Growth, and Survival: Limited Dependent Variables, Heteroscedasticity and Sample Selection

J Paul Dunne

March 1, 2005

1 The Law of Proportionate Effects

A useful approach for analysing the growth of companies is an empirical analysis within the farmework of testing Gibrat's law (Sutton, 1997 provides a review). This approach was first used in the 1970s to analyses the reasons for an observed inexorable rise in concentration of manufacturing industry,. There was a concern that this would continue and lead to increasing monoopoly poere Hannah and Kay (1977). In fact as this problem was identified things were changing and there was a steady rise in the share of smaller firms in total output taking place. Gibrat's law stated that the probability distribution of growth rates was the same for all sizes of firms.

$$\frac{S_{it}}{S_{it-1}} = \varepsilon_{it}$$

If this held it suggested that the growth rate of companies would be random across size classes, a feature that would lead to the inexorable rise of concentartion within industry and the economy. This can be written in estimable form as:

$$logS_{it} = \alpha \ + \ \beta \log S_{it-1} + \varepsilon_{it}$$

This implies that firms of different sizes have the same proportional growth rates, have the same variances (homoscedastic) and have no serial correlation. Estimating this equation allows a test of Gibrats law by testing if $\beta = 1$ as this implies:

$$logS_{it} - \log S_{it-1} = \alpha + \varepsilon_{it}$$

if $\beta < 1$ smaller firms are growing faster than the larger firms and if $\beta > 1$ the larger firms are growing faster than the smaller firms. This can also be reparametised as a grwoth rate equation

$$\Delta \log S_{it} = \alpha + (\beta - 1) \log S_{it-1} + \varepsilon_t$$

in this case the test is for the coefficient on $\log S_{it-1}$to be zero.

This model has been used frequently in the industrial economics literature. Sutton (1997) Caves (1998) provide reviews.

There are a number of econometric issues that arise:

Growth persistence which can lead to serial correlation

Heteroscedasticity: if the variance iof the growth decreases with size, as seems to be he case this can be a problem, leading to unbiasded but incosnsistet estimates. Will be using heteroscedatsic robust standard error estimates to deal with this.

Outliers: not clear what an outlier is and certainky would be loathe to drop any companies form the sample, but have to be aware that the resulst may be sensitive to extreme values.

Sample selection

A particularly important concern is the issue of sample selction. The way the model is formaulated it is only possible to include companies thatsuvivie over the whole period. However, if the non surviving companies share certain characteristics, such as they are slow growing then this can obviously bias the estimation results. More frmally, what we have is:

$$\begin{aligned} logS_{it} &= \alpha + \beta \log S_{it-1} + \varepsilon_{it} \text{ if } S_{it} > 0 \\ &= 0 \text{ otherwise} \end{aligned}$$

thus

$$E(\log S_{it} \mid \log S_{it-1}, S_{it} > 0) = \alpha + \beta \log S_{it-1} + E(\varepsilon_{it} \mid S_{it} > 0)$$

with

$$\varepsilon_t \sim N(0, \sigma^2)$$

This can be written as

$$E(\log S_{it} \mid \log S_{it-1}, S_{it} > 0) = \alpha + \beta \log S_{it-1} + \sigma\lambda_i$$

where

$$\lambda_i = \frac{f(V_i)}{1 - F(V_i)} \text{ and } V_i = [\frac{\alpha + \beta \log S_{it-1}}{\sigma}]$$

with f(.) the density fuction for the standard normal and F(.) the cumulative density. If we were to estimate a simple OLS regression then we would omit $\sigma\lambda_i$ giving biased and inconsistent estimators.

Another way of interpreting these regressions is to consider the model in log deviations form. Allow

$$y_{it} = \log(S_{it} - \overline{S})$$

then

$$\begin{aligned} y_{it} &= \beta y_{it-1} + \varepsilon_{it} \\ \text{or } y_{it} - y_{it-1} &= (\beta - 1) y_{it-1} + \varepsilon_{it} \end{aligned}$$

Now

$$\frac{\sum y_{it}^2}{N} = \beta^2 \frac{\sum y_{it-1}^2}{N} + \frac{\sum \varepsilon_{it}^2}{N}$$

meaning

$$\sigma_t^2 = \beta^2 \sigma_{t-1}^2 + \sigma_\varepsilon^2$$

which means

$$1 = \beta^2 \frac{\sigma_{t-1}^2}{\sigma_t^2} + \frac{\sigma_\varepsilon^2}{\sigma_t^2}$$

or

$$\beta^2 \frac{\sigma_{t-1}^2}{\sigma_t^2} = 1 - \frac{\sigma_\varepsilon^2}{\sigma_t^2}$$

now the right hand side of this equation is the formula for the R^2 so

$$\beta^2 \frac{\sigma_{t-1}^2}{\sigma_t^2} = R^2$$

or

$$\frac{\sigma_{t-1}^2}{\sigma_t^2} = \frac{R^2}{\beta^2}$$

This means that the larger the ratio of the R^2 and β^2 the larger is the ratio of the previous period variance to the present period. Ths is clearly a measure of convergence.

Moving back to the sample sample selection model, this has become familiar in applied microeconometrics and can be estimated using the Heckman two stage procedure or a maximum likelihood method, both available in the LIMDEP package.

For the two stage procedure let

$$\begin{aligned} d_i &= 1 \text{ when } S_{it} > 0 \\ d_i &= 0 \text{ otherwise} \end{aligned}$$

Then we can set up a likelihood function

$$\begin{aligned} L &= \Pi_{i=1}^{N} \left[\Pr(\varepsilon_i \leqslant -V_i\right]^{1-d_i} \left[\Pr(\varepsilon_i \leqslant -V_i\right]^{d_i} \\ &= \Pi \; F \left[\frac{V_i}{\sigma}\right]^{d_i} \; \left\{1 - F\left[\frac{V_i}{\sigma}\right]\right\}^{1-d_i} \end{aligned}$$

As $F(-t) = 1 - F(t)$ this is the likelihood function for the probit estimation on d_i and $E(d_i) = V_i/\sigma$. So we estimate a probit:

$$\Pr(d_i = 1) = P(V_i)$$

compute V_i and

$$\lambda_i = \left[\frac{f(V_i)}{1 - F(V_i)}\right]$$

For the second stage we use the consistent estimator of λ_i, $\widehat{\lambda}_i$ to estimate

$$E(S_{it} \mid S_{it-1}, S_{it} > 0) = \alpha + \beta S_{it-1} + \sigma \widehat{\lambda}_i$$

giving $\widehat{\beta}$ a consistent estimator.

It is also possible to use a maximum likelihood method , that uses this consistent estimator as a starting value (Judge et al, Limdep) to search for a solution on the highly non-linear likelihood function;

$$L = \Pi_{d_{i=0}} F\left(-V_i, \sigma^2\right)\ \Pi_{d_i=1}\left(S_{it} - V_i, \sigma^2\right)$$

now as $1 - F(-V_i, \sigma^2) = 1 - F(V_i, \sigma^2)$ which we call $1 - F_i$

$$L = \sum_{d_{i=0}} \ln(1 - F_i) - \frac{N - S}{2} ln\sigma^2 - \frac{1}{2\sigma^2} \sum_{d_i=1} \left(S_{it} - V_i, \sigma^2\right)$$

which can be solved using an iterative pocess such as Newton Raphson.