Assicurazioni Generali
GROUP INSURANCE RESEARCH

# The S-Curve and Reality

2016

*Giovanni Millo*

WORKING PAPER WP1/16

# The S-curve and reality

Giovanni Millo
Group Insurance Research, Assicurazioni Generali SpA

November 10, 2015

**Abstract**

We challenge the common wisdom that the income elasticity of insurance be higher, ceteris paribus, in developing countries (the so-called S-curve hypothesis). Focusing on non-life insurance, we show that the available evidence is contradictory and heavily dependent on methodology. Based on a recent approach to consistent inference on the income elasticity of insurance, we show counterexamples to the theory. Although not supporting it in general, we argue that it could still be relevant for explaining the behaviour of particular lines of business.

## 1 Introduction

In this paper we challenge what is arguably the most popular model of the evolution of insurance markets, the S-curve by Rudolf Enz (2000). Based on earlier work by Carter and Dickinson (1992), Enz (2000) put forth the hypothesis that insurance penetration may be approximated, country by country, by a logistic function of economic development, so that the scatterplot of current insurance penetration versus GDP per capita in the countries of the world may be effectively interpolated by means of a logistic (or sigmoid) curve. From this relationship one can also calculate the income elasticity of insurance at a given level of per-capita income. An implication is that the development of insurance is slower in the first stages of economic development, then grows more than proportionally and ultimately slows down again.

The idea that the evolution of insurance may be stylized this way is a powerful and elegant hypothesis which has become very popular among practitioners. It bears resemblances to a number of typical applications of the logistic function in various sciences – from models of population growth (Verhulst, 1845, 1847) to the diffusion of innovations in the economy (Ayres, 1990a,b) – which add to its appeal. Moreover, such a relationship, if confirmed by the empirical evidence, would provide a natural characterization of the evolutive pattern of insurance penetration in time and a solid link with that of income, allowing consistent forecasting of the market conditional (solely) on economic development at large.

Despite a number of weaknesses, the hypothesis has become so ingrained in the common sense of both academics and practitioners as to be often taken for granted. In time, while this theory was gaining widespread acceptance, any evidence of non-linearity has been considered by-and-large consistent with it[1]. We will instead argue that most of the evidence actually suggests that – net of individual idisyncracies, deterministic trends, international common factors and substantial statistical uncertainty – linearity can still be the best bet.

Enz (2000) underlined that the often-used linear approximation to the income-insurance relationship is unlikely to hold too far out of any given sample, as greater-than-one elasticities would eventually lead to indefinitely high values of the penetration rate. Therefore, the use of non-linear functional forms capping the elasticity over a certain value of income is warranted. From this viewpoint, as others have observed (Lee and Chiu, 2012), the logistic curve, one among many functional forms downweighing the elasticity at extreme values of income, is not necessarily the most appropriate one: functional form has to be tested for. Although in principle we agree, we will ultimately argue that the argument might not be relevant if, as per some recent evidence, elasiticities are actually not greater than one.

Moreover, the S-curve of Enz (2000) is illustrated on cross-sectional data but estimated on pooled cross-section and time series data. The author actually recognizes the importance of allowing for country-level heterogeneity and therefore considers the residuals from the curve's fit regressing them against individual, time-invariant effects and time trends. Given these concerns, the interpretation of the standard S-curve – a statistical interpolation of the existing penetration rates in a pool of countries over different periods – in marginal terms is not straightforward. Also, from a modelling viewpoint, heterogeneity has important consequences on the consistency of estimators. For this reason, following Millo (2014), we will change perspective estimating the average income elasticity of insurance from individual time series in the framework of Pesaran (2006), which provides consistency vs. the above issues; then we will revert to testing the predictions of the S-curve hypothesis on the results.

In order to test the implications of the S-curve theory, we will try to be precise about its meaning. In fact, the literature is often unclear about what the S-curve is meant to be: a descriptive graph depicting a regular pattern in the space spanned by GDP and insurance penetration, or a theory of insurance elasticity as a function of income (alone)? This distinction between statistical fitting of positive data and econometric modelling of a (partial?) elasticity as a conditional expectation has, in our opinion, been neglected in previous studies, where the two things often confuse. If one assumes that $p/y = f(y) + \varepsilon$ is a sufficient model, then fitting makes sense, but still the descriptive curve as a snapshot of a moment in time has cross-sectional nature, while fitting it as a panel requires both a model of its evolution in time and consideration for the individual dimension, both in terms of unobservable individual heterogeneity (individual effects) and of error clustering in the sense of Moulton (1986, 1990). "Pooled" S-curve plots where snapshots across space, or equivalently trajectories in time, are superposed and fitted all together, are common but in our view problematic.

Obviously, while the descriptive evidence is to be taken per se as an interesting stylized fact, the model interpretation has the benefit of providing testable hypotheses. In particular, as Enz (2000) himself shows, the characterization of

[1] See e.g., for the case of Life insurance, Chang and Lee (2012, comments to Figure 1 and Table 2), where most of the evidence presented (only one threshold, income elasticity three to five times higher in the richer half) actually points towards a strongly hyperbolic shape.

penetration ratios as a logistic function of income implies a hump-shaped distribution of elasticities along the income scale. Individual estimates of elasticities from time-series models will be the observable feature on which we will test the predictions of the theory. Attempts at falsification, in a Popperian sense, will depend on how strict a definition of said distribution one employs. By the nature of the hypothesis considered, "falsification" will have to proceed by looking for evidence of either a particular non-linear shape, or generic nonlinearity, against the null of linearity. Statistical uncertainty will therefore be on the side of the simpler hypothesis, in a sense leaving the burden of proof to the defendant: which is somewhat unfair. For this reason, in the following we will also consider very loose versions of the hypothesis.

We will focus on the non-life sector, although our methods of analysis can be carried over to life insurance. The evidence will turn out not to fully support even the weakest nonlinearity hypothesis. Hence we will conclude against the empirical plausibility of the S-curve theory, when applied to the non-life sector as a whole. We will nevertheless close the paper arguing that, if considered separately, some individual lines of business (property, liability, non-motor as a whole) do actually show signs of nonlinearity compatible with an S-curve-like behaviour, and that the seemingly linear relationship between total non-life insurance and income might actually result from the compensation between different lines developing more or less quickly at different stages of economic development. This conjecture, which is based on incomplete and insufficient evidence, will be reported only as a possible line of further research.

# 2 Insurance penetration in the world: the Sigma dataset

The standard source for premium income in the world's markets has long been the Sigma dataset, by Swiss Re Research and Consulting (sigma, various issues). This work is no exception. The dataset regards non-life annual premium volume for direct business, and covers a maximum of 95 countries 1970-2010. It is highly unbalanced, as its geographic scope vastly increased through time: in 1970 there are 35 countries; 56 in 1980, 68 in 1990; since the Nineties Eastern Europe is included and in 2000 only Serbia and Liechtenstein are missing (the latter two have been dropped from the dataset as having less than nine observations in time). Bar Angola and Botswana, all other 91 countries have at least 15 observations.

The Sigma dataset also provides statistics on insurance penetration over GDP.[2] At world level, insurance penetration over GDP has remained approximately constant in the 2.4-2.8 range throughout the last three decades. This is largely the result of the overwhelming weight of developed markets, of which only European ones have shown a recognizable upward trend, compensated for by a slight contraction in the US and especially in Japan during its two "lost decades" (see Table 2). So while there is little doubt that the share of insurance over GDP has been on an increasing pattern throughout the developing world

[2] In the following, also for modelling purposes, we take GDP data from the Version 7 of Heston et al (2011), as in the R package 'pwt' (Zeileis and Yang, 2012). Transformations to constant prices and purchasing power parities for GDP and premiums have been done as in Millo (2014).

| | 1970 | 1980 | 1990 | 2000 | 2010 |
|---|---|---|---|---|---|
| North America | 4.02 | 4.46 | 5.03 | 4.26 | 4.46 |
| Latin America and Caribbean | - | 0.90 | 1.10 | 1.29 | 1.53 |
| Western Europe | 2.12 | 2.36 | 2.60 | 2.86 | 3.18 |
| Eastern Europe | - | - | - | 1.49 | 2.05 |
| Middle East and Central Asia | - | - | 0.42 | 0.78 | 1.11 |
| Japan | - | 1.73 | 2.52 | 2.26 | 2.11 |
| South-East Asia | - | 0.34 | 0.58 | 0.71 | 1.07 |
| Africa | - | 0.59 | 1.09 | 1.12 | 1.13 |
| Oceania | 2.27 | 2.75 | 3.05 | 3.25 | 2.99 |

Table 1: Insurance penetration by macroarea, 1970-2010. Source: Swiss Re, Sigma database.

for the last 40 years, in developed countries it seems to have been stabilizing for a long time, the exception of Europe being probably attributable to the need to surrogate in the ongoing retreat of the welfare state. Therefore, at a first descriptive glance, the intuition of rich, saturated markets as opposed to younger ones where there is plenty of opportunity left for growth seems to be confirmed. Nevertheless, it remains to be ascertained whether this is the effect of rising income or of other factors, perhaps institutional.

Returning to our main research question, however, it is clear that the situation is unlikely to be uniform across the world. As observed, it has been argued (Enz, 2000) that the income elasticity of insurance follows a nonlinear pattern across countries with different levels of development, being comparatively low in the less developed markets, then growing together with the economy and finally, beyond a certain stage of development, falling again. In the following, we provide timelines linking all subsequent points in the space defined by penetration and development. If the descriptive S-plot is a snapshot of the position of different countries in this space at a given moment *in* time, the timeline depicts instead the evolution of one single market *through* time. Without the need to assume a steady evolution towards development, for any growth pattern of per-capita income through time, if the S-curve hypothesis holds then the timelines should mimick the sigmoid shape of the original S-curve, although limited to the domain spanned by the evolution of the country over the observed sample. In other words, a developed country can only be expected to cover the rightmost, convex-to-flat part of an ideal S-curve, and so on.

Descriptive evidence is mixed, as each case study is heavily influenced by idiosyncratic factors. Looking at the patterns of insurance penetration versus real per capita GDP (PPP weighted) in Figure 1, we can see examples of a mature market undergoing various economic cycles (USA), another one still with a clear upward trend (Germany) and one with the sector first rising in importance, then shrinking as the economy as a whole fails to grow (Japan); lastly, one pertaining to a country which can be said to have passed most of the stages of economic development in the last 40 years (South Korea), showing an ever-rising tendency interrupted only by the setback of the 1997 Asian crisis.

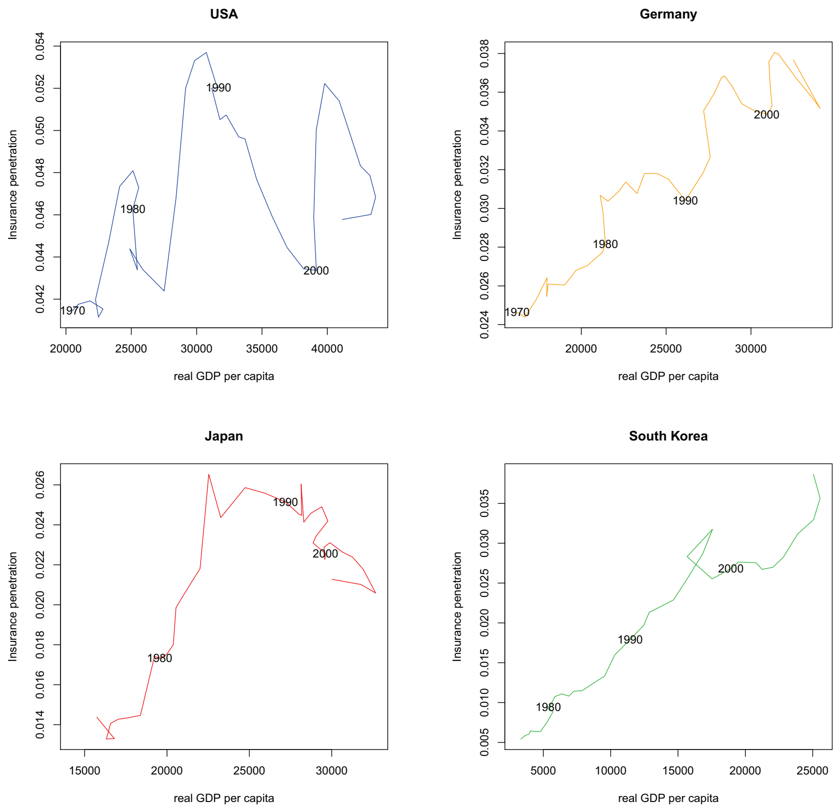


Figure 1: Relative patterns of insurance penetration and economic development (measured as real GDP per capita at PPP in international dollar) in four developed countries, ca.1970-2009.

# 3 The S-curve: predictions and evidence

The original S-curve hypothesis states that insurance penetration may be approximated, country by country, by a logistic function of economic development:

$$\frac{P}{Y} = \frac{1}{c_1 + c_2 c_3^Y} \tag{1}$$

where $P$ are insurance premiums per capita, $Y$ is real GDP per capita, so that the scatterplot of current insurance penetration versus GDP per capita in the countries of the world may be effectively interpolated by means of a logistic (or sigmoid) curve. From this relationship one can also calculate the income elasticity of insurance at a given level of per-capita income as:

$$\eta_{P,Y} = 1 - \frac{c_2(c_3^Y)Y\dot{l}n(c_3)}{c_1 + (c_2 c_3)^Y} \tag{2}$$

where $P$ are insurance premiums and $Y$ is real GDP per capita. Hence, if $c_3 \leq 1$ (which is considered the normal case), penetration starts from a minimum of $1/(c_1+c_2)$ and grows towards the asymptotic value of $1/c_1$, first with steepness growing up to the inflection point, then decreasing. Translating in terms of elasticity, this latter grows with income up to reaching a maximum at $Y^*$ : $1 + Y^* ln(c_3) + \frac{c_2 c_3^{Y*}}{c_1} = 0$, then decreases reaching the same value it started from, meaning that the development of insurance is slower in the first stages of economic development, then grows more han proportionally and ultimately slows down towards the same elasticity it begun with. Enz (2000) estimates this model on pooled cross-section and time series data and shows how it can predict the elasticity of one country at one given level of development.

In the following we will question the logistic interpolation. The biggest challenge to the S-curve hypothesis, nevertheless, is its incompleteness: if it is meant to be a model, i.e. it is to be interpreted as a conditional expectation of premiums given income, then all statistically relevant information must have been included. In other words, given income, insurance premiums must be conditionally independent from every other characteristic of a country (and, if considering panel data, time period). If this condition is violated then estimates are inconsistent. For this reason, we will estimate the income elasticity of insurance controlling for such features by means of a recently developed estimator, and then come back to test the predictions of the S-curve hypothesis on the results. But first we will address the logical argument underlying the logistic hypothesis and the descriptive evidence that inspired it.

## 3.1 The empirical S-curve as a stylized fact

The S-curve hypothesis was born as a stylized fact and from a theoretical consideration. The stylized fact is based on the observation of the scatterplot of insurance penetrations versus per capita GDP in the cross-section of the World's countries. Moreover, Enz (2000) observes that insurance penetration cannot grow forever, as it is naturally capped by reasonable limits to the importance of insurance in the economy.

**The share of GDP cannot grow forever** The logical argument in Enz (2000) that the share of insurance on GDP must admit an upper limit scarcely allows discussion, but its relevance at the current time is far less evident, especially when considering that premium income is not an appropriate measure of the sector's contribution to GDP. In fact, only the share of premiums related to intermediation and risk-bearing services enters the value added of (non-life) insurance, which can be measured as the sum of employee compensation and industry profits, or as total premium income minus claims[3]. In any case, the expense ratio (usually between 10-40% of premiums) can be considered as a proxy for the magnitude. Therefore, the actual share of value added for even the most developed non-life insurance markets stands much lower than the corresponding penetration ratio. Consider the case of the European Union. In fact, in 2011 the share of value added of the *financial and* insurance sector as a whole (5.7%, source: Eurostat) was much lower than the ratio of insurance premiums alone over GDP (7.9%, source: Insurance Europe). A penetration ratio of 3.2% therefore probably puts the share of non-life premiums in value added in the region of one percent. Doubling or even tripling it would hardly displace the rest of the economy.

Moreover, Enz (2000, Introduction) starts from the assumption that "income elasticity [is] generally greater than one" to conclude that under a linear model "there are no limits to insurance penetration"; his point is true but loses relevance if elasticity is actually not different from one, as per the conclusions of Millo (2014).

**Functional form** The logistic has a number of features which might be inappropriate, first of all its symmetry, meaning that markets will be as fast in picking up from low penetration at the earliest stages of development as they will be in saturating after having reached maturity. For the sake of example, the evolution of markets might be better described by an asymmetric function, as long as its slope eventually becomes decreasing at the right end. The behaviour in the leftmost range seems to be the easiest part to verify empirically, as there is plenty of developing countries in the potential sample; the hardest part is, instead, to infer on the possible shape of the curve outside the right boundary of the sample on the basis of observations on a few big industrialized countries and some small oil exporters and city states: has the turning point been reached already, or will penetration grow further as the economy does?

At first consideration, the empirical support for a logistic-shaped relationship looks less robust than the theoretical a-priori grounds suggesting it. From a purely descriptive viewpoint, plotting insurance penetration versus per capita GDP at market prices rather suggests, if any, a logarithmic shape (see Figure 3, left); doing the same with PPP-weighted per capita GDP (idem, right) suggests a broken linear interpolant with two kinks, one after the cluster of poorer countries, one before developed ones, with a gentler sloping part in the middle. Spline smoothers (see, again, Figure 3) point at the symmetric of a logistic curve, with slopes actually higher at both ends of the data range.[4]

[3]On the subject, see e.g Hornstein and Prescott (1991a,b) and Sherwood (1999).

[4]The striking dissimilarity with Figure 3 in Enz (2000) is due to his using a logarithmic scale for the horizontal axis only, while we have preferred to preserve the original proportions. In this last form, the scatterplot assumes an hyperbolic shape.

For all the strength of the argument, the evidence that the rightmost bend has been reached and passed by developed countries is far from compelling. A look at Enz's scatterplot (see the original as Enz, 2000, Fig.3) already reveals that most points follow a hyperbolic shape rather than a logistic one, the penetration in high-income countries showing scant evidence of moderation[5]; see also the comments to Figure 2 in Outreville (2013) (although these are referred to total insurance). Moreover, the apparent curvature is due to taking the log of GDP against the linear scale for penetration[6].

All in all, the issue of functional form may look unresolved given this preliminary evidence. Hence one may want to avoid imposing a logistic shape from the beginning (see also Lee and Chiu, 2012). Fitting a non-parametric spline on the very same cloud (Sigma database, 1998) yields an almost-linear shape in a standard graph of penetration versus GDP per capita, an hyperbolic one when taking GDP on a log scale (Figure 3.1, respectively left and right panel). The

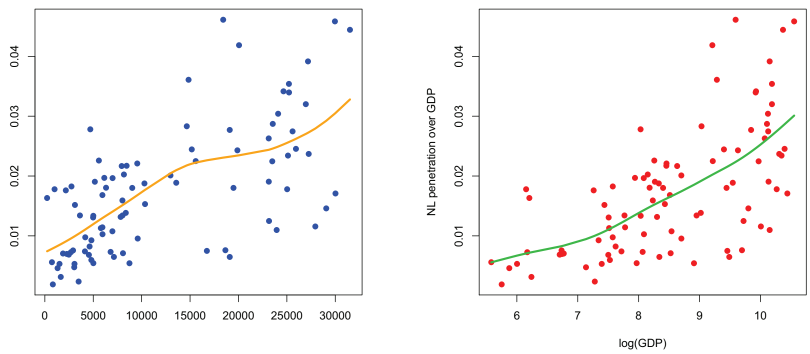


Figure 2: S-curve plots on same data as Enz (2000), linear scale (left) and log-linear scale (right), interpolated by Spline smoothers.

idea that elasticity moderates in mature markets, although logically sound, is unsupported by descriptive evidence.

## 3.2 Graphical tools: descriptive vs. marginal S-plots

In this section we have investigated the shape of the penetration - GDP per capita relationship at a static level, simply plotting the available data. In subsequent ones, we will argue that the correct perspective from which to look at the elasticity is a time series one, and consequently bring the focus of the analysis from pooled premiums-to-GDP ratios towards the coefficient of income in a by-country regression model of insurance expenditure. We will also highlight the importance of considering determinants other than income for obtaining an unbiased estimate of income elasticity. But first we will argue that graphical

[5]The few high-income, low-penetration points on the right end of the graph are from "untypical" countries: either Islamic oil-exporters, where income is high but insurance penetration is lower, not last for religious reasons (see Grace and Skipper, 1991), or small city-states and financial centers like Luxembourg, Singapore and Hong Kong.

[6]Notice that in the model, unlike in the plots, GDP is not logged.

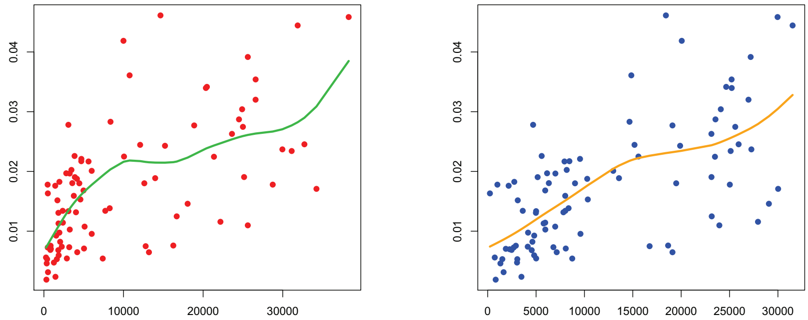


Figure 3: Enz's (2000) S-curve plot: insurance penetration versus 1998 per-capita GDP at market prices (left) and at PPP (right), linear scale on both axes and interpolated by spline smoothers. On the far right, Luxembourg is excluded as an outlier and to preserve the ideal banking ratio.

assessments too have to be based on the marginal effect of income on insurance development rather than on snapshots of the existing levels, which last might have been arrived at in different, although observationally equivalent, ways than through an S-curve-like evolutionary pattern.

To this end, we will distinguish between two possible graphical tools, which we will both call *S-plots*: a "descriptive S-plot" plotting insurance penetration (premiums over GDP, on the vertical axis) against the level of development, measured as GDP per capita; and a "marginal" S-plot where the (partial) elasticity of insurance premiums to income is, again, plotted versus the level of development.

As already observed, under the S-curve hypothesis the points in the former graph should follow a sigmoid shape, those in the latter a humped one (like, respectively, Figures 3 and 4 in Enz, 2000, but for the log scale). Of course, while given the data the descriptive S-plot is uniquely determined (up to transformations) the shape of the marginal S-plot will be dependent on the underlying model employed to estimate the individual coefficients. For this reason, we will now address the issue of consistent modelling of the insurance-income relationship and proceed to formal statistical testing, before coming back to graphical assessments in the last part of the paper.

# 4 Model-based evidence

We will now set the S-curve theory in the form of testable hypotheses. We will distinguish a stricter version of the theory from weaker ones, ranging from imposing a logistic form to simply checking for non-linearity.

We will then review the existing literature, commenting on the consistency of the results with the above hypotheses and on the empirical issues of the estimators employed, before setting out our preferred specification and choice of estimator, on which the main part of the paper will be based. Consistent estimation of the elasticity of premiums to income will be done country by country and at a panel level, as done in Millo (2014), using a common-factor augmented pool of individual time series. A reassessment of the relationship between the income elasticity of insurance and the level of economic development of a country will follow on this new basis, against which the predictions of the S-curve model will be tested.

## 4.1 The S-curve hypothesis in broad and strict sense

*Stricto sensu*, the (logistic) S-curve hypothesis implies that the elasticities be a precise function of income levels: that the function be bell-shaped and symmetric, so that at the beginning and at the end of the income spectrum the elasticity be lowest, while highest in the middle. This, which we can label "S-curve hypothesis in strict sense", can be specified as follows:

- *H1: the relation between the (per capita) income level and the income elasticity of insurance is well described as $dp/dy = f(y) + \varepsilon$, where $f$ is the derivative of the logistic function and $\varepsilon$ a well-behaved stochastic disturbance*

Yet this is in all likelihood too strict a requirement, even allowing for a good deal of (non-systematic) statistical uncertainty. A more reasonable, and natural, benchmark requirement for consistency with the S-curve theory is that the elasticity be *some* non-linear function of income levels that be at least qualitatively consistent with the strict S-curve hypothesis. Let us specify an "S-curve hypothesis in broad sense" as follows:

- *H2: The income elasticity of insurance is systematically higher in developing countries both with respect to developed and to underdeveloped ones*

Lastly, a minimal necessary – but not sufficient – condition for compliance with any version of the S-curve theory is the following

- *H3: There is some systematic relation between the income level and the income elasticity of insurance*

where of course some systematic behaviour may also be inconsistent with $H1-2$: e.g., a linear (increasing or decreasing) relationship, or different kinds of nonlinearity.

These benchmarks have been presented from stricter to looser, i.e. ordered by logical implication, so that if the evidence is not consistent with the latter, it is against the former as well. In the following we will review the previous evidence for consistence with $H1-3$, then present our own attempt at measuring the income elasticity conditional on income levels and evaluate $H1-3$ on this new basis.

## 4.2 The S-curve as an econometric model: a survey

In the following we review the previous literature for consistency with $H1-3$.

There have been two main approaches at econometric estimation of S-curve-like relationships: either parametric estimation of a logistic function or of threshold models allowing for regime changes. Enz (2000)'s original contribution fits a logistic model to pooled data on 88 countries over the years 1970-98. His estimates yield a maximum elasticity of 1.7 for non-life at 8900 PPP weighted 1997 US dollar (2.3 for life insurance). While the functional form is postulated to be a logistic on theoretical grounds and not compared against viable alternatives, he acknowledges the possible misspecification from not controlling for country fixed effects (see Enz, 2000, Footnote 4). Zheng et al (2008) apply Enz's model to more recent data (95 countries, 1980-2006) in order to draw long-term predictions on the development of the Chinese market based on its path relative to the benchmark of the S-curve. In fitting the logistic curve, they do not test the functional form neither do they control for any heterogeneity, serial or cross-sectional correlation, nor do they discuss stationarity. Maximum elasticity is 1.425 at 7531 constant 1990 USD.

Lee and Chiu (2012, Introduction) criticize Enz (2000) and Zheng et al (2008) for imposing a logistic instead of testing for the optimal functional form. They employ a smooth transition model discriminating between linear, hyperbolic and logistic functional forms on the basis of data. The model for non-life premiums turns out to be nonlinear with two regime changes, but the difference with respect to a linear form is very small in magnitude. Moreover, regimes are inverted with respect to the logistic of Enz (2000): see their Figures 1 and 2. Importantly, unlike the previous work, they allow for country heterogeneity through fixed effects; elasticities fall, which is unsurprising considering that fixed effects are probably going to account for a number of development indicators which are otherwise omitted from the model and are positively related to income. Non-life premiums elasticity is, respectively, 1.039 and 1.08 in the two regimes, hence they claim that NL is a luxury good and that this is consistent with Enz (2000) (at 1.5) and Zheng et al (2008) (1.425). While formally correct, such claims neglect the different magnitude of the former estimates and the latter, and rely heavily on the precision of the estimates. As will be discussed below, there are reasons to advocate the use of more tolerant standard errors, leading in all likelihood to wider confidence bands which would not allow to exclude 1 (i.e., the hypothesis of insurance as a normal good) from the confidence interval of the estimated elasticity. Moreover, the pattern of elasticity vs. GDP levels follows an inverted "U" shape (see p.252 and Fig.2), which – apart from the minimal range spanned (1.055 to 1.08) – is the opposite with respect to the predictions of the S-curve model. Hence, despite having the expected double threshold, the functional form is ultimately *not* consistent with the S-curve hypothesis (it is, though, with the specular shape in Figure 3.1, left panel). As a robustness check, Lee and Chiu (2012, p.254) also do a separate estimation by (linear) 2SLS on subsamples of developing and developed countries, finding that the elasticity of NL premiums in developed countries, at 1.073, is higher than those for emerging countries, at 1.016: again, a 0.057 difference.

Although considering individual heterogeneity and testing for the functional form can be considered big steps forward with respect to the earlier literature, there are problematic aspects in Lee and Chiu (2012)'s work too, testified by their own diagnostics.Cross-sectional dependence is found in the descriptive statistics (see Table A2) and considered when using robust unit root tests (Table A3), but not when estimating the model. Hence, the latter is misspecified to some extent: at a minimum, if there is neglected cross-sectional dependence in errors but no (endogenous) time effects or common factors, such misspecification is probably leading to overoptimistic standard errors and hence to exceedingly narrow confidence bands for the parameters, see above. Secondly, their claim that real GDP and premia are stationary is problematic at best. The critical values for the CIPS test given in Lee and Chiu (2012, Table A3) are those excluding both trend and intercept (see Pesaran, 2007, Table 3a); allowing for an intercept, let alone for trends, already reverses Lee and Chiu (2012)'s results, supporting nonstationarity (compare their Table A3 to Pesaran, 2007, Tables 3b, 3c). In fact, Millo (2014, p.14) finds evidence of unit roots in premiums even when allowing for trends. As for GDP, it is common wisdom that it can at most be considered *trend-stationary*[7]: but then, one should allow for (individual) time trends in the model too.

Summing up – and skipping the unit root issue altogether – the evidence of nonlinearity, although statistically significant, is very small in absolute value. The statistical significance comes in all likelihood from the narrow confidence interval estimates associated with the assumptions of homogeneous coefficients (pooling) and independent and homoskedastic errors, not allowing either for cross-sectional correlation in the error terms – yet testified by the CD tests in their Table A2 – or for the serial correlation induced by the time-demeaning of variables they employ to eliminate fixed effects[8] (as mentioned in 3.3 on page 250). The statistical significance of the small difference between the two regimes is unlikely to survive if taking these features of the data into account.

All that said, we tend to attribute the sharp evidence of non-linearity in the literature to neglected heterogeneity, common factors and individual trends. It is telling that when controlling for fixed country effects the difference between country groups (regimes) becomes so small even in Lee and Chiu's model. Pooled homogeneous panel models usually deliver sharp results (narrow confidence bands, well defined functional forms); nevertheless, they are at far greater risk of misspecification than can be a relatively innocuous linear approximation in the context of a heterogeneous and factor-augmented model.

In the following, we will adopt a consistent approach to the estimation of the elasticity of premiums to income, country by country and at a panel level, based on a heterogeneous linear panel model augmented with common correlated effects (Pesaran, 2006) as done in Millo (2014). Such model produces a population of estimated elasticities $\beta_i$ for each individual unit (here: country) in the sample (see the details in Millo, 2014, Section 3), so that the distribution of the elasticities can be assessed, *a la* Enz (2000), versus per capita income.[9]

### 4.3 The income elasticity of insurance in a pooled time series perspective

Enz (2000) himself acknowledges the influence of *all other factors*, the role of individual heterogeneity and the possibility of (individual) trends. For all these reasons, a multiple regression framework is in order if we want to infer about the relationship between income and insurance net of other potentially confounding

[7]The question whether there is a unit root in real GDP has been debated since Nelson and Plosser (1982); see the review in Papell and Prodan (2004). Ever since, the debate has been between unit roots or *trend*-stationarity.

[8]See Wooldridge (2002) on how subtracting the time means induces serial correlation with a coefficient of $-\frac{1}{T-1}$ if the original errors were serially independent.

[9]The short section summarizing the methodology which follows is based on Millo (2014, Section 3).

factors. Based on a regression of (per capita) insurance consumption on GDP *and other control variates* one can directly estimate the *partial* income elasticity of insurance as the regression coefficient of GDP.

Given that our subject of interest will be a regression coefficient, we are left with the choice of the relevant empirical setting. To infer about income elasticity across different stages of economic development, we must first choose a sample that spans them all, either across countries or in time. Being able to observe the behaviour of the insurance market during the transition a country from underdeveloped to developing, and then developed, is far less likely to be feasible, given that we can only rely on data from the last forty years and precious little countries can be thought of as having gone the whole path.

In order to distinguish between countries at different stages of economic development, we must either: estimate a nonlinear relationship in income; or estimate a separate coefficient for countries or country groups, ordered by income. The traditional way of assessing the elasticity is to start from a cross-sectional perspective and then either use one cross-section only, pool some cross-sections with or without adding panel features (individual effects). We change perspective, a) taking the time series as our primary perspective and b) augmenting it with common factors, so as to 1) obtain one elasticity for each country, and then 2) assign that country to one development class.

We consider the following linear heterogeneous panel model:

$$p_{it} = \alpha_i + d_t + \beta_i' \mathbf{x}_{it} + u_{it} \tag{3}$$

where $p_{it}$ indicates nominal per-capita insurance consumption in current dollars in country $i$ at time $t$, $\mathbf{x}_{it}$ is, in general. a $k \times 1$ set of regressors (here: GDP), $\alpha_i$ is a country-specific intercept, and $u_{it}$ ia an error term. Premiums and GDP are expressed in natural logs, so that the coefficient can be directly read as an elasticity. The error term is in turn specified according to a multifactor structure as the sum of $m$ unobserved common effects and an idiosyncratic remainder error term:

$$u_{it} = \gamma_i' \mathbf{f}_t + \epsilon_{it} \tag{4}$$

The errors $u_{it}$ can be cross-sectionally correlated because of the similar, albeit not identical, response across countries to modifications in the common factors, measured by the factor loadings $\gamma_i$. The common factors are allowed to be correlated with the regressors, as is most likely to be the case, so their effect comes both through factor loadings and through the indirect effect on the observed regressors. The common factors are also allowed to be nonstationary.

Pesaran's Common Correlated Effects (CCE) estimators can be used to consistently estimate (3) with errors as in (4) and, possibly, also weak-form spatial correlation. The CCE estimators work by augmenting the basic model with cross-sectional averages of both the response and regressors, which pick up the effect of the common factors. Slope parameters $\beta_i$ are estimated by applying least squares to the augmented regression

$$p_{it} = \alpha_i + d_i t + \beta_i' \mathbf{x}_{it} + \mathbf{g}_i' \bar{\mathbf{z}}_t + e_{it} \tag{5}$$

where $\bar{\mathbf{z}}_t = (\bar{p}_t, \bar{\mathbf{x}}_t)'$. The estimator for each individual slope coefficient can then be written compactly as

$$\beta_{CCE,i} = (\mathbf{x}_i' \bar{\mathbf{M}} \mathbf{x}_i)^{-1} \mathbf{x}_i' \bar{\mathbf{M}} p_i \tag{6}$$

with $\bar{\mathbf{M}} = \mathbf{I}_T - \bar{\mathbf{H}}(\bar{\mathbf{H}}'\bar{\mathbf{H}})^{-1}\bar{\mathbf{H}}'$, where $I_t$ is an identity matrix of dimension $T$ and $\bar{\mathbf{H}}$ contains: the $T \times K$ matrix of cross-sectional averages $\mathbf{z}_t$, $t = 1, \dots T$; and a deterministic component comprising individual intercept and time trend (Pesaran, 2006, p.974).

CCE estimation can be performed either imposing parameter homogeneity (but maintaining heterogeneity in intercepts, factor loadings and possibly time trends) which is appropriate under the assumption that $\beta_i = \beta$; or parameters $\beta_i$ can be left free to vary, and the average elasticity $E(\beta)$ is estimated by the mean groups method as $\hat{\beta}_{CCEMG} = \frac{1}{N}\sum_{i=1}^{N}\hat{\beta}_{CCE,i}$. We will employ this last estimator, known as CCEMG, because we are interested in the variation between groups of individual elasticities.

Omitting the individual averages from $\bar{\mathbf{M}}$ so that each individual regression is just OLS on the single time series yields the mean groups (MG) estimator of Pesaran and Smith (1995), which does not account for common factors. In the following, the distribution of individual coefficients $\hat{\beta}_{OLS,i}$ will be compared with that of the $\hat{\beta}_{CCE,i}$ from our preferred specification to give an intuition of the influence of common international factors on estimated elasticities.

**Controlling for omitted variables** Like many others, we regress premiums on GDP too. How is this a complete model? In the following, we explain how the peculiar features of the chosen estimator can account for the other potentially relevant omitted regressors in terms of common factors, individual intercepts and trends.

Country-specific (real) interest rates were added to this very specification in Millo (2014) and deemed insignificant. Thje reason is probably to be found in the substantial comovement of safe bonds and listed equity across the world's markets. Considerable cross-sectional correlation can be observed in equities' markets (Longin and Solnik, 1995; Forbes and Rigobon, 2002; Bekaert et al, 2009) and *a fortiori*, via the real interest rate parity condition (Dooley and Isard, 1980), in fixed income markets (Gagnon and Unferth, 1995).This is not a feature of the increased level of financial market integration in recent years but it has been present throughout our sample period. Real interest rates, in particular, show a common component, the *world*, or *global, interest rate*, "arguably the most important price in financial markets" (Helbling and Wescott, 1995), determined mainly by stockmarket booms and oil shocks, and to a lesser extent by (world aggregate) monetary growth and public debt, around which national rates fluctuate as the result of "substantial and often persistent [...] individual-country components" (Barro, 1991). Equity markets across countries and industrial sectors, too, turn out to be well approximated by global factors related to market momentum, (average) cash flow to price ratios and global risk factors (Hou et al, 2011). The world real interest rate is represented in this specification by a common factor, varying in time but not over the cross-section, to which each country's insurance market is allowed to react in its idiosyncratic way, according to the factor loading $\gamma_i$.

We should account for risk conditions too, which in cross-sectional studies are usually accounted for by population density. In this setting, the time-persistent differences in population density are absorbed by the intercept and by the deterministic trend in each time series regression. The changes in risk conditions along the time dimension can instead be considered as common unobserved factors, as they are usually of global nature: the rise of product liability, the boom in world commerce, the emergence of terrorism after 2001 etc.. Standard panel models take them into account through time fixed effects, which constrain the factor loadings to be equal; a common factor model allows instead for the reaction of each domestic market to be different.

The international price of reinsurance is another very important common factor in insurance, as determining the conditions at which direct insurers can transfer excess risk to reinsurers. As such, increases in the reinsurance price will readily, although partially, be reflected in insurance prices. The unavailability of reinsurance price indices over sufficiently long timespans is another problem to be tackled when analyzing our subject. Time fixed effects are in fact too restrictive, as forcing the factor loading on each country to be equal, which is not realistic: bigger or more developed countries will often have bigger insurers with more capacity, less need to reinsure and hence a lower sensitivity to changes in international reinsurance tariffs. Fortunately, again, the CCE estimator allows for an unobserved factor to affect countries to different degrees.

The inclusion of an individual time trend in each separate time series regression accounts for those characteristics that are indeed time-variant but usually follow a regular, linear pattern, as is the case for urbanization, the share of agriculture, the literacy rate or income inequality. The final reality check for the completeness of the model is given by its cointegration properties, testified by the stationarity of the residuals, in the light of the property of invariance of cointegration spaces (see, again, Millo, 2014).

**Asymptotics** It must be borne in mind, nevertheless, that the good statistical properties of heterogeneous estimators of this kind depend on pooling a reasonably big number of individual coefficient estimates (Pesaran, 2006); which last are indeed consistent, but being based on the time series only they are often unstable and overdispersed. For this reason, we can only assess the behaviour of the average $\bar{\beta}_{i \in S}$ over relatively big subsets $S$ of countries.

A reassessment of the relationship between the income elasticity of insurance and the level of economic development of a country will follow on this new basis, against which the predictions of the S-curve model in the weakest form of hypothesis $H3$ will be tested. Before formal testing, now that we have estimated individual coefficients for each State, in the next paragraph we employ marginal S-plots to get a first intuition of the behaviour of partial elasticities of insurance to income across the economic development spectrum.

## 4.4 Graphical assessment

The specification we employ, allowing for individual heterogeneity in coefficients, intercepts and trends, can in principle be compatible with any ex-post distribution of countries in the descriptive S-graph. Even a pooled specification with fixed or common correlated effects, although imposing homogeneity in the elasticity, would be consistent with it, explaining the different positions by (historical accumulation of) shift factors like individual effects and trends. Only in absence of these would it in fact predict a flat, horizontal S-curve. Therefore, if we allow for different starting points and deterministics, a descriptive S-plot tells us little about the behaviour of insurance at different levels of de-

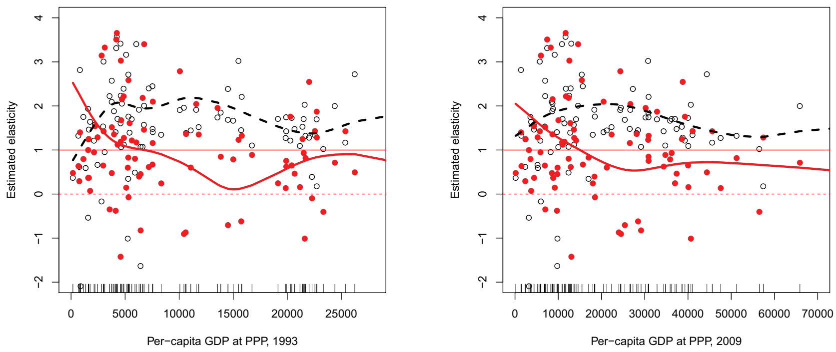


Figure 4: Individual estimated income elasticities of insurance from mean groups (MG) (hollow points) and common correlated effects mean groups (CCEMG) with individual intercept and trend (full points).

velopment. To graphically check $H1-3$ against linearity we must resort to the partial elasticities, i.e. to the *marginal* plot.

In this paragraph we report marginal S-plots for all countries in the sample, with economic development measured in purchasing power parity (PPP) terms in two different years: 1993, the "gravity centre" of the distribution of available data, and 2009, the last reasonably complete year in the sample. We concentrate on the importance of estimating out individual intercepts, trends and common cross-sectional factors, contrasting the results from a mean groups (MG) estimator (hollow black points) with those of our preferred common correlated effects mean groups (CCEMG) estimator (full red points). Spline smoothers are added to the two points clouds to try and identify an overall tendency; the latter are of course dependent on the chosen bandwidth, and only as reliable as the number of available observations in each "region" on the horizontal axis. To help the reader assess the sample size at each level of development, a "rug" graph highlighting the position of each data point is added to the horizontal axis.

As can be seen, individual elasticities scatter widely, often assuming implausibly high or low – even negative – values. This is typical of MG estimators, which are not meant to be consistent pointwise, for each individual/State $i$, but rely instead on averaging the individual coefficients $\beta_i$ over the cross-sectional dimension to get a consistent estimate of the average $\beta$ (see Baltagi et al, 2000). By extension of this line of reasoning, the smoothing spline can be seen as depicting a moving average of the individual coefficients in the "vicinity" of each income level.

While the point cloud from the MG estimator is generally higher and, by and large, consistent with the humped shape predicted by the S-curve theory, accounting for individual heterogeneity and common factors gives rise to a cloud of partial elasticities that are, on average, lower in value and do not follow the predicted shape any more. Rather, CCEMG elasticities tend to be higher in the lower regions of per-capita income, the majority of them staying below one with a tendency to rise in the rightmost part of the income spectrum: i.e., for rich countries.

The main message we get from this marginal version of the S-plots regards the general importance of controlling for idiosyncratic country-specific elements and common factors. As for the S-curve theory, the shape of the smoothing spline actually departs from the predicted one after controlling for common correlated effects; yet this graphical evidence can, at most, be looked at as a very rough indication.[10].

In the following, we elaborate on statistical tests from Millo (2014) in order to formally assess the linearity hypothesis $H3$ first over the cross-section, and then – for the average coefficients – over time.

## 4.5 The S-curve hypothesis vs. linearity: formal testing

In this paragraph we address the robustness of a linear model to the implications of the S-curve hypothesis, i.e. to potential nonlinearity of the income elasticity of insurance and in particular to the hypothesis that $\beta_i = f(y)$. Observationally, this has two possible meanings: elasticity changing with levels of development *across the sample of countries* at a given point in time, or *within a single country*, as it transitions through different levels of development itself.

### 4.5.1 Cross-section "homogeneity"

Our reference model does not assume a constant elasticity for each country, but makes the milder assumption that individual elasticities be drawn randomly from a distribution of mean $\beta$, to be estimated as the average of individual coefficients. Hence if the pattern of elasticity in the cross section were sistematically related to income groups, it would invalidate the hypothesis of random drawing from the same population. As a consequence, the individual estimates would still be consistent but the average estimate would be meaningless.

In Table 2 we report t-tests for difference in means over subsamples of individual coefficients from CCEMG estimates as in Millo (2014, Table 3), separating country groups either according to their income level or to their OECD membership. As a robustness check, we consider coefficients from progressively reduced samples according to the length of the available time series.

The difference between coefficients' populations is marginally significant only in one single case: in the comparison of High-income versus Low- plus Mid-income countries, in the sample of all countries with at least 15 observations in time. According to the "OECD members" criterion, which is the preferred one in most of the literature, it is never significant in any of the four samples considered.

### 4.5.2 Timewise linearity

By contrast, in our model the elasticity is indeed assumed to be constant over time, within a single country. The CCEMG assumes invariance of individual

[10]Apart from the slight rightmost hump, one could even argue that this distribution could still be interpreted as depicting only the right part of an hypothetical S-curve, where the leftmost countries represent developing ones and the truly underdeveloped world is not even represented: but see below, Section 5.

| | Low-income | Mid-income | High-income | OECD members |
|---|---|---|---|---|
| T>14 | 0.80 | 1.64 | -2.04* | -0.70 |
| | (0.43) | (0.11) | (0.05) | (0.49) |
| T>19 | -0.33 | 1.26 | -0.40 | -0.23 |
| | (0.74) | (0.22) | (0.69) | (0.82) |
| T>24 | 1.29 | 0.45 | -1.52 | -1.02 |
| | (0.22) | (0.66) | (0.14) | (0.32) |
| T>29 | 0.57 | 0.50 | -0.90 | -0.45 |
| | (0.58) | (0.63) | (0.38) | (0.66) |

Table 2: Pairwise t-tests for difference of each subset of individual coefficients from rest of population: low-, mid- or high-income countries according to quantiles in the 2000 distribution of per-capita GDP at PPP, and OECD vs. non-OECD. p-values in brackets (reproducing Table 3 in Millo, 2014).

coefficients in time. Hence – under the S-curve hypothesis – the fact that a country transitions through different levels of development inside the sample's time horizon would invalidate the estimate for that single country because of incorrect functional form. Intuitively, not many countries will have transitioned across broad stages of economic development within the timespan of our sample: according to the OECD membership criterion, after Australia and New Zealand (1971, 1973) only the Czech Republic (1995), Hungary (1996), South Korea (1996), Mexico (1994), Poland (1996) and Slovakia (2000) did, while Chile, Estonia, Israel and Slovenia all joined the Organization in 2010. Nevertheless, some formal diagnostic testing is in order. Given that the available sample length over time is too short to test for linearity in time on a by-country basis, Millo (2014) resorts to the pooled linearity test proposed in Lee and Chiu (2012), which amounts to adding squares and cubes of log income to the CCEMG model, testing their joint significance, and is analogous to a pooled RESET test. The $\chi^2(2)$ test statistic takes a value of 0.0824 (p-value: 0.96) hereby accepting the linearity hypothesis; for comparison we run the same test on a two-way fixed effects (homogeneous) model, which strongly rejects ($\chi^2(2) = 142.49$). We attribute this result to the overly restrictive nature of the fixed effects specification, and especially to its inability to account for individual trends. A visual inspection of by-country residuals (not shown) supports this view. Neither of the hypotheses of heterogeneous coefficients drawn from the same population and of linearity in time are rejected.

# 5 Saving the S-curve?

We have shown how the S-curve theory, even if taken in a rather broad sense, is unsupported by statistical evidence based on the observation in time of large international samples of national non-life markets.

As observed, nevertheless, assessing the S-curve theory against linearity leaves the burden of proof to the defendant: the power of statistical tests in detecting departures from linearity as $H3$ will depend on the signal to noise ratio in the given sample. "Noisy" data, full of idiosyncratic variation – and so insurance data use to be – might overshadow a nonlinear data generating process.

Advocates of the theory might find another explanation for the paths observed in the marginal S-plots. While the Sigma dataset is probably the most comprehensive insurance database at hand, it still comprises almost exclusively countries already endowed with a functional financial market and a reasonable level of economic development. India, Indonesia, Ecuador and Vietnam, for example, all fall into the lowest development quintile. Therefore, we can think of our working sample as starting out at income levels already higher than those at which, according to the S-curve hypothesis, an insurance market would slowly come into existence. Putting it more generally, inconsistent evidence might stem from partial observation: we might be looking at just one part of the full S-curve of world insurance. This line of argument, which has been taken quite far in the literature[11], does not seem convincing, as the Sigma sample covers countries with very diverse levels of development, starting at very low levels of per-capita GDP. Moreover, if we conceded that nonlinearity be present, but too much drowned in random variation to be detectable by formal testing; and took the shape of the smoother splines to represent the average behaviour of income elasticities as per-capita income varies; and also conceded to be observing only the rightmost part of the marginal S-curve; still, Enz's device would leave unexplained the second hump which is apparent in the distribution of elasticities towards the right end of the income distribution in Figure 4.4.

We rather attribute the empirical failures documented in this paper to the hypothesis of product homogeneity underlying most of the literature on the S-curve. In other words, we argue that if we were able to observe a large panel sample of national markets for more specific lines of business than total non-life, then we might well observe a behaviour consistent with Enz's theory. In the following we will provide arguments, and some partial evidence, in favour of this conjecture.

## 5.1 The many waves of insurance growth

It is common wisdom that markets for traditional insurance lines (motor third party liability, fire, theft) tend to saturate in developed countries, with products commoditizing and competition putting pressure on premium revenue: so that, if we were able to observe income for these lines alone, we might perhaps confirm the S-curve's predictions as regards the moderation of income elasticity in the rightmost part of the development spectrum. Yet typically the mature markets of rich countries also develop new business in the insurance lines associated with modern service economies: professional and product liability, legal protection, travel assistance, business interruption, long term care and so on. A changing product mix might therefore be the explanation for the apparently flat distribution of elasticities and also for the "second youth" in the life cycle of insurance markets hinted at in Figure 4.4, in the spirit of the "two waves of service-sector growth" of Eichengreen and Gupta (2013). In turn, the general failure of the

[11] E.g., Chang and Lee (2012, p.242) dismiss the inconsistency between their results (elasticity is at least three times *higher* for developed countries) and those of Ward and Zurbruegg (2002, p.405) (elasticity is three times *lower* for the OECD with respect to developing Asia): "it could be interpreted that our findings here portray the former half segment of the S-curve, while Ward and Zurbruegg characterise the latter-half one."; but Ward and Zurbruegg (2002)'s sample of countries is a proper subset of theirs.

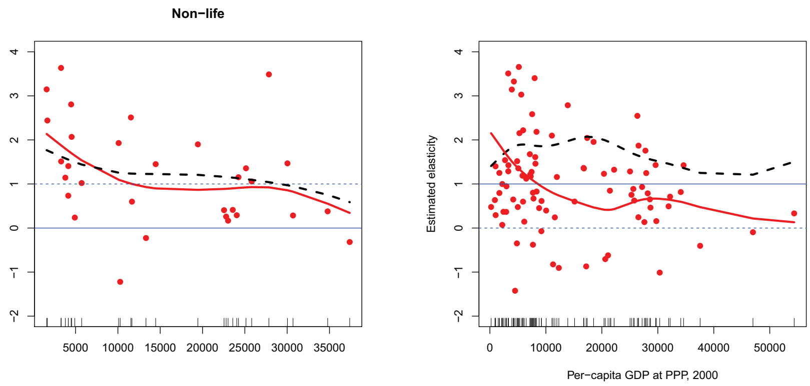


Figure 5: Individual estimated income elasticities of insurance from CCEMG with individual intercept and trend (full points), plotted against per-capita GDP at PPP in 2000 for European sample (left) and World (right). Spline smoothers for CCEMG (red) and for MG (dashed black line, points not shown) are superposed to the graph.

S-curve theory in describing the available evidence – and, more in general, the tendency of nonlife insurance markets to grow in line with GDP at any level of development documented by Millo (2014) – might stem from compensations between mature, commoditized business lines like motor TPL, with compressed margins and lower revenues, and "young", innovative lines taking their place in the total business mix of non-life.

We now ask ourselves whether there is any evidence that the S-curve theory can explain the evolution of individual branches of non-life. Data limitations will be more binding than above, because of the unavailability of databases of geographical breadth and time depth comparable to that of the Sigma dataset. Given the smaller and shorter sample available, we will limit ourselves to a graphical analysis as a first assessment of compliance with the theory. If encouraging, this will provide directions for future work.

To address the issue of composition within the non-life sector, we must resort to a different dataset. Insurance Europe (various issues) has been publishing data over European insurance premiums since 1992, divided into (life and) some standard subsets of non-life: motor, property, liability, accident and health, marine aviation and transit (MAT) and legal expenses. Despite the European focus of the dataset, the development spectrum covered is comparable with that of the Sigma World dataset. Comparing marginal S-plots for both in the year 2000 (see Figure 5.1), the behaviour of the European subset is largely consistent with the rest, but for the fact that - consistently with European integration - idiosyncratic and common factors seem to play a lesser role (witness the closeness between the MG and CCEMG splines).

Moving towards a by-line analysis, we start with descriptive S-plots of non-life vs. each subsector (Figure 5.1). Despite dispersion, the tendency of motor to have the highest penetration in the middle section is rather evident; even clearer the tendency of countries to gather in two distinct groups for property insurance, the richer ones having, on average, about double the insurance penetration of the poorer ones. Although not so clear-cut, the tendency is similar for the other non-motor branches, perhaps with the exception of MAT which is very much dependent on the peculiar characteristics of the country, and therefore has a few outliers (the UK, Norway) in an otherwise flat point cloud. Accident and health, and legal expenses, are also clearly idiosyncratic.[12]

Turning to the marginal S-plots, by comparing those of motor and non-motor (respectively, left and right panels of Figure 5.1) we can clearly see how the rightmost hump in the distribution of total non-life elasticities in Figure 4.4 is due to the contribution of non-motor, motor showing instead a set of very high elasticities at the lowest end of the development range and then a very flat distribution, mostly concentrated in the zero-one range. The high elasticities in the leftmost part of the figure are relative to Eastern European countries, where the time period of our sample witnessed the contemporaneous development of private insurance and the surge in private car transport after the fall of the Berlin Wall. Once more, an individual idiosincracy.

# 6 Conclusions

The S-curve hypothesis has been taken for granted by much of the recent literature on insurance growth. Various kinds of inconsistent evidence have been accommodated by basically choosing some part of the curve which best fitted the available data, or more generally by considering any nonlinearity as evidence in its favour. This does a bad service to the theory itself. To paraphrase the incipit of a famous paper in economic growth theory, this paper "takes Rudolf Enz seriously", first by trying a precise understanding of what the S-curve is meant to be, then by checking whether its predictions are confirmed by the data.

As we have shown drawing on a) a critical review of previous empirical studies, and on b) the specific results of a recent paper providing a consistent method to estimate the average elasticity to income of a set (or a "big enough" subset) of countries, the S-curve hypothesis does not generally hold, even in very mild form. There is scant evidence that income elasticities depend on the level of development in any systematic way.

The S-curve hypothesis remains nevertheless an appealing and intuitively plausible description of the evolution of insurance markets, and it would be a very useful forecasting model. All hope of reconciling it with empirical evidence is not lost. It may well hold for specific lines of business, and composition effects, i.e. an excessive aggregation of data into the big categories of life and non-life, could have prevented us from finding any evidence in its favour. An assessment of this conjecture in more disaggregated settings is left for future work, the preliminary results of which, reported in the final section of this paper, are encouraging.

[12] It must be borne in mind, e.g., that some European welfare systems allow (some categories of) citizens to opt out of the public health system, providing a decisive boost to private health insurance.

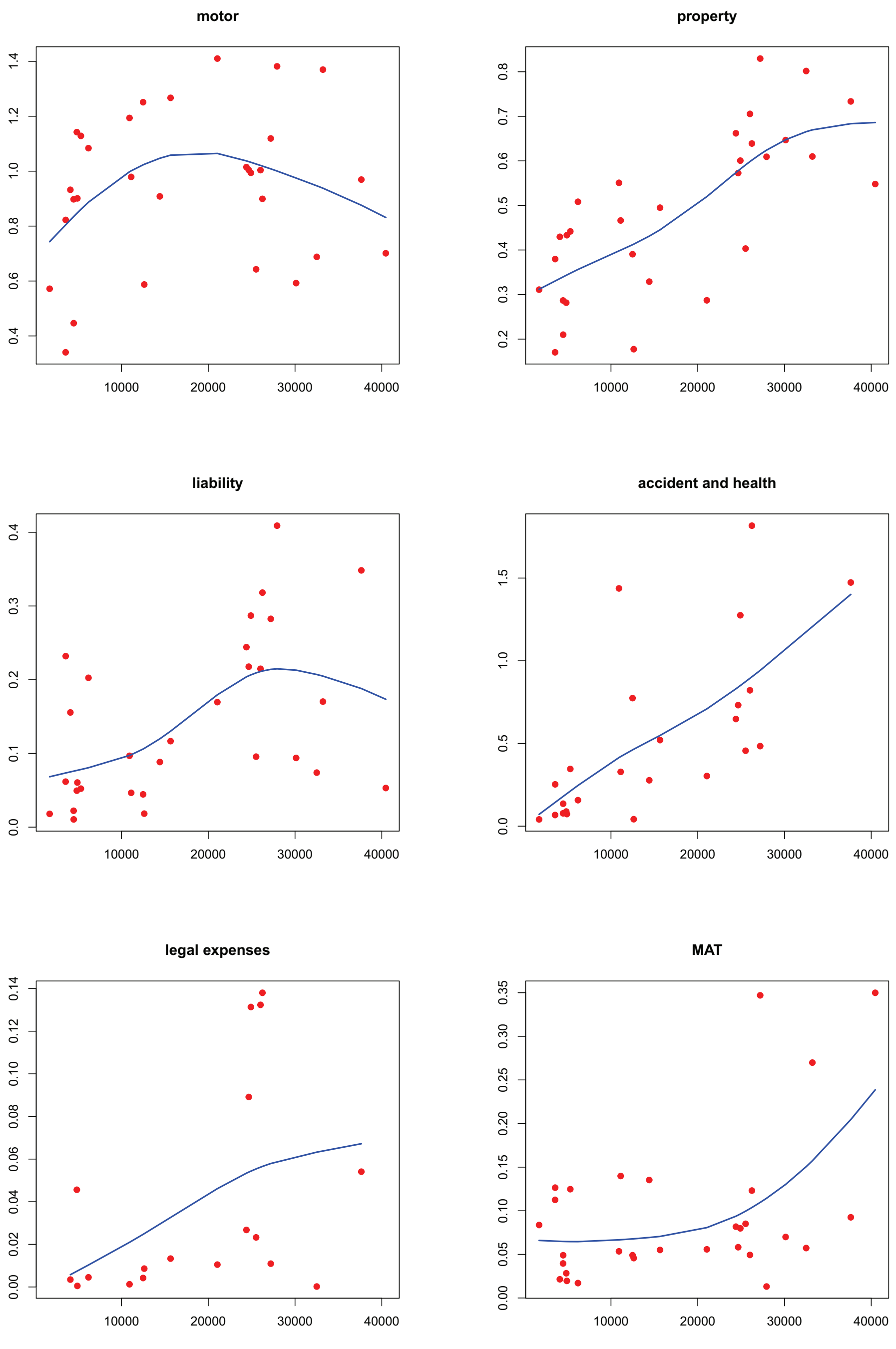


Figure 6: By-line descriptive S-plots: insurance penetration vs. GDP per capita at PPP. Spline smoothers added.

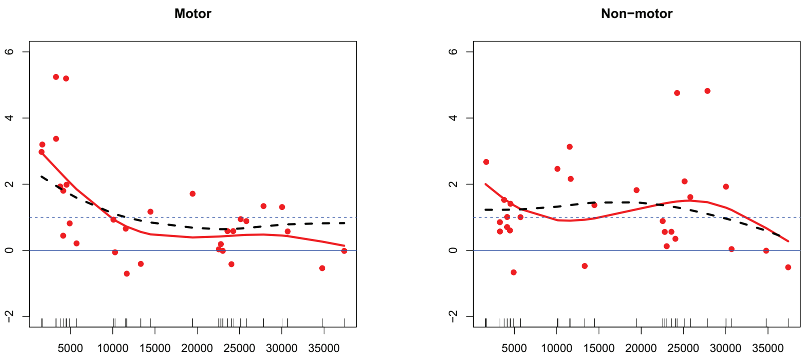


Figure 7: Individual estimated income elasticities of Motor (left) and Non-motor insurance (right) from Common Correlated Effects Mean Groups with individual intercept and trend plotted against per-capita GDP at PPP in 2000.

# References

Ayres R (1990a) Technological transformations and long waves. part i. Technological Forecasting and Social Change 37(1):1–37

Ayres R (1990b) Technological transformations and long waves. part ii. Technological Forecasting and Social Change 37(2):111–137

Baltagi BH, Griffin JM, Xiong W (2000) To pool or not to pool: Homogeneous versus heterogeneous estimators applied to cigarette demand. Review of Economics and Statistics 82(1):117–126

Barro RJ (1991) World interest rates and investment. Tech. rep., National Bureau of Economic Research

Bekaert G, Hodrick RJ, Zhang X (2009) International stock return comovements. The Journal of Finance 64(6):2591–2626

Carter R, Dickinson G (1992) Obstacles to the Liberalization of Trade in Insurance. Harvester Wheatsheaf, London

Chang CH, Lee CC (2012) Non-linearity between life insurance and economic development: A revisited approach. The Geneva Risk and Insurance Review 37(2):223–257

Dooley MP, Isard P (1980) Capital controls, political risk, and deviations from interest-rate parity. The journal of political economy 88(2):370

Eichengreen B, Gupta P (2013) The two waves of service-sector growth. Oxford Economic Papers 65(1):96–123

Enz R (2000) The s-curve relation between per-capita income and insurance penetration. Geneva Papers on Risk and Insurance: Issues and Practice 25(3):396–406

Europe I (various issues) European insurance in figures. Tech. rep., Insurance Europe

Forbes KJ, Rigobon R (2002) No contagion, only interdependence: measuring stock market comovements. The journal of finance 57(5):2223–2261

Gagnon J, Unferth M (1995) Is there a world interest rate? Journal of International Money and Finance 14(6):845–855

Grace MF, Skipper HD (1991) An analysis of the demand and supply determinants for non–life insurance internationally. Technical report, CRMIR, Georgia State University

Helbling T, Wescott R (1995) Ii the global real interest rate. Staff Studies for the World Economic Outlook (EPub)

Heston A, Summers R, Aten B (2011) Penn World Table Version 7.0. Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania, URL http://pwt.econ.upenn.edu/

Hornstein A, Prescott E (1991a) Insurance contracts as commodities: a note. The Review of Economic Studies 58(5):917–928

Hornstein A, Prescott E (1991b) Measures of the insurance sector output. Geneva Papers on Risk and Insurance 16:191–206

Hou K, Karolyi GA, Kho BC (2011) What factors drive global stock returns? Review of Financial Studies 24(8):2527–2574

Lee C, Chiu Y (2012) The impact of real income on insurance premiums: Evidence from panel data. International Review of Economics and Finance 21:246–260

Longin F, Solnik B (1995) Is the correlation in international equity returns constant: 1960–1990? Journal of international money and finance 14(1):3–26

Millo G (2014) The income elasticity of nonlife insurance: A reassessment. Journal of Risk and Insurance forthcoming

Moulton BR (1986) Random group effects and the precision of regression estimates. Journal of econometrics 32(3):385–397

Moulton BR (1990) An illustration of a pitfall in estimating the effects of aggregate variables on micro units. The review of Economics and Statistics pp 334–338

Nelson C, Plosser C (1982) Trends and random walks in macroeconmic time series: some evidence and implications. Journal of monetary economics 10(2):139–162

Outreville J (2013) The relationship between insurance and economic development: 85 empirical papers for a review of the literature. Risk Management and Insurance Review 16(1):71–122

Papell D, Prodan R (2004) The uncertain unit root in us real gdp: Evidence with restricted and unrestricted structural change. Journal of Money, Credit and Banking 36(3):423–427

Pesaran M (2006) Estimation and inference in large heterogeneous panels with a multifactor error structure. Econometrica 74(4):967–1012

Pesaran M (2007) A simple panel unit root test in the presence of cross-section dependence. Journal of Applied Econometrics 22(2):265–312

Pesaran MH, Smith R (1995) Estimating long-run relationships from dynamic heterogeneous panels. Journal of econometrics 68(1):79–113

Sherwood M (1999) Output of the property and casualty insurance industry. The Canadian Journal of Economics 32(2):518–546

sigma (various issues) World insurance in (year). Tech. rep., Swiss Re

Verhulst P (1845) Recherches mathématiques sur la loi d'accroissement de la population. Nouv mém de l'Académie Royale des Sciences et Belles-Lettres de Bruxelles 18:–..41

Verhulst P (1847) Deuxième mémoire sur la loi d'accroissement de la population. Mém de l'Académie Royale des Sci, des Lettres et des Beaux-Arts de Belgique 20:1–32

Ward D, Zurbruegg R (2002) Law, politics and life insurance consumption in asia. Geneva Papers on Risk and Insurance Issues and Practice pp 395–412

Wooldridge JM (2002) Econometric analysis of cross-section and panel data. MIT Press

Zeileis A, Yang G (2012) pwt: Penn World Table. URL http://CRAN.R-project.org/package=pwt, r package version 7.0-1

Zheng W, Liu Y, Dickinson G (2008) The chinese insurance market: estimating its long-term growth and size. The Geneva Papers 33:489–506