Attrition rate of UK companies registered at a similar location

November 16, 2024

We are interested in understanding the dynamics of change in the number of companies active at some location in UK. In particular the focus is on the rate at which active companies become not active.

Companies House data

Data on UK companies is available from the companies house website for free[1]. Among other things it contains the name of the company, the address, including the postcode, and the incorporation date. Companies House data includes a flag to indicate whether company is active. The data also includes dissolution date, but this column is not well populated, so there seems to be no accurate way of determining when company has seized to be active.

One can determine when company was incorporated, where it is registered, and whether it is active now.

Location proxy - simplified postcode

As a convenient proxy for location we shall use postcode of company address, but without the last two characters. For example, `SO17 1BJ` is the postcode of the University of Southampton (UK), whilst `SO17 1XJ` is a postcode of St Edmund's Roman Catholic Church, also in Southampton, located within 10min drive from the University. Thus grouping companies using *simplified postcode*, created by removing the last two characters, allows gathering addresses that are geographically close.

Since of particular interest here are locations with nontrivial changes in the number of companies, one can consider postcode locations with many registered companies, a sizeable proportion of which is no longer active (e.g. 20%). Limiting query to companies that have been incorporated after `2000-01-01` one finds[2]:

simplified post code	active company count	full company count	comment
`E14 9??`	6793	8575	Canary Wharf, London, UK
`NR1 1??`	2494	3255	Norwich, Norfolk, UK
`PO15 7??`	2395	3505	Fareham/Swanwick, Hampshire, UK

It is convenient to focus on a specific location as a first example. Canary Wharf is not representative of most of the UK, so let us focus on simplified postcode `NR1 1??` (Norwich). One finds that number of companies incorporated at that group of postcodes in the last 5 months was: 19, 22, 20, 21, 15. Comparing with deeper past one can see that number of companies incorporated monthly has remained relatively stable, at around 15 per month, during the period 2014...2020, and then grew over the period 2020...2022, to around 20 companies incorporated per month, then remained stable. See Fig. 1

Model for active company attrition

The aim of this work is to understand at which rate an active company becomes not-active. In reality, this is likely to be a complex process with multiple stages but for purposes of analysis on the level of common location we shall model the process as follows:

1. Company is incorporated and becomes active. The number of companies incorporated each month is known (e.g. Fig. 1)

[1] Guidance for downloading the data and doing initial preprocessing is available in the `README.md`

[2] All processing for this summary is contained in `company_count_growth_model.ipynb`

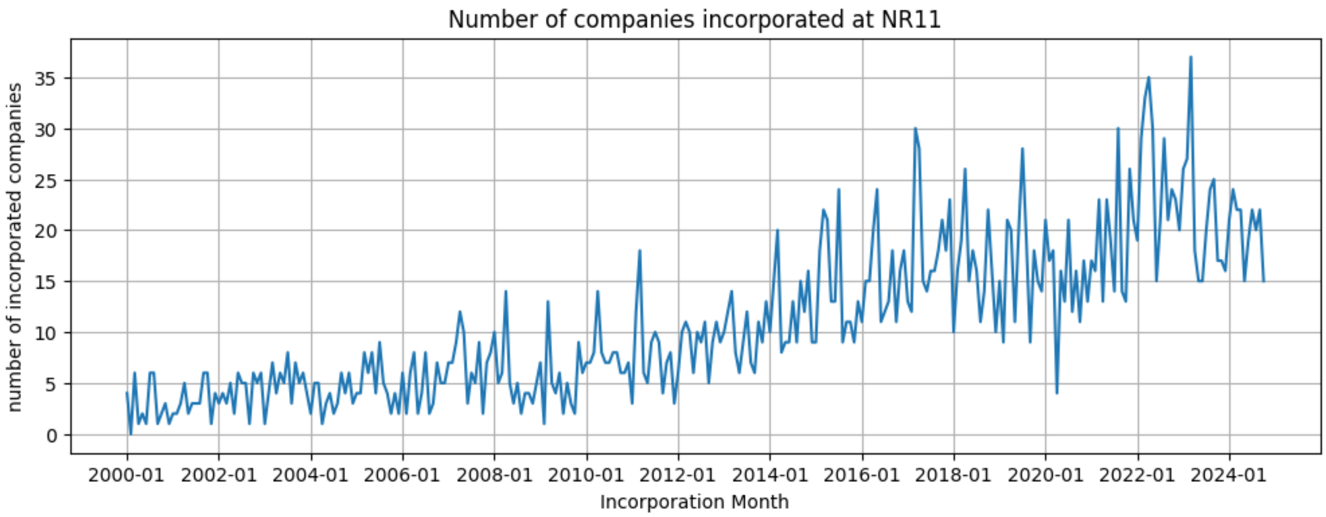

Figure 1: Number of companies incorporated at `NR1 1??` during each month from year 2000 up to present period. Includes companies that are no longer active.

2. Each month the probability of company to become not active is r, so if there were 1000 active companies in January, and $r = 0.1$, then 900 of them would be expected to remain active in February (on average)
3. Once company becomes not active it is no longer considered (no way back)

The inference we would like to derive is what is r, i.e. at which rate the become not-active (attrition rate). One complicating factor is that the longer the company exists the more time it has to become not active, but we only know the full number of not active companies. To express this process, it is convenient to use operator of binomial thinning ($\circ$). If n_{i-1} is the number of active in month $(i-1)$th, then:

$$(1-r) \circ n_{i-1} \sim Binomial\left(n_{i-1},\, 1-r\right)$$

Is a random variable that expresses the number of companies one would expect to remain active in (i)th month if the attrition rate is r. This way the number of companies active in each month will be:

$$\begin{aligned}
n_1 &= v_1 \\
n_2 &= v_2 + (1-r) \circ n_1 = v_2 + (1-r) \circ v_1 \\
n_3 &= v_3 + (1-r) \circ n_2 = v_3 + (1-r) \circ v_2 + (1-r)^2 \circ v_1 \\
&\vdots \\
n_k &= v_k + (1-r) \circ n_{k-1} = v_k + (1-r) \circ v_{k-1} + \cdots + (1-r)^{k-1} \circ v_1
\end{aligned}$$

Where v_i is the number of companies incorporated in (i)th month. We only observe the very last step, n_k. Expression $n_k = v_k + (1-r) \circ v_{k-1} + \cdots + (1-r)^{k-1} \circ v_1$ may be understood as:

- v_k: active companies incorporated this month
- $(1-r) \circ v_{k-1}$: active companies incorporated last month, which survived one month of attrition
- $(1-r)^2 \circ v_{k-2}$: active companies incorporated month before the last month, which survived two months of attrition
- etc...

The number of companies active in k-th month is therefore a sum of binomial variables. In general, this type of variable can be shown to follow Poisson binomial distirbution. Accurate methods for dealing with this distribution tend to be numerically ill-behaved. One can work around this by approximating each of the binomial variables in the sum by a normally distributed variable, and then using the properties related to summation of normal variables. Thus one can use

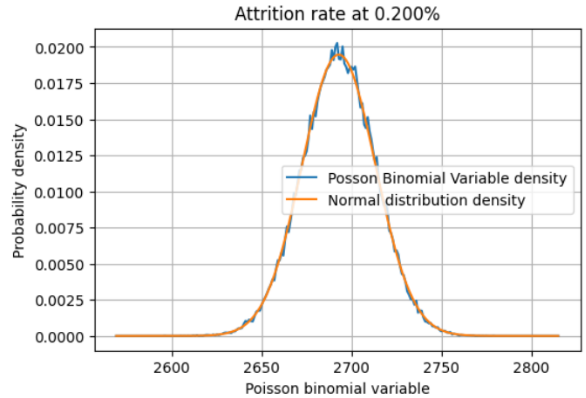

Figure 2: Fit between the approximate normal distribution, and the sampled distribution of the Posson binomial variable, for companies incroporated at `NR1 1` and for attrition rate of 0.2%.

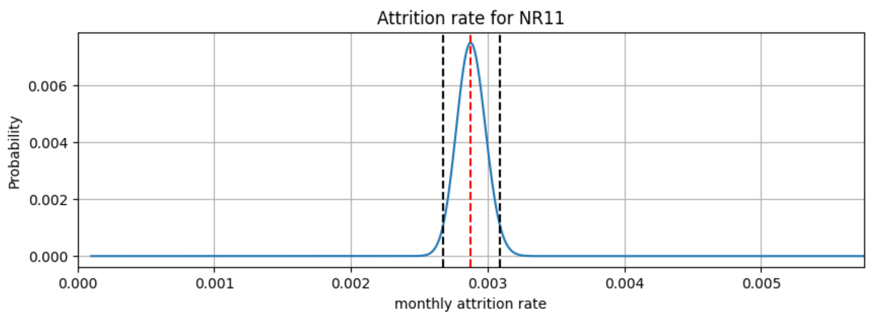

Figure 3: Posterior probability distribution for the attrition rate of companies at `NR1 1`. The maximum likelihood estimate is 0.288% and the credible interval is [0.268%, 0.309%].

an approximation:

$$n_k = v_k + (1-r) \circ v_{k-1} + \cdots + (1-r)^{k-1} \circ v_1$$
$$n_k \overset{approx}{\sim} Normal\left(\mu_k, \sigma_k^2\right)$$
$$\mu_k = v_k + (1-r) \cdot v_{k-1} + \cdots + (1-r)^{k-1} \cdot v_k$$
$$\sigma_k^2 = r \cdot (1-r) \cdot v_{k-1} + \left(1-(1-r)^2\right) \cdot (1-r)^2 \cdot v_{k-1} \cdots + \left(1-(1-r)^{k-1}\right) \cdot (1-r)^{k-1} \cdot v_k$$

Fortunately, it is quite simple to check whether the approximation makes sense. One simply assumes some attrition rate r, then samples from the underlying binomial distributions, and adds the results. The following Python code would accomplish this: Algo. 1.

One can plot normal distribution density and overlay it with sampled density. For example, Fig. 2 shows the fit for attrition rate $r = 0.002$. The fit is quite good. Similar results were obtained for other attrition rates in the range from 50% to 0.1%.

Having a tractable probability distribution for the number of companies at the final month allows one to estimate both the best fit of the attrition rate, as well as its probability distribution (treating the problem within Bayesian framework). For the chosen example of `NR1 1` the posterior probability distribution of the attrition rate is shown in Fig. 3

Algorithm 1 Sampling from distribution of a variable that is a sum of binomial random variables (Python)

```
attrition_r = 0.002 # assumed attrition rate
# company_count_arr is the number of companies incorporated in
# consequtive months

# compute individual terms in the sum for the approximate distribution
# for debugging, it is useful to keep things separated
loc_sum_arr = np.zeros(len(company_count_arr))
var_sum_arr = np.zeros(len(company_count_arr))
keep_prob_arr = np.zeros(len(company_count_arr))

# go over the months during which the companies were incorporated
# loop in reverse order
for i_month, cur_add in enumerate(company_count_arr[::-1]):
        # probability of keeping the company active
        # (given that it was incroporated in i-th month)
        keep_prob = (1 - attrition_r) ** (i_month)

        keep_prob_arr[i_month] = keep_prob
        loc_sum_arr[i_month] = cur_add * keep_prob

        if i_month > 0:
                var_sum_arr[i_month] = cur_add * keep_prob * (1-keep_prob)

#########

# pre-allocate arrays of zeros
approx_loc = np.sum(loc_sum_arr)
approx_var = np.sum(var_sum_arr)
approx_scale = np.sqrt(approx_var)

print(f'Approx location {approx_loc:.3f}; approx scale {approx_scale:.3f}')

#### sample poisson binomial variable
rng = npr.default_rng() # initialize random number generator from numpy.random
poibin_sample_count = 100000
poibin_sum_arr = np.zeros(poibin_sample_count, dtype=int)

# sample the from the poisson binomial distribution
for i_sample in range(poibin_sample_count):
        # sample from individual binomial distribution
        cur_sample_arr = rng.binomial(n=company_count_arr[::-1], p=keep_prob_arr)
        poibin_sum_arr[i_sample] = np.sum(cur_sample_arr)
```

Attrition rates for some simplified postcodes

Using the procedure described in the previous section one can now estimate the attrition rates for companies in various address groups encoded by simplified postcodes. Ordering simplified postcodes by the number of incorporated companies:

simplified post code	company count	active company count	rank	attrition rate (%)	credible interval (%)
`WC2H9??`	89,152	89,094	1	0.00230	[0.00180, 0.00302]
`EC1V2??`	64,477	62,363	2	0.00516	[0.00434, 0.00626]
`N17??`	45,799	45,701	3	0.00470	[0.00390, 0.00578]
...	...	...	...	...	...
`HR53??`	6,580	6,563	20	0.00635	[0.00421, 0.0107]
`HA11??`	6,681	6,451	21	0.0392	[0.0346, 0.0448]
`NW107??`	6,073	6,054	22	0.00698	[0.00471, 0.0114]
...	...	...	...	...	...
`NW110??`	2,682	2,678	100	0.00136	[0.000677, 0.004276]
`PE26??`	2,690	2,665	101	0.00997	[0.00702, 0.0152]
`E59??`	2,654	2,646	102	0.00396	[0.00227, 0.00869]

Limiting selected locations to those where the difference between incorporated and active companies is at least 20%:

simplified post code	company count	active company count	rank	attrition rate (%)	credible interval (%)
`E149??`	8,575	6,793	1	0.332	[0.317, 0.348]
`NR11??`	3,255	2,494	2	0.288	[0.268, 0.309]
`PO157??`	3,505	2,395	3	0.333	[0.314, 0.354]
...	...	...	...	...	...
`BN14??`	1,861	1,403	10	0.302	[0.276, 0.332]
`BR60??`	1,801	1,386	11	0.331	[0.300, 0.365]
`N200`	1,652	1,311	12	0.259	[0.233, 0.288]

One would expect there to be a relationship between the attrition rate and the percentage of active companies at some postcode location. Table below provides a summary[3] for few cutoffs:

share of incorporated companies that remain active[4]	typical monthly attrition rate
90%-100%	0.0025%
80%-90%	0.17%
70%-80%	0.29%
60%-70%	0.43%
50%-60%	0.73%

Conclusion

A method for estimating the monthly rate at which active companies become not active, the attrition rate, has been developed. Typical attrition rate for companies grouped by postcode can range between 0.002% per month up to nearly 1%.

[3]Summary for 80-90% share of active companies has been obtained by randomly selecting 10 postcode locations with more than 100 incorporated companies, and share of active companies that meet the 80-90% criterion, extracting 10 individual attrition rates, and then taking the average. Equivalent procedure has been applied for other rows.

[4]Incorporated after 2000-01-01 and remain active today