, whereas I find only a slight fall in absolute value in the coefficient in going from one to seven years (about .05 per year).

The addition of industry dummies do not change the coefficients much, although they are significant at conventional levels (the 1% critical level is about 2.2). Since this result held for most of the results reported in this paper, we have not presented estimates with industry dummies in the rest of the tables; they almost invariably were moderately significant but had little or no effect on the other coefficients. The study of interindustry differences in these data appears to be warranted but is beyond the scope of the paper.

The tentative conclusion is that uncorrelated errors of measurement in employment cannot be responsible for more than about ten percent of the observed negative relationship between size and growth. That is, the noise to signal ratio for the employment variable is about one-tenth of a percent in levels and about ten percent in first differences. However, we should note that this does correspond to a standard deviation of about fifteen percent of the level of employment in any one year, a not inconsiderable amount. It is simply that a measurement error of this magnitude is swamped by the large variance in size across our firm population and introduces very little bias in the estimating

equation.

Repeating the exercise using sales produced much the same result in the growth-size regressions, although the standard deviation of the measurement error in this case could be about half again as large. For the instrumental variable estimates corresponding to column 4 of Table 2, the size coefficient was -0.90 (.21).

3. The Time Series Behavior of Employment Growth

In Appendix A I present the results of a time series analysis of the three sets of firms drawn from my sample: 1) 1349 firms from 1972 to 1979, 2) 1098 firms from 1976 to 1983, and 3) 962 firms from 1972 to 1983. Obviously, there is substantial overlap in these samples; the earlier period was chosen to maximize the number of firms in a balanced panel, while the latter has the feature that it begins in the year that our universe was chosen, so that selection only goes in one direction. I use all three samples in order to obtain some information about parameter stability. In this section I interpret these time series results in the context of several hypotheses of interest.

The evidence of the regressions in Table 2 suggested that the simple measurement error model I was considering be modified to include an autoregressive component. This expanded model can be written as

$$(1) \quad \begin{array}{ll} X_t = \beta X_{t-1} + u_t & Eu_t^2 = \sigma_u^2 \quad EX_{t-1}u_t = 0 \\ y_t = X_t + w_t & Ew_t^2 = \sigma_w^2 \quad Eu_s w_t = 0 \quad \forall s, t \end{array}$$

I have allowed the variance of employment growth to vary from year to year since this is a strong implication of the patterns observed in

the covariance matrix of differences. The model is above equivalent to a standard ARMA(1,1) model, but the latter is valid over a larger parameter space; this turns out to be important. The ARMA(1,1) model is written as

$$(2) \quad (1-\alpha L) y_t = (1-\mu L) \varepsilon_t \quad \varepsilon_t \text{ white noise}$$

whereas the AR(1) model with measurement error was

$$(3) \quad (1-\beta L) y_t = u_t + (1-\beta L)w_t = u_t + w_t - \beta w_{t-1}$$

If the disturbances are normally distributed, it is easy to show that the two models are equivalent with

$$\begin{aligned} \alpha &= \beta \\ (1+\mu^2)\sigma_\varepsilon^2 &= \sigma_u^2 + \sigma_w^2(1+\beta^2) \\ \sigma_w^2 &= (\mu/\beta)\sigma_\varepsilon^2 \end{aligned}$$

However, the measurement error model requires that σ_w^2 be positive, which imposes the constraint that μ and α are of same sign in the ARMA(1,1) model and restricts the parameter space. It turns out that when I estimate the ARMA(1,1) model using these data that the constraint is not satisfied, which implies a slightly negative σ_w^2 . The estimates are shown in the top half of Table 3. They are quite stable across the periods and are consistent with the IV estimates in Table 2, since they imply a coefficient of 100 $(\beta-1) = 1.0$ percent in the growth rate equation together with a slightly positive measurement error bias (of the order of 0.001, or one tenth of a percent).

Since the time series analysis in the appendix shows that an ARMA(2,1) model fits the data significantly better than an ARMA(1,1)

model, we also choose to explore what Leonard calls a "mean-reverting" model, which is familiar from the investment literature as a flexible accelerator model. This model rests on the idea that the number of employees is a kind of stock which is not instantaneously adjustable at zero cost to the firm. For a constant returns to scale firm which has quadratic adjustment costs, there is a linear relationship between employment changes and the current and desired levels of employment:

$$\Delta y_t = (1-\lambda)(y_t^* - y_{t-1}) + w_t$$

In terms of the ARMA model type of formulation, this can be written as

$$(1-\lambda L)y_t = (1-\lambda) y_t^* + w_t$$

The time series process implied by this model depends on what is assumed about the process generating the desired level of employment. If the process is constant for each firm, it becomes

$$(1-\lambda L)y_t = \alpha_i + w_t \quad \text{where } \alpha_i = (1-\lambda)\bar{y}_i$$

Because of the short panel, this cannot be estimated consistently in levels, so I write it in first differences:

$$(1-\lambda L)(1-L)y_t = (1-\lambda)\Delta y_t = (1-L)w_t$$

This is an ARMA(1,1) process with μ constrained to be equal to one.

If the desired level of employment is a Martingale process, which seems somewhat more reasonable, since we might expect that the target size evolves as the firm receives random shocks each year about demand, cost, and so forth, then

$$(1-L)y_t^* = u_t \quad \text{with } Eu_t^2 = \sigma_t^2$$

and the process becomes

$$(1-\lambda L)\Delta y_t = (1-\lambda)u_t + (1-L)w_t$$

which is equivalent to an ARMA(1,1) process with both α and μ free.⁵ Since the estimated μ for this model is not unity, it is easy to reject the first version (constant target size). The estimates for the second version are shown in the bottom of Table 3. They are not very stable across the time periods; in fact, those for the first period not really make much sense in the context of this model since they imply that the firm adjusts its size away for the desired level of employment ($1-\lambda = -0.745$).

The conclusions from this exercise are quite strong, if somewhat negative: 1) the growth of firms as measured by employment is quite random from year to year, with a standard deviation corresponding to about twenty to thirty percent of the level. There is little evidence of a systematic relationship of growth rates across time, at least for

5. This can be shown in the same way we showed the equivalence of (2) and (3). Since the order of the AR part is the same, $\alpha = \lambda$, and we have

$$2\sigma_w^2 + (1-\alpha)^2 \sigma_u^2 = \sigma_\varepsilon^2 (1+\mu^2)$$

$$\sigma_w^2 = \mu\sigma_\varepsilon^2$$

which implies

$$\sigma_u^2 = \sigma_\varepsilon^2 (1-\mu)^2 / (1-\alpha)^2$$

where α , μ , σ_ε^2 are the parameters of the ARMA(1,1) model.

the majority of the firms. 2) Measurement error accounts for very little (about ten percent) of these random year-to-year movements; most of the change represents a permanent change in the level of employment. 3) Small firms do indeed grow faster, but not by much, and not by enough to have a significant effect on the size distribution over a ten year period.

4. Correcting for Sample Attrition

In obtaining the previous time series results, I used a balanced sample of firms, ignoring the possible biases introduced by entry and exit into the sample during the time period. In this section I explore the consequences of sample attrition on the estimates of a growth equation. The initial exploration is reduced form in nature, since it does not use an explicit model for exit and entry, other than the simple fact that most of the action takes place among the smaller firms in the sample. It turns out that the power of this kind of model to control for selection is weak, since the identification comes from the functional form, and I intend in future work to experiment with a more explicit model of bankruptcy and merger. In the final section of this paper, I make a first attempt along these lines.

Any attempt to regress the growth rate of a firm from period to period on its size in the initial period will be subject to biases arising from the selection of the sample. In order to measure the growth rate, we require that data be available for the firm in both the beginning and the ending period. Even if we are able to draw a sample of firms which are representative of the population in the initial period, by the time we reach the final period, the smaller and more slowly growing firms are those most likely to have dropped out of the

sample.

In terms of a sample selection model, let γ_i be the growth rate of the i th firm, and let Y_i and \tilde{Y}_i be the initial and final period logarithm of size. Then the observed growth rate is

$$\Delta Y_i = \tilde{Y}_i - Y_i = \gamma_i + u_i$$

where u_i is an i.i.d. random variable, with $E(Y_i u_i) = E(\gamma_i u_i) = 0$. I hypothesize that the firm will drop out of the sample when its size drops below some cutoff value, denoted as a_i in order to capture the notion that it may vary in some systematic way across firms (for example, by industry). Then the model which I actually estimate becomes

$$\begin{aligned} \Delta Y_i &= \gamma_i + u_i && \text{if } Y_i + u_i + \gamma_i > a_i \\ \Delta Y_i &\text{ not observed} && \text{if } Y_i + u_i + \gamma_i \leq a_i \end{aligned}$$

Under normality of the u_i , for the observed sample the regression equation becomes

$$\begin{aligned} (4) \quad E(\Delta Y_i | \Delta Y_i \text{ observed}) &= E(\gamma_i + u_i | u_i > a_i - Y_i - \gamma_i) \\ &= \gamma_i + \sigma \lambda(Y_i + \gamma_i - a_i) \end{aligned}$$

where $\lambda(\cdot)$ is the inverse Mills ratio, the ratio of the normal density to the cumulative normal. Since $\lambda > 0$, the disturbance for the estimating equation is no longer mean zero, conditional on the firm surviving. Since λ is monotonically decreasing in its argument, the bias goes to zero as $Y_i + \gamma_i - a_i$ becomes large, that is for large firms with positive growth rates and a low dropout size.

The model above is a variant of the standard censored regression model with a stochastic threshold as first presented by Nelson (1977). It assumes we know or can estimate the cutoff a_i for each firm and that

firms never drop out for reasons other than becoming too small. To allow for more ignorance about the reasons for exit, I have chosen instead to use a general sample selection model in which the probability of survival equation is not explicitly linked to a threshold value of the dependent variable. I show in Appendix B that this model is observationally equivalent to the stochastic censoring model of Nelson, but that in return for gaining a slightly more general interpretation of the survival equation, I give up the ability to identify the coefficients of the a_i equation.

With this in mind, the model I use for estimation is a standard generalized Tobit model of the form

$$\begin{aligned}
 y_{1i} &= X_i \beta + v_{1i} && \text{if } y_{2i} > 0 \\
 y_{1i} &\text{ not observed} && \text{if } y_{2i} < 0 \\
 y_{2i} &= Z_i \delta + v_{2i}
 \end{aligned}$$

with a covariance matrix

$$\text{E}v v' = \begin{bmatrix} \sigma_1^2 & \rho \sigma_1 \\ & 1 \end{bmatrix}$$

where I have normalized the residual variance of the unobserved latent variable z_{2i} to be unity. A discussion of this model and its estimation by the method of maximum likelihood is given in Griliches, Hall, and Hausman 1978.

My sample is drawn from the universe of Compustat firms in 1976. Hence there is selection in both directions: small fast growing firms may exist in 1976 but not in 1972 and some firms exit by 1979 or 1983. If each firm is allowed to have its own growth rate, constant over the

period, and the remaining shocks to growth are serially uncorrelated for each firm, I can easily generalize the above model to allow for bias due to sample entry and exit. I write the models for E_{76} and E_{79} as

$$E_{76} = E_{72} + 4 \gamma_i + u_1$$

$$E_{79} = E_{76} + 3 \gamma_i + u_2$$

where γ_i is now the growth rate per annum.⁶ The model for the growth rate from 1976 to 1979 is identical to the one I presented above, so that equation 4 applies.

For the growth rate equation from 1972 to 1976, the relationship is

$$\Delta Y_i = 4 \gamma_i + u_1 \quad \text{if } E_{72} > a_i$$

so that the regression equation becomes

$$\begin{aligned} E(\Delta Y_i | \Delta Y_i \text{ observed}) &= 4 E(\gamma_i | E_{72} > a_i) + E(u_1 | E_{72} > a_i) \\ &= 4 E(\gamma_i | E_{72} > a_i) \end{aligned}$$

Thus there is no selection bias for this equation under the null (when γ_i does not depend on E_{72}). Since the ordinary least squares estimate of the coefficient of E_{72} in a regression of employment growth from 72 to 76 on E_{72} is $-1.34(.16)$, there is reason to suspect that selection bias may be a factor here. However, unlike the previous case, I do not observe the independent variable E_{72} for the firm when the growth rate

6. We are now treating γ_i as a fixed growth rate associated with the firm and have subsumed the stochastic part in the disturbances u . This formulation implies that u_1 and u_2 have different variances, proportional in a known way under a stationary model, but since I do not wish constrain coefficients to be constant over the two subperiods, there is no loss in writing the model this way and treating u_1 and u_2 as independent disturbances with different variances.

from 72 to 76 is not observed, so I need to use a truncated Tobit model (see, for example, Maddala 1983, pp. 176-177). This model can also be estimated by maximum likelihood, but since identification of the parameters of the selection equation come only from the probability term in the denominator of the likelihood function, the estimates may be rather imprecise. In practice, I found this to be the case, so I do not report these estimates here.

In Table 5 I present the results of estimation of the growth rate equations for the two periods first using ordinary least squares and then maximum likelihood of the sample selection models. In neither period was the sample selection correction significant; the estimated ρ is essentially zero and the coefficients do not change. In the case of the first period, the selection model is close to not identified since ρ is zero and the standard errors are not really computable. I should note that this is true even though I have excluded a quadratic size term from the model, so that spurious collinearity with a Mills ratio term is not the problem. In the table I show the χ^2 (4) statistic for the inclusion of quadratic and cubic terms in both equations in the presence of correlated sample selection. There is evidence of nonlinearity in the relationship of growth and size, and in the next section I attempt to disentangle this nonlinearity from size-related heteroskedasticity coupled with sample selection.

I note in passing that adding industry dummies improved the explanatory power of the survival equation (from 76 percent correct to 84 percent correct) but did not change the size coefficient in the growth rate equation very much. The conclusion is that selection bias of this simple kind does not seem to account for the negative

relationship between growth and size.

5. Correcting for Heteroskedasticity

It is well-known that estimates of limited dependent variable models are not robust to departures from normality or heteroskedasticity of the disturbances.⁷ This seems likely to be a problem here from the evidence of the plots in Figures 1 and 2, which suggest that the variance of growth rates is size-related. A simple Lagrange multiplier test due to White (1980) for heteroskedasticity of the disturbance in the regression equation in columns 1 and 2 of Table 5 yields chi-squares of 60.6 and 43.0 respectively with two degrees of freedom.⁸ Also, the heteroskedastic-consistent standard error estimates which I show for the OLS estimates in Table 5 differ from the conventional estimates. (Note that the standard errors for the Probit equation are almost identical, however.)

Of course, the test based on the OLS residuals is not quite appropriate for the sample selection estimates in columns 3 and 4, so I use a simple variation of a test developed by Lee and Maddala (1985) for

7. The tests and corrections applied in this section are designed to correct for heteroskedasticity which is correlated with the regressors since this seemed to be the most serious failure of the homoskedastic normality assumption required for consistency of the maximum likelihood estimates. I also investigated the normality properties of the basic dependent variable, the growth rate, with the following result: skewness did not appear to be a problem, with a coefficient of about -0.1 both for the whole sample and for two size classes. But the coefficient of kurtosis ($\mu_4/\sigma^4 - 3$) was 4.4 for the whole sample, and about three when I standardized the variable by the estimate of σ_1^2 described below. These results seem to suggest that the distribution of the disturbance has somewhat longer tails than would be predicted by the normal distribution, even after heteroskedasticity is corrected for.

8. This test is the TR^2 from a regression of the residuals squared on all the variables in the regression and their cross products.

heteroskedasticity in the Tobit model. In Appendix C I extend their results to the regression equation of the general sample selection model; this yields a simple LM test, which consists of regressing a function of the squared residuals and the estimated correlation coefficient from the sample selection model on the variables of the model. The value of the test statistic when the heteroskedasticity is modelled as a function of size and size squared is shown in columns 3 and 4 of Table 5. The test statistic is almost the same as that given in the first two columns, which is not surprising given the low estimated value of ρ . We would not generally expect the statistics to be the same if ρ were significantly different from zero, however.

If I were willing to maintain that the error in the selection (survival) equation was homoskedastic and normally distributed, it would be possible to compute consistent estimates of the coefficients of the regression equation and their standard errors using the methodology of Heckman. It has been shown by Olsen (1980) that in this case consistent estimates can be computed by including the estimated Mills ratio in the regression and using White's formula for heteroskedastic-consistent standard errors. However, in spite of the fact that the probit disturbance appears to be homoskedastic this assumption seems unwarranted here, since the selection equation itself arises from much the same process as generated the heteroskedastic disturbances in the regression equation. A more promising avenue to explore would be the modelling of the heteroskedasticity in some simple fashion depending on size.

Accordingly, I constructed a simple model for the variance σ_1^2 of the disturbances in the growth equation by regressing the

estimated residuals squared on size and size squared in the initial period. A typical regression of this sort had a negative coefficient on size and a small positive coefficient on size squared. The predicted standard deviation of the growth rate disturbance fell from about 17 percent for small firms to seven percent for the larger firms. I assumed that the heteroskedasticity in the selection equation is proportional to that in the growth equation, and used these estimated σ_i^2 as weights in both equations to induce approximate homoskedasticity of the disturbances. Note that this procedure performs the estimation of the model in two stages, and the maximum likelihood estimates are no longer fully efficient, but are conditional on the model chosen for σ_i^2 . It would be possible, but difficult due to the high nonlinearity involved, to estimate this new model by maximum likelihood by including the model for σ_i^2 explicitly in the denominator of the residual functions, but I have chosen not to do this in order to simplify the estimation.

The results of this procedure are shown in Table 6. Focusing for the moment on columns 1 and 4, which are comparable to the sample selection estimates in the previous table, we can see that the size coefficient has fallen by one half, and the estimate of ρ is now positive, but insignificant. The LM test for heteroskedasticity of the disturbances of the weighted model no longer rejects after the weighting has been performed. However, the results now show that size has an opposite effect on the probability of survival during the two periods, which seems highly unlikely, given that one sample is a subset of the other. This turns out to be due to a combination of the weighting scheme used and the nonlinearity of the probit index with respect to size but it shows how sensitive this type of estimate can be to

weighting.

Because of this problem and because my goal in performing this test in the sample selection setting was to sort out the different effects of size-related heteroskedasticity, size-related sample attrition, and nonlinearity in the relationship of growth and size, in the other columns of Table 6 I present estimates of the growth rate equation with quadratic and cubic size terms. Note first that the LM test statistic is still insignificant, so heteroskedasticity of a size-related kind is not a problem here. The quadratic term is significant in both periods and the cubic term significant only in the first. The estimates for the probit equation imply a probability of survival which is roughly constant (about 0.88 in 1979 and 0.65 in 1983) until a size of around 10,000 employees and then rises fairly quickly to near one. This is consistent with the observed survival rates.

In both periods, the estimated ρ is quite negative. The fact that the estimate of ρ is robust to the order of the polynomial expansion of the size equation is evidence that the Mills ratio term is not simply proxying for some higher order function of size (in fact, a quartic does not enter this equation significantly). However, a negative correlation between the disturbances of the growth equation and the survival equation does call into question the basis for my original model of exit from the sample, since it seems to imply that firms which grow faster than predicted by their size are more likely to exit from the sample, holding size constant. I will explore this puzzle further in the next section when I look at the reasons for exit from the sample.

6. Sample Attrition as a Function of Tobin's Q

The preceding highlights a problem with this approach to sample

selection correction. Many before us have pointed out that in the absence of exclusion restrictions in the selection equation the identification in such sample selection models comes through the nonlinearity of the Mills' ratio, i.e., the exact functional form of the distribution function. In principle, as we add higher order terms to the regression equation, these terms become more and more collinear with the Mills ratio variable, which itself can be well approximated by a particular polynomial expansion in the Z's. When there are additional variables in the selection equation, this near collinearity disappears and it becomes possible to include nonlinear terms without having them proxy for the selection bias correction.

However, when correcting for selection due to growth, it is extremely difficult to think of variables which belong in a selection equation and not in the growth equation. One possible avenue to pursue is a more explicit modelling of the reasons for exit, about which we have some information. Of the 1778 firms in 1976, 225 exit from the sample by 1979 and another 369 exit by 1983. Both Compustat and the CRSP files (which include many but not all of these firms) contain a code giving the reason for deletion when the data for the firm is removed from the file. Using these codes, Addanki (1985) and I, in parallel work, were able to establish that approximately sixty percent of the firms were dropped due to merger or acquisition, eight percent because of bankruptcy or liquidation, and the remainder for reasons unknown. The last category includes smaller firms, many of which were probably acquired.

We hypothesize that a firm will be acquired and disappear from the sample when the existing assets of the firm are not being employed in

the optimal way; a prospective buyer is willing to buy the firm at the current stock price in the hopes of producing an above average return on the stock by redeploying the assets in some way. That is, the probability of a firm's being acquired is a function of the average Tobin's Q for the firm, the ratio of the market value to the book value of the assets. The market value is assumed to be the current capitalized value of the future earnings potential of the firm's assets. The higher is Q, the less likely that the firm will be acquired and disappear from our sample. This is a fairly crude story which leaves unexplained why the market is undervaluing the assets in this way; it simply posits that if they are undervalued, an opportunity exists for a potential purchaser. What is surprising is that this variable turns out to be a fairly good predictor of survival, somewhat better than the pure size variable we have been using.

Of course, the assets of a firm include more than the physical assets; in particular, we are interested in the value of the assets represented by the firm's technological position, or knowledge stock, as proxied by its R&D history. Thus we would like to use a Q variable which contains a measure of R&D stock as well as physical capital in the denominator. Following Hayashi (1982), Wildasin (1984) has shown that the market value of a firm which maximizes discounted cash flow using more than one stock of capital is given by a weighted sum of the value of the capital stocks:

$$V = \sum_{i=1}^n \lambda_i K_i$$

where K_i are the capital stocks in physical units and the λ_i are the shadow prices of these stocks, which depend on taxes, depreciation, and adjustment costs, and are not necessarily equal over different kinds

of stocks. Unfortunately, we do not have a measure of these shadow prices, so we do not know how to weight the physical assets and R&D stock appropriately in computing Q. Denoting the physical assets by A and the knowledge stock by R, we can write Q as

$$Q = \frac{V}{\lambda_1 A + \lambda_2 R} = \frac{V}{\lambda_1 A} \left(\frac{1}{1 + \gamma R/A} \right)$$

where γ is the ratio of the two shadow prices. Because the measured Q variable in these data exhibits a very longtailed distribution, which tends to give extreme weight to a few out outliers, I chose to use the logarithm of the variable in the selection equation, so that the variable becomes

$$\log Q - \log (V/A) - \log(1 + \gamma(R/A))$$

I approximate $\log (1 + \gamma(R/A))$ by (R/A) since I expect $\gamma(R/A)$ to be small, so that the variables actually used in the selection equation are $\log(V/A)$ and R/A . Firms with no R&D program have an R/A stock equal to zero; the inclusion of a separate dummy for these firms in the selecting equation had no effect on the results. Because of work by Addanki (1985), who found that the valuation of a firms R&D program at the time of acquisition differed depending on whether a firm was a patenter, I allowed for a separate coefficient on R/A for those firms which filed successful patent applications in 1976.

As a measure of the Q of physical assets, V/A , I use the total market value of the firm (common stock, debt, and preferred stock) divided by the sum of net capital stock, inventories, and other assets (including subsidiaries). The value of the components of both V and A have been adjusted for the effects of inflation using the methodology of

Brainard, Shoven, and Weiss; the computations are more fully described in Cummins, Hall, Laderman, and Mundy (1984).⁹

The estimates for a probability of survival equation using these variables are shown in Table 7, along with growth rate equations augmented by the two investment variables (these will be discussed later). What the probit equations show is that both the V/A and R/A variables are more important in predicting survival than the raw size variable employment, although the R/A variable has a large standard error.¹⁰ At the sample means these estimates imply that a doubling of employment increases the probability of survival to 1983 by .03, a doubling of Q increases the probability .10, and a doubling of R&D increases it .03 for non-patenters and .01 for patenters, ceteris paribus. Firms with a larger portion of their assets in R&D are less likely to disappear from the sample, while having patents makes them somewhat more likely to exit than firms with R&D and no patents. This last result is consistent with Addanki.

The use of these variables to help predict the probability of survival has had some effect on the estimates of the growth rate equation. The size coefficient has increased substantially over the estimates in columns 1 and 4 of Table 6 and the estimated correlation

9. For comparison, I also used an unadjusted Q based solely on the raw numbers on the Compustat files. In 1976, the value of this Q was lower, and the dispersion less. The qualitative results of the Probit equation were unchanged, and the coefficients were more significant, suggesting that the process of adjusting for inflation bias also introduces more measurement error into the variable.

10. The other variables in the growth equation were also included in the selection equation, but they had insignificant coefficients so the estimates reported do not include them.

between the residuals of the selection equation and the growth rate equation is quite negative. The results are not sensitive to the exact specification of the selection equation. Inclusion of the Q variable as a predictor seems to be enough to produce a rather anomalous result: a firm which grows faster than predicted by its size and level of investment is somewhat more likely to exit from the sample, controlling for size and Q. This implies that the average growth rate for the smaller firms is underestimated, and hence that the size coefficient in the growth rate equation is biased toward zero when we do not correct for selection.

7. Investment and Firm Growth

This section reports on some descriptive regressions which relate the firm growth rates to the level of investment, both physical and R&D, in 1976. These results reported in this section are in no sense derived from a structural model; we are merely documenting the magnitude of the correlation between investment and growth in the manufacturing sector.

In Table 7 we have added three variables to the standard growth rate equation: the logarithm of capital expenditures in 1976, the logarithm of R&D investment in 1976, and a dummy equal to one for those firms who do no or negligible R&D. Both of the expenditure variables have been scaled by subtracting the logarithm of 1976 employment so that the total size effect still appears in the coefficient of $\log E_{76}$. The investment coefficients are quite substantial: at the mean level of investment for these firms, an increase of four million dollars in physical investment is associated with a one percent increase in the annual growth rate from 1976 to 1979, while it takes only two million

dollars of R&D investment to achieve the same effect for those firms which do R&D. In the second period the effects are the same, which implies considerable persistence in the correlation of growth and investment. Firms which have no R&D program grow on average about one to two percent more slowly than those which do.

Earlier work in this area (Mansfield 1962 and Hymer and Pashigian 1962) found that two results seem to hold when firm growth is examined over a large size range of firms: 1) the variance (in logarithms) is larger at the lower end of the size distribution, and 2) Gibrat's Law is closer to holding for large firms than for small. We have already seen that the first result holds in this sample and the nonlinear estimates in Table 6 suggest that the second one probably holds as well. To check this result I divided the sample into two size classes and reestimated the equations in Table 5. The size cut I chose was 2500 employees in 1976. The median number of employees in 1976 is 2300 and the geometric mean is 2700 (based on the 1349 firms which survive from 1972 to 1979), so there are roughly equal numbers in each class for the observed samples.

A summary of the results for these two size classes is presented in Table 6; these estimates are also computed with corrections for heteroskedasticity and sample selection. The results for the larger firms do not look that different from those for the smaller firms, although they are somewhat attenuated. A noteworthy feature of the estimates is the substantial difference in the variance of the growth rates across the two samples: in 1979, the ratio of the mean variances (after weighting by weights normalized to be unity on the average) is about 0.4. The estimated investment coefficients are not that different from those for the whole sample although they have larger standard

errors. The finding that a dollar of R&D expenditures is a more important predictor of growth in the immediate future than expenditures on physical capital is robust across size classes: the ratio of the amount required to obtain an increase in annual growth rates of one percent is 1.6 for the smaller firms and three for the larger firms. This is suggestive for future work: is it because R&D is far less volatile and therefore a better indicator of firms on a "successful" trajectory, or because the rates of return to R&D are on average somewhat higher?

8. Conclusion

This goal of this paper was to investigate several econometric explanations which have been suggested for the finding of a negative correlation between firm size and growth and to lay some groundwork for a more careful modelling of firm dynamics. With respect to the first question, we have negative results in the sense that neither measurement error in employment nor sample attrition can account for the negative coefficient on firm size in the growth rate equation. There are large random changes in employment at any one firm from year to year, but these changes are largely permanent, and do not reflect a non-serially correlated measurement error. Substantial differences in the variance of growth rates across size classes was also observed with smaller firms having a variance at least twice as large.

With qualifications due to the difficulty of constructing an adequate model of sample attrition, it does appear that the smaller firms in the sample grow faster, with a four percentage point difference in annual growth rates between firms in the 25th and 75th percentiles in size. Because of the large element of randomness in growth rates across

firms from year to year, however, this difference is not enough to cause firms to move very far in the size distribution over a ten year period.

With respect to the second question, I have found that the obvious systematic differences among firms, such as industry and the level of investment, do very little to reduce the variance of growth rates. The best I could do was a reduction in the standard error from 12.6 percent to 12.1 percent (this conclusion is based on the unweighted data, since it is difficult to interpret the standard error after weighting). In future I hope to explore this further using an intertemporal dynamic optimizing model of the firm. The results of this paper give some confidence that this can be done, at least over short periods, without worrying too much about those firms which exit from the sample.

Table 1

Employment Growth by Industry

Industry	Number of Firms	Log E 1976	E 76 (1000s)	R/S 76		I/S 76	
				Firm	Ind.	Firm	Ind.
Food & kindred prod.	144	1.247	3.5	.004	.003	.047	.039
Textiles & apparel	135	.857	2.3	.008	.004	.029	.033
Chemicals excl. drugs	98	.954	2.6	.022	.023	.066	.102
Drugs & med. inst.	87	.581	1.8	.053	.042	.053	.044
Petroleum ref. & ext.	43	1.911	6.8	.005	.003	.125	.103
Rubber & misc. plast.	47	.636	1.9	.016	.017	.050	.040
Stone, clay & glass	60	1.026	2.8	.010	.011	.070	.064
Primary metals	83	1.032	2.8	.010	.006	.055	.080
Fabricated metals	135	.349	1.4	.009	.006	.049	.043
Engines, farm & const.	49	1.465	4.3	.017	.018	.114	.059
Office, comp. eq.	61	.465	1.6	.090	.054	.224	.129
Other mach., not elec.	143	.448	1.6	.020	.015	.049	.041
Elec. equip. & supplies	65	1.178	3.2	.021	.020	.043	.038
Communication equip	148	.0002	1.0	.049	.029	.045	.071
Motor veh. & trans.	80	1.430	4.2	.011	.022	.063	.039
Aircraft & aerospace	29	1.397	4.0	.021	.031	.029	.024
Prof. & sci. equip.	75	.232	1.3	.050	.053	.040	.062
Lumber, wood & paper	120	.831	2.3	.007	.004	.056	.093
Misc. consumer goods	153	.740	2.1	.014	.003	.039	.061
Conglomerates	23	3.082	21.8	.014	.011	.048	.053
Total	1778	.811	2.3	.027		.058	

Table 1 (continued)

Industry	1972		Log E 1976	dlog E 72-79	1983		Log E 1976	dlog E 76-83
	Number	%			Number	%		
Food & kindred prod.	112	78	1.409	2.2	62	43	1.487	1.3
Textiles & apparel	101	75	1.089	-0.02	67	50	1.269	-3.4
Chemicals excl. drugs	72	73	1.391	3.8	68	69	1.257	0.8
Drugs & med. inst.	66	69	1.034	5.3	63	72	0.976	4.1
Petroleum ref. & ext.	37	43	2.201	4.4	26	60	1.980	0.7
Rubber & misc. plast.	33	77	0.885	1.6	28	60	0.936	-0.3
Stone, clay & glass	46	77	1.069	-0.8	37	62	1.330	-2.0
Primary metals	66	80	1.174	2.2	47	57	1.345	-1.1
Fabricated metals	103	76	0.474	2.5	82	61	0.482	-0.4
Engines, farm & const.	39	80	1.791	5.8	40	82	1.502	-2.5
Office, comp. eq.	37	61	1.206	11.2	40	66	0.535	7.5
Other mach., not elec.	109	76	0.560	4.3	82	57	0.404	-0.2
Elec. equip. & supplies	50	77	1.300	2.9	45	69	1.099	0.4
Communication equip.	107	72	0.232	5.1	95	64	-0.067	3.5
Motor veh. & transp.	61	76	1.613	0.2	51	64	1.787	-3.5
Aircraft & aerospace	27	73	1.480	2.8	21	72	1.783	0.0
Prof. & sci. equip.	59	79	0.504	6.6	53	71	0.341	5.0
Lumber, wood & paper	92	77	1.045	4.5	78	65	1.013	1.1
Misc. consumer goods	116	76	0.861	2.6	98	64	0.935	2.2
Conglomerates	16	70	3.522	2.1	15	65	3.336	-1.0
Total	1349		1.044	2.9	1098		0.983	0.8

Notes:

The column E 76 is the geometric mean of 1976 employment.

The growth rates are per year.

For R/S and I/S, both the average firm ratio (labelled Firm) and the aggregate industry ratio (labelled Ind.) are shown.

Table 2

Growth Rate Regressions

1972-1979: 1349 Firms

Dependent Variable	Annual Growth Rate in Percentage Terms			
	72-79 OLS	73-79 OLS	73-79 OLS	73-79 Inst. Var.
Intercept	4.0 (0.2)	2.9 (0.3)	2.9 (0.3)	3.0 (0.3)
Logarithm of size in year	-1.14 (.15) 72	-0.98 (.14) 72	-0.92 (.14) 73	-0.99 (.14) 73
Standard error	8.4	9.4	9.5	9.5

Dependent Variable	20 Industry Dummies			
	72-79 OLS	73-79 OLS	73-79 OLS	73-79 Inst. Var.
Intercept				
Logarithm of size in year	-1.09 (.13) 72	-0.95 (.14) 72	-0.90 (.14) 73	-0.97 (.14) 73
Standard error	7.6	8.7	8.7	8.7
F-statistic for industry dummies	7.20	6.36	6.44	6.44

1976-1983: 1098 Firms

Dependent Variable	Annual Growth Rate in Percentage Terms			
	76-83 OLS	77-83 OLS	77-83 OLS	77-83 Inst. Var.
Intercept	1.89 (.30)	1.07 (.32)	1.06 (.33)	1.13 (.33)
Logarithm of size in year	-1.06 (.15) 76	-0.99 (.16) 76	-0.93 (.16) 77	-1.00 (.16) 77
Standard error	8.6	9.2	9.2	9.2

Dependent Variable	20 Industry Dummies			
	76-83 OLS	77-83 OLS	77-83 OLS	77-83 Inst. Var.
Intercept				
Logarithm of size in year	-0.89 (.16) 76	-0.79 (.17) 76	-0.72 (.17) 77	-0.80 (.17) 77
Standard error	8.3	8.9	8.9	8.9
F-statistic for industry dummies	4.83	5.24	5.30	5.30

Table 3

Time Series Behavior of Log Employment

Autoregressive Model with Measurement Error¹

	1972-79	1976-83	1972-83
β	.991(.001)	.990(.002)	.991(.001)
Var. of Meas. Error (σ_w^2)	-.0018(.0004)	-.0036(.0005)	-.0027(.0012)
Var. of Shock (σ_u^2) ³	.0368(.0015)	.0426(.0023)	.0406(.0174)

Adjustment Cost Model²

	1972-79	1976-83	1972-83
λ	1.745(.147)	.553(.097)	.878(.051)
Var. of Meas. Error (σ_w^2)	.057(.005)	.015(.004)	.029(.002)
Var. of Shock (σ_u^2) ³	.032(.001)	.057(.006)	.075(.019)

Notes:

¹These estimates are derived from the ARMA(1,1) estimates which were based on the covariance of the levels of log employment described in Appendix A. See the text for a definition of the model and an explanation of the negative variance estimate.

²These estimates are derived from the ARMA(1,1) estimates which were based on the covariance of the first differences of log employment described in Appendix A.

³These are derived for a representative estimate of σ_ε^2 (about 0.035). They actually will change slightly each year.

Table 4
Annual Growth Rates 1972-1983

962 Firms

	Average Growth Rate	Standard Deviation	Estimated S.D.* of Shock
1972-73	8.8	16.4	16.2
73-74	0.5	18.5	18.4
74-75	-4.2	15.9	15.8
75-76	5.3	18.3	18.3
76-77	4.6	17.1	17.0
77-78	6.7	14.4	14.2
78-79	4.4	17.1	16.9
79-80	-2.7	19.8	19.7
80-81	-1.5	17.9	17.6
81-82	-8.7	22.3	22.2
82-83	0.6	24.1	24.1

*These estimates are from the ARIMA(1,1,1) model.

Table 5

Growth Rate Regressions with Selection Correction

1778 Firms

Dependent Variable	OLS and Probit		Sample Selection	
	1979	Annual Growth Rate from 1976 to 1983	1979	1983
Number of Firms	1551	1184	1551	1184
Intercept	5.54(.40)* (.33)	1.81(.34) (.29)	5.54(**) (.69)	1.69(.33) (1.05)
Log E ₇₆	-1.06(.18) (.18)	-0.92(.16) (.15)	-1.06(**) (.18)	-0.91(.15) (.17)
Standard error	11.7	8.6	11.7	8.6
Dependent Variable	Probability of Survival			
Intercept	1.10(.041) (.041)	.352(.034) (.034)	1.10(**) (.041)	.352(.034) (.034)
Log E ₇₆	.057(.021) (.024)	.104(.019) (.019)	.057(**) (.024)	.104(.019) (.020)
$\hat{\rho}$	-	-	0.0 (**) (.204)	.022(.011) (.197)
Log of Likelihood	446.3	105.1	446.3	105.1
χ^2 for squared and cubic size terms (DF=4)			60.4	32.8
LM Test for heteroskedasticity (DF=2)	60.6	43.0	60.6	43.2

*The first set of numbers in parentheses are heteroskedastic consistent standard error estimates and the second set are ordinary estimates.

**The HS-consistent standard errors are not computable since the maximum likelihood estimate of ρ is exactly zero.

Table 6
Growth Rate Regressions with Corrections
for Heteroskedasticity and Selection

1778 Firms

Dependent Var.	Annual Growth Rate from 1976 to					
	1979	1979	1979	1983	1983	1983
Intercept	4.51(.45)	6.73(.46)	5.73(.76)	0.25(1.00)	6.37(.57)	6.10(.69)
Log E ₇₆	-.53(.16)	-1.52(.29)	-1.54(.29)	-.49(.22)	-2.12(.30)	-2.10(.31)
(Log E ₇₆) ²		0.16(.07)	0.64(.14)		0.16(.07)	0.26(.16)
(Log E ₇₆) ³			-.10(.03)			-.02(.03)
Slope(E=700)	-.53(.16)	-1.63(.35)	-2.03(.37)	-.49(.22)	-2.23(.34)	-2.29(.35)
Slope(E=17000)	-.53(.16)	-.77(.20)	-.14(.29)	-.49(.22)	-1.39(.21)	-1.24(.26)
Std error (wtd)	9.32	9.60	9.51	5.54	6.46	6.41

Dependent Variable	Probability of Survival					
	Intercept	13.4(.53)	13.0(.57)	13.4(.66)	2.85(.33)	2.66(.33)
Log E ₇₆	-1.01(.21)	-2.91(.50)	-2.96(.53)	.57(.15)	-.33(.28)	-.32(.30)
(Log E ₇₆) ²		.70(.17)	.47(.28)		.32(.09)	.31(.15)
(Log E ₇₆) ³			.058(.080)			.0004(.039)
$\hat{\rho}$.16(.10)	-.49(.12)	-.41(.22)	.22(.16)	-.74(.06)	-.73(.07)
Log likelihood	-2872.4	-2855.9	-2850.0	-2801.1	-2781.1	-2780.9
LM test for heteroskedasticity (DF=2)	0.17	0.06	0.08	2.51	2.55	2.50

All standard error estimates are heteroskedastic consistent estimates; they are the same as the conventional estimates to two digits.

The weights are inversely proportional to size and size squared (see the text for an explanation of the heteroskedasticity correction).

Table 7
Growth Rate Regressions with Q

1753 Firms

OLS and Probit

Sample Selection

Dependent Variable	Annual Growth Rate from 1976 to			
	1979	1983	1979	1983
Number of Firms	1529	1171	1529	1171
Intercept	6.68(.46)	2.59(.38)	8.38(.51)	6.83(.45)
Log E ₇₆	-1.08(.16)	-1.14(.13)	-1.46(.18)	-1.95(.14)
Log (I/E) ₇₆	1.26(.28)	1.22(.24)	1.41(.29)	1.36(.24)
Log (R/E) ₇₆	1.31(.35)	1.33(.20)	1.26(.35)	1.10(.22)
D(R=0)	-3.50(.62)	-2.30(.53)	-2.71(.63)	-0.81(.51)
Standard error (wtd)	8.98	5.17	9.55	6.22

Dependent Variable	Probability of Survival to			
	1979	1983	1979	1983
Intercept	11.6(.56)	1.97(.34)	10.8(.60)	1.54(.33)
Log E ₇₆	-.92(.23)	.62(.15)	-.62(.25)	.60(.14)
Log Q ₇₆	1.66(.64)	2.17(.37)	2.47(.70)	2.75(.33)
(R/A) ₇₆ • (Patents=0)	2.72(4.82)	2.25(3.13)	8.07(5.14)	4.61(2.77)
(R/A) ₇₆ • (Patents>0)	2.59(4.10)	1.25(2.66)	3.60(3.30)	2.83(2.65)
$\hat{\rho}$	-	-	-.61(.09)	-.82(.03)
Log likelihood	-2918.1	-2740.0	-2903.2	-2702.3
LM test for heteroskedasticity (DF=4)			20.6	0.11

All standard error estimates are heteroskedastic-consistent estimates.

Estimates are obtained by maximum likelihood of the sample selection model with the disturbances weighted to correct for heteroskedasticity.

Table 8
Growth Rate Regressions by Firm Size

Dependent Variable	Small Firms		Large Firms	
	1979	Annual Growth Rate from 1976 to 1983	1979	1983
Number of Firms	832	604	697	567
Intercept	9.08(.63)	7.29(.61)	6.92(.96)	6.02(.93)
Log E_{76}	-1.58(.56)	-1.49(.49)	-0.94(.30)	-1.75(.28)
Log $(I/E)_{76}$	2.74(.43)	1.46(.38)	1.86(.33)	1.22(.30)
Log $(R/E)_{76}$	2.41(.38)	1.28(.37)	1.01(.35)	0.85(.27)
D(R=0)	-4.15(.88)	-1.71(.77)	-1.48(.81)	-0.30(.68)
Standard error (wtd)	11.6	6.60	7.50	5.63

Dependent Variable	Probability of Survival to			
	1979	1983	1979	1983
Intercept	10.9(.69)	15.8(.39)	7.99(1.21)	.494(.81)
Log E_{76}	-2.03(.81)	-.030(.44)	.703(.48)	.951(.30)
Log Q_{76}	1.66(.82)	2.01(.43)	3.95(.89)	3.42(.50)
$(R/A)_{76} \cdot (\text{Patents}=0)$	15.6(9.6)	5.70(2.92)	-40.6(17.8)	-18.5(13.1)
$(R/A)_{76} \cdot (\text{Patents}>0)$	4.8(4.0)	2.75(2.87)	11.0(17.6)	0.2(3.8)
$\hat{\rho}$	-.77(.06)	-.80(.05)	-.57(.10)	-.82(.05)
Log likelihood	-1566.6	-1477.8	-1253.2	-1238.9
LM test for heteroskedasticity (DF=4)	0.64	0.01	0.49	0.23

All standard errors are heteroskedastic-consistent estimates.

Estimates are obtained by maximum likelihood of the sample selection model with the disturbances weighted to correct for heteroskedasticity.

Figure 1

1976-1979 Growth Rate versus Size in 1976

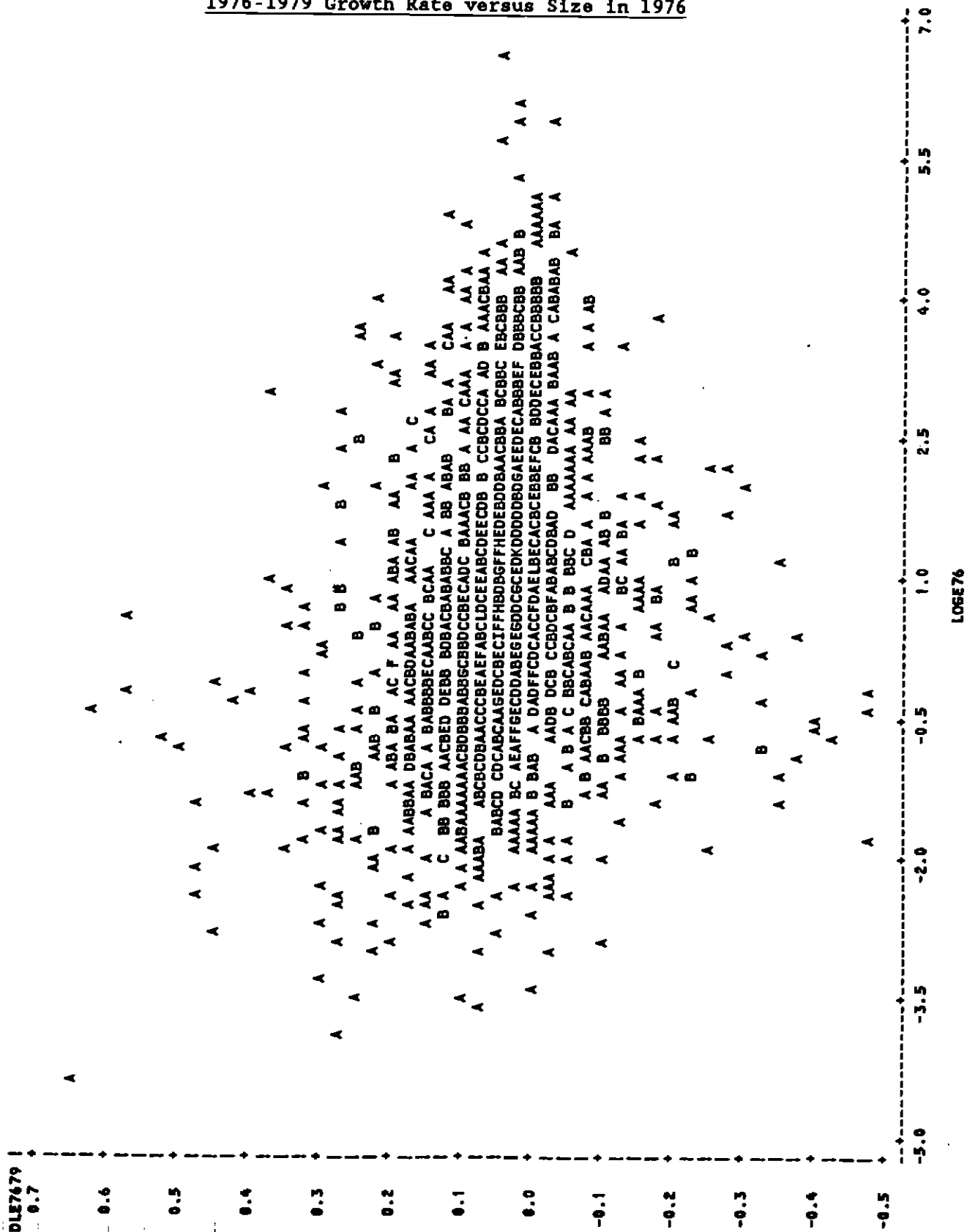
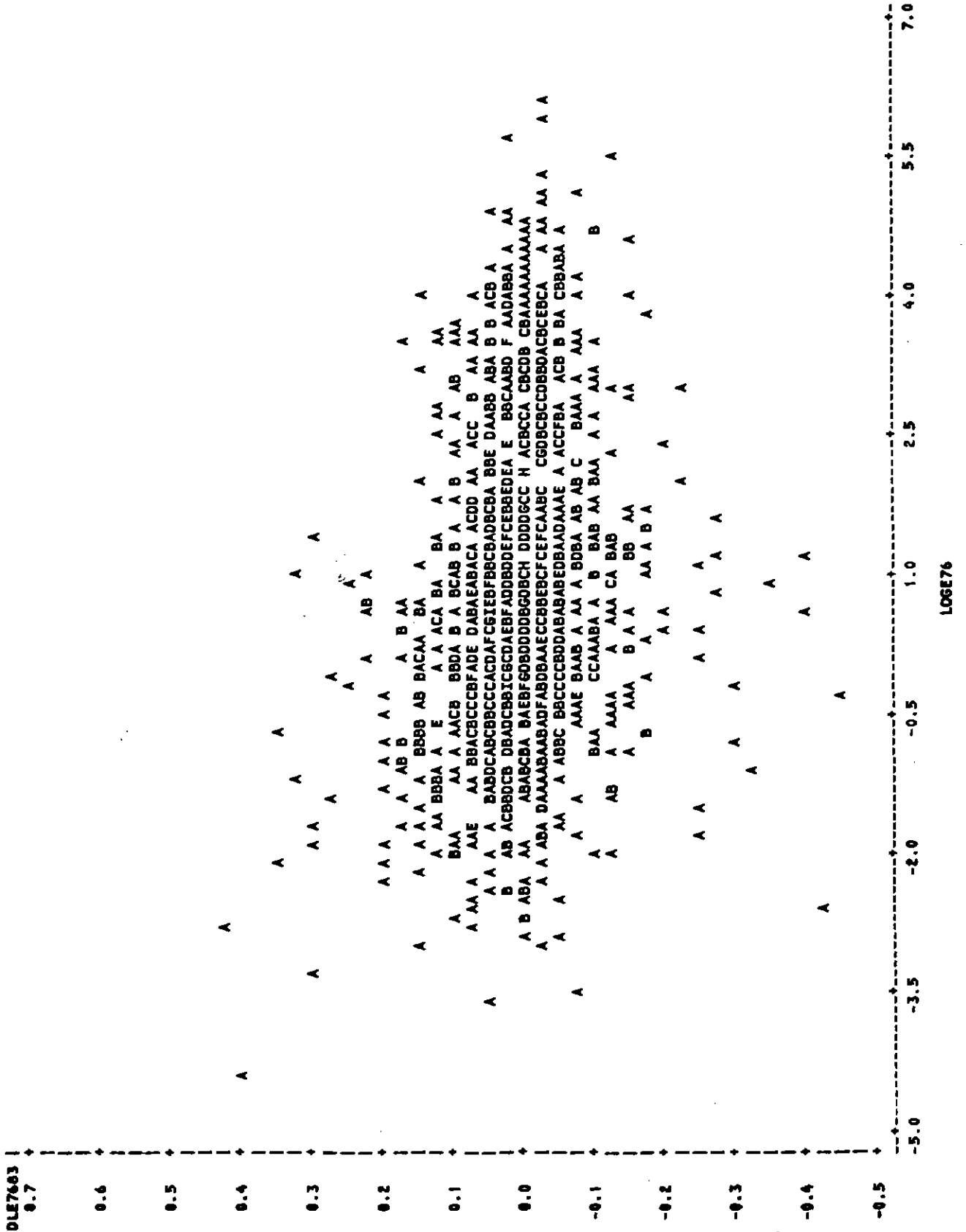


Figure 2

1976-1983 Growth Rate versus Size in 1976



References

- Addanki, Sumanth. 1985. mimeo, National Bureau of Economic Research.
- Amemiya, Takeshi. 1984. "Tobit Models: A Survey." Journal of Econometrics 24: 3-61.
- American Bar Association. 1968. Selected Articles on Federal Securities Laws. Chicago: R. R. Donnelly & Sons.
- Anderson, T. W. and C. Hsiao. 1981. "Estimation of Dynamic Models with Error Components." JASA 76:598-606.
- Armington, Catherine, and Marjorie Odle. 1982. "Sources of Job Growth: A New Look at the Small Business Role." Economic Development Commentary. 6:no. 3.
- Bound, J., C. Cummins, Z. Griliches, B. H. Hall, and A. Jaffe. 1984. "Who Does R&D and Who Patents?", in R&D, Patents, and Productivity, Z. Griliches (ed.), Chicago: University of Chicago Press, 1984.
- Brainard, W., J. Shoven, and L. Weiss. 1980. "The Financial Valuation of the Return to Capital." Brookings Papers on Economic Activity, 2.
- Breusch, T.S., and A.R. Pagan. 1979. "A Simple Test for Heteroskedasticity and Random Coefficient Variation." Econometrica 47:1287-1294.
- Brock, William A. and David S. Evans. 1986. The Economics of Small Businesses: Their Role and Regulation in the U.S. Economy. New York: Holmes and Meier.
- Bureau of Labor Statistics. 1980. Handbook of Labor Statistics, Bulletin 2070, Washington, D. C.: United States Government Printing Office.
- Cummins, C., Bronwyn H. Hall, Elizabeth S. Laderman and Joy Mundy. 1984. "The R&D Master File: Documentation," National Bureau of Economic Research. Mimeo.
- Evans, David S. 1985. "The Relationship between Firm Size, Growth, and Age: U.S. Manufacturing 1976-1982." Fordham University and CERA. Mimeo.
- Friedman, Stanley J. 1977. SEC Reporting Requirements, New York City: Practising Law Institute.
- Greene, Richard. 1982. "Tracking Job Growth in Private Industry." Monthly Labor Review, 1982: September.
- Griliches, Zvi. 1981. "Market Value, R&D, and Patents." Economic

Letters 7.

Griliches, Zvi, and Jacques Mairesse. 1981. "Productivity and R&D at the Firm Level." National Bureau of Economic Research Working Paper No. 826.

Griliches, Zvi. 1984. R&D, Patents, and Productivity, NBER Conference Volume. Chicago: Chicago University Press.

Griliches, Zvi, Bronwyn H. Hall, and Jerry A. Hausman. 1978. "Missing Data and Self Selection in Large Panels." Annals de l'INSEE 30-31: 137-176.

Griliches, Zvi, and Jerry A. Hausman. 1985. "Errors in Variables in Panel Data," National Bureau of Economic Research Working Paper no.

Hall, Bronwyn H. 1986. MOMENTS User's Manual. Stanford, California: TSP International.

Hall, Bronwyn H., Zvi Griliches, and Jerry A. Hausman. 1984. "Patents and R&D: Is There a Lag?" IER 27:265-285.

Harris, Candee. 1982. "Methodological Differences in Job Generations Studies of David Birch and of Catherine Armington and Marjorie Odle." Brookings Institution. Mimeo.

Hart, P. E., and S. J. Prais. 1956. "The Analysis of Business Concentration: A Statistical Approach." Journal of the Royal Statistical Society, Series A: 150-191.

Hayashi, Fumio. 1982. "Tobin's Marginal q and Average q : A Neoclassical Interpretation." Econometrica, 50:213-224.

Hymer, Stephen and Peter Pashigian. 1962. "Firm Size and Rate of Growth." Journal of Political Economy, 52: 556-569.

Jovanovic, Boyan. 1982. "Selection and Evolution of Industry." Econometrica, 50: 649-670.

Lee, Lung-Fei, and G.S. Maddala. 1985. "The Common Structure of Tests for Selecting Bias, Serial Correlation, Heteroskedasticity, and Non-normality in the Tobit Model." IER 26:1-20.

Leonard, Jonathan S. 1984. "On the Size Distribution of Employment and Establishments." Institute of Industrial Relations, University of California at Berkeley. Mimeo.

Lucas, Robert E. 1978. "On the Size Distribution of Business Firms." Bell Journal of Economics, 9:508-523.

MaCurdy, Thomas E. 1981. "Asymptotic Properties of Quasi-Maximum Likelihood Estimators and Test Statistics." National Bureau of Economic Research, Technical Paper No. 14.

_____. 1985. "A Guide to Applying Time Series Models to Panel Data." Stanford University. Mimeo.

Maddala, G. S. 1983. Limited-dependent and Qualitative Variables in Econometrics. Cambridge, England: Cambridge University Press.

Mandelbrot, Benoit. 1963a. "Oligopoly, Mergers, and the Paretian Size Distribution of Firms." Research Note #NC-246. Yorktown Heights: Thomas J. Watson Research Center.

Mandelbrot, Benoit. 1963b. "A Survey of Growth and Diffusion Models of the Law of Pareto." Research Note #NC-253. Yorktown Heights: Thomas J. Watson Research Center.

Mansfield, Edwin. 1962. "Entry, Innovation, and the Growth of Firms." American Economic Review, 52: 1023-1051.

Marris, Robin. 1979. The Theory and Future of the Corporate Economy and Society. Amsterdam: North-Holland Publishing Company.

Nelson, Forrest. 1977. "Censored Regression Models with Unobserved Stochastic Censoring Thresholds." Journal of Economics, 6:309-22.

Olsen, R. J. 1980. "A Least Squares Correction for Selectivity Bias." Econometrica 48: 1815-1820.

Penrose, Edith T. 1959. The Theory of the Growth of the Firm. Whiteplains, New York: M. E. Sharpe, Inc.

Quandt, Richard E. 1966. "On the Size Distribution of Firms." Econometrica, 34: 416-432.

Scherer, F. M. 1984. Innovation and Growth: Schumpeterian Perspectives. Cambridge: MIT Press.

Scherer, F. M. 1980. Industrial Market Structure and Economic Performance. 2d ed. Boston: Houghton Mifflin Company.

Simon, Herbert A., and Charles P. Bonini. 1958. "The Size Distribution of Firms." American Economic Review, 48: 607-617.

White, Halbert. 1980. "A Heteroskedastic Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity." Econometrica 48: 817-838.

Wildasin, David E. 1984. "The q Theory of Investment with Many Capital Goods." American Economic Review 74:203-210.

Williamson, Oliver E. 1975. Markets and Hierarchies. London: The Free Press.

Appendix A

The Time Series Behavior of Employment Growth

In this paper we present evidence that sample selection or attrition introduces very little bias into growth rate equations over time periods of approximately five to ten years. Therefore, in this appendix we present the results of a time series analysis of three sets of firms (those in the sample from 1972 to 1979, from 1976 to 1983, and from 1972 to 1983) with some confidence that these results are not biased by the exclusion of entrants and exiters.

In Tables A1 and A2 we show the covariance matrix of the logarithm of employment over time for two samples of firms, both in levels and in first differences. In both cases, the overall mean for each year has been removed.¹ These tables indicate that the log employment time series process has the following characteristics: it has an AR component with a root near one, and possibly a small MA component or higher order AR terms. The hypotheses that the variance of growth rates is equal across years can be rejected. Accordingly, we parametrize the process as a standard ARMA model with the variance of the innovation changing over time:

$$(1 - \alpha_1 L - \alpha_2 L^2) Y_t = (1 - \mu_1 L) \varepsilon_t \quad E \varepsilon_t^2 = \sigma_t^2$$

1. The differenced matrix was also estimated with industry means removed for each year to control for possible industry effects of the oil price shock, but this made little difference, reducing the diagonal elements by about five percent and leaving the off-diagonal elements essentially unchanged.

Under the assumption of multivariate normality of the ε_t , it is possible to estimate the parameters of this process by maximum likelihood; the covariance matrices are a sufficient statistic for the problem. The method by which I perform this estimation is described in Hall (1979)². Macurdy (1981) has shown that these estimates are consistent even if the disturbances are not multivariate normal, although the estimated standard errors are no longer correct.

Before describing the results of my estimation of the model, I need to say something about the treatment of initial conditions. I have assumed that the process for each firm began at a random time in the past and at a random level, and accordingly, have estimated the initial variance as a free parameter (in the case of AR(2), two initial variances and a covariance are free). Justification for this procedure is provided both by Anderson and Hsiao (1981) and MaCurdy (1985). It will not be correct if the unknown initial condition is a fixed constant.³ It is difficult to conceive of an experiment with this data that would distinguish the two possibilities, although the smooth lognormality of the size distribution gives me some confidence that the

2. The likelihood function being maximized can be written

$$\log L = -\frac{T}{2} \left[\log 2\pi + \log |\Omega(\theta)| \right] - \frac{N}{2} \text{tr} \left[Y' Y \Omega(\theta)^{-1} \right]$$

where N is the number of firms, T is the number of time periods and Y'Y and $\Omega(\theta)$ are the observed and predicted covariance matrix of the data respectively.

3. I am grateful to Jerry Hausman for pointing this out to me.

first assumption is not unreasonable. The consequence of this treatment of initial conditions is to add two or three more parameters when the model is expanded to include a second order term, rather than only one. This has a tendency to increase the log likelihood by more than is accounted for by the additional AR parameter, due to the fact that the first two variances and the associated covariance are now estimated freely. This accounts, for example, for the fact that the 72-79 data prefers the ARMA (2,1) model strongly, in spite of the fact that it seems to have redundant roots.

Using these assumptions about initial conditions, I estimated the ARMA model on three sets of data: the two samples from 1972 to 79 and 1976 to 83 shown in Tables A1 and A2, and finally on a longer sample from 1972 to 1983 containing 962 firms. The results are essentially the same across the three samples and they are shown in Table A3. There I show the value of the log likelihood obtained for six different models, where I have imposed the constraints $\alpha_1 = 1$, $\alpha_2 = 0$, and $\mu_1 = 0$ separately and combined. In the table it can be seen that the gain in the likelihood per degree of freedom is vastly greater going from a simple Martingale to an ARMA(2,1) model than from the ARMA(2,1) to the unconstrained model. The Akaike information criterion suggest that either AR(2) or ARMA(2,1) are to be preferred in levels, while ARMA (1,1) is preferred in first differences.

In Table A4 I show the estimated value of the roots of the different processes. For example, the ARMA(2,1) estimates for 1972-1979 suggest that the employment process be described as

$$(1-1.757L)(1-.984L)Y_t = (1-1.748L) \varepsilon_t$$

It can be seen from the table that the estimates obtained with first

differences are entirely consistent with those obtained with levels, since the dominant effect in the latter case is one autoregressive root near unity. It is also the case that both the ARMA(2,1) in levels and ARMA(1,1) in differences have near redundant roots for the 1972 to 1979 period (the t-statistic for equality is 0.9), while in the later period the roots are stable and significantly different from each other. I conclude that an adequate representation of the times series behavior of the data is ARIMA(1,1,1), with a possible preference for a slightly simpler model in the case of the earlier period because of the unstable and near redundant roots. In the paper I interpret these time series results in the context of several slightly more informative models.

Table A1
Log Employment Covariance over Time

1349 Firms

Levels

	72	73	74	75	76	77	78	79
72	2.762							
73	2.704	2.678						
74	2.695	2.670	2.705					
75	2.655	2.631	2.666	2.655				
76	2.610	2.587	2.623	2.612	2.602			
77	2.575	2.559	2.594	2.586	2.577	2.582		
78	2.546	2.531	2.568	2.600	2.552	2.560	2.566	
79	2.541	2.530	2.573	2.563	2.557	2.570	2.582	2.643

First Differences

	73-72	74-73	75-74	76-75	77-76	78-77	79-78
73-72	.0314						
74-73	.0019	.0427					
75-74	.0012	-.0001	.0273				
76-75	-.0013	-.0006	-.0007	.0323			
77-76	.0038	-.0002	-.0029	-.0010	.0295		
78-77	.0030	.0018	-.0000	-.0012	-.0035	.0275	
79-78	.0041	.0066	-.0023	-.0023	-.0039	-.0059	.0466

Notes:

Overall year means have been removed before computing these matrices. The asymptotic standard error is approximately 0.09 for the levels and 0.002 for the first differences.

Table A2

Log Employment Covariance Over Time

1098 Firms

Levels

	76	77	78	79	80	81	82	83
76	2.95							
77	2.91	2.90						
78	2.87	2.86	2.84					
79	2.83	2.81	2.81	2.82				
80	2.81	2.81	2.79	2.81	2.83			
81	2.80	2.79	2.78	2.80	2.83	2.87		
82	2.78	2.78	2.77	2.79	2.82	2.86	2.91	
83	2.73	2.74	2.73	2.75	2.78	2.82	2.87	2.89

First Differences

	77-76	78-77	79-78	80-79	81-80	82-81	83-82
77-76	.0333						
78-77	.0035	.0234					
79-78	.0040	.0013	.0347				
80-79	-.0012	.0028	.0054	.0405			
81-80	.0000	.0014	.0040	.0086	.0345		
82-81	.0029	.0011	-.00002	.0064	-.0005	.0561	
83-82	.0042	.0035	-.0016	-.0052	-.0001	.0058	.0597

Notes:

Overall year means have been removed before computing these matrices. The asymptotic standard error is approximately 0.11 for the levels and 0.0025 for the first differences.

Table A3

Time Series Estimates for Log Employment

Levels

Model	1972-79		1976-83		1972-83	
	# Params	LogL ⁺	# Params	LogL	# Params	LogL
$\alpha_1=1, \alpha_2=0, \mu_1=0$	8	-151.3	8	-140.9	12	-207.0
AR(1): $\alpha_2=0, \mu_1=0$	9	-107.0	9	-115.7	13	-165.9
$\alpha_1=1, \alpha_2=0, MA(1)$	10	-111.9	10	-96.3	14	-162.7
ARMA(1,1): $\alpha_2=0$	11	-86.0	11	-76.1	15	-133.7
AR(2): $\mu_1=0$	11	-85.2	11	-69.9	15	-132.2
ARMA(2,1)	13	-36.4	13	-55.8	17	-117.2
Unconstrained	36	0.0 (225.2)	36	0.0 (-406.6)	78	0.0 (892.0)

First Differences

	1972-79		1976-83		1972-83	
	# Params	LogL	# Params	LogL	# Params	LogL
ARMA(0,0)	7	-87.6	7	-127.9	11	-193.4
MA(1)	9	-74.7	9	-89.9	13	-162.9
ARMA(1,1)	10	-29.9	10	-73.2	14	-142.6
Unconstrained	28	0.0 (2757.4)	28	0.0 (1729.6)	66	0.0 (3226.0)

Notes:

⁺The logarithm of the likelihood is measured relative to the unconstrained model, which freely fits each covariance to a separate parameter. The actual unconstrained log likelihoods are shown in parentheses at the bottom of the table.

Table A4

Parameter Estimates for Time Series Models

		Roots of AR Process		Roots of MA Process	
		Levels			
AR(1)	72-79	.989(.001)	0		0
	76-83	.991(.001)	0		0
	72-83	.990(.001)	0		0
MA(1)	72-79	1	0		-.0533(.0109)
	76-83	1	0		-.0951(.0119)
	72-83	1	0		-.0735(.0103)
ARMA(1,1)	72-79	.991(.001)	0		-.0553(.0111)
	76-83	.990(.002)	0		-.1025(.0125)
	72-83	.991(.001)	0		-.0765(.0106)
AR(2)	72-79	.991(.017)	.0601(.0182)		0
	76-83	.990(.018)	.1196(.0223)		0
	72-83	.990(.003)	.0816(.0055)		0
ARMA(2,1)	72-79	.984(.414)	1.757(.294)		1.748(.138)
	76-83	.991(.175)	.574(.388)		0.449(.096)
	72-83	1.000(.767)	.939(.012)		0.919(.057)
		First Differences			
MA(1)	72-79	1	0		-.0542(.0120)
	76-83	1	0		-.1051(.0146)
	72-83	1	0		-.0756(.0108)
ARMA(1,1)	72-79	1	1.745(.147)		1.737(.150)
	76-83	1	.553(.097)		.432(.101)
	72-83	1	.878(.051)		.821(.056)

Appendix B

In this appendix we demonstrate the relationship of the stochastic threshold model to the generalized Tobit (sample selection) model and discuss the consequences of the identifying assumptions used in estimating each model. Although there is nothing new here, the literature on Tobit models (see Amemiya 1984, Maddala 1983) does not seem to contain a discussion of the connection between the two models. Such a connection is useful, since it implies that the same computer program can be used to estimate both models.

First we present the standard censored regression model with a stochastic threshold due to Nelson (1977). Denote the size of the firm in the second period as y_{1i} and the unobserved threshold below which the firm will drop out of the sample as y_{2i} . Then we have the model

$$y_{1i} = X_i \beta_1 + u_{1i} \quad \text{if } y_{1i} > y_{2i}$$

$$y_{1i} \quad \text{not observed} \quad \text{if } y_{1i} < y_{2i}$$

$$y_{2i} = X_i \beta_2 + u_{2i}$$

The disturbance vector $u = (u_{1i} \ u_{2i})$ has a bivariate normal distribution with mean zero and variance

$$Euu' = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ & \sigma_2^2 \end{bmatrix}$$

The X_i include all the exogenous and predetermined variables for the model, including the size in the initial period. Some of the β 's may be zero if there are exclusion restrictions. Nelson shows that this model requires at least one exclusion restriction or the restriction $\rho=0$ in order to identify all the parameters.

We can rewrite this model as a standard generalized Tobit model of

the form

$$\begin{aligned}
 y_{1i} &= X_i \beta_1 + v_{1i} && \text{if } z_{2i} = y_{1i} - y_{2i} > 0 \\
 y_{1i} &\text{ not observed} && \text{if } z_{2i} = y_{1i} - y_{2i} < 0 \\
 z_{2i} &= X_i \delta + v_{2i} && \text{where } \delta = \beta_1 - \beta_2
 \end{aligned}$$

with a covariance matrix

$$\text{E}v v' = \begin{bmatrix} \omega_1^2 & \lambda \omega_1 \omega_2 \\ & \omega_2^2 \end{bmatrix}$$

It is customary in estimating this model to normalize the residual variance of the unobserved latent variable z_{2i} to be unity so that the disturbance is $v_{2i} \omega_2$ and we actually estimate δ/ω_2 . Since z_2 is completely unobserved, this normalization is innocuous, and still allows us to estimate λ .

However, this is not sufficient to identify the parameters of the stochastic threshold model. It can be easily shown that the relationship between the covariance matrices of the two sets of disturbances is

$$\text{E}u u' = \begin{bmatrix} \omega_1^2 & \omega_1^2 - \lambda \omega_1 \omega_2 \\ & \omega_1^2 - 2\lambda \omega_1 \omega_2 + \omega_2^2 \end{bmatrix}$$

so that $\sigma_1^2 = \omega_1^2$ and

$$\begin{aligned}
 \beta_2 &= \beta_1 - \omega_2 (\delta/\omega_2) \\
 \rho &= (\sigma_1 - \lambda \omega_2)/\sigma_2 \\
 \sigma_2^2 &= \sigma_1^2 - 2\lambda \omega_2 \sigma_1 + \omega_2^2
 \end{aligned}$$

Given estimates of δ/ω_2 , λ , and σ_1 , we will need ω_2 in order to identify the parameters of the stochastic threshold model. As Nelson showed, this can be achieved either with an exclusion restriction on one of the β 's or by setting the correlation between the two equations, ρ , to zero. Therefore, the identifying assumption we used in the sample selection model is not sufficient to identify the parameters of this model.

On the other hand, in the presence of one of the identifying assumptions for the Nelson model, we no longer need to normalize the variance of v_{2i} to be unity. Thus the stochastic threshold model is in some sense a special case of the more general sample selection model. In this paper we have chosen to use the more general model in order to capture the notion the firms may drop out of the sample for reasons other than a size threshold.

Appendix C

Testing for Heteroskedasticity in the Sample Selection Model

This Appendix develops a Lagrange Multiplier Test for heteroskedasticity of the disturbance in the regression equation of the sample selection model, following a test suggested by Lee and Maddala (1985) for the Tobit model. The alternative being considered is that the disturbance v_{1i} of the growth rate equation has a variance of σ_i^2 which is a (possibly nonlinear) function of the regressors (in particular, as we have seen, of size):

$$\sigma_i^2 = G(\alpha + X_i\gamma)$$

Under the null hypothesis, the vector γ is zero and the disturbance is homoskedastic. This test has the usual properties of an LM Test: it is asymptotically locally most powerful under the alternative being considered. As in Lee and Maddala, it turns out that the exact form of G does not matter, since I am approximating it by linear functions of X_i near $\gamma = 0$.

I write the likelihood function for the generalized Tobit model outlined in the text as

$$\begin{aligned} \log L = & \sum_0 \log[1 - \Phi(Z_i\delta)] + \sum_1 \log \Phi(Z_i\delta + \rho v_{1i}/\sigma) \\ & - \sum_1 (v_{1i}^2/2\sigma^2) - \sum_1 \log \sigma - \sum_1 (1/2) \log(1-\rho^2) \end{aligned}$$

where the summation over 0 and 1 denotes the sum over not observed and observed data respectively. $\Phi(\cdot)$ denotes the standard normal CDF.

Differentiating with respect to σ^2 (under the null of homoskedasticity), I obtain

$$(C.1) \quad \frac{\partial \log L}{\partial \sigma^2} \Big|_{H_0} = (2\sigma)^{-2} \sum_1 \left[-\rho(v_{1i}/\sigma) \lambda(Z_i \delta + \rho v_{1i}/\sigma) + (v_{1i}^2/\sigma^2 - 1) \right]$$

where $\lambda(\cdot)$ denotes the inverse Mills ratio, $\phi(\cdot)/\Phi(\cdot)$. The LM (score) test for $\gamma = 0$ is then a test for

$$\frac{\partial \log L}{\partial \sigma^2} \Big|_{H_0} G'(\alpha) X_i = 0$$

where the degrees of freedom for the test are the number of regressors in X_i and all quantities are evaluated at the maximum likelihood estimates obtained under the null hypothesis.

Note that because we are testing for heteroskedasticity of v_{1i} only and not of v_{2i} , only the observations for which y_{1i} is observed enter the test statistic, in contrast to the Tobit model case, where the disturbance of the selection equation and regression equation are the same. To perform the actual test, I use the regression methodology of Breusch and Pagan (1979), which implicitly estimates the variance of this statistic from its sample variance. This computation is invariant to any renormalization which does not depend on the observations so the $G'(\alpha)$ term drops out. The quantity which I regress on the X_i to perform the test is given in the square brackets in equation (C.1). Note that if the estimated ρ is zero, this is the conventional LM test for heteroskedasticity, where v_{1i}^2 is regressed on a constant and the X_i .