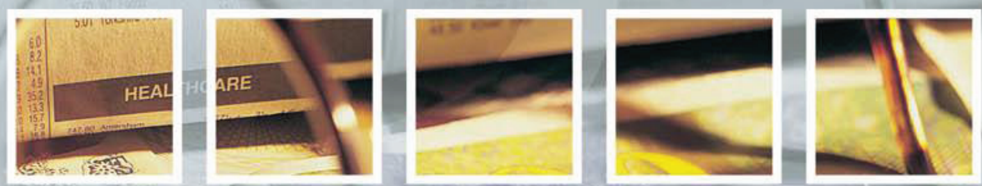
Assicurazioni Generali

RESEARCH DEPARTMENT

The Income Elasticity of Non-Life Insurance:
A Reassessment



2014

Giovanni Millo

WORKING PAPER WP1/14

The income elasticity of non-life insurance: a reassessment*

Giovanni Millo†

June 14, 2013

Abstract

In aggregate insurance regressions at country level, the question whether insurance is a normal or superior good translates into whether income elasticity is significantly greater than one or not. 25 years after a seminal paper, we reassess the income elasticity of non-life insurance by means of homogeneous and heterogeneous versions of the Common Correlated Effects estimator, controlling for common factors and individual trends and characterizing the average behaviour of insurance markets while allowing for individual heterogeneity. The evidence supports the existence of a cointegrating behaviour between insurance consumption and GDP and the view of non-life insurance as a normal good.

1 Introduction

In aggregate insurance regressions at country level, GDP shows up as by far the most important driver of growth, proving positive and significant e.g. in all studies considered in Outreville (2012)'s review. Hence, characterizing the long-run elasiticity of aggregate premiums to GDP is of great value to forecasters trying to gauge the future development trends in the medium-to-longer term. On the other hand the theoretical debate has long been going on whether insurance is a normal or a superior good, translating in an aggregate setting into whether said elasticity is significantly greater than one or not. We address the question of whether insurance is a normal or a superior good from an aggregate perspective, i.e., do market premiums grow less or more than proportionally with economic development?

*The paper has benefited greatly from discussions with Gaetano Carmeci; nevertheless all the errors are the author's responsibility. The views expressed are solely his own and do not necessarily reflect those of his employer. The author is grateful to Swiss Re Research and Consulting for providing missing data from back issues of *sigma*. All the computations in the paper are done inside the R open-source environment for statistical computing (R Development Core Team, 2012), generally using the *plm* add-on package for panel data econometrics. This paper has been prepared as a dynamic document with the *Sweave* utility (Leisch, 2002) according to the principles of literate statistical practice.

†Generali Research & Development at Generali S.p.A., via Machiavelli 3, 34132 Trieste.Tel.: +39-040-671184, Fax: +39-040-671160, email: giovanni_millo@generali.com.

In two seminal papers, Beenstock et al (1986, 1988) were the first to consider the behaviour of, respectively, life and non-life[1] insurance with respect to economic growth by pooling a number of time series from different countries. 25 years from Beenstock et al (1986, 1988)'s influential twin papers, methodology has progressed to the point of successfully tackling a number of then-unresolved problems, already acknowledged by the original authors: in particular, cross-sectional between-countries heterogeneity in the coefficients of interest (Beenstock et al, 1988, p. 259) and serial correlation (Beenstock et al, 1988, p. 267)[2]. New methodological concerns have emerged since, to which the scientific community paid scant attention or of which it was even scarcely aware back then: most notably, cross sectional and spatial correlation and nonstationarity of variables in panel data. The latter is possibly leading to spurious regressions in the sense of Granger and Newbold (1974), so that results of regressions between nonstationary data must be taken with care, at least unless cointegration is proved (see Phillips and Moon, 1999). The former, cross-sectional correlation, can assume different forms, essentially based on whether its scope is local and distance-decaying (*spatial* correlation) or globally affecting every country in the cross-section (as in *common factor* models), as formalized in Pesaran and Tosetti (2011); this second case being the most problematic, as symptomatic of a specification flaw possibly leading to inconsistent estimates if the common factors' influence is not accounted for. National insurance markets are notoriously affected by common, international factors, like shifts in the global price of reinsurance, single catastrophal losses of more-than-national scope or global changes in risk conditions, as most notably happened in 2001 after the World Trade Center attack, and therefore controlling for this kind of dependence is of the utmost importance. Fortunately, new estimators are available that are able to effectively account for unobserved common factors, as will be detailed in the following.

From the point of view of information, time has again healed some problematic aspects acknowledged in the original study, as the short timespan available (Beenstock et al, 1988, p. 260). Today we draw on a new Sigma dataset, beginning from the year 1970 as in Beenstock et al (1988) but now extending until 2010, thus spanning 40 years of insurance history (sigma, various issues). In turn, the geographical scope of the database has been constantly extended, so that now with respect to the 1981 version in Beenstock et al (1988) it comprises a much larger number of countries and completely new areas like Eastern Europe. Although many "new" countries have been added only recently, on average the Sigma dataset has become "long" enough to allow employing modern panel time series methods.

By contrast, despite many later efforts (for the non-life sector, see Outreville, 1990; Browne et al, 2000; Esho et al, 2004; Feyen et al, 2011) the clarity of Beenstock et al (1988)'s economic approach to specification remains a benchmark for most studies of insurance development. Starting from a microeconomic perspective, they define market turnover, the only observable variable of interest, as the equilibrium outcome of supply and demand, drawing theoretical predic-

[1] They use the term *property-liability* as in standard insurance parlance, meaning total non-life premiums and hence including other lines, notably – but not only – accident, health and transport.

[2] The use of the Durbin Watson statistic in a dynamic model is problematic, see Dezhbakhsh (1990).

tions from a formalization of (unobservable) demand and supply schedules and translating them in terms of (observable) premium volume. We therefore felt it was time for a reassessment of their original economic approach to the relationship between insurance and economic development at the country level with new tools and improved data. For the above considerations, we will maintain Beenstock et al (1988)'s economic framework while employing modern econometric techniques. In this latter respect, we will essentially borrow from the methodological work of Hashem Pesaran and coauthors (Pesaran, 2004, 2006, 2007; Pesaran and Tosetti, 2011; Holly et al, 2010; Kapetanios et al, 2011). Our empirical approach will in turn be inspired by the recent literature on health expenditure, which has been applying these methods to an analogous research question (Baltagi and Moscone, 2010; Moscone and Tosetti, 2010).

In the following section, we will discuss the main research question, providing some descrptive evidence and motivating our approach. Section 2.1 will be dedicated to a methodology-oriented critical review of existing literature and the motivation for a time-series perspective. The main part of the paper will then follow as Section 3, where we will discuss both the economic and the econometric specification with particular attention to some hitherto unanswered methodological issues: controlling for omitted common factors and for spatial correlation. At the end of Section 3 a preliminary analysis of data will demonstrate the relevance of these issues for the case at hand, together with data nonstationarity and hence the need for cointegration methods. Section 4 will be devoted to the presentation and discussion of empirical results, Section 5 to the conclusions.

2 Insurance: luxury or necessity?

Among insurance theorists there has been a long-standing debate on the possibility of insurance as an inferior good, dating back to Mossin (1968). Hoy and Robson (1981) examined conditions for insurance to even be a Giffen good, a possibility later dismissed by Borch (1986) while the Hoy and Robson (1981) model has been generalized by Briys et al (1989) and recently by Hau (2008). The empirical counterpart at the macroeconomic level of this research question roughly translates onto the magnitude of the elasticity of insurance consumption to income, as in Beenstock et al (1988) and Grace and Skipper (1991). A negative value would be considered consistent with the hypothesis of insurance as an inferior good. However, in empirical studies only largely positive values have ever been found (see the review in Outreville, 2012). Therefore, in the empirical literature the debate has rather been focused on whether insurance is a superior or a normal good; in perhaps more evocative terms, a *luxury* or a *necessity*. If income elasticity is (positive and) not greater than one, i.e. insurance is a normal good, then from a macroeconomic viewpoint, ceteris paribus, its share in the total economy shrinks – or at most remains unchanged – along with economic development. On the contrary, an elasticity greater than one (superior good) means that the sector is growing in importance.

Approaching the relationship between insurance and income from a bivariate perspective, nevertheless, misses the point, as all other characteristics of the economic environment are likely to move together with income, as will demography and so on. In fact, while the evidence of a largely positive statistical relationship between insurance premiums and GDP is overwhelming, the question is whether the development of the insurance market is actually due to income growth or to other correlated factors. We argue that the most challenging source of heterogeneity is in more or less persistent individual characteristics of a country, and therefore an approach based on cross-sectional or "short panel" data faces great difficulties from the beginning. Hence, in line with recent research on health expenditure and supported by the availability of a unique database dating back to 1970, we aim at measuring the income elasticity of insurance from a panel time series perspective. But first let us go through a succinct review of the existing literature.

2.1 A methodological review of the literature

In this section we review methodological approaches and empirical results from the previous literature on the income elasticity of insurance. Cross-country comparisons in the literature have been focusing on structural economic, cultural and social differences as well as the different level of economic development across countries. Beenstock et al (1988) analyze the effect of income without controlling for other determinants on a cross-section of 45 countries observed in 1981, finding a strong positive correlation between income and property-liability insurance revenues and estimating a coefficient of 1.34. They also estimate a more complete LSDV dynamic model on a pool of time series for 12 developed countries, where the country fixed effects are explicitly meant to control for the cross-country differences in loss probability; they find positive effects of income and real interest rates, which last they attribute to the supply of underwriting capital attracted by the higher returns dominating the opposite effect on demand. The estimated elasticity to income for the 12 countries in the restricted sample is always greater than one with the exception of Italy and Japan. They ultimately characterize insurance as a superior good.

Outreville (1990) analyzes a cross-section of 55 developing countries observed in 1983-84 by UNCTAD. He first estimates a simple log-log model with income as the only regressor, finding the very same coefficient of 1.34 as Beenstock et al (1988) did. He then adds different measures of financial development and of insurance prices, finding positive effects from the former and not significant ones from the latter; in all the augmented models the coefficient on log income remains between 1.24 and 1.34 (see his Table 1). He then hypothesizes the association of low insurance development with high inflation and the prevalence of agriculture and estimates another specification including the latter together with a measure of literacy, failing to find a statistically significant relationship for the former two and instead finding literacy to be negative and marginally significant (his Table 3, where the coefficient of log income is 1.37). Lastly, he estimates two regional models (with two alternative specifications each) for 20 Latin American and Caribbean countries and, respectively, for 24 African countries, with the income elasticity turning out, respectively, at 1.01-1.17 and 1.14-1.18, somewhat lower than the sample average but still supporting the "superior good" view.

Grace and Skipper (1991) analyze a sample of developing as well as developed countries, finding a positive and significant influence on non-life insurance demand by income and literacy. Islamic countries have, ceteris paribus, lower values while the share of government consumption over GDP is in turn associated with higher insurance density. They also assess the influence of a monopolistic market and other institutional differences, all going in the expected direction. They estimate different income elasticities for developing and developed countries (identified as OECD ones), characterizing insurance as a superior good in both samples but finding a higher elasticity for the OECD one.

It must be noted that both these latter studies do not explicitly define market equilibrium: Outreville (1990) only specifies the demand equation, but then observes equilibrium consumption; Grace and Skipper (1991) specify a demand and supply system omitting price and they estimate it as two seemingly unrelated regressions. Therefore we believe their findings are better characterized as describing insurance consumption, rather than insurance demand as stated in the papers. This lack of distinction between insurance demand and equilibrium consumption, often used as synonyms, persists in some of the current literature (see e.g. Hussels et al, 2005); a possible reason is the underlying belief that insurance supply be very elastic with respect to price, so that the equilibrium quantity be determined mainly by demand factors.

Browne et al (2000) instead resort to explicitly estimating a single-equation relating (equilibrium) insurance consumption to a set of regressors, most of which are demand-related. They address the issue of heterogeneity by estimating both a pooled and a fixed effects regression[3] on an unbalanced panel of 22 OECD countries over the period 1987 to 1993. The fixed effects are explicitly meant to control for "accident rates, motorization rates, alcohol consumption, capital stock and attitudes towards litigation" and any other unobservable country-specific factor.[4] Unlike previous studies, they focus on the subset of motor and liability insurance, most of which is compulsory. Their findings are rather diverse both across sectors and across the two specifications, with coefficients often changing sign, which can be seen as evidence in favour of the need to control for unobserved heterogeneity in cross country studies. Concentrating on the fixed effects results, they find income to be a postive determinant for both sectors, although their specification doesn't allow to read it directly as an elasticity. They also claim that insurance consumption is decreasing with wealth, thus supporting the view of wealth as a substitute for coverage; and that insurance consumption turns out to be higher in common law countries[5].

Further work by Esho et al (2004), drawing on research on the effect of property rights enforcement on the development of financial markets, investigates the influence of the type of legal system on insurance, this time distinguishing between English, French, German and Scandinavian. They consider a sample of 44 developed and developing countries, focusing on the legal origin rather than the degree of development. They perform both a cross-section and a panel analysis, the latter on data from 1984 to 1998. The panel model is estimated first by fixed effects over three-year averages, then in a dynamic specification employing the generalized method of moments (GMM) system estimator of Arellano and Bond (1991) and Arellano and Bover (1995). They find the legal origin to be not significant, while controlling for the level of enforcement of property rights they find a positive effect. As for the income elasticity, in the cross-sectional

[3]They also estimate, but do not report, a random effects regression with results similar to the fixed effects one.

[4]The authors are ostensibly assuming all these factors to be time-invariant.

[5]Having been forced to omit these two time-invariant variables from the fixed effects analysis, these two claims are based only on pooled models, so that the abovementioned inconsistencies between the two specifications may cast doubt on this finding.

model their estimates range from 0.83 to 1.77; in the panel ones, the long-run elasticity[6] turns out to be 1.16 for the FE model, 0.80 for the GMM model on three-year averages and 1.35 for the GMM on yearly data (see their Table 4).[7]

Lastly, and most recently, Feyen et al (2011) analyze another dataset of 90 developed and developing countries over the years 2000 to 2008 (source is AXCO Insurance Services) in order to identify the drivers of insurance development. To this end, they control for a number of covariates ranging from inflation to population density and religion to legal rights protection and car density. Translating their specification into income elasticities,[8] the resulting values range from 0.94 to 1.30 (see their Table 6).

2.2 From cross-sections of countries to time series

The earlier studies have been subject to a number of potential weaknesses related to the nature of the samples involved, which were at the time difficult to tackle both for lack of adequate data and of appropriate econometric techniques, then still to be developed. We have seen how the literature has progressed from cross-sectional analysis towards panel datasets in order to control for the effect of unobservables. As Millo and Carmeci (2011) argue, these regressions are still threatened by unobserved heterogeneity - mostly related to institutional factors - which is not guaranteed to be absorbed by time-invariant fixed effects, and whose effect income is likely to pick up, as witnessed by the fact that in the literature most of the time adding further regressors has decreased its coefficient.[9] Moreover, the very coefficients of interest may be heterogeneous across countries. Their solution is to investigate some determinants of insurance expenditure, and especially income elasticity, in the most institutionally homogeneous setting available, that of regions within a single country. Yet this doesn't provide evidence on how the elasticity of income should behave in conditions different from those spanned by the regions of the only country considered. Therefore our analysis here cannot but take a country-level perspective.

The cross-sectional dimension on which most studies have hitherto relied in order to identify income elasticity suffers, in panel data parlance, of the *incidental parameters problem*: i.e., you cannot increase the sample without increasing the heterogeneity as well. With the increasing availability of "long" panels of insurance data, i.e. panels where the time dimension exceeds 20-30 yearly data points, it is natural to resort to time series variability in order to identify income elasticity. Standard panel data methods, though, are prone to a number of futher issues when analyzing the typical "long" panel, or *pooled time series*, of countries. The most important issues in this respect are nonstationarity in the time dimension, which unless there is cointegration may lead to *spuri-*

[6]The long-run elasticity has been calculated from their estimates of the short-run coefficient on income β and the autoregressive coefficient on premiums γ as $\beta/(1-\gamma)$.

[7]It must be observed, though, that while the results are clearly consistent with a log-log specification, no mention of this is to be found in the paper.

[8]The original specification employs the log of premiums over GDP as the dependent variable, per capita GDP as a regressor. By the properties of logarithms, therefore, reparameterizing into our desired specification the original coefficient β of income translates into $1+\beta$.

[9]As is well known, omitted variable bias has the same sign of the correlation between each included regressor and the omitted variable. In cross-country regressions, income ends up being correlated with practically every other potential regressor, most of the time positively. On the subject see also Zingales (2003).

ous regression in the sense of Granger and Newbold (1974), and correlation in the cross-sectional one, affecting the properties of estimators and possibly invalidating unit root tests; these issues have spurred a substantial amount of methodological literature in the last fifteen years, and while this line of research is still very much in development it is safe to say that many results are well established, although most of them haven't found their way into econometrics textbooks yet. A whole toolbox of appropriate methods is now available, which was not at the time when most of the literature cited above was produced, and against which many findings are to be reconsidered.

In the next section we will employ econometric techniques from the recent literature on nonstationary panels, on spatial panels and on panels with (unobserved) common factors in order to consistently estimate the long-run income elasticity of insurance.

3 Consistent estimation of the income elasticity of insurance

In this section we will outline the economic specification that will be estimated, discussing the limitations of the observational context with particular attention to the possible sources of heterogeneity and unobservable common factor influences, and the empirical econometric framework employed in the estimation.

We will turn to the empirical framework first, shortly reviewing the idea behind Pesaran's Common Correlated Effects (CCE) estimator and its properties, motivating its use in terms of the characteristics of the problem at hand, and especially the need to account: for unobserved country-level heterogeneity in both intercepts *and slopes*; for the influence of unobserved, time-varying and nonstationary common factors potentially correlated with the included regressors, influencing each country according to a different, although time-constant, coefficient; and for global shocks and potential spatial spillovers at relatively local level. In particular, we will explain how the CCE estimator is expected to successfully account for unobservables among the variables suggested by the theory for inclusion. Then we will outline the equilibrium model of insurance consumption of Beenstock et al (1988), a supply and demand system where premiums depend on income, interest rates and loss probability. As only income and interest rates are readily observable, these will be included in the specification; yet loss probability and other influential common or idiosyncratic factors will be accounted for through intrinsic characteristics of the estimator employed, as detailed below.

We will closely follow the empirical strategy in Holly et al (2010), Moscone and Tosetti (2010) and Baltagi and Moscone (2010), with the exception that while the latter, drawing on an OECD sample, concentrate ex ante on pooling estimators, for reasons of sample heterogeneity we will maintain heterogeneous estimators as our preferred choice unless disproven by the data.

3.1 Econometric specification

The extreme heterogeneity of our sample, comprising the majority of relevant insurance markets in the world, suggests to avoid imposing pooling restrictions in the basic econometric model. We therefore consider the following linear heterogeneous panel model:

$$p_{it} = \alpha_i + d_i t + \beta_i' \mathbf{x}_{it} + u_{it} \tag{1}$$

where p_{it} indicates real per-capita insurance consumption in constant 2005 dollars in country i at time t, $\mathbf{x}_{it}$ is a $k \times 1$ set of regressors including real GDP and controls, α_i is a country-specific intercept, $d_i t$ is a country-specific time trend 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:

$$u_{it} = \gamma_i' \mathbf{f}_t + \epsilon_{it} \tag{2}$$

where γ_i and $\mathbf{f}_t$ are $m \times 1$ vectors of, respectively, factor loadings and common factors. Such structure is capable of generating cross-sectional correlation in case of a similar, albeit not identical, response across countries to modifications in the common factors, measured by the factor loadings γ_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. Moreover, the remainder error term ϵ is allowed to be spatially correlated as in

$$\epsilon_{it} = \rho \sum_{j=1}^{N} w_{ij} \epsilon_{jt} + \nu_{it} \tag{3}$$

where w_{ij} is the generic element of an $N \times N$ neighbourhood matrix W in which nonzero elements correspond to pairs of neighbouring countries. According to standard practice W is row-standardized so that each row sums to one; hence each error is correlated with the average of the errors in neighbouring countries according to the parameter ρ, in what is known in the spatial econometrics literature as the Spatial Autoregressive Model (SAR) (see Anselin, 1988).[10]

The two kinds of error dependence induced by omitted common factors (1) and by spatial error correlation (2) have serious consequences on the properties of estimators if they are neglected. The former induces cross-sectional correlation of a pervasive type, not dying out with distance, characterized by Pesaran and Tosetti (2011) as *strong*; moreover, if the omitted common factors are correlated with the regressors, the latter become endogenous and estimators become inconsistent (for an assessment of the properties of panel time series estimators under different omitted factors scenarios, see Coakley et al, 2006). The latter type of dependence, dubbed *weak* because it dies out with distance, has less serious consequences on estimation but can still cause inefficiency (and hence inflated standard errors and invalid inference); moreover, as discussed in the next section, it weakens consistency in the particular case of spurious panel regression.

Common Correlated Effects (CCE) estimators can be used to consistently estimate (1) with errors as in (2) and/or (3). The CCE estimators work by

[10]With reference to the notation in Anselin (1988), error spatial dependence can be written more compactly, stacking observations by time first, in the standard "spatial lag" notation as $\epsilon = \rho(W \otimes I_N)\epsilon + \nu$ where $\otimes$ is the Kronecker product.

augmenting the basic model with cross-sectional averages of both the response ($\bar{p}_t$) and regressors ($\bar{\mathbf{x}}_t$) , which pick up the effect of the common factors (see Pesaran, 2006) so that the individual slope parameters β_i can be consistently estimated by applying least squares to the augmented regression

$$p_{it} = \alpha_i + d_i t + \beta_i' \mathbf{x}_{it} + \mathbf{g}_i' \bar{\mathbf{z}}_t + e_{it} \tag{4}$$

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

$$\beta_{CCE,i} = (\mathbf{x}_i' \bar{\mathbf{M}} \mathbf{x}_i)^{-1} \mathbf{x}_i' \bar{\mathbf{M}} p_i \tag{5}$$

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 $T \times 2$ matrix of cross-sectional averages $\mathbf{z}_t$, $t = 1, \ldots T$; and a deterministic component comprising individual intercept and time trend (Pesaran, 2006, p.974). The idea of the estimator is based on cross-sectional averages as N-consistent estimators of the unobserved common factors; in a partitioned regression perspective, each individual regression (5) controls for the common deterministic component $(\alpha_i, d_i t)'$ and for the estimated common factors $\mathbf{z}_t$ through the residual operator $\bar{\mathbf{M}}$. Being robust to strong forms of cross-sectional dependence, the CCE estimator is also to weak ones like spatial correlation (see Pesaran and Tosetti, 2011). Although some alternatives are available, the CCE strategy has proved most effective in a number of simulation studies, e.g. Coakley et al (2006); Pesaran and Tosetti (2011); Kapetanios et al (2011).

CCE estimation of the overall elasticity can be performed either imposing parameter homogeneity (but maintaining heterogeneity in intercepts, factor loadings and time trends) which leads to the CCEP (pooled) estimator

$$\hat{\beta}_{CCEP} = (\sum_{i=1}^{N} \mathbf{x}_i' \bar{\mathbf{M}} \mathbf{x}_i)^{-1} \sum_{i=1}^{N} \mathbf{x}_i' \bar{\mathbf{M}} p_i \tag{6}$$

and is to be preferred on efficiency grounds when the underlying assumption that $\beta_i = \beta$ is reasonable; or parameters β_i can be left free to vary, and the average elasticity $E(\beta)$ is estimated by the Mean Groups (MG) method,

$$\hat{\beta}_{CCEMG} = \frac{1}{N} \sum_{i=1}^{N} \hat{\beta}_{CCE,i} \tag{7}$$

this last estimator being known as CCEMG.[11]

We will employ both the CCEP and CCEMG estimators[12], complementing them with pooled ordinary least squares (POLS) and two-way fixed effects (FE2) for comparison purposes, especially as regards previous results in the literature. As a rule, we will present them in order of increasing robustness, from POLS

[11] The standard pooled estimators can be seen an special cases of this more general formulation where augmentation is eliminated or reduced: pooled OLS as CCEP with $\bar{\mathbf{M}} = \mathbf{I}_T$, individual fixed effects as CCEP with $\bar{\mathbf{H}}$ containing only individual dummies. The Mean Groups (MG) estimator (see Hsiao and Pesaran, 2008, 6.4) can in turn be seen as CCEMG where $\bar{\mathbf{M}} = \mathbf{I}_T$.

[12] This is in line with Holly et al (2010) and Moscone and Tosetti (2010), while Baltagi and Moscone (2010) focus on the CCEP considering likely homogeneity of their OECD sample.

which makes the most restrictive assumptions ($\alpha_i = \alpha$, $\beta_i = \beta$, $f_t = 0$ or $\gamma_i = 0$ for each i, t) to FE2, which allows the intercept to vary in both space and time but constrains both the coefficients and factor loadings to be uniform across individuals ($\beta_i = \beta$ and $\gamma_i = \gamma$ for each i), to the CCEP which in turn allows for the full common factor specification with idiosyncratic loadings in the errors, but constrains the coefficients of interest to be homogeneous across countries ($\beta_i = \beta$ for each i) and therefore produces an estimate for the general coefficient vector β. Lastly, we will present the results for the CCEMG specification, which places no restrictions on the model in (1) and (2), considering the coefficient vector β as a random variate and hence estimating separate (augmented) regressions over time for each country and averaging the coefficients β_i to obtain an estimate of the expected value of β.

3.1.1 Nonstationarity and cointegration

The time series dimension of panel datasets raises the issue of possible nonstationarity and cointegration. In particular, should insurance premiums and GDP be nonstationary, then two situations can occur. If there exists a stationary linear combination (i.e., they are cointegrated), then this is evidence of a long-run economic relationship between them. From an econometric viewpoint, if two (single) nonstationary time series are cointegrated, then the least squares estimator of the regression parameter characterizing the relationship is superconsistent and converges to the true value faster than its stationary counterpart (Stock, 1987). If on the contrary premiums and GDP are nonstationary but not cointegrated, the statistical relationship is spurious and least squares estimates do not converge to their true values at all, while fit and significance diagnostics yield the false positive results famously discussed by Granger and Newbold (1974).

In a panel time series context, there is one more dimension available for inference: the cross section. Under certain conditions, as shown by Phillips and Moon (1999), a spurious panel data regression can still deliver a consistent estimate of long run parameters, although its convergence properties will be weaker than those of a cointegrating one. In particular, the coefficients of a spurious panel regression will still converge to their true values, although at a much slower rate $\sqrt{N}$ than that of a cointegrating panel, which is $T\sqrt{N}$. This result, however, which depends on an assumption of cross-sectional independence, is weakened if the errors are cross-sectionally weakly correlated, as with a spatial process, and can be expected to fail in presence of strong cross-sectional dependence, as would arise when omitting to control for common factors (Phillips and Moon, 1999, pages 1091-1092). Both pooled OLS (Phillips and Sul, 2003) and mean groups estimators (Coakley et al, 2006) lose their advantage in precision from pooling when cross-sectional dependence is present.

As discussed above, cross-sectional independence is unlikely in our case and the cross-sectional dimension N of the study is only moderate: hence either stationarity of all variables or cointegration are the necessary requirement to obtain reliable estimates of the income elasticity of insurance. Pesaran (2007) proposes a panel unit root test robust to cross-sectional dependence, based on appliying the same factor augmentation principle discussed above to a Dickey-Fuller regression:

$$\Delta q_{it} = \alpha_i + \delta_i t + b_i q_{i,t-1} + \sum_{j=1}^{p} d_{ij} \Delta q_{i,t-j} + \mathbf{g}_i' \bar{\mathbf{z}}_t + e_{it} \tag{8}$$

where $\bar{\mathbf{z}}_t = (\bar{q}_{t-1}, \Delta\bar{q}_t, \Delta\bar{q}_{t-1}, \ldots, \Delta\bar{q}_{t-p})$ is the matrix of cross-section averages of response and regressors, as above. Pesaran's CIPS test for a unit root in N_1 of the N time series q_i (with N_1/N tends to a fixed nonzero constant as N diverges) is based on the average of the t-ratios of the OLS estimates of the coefficients b_i in (8).

3.1.2 Cross-sectional and spatial correlation

While using robust methods from the beginning, we will nevertheless assess cross-sectional correlation both ex ante, in the preliminary statistical analysis of our sample, and ex post for critical evaluation of estimation results. Cross-sectional and spatial correlation tests can be based on a family of statistics which are all constructed as transformations of the product-moment correlation coefficient of a model's residuals, defined as

$$\hat{\rho}_{ij} = \frac{\sum_{t=1}^{T} \hat{\varepsilon}_{it} \hat{\varepsilon}_{jt}}{(\sum_{t=1}^{T} \hat{\varepsilon}_{it}^2)^{1/2} (\sum_{t=1}^{T} \hat{\varepsilon}_{jt}^2)^{1/2}} \tag{9}$$

The original LM test of Breusch and Pagan (1980) is based on the squares of $\bar{\rho}$ and is appropriate in T-asymptotic settings. The CD test is a variant by Pesaran (2004) which is based on N-asymptotics:

$$CD = \sqrt{\frac{2T}{N(N-1)}} \left(\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \hat{\rho}_{ij} \right) \tag{10}$$

Both the LM and CD tests have power against general cross-sectional dependence of either the strong type, as generated from a factor structure as in (2), or the weak type as in (3). If dependence is indeed weak, a better test can be employed which has been designed for this situation: the *local* CD or $CD(p)$ test, which is CD restricted to pairs of neighbouring observations[13]

$$CD = \sqrt{\frac{T}{\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} w(p)_{ij}}} \left(\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} [w(p)]_{ij} \hat{\rho}_{ij} \right) \tag{11}$$

where $[w(p)]_{ij}$ is the (i,j)-th element of the p-th order proximity matrix, so that if h, k are not neighbours, $[w(p)]_{hk} = 0$ and $\hat{\rho}_{hk}$ is not taken into account.[14] The $CD(p)$ test obviously improves on the power of CD under spatial correlation, as it considers only the sub-sample of most correlated pairs; the opposite happens under global cross-sectional correlation where the global CD test can make use of the full cross-sectional sample of pairs. In the following we will use CD when testing for global correlation of model residuals and the local test on first-order neighbours $CD(1)$ to detect spatial correlation as in (3) in the residuals of (4).

[13]This adaptation to irregular lattices, although straightforward, is undocumented in the original source, where a regular, "circular" world is considered.

[14]Notice that W is employed here as a binary selector for neighbouring pairs: the value of nonzero elements is irrelevant, unlike in (3) where they act as weights.

3.2 Economic specification

In this section we outline our model specification, which closely follows the framework of Beenstock et al (1988), who synthesize the findings of previous research formalizing a demand and supply system for the non-life insurance market based on theoretical predictions in terms of observable variables.

The observable quantity of interest is the equilibrium premium revenue $V = PQ$. Demand is assumed to depend positively on income Y (as the observable counterpart of wealth) and probability of loss π; and negatively on the interest rate r and the premium rate (the price of coverage) P: $Q_D = F_1(Y, \pi, r, P)$ Supply is hypothesized to vary positively with the interest rate, measuring the return on invested reserves, and the premium rate; negatively with the probability of loss, which can be seen as a measure of unit production cost: $Q_S = F_2(r, \pi, P)$ Imposing the market equilibrium condition $Q_D = Q_S$ and solving for quantity and price, Beenstock et al (1988) get equilibrium solutions for both; hence, they express total revenue V as a function of Y, π and r, where the sign of the effect is assumed positive for Y (which only shifts demand) and ambiguous for π and r, which shift both supply and demand in opposite directions. Thus premium volume is expected to depend positively from income, and with uncertain sign from interest rates and loss probability.

3.2.1 Omitted individual and common factors

Now to some potential omissions, and how they are handled in the econometric specification. While income (measured as GDP) and interest rates can be readily observed, how to proxy the effect of loss probability has proved far more controversial. On one hand, loss probability may be related to income as a measure of economic activity; urbanization has also been used for this purpose (Outreville, 1990; Browne et al, 2000; Esho et al, 2004), and more recently population density (Millo and Carmeci, 2011; Feyen et al, 2011). Loss probability, on the other hand, impacts insurance demand – not supply – through the general level of risk aversion prevailing in the population. Previous studies have tried to account for this somewhat elusive variable through synthetic indices (Esho et al, 2004). Aspects of risk aversion may be captured by education or the age structure of the population, even though the expected sign of the effect is unclear: better educated people could be more risk-conscious and therefore purchase more insurance; or they could be more efficient in managing and diversifying risk, which would lead in the opposite direction (see the discussion in Browne et al, 2000).

In this setting, the time-persistent differences in cultural values, urbanization or population density are already absorbed by fixed country effects in the homogeneous fixed effects model; additionally, long-term trends in education or in demography by the intercept and by the deterministic trend in each time series regression both in the augmented homogeneous CCEP and in the heterogeneous CCEMG models.

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 constrains the factor loadings to be equal; a CCE model (irrespective of whether CCEP or CCEMG) 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 again 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; by contrast, both CCE estimators allow for an unobserved factor affecting countries to different degrees.[15]

The inclusion of an individual time trend in each separate time series in the model augmentation 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. One-off policy measures of supernational scope, like for example European directives, are subsumed into the common factors.

3.2.2 Spatial correlation

The literature has paid scant attention to spatial diffusion processes in insurance. At first, one might consider that while reinsurance is largely globalized, direct insurance markets are essentially national and the scope of most transactions may be expected to remain largely confined inside national borders. Nevertheless, at the determinants level the role of space is substantial. Risk factors, from weather to earthquakes, are largely clustered in space. Personal characteristics of consumers, like risk aversion and awareness, behaviour and habits are geographically persistent.

We will assess whether there is any cross-section dependence and, if any, whether it has a spatial nature, which may arise from the effect on expenditure of unobservable general characteristics of neighbouring countries or from the diffusion of technology, here mainly in the design of products and of distribution channels, a notable example of the latter being the spread of the direct sales channel (telephone, Internet) from the UK to continental Europe with dramatic effects on some Motor markets. Moreover, some risk conditions (storms, hail, ice and especially earthquakes) are likely to be geographically concentrated even at the national scale.

While interesting in itself, we observe that the results of this ancillary question do not bear on the main findings of the paper regarding consistent estimation of the income elasticity, since our preferred estimators and tests on which the main empirical findings will be based are robust to spatial correlation.

3.3 The data

Insurance data on non-life annual premium volume for direct business come from Swiss Re's Sigma database (sigma, various issues), covering a maximum of 95 countries over the period 1970 to 2010. Coverage extends from 35 countries

[15] In principle, although it can hardly be the case here, loadings can have opposite signs.

in 1970 to 56 in 1980, 68 in 1990; in the Nineties coverage gradually extends to Eastern Europe, so that in 2000 only two countries (Serbia and Liechtenstein) are still missing. On the recent side, new missing values appear for Zimbabwe (since 2008) and Botswana (in 2010). Hence the length of the time series for each country is extremely variable. In this respect, Serbia and Liechtenstein have been eliminated from the beginning as having less than nine observations. Angola and Botswana are next at 12 and 13, all other countries having at least 15 observations.

GDP data have been taken from the Penn World Tables, Version 7 (Heston et al, 2011), as available in the R package 'pwt' (Zeileis and Yang, 2012). As a rule, 1970-2010 coverage is complete for most countries excepted those of the former Communist Bloc, where it begins from 1990. Therefore the availability of GDP data is no limiting factor for the length of time series. By contrast, the interest rate on deposits, taken from the World Bank database and integrated with the International Financial Statistics Yearbook of the International Monetary Fund, has more limited coverage, often not overlapping with the insurance sample.

Ex-post real interest rates are defined according to the Fisher equation (see Mishkin, 1985, p.2-3) as nominal interest rates minus (realized) inflation. This implicitly assumes that inflationary expectations are realised each period (Tease et al, 1991). The nominal interest rates on deposits are deflated by consumer price indices, as in Gagnon and Unferth (1995). The source for consumer price indices is the International Monetary Fund.

Hyperinflation periods which have been pruned out from the data are, in alphabetical order, Angola 1995-96, Argentina 1977-90, Brazil 1980-94, Croatia 1992-93, Israel 1983-85, Kazakhstan 1994, Peru 1988-91, Poland 1989-90, Russia 1995, Ukraine 1993-94, Uruguay 1976-91, Zimbabwe 2004-07.

Insurance data have been expressed in International Dollars at 2005 constant prices by deflating them through the implicit deflation factors in the Penn World Tables, transforming them to purchasing power parity (PPP) and in per-capita terms using PPP and population values from the same source.

In the next section, preliminary to estimation, we will investigate two essential characteristics of the time series at hand: cross-sectional or spatial correlation and stationarity. We will do it in this very order, as robust, so-called *second-generation* unit root tests have to be used in the presence of cross-sectional correlation (see Pesaran, 2007). Should all the variables of interest turn out to be stationary, one can consistently estimate a model on levels; else, one must proceed to assessing the stationarity of residuals and the possibility of cointegration between the variables.

3.3.1 Preliminary data analysis

As detailed above, we concentrate henceforth on the log of real per-capita premiums p, the log of real per-capita GDP y and the log c of the capitalization factor $1 + r$, where r is the real interest rate on deposits[16] Cross-sectional and spatial correlation tests are carried out in the form of CD and CD(p) tests on the

[16]The capitalization factor is chosen over the simple interest rate in order to avoid negative values, not compatible with the log specification. On grounds of data availability, the models with the interest rate dispose with observations from the countries listed in Section 3.3 and in general with substantial non-overlapping parts of the sample.

residuals of AR(2) univariate models, in order to eliminate serial dependence, as suggested by Pesaran (2004). Local (spatial) CD(p) tests are carried out using a first-order neighbourhood matrix.

	p		y		c	
CD test	17.75	***	57.13	***	31.17	***
CD(1) test	6.73	***	19.49	***	6.98	***

Table 1: CD test for global cross-sectional correlation (top) and CD(1) test for spatial correlation (bottom), both on the residuals from AR(2) models. Null hypothesis is no correlation. Test statistics are distributed as standard Normal. Significance levels corresponding to stars are: '.' 0.1; '*' 0.05; '**' 0.01; '***' 0.001.

All three variables of interest show considerable dependence both on the entire cross-section and when considering neighbouring spatial units only. This finding would invalidate commonly used unit root tests, and calls for second-generation testing procedures.

	p		y		c	
levels	-2.54		-2.20		-2.95	*
1st diff	-3.55	*	-3.50	*	-4.64	*

Table 2: CIPS test for stationarity of model variables in levels (top) and first differences (bottom). Null hypothesis is nonstationarity. The distribution of test statistics is nonstandard; (approximate) 5 percent critical values are, respectively, -2.62 for the levels test and -2.08 for the test on first differences. Test results are reported for one lag but do not change substantially for two lags. Significance levels corresponding to stars are: '.' 0.1; '*' 0.05; '**' 0.01; 0.001 critical values are not tabulated.

The variables are then tested for stationarity using Pesaran (2007)'s CIPS test which is robust to cross-sectional correlation. As expected, according to the results of the CIPS test both premiums and income can be considered integrated of order one (I(1)), their first difference being stationary, while interest rates are stationary from the beginning. A regression model relating the levels of premiums and income will therefore yield either spurious or superconsistent estimates, depending on whether the residuals turn out to be, respectively, I(1) or stationary.

4 Empirical results and discussion

In this section we review our empirical results. The ultimate goal of the exercise being to estimate the long-run behaviour of premiums w.r.t. income, we divide this task into three parts: first we estimate the static long-run specifications on levels, then we assess cointegration by testing for unit roots in the residuals from the long-run relationship; lastly, having established cointegration, we estimate an error correction model (ECM) in order to investigate the short-term dynamics together with the adjustment speed towards equilibrium.

The section is divided in two subsections, each of which answers a different research question. The first is dedicated to the estimation and diagnostic appraisal of the long-run static model, from which we will gather whether our static specification can be considered a cointegration relationship, and hence whether the estimate of the long-run income elasticity of insurance is superconsistent. The second subsection is dedicated to the short-run, dynamic error correction model, which will describe the way insurance consumption responds to income shocks and how quickly it reverts towards long-run equilibrium paths.

4.1 Long-run static model

The long-run static model is initially specified as the log of per-capita premiums at PPP, expressed in 2005 constant international dollars, (p) regressed on the log of current per-capita GDP at PPP, idem (y) and on the log of the capitalization factor (c).

In Table 3 different estimators are compared, ordered from the most to the least restrictive in the underlying assumptions: the pooled ordinary least squares (POLS) assuming homogeneity of all parameters, the two-ways fixed effects (2FE) allowing for heterogeneous time and individual effects, the common correlated effects pooled (CCEP) controlling for individual and time heterogeneity, individual trend and unobserved common factors, and lastly the common correlated effects mean groups (CCEMG) which relaxes parameter homogeneity in the CCEP framework. Hence, results are to be read left to right in order of increasing robustness, with the CCEMG being our maintained estimator, in view both of the extreme heterogeneity of the sample and of all the above mentioned concerns about omitted unobservables, and the others being reported for comparison purposes.

Income is always significant, with elasticities significantly greater than one for POLS and FE2. The CCEP and CCEMG estimates are lower, and both CCE-type estimators fail to reject the hypothesis of insurance being a normal good. Cross-sectional dependence tests point to considerable dependence, which the CCE augmentation is unable to account for completely (as witnessed by the comparison between the statistics for defactored residuals $\hat{e}_{it}$ and those for $\hat{u}_{it}$).

Testing model residuals for stationarity will reveal spurious regressions. It must be noted that in this step the defactored residuals $\hat{e}_{it}$ from the CCEP and CCEMG models are used, because for the moment we are interested in the statistical properties of the augmented regression model (4) rather than of the base specification (1). Apart from those of the pooled OLS model, tests on all other residuals support stationarity, the FE2 only if considering one lag, those of the CCE models in both cases. This finding means that we can safely rely upon the static long-run estimates from CCEP and CCEMG models.

It is noteworthy how the standard pooled estimators, both with and without fixed effects, would seem to confirm the results Beenstock et al (1988) drew from the 1970-1981 version of the dataset: income elasticity significantly greater than one (i.e., insurance as a superior good) and a positive effect of real interest rates. Yet the stationarity diagnostics make us reject the relationships as spurious, forcefully for the pooled model without fixed effects, while more marginally for the FE2, as witnessed by the CIPS tests in 3.Moreover, the CD tests show evidence of cross-sectional correlation, casting further doubt on the consistency of these estimators, although the latter problem is greatly alleviated by the

	POLS		FE2		CCEP		CCEMG	
y	1.146	***	1.491	***	0.908	*	0.837	***
	(0.02)		(0.04)		(0.35)		(0.17)	
c	0.255	.	0.073	.	-0.009		0.118	
	(0.14)		(0.04)		(0.18)		(0.18)	
CD test on u	48.97	***	2.03	*	77.39	***	31.16	***
CD test on e	48.97	***	2.03	*	23.89	***	20.41	***
CD(1) test on u	13.88	***	2.49	*	20.93	***	7.08	***
CD(1) test on e	13.88	***	2.49	*	7.05	***	6.33	***
CIPS(1) test	-0.93		-1.68		-2.85	*	-3.14	*
CIPS(2) test	-0.65		-1.38		-2.4	*	-2.69	*
Test b(y)=1	45.5	***	124.29	***	0.07		0.93	
Obs.	2234		2234		2234		2234	
Countries	84		84		84		84	
T min./max.	13–40		13–40		13–40		13–40	

Table 3: Long-run models of per-capita premiums vs. per-capita income, both at PPP, and deposit yields; all variables in logs. Left to right: pooled ordinary least squares (POLS), two-way fixed effects (FE2), common correlated effects pooled (CCEP) and common correlated effects mean groups (CCEMG). Top to bottom: coefficient estimates; CD test for cross-sectional dependence on standard residuals (u), CD test on defactored residuals (e), CD(1) test for spatial dependence: null hypothesis is no dependence; CIPS test for stationarity of model residuals with augmentation orders 1 and 2: null hypothesis is nonstationarity, the (approximate) 5 percent critical value is -1.53; Wald test for b(y)=1 (income elasticity of insurance is unity); total number of observations; number of countries; minimum and maximum length of time series. Significance levels corresponding to stars are: '.' 0.1; '*' 0.05; '**' 0.01; '***' 0.001. The 0.001 critical values for the CIPS test are not tabulated.

inclusion in FE2 of time effects. By contrast, accounting for the influence of unobserved common factors makes us change our empirical conclusions drastically: the income elasticity is lower than, but not significantly different from, one; and the effect of real interest rates is not significant.

As far as the effect of interest rates is concerned, there is extreme variability in the models' results. Three models give positive values of much different magnitude, those of the POLS and FE2 marginally significant, that of the CCEMG not significant; the CCEP yields a negative but non significant estimate. The economic role of the interest rate is therefore unclear; but fortunately, this is not part of our main research question, for the purpose of which we conclude that it is statistically admissible to eliminate the interest rate from the CCE models. For this reason we henceforth concentrate on the model

$$p_{it} = \alpha_i + d_i t + \beta_i' y_{it} + \mathbf{g}_i' \bar{\mathbf{z}}_t + u_{it} \tag{12}$$

with $\bar{\mathbf{z}}_t = (\bar{p}_t, \bar{y}_t)'$, i.e. with income as the only regressor, and errors as in (2) and (3). Restricting our attention to premiums and income has important additional advantages, allowing us to draw on a larger and more stable sample; moreover, being stationary, interest rates are guaranteed not to play any role in the cointegration analysis.

The results from model (12) reported in Table (4) do not qualitatively change our findings: POLS and FE2 models would strongly indicate an elasticity significantly greater than one; yet, as they have nonstationary and cross-sectionally correlated residuals, these estimates are spurious and unlikely to be consistent. CCEP and CCEMG estimates are very close between themselves and not far from POLS (while FE2 is larger); but their estimated dispersion is much bigger than that of pooled models. The statistical qualities of both CCE estimates are satisfactory, with stationary (defactored) residuals; cross-sectional correlation (global and, to a lesser extent, local) in residuals is greatly reduced through defactoring and, although still present, not a concern for these estimates. The hypothesis of insurance as a normal good is not rejected in either of CCE models at any confidence level. We conclude maintaining the hypothesis $\beta = 1$.

	POLS		FE2		CCEP		CCEMG	
y	1.098	***	1.181	***	1.097	**	1.092	***
	(0.02)		(0.05)		(0.36)		(0.29)	
CD test on u	69.16	***	10.78	***	69.36	***	42.82	***
CD test on e	69.16	***	10.78	***	20.79	***	15.94	***
CD(1) test on u	20.14	***	5.62	***	20.18	***	11.99	***
CD(1) test on e	20.14	***	5.62	***	3.06	**	3.92	***
CIPS(1) test	-1.06		-1.54		-3.2	*	-3.42	*
CIPS(2) test	-0.87		-1.31		-2.76	*	-2.89	*
Test b(y)=1	22.07	***	12.37	***	0.07		0.1	
Obs.	2598		2598		2598		2598	
Countries	84		84		84		84	
T min./max.	14–40		14–40		14–40		14–40	

Table 4: Long-run models of per-capita premiums vs. per-capita income, both at PPP; all variables in logs. Left to right: pooled ordinary least squares (POLS), two-way fixed effects (FE2), common correlated effects pooled (CCEP) and common correlated effects mean groups (CCEMG). Top to bottom: coefficient estimates; CD test for cross-sectional dependence on standard residuals (u), CD test on defactored residuals (e), CD(1) test for spatial dependence: null hypothesis is no dependence; CIPS test for stationarity of model residuals with augmentation orders 1 and 2: null hypothesis is nonstationarity, the (approximate) 5 percent critical value is -1.53; Wald test for b(y)=1 (income elasticity of insurance is unity); total number of observations; number of countries; minimum and maximum length of time series. Significance levels corresponding to stars are: . 0.1; * 0.05; ** 0.01; *** 0.001. The 0.001 critical values for the CIPS test are not tabulated.

We now turn to an ex-post assessment of the estimated relationship between premiums and income from the point of view of cointegration.

4.1.1 Cointegration analysis

Given the nonstationary nature of both (log real) premiums and (log real) income ascertained in Section 3.3.1, in this section we test the hypothesis that there exists a cointegrating vector $(1, -\beta)$ whereby their linear combination is stationary. From an economic viewpoint, this would imply the existence of a stable long-run relationship between premiums and income and the possibility of characterizing short-run behaviour by an error correction representation according to which the system reverts towards the long-run equilibrium in response to temporary deviations. In particular, given that the hypothesis of interest $\beta = 1$ was not rejected, we test for the cointegrating vector $(1, -1)$.

A number of different approaches to panel cointegration testing have emerged in the recent literature, usually based on two-step procedures à la Engle and Granger (1987) whereby the residuals from a first-stage regression on levels between the variables of interest are tested for unit roots (see the review in Holly et al, 2010, 3.4). We follow the approach in Moscone and Tosetti (2010) and Holly et al (2010), which allows for common factors and spatial correlation in both steps of the test. Accordingly, we base our test on the results of the CCEMG and of the CCEP regression, which last, notably, is a consistent estimator for β even under the heterogeneity hypothesis (Holly et al, 2010, p.165 and 5.2). As the hypothesis that $\beta = 1$ cannot be rejected based on either set of estimates, an equivalent relationship to be tested is given by

$$\hat{u}_{it} = p_{it} - \hat{\alpha}_i - y_{it}$$

with $\hat{\alpha}_{it} = T^{-1} \sum_{t=1}^{T} (p_{it} - y_{it})$.

Again in accordance with Holly et al (2010, 5.3), we apply the CIPS test (8) with one and two lags and no deterministic component to $\hat{u}_{it}$ obtaining $CIPS(1) = -2.25$ and $CIPS(2) = -2.15$, both significant at the 1 percent level.[17] We conclude in favour of the hypothesis of panel cointegration between premiums and income with cointegrating vector $(1, -1)$. Having established cointegration, we can now proceed to estimating an error correction model in order to investigate the dynamic behaviour of the system.

4.2 Dynamic error correction model

The error correction (ECM) specification

$$\Delta p_{it} = \delta_{0i} + \delta_{1i} \Delta y_{it} + \phi_i (p_{i,t-1} - y_{i,t-1}) + \delta_{2i} \Delta p_{i,t-1} + \delta_{3i} \Delta y_{i,t-1} + \eta_{it} \quad (13)$$

will allow us to look into two further aspects: the short-term dynamics and the speed of adjustment towards the long-run equilibrium.[18] Results are reported in Table 5.

The POLS and FE2 models are again reported for completeness, although they are more appropriate in this difference specification where all regressors are stationary. Unsurprisingly, all signs of spurious regression are gone: all CIPS tests (of which only CIPS(2) is reported) reject the hypothesis of unit roots in the residuals. Unlike the POLS estimator, the FE2 also seems to control quite

[17]As a robustness check, we perform the same test using the exact estimate of $\hat{\beta}_{CCEP}$ as in Moscone and Tosetti (2010): we extract the (non-defactored) residuals $\hat{u}_{it} = p_{it} - \hat{\alpha}_i - \hat{\beta}_{CCEP} y_{it}$ and apply the CIPS test, obtaining $CIPS(1) = -1.71$ and $CIPS(2) = -1.53$, the first significant at one percent level, the second roughly corresponding to the five percent critical value.

[18]We have experimented with adding lags to the ECM, to control for the possibility of a richer dynamics. Further lags up to 3 were not significant; results are not reported and are available upon request.

	OLS		FE2		CCEP		CCEMG	
dy	0.873	***	0.983	***	0.968	**	0.711	***
	(0.06)		(0.06)		(0.35)		(0.15)	
EC(-1)	-0.098	***	-0.119	***	-0.164		-0.267	***
	(0.01)		(0.01)		(0.15)		(0.03)	
dp(-1)	0.268	***	0.178	***	0.132		0.151	***
	(0.02)		(0.02)		(0.12)		(0.03)	
dy(-1)	-0.236	***	-0.121	.	-0.188		0.278	**
	(0.06)		(0.07)		(0.49)		(0.1)	
Half-life	6.72		5.47		3.87		2.23	
CD test on u	20.61	***	0.49		30.95	***	30.48	***
CD test on e	20.61	***	0.49		0.43		2.1	*
CD(1) test on u	7.11	***	4.11	***	9.58	***	7.77	***
CD(1) test on e	7.11	***	4.11	***	3.23	**	2.05	*
CIPS(2) test	-2.99	*	-2.96	*	-3.44	*	-3.07	*
Obs.	2169		2169		2169		2169	
Countries	67		67		67		67	
T min./max.	18–38		18–38		18–38		18–38	

Table 5: Error Correction models of insurance and income; all variables in logs. Left to right: two-way fixed effects (FE2), common correlated effects pooled (CCEP) and common correlated effects mean groups (CCEMG).

successfully for cross-sectional correlation, with the CD statistic not significant, as that of the CCEP, while that of CCEMG is marginally significant.

The homogeneous CCEP estimator, restricting all coefficients to be equal across countries, suffers from the lowest precision in the lot. The results are qualitatively rather similar to those of the FE2, but the dispersion of the estimates is wider by an order of magnitude, signalling a heterogeneity problem. The more robust heterogeneous CCEMG estimator, by contrast, has much lower variance; robustness checks in Tables (8) and (9) on samples restricted to countries with a minimum length of available data, reveal that while the CCEMG is relatively stable throughout, the CCEP estimates tend to vary considerably as the sample is extended to countries with “shorter” time coverage, which can also be expected to have the most heterogeneous behaviour. As the sample is restricted to countries with 30 years or more of data, thus excluding smaller countries and “young” markets like Eastern Europe’s ones, the CCEP estimates tend to converge towards the CCEMG ones. Once again, this is evidence in favour of the appropriateness of the heterogeneous specification: our conclusions will therefore be based on CCEMG estimates.

On average, the short-run income elasticity of non-life insurance is significantly lower than one, market premiums being also positively affected by lagged changes in income. This behaviour is consistent with the institutional lag due to the yearly timespan of most non-life insurance contracts overlapping with the calendar year. The attractor term estimating the strength of the tendency towards the long-term equilibrium has the expected negative sign and considerable magnitude, implying a half-life of about 27 months until 50 percent absorption of any income shock. All in all, the short-term dynamics of the insurance market seems to follow income variations rather closely.

5 Conclusions

This article has been dedicated to a reappraisal of the results from a seminal paper of the Eighties using modern econometric techniques and drawing on a much improved version of the same dataset.

Beenstock et al (1988) were the first to approach the relationship between income and non-life insurance consumption from an international perspective, drawing on a database by Swiss Re which has since become the standard for international insurance comparisons. They conducted both a cross-sectional study on 45 countries in 1981 and a pooled time series analysis on 1970-1981 data for 12 developed countries, concluding that "[non-life] insurance is a superior good and is disproportionately represented in economic growth", which leads to the prediction that insurance penetration, ceteris paribus, be in turn rising with income. They also concluded that insurance consumption grows with real interest rates, while they did not find evidence of short-run effects of the economic cycle.

We still consider their theoretical model to be the most appropriate foundation for empirical work. From a methodological viewpoint, while their cross-sectional results are likely to suffer from unobserved heterogeneity, we support their use of pooled time series as the most promising approach to the research question at hand. Standing on the shoulders of researchers from the last two decades, and drawing on a modern version of the same database now counting 93 countries, the better part of which observed over 20 to 40 years, we reassessed their findings with particular care for methodological issues as unobserved heterogeneity and common factors, nonstationarity and spatial correlation. In our exposition we present pooled models with and without fixed country and time effects, comparing them with augmented common correlated effects (CCE) estimators and supporting the use of the latter both from a theoretical, ex ante viewpoint and based on ex-post diagnostics. Applied to the new dataset, the traditional estimators confirm the original findings of Beenstock et al (1988) while the consistent CCE estimators do not.

The average long-run income elasticity of non-life insurance turns out to be statistically not different from one. 25 years from the original paper, we therefore reverse Beenstock et al (1988)'s conclusions, characterizing non-life insurance as a normal good. In other words, we find no support for the prediction that, holding all other factors equal, insurance penetration on GDP, and hence the role of the insurance sector in the economy, be growing with economic development. As testified by the pooled estimators we report for comparison purposes, we rather attribute Beenstock et al (1988)'s findings to neglected heterogeneity in coefficients, to the influence of unobserved common factors and country-specific effects and trends.

Moreover, the evidence of a cointegrating relationship between non-life insurance consumption and income means that world insurance markets tend, on average, to grow in line with the general economy, reacting in a less-than-proportional way to income shocks in the short run but then reverting to its long-run path according to an error-correction mechanism. The attractor coefficient governing the return of the system to long-run equilibrium turns out to be quite high in absolute value, implying a relatively short half-life of approximately

27 months.

Lastly, and again differently from Beenstock et al (1988)'s findings, real interest rates do not play a statistically significant role, which we take as evidence for the conflicting influence of investment yields on both insurance demand and supply, rather than for a lack of importance of financial aspects.

The main result of the paper is that non-life insurance markets cannot, according to our evidence, be expected to benefit more than proportionally from economic growth, but rather to follow it quite closely. Further research directions, which we do not pursue here, would involve the disaggregation of non-life premium income into individual lines of business in order to verify whether this behaviour is uniform across lines of business or, on the contrary, resulting from a compensation between different lines.

References

Anselin L (1988) Spatial Econometrics: Methods and Models. Kluwer Academic Publisher, University of California, Santa Barbara

Arellano M, Bond SR (1991) Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. Review of Economic Studies 58:277–297

Arellano M, Bover O (1995) Another look at instrumental-variable estimation of error components models. Journal of Econometrics 68:29–52

Baltagi B, Moscone F (2010) Health care expenditure and income in the oecd reconsidered: Evidence from panel data. Economic Modelling 27(4):804–811

Beenstock M, Dickinson G, Khajuria S (1986) The determinants of life premiums: an international cross-section analysis 1970-1981. Insurance: Mathematics and Economics 5:261–270

Beenstock M, Dickinson G, Khajuria S (1988) The relationship between property-liability insurance premiums and income: an international analysis. The Journal of Risk and Insurance 55(2):259–272

Borch K (1986) Insurance and giffen's paradox. Economics Letters 20(4):303–306

Breusch T, Pagan A (1980) The lagrange multiplier test and its application to model specifications in econometrics. Review of Economic Studies 47:239–253

Briys E, Dionne G, Eeckhoudt L (1989) More on insurance as a giffen good. Journal of Risk and Uncertainty 2(4):415–420

Browne MJ, Chung J, Frees EW (2000) International property-liability insurance consumption. The Journal of Risk and Insurance 67(1):73–90

Coakley J, Fuertes AM, Smith RP (2006) Unobserved heterogeneity in panel time series models. Computational Statistics and Data Analysis 50(9):2361–2380

Dezhbakhsh H (1990) The inappropriate use of serial correlation tests in dynamic linear models. The Review of Economics and Statistics 72(1):126–132

Engle R, Granger C (1987) Co-integration and error correction: representation, estimation, and testing. Econometrica 55(2):251–276

Esho NA, Kirievsky A, Ward D, Zurbruegg R (2004) Law and the determinants of property-casualty insurance. The Journal of Risk and Insurance 71(2):265–283

Feyen E, Lester R, Rocha R (2011) What drives the development of the insurance sector. Policy Research Working Paper 5572, World Bank

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

Granger CW, Newbold P (1974) Spurious regressions in econometrics. Journal of Econometrics 2

Hau A (2008) When is a coinsurance-type insurance policy inferior or even giffen? The Journal of Risk and Insurance 75(2):343–364

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/

Holly S, Pesaran M, Yamagata T (2010) A spatio-temporal model of house prices in the usa. Journal of Econometrics 158(1):160–173

Hoy M, Robson AJ (1981) Insurance as a giffen good. Economics Letters 8(1):47–51

Hsiao C, Pesaran M (2008) Random coefficient models. In: Matyas L, Sevestre M (eds) The Econometrics of Panel Data, 3rd Ed., Springer Netherlands, chap 6, pp 185–213

Hussels S, Ward D, Zurbruegg R (2005) Stimulating the demand for insurance. Risk Management and Insurance Review 8(2):257–278

Kapetanios G, Pesaran M, Yamagata T (2011) Panels with non-stationary multifactor error structures. Journal of Econometrics 160(2):326–348

Leisch F (2002) Sweave: Dynamic generation of statistical reports using literate data analysis. In: Härdle W, Rönz B (eds) Compstat 2002 - Proceedings in Computational Statistics, Physica Verlag, Heidelberg, pp 575–580

Millo G, Carmeci G (2011) Non-life insurance consumption in italy. a subregional panel data analysis. Journal of Geographical Systems 13(3):273–298

Mishkin F (1985) The real interest rate: a multi-country empirical study. Working Paper 1047, NBER

Moscone F, Tosetti E (2010) Health expenditure and income in the united states. Health Economics 19(12):1385–1403

Mossin J (1968) Aspects of rational insurance purchasing. Journal of Political Economy 76(4):553–568

Outreville J (2012) The relationship between insurance and economic development: 85 empirical papers for a review of the literature. Risk Management and Insurance Review 00(0):1–52

Outreville JF (1990) The economic significance of insurance markets in developing countries. The Journal of Risk and Insurance 18(3):487–498

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 M, Tosetti E (2011) Large panels with common factors and spatial correlation. Journal of Econometrics 161(2):182–202

Pesaran MH (2004) General diagnostic tests for cross section dependence in panels. Working Paper 1229, CESifo

Phillips P, Moon H (1999) Linear regression limit theory for nonstationary panel data. Econometrica 67(5):1057–1111

Phillips P, Sul D (2003) Dynamic panel estimation and homogeneity testing under cross section dependence. The Econometrics Journal 6:217–259

R Development Core Team (2012) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, URL http://www.R-project.org/, ISBN 3-900051-12-7

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

Stock J (1987) Asymptotic properties of least squares estimators of cointegrating vectors. Econometrica 55(5):1035–1056

Tease W, Dean A, Elmeskov J, Hoeller P (1991) Real interest rate trends: the influence of saving, investment and other factors. Economic Studies 17, OECD

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

Zingales L (2003) The weak links. Federal Reserve of S Louis Review 7:47–52

Appendix 1: sensitivity analysis

In this appendix we perform some checks for the sensitivity of our two basic models, the long run relationship and the ECM, estimated by CCEMG, to variations in sample composition. We consider four different, mutually inclusive subsamples containing respectively only those countries which have at least 15, 20, 25 and 30 observations in time.

Sample sensitivity check, long run model

	T > 14		T > 19		T > 24		T > 29	
log(rgdpl)	0.701	***	0.651	**	1.03	***	1.195	***
	(0.16)		(0.22)		(0.15)		(0.17)	
log(1+r)	0.043		-0.002		0.157		0.114	
	(0.19)		(0.19)		(0.18)		(0.19)	
CD test	15.96	***	20.02	***	14.6	***	2.75	**
CIPS(2) test	-2.6	*	-2.67	*	-2.65	*	-2.83	*
Test rgdpl=1	3.41	.	2.5		0.04		1.28	
Obs.	2137		1793		1527		1308	
Countries	77		57		45		37	
T min./max.	15–40		20–40		25–40		30–40	

Table 6: Sensitivity analysis to sample composition of CCEMG long-run models of per-capita premiums vs. per-capita income, both at PPP, and interest rates; all variables in logs. Left to right: countries with respectively at least 15, 20, 25 and 30 observations in time.

	<0	[0,1]	>1
T>14	0.12	0.44	0.44
T>19	0.18	0.46	0.37
T>24	0.11	0.51	0.38
T>29	0.11	0.27	0.62

Table 7: Sensitivity analysis to sample composition of coefficients' dispersion: CCEMG long-run models of per-capita premiums vs. per-capita income, both at PPP; all variables in logs. Top to bottom: countries with respectively at least 15, 20, 25 and 30 observations in time. Left to right: share of individual coefficient estimates in each class.

Sample sensitivity check, ECM model

	T > 14		T > 19		T > 24		T > 29	
dy	0.91	*	0.917	*	0.761	***	0.796	***
	(0.41)		(0.36)		(0.13)		(0.14)	
EC(-1)	-0.246		-0.177		-0.22	***	-0.235	***
	(0.16)		(0.16)		(0.07)		(0.06)	
dp(-1)	0.019		0.037		0.06		0.043	
	(0.12)		(0.12)		(0.06)		(0.06)	
dy(-1)	-0.049		-0.103		0.152		0.145	
	(0.53)		(0.46)		(0.1)		(0.11)	
Half-life	2.45		3.56		2.79		2.59	
CD test	-0.69		-2.33	*	-3.05	**	-3.02	**
CIPS(2) test	-3.04	*	-3.16	*	-3.03	*	-3.15	*
Obs.	2314		1921		1638		1375	
Countries	77		57		45		37	
T min./max.	14–38		18–38		23–38		28–38	

Table 8: Sensitivity analysis to sample composition of CCEP dynamic error correction models. Left to right: countries with respectively at least 15, 20, 25 and 30 observations in time.

	T > 14		T > 19		T > 24		T > 29	
dy	0.589	***	0.631	***	0.661	***	0.661	***
	(0.09)		(0.09)		(0.09)		(0.1)	
EC(-1)	-0.399	***	-0.287	***	-0.252	***	-0.251	***
	(0.04)		(0.03)		(0.03)		(0.03)	
dp(-1)	0.131	***	0.132	***	0.1	**	0.101	**
	(0.03)		(0.03)		(0.03)		(0.04)	
dy(-1)	0.263	*	0.171	*	0.24	**	0.239	**
	(0.11)		(0.09)		(0.08)		(0.09)	
Half-life	1.36		2.05		2.39		2.4	
CD test	1.07		-0.63		-1.91	.	-1.92	.
CIPS(2) test	-2.85	*	-2.98	*	-3.04	*	-3.16	*
Obs.	2314		1921		1638		1375	
Countries	77		57		45		37	
T min./max.	14–38		18–38		23–38		28–38	

Table 9: Sensitivity analysis to sample composition of CCEMG dynamic error correction models. Left to right: countries with respectively at least 15, 20, 25 and 30 observations in time.