Unpublished appendices from
The Relationship between Firm Size and Firm Growth in the U.S. Manufacturing Sector

Bronwyn H. Hall

Journal of Industrial Economics 35 (June 1987): 583-606.

Appendix A: The time series behavior of employment growth

In this paper I present evidence that sample selection or attrition introduces very little bias into growth rate equations over time periods of approximately five to ten years. In this appendix I 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 the combined sample from 1972 to 1983) with some confidence that these results are not very biased by the exclusion of entrants and exiters.

In Tables A1 and A2 I show the covariance matrix of the logarithm of employment over time for the first two samples of firms, both in levels and in first differences. Note that in both cases, the overall mean for each year has been removed from the data.[1] These tables suggest that the log employment time series 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. In addition, the hypothesis that the variance of growth rates is equal across the years can be rejected. Accordingly, I parametrize the process as a standard ARMA model with the variance of the innovation changing over time:

$$(1-\alpha_1 L-\alpha_2 L^2)Y_t=(1-\mu_1 L)\varepsilon_t \qquad E\varepsilon_t^2=\sigma_t^2 \tag{1}$$

Under the assumption of multivariate normality of the ε_t, it is possible to estimate the parameters of this process by maximum likelihood and 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 consistent even if the disturbances are not multivariate normal, although the estimated standard errors will no longer be correct.

---

[1] The differenced matrix was also estimated with industry means removed for each year to control for possible industry effects of the oil price shocks in 1973-74 and 1978-79, but this made little difference, reducing the diagonal elements by about five percent and leaving the off-diagonal elements essentially unchanged.

[2] The likelihood function being maximized can be written as follows:

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

N is the number of firms, *T* is the number of time periods, and *Y'Y* and *Ω(θ)* are the observed covariance matrix of the data and predicted covariance matrix from the model respectively.

Before describing the result 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 (and, in the case of AR(2) or ARMA(2,1), 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. It is difficult to conceive of an experiment with these data that would distinguish the two possibilities, although the smooth lognormality of the size distribution gives me some confidence that the 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 procedure 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. For example, this accounts for the fact that the 1972-79 data prefers the ARMA(2,1) strongly, even though this model seems to have redundant roots (compare 1.757 and 1.748).

Using these assumptions about initial conditions, I estimated the ARMA model on three sets of data: the two samples from 1972 to 1979 and 1976 to 1983 shown in Tables A1 and A2, and a longer sample from 1972 to 1983 that contained 962 firms. The results are shown in Tables A3 and A4 and they are essentially the same across the three samples. In Table A3 I show the value of the log likelihood obtained for six different ARMA models, including ARMA (0,0) or Martingale/random walk, as well as an unconstrained model that allows for a free covariance matrix across time. In the table it can be seen that the gain in the likelihood per degree of freedom is vastly greater going from a simple random walk to an ARMA(2,1) model than from the ARMA(2,1) to the unconstrained model. The Akaike information criteria suggest that either AR(2) or ARMA(2,1) are to be preferred among the levels models, while ARMA(1,1) is preferred for the first differenced data.

In Table A4 I show the estimated values 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 follows:

$$(1-1.76L)(1-0.98L)Y_t = (1-1.75L)\varepsilon_t$$

It can be seen from this table that the estimates obtained with first differenced data are entirely consistent with those obtained using levels, because 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 time 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.

Appendix B: Sample selection and stochastic threshold models

In this appendix I derive 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 (Amemiya 1984, Maddala 1983) does not seem to contain a discussion of the connection between the two models. Such a connection is useful, because it implies that the same computer program can be used to estimate either model.

First I 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 the model can be written as follows:

$$
\begin{aligned}
&y_{1i} = X_i\beta_1 + u_{1i} \quad \text{if} \quad y_{1i} \geq y^*_{2i} \\
&y_{1i} \ \text{not observed} \ \text{if} \quad y_{1i} < y^*_{2i} \\
&y^*_{2i} = X_i\beta_2 + u_{2i}
\end{aligned}
$$

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 includes 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.

The above model can be rewritten as a standard generalized Tobit model of the following form:

$$
\begin{aligned}
&y_{1i} = X_i\beta_1 + \nu_{1i} \quad \text{if} \quad z_{2i} = y_{1i} - y^*_{2i} \geq 0 \\
&y_{1i} \ \text{not observed} \ \text{if} \quad z_{2i} = y_{1i} - y^*_{2i} < 0 \\
&z_{2i} = X_i\delta + \nu_{2i} \quad \text{where} \quad \delta = \beta_1 - \beta_2
\end{aligned}
$$

The covariance matrix of the disturbances is now the following:

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

It is customary when estimating this model to normalize the residual variance of the unobserved latent variable z_{2i} to be unity so that its disturbance is $v_{2i}\omega_2$ and the coefficient vector estimated is δ/ω_2. Because z_2 is completely unobserved, this normalization is innocuous and β_1 is still completely identified.

However, estimation of the sample selection model 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 the following:

$$Euu' = \begin{bmatrix} \omega_1^2 & \\ \omega_1^2 + \lambda\omega_1\omega_2 & \omega_1^2 + 2\lambda\omega_1\omega_2 + \omega_2^2 \end{bmatrix}$$

So that we have

$$\begin{aligned} \sigma_1^2 &= \omega_1^2 \\ \beta_2 &= \beta_1 - \omega_2(\delta / \omega_2) \\ \rho &= (\sigma_1 + \lambda\omega_2) / \sigma_2 \\ \sigma_2^2 &= \sigma_1^2 + 2\lambda\omega_2\sigma_1 + \omega_2^2 \end{aligned}$$

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, identification 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 I used in the sample selection model is not sufficient to identify the parameters of the stochastic threshold model.

On the other hand, in the presence of one of the identifying assumption for the Nelson model, it is no longer necessary to normalize the variance of v_{2i} to unity. Thus the stochastic threshold is in some sense a special case of the more general sample selection model. In this paper I chose to use the more general model in order to capture the notion that the firm 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 that is a (possibly nonlinear) function of the regressors, in particular, of firm size:

$$\sigma_1^2 = G(\alpha + X_i\gamma)$$

Under the null hypothesis, the vector γ is zero and the disturbance therefore 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* goes not matter, since it is being approximated by linear functions of X_i near $\gamma=0$.

The likelihood function for the generalized Tobit model outlined in the text is the following:

$$\log L = \sum_0 \log[1-\Phi(Z_i\delta)] + \sum_1 \left[\log\Phi(Z_i\delta + \rho v_{1i}/\sigma) - v_{1i}^2/2\sigma^2 - \log\sigma - \frac{1}{2}\log(1-\rho^2) \right]$$

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

Differentiating with respect to σ^2 (under the null of homoskedasticity), I obtain the following expression for the score:

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

where $\lambda(.)$ denotes the inverse Mills' ratio, $\varphi(.)/\Phi(.)$. The LM (score) test for γ = 0 is then a test of the following form:

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

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 that the $G'(\alpha)$ term drops out. The quantity which is regressed on the X_i to perform the test is given in the square brackets of equation (2). 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.

Table A1: Log Employment Covariance Matrix

1349 Firms

	Levels							
	1972	*1973*	*1974*	*1975*	*1976*	*1977*	*1978*	*1979*
1972	2.762							
1973	2.704	2.678						
1974	2.695	2.670	2.705					
1975	2.655	2.631	2.666	2.655				
1976	2.610	2.587	2.623	2.612	2.602			
1977	2.575	2.559	2.594	2.586	2.577	2.582		
1978	2.546	2.531	2.568	2.600	2.552	2.560	2.566	
1979	2.541	2.530	2.573	2.563	2.557	2.570	2.582	2.643

	First differences						
	1973-72	*1974-73*	*1975-74*	*1976-75*	*1977-76*	*1978-77*	*1979-78*
1973-72	0.0314						
1974-73	0.0019	0.0427					
1975-74	0.0012	-0.0001	0.0273				
1976-75	-0.0013	-0.0006	-0.0007	0.0323			
1977-76	0.0038	-0.0002	-0.0029	-0.0010	0.0295		
1978-77	0.0030	0.0018	0.0000	-0.0012	-0.0035	0.0275	
1979-78	0.0041	0.0066	-0.0023	-0.0023	-0.0039	-0.0059	0.0466

Overall year means removed.

The asymptotic s.e. is approximately 0.09 for the levels and 0.002 for the first differences.

Table A2: Log Employment Covariance Matrix

1098 Firms

	Levels							
	1976	*1977*	*1978*	*1979*	*1980*	*1981*	*1982*	*1983*
1976	2.95							
1977	2.91	2.90						
1978	2.87	2.86	2.84					
1979	2.83	2.81	2.81	2.82				
1980	2.81	2.81	2.79	2.81	2.83			
1981	2.80	2.79	2.78	2.80	2.83	2.87		
1982	2.78	2.78	2.77	2.79	2.82	2.86	2.91	
1983	2.73	2.74	2.73	2.75	2.78	2.82	2.87	2.89

	First differences						
	1977-76	*1978-77*	*1979-78*	*1980-79*	*1981-80*	*1982-81*	*1983-82*
1977-76	0.0333						
1978-77	0.0035	0.0234					
1979-78	0.0040	0.0013	0.0347				
1980-79	-0.0012	0.0028	0.0054	0.0405			
1981-80	0.0000	0.0014	0.0040	0.0086	0.0345		
1982-81	0.0029	0.0011	0.0000	0.0064	-0.0005	0.0561	
1983-82	0.0042	0.0035	-0.0016	-0.0052	-0.0001	0.0058	0.0597

Overall year means removed.

The asymptotic s.e. is approximately 0.011 for the levels and 0.0025 for the first differences.

Table A3: Time Series Estimates for Log Employment

	1972-79		1976-83		1972-83	
Model	*# params*	*Log L**	*# params*	*Log L**	*# params*	*Log L**
Levels						
Random walk	8	-151.3	8	-140.9	12	-207.0
AR(1)	9	-107.0	9	-115.7	13	-165.9
MA(1)	10	-111.9	10	-96.3	14	-162.7
ARMA(1,1)	11	-86.0	11	-76.1	15	-133.7
AR(2)	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	36	0.0	78	0.0
		(225.2)		(406.6)		(892.0)
First differences						
Random walk	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	28	0.0	66	0.0
		(2757.4)		(1729.6)		(3226.0)

*The logarithm of the likelihood is measured relative to the unconstrained model, which freely fits each covariance to a separate parameter. The actual values of the unconstrained log likelihoods are shown in parentheses below each column.

Table A4: Parameter Estimates for the Time Series Models

		Roots of AR process				*Roots of MA process*	
				Levels			
AR(1)	1972-79	0.989	(0.001)	0		0	
	1976-83	0.991	(0.001)	0		0	
	1972-83	0.990	(0.001)	0		0	
MA(1)	1972-79	1		0		-0.053	(0.011)
	1976-83	1		0		-0.095	(0.012)
	1972-83	1		0		-0.074	(0.010)
ARMA(1,1)	1972-79	0.991	(0.001)	0		-0.0553	(0.011)
	1976-83	0.990	(0.002)	0		-0.1025	(0.013)
	1972-83	0.991	(0.001)	0		-0.0765	(0.011)
AR(2)	1972-79	0.991	(0.017)	0.060	(0.018)	0	
	1976-83	0.990	(0.018)	0.120	(0.022)	0	
	1972-83	0.990	(0.003)	0.082	(0.006)	0	
ARMA(2,1)	1972-79	0.984	(0.414)	1.757	(0.294)	1.748	(0.138)
	1976-83	0.991	(0.175)	0.574	(0.388)	0.449	(0.096)
	1972-83	1.000	(0.767)	0.939	(0.012)	0.919	(0.057)
				First differences			
MA(1)	1972-79	1		0		-0.054	(0.012)
	1976-83	1		0		-0.105	(0.015)
	1972-83	1		0		-0.076	(0.011)
ARMA(1,1)	1972-79	1		1.745	(0.147)	1.737	(0.150)
	1976-83	1		0.553	(0.097)	0.432	(0.101)
	1972-83	1		0.878	(0.051)	0.821	(0.056)