Initial empirical results from SCOMP data

September 20, 2025

In this document I will present what we are learnign from out empirical work. This is the continuation of the file which presents the initial datawork.

1 IE 5

1.1 Granularity of the insurer mortality tables

The logic in table 1 is that more sophisticated firms should have a statistically significant value of savings (val_uf_prima) because they include it in the mortality tables. Whereas less sophisticated firms do not include it and therefore the coefficient should be statistically insignificant. We rank firms according to sales and we show the firms with odd ranks (1,3,5,7,9, ..., 15), and we assume is a proxy of sophistication[1] There does not appear to be a trend in which firms with lower ranks have a smaller value of the savings coefficient. This could be caused by the fact that the model is misspecified, we are trying to approximate the pricing formula which contains thousands of coefficients (mortality rates) with 5 coefficients.

In table 2 we repeat the excercise but without controlling for savings under the logic that the R2 of more sophisticated firms should be lower because we are missing a key variables in the pricing formula. There does not appear to be a trend in which firms with lower ranks have a higher R2. Again, this could be caused by misspecification.

If more sophisticated firms are using t

[1] If creating better mortality tables is a fixed cost then it is natural to expect firms with more sales to have more granular mortality tables.

Table I: ratio on controls by rank_sales

	(1) ratio	(2) ratio	(3) ratio	(4) ratio	(5) ratio	(6) ratio	(7) ratio	(8) rati
val_uf_prima2	0.208***	1.362***	0.389***	0.300***	0.260***	0.998***	0.149*	0.06
	(0.052)	(0.054)	(0.051)	(0.056)	(0.053)	(0.069)	(0.081)	(0.06
male	-8.811***	-10.681***	-9.868***	-7.526***	-8.201***	-10.718***	-7.729***	-6.605
	(0.299)	(0.287)	(0.292)	(0.323)	(0.302)	(0.406)	(0.387)	(0.35
age_years	-5.721***	-5.504***	-5.792***	-6.071***	-5.923***	-5.478***	-5.554***	-5.674
	(0.045)	(0.046)	(0.044)	(0.046)	(0.045)	(0.057)	(0.073)	(0.05
A	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.00
	(.)	(.)	(.)	(.)	(.)	(.)	(.)	(.)
D	-1.028***	-1.296***	-1.509***	-0.689**	-0.690**	-0.404	-0.176	0.19
	(0.282)	(0.263)	(0.272)	(0.303)	(0.286)	(0.357)	(0.313)	(0.35
P	-0.057	-0.311	-0.207	-0.975**	0.101	-0.173	0.113	0.00
	(0.376)	(0.348)	(0.358)	(0.414)	(0.377)	(0.469)	(0.410)	(0.45
Constant	577.136***	563.811***	587.362***	601.633***	587.724***	559.175***	564.035***	572.95
	(2.971)	(2.870)	(2.771)	(2.962)	(2.815)	(3.624)	(4.579)	(3.22
N	18266.000	20151.000	21324.000	13485.000	19945.000	11870.000	10148.000	9475.
R-sq	0.658	0.656	0.646	0.709	0.655	0.649	0.662	0.71

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table II: ratio on controls by rank_sales

	(1) ratio	(2) ratio	(3) ratio	(4) ratio	(5) ratio	(6) ratio	(7) ratio	(8) ratio
male	-8.882***	-11.263***	-10.035***	-7.570***	-8.290***	-9.611***	-7.764***	-6.569***
	(0.299)	(0.290)	(0.291)	(0.323)	(0.302)	(0.402)	(0.387)	(0.356)
age_years	-5.684***	-5.271***	-5.722***	-6.027***	-5.879***	-5.338***	-5.532***	-5.664***
	(0.044)	(0.045)	(0.043)	(0.046)	(0.044)	(0.057)	(0.072)	(0.050)
A	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	(.)	(.)	(.)	(.)	(.)	(.)	(.)	(.)
D	-0.970***	-0.944***	-1.382***	-0.611**	-0.614**	-0.248	-0.164	0.206
	(0.282)	(0.267)	(0.271)	(0.303)	(0.285)	(0.360)	(0.313)	(0.357)
P	0.014	0.132	-0.080	-0.873**	0.186	-0.005	0.138	0.019
	(0.375)	(0.353)	(0.358)	(0.414)	(0.377)	(0.473)	(0.410)	(0.456)
Constant	575.315***	552.213***	583.787***	599.686***	585.535***	552.407***	563.024***	572.529***
	(2.937)	(2.879)	(2.734)	(2.943)	(2.781)	(3.625)	(4.546)	(3.205)
N	18266.000	20151.000	21324.000	13485.000	19945.000	11870.000	10148.000	9475.000
R-sq	0.657	0.645	0.645	0.708	0.655	0.643	0.662	0.717

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

A second approach is to:

1. Generate groups based on covariates $x = (age, savings, year)$ and $\hat{x} = (x, savings)$
2. Calculate the coefficient of variation for groups within the groups based on x and $\hat{x}$.
3. Variation in groups using x should be higher for more sophisticated firms since we are not accounting for savings. Moreover if we calculate the decrease in variation when using $\hat{x}$ instead of x then we should see a decrease for sophisticated firms but not for the other firms.

2 IE 6: Differences in Hazard rates across firms.

In this section we prove that there are diffferences in hazard rates between companies and they are not only explained by differences in observables.

Figure 1 shows the distribution of age at death.

Figure 1

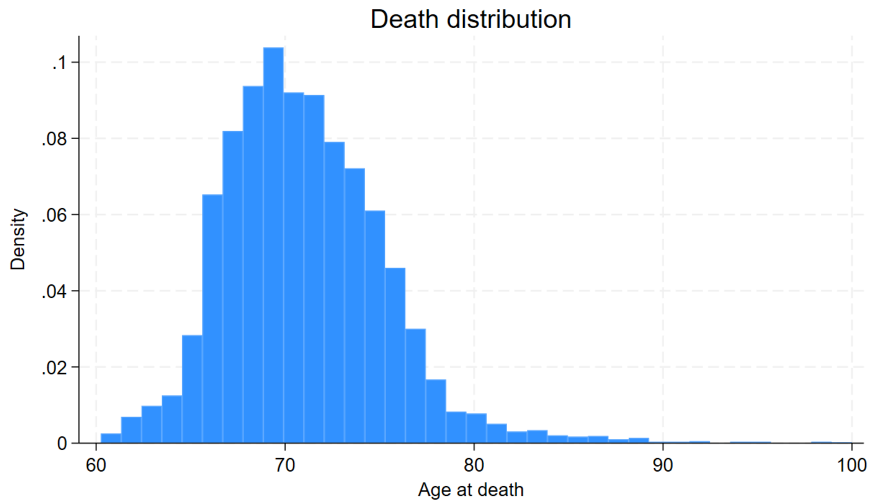
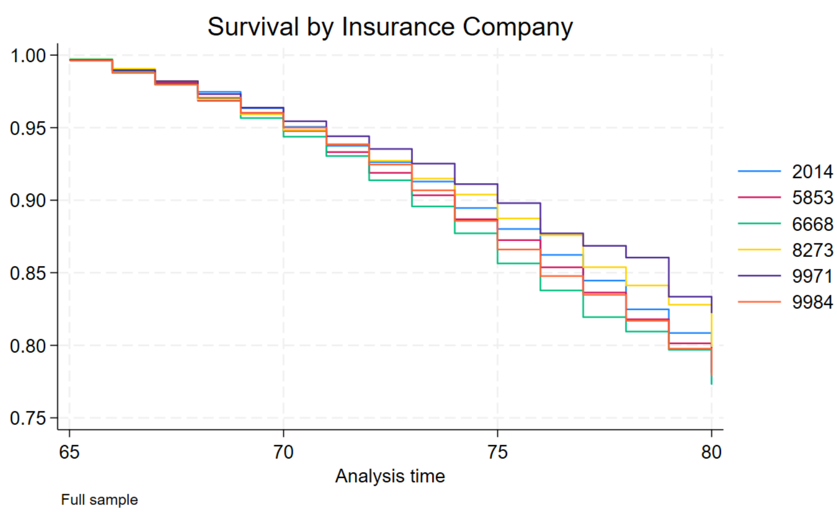


Figure 4 shows the survival function by each firm. There are differences among firms.

Figure 2



One possibility is that the differences represent differences in observable characteristics, not on unobservables.

As a first step we look only at the sample of men (to control for gender differences) and we also group buyers in 3 income groups (to control for income differences).

Figure 3

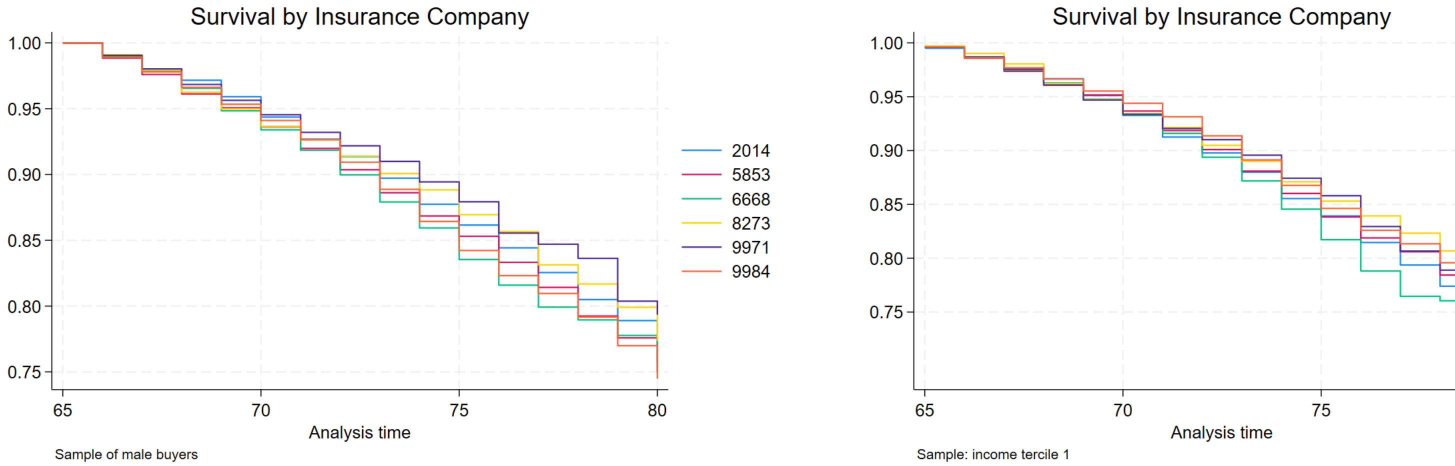
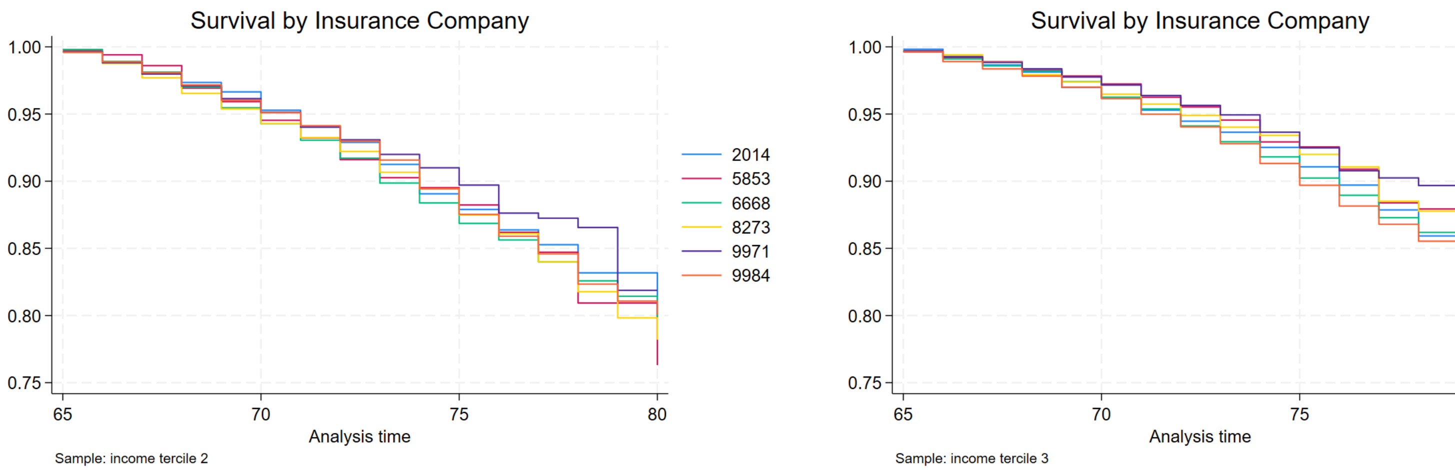


Figure 4



The differences once we use this subsamples continue being significant, the exception is income group 2 for which they are not significant any more (p-value 0.15).

To do a more formal test we use a Cox proportional hazard model.

We estimate Cox PH models of the form

$$h_i(t \mid X_i) \;=\; h_0(t)\,\exp\bigl(X_i^{\top}\beta\bigr),$$

where $h_0(t)$ is an unspecified baseline hazard and X collects the screened covariates (sex, premium level, intermediation type, guaranteed-months dummies). We then add company indicators to test for across-firm differences:

$$\textbf{Baseline PH:} \qquad h_i(t \mid X_i, \mathrm{Firm}_i) = h_0(t)\,\exp\bigl(X_i^{\top}\beta + \gamma_{\mathrm{Firm}_i}\bigr).$$

To relax proportionality for the *covariates* (but not for firm dummies), we also estimate a model with time-varying coefficients using $\ln t$ interactions (Stata's `tvc()` with `texp(ln(_t))`):

$$h_i(t) = h_0(t)\,\exp\bigl(X_i^{\top}\beta + (\ln t)\,X_i^{\top}\delta + \gamma_{\mathrm{Firm}_i}\bigr).$$

This allows the effect of X to change with time while still testing whether firms differ by a *proportional* shift.

After each Cox fit with firm dummies, we run a Wald test of

$$H_0: \ \gamma_2 = \cdots = \gamma_J = 0,$$

we get 0.01 and 0.01 as the p-values.