Cass Business School
CITY UNIVERSITY LONDON

Faculty of Actuarial Science and Insurance

Actuarial Research Paper
No. 197

Are the Dimensions of Private Information More Multiple than Expected?
Information Asymmetries in the Market of Supplementary Private Health Insurance in England*

Martin Karlsson
Florian Klohn
Ben Rickayzen

April 2012

ISBN 978-1-905752-31-7

Cass Business School
106 Bunhill Row
London EC1Y 8TZ
Tel +44 (0)20 7040 8470
www.cass.city.ac.uk

Are the Dimensions of Private Information More Multiple than Expected? Information Asymmetries in the Market of Supplementary Private Health Insurance in England*

Martin Karlsson† Florian Klohn‡ Ben Rickayzen§

April 16, 2012

†Technische Universität Darmstadt (TU Darmstadt)
‡Corresponding author: Technische Universität Darmstadt (TU Darmstadt), Marktplatz 15 - Residenzschloss, 64283 Darmstadt, e-mail: klohn@vwl.tu-darmstadt.de
§Cass Business School, London
*We would like to thank Astrid Kiil, Stefan Pichler, Robert Plumb, Hendrik Schmitz, Udo Schneider and participants of the 2011 DIBOGS Workshop in Hamburg) for helpful comments on previous versions of this paper.

Abstract

Our study reexamines standard econometric approaches for the detection of information asymmetries on insurance markets. We claim that evidence based on a standard framework with 2 equations, which uses potential sources of information asymmetries, should stress the importance of heterogeneity in the parameters. We argue that conclusions derived from this methodology can be misleading if the estimated coefficients in such an 'unused characteristics' framework are driven by different parts of the population.

We show formally that an individual's expected risk from the perspective of insurance, conditioned on certain characteristics (which are not used for calculating the risk premium), can equal the population's expectation in risk – although such characteristics are both related to risk and insurance probability, which is usually interpreted as an indicator of information asymmetries.

We provide empirical evidence on the existence of information asymmetries in the market for supplementary private health insurance in the UK. Overall, we found evidence for advantageous selection into the private risk pool; ie people with lower health risk tend to insure more. The main drivers of this phenomenon seem to be characteristics such as income and wealth. Nevertheless, we also found parameter heterogeneity to be relevant, leading to possible misinterpretation if the standard 'unused characteristics' approach is applied.

Keywords: Information Asymmetries; Insurance markets; Applied Econometrics

JEL classification: I11; I13; C18

1 Introduction

More than 15 per cent of UK citizens have either full private health cover, or partial private health cover, in addition to their use of the National Health Service (NHS)[1]. 41 per cent of the adult population have life insurance; by contrast, as many as 68 per cent play the national lottery or gamble[2]. There are several possible explanations for differences in peoples' private insurance requirements and they can be explained by the supply and demand side of the insurance market. One commonly given explanation is that individual preferences influence the probability that a person buys insurance. Another explanation regards the heterogeneity in risk which the insurance company does not take into account when calculating the risk premium, but is taken into account by customers making decisions to buy insurance. However, the converse may also be true: some risk-related characteristics of customers are used by insurance companies to lower their exposure to high risk policyholders.The evaluation of information asymmetries (IA) is still a big issue in economic research. However, certainly within the UK, very little research has been done in this area up to now.

The level of health experienced by individuals clearly directly affects their happiness, contribution to the economy and their ability to participate within society at large. There will always be a disparity in health status between members of a population. Furthermore, providing healthcare for a population is becoming increasingly expensive; in part due to technical advances in treatments and also due to demographic change (ie an ageing population). It is therefore becoming even more important to have a health insurance market which allocates its resources as efficiently as possible.

Since Akerlof (1970) and Rothschild and Stiglitz (1976) wrote their seminal works on market efficiency in insurance markets, there has been a preponderance of theoretical and empirical research in this field. A key area of research remains, and that is how information asymmetries and their consequences can be accurately and adequately measured. This area is so important as it directly affects the funding of a country's health and welfare system.

The aim of this paper is twofold: Firstly, after discussing some commonly used tests for the detection of information asymmetries, we show formally that an approach which allows the detection of information asymmetries due to specific characteristics can lead to the wrong conclusions being made. This is the case if estimated coefficients for a potential source of IA in a framework with two equations are driven by different parts of the population. Secondly, accounting for this issue, we provide empirical evidence on the existence of information asymmetries on the market for private health insurance for the English age 50+ population.

In this paper, we start with an introduction of the institutional background to the current health care provision in the UK. We then provide a literature overview with an emphasis on several commonly used tests to identify information asymmetries. Following this, we show that under specific circumstances a test based on two equations is misleading. In the empirical section of this paper we compare evidence based on such a test, with specifications allowing for parameter heterogeneity. Our empirical analysis is based on the English Longitudinal Survey of Ageing (ELSA) survey; an individual level dataset representative of the UK's population over age 50. After discussing our findings, we draw conclusions and make suggestions for future economic research in this field.

[1]http://www.ess-europe.de/en/uk_health_insurance.htm

[2]http://www.principlefirst.co.uk/insurance-news/more-play-lotto-than-buy-life-insurance-cover/

2 Theoretical considerations

2.1 Institutional Background

The population of the United Kingdom is entitled to free healthcare, and this is provided by the National Health Service (NHS) through primary care (general practice) and secondary care (hospital based care given through both NHS and Foundation trusts).[3] The main principle of the NHS is to make health services available to every single UK citizen who is in need. However, in practice, there are a number of treatments which are not available within the system. Most of the latter are excluded because they are viewed as being non-essential, but some are excluded for financial reasons.[4] In addition to the public provision of healthcare via the NHS, individuals can choose to top up their provision through the purchase of private health insurance. This might be done on an individual basis or as part of the benefits package offered by the employer. The insurance usually provides cover for enhanced services such as faster access and a wider consumer choice, compared to what is offered by the NHS. Insurers can freely determine their services, but most packages cover surgery as an inpatient or day case, hospital accommodation, nursing care and inpatient tests. Since there is no regulation on products or pricing (Boyle, 2011), we can assume that the market for private supplementary health insurance in the UK is competitive. Although a competitive market should result in an actuarially fair risk premium due to the possibility of consumers switching between different health care plans, we cannot infer any conclusions about the efficiency in this market, i.e selection effects due to information asymmetries that are not accounted for in the risk premium. There is a broad range of literature on theoretical aspects of IA and how to measure them.

2.2 The Detection of Information Asymmetries

The empirical evidence on the existence of IA is mixed. In a recent article, Cohen and Siegelman (2010) provide a metastudy on testing for adverse selection on a wide range of insurance markets. They focus on the positive correlation approach and find that a risk-insurance-correlation exists in some studies but not in others. For example, the evidence in the market for health insurance seems to be strongly heterogeneous. Looking at studies which focus on the US market, they find evidence for both the existence of information asymmetries and market efficiency. Furthermore, they claim that one has to distinguish between different kinds of IA. They conclude that it might be of interest to evaluate the circumstances under which adverse selection seems to or does not seem to arise. This point of view is especially relevant from a policy perspective, given that if any institutional changes were made, we would like to predict efficiency changes which occur from the initial market conditions. In their work, Cohen and Siegelman (2010) emphasize mainly on an approach for the detection of IA, developed by Chiappori and Salanié (1997). This test is still widely used today, despite ongoing developments in the area. The main thrust of the test is to jointly estimate two different equations. The first one captures the probability of buying an insurance contract, given the information about an individual which an insurance company will use for calculating the risk premium. The second equation measures the correlation of these variables with the probability of the insurer making a loss on the contract. The error term of both equations will cover all the information in respect of both events which are not used for pricing purposes. If risk and insurance coverage are correlated, this is usually interpreted as indicating that a self selection process is occurring. Hence it is useful to estimate the correlation

[3] http://www.nhshistory.com/
[4] http://www.londonhealth.co.uk/nhs/index.html

between both the equations' error terms. This approach is often called the 'positive correlation test'.

Formally this approach can be described by the following equations, where I is an indicator for insurance status and R and indicator for being at risk:

$$I = X\delta + \epsilon \tag{1}$$

and

$$R = X\beta + \eta \tag{2}$$

Another approach which is instead based on a single equation estimate is suggested by Dionne et al. (2001) which evaluates IA in the automobile insurance market. Their approach is based on the relation

$$I = X\delta_1 + A\delta_2 + \hat{A}\delta_3 + \epsilon \tag{3}$$

Here, insurance probability is the dependent variable and items within the consumer information held by the insurance company are the explanatory variables. As further control variables the authors use the number of accidents (A) as well as the expected number of accidents ($\hat{A}$). These are calculated based on consumer demographics and will capture nonlinearities in the initial exogenous control variables which the insurance company uses for pricing.

Finkelstein and McGarry (2006), however, argue that such a positive correlation of the error terms in the 'Chiappori approach' is neither a necessary nor a sufficient condition for the existence of information asymmetries. They suggest that misleading results of such a test might arise if more than one characteristic has an impact on both dependent variables, and that both effects cancel out on average. For example, in additional to an individual's risk class, risk preference heterogeneity of the consumers might offset the correlation of the two equations' error terms. The authors state that if an econometrician can observe such relevant information, and this information is not used for pricing by the insurer, then an inclusion of this variable as an additional explanatory variable into equations (1) and (2) above will make it possible to detect this kind of self selection separately – despite the second relevant variable having the 'offsetting' effect. This approach, which we call the 'unused characteristics approach', is based on the equations:

$$I = X\delta_1 + Z\delta_2 + \epsilon \tag{4}$$

and

$$R = X\beta_1 + Z\beta_2 + \eta. \tag{5}$$

Z represents a matrix containing additional information about the insured but which is not used for pricing. The condition for recognising information asymmetries would be that any new variable being included in the model has an impact both on insurance probability and of 'being at risk'. In their study, Finkelstein and McGarry are able to use information that is assumed to be unknown to the insurer in the market for long term care (LTC) in the US. In their econometric model, they chose the self stated probability of utilising nurse help to be a possible source for adverse selection. For advantageous selection they used investment in preventative health care and the usage of seatbelts, both of which can be interpreted as private information not used for pricing purposes by an insurance company. In addition, they compare their findings with those of two other approaches to detect information asymmetries – whose results are quite similar. The first approach is the classical test of correlating the error terms of the two equations; the second is a one-equation model where risk class is the dependent variable and insurance status is an independent variable. It is a slightly modified version of the positive correlation test suggested by Dionne et al. (2001).

The 'unsused characteristics' approach just described can be also useful if we are interested in market efficiency with regard to pricing, without using *observables*. Compared to the case mentioned above, Finkelstein and Poterba (2006) focus on a scenario when insurance companies observe, or could observe, relevant characteristics of their customers, but do not use this information when calculating the risk premium. When analysing the UK annuity market, they show that annuity purchases and the annuitant's mortality are regionally correlated. Making the assumption that regional information is not used for calculating the risk premium, this is interpreted as an indicator of adverse selection. Finkelstein and Poterba's results raise the question as to why insurance companies do not use such relevant information.

Cutler et al. (2008) look for information asymmetries within several insurance markets in the US, based on data from the Health and Retirement Study. They also use a 2-equation model. One equation has insurance status and the other has risk occurrence as dependent variables. While conditioning on variables which are used for insurance pricing, they also include some behavioural variables which are used to measure heterogeneity in risk preference (seatbelt usage, preventive activities) and individual risk behaviour (eg smoking, drinking). Their findings show that individuals with characteristics that imply low risk have a higher probabilty of buying an insurance product. The expected claims of these individuals suggest advantageous selection in the market for life and LTC insurance, but adverse selection for annuities. For Medigap and acute health insurance no systematic relationship between risk-characteristics and expected claims was found.

Dardanoni and Li Donni (2008) use the same type of test but implement it in a different way. They use a finite mixture model to identify types which are unobservable due to heterogeneous risk behaviours and risk classes, and they impose the assumption that these unobserved factors are constant. Their findings suggest both asymmetric information and multidimensional unobserved heterogeneity in the LTC market in the US.

There are some studies on IA in insurance markets using 'self-assessed health' for capturing individual risk, as this study does. For example, Doiron et al. (2008) analysed the relationship between health status and private health insurance in Australia. They used self assessed health as their health risk indicator. One feature of their study, and a difference from the institutional background in our study, is that insurance companies were not allowed to calculate premiums based on individual characteristics, and furthermore all applicants had to be accepted for cover. They therefore argue that the correlation between explanatory variables is due to consumer preferences rather than insurance actions. Given this assumption, they interpret the correlation of objective health measures and health risk, and the correlation between health risk and insurance coverage, as pure indicators of adverse selection. Furthermore, they find risky behaviour to be negatively correlated with risk and insurance probability, suggesting that heterogeneous risk preferences are a source of advantageous selection.

Bolin et al. (2010) have also used self assessed health as an indicator of being at risk. They evaluate asymmetric information and the demand for voluntary health insurance in Europe using the Survey of Health, Ageing, and Retirement in Europe (SHARE) dataset. The dataset is comparable to the one we use in our study, since it is representative of the age 50+ population. Their results support those of previous studies, which suggest a negative correlation between health-risk and voluntary health insurance. The authors focus on the correlation between self-assessed health (as a measure for health risk) and the probability of taking out insurance. Furthermore, they distinguish between heterogeneous risk preferences and screening. They conclude that there seem to be no heterogeneous risk preferences behind the negative correlation between health status and the probability of taking out insurance. Interestingly, they also find a negative correlation between risk and insurance but find no evidence that this is an indicator for heterogeneous risk preferences.

There is little evidence on this issue regarding the market for supplementary private health insurance in the UK. Propper et al. (2001) analyse the dynamics in the demand for private medical insurance (PMI) between 1978 and 1996, using the Family Expenditure Survey. Accounting for consumer characteristics and health service quality measures, they find that the availability of private healthcare facilities and cohort effects, which might indicate changes in tastes/attitude, to be relevant determinants of the purchase of PMI.

Wallis (2003) also looks for the determinants of demand for PMI in the UK, based on British Household Panel Survey (BHPS) data. He evaluates the switching behaviour of individuals and focuses both on characteristics influencing the probability of purchasing insurance and the individual cost of PMI (ie the risk premium). He differentiates between the demand side characteristics of consumers and the supply side aspects that can influence insurance status, eg quality of service.

Another study, which uses the BHPS data, was done by Olivella and Vera-Hernandez (2006). They focus on adverse selection in the market for PMI, using hospitalisation as a measure for being at risk. Assuming that the health status of individuals is independent of receiving PMI as a fringe employment benefit, their results suggest the existence of adverse selection in the PMI market in England.

Until now, no empirical evidence on specific sources of information asymmetries in the UK private health insurance market has been found. One of the aims of this study is to analyse this further. In addition, it is of great interest to determine whether adverse selection is still an issue in the UK if another method of measuring information asymmetries is used.

2.3 Using Unused Variables

The starting point of our discussion is the positive correlation test of Chiappori and Salanié (1997), who suggest that if the hypothesis of a correlation between the 2 equations' error terms can be rejected, both equations are determined independently – ie insurance coverage and risk are uncorrelated. It is worth mentioning that this test allows us to compare each individual i in the sample with itself ($\mathrm{Cov}(\eta_i, \epsilon_i) = \mathbb{E}(\eta_i - \bar{\eta})(\epsilon_i - \bar{\epsilon})$) since the estimated error terms are given at the individual level.

Finkelstein and McGarry (2006) argue that the standard positive correlation test is going to fail if we find more than one individual characteristic that is not used for pricing purposes, but which affect insurance status and risk of loss to the insurance company. Both characteristics can lead to an offsetting of the correlation between the equations' error terms.

The advantage of such an 'unused characteristics' approach is that it enables one to identify concrete characteristics which can, from a theoretical perspective, be assumed to be a source of IA, even if the positive correlation test does not lead to such a revelation. The interpretation of the coefficients in these models is perfectly reasonable if we have an underlying theory as to why these characteristics should be correlated with both insurance and health status.

However, a disadvantage of this approach compared to the standard positive correlation test is that we measure correlations of included variables with both risk situation and insurance status, without relating both equations to one another from a statistical perspective (as the positive correlation test does).

Here, the conclusion about selection effects is derived based on the significance of the added variables in both equations, which is reasonable in some cases. This approach might, however, be problematic if we find an additional variable to be relevant in our framework and we, a priori, do not have a good theory about the mechanism which relates this variable to risk and insurance status. This is the case since a variable might, on average, control for relevant factors which explain, for example, insurance probability and risk situation on the one hand. But, on the other hand, it is possible that these estimated coefficients are driven by different parts of the sample. We suggest that if individuals with a common characteristic are heterogeneous in outcomes of this characteristic (ie marginal changes in risk and insurance probability due to an unused variable are negatively related) the 'unused characteristics' approach can lead to the wrong conclusions being made.

To describe this issue formally, we simplify equations 4 and 5 without loss of generality in a way that we exclude all 'observable' information from both equations, allowing just for one 'unused' variable z and rewrite them as follows:

$$\mathbb{E}(R|z) = \mathbb{E}(\beta z + \eta)$$

and

$$\mathbb{E}(I|z) = \mathbb{E}(\delta z + \epsilon)$$

where estimates $\hat{\beta} > 0$ and $\hat{\delta} > 0$ would usually be interpreted as an indicator of adverse selection. Now let's assume that every parameter is defined for every single individual i with the population means $\mathbb{E}(\beta_i) = \mu_\beta > 0$ and $\mathbb{E}(\delta_i) = \mu_\delta > 0$. Furthermore let's assume for simplification that $\text{Cov}(z_i, \beta_i) = 0$, $\text{Cov}(z_i, \delta_i) = 0$ and $\text{Cov}(\eta_i, \epsilon_i) = 0$.

Based on this model we now evaluate the risk position of the subpopulation being insured $(\delta_i z_i + \epsilon_i > 0)$:

$$\mathbb{E}(R_i|I_i > 0) = \mathbb{E}(\beta_i z_i|\delta_i z_i > -\epsilon_i)$$

Given the population means μ_β and $\mathbb{E}(z_i) = \mu_z$, this equation can be rewritten as

$$\begin{aligned}\mathbb{E}(\beta_i z_i|\delta_i z_i > -\epsilon_i) =& \mu_\beta \mu_z \\ &+ \mu_\beta \times \mathbb{E}(z_i - \mu_z|\delta_i z_i > -\epsilon_i) \\ &+ \mu_z \times \mathbb{E}(\beta_i - \mu_\beta|\delta_i z_i > -\epsilon_i) \\ &+ \mathbb{E}((\beta_i - \mu_\beta)(z_i - \mu_z)|\delta_i z_i > -\epsilon_i).\end{aligned} \tag{6}$$

We want to show under which circumstances $\mathbb{E}(R_i) = \mathbb{E}(R_i|I_i > 0)$, since in that case the 'unused characteristics' approach would have falsely detected adverse selection. The first and second term in decomposition (6) is positive due to our assumptions. Hence, $\mathbb{E}(\beta_i z_i|\delta_i z_i > -\epsilon_i) = \mathbb{E}(R)$ can be true if $\mathbb{E}(\beta_i - \mu_\beta|\delta_i z_i > -\epsilon_i) < 0$ or $\mathbb{E}((\beta_i - \mu_\beta)(z_i - \mu_z)|\delta_i z_i > -\epsilon_i) < 0$. Therefore, we would need $\text{Cov}(\beta_i, \delta_i) < 0$, and thus $\mathbb{E}(\beta_i|\delta_i z_i > -\epsilon_i) < \mu_\beta$ to allow for an offsetting of the other terms in the decomposition.

This means that in expectation, for an individual whose β_i is smaller than the population mean μ_β, the individual coefficient δ_i will be higher than the estimated population mean μ_δ, and vice versa.

Obviously our framework requires $\mathrm{Cov}(\beta_i, \delta_i) > 0$ in the case of $\mu_\beta < 0$ and $\mu_\delta > 0$, ie wrongly detected advantageous selection with the 'unused characteristics' approach. If the correlation of β_i and δ_i is strong enough, the suggested direction of selection may even change to opposite of what it should be.

To solve this problem, one could implement a comparison of insurance and risk status at the individual level, as done by Chiappori and Salanié (1997), by means of calculating the correlation of the error terms. As Finkelstein and McGarry (2006) argue, this approach is not helpful in all contexts, since some correlations might cancel out.

The method suggested by Dionne et al. (2001) is also not of help if we want to know if one specific variable generates IA. This is because we would only find this variable if it were significantly correlated with *both* insurance status and health risk. Hence we need to find another way to deal with this issue. A commonly used way to account for the possibility of heterogeneous parameters is to split the sample and calculate the probabilities of being privately insured, separately for groups with different health risk. Thus, the model we implement is based on the equation

$$\mathbb{E}(I|R = r) = \mathbb{E}(X\delta_1 + Z\delta_2 + \epsilon|R = r) \tag{7}$$

where r is an indicator variable for being in a specific risk class.

Hence, the first step is to measure the impact of potential sources of IA on the probability of risk and taking out insurance. The second step is to run separate estimations for people with contrasting health-risk states and then compare how an added variable, that is assumed not to be used for calculating risk premiums, affects the insurance probability in both groups. If the coefficient is similar for the two different groups, then subpopulations with heterogeneous coefficients should not lead to misleading results within the 'unused characteristics'framework.

As far as we know, there is scant literature on the detection of information asymmetries with a focus on differing subpopulations. Having said that, we should mention a study by Cohen (2005) who finds heterogeneous results for several subsets in the population of policyholders in the car insurance market. However, the background to Cohen's work is not the same since she uses different information about the customers' contract choices to inform the analysis whereas we focus on whether the detection of selection effects which are given by commonly used tests can be misleading as a result of having diverging subpopulations. This seems to be especially important if we, a priori, do not have an assumption about the correlation between a variable and health status or insurance status from a theoretical point of view.

3 Empirical Implementation

3.1 Data and Econometric Model

Our empirical illustration of this idea is based on ELSA, which is a representative individual level dataset for England's age 50+ population. No study has used this data for detecting information asymmetries before, and we know from previous research that tax subsidies which existed up until the end of the 1990s induced economic incentives for older people to buy private medical insurance in England (King and Mossialos, 2005). Hence, it is of interest to focus on this subpopulation to reveal whether the decision to buy supplementary health insurance was accompanied by a self-selection of good or bad risks into the insurance market via information asymmetries. The ELSA dataset contains a broad range of information on an individual's health and financial circumstances, together with the demographics, which makes it an ideal source to model both economic decisions and health related characteristics.

For our analysis we use the cross sectional dimension of Wave 2 as it gives us the richest source of available and relevant information. Since our model is based on a static model of self selection, we do not need to allow for the time dimension of the dataset.

As a main dependent variable, we use self assessed health[5] as a measure for being at risk. Although this information can be subjective, we assume it to be a reasonable indicator since it is not just capturing observable information (which we control for in our analysis) but also all information that can affect future demand for health care which cannot be accounted for from using observable and objective health data. Based on their findings, Idler and Benyamini (1997) argue that a global health rating "... represents an irreplaceable dimension of health status and in fact that an individual's health cannot be assessed without it." Hence, we conclude that self assessed health is a suitable measure for our purposes.

ELSA provides the commonly used 5-point scale SAH measure which covers health as a factor, varying between 'Excellent' and 'Very Bad'. We collapse this information into a binary variable which we call 'Good', which represents people answering 'Excellent' and 'Very Good'. All other responses are scored zero. This way we are able to achieve an average health status in our sample that is near the median. Given that the information used is subjective, we postulate that collapsing the health variable as described will provide much more robust findings about the population's health than if we had used the ordinal scaled information divided into 5 health categories. The second main dependent variable is a dummy variable that equals 1 if someone has private health insurance and 0 otherwise.

As previously mentioned, it is important for our research question to assume that the econometric model contains all the relevant information used by an insurance company in order to calculate the risk premium. Since the people in our sample have private health care insurance with different suppliers, we cannot provide a general framework for calculating the premiums. We therefore have to make assumptions which are as realistic as possible. We know that insurers price risk mainly on age, but other factors such as sex, smoking status and occupational status can be also taken into account. Treatment for specific illnesses (eg chronic conditions) can be excluded from a private health insurance policy. However, we know that insurers require the applicants to provide detailed information on their past and present health status (Boyle, 2011) and it is therefore reasonable to assume that this information is used by insurance companies for pricing purposes. Hence, in our model, we assume that the following variables are used by insurance companies: Age, sex, smoking, employment status, education (where 'no qualification' is the reference group), race, family status (indicator for being married or cohabiting), whether the person has children and whether the person has children living at home. Since we have detailed information on each individual's health, we can also include variables capturing hypertension, previously diagnosed angina, diabetes, stroke and cancer. Heart problems due to myocardial infarction, congestive heart failure, heart murmur and arrhythmia, are controlled for as well. In addition, we collapse hedibonic lung disease and asthma into one illness category, and arthritis and osteoporosis into another. Parkinson's disease, psychiatric disorder, Alzheimer's disease and dementia are also taken into account together. The final health category concerns eye problems such as glaucoma, diabetic eye disease, macular degeneration and cataract. All these health variables are self stated, so it might be possible that we have a bias due to measurement error.

We also require additional control variables that affect both health status and the probability of purchasing insurance but which we assume are not used by an insurance company for pricing purposes. We include the following as such variables: weekly income and financial wealth (equivalised on the basis of the number of people in the household who benefit from the income),

[5] Variable definitions and descriptive statistics in Table 1 and Table 2

and lifestyle variables which are usually used to control for heterogeneous health risk such as drinking behaviour, and being overweight.

Another characteristic which could capture differences in risk is the region where the respondent lives. Finkelstein and Poterba (2006) found this information relevant in the insurance market for annuities. In our context, regional information could be important too, since it is widely known that the health status of people strongly varies by locality. For example, mortality rates from cardiovascular disease are much higher in the northern parts than the southern parts of England (Müller-Nordhorn et al., 2008). If such health differences also correlate with insurance purchase probability, then the existence of selection effects might be indicated. To control for this, we group regions into a northern part (North, North West, Yorkshire, Humberside and Durham), a midland part (East Midlands, West Midlands and East Anglia) and the South West (with the South East used as the reference category). Since we know that health is on average better in urban than in rural areas (Dye, 2008) we also use a dummy variable to capture whether a respondent is living in a highly dense region (5th quintile).

As a benchmark we apply the classical correlation test of Chiappori and Salanié and estimate a bivariate probit model, including as explanatory variables all the information known to the insurer. We then calculate the correlation of residuals. The dependent variables in this model are insurance status and being in good health (ie low risk). Furthermore, we apply univariate probit specifications within the 'unused characteristics' framework to determine whether we can identify some specific variables which are not used for pricing, but which do affect both health risk and insurance probability. Such a finding is usually interpreted as indicating a selection mechanism due to IA, which would not necessarily be revealed with the methodology of Chiappori and Salanié. We use sample weights for the estimations and cluster the standard errors at household level. All estimation outputs show marginal effects.

To see whether conclusions based on the 'unused characteristics' approach are misleading due to having a heterogeneous population, we estimate equation (6) for the subpopulations for which our health indicator variable r equals 1 and 0 separately, and compare the results. This is a simplification of the framework provided in the last section, since our conditioning variable is not defined continuously any more. Furthermore, we test for the relevance of parameter heterogeneity via estimating a pooled, fully interacted specification, ie we calculate marginal effects for a specification where the insurance status is explained by all explanatory variables together with their interaction with good health status. In this framework, both high and low risk individuals are assumed to have the same constant, which is in line with the idea of the unused charactersitics approach. This procedure allows us to directly test the statistical significance of heterogeneous parameters between the two risk groups. Due to the nonlinear relationship in our setup the marginal effects of the interaction terms were calculated manually[6]. In addition, we check whether the findings of this approach are robust, by comparing it with the results of a linear probability model (LPM).

3.2 Results and Discussion

The application of the approach of Chiappori and Salanié, explaining insurance status and 'low risk', strongly suggests the existence of advantageous selection (ie that healthier people are more likely to take out private health insurance). Assuming no offsetting effects due to multiple information, a correlation coefficient of 0.11 (p-value: 0.00) suggests that there is a relevant mechanism that relates the insurance decision and an individual's health positively, conditioned on the observables used by the insurer. Findings, based on a specification where the

[6] For an insightful discussion see Ai and Norton (2003).

health measure is representing one SAH-category more, yield quite similar results (correlation coefficient 0.10; p-value: 0.01). A specification which excludes all personal health information from the model reveals a correlation coefficient of 0.12 with p-value: 0.00. Knowing that this finding is neither a necessary nor sufficient condition for IA, we now focus on the results using the 'unused characteristics' approach.

The marginal effects of a probit regression based on equation (4)[7] show that income and wealth, conditioned on all other information used in our model, both have a significant positive impact on being privately insured. Our regional variables suggest that living in the regions away from the South East of England, is negatively related to the probability of being privately insured. This effect is strongest for the North of England. In other words, it seems to be the case that the further an individual lives from the South East of England, the less likely they are to have purchased private health insurance.

The results for the correlation of our unused characteristics with being at low risk, based on Equation (5), show that income and wealth are positively related to low risk, which seems to be intuitively reasonable (ie the higher the income, the more likely the individual is to be healthy). Drinking is positively related to the dependent variable, which is a little puzzling since one might expect that heavier drinking is more likely to be associated with poor health. One possible explanation is that it is capturing factors such as social relationships. Being overweight is negatively correlated with low risk. All regional variables are negatively correlated with being at low risk, although the coefficients are not statistically significant different from 0. Nevertheless the sign of the coefficient is what we would expect since we know that the population in the North of England tends, on average, to be less healthy than people living in England's southern regions.

No other variables are of interest in our analysis since we assume that they are used in the calculation of the individual's risk premium. Hence, they are unable to give us further information on the existence of information asymmetries in the market for health insurance. Comparing the coefficients of both equations, we now derive conclusions about possible selection effects due to information asymmetries.

The income variable suggests that more healthy individuals buy supplementary private health insurance, so it should be interpreted as a source of advantageous selection in our framework. Also financial wealth seems to drive the selection of healthier individuals into the private insurance market. The impact of wealth is stronger than the income effect, which makes intuitive sense, since we are analysing the behaviour of a population from which a minority of 33 percent is still working. Overall, there seem to be clear indications of the existence of advantageous selection, which is a widely found phenomenon in the literature and, in our case, also in line with empirical evidence based on 'Chiappori and Salanié'.

In the next step we estimate a regression of being privately insured for the people in the 'good' health category (indicating low risk) and for the reference group (bad risk) separately[8]. For some of the 'unused characteristics' in our model this procedure reveals very interesting findings. Firstly the variable controlling for a high population density reveals opposite signs: a positive relationship with insurance status for people of the good risk group, and a negative one for the group with bad risks. Both coefficients are not statistically different in this split sample approach. Secondly, the variable representing the highest wealth quantile is less important for individuals within the low risk group, but stays relevant and statistically significant for both groups. This finding suggests that, although wealth is the driving force of advantageous selection overall, the probability of taking out insurance is even higher for the bad risks, suggesting that

[7] In Table 3
[8] In Table 4

wealth is a source of adverse selection as well.

This suggests that a common interpretation of this variable in the 'unused characteristics' approach leads to a wrong conclusion being made. The variable should not just simply be interpreted as an indication of advantageous selection any more, since the positive correlation with insurance probability is mainly driven by individuals with poorer health.

To support this hypothesis we estimate a regression model including interaction terms of all our explanatory variables with our low risk variable 'good' included in the specification. The marginal effects shown in Table 5 provide evidence for the statistical significance of parameter heterogeneity, ie we are testing whether the difference of the estimated coefficients is statistically different from 0. Interestingly, in this setup, population density is an important driver of the insurance decision of the good risks. One explanation for this finding is that the coefficients can be interpreted with regard to the overall population used in this study. This suggests that, people living in highly dense areas who are in good health, have an above-average probabilty of taking out insurance. Therefore, population density can also be interpreted as a source of advantageous selection.

As expected, the interaction term of the 4th wealth quantile was found to be negativly related to the insurance decision, suggesting some kind of offsetting of the overall advantageous selection effect by adverse selection. In this setup the interaction is not statistically significant at standard levels.

In comparison with the nonlinear model, the LPM suggests[9] both interaction terms to be statistically significant. Despite theoretical concerns over the interpretation of the coefficients in a LPM (eg Wooldridge (2003)), we interpret this finding to support our hypothesis about the relevance of parameter heterogeneity.

Hence we conclude that, although a variable can be correlated with insurance and health status separately, the relationship does not necessarily tell us anything about the importance of information aysmmetries. This is because the estimated coefficients of such a variable (indicating IA) can be caused mainly by heterogeneous parts of the population. Also, if no correlation is found, there might still a selection mechanism driven by a variable included in the specification, which is incorrectly not revealed when the unused-characteristics approach is applied.

This suggests that, as 2 different sources of private information can offset the correlation of the error in the approach of Chiappori and Salanié (Finkelstein and McGarry, 2006), the inclusion of private information within an 'unused characteristics' framework can lead to incorrect conclusions if the estimated coefficients in both equations are driven by different parts of the population.

Using different sources of unused information, overall we found advantageous selection to be the most relevant selection mechanism in the supplementary private health insurance market in England. This finding conflicts with evidence on the English market for private medical insurance by Olivella and Vera-Hernandez (2006) who claim the existence of adverse selection. The discrepancy might be explained by the fact that we use a different methodology and focus on the age 50+ population as opposed to the whole population. If the latter point is important it would be of interest to reveal why the selection mechanism differs so strikingly between the age groups. Our findings do not purport to make any claims about welfare, ie we do not know whether the advantageous selection is an indicator of market inefficiency due to overinsurance, or whether marginal costs equal willingness to pay in equilibrium (Einav et al., 2010).

[9]Results provided in Table 6

4 Conclusion

After introducing the institutional background of the UK health system, we provided an overview about commonly used testing procedures for the detection of information asymmetries, focusing on its strengths and weaknesses. We also summarized some empirical findings for the UK private health insurance market.

We argued that, although a classical positive correlation test might lead to incorrect conclusions, it still has the advantage that it compares each individual in the sample with itself, through the correlation of the residuals. To account for that advantage within the often adopted 'unused characteristics' approach (eg Finkelstein and McGarry (2006)) we controlled for the possibility of parameter heterogeneity to avoid possible misinterpretation. We provided separate estimates for different groups and fully interacted specifications to show that parameter heterogeneity is a relevant issue if we want to measure the probability of taking out insurance for the total population. This can be problematic within a framework using 2 equations if evidence is wrongly based on single coefficients to derive conclusions about the existence of information asymmetries in insurance markets.

Our findings suggest that, due to the existence of heterogeneous subpopulations, some kinds of selection mechanisms cannot be revealed by using just two equations and calculating marginal effects. The idea that the estimated mean from such an approach can be driven by a different part of the population suggests even that the opposite interpretation is possible (ie claiming that a detected source of adverse selection is in fact advantageous selection, or vice versa). Nevertheless, in our empirical implementation of this issue, the findings with the 'unused characteristics' concept are robust overall – even if we allow for heterogeneous coefficients. However, the relevance of parameter heterogeneity is an empirical question and a general conclusion for other markets cannot be provided. Hence, we suggest accounting for parameter heterogeneity when using this approach to detect specific sources of information asymmetries.

Another interesting finding of this study is that, overall, there is a selection of good risks into the market for private health insurance in England which is mainly driven by variation of income and wealth.

Our findings are important for analysing the efficiency of insurance markets. It is of interest to both the insurance industry and policy makers, and should be allowed for whenever structural changes in contract design or institutions are implemented. Our findings are particularly relevant in the case of unused variables for which, a priori, we cannot assume any specific relationship with insurance status and health status. In this case, it is necessary to be very careful about potential parameter heterogeneity.

Since we do not have any information about costs arising due to information asymmetries, and do not assume any specific utility function, we cannot make any claim about welfare in the market for private health insurance. Nevertheless, our findings should be taken into account in future research since they directly affect the interpretation of which kind of selection is measured in the observed market. Assuming that our subjective health-risk variable is a good indicator for the individual health status, we still do not know if it is also a good measure for health care utilization. Hence, we make the implicit assumption in our analysis that people with a relatively low (or relatively high) self assessed health status are correlated with a higher (lower) probability of making a claim in respect of health insurance. Future research should evaluate if the relevance of our findings can be supported when using objective data about health care utilization, eg number of visits to the doctor or, even better, treatment costs.

An additional feature of our study is that there is very little regulation in relation to the market for private health insurance in the UK. Given the competitive nature of the UK market, the findings could be interpreted as a benchmark compared to analysis which is based on populations with different institutional backgrounds. For future research it would be of interest to analyze deviations with regard to market regulation. In particular, it would be interesting to compare our findings with outcomes in more regulated markets, eg the US.

References

Ai, C., Norton, E., 2003. Interaction terms in logit and probit models. Economics Letters 80 (1), 123–129.

Akerlof, G., 1970. The market for" lemons": Quality uncertainty and the market mechanism. The quarterly journal of economics, 488–500.

Bolin, K., Hedblom, D., Lindgren, A., Lindgren, B., 2010. Asymmetric information and the demand for voluntary health insurance in europe. Tech. rep., National Bureau of Economic Research.

Boyle, S., 2011. United kingdom (england) health system review. Health systems in transition 13 (1), 1.

Chiappori, P., Salanié, B., 1997. Empirical contract theory: The case of insurance data. European Economic Review 41 (3-5), 943–950.

Cohen, A., 2005. Asymmetric information and learning: evidence from the automobile insurance market. Review of Economics and Statistics 87 (2), 197–207.

Cohen, A., Siegelman, P., 2010. Testing for adverse selection in insurance markets. Journal of Risk and Insurance 77 (1), 39–84.

Cutler, D., Finkelstein, A., McGarry, K., 2008. Preference heterogeneity and insurance markets: Explaining a puzzle of insurance. Tech. rep., National Bureau of Economic Research.

Dardanoni, V., Li Donni, P., 2008. Testing for asymmetric information in insurance markets with unobservable types. Health, Econometrics and Data Group (HEDG) Working Papers 8, 26.

Dionne, G., Gouriéroux, C., Vanasse, C., 2001. Testing for evidence of adverse selection in the automobile insurance market: A comment. Journal of Political Economy 109 (2), 444–453.

Doiron, D., Jones, G., Savage, E., 2008. Healthy, wealthy and insured? the role of self-assessed health in the demand for private health insurance. Health economics 17 (3), 317–334.

Dye, C., 2008. Health and urban living. Science 319 (5864), 766.

Einav, L., Finkelstein, A., Cullen, M. R., 2010. Estimating welfare in insurance markets using variation in prices. The Quarterly Journal of Economics 125 (3), 877–921.

Finkelstein, A., McGarry, K., 2006. Multiple dimensions of private information: evidence from the long-term care insurance market. American Economic Review 96 (4), 938–958.

Finkelstein, A., Poterba, J., 2006. Testing for adverse selection with" unused observables". Tech. rep., National Bureau of Economic Research.

Idler, E., Benyamini, Y., 1997. Self-rated health and mortality: a review of twenty-seven community studies. Journal of health and social behavior, 21–37.

King, D., Mossialos, E., 2005. The determinants of private medical insurance prevalence in england, 1997–2000. Health services research 40 (1), 195–212.

Müller-Nordhorn, J., Binting, S., Roll, S., Willich, S., 2008. An update on regional variation in cardiovascular mortality within europe. European heart journal 29 (10), 1316.

Olivella, P., Vera-Hernandez, M., 2006. Testing for adverse selection into privatemedical insurance.

Propper, C., Rees, H., Green, K., 2001. The demand for private medical insurance in the uk: a cohort analysis. The Economic Journal 111 (471), 180–200.

Rothschild, M., Stiglitz, J., 1976. Equilibrium in competitive insurance markets: An essay on the economics of imperfect information. The Quarterly Journal of Economics, 629–649.

Wallis, G., 2003. The determinants of demand for private medical insurance: evidence from the british household panel survey.

Wooldridge, J., 2003. Introductory econometrics. A Modern Approach. 2nd edition. Thomson South-Western.

Table 1: Data description

Variable	Description
privins	owner of private health insurance
good	excellent and very good health
obese	body mass index >35
alcohol	daily drinker
north	North, North West, Yorkshire, Humberside and Durham
middle	East Midlands, West Midlands and East Anglia
south west	South West
highdens	5th density quintile
inc2	second income quantile
inc3	third income quantile
inc4	fourth income quantile
wealth2	second wealth quantile
wealth3	third wealth quantile
wealth4	fourth wealth quantile
smoke	smoking
women	female
age6070	age 60 to 69
age7080	age 70 to 79
age8090	age 80 to 89
age90100	age 90+
work	respondent is working
educ	more than no qualification
famstat	married or cohabit
nwhite	ethnicity recoded to non white
children	respondent has children
childrenHH	has children in household
eye	diagnosed glaucoma, diabetic eye disease, macular degeneration or cataract
angina	respondent had angina
psychic	Parkinson's disease, psychiatric disorder, Alzheimer's disease or dementia
cancer	cancer
bones	arthritis or osteoporosis
breath	hedibonic lung disease or asthma
stroke	stroke
diab	diabetes
heart	myocardial infarction, congestive heart failure, heart murmur or arrhythmia
bloodp	hypertension

Table 2: Summary statistics

Variable	Mean	Std. Dev.	Min.	Max.
privins	0.141	0.348	0	1
good	0.436	0.496	0	1
obese	0.076	0.264	0	1
alcohol	0.181	0.385	0	1
north	0.299	0.458	0	1
middle	0.249	0.432	0	1
southwest	0.122	0.328	0	1
highdens	0.143	0.35	0	1
inc2	0.246	0.43	0	1
inc3	0.261	0.439	0	1
inc4	0.26	0.439	0	1
wealth2	0.247	0.431	0	1
wealth3	0.27	0.444	0	1
wealth4	0.281	0.449	0	1
smoke	0.136	0.343	0	1
women	0.554	0.497	0	1
age90100	0.006	0.075	0	1
age8090	0.086	0.28	0	1
age7080	0.241	0.428	0	1
age6070	0.355	0.479	0	1
work	0.329	0.47	0	1
educ	0.663	0.473	0	1
famstat	0.71	0.454	0	1
nwhite	0.012	0.109	0	1
children	0.861	0.346	0	1
childrenHH	0.172	0.378	0	1
eye	0.218	0.413	0	1
angina	0.099	0.298	0	1
psychic	0.105	0.307	0	1
cancer	0.08	0.271	0	1
bones	0.398	0.489	0	1
breath	0.176	0.381	0	1
stroke	0.039	0.195	0	1
diab	0.078	0.268	0	1
heart	0.172	0.377	0	1
bloodp	0.427	0.495	0	1
N	5851			

Table 3: Unused Characteristics

	(1) privins	(2) good
obese	0.009	-0.101***
	(0.616)	(0.000)
alcohol	0.010	0.049**
	(0.396)	(0.012)
north	-0.052***	-0.022
	(0.000)	(0.261)
middle	-0.038***	-0.027
	(0.001)	(0.187)
southwest	-0.037***	-0.024
	(0.005)	(0.309)
highdens	0.007	0.001
	(0.632)	(0.975)
inc2	0.016	0.032
	(0.358)	(0.153)
inc3	0.039**	0.029
	(0.030)	(0.201)
inc4	0.113***	0.071***
	(0.000)	(0.004)
wealth2	0.056***	0.052**
	(0.005)	(0.028)
wealth3	0.087***	0.076***
	(0.000)	(0.002)
wealth4	0.146***	0.170***
	(0.000)	(0.000)
smoke	-0.039***	-0.114***
	(0.001)	(0.000)
women	0.015**	0.062***
	(0.025)	(0.000)
age90100	0.043	-0.061
	(0.553)	(0.500)
age8090	-0.034*	0.034
	(0.055)	(0.342)
age7080	-0.031**	-0.010
	(0.024)	(0.679)
age6070	-0.018	0.008
	(0.115)	(0.695)
work	-0.004	0.080***
	(0.711)	(0.000)
educ	0.042***	0.035**
	(0.000)	(0.035)
famstat	0.002	-0.068***
	(0.874)	(0.000)
nwhite	0.068	-0.142**
	(0.190)	(0.023)
children	-0.009	0.067***
	(0.512)	(0.002)
childrenHH	-0.018	-0.051**
	(0.167)	(0.012)
eye	0.001	-0.061***
	(0.918)	(0.002)
angina	0.019	-0.178***
	(0.279)	(0.000)
psychic	-0.022*	-0.148***
	(0.084)	(0.000)
cancer	-0.002	-0.119***
	(0.885)	(0.000)
bones	0.018*	-0.166***
	(0.059)	(0.000)
breath	-0.004	-0.158***
	(0.706)	(0.000)
stroke	-0.009	-0.154***
	(0.731)	(0.000)
diab	-0.023	-0.209***
	(0.140)	(0.000)
heart	-0.012	-0.117***
	(0.285)	(0.000)
bloodp	0.002	-0.121***
	(0.860)	(0.000)
N	5851	5851

Finite Differences Marginal Effects,
Clustered standard errors at household,
p-values in parathesis
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4: Split Sample

	(1) privins—good=1	(2) privins—good=0
obese	0.013	0.009
	(0.714)	(0.616)
alcohol	-0.001	0.020
	(0.974)	(0.166)
north	-0.062***	-0.043***
	(0.001)	(0.001)
middle	-0.039**	-0.038***
	(0.042)	(0.001)
southwest	-0.050**	-0.029**
	(0.022)	(0.041)
highdens	0.039	-0.013
	(0.148)	(0.400)
inc2	0.036	0.005
	(0.289)	(0.759)
inc3	0.068**	0.023
	(0.042)	(0.205)
inc4	0.167***	0.069***
	(0.000)	(0.003)
wealth2	0.033	0.066***
	(0.360)	(0.002)
wealth3	0.092**	0.078***
	(0.014)	(0.001)
wealth4	0.115***	0.174***
	(0.002)	(0.000)
smoke	-0.051**	-0.027**
	(0.020)	(0.028)
women	0.017	0.010
	(0.164)	(0.267)
age90100	0.011	0.076
	(0.920)	(0.413)
age8090	-0.047	-0.027
	(0.160)	(0.145)
age7080	-0.038	-0.022
	(0.116)	(0.139)
age6070	-0.027	-0.014
	(0.162)	(0.293)
work	-0.021	0.009
	(0.261)	(0.534)
educ	0.041**	0.039***
	(0.020)	(0.000)
famstat	0.032*	-0.017
	(0.093)	(0.233)
nwhite	0.044	0.072
	(0.611)	(0.224)
children	-0.015	-0.006
	(0.541)	(0.708)
childrenHH	-0.056***	0.009
	(0.003)	(0.559)
eye	0.024	-0.007
	(0.356)	(0.575)
angina	-0.047	0.034*
	(0.155)	(0.060)
psychic	-0.049**	-0.004
	(0.030)	(0.800)
cancer	0.026	-0.008
	(0.391)	(0.589)
bones	0.023	0.019*
	(0.189)	(0.061)
breath	-0.016	-0.001
	(0.432)	(0.940)
stroke	-0.021	-0.006
	(0.731)	(0.811)
diab	-0.030	-0.015
	(0.436)	(0.313)
heart	-0.031	-0.001
	(0.143)	(0.921)
bloodp	0.009	-0.000
	(0.564)	(0.959)
N	2551	3300

Finite Differences Marginal Effects,
Clustered standard errors at household,
p-values in parathesis
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 5: Interacted Probit Model

	(1) privins
alcohol	.0126
	(0.292)
obese	.0107
	(0.552)
north	-.0507***
	(0.001)
middle	-.0394***
	(0.004)
southwest	-.0370
	(0.013)**
highdens	.0036
	(0.802)
inc2	.0159
	(0.343)
inc3	.0396
	(0.034)**
inc4	.1066***
	(0.000)
wealth2	.0554***
	(0.006)
wealth3	.0841***
	(0.000)
wealth4	.1478***
	(0.000)
alcohol_good	-.0224
	(0.327)
obese_good	.0010
	(0.976)
north_good	-.0002
	(0.990)
middle_good	.0137
	(0.455)
southwest_good	-.0058
	(0.794)
highdens_good	.0506*
	(0.088)
inc2_good	.0281
	(0.410)
inc3_good	.0349
	(0.339)
inc4_good	.0798*
	(0.087)
wealth2_good	-.0409
	(0.301)
wealth3_good	-.0004
	(0.991)
wealth4_good	-.0705
	(0.156)
N	5851

Finite Differences Marginal Effects,
Clustered standard errors at household,
p-values in parathesis
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 6: Interacted LPM

	(1) privins
alcohol	0.025
	(0.143)
obese	0.004
	(0.826)
north	-0.050***
	(0.002)
middle	-0.047***
	(0.005)
southwest	-0.040*
	(0.066)
highdens	-0.013
	(0.419)
inc2	0.001
	(0.935)
inc3	0.018
	(0.288)
inc4	0.081***
	(0.000)
wealth2	0.039***
	(0.003)
wealth3	0.047***
	(0.002)
wealth4	0.148***
	(0.000)
alcohol_good	-0.029
	(0.281)
obese_good	0.009
	(0.824)
north_good	-0.015
	(0.551)
middle_good	0.004
	(0.874)
southwest_good	-0.021
	(0.540)
highdens_good	0.048*
	(0.099)
inc2_good	0.021
	(0.403)
inc3_good	0.028
	(0.318)
inc4_good	0.075**
	(0.026)
wealth2_good	-0.025
	(0.296)
wealth3_good	0.008
	(0.754)
wealth4_good	-0.056*
	(0.081)
N	5851

Clustered standard errors at household, p-values in parathesis
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Cass Business School
CITY UNIVERSITY LONDON

FACULTY OF ACTUARIAL SCIENCE AND INSURANCE

Actuarial Research Papers since 2001

Report Number	Date	Publication Title	Author
160.	January 2005	Mortality Reduction Factors Incorporating Cohort Effects. ISBN 1 90161584 7	Arthur E. Renshaw Steven Haberman
161.	February 2005	The Management of De-Cumulation Risks in a Defined Contribution Environment. ISBN 1 901615 85 5.	Russell J. Gerrard Steven Haberman Elena Vigna
162.	May 2005	The IASB Insurance Project for Life Insurance Contracts: Impart on Reserving Methods and Solvency Requirements. ISBN 1-901615 86 3.	Laura Ballotta Giorgia Esposito Steven Haberman
163.	September 2005	Asymptotic and Numerical Analysis of the Optimal Investment Strategy for an Insurer. ISBN 1-901615-88-X	Paul Emms Steven Haberman
164.	October 2005.	Modelling the Joint Distribution of Competing Risks Survival Times using Copula Functions. ISBN 1-901615-89-8	Vladimir Kaishev Dimitrina S, Dimitrova Steven Haberman
165.	November 2005.	Excess of Loss Reinsurance Under Joint Survival Optimality. ISBN1-901615-90-1	Vladimir K. Kaishev Dimitrina S. Dimitrova
166.	November 2005.	Lee-Carter Goes Risk-Neutral. An Application to the Italian Annuity Market. ISBN 1-901615-91-X	Enrico Biffis Michel Denuit
167.	November 2005	Lee-Carter Mortality Forecasting: Application to the Italian Population. ISBN 1-901615-93-6	Steven Haberman Maria Russolillo
168.	February 2006	The Probationary Period as a Screening Device: Competitive Markets. ISBN 1-901615-95-2	Jaap Spreeuw Martin Karlsson
169.	February 2006	Types of Dependence and Time-dependent Association between Two Lifetimes in Single Parameter Copula Models. ISBN 1-901615-96-0	Jaap Spreeuw
170.	April 2006	Modelling Stochastic Bivariate Mortality ISBN 1-901615-97-9	Elisa Luciano Jaap Spreeuw Elena Vigna.
171.	February 2006	Optimal Strategies for Pricing General Insurance. ISBN 1901615-98-7	Paul Emms Steve Haberman Irene Savoulli
172.	February 2006	Dynamic Pricing of General Insurance in a Competitive Market. ISBN1-901615-99-5	Paul Emms
173.	February 2006	Pricing General Insurance with Constraints. ISBN 1-905752-00-8	Paul Emms
174.	May 2006	Investigating the Market Potential for Customised Long Term Care Insurance Products. ISBN 1-905752-01-6	Martin Karlsson Les Mayhew Ben Rickayzen

Report Number	Date	Publication Title	Author
175.	December 2006	Pricing and Capital Requirements for With Profit Contracts: Modelling Considerations. ISBN 1-905752-04-0	Laura Ballotta
176.	December 2006	Modelling the Fair Value of Annuities Contracts: The Impact of Interest Rate Risk and Mortality Risk. ISBN 1-905752-05-9	Laura Ballotta Giorgia Esposito Steven Haberman
177.	December 2006	Using Queuing Theory to Analyse Completion Times in Accident and Emergency Departments in the Light of the Government 4-hour Target. ISBN 978-1-905752-06-5	Les Mayhew David Smith
178.	April 2007	In Sickness and in Health? Dynamics of Health and Cohabitation in the United Kingdom. ISBN 978-1-905752-07-2	Martin Karlsson Les Mayhew Ben Rickayzen
179.	April 2007	GeD Spline Estimation of Multivariate Archimedean Copulas. ISBN 978-1-905752-08-9	Dimitrina Dimitrova Vladimir Kaishev Spiridon Penev
180.	May 2007	An Analysis of Disability-linked Annuities. ISBN 978-1-905752-09-6	Ben Rickayzen
181.	May 2007	On Simulation-based Approaches to Risk Measurement in Mortality with Specific Reference to Poisson lee-Carter Modelling. ISBN 978-1-905752-10-2	Arthur Renshaw Steven Haberman
182.	July 2007	High Dimensional Modelling and Simulation with Asymmetric Normal Mixtures. ISBN 978-1-905752-11-9	Andreas Tsanakas Andrew Smith
183.	August 2007	Intertemporal Dynamic Asset Allocation for Defined Contribution Pension Schemes. ISBN 978-1-905752-12-6	David Blake Douglas Wright Yumeng Zhang
184.	October 2007	To split or not to split: Capital allocation with convex risk measures. ISBN 978-1-905752-13-3	Andreas Tsanakas
185.	April 2008	On Some Mixture Distribution and Their Extreme Behaviour. ISBN 978-1-905752-14-0	Vladimir Kaishev Jae Hoon Jho
186.	October 2008	Optimal Funding and Investment Strategies in Defined Contribution Pension Plans under Epstein-Zin Utility. ISBN 978-1-905752-15-7	David Blake Douglas Wright Yumeng Zhang
187.	May 2008	Mortality Risk and the Valuation of Annuities with Guaranteed Minimum Death Benefit Options: Application to the Italian Population. ISBN 978-1-905752-16-4	Steven Haberman Gabriella Piscopo
188.	January 2009	The Market Potential for Privately Financed Long Term Care Products in the UK. ISBN 978-1-905752-19-5	Leslie Mayhew
189.	June 2009	Whither Human Survival and Longevity or the Shape of things to Come. ISBN 978-1-905752-21-8	Leslie Mayhew David Smith
190	October 2009	ilc: A Collection of R Functions for Fitting a Class of Lee Carter Mortality Models using Iterative fitting Algorithms* ISBN 978-1-905752-22-5	Zoltan Butt Steven Haberman
191.	October 2009	Decomposition of Disease and Disability Life Expectancies in England, 1992-2004. ISBN 978-1-905752-23-2	Domenica Rasulo Leslie Mayhew Ben Rickayzen

Report Number	Date	Publication Title	Author
192.	October 2009	Exploration of a Novel Bootstrap Technique for Estimating the Distribution of Outstanding Claims Reserves in General Insurance. ISBN 978-1-905752-24-9	Robert Cowell
193.	January 2010	Surplus Analysis for Variable Annuities with a GMDB Option. ISBN 978-1-905752-25-6	Steven Haberman Gabriella Piscopo
194.	January 2010	UK State Pension Reform in a Public Choice Framework. ISBN 978-1-905752-26-3	Philip Booth
195.	June 2010	Stochastic processes induced by Dirichlet (B-) splines: modelling multivariate asset price dynamics. ISBN 978-1-905752-28-7	Vladimir Kaishev
196.	August 2010	A comparative Study of Parametric Mortality Projection Models. ISBN 978-1-905752-29-4	Zoltan Butt Steven Haberman
197.	April 2012	Are the Dimensions of Private Information More Multiple than Expected? Information Asymmetries in the Market of Supplementary Private Health Insurance in England* ISBN 978-1-905752-31-7	Martin Karlsson Florian Klohn Ben Rickayzen

Statistical Research Papers

1.	December 1995.	Some Results on the Derivatives of Matrix Functions. ISBN 1 874 770 83 2	P. Sebastiani
2.	March 1996	Coherent Criteria for Optimal Experimental Design. ISBN 1 874 770 86 7	A.P. Dawid P. Sebastiani
3.	March 1996	Maximum Entropy Sampling and Optimal Bayesian Experimental Design. ISBN 1 874 770 87 5	P. Sebastiani H.P. Wynn
4.	May 1996	A Note on D-optimal Designs for a Logistic Regression Model. ISBN 1 874 770 92 1	P. Sebastiani R. Settimi
5.	August 1996	First-order Optimal Designs for Non Linear Models. ISBN 1 874 770 95 6	P. Sebastiani R. Settimi
6.	September 1996	A Business Process Approach to Maintenance: Measurement, Decision and Control. ISBN 1 874 770 96 4	Martin J. Newby
7.	September 1996.	Moments and Generating Functions for the Absorption Distribution and its Negative Binomial Analogue. ISBN 1 874 770 97 2	Martin J. Newby
8.	November 1996.	Mixture Reduction via Predictive Scores. ISBN 1 874 770 98 0	Robert G. Cowell.
9.	March 1997.	Robust Parameter Learning in Bayesian Networks with Missing Data. ISBN 1 901615 00 6	P.Sebastiani M. Ramoni
10.	March 1997.	Guidelines for Corrective Replacement Based on Low Stochastic Structure Assumptions. ISBN 1 901615 01 4.	M.J. Newby F.P.A. Coolen
11.	March 1997	Approximations for the Absorption Distribution and its Negative Binomial Analogue. ISBN 1 901615 02 2	Martin J. Newby
12.	June 1997	The Use of Exogenous Knowledge to Learn Bayesian Networks from Incomplete Databases. ISBN 1 901615 10 3	M. Ramoni P. Sebastiani

Report Number	Date	Publication Title	Author
13.	June 1997	Learning Bayesian Networks from Incomplete Databases. ISBN 1 901615 11 1	M. Ramoni P.Sebastiani
14.	June 1997	Risk Based Optimal Designs. ISBN 1 901615 13 8	P.Sebastiani H.P. Wynn
15.	June 1997.	Sampling without Replacement in Junction Trees. ISBN 1 901615 14 6	Robert G. Cowell
16.	July 1997	Optimal Overhaul Intervals with Imperfect Inspection and Repair. ISBN 1 901615 15 4	Richard A. Dagg Martin J. Newby
17.	October 1997	Bayesian Experimental Design and Shannon Information. ISBN 1 901615 17 0	P. Sebastiani. H.P. Wynn
18.	November 1997.	A Characterisation of Phase Type Distributions. ISBN 1 901615 18 9	Linda C. Wolstenholme
19.	December 1997	A Comparison of Models for Probability of Detection (POD) Curves. ISBN 1 901615 21 9	Wolstenholme L.C
20.	February 1999.	Parameter Learning from Incomplete Data Using Maximum Entropy I: Principles. ISBN 1 901615 37 5	Robert G. Cowell
21.	November 1999	Parameter Learning from Incomplete Data Using Maximum Entropy II: Application to Bayesian Networks. ISBN 1 901615 40 5	Robert G. Cowell
22.	March 2001	FINEX : Forensic Identification by Network Expert Systems. ISBN 1 901615 60X	Robert G.Cowell
23.	March 2001.	Wren Learning Bayesian Networks from Data, using Conditional Independence Tests is Equivalant to a Scoring Metric ISBN 1 901615 61 8	Robert G Cowell
24.	August 2004	Automatic, Computer Aided Geometric Design of Free-Knot, Regression Splines. ISBN 1-901615-81-2	Vladimir K Kaishev, Dimitrina S.Dimitrova, Steven Haberman Richard J. Verrall
25.	December 2004	Identification and Separation of DNA Mixtures Using Peak Area Information. ISBN 1-901615-82-0	R.G.Cowell S.L.Lauritzen J Mortera,
26.	November 2005.	The Quest for a Donor : Probability Based Methods Offer Help. ISBN 1-90161592-8	P.F.Mostad T. Egeland., R.G. Cowell V. Bosnes Ø. Braaten
27.	February 2006	Identification and Separation of DNA Mixtures Using Peak Area Information. (Updated Version of Research Report Number 25). ISBN 1-901615-94-4	R.G.Cowell S.L.Lauritzen J Mortera,
28.	October 2006	Geometrically Designed, Variable Knot Regression Splines : Asymptotics and Inference. ISBN 1-905752-02-4	Vladimir K Kaishev Dimitrina S.Dimitrova Steven Haberman Richard J. Verrall

Report Number	Date	Publication Title	Author
29.	October 2006	Geometrically Designed, Variable Knot Regression Splines : Variation Diminishing Optimality of Knots. ISBN 1-905752-03-2	Vladimir K Kaishev Dimitrina S.Dimitrova Steven Haberman Richard J. Verrall
30.	November 2008	Scheduling Reentrant Jobs on Parallel Machines with a Remote Server. ISBN 978-1-905752-18-8	Konstantin Chakhlevitch Celia Glass
31.	November 2009	Probabilistic Expert Systems for Handling Artifacts in Complex DNA Mixtures. ISBN 978-1-905752-27-0	R.G.Cowell S.L. lauritzen J. Mortera
32.	April 2012	Detecting a Family Trio in a DNA Mixture. ISBN 978-1-905752-30-0	Robert G. Cowell

Papers can be downloaded from

http://www.cass.city.ac.uk/facact/research/publications.html

Faculty of Actuarial Science and Insurance

Cass Business School

ISBN 978-1-905752-37-1