EXERCISE 1. GROWTH AND EXIT OF FIRMS

(To be handed in on Nov. 2)

Firms exit through acquisition, merger, buyout, and dissolution. Economists are interested in the factors that contribute to exit. For example, does increased R&D investment make exit more or less likely? This exercise (and the following Exercises 2 and 3) uses a data set extracted by Bronwyn Hall from Compustat for the years 1988 through 1994. The data can be found in the file ex1-1.dat. A second *choice-based* sample with the same variables is found in the file ex1-2.dat. The variables are described below:

id	firm id
year	4 digit year, between 1986 and 1995
sic	4 digit sic code
ind	2 digit industry code
sales	annual sales (mill. dol.)
emply	employment (1000s)
invest	investment (mill. dol.)
rnd	R&D spending (mill. dol.)
cashfl	cash flow (= retained earnings + depreciation allowances) (mill. dol.)
kstock	knowledge stock (mill. dol.)
netcap	net capital stock (mill. dol.)
debt	long term debt (mill. dol.)
q	Tobin's q
loge	log (employment in 1000s)
rs	ratio of R&D invest to sales
cc	ratio of cashflow to net capital stock
drnd	dummy: zero R&D investment
exit	dummy: firm exits between year and year+1
grsales	growth rate in sales (percent) between year and year+1

There are 8031 observations in the file ex1-1.dat, and 1050 observations in the file ex1-2.dat. Both files can be read into a statistical program using an unformatted read that recognizes blanks as separators, or using the format 'f12.0,3f10.0,3g11.0,f9.0,11g11.0'. The data have a panel structure, with the same firm (identified by id) appearing in successive years until it exits or the last observed year is reached. However, for this exercise, ignore the panel structure and treat the observations across years *as if* they were independent.

A. Choice-Based Sampling. The file ex1-2.dat contains all 444 firm/year observations in which an exit between year and year+1 is observed, and a random sample of 606 firm/year observations from the population in which an exit between year and year+1 is not observed. The empirical probability that a non-

exiting firm/year in the population appears in this sample is 435/7587; thus this is a choice-based sample in which the qualification factors are 0.0553 for exit and 0.9447 for non-exit. In the choice-based sample, the sampling rates are 0.4229 for exit and 0.5771 for non-exit.

a. Using the Conditional Maximum Likelihood method for choice-based samples, estimate a logit model for exit as a function of loge, rs, drnd, cc, q, and an intercept.

b. Test the hypothesis that R&D does not influence exit (i.e., the true coefficients of rs and drnd are zero).

c. Estimate the population intercept term by correcting for sampling rates.

B. Exit Models in the Random Sample. The file ex1-1.dat contains a random sample of 8031 firm/years.

a. In this random sample, estimate a logit model for exit as a function of loge, rs, drnd, cc, q, and an intercept. Compare the results with the logit estimates in part A with a corrected intercept term. What might cause these estimates to differ, other than sampling error?

b. Estimate the logit model with separate intercepts for each year. Does the resulting pattern of coefficients make *economic* sense? Is there *statistically significant* variation across years?

c. Estimate a *probit* model with the same variables as in part a. Coefficients obviously differ by a scaling factor because the standard probit and standard logit formulas are scaled differently. Do you see evidence of differences in coefficient *ratios* in the two models?

C. Selection. Growth rate in sales is observed for a firm/year only if the firm does not exit before the following year. This creates a *selection problem* in trying to estimate the factors that determine the growth rate in sales in the whole population (before selection).

a. Regress grsales on loge, rs, drnd, cc, q, and an intercept on the sub-sample of non-exiting firms, ignoring the selection problem. What factors appear to be significant? Do they make economic sense?

b. Using results from the probit model in part B, add to the previous regression an estimated inverse Mills ratio term. Is there a statistically significant selection problem? Does it change the economic interpretation of the importance of the factors in the regression?

c. (extra credit) Correct the standard errors in the augmented regression b. for the impact of heteroskedasticity and an estimated explanatory variable.

d. (extra credit) Estimate the selection model, assuming normality, by maximum likelihood.