City Research Online

The Role of Heterogeneous Parameters for the Detection of Selection in Insurance Contracts

Martin Karlsson[†] Florian Klohn[‡] Ben Rickayzen[§]

September 6, 2018

[†]CINCH, University of Duisburg-Essen, and Lund University. E-mail: martin.karlsson@uni-due.de

[‡]Corresponding author: GWQ ServicePlus AG, Tersteegenstr. 28, 40474 Düsseldorf, phone: +49211 758498-660, fax: +49211 758498-48, e-mail: florian.klohn@uni-due.de

[§]City, University of London, Cass Business School, 106 Bunhill Row, EC1Y 8TZ London, e-mail: b.d.rickayzen@city.ac.uk

[*]We would like to thank Martin Fischer, Astrid Kiil, Stefan Pichler, Robert Plumb, David Smith, Christoph Strupat, anonymous reviewers and participants of different events for helpful comments: the 2011 DIBOGS Workshop, the European Conference on Health Economics (2012), the Health Economics Seminar at the University Duisburg-Essen. All views expressed in this article should be considered those of the authors alone, and do not necessarily represent the position of the GWQ ServicePlus AG.

Abstract

This study re-examines standard econometric approaches for detecting adverse and advantageous selection in insurance contracts based on variables that are not used for calculating the insurance premium. We formally demonstrate that existing strategies for detecting selection based on such 'unused characteristics' can lead to incorrect conclusions if the estimated coefficients of interest are driven by different parts of the population. We show that this issue can empirically be accounted for by allowing for heterogeneous parameters. We compare existing approaches by using simulated data with different selection regimes and test for parameter heterogeneity within the data. We further provide empirical evidence about selection into the market for private health insurance in England. Both our simulations, and the findings using real data, suggest that parameter heterogeneity is a relevant issue that can confound the interpretation of standard 'unused characteristics' approaches. Our findings are important for analysing the efficiency of insurance markets. They are of interest to both the insurance industry and policymakers, and should be accounted for when selection based on specific characteristics needs to be detected or the effects of structural changes of insurance policies/markets are to be predicted.

Keywords: Insurance Markets; Information Asymmetries; Selection; Applied Econometrics

JEL classification: C18; G22; I13

1 Introduction

There has been a preponderance of theoretical and empirical research in the area of selection based on information asymmetries ever since Akerlof (1970) and Rothschild and Stiglitz (1976) wrote their respective seminal works on information asymmetries and market efficiency. Contemporary economic literature suggests that both adverse and advantageous selection in insurance plans can induce inefficiencies (e.g. De Meza and Webb, 2001; Einav et al., 2010) and that they depend on the actual circumstances, e.g. the kind of market at hand and the institutional arrangements involved (Cohen and Siegelman, 2010). However, in order to discuss implications for welfare in a specific market a proper empirical detection of selection in the first place is key, and this is what our paper is emphasizing. From a policy perspective this is of tremendous importance both for judging the actual performance of an insurance market, and also in terms of predicting outcomes, when new policies (e.g. regulation of the contracting of health services) are planned for implementation.

Since the seminal contribution of Rothschild and Stiglitz (1976), which represents the classical case of pure adverse selection, other types of selection have received attention in the literature. The most prominent of the latter is advantageous selection,[1] which entails a situation where low-risk customers purchase more insurance coverage because risk is negatively correlated with some other characteristic that is not used for pricing (e.g. risk aversion) but which determines demand for insurance (Finkelstein and McGarry, 2006; Fang et al., 2008; Buchmueller et al., 2013).

It has been shown that adverse selection can appear in a range of different market settings, whereas advantageous selection driven by asymmetric information is typically not compatible with competitive markets, since no separating equilibrium is possible under such circumstances (Chiappori and Salanié, 2013). If there are sufficiently high administrative costs, however, a negative coverage-risk correlation driven by multidimensional private information is possible even in a perfectly competitive insurance market (Fang and Wu, 2018). The methodological contribution

[1] In long-term contracts, selection effects can also arise due to one-sided commitment and learning (Hendel and Lizzeri, 2003).

made in this paper is relevant in any setting where standard testing procedures are used to detect sources of selection, regardless of the actual mechanism giving rise to selection effects.

The empirical strategies on selection which are discussed in this paper (e.g. Finkelstein and McGarry, 2006; Fang et al., 2008) are usually applied to indirectly derive evidence about asymmetric information, i.e. coefficients are interpreted as indicating asymmetric information, depending also on the market structure at hand. Furthermore, they do not allow us to distinguish between adverse selection and moral hazard (e.g. Finkelstein and Poterba, 2014; Browne and Zhou-Richter, 2014), although this distinction is beyond the scope of this study. Our contribution is mainly a technical one, discussing whether the findings of empirical methods that apply variables which are not used by insurance companies for underwriting purposes can be used to test if such a variable is affecting the risk pool in a positive or negative way. Bearing this in mind, we use the term 'selection' throughout the paper. Furthermore, we use the term 'selection' from the perspective of an expected loss/benefit of the insurance company, i.e. we actually mean adverse or advantageous selection.

The early empirical literature (e.g. Chiappori and Salanié, 1997; Chiappori and Salanie, 2000) was, in particular, concerned with finding evidence of information asymmetry based on the prediction that private information will lead to a positive correlation between risk and insurance coverage. This analysis is typically done by means of a 'positive correlation test', which shows how an individual's insurance coverage and *ex post* risk are correlated after all individual characteristics used in the pricing of contracts have been partialled out. Such tests can indeed carry information regarding the relevance of information asymmetries in the aggregate. However, when there are several types or sources of selection, their usefulness for understanding individual behaviour and the functioning of the market becomes quite limited. The very mixed evidence that comes out of this early literature suggests that a more detailed analysis is required. For example, the evidence in the market for health insurance seems to be strongly heterogeneous. Looking at studies which focus on different insurance markets, Cohen and Siegelman (2010) find evidence for both the existence of asymmetric information and market efficiency.

During the last decade a large volume of literature has emerged which deals with these issues by considering the specific origins of selection in insurance markets (e.g. Finkelstein and McGarry, 2006; Fang et al., 2008). The underlying rationale is that there can be different types of selection which may cancel out in the aggregate. The strategy typically used is to study variables which we refer to as 'unused characteristics' in the rest of this paper, i.e. variables that are not used in the calculation of the insurance premium by an insurance company, but which are available as observations for an empirical analysis. Such unused characteristics may either be private information, and thus unobserved by the insurance company, or 'unused observables' – i.e. variables that are known to the insurance company but are not used for pricing purposes (Finkelstein and Poterba, 2014; Kesternich and Schumacher, 2014). From a methodological point of view, the two types of unused information are equivalent, but the policy implications can be very different since 'unused observables' could be used in pricing and thus could potentially improve market efficiency.

The ability to attribute selection in insurance markets to *specific variables* is of high theoretical and empirical importance, in particular when the evidence regarding the overall direction of selection is unclear or conflicting. This approach has been fruitfully applied by several studies that used different markers for selection in insurance markets. Examples of such sources of selection include a variety of factors such as preferences (Finkelstein and McGarry, 2006; Cutler et al., 2008; Bauer et al., 2017), cognitive abilities (Fang et al., 2008), variables that reflect the economic situation, such as financial wealth or education (Bolhaar et al., 2012; Gao et al., 2009; Buchmueller et al., 2013; Browne and Zhou-Richter, 2014) and place of residence (Finkelstein and Poterba, 2014; Bauer et al., 2017).

Despite the aforementioned intense and ongoing usage of such 'unused characteristics' approaches regarding sources of selection, no systematic assessment of the properties of the tests has taken place to date. This is quite remarkable considering that the identification of sources of selection has been used to draw strong conclusions regarding both individual behaviour and the functioning of insurance markets. Therefore, one of the main contributions of this paper is to evaluate, both theoretically and empirically, the impact of parameter heterogeneity on the properties of testing approaches which use 'unused characteristics' to identify selection into insurance markets. Parameter heterogeneity is present whenever such a characteristic has a different impact on the key dependent variables (typically risk type and insurance coverage) for different individuals. For example, the unused characteristic could be an important determinant of health only for a specific part of the population, and an important determinant of insurance coverage for a completely different part of the population. In such a scenario, the standard testing procedures would suggest that the unused characteristic is a source of selection, even though it is not. Using a simulation exercise where we are able to control the relationships between the key variables, we show that this is more than a theoretical possibility and actually quantitatively relevant in many plausible settings.

In addition, parameter heterogeneity is a real and relevant challenge given the growing literature documenting heterogeneity in the effects of various determinants of health such as time availability (Eibich, 2015; Berniell and Bietenbeck, 2017); wealth (Cesarini et al., 2016); public health interventions (Bhalotra et al., 2017); and lifestyles (Felfe et al., 2016). A large literature in medicine on gene-environment interactions delivers a further rationale for why we should expect environmental determinants of health to be heterogeneous in their impact. This does indeed appear to be the case e.g. regarding the effect of diet on obesity (Qi et al., 2014) or the effects of various lifestyle factors on the risk of multiple sclerosis (Olsson et al., 2017). There are also reasons to expect that the impact of unused characteristics on the demand for insurance is heterogeneous. For example, Brown et al. (2016) find substantial heterogeneity in the state-dependence of utility. In general, people tend to value consumption in the healthy state more; however, this tends to vary with the type of health shocks considered. In addition, there is substantial heterogeneity between individuals in their valuation of consumption in the unhealthy state, and this heterogeneity is largely orthogonal to personal characteristics. It follows that any such personal characteristic which predicts demand for insurance will be heterogeneous in its impact.

In this paper, to further illustrate the interplay between heterogeneity and unused characteristics, we consider the variable "time availability" in the context of private health insurance in England. We might ask, for example, what role heterogeneity may play when considering time availability as a determinant of selection into private health insurance? It could well be the case that some very impatient individuals increase their likelihood of purchasing health insurance (e.g. if they want to reduce waiting times), whereas impatience in other cases might reduce the same likelihood (e.g. if it leads to less forward-looking behaviour and lower demand for health insurance). Hence, it is easy to see how heterogeneity may play an important role in the interpretation of 'unused characteristics' approaches. Therefore, it is important to compare the outcomes of these approaches in detecting selection in the real world. As a result, we use individual level panel data from England in order to assess whether the 'unused characteristic' availability of time is a determinant of selection into the private English health insurance market, allowing for heterogeneity at the individual level.

The aims of this paper are threefold: Firstly, the main contribution of this study is a technical one. We formally demonstrate that standard 'unused characteristics' approaches which allow the detection of selection due to specific characteristics can lead to the wrong conclusions being made. Secondly, we create artificial data to emphasize this issue allowing for different correlation structures that reflect selection effects into the insurance market. We use a multilevel model to take this phenomenon into account, allowing for individual parameter heterogeneity in our estimates. Thirdly, we provide further empirical evidence, using the 'availability of time' as a potential source of risk-based selection into the market for private health insurance, using longitudinal data for the English population over the age of 50.

In the next section we provide a literature overview with an emphasis on commonly used tests which empirically identify information asymmetries and selection in insurance markets. Following this, we show first theoretically, and then by simulation, that under specific circumstances tests based on two equations which are often applied to detect information asymmetries with 'unused characteristics' can give misleading results. We then give a brief overview of the institutional framework in the English health care system and provide additional empirical evidence on selection and parameter heterogeneity in the market for private health insurance in England using the English Longitudinal Survey of Ageing (ELSA). After discussing our findings we draw conclusions and make suggestions for future research in this field.

2 Theoretical Considerations

2.1 The Detection of Selection in Insurance Markets

Our starting point will be the 'positive correlation test', originally developed by Chiappori and Salanié (1997). Formally the approach can be described by the following equations,

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

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

where I is an indicator for insurance status and R is an indicator for being at risk, while X is a matrix containing variables which an insurance company uses for calculating the insurance premium. The vectors ϕ and ψ contain parameters to be estimated and ϵ and η reflect error terms of the equations. Under the null hypothesis of no asymmetric information, the residuals ϵ and η should be uncorrelated. Conversely, a significant correlation between the two is indicative of asymmetric information.

Subsequent literature assesses the issue of multiple dimensions of private information in the context of detecting selection. As Finkelstein and McGarry (2006) argue, the correlation of error terms in the 'Chiappori approach' is neither a necessary nor a sufficient condition for the existence of information asymmetries about the risk type.

The authors suggest that misleading results may arise if several characteristics have an impact on both dependent variables (some negatively, some positively), and that, overall, the effects cancel each other out on average. For example, in addition to an individual's risk class, the 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 the inclusion of these variables as additional explanatory variables into equations (1) and (2) will make it possible to detect and separate out the particular source of selection. This approach, which we are calling an 'unused characteristics' approach, is based on the equations

$$I = X\phi + Z\delta + \epsilon \tag{3}$$

$$R = X\psi + Z\beta + \eta \tag{4}$$

where Z represents a matrix containing additional information about the insured which is not used for pricing. The condition for attributing selection would in this case 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. As mentioned above, the test can also be applied to characteristics which are observable but not used in pricing (Finkelstein and Poterba, 2014).

Fang et al. (2008) develop a similar approach in order to reveal 'unused characteristics' that drive selection in insurance markets. Assuming 'used observables' are already partialled out, their approach is based on the regression model:

$$I = \alpha_1 + \alpha_2 R + \nu \tag{5}$$

followed by a regression where an 'unobserved' variable z is included:

$$I = \gamma_1 + \gamma_2 R + \gamma_3 z + \mu. \quad (6)$$

It follows directly from the formula of omitted variable bias that the OLS estimator $\hat{\alpha}_2$ has expectation $\mathbb{E}(\hat{\alpha}_2 \mid R) = \gamma_2 + \gamma_3 (R'R)^{-1} \mathbb{E}[R'z \mid R] = \gamma_2 + \gamma_3\hat{\beta}$, where $\hat{\beta}$ in the second term is the OLS coefficient of z when regressed on R. The detection of selection is based on the difference between the estimates of α_2 and γ_2. Hence, the 'detected' source of selection induced by z is defined as $\mathbb{E}(\hat{\alpha}_2) - \mathbb{E}(\hat{\gamma}_2) = \gamma_2 + \gamma_3\hat{\beta} - \gamma_2 = \gamma_3\hat{\beta}$. In the case of detected advantageous selection driven by z this difference will be negative i.e. $(\gamma_2 > \alpha_2)$ if $\gamma_3 < 0$ and z is positively partially correlated with R. As compared to the approach suggested by Finkelstein and McGarry (2006), evidence is not based directly on the comparison of how z associates with two different outcomes, but on one coefficient and the partial correlation of z and R.

2.2 Introducing Heterogeneity

So far, we have followed the convention in the literature of assuming *parameter homogeneity*, i.e. that the 'unused characteristics' have the same relationship with R and I in all parts of the population. In the light of the abundance of evidence suggesting that heterogeneity is a relevant phenomenon for many health risks, we now consider the consequences of relaxing the homogeneity assumption. In order to simplify the notation, we now suppress the used observables X in all equations. All derivations that follow are thus implicitly conditional on X. This leads to the simplified system of equations

$$I_i = \delta_i z_i + \epsilon_i \quad (7)$$

$$R_i = \beta_i z_i + \eta_i \quad (8)$$

where heterogeneity is captured by the individual-specific coefficients (δ_i, β_i) with means (μ_δ, μ_β). If the system is identified – which requires $\mathrm{Cov}(z_i, \beta_i) = \mathrm{Cov}(z_i, \delta_i) = 0$ – and imposing homogeneity, the OLS estimators $\hat{\delta}$ and $\hat{\beta}$ would provide unbiased and consistent estimates of the population parameters μ_δ and μ_β.

A finding that $\hat{\delta} > 0$ and $\hat{\beta} > 0$ would be interpreted as evidence of adverse selection.

In order to assess the consequences of heterogeneity, we impose the assumptions that $\mu_\delta > 0$ and $\mu_\beta > 0$ so that we expect the test based on unused characteristics to suggest adverse selection. The critical issue is whether these assumptions imply $\mathbb{E}(R_i) \geq \mathbb{E}(R_i|I_i > 0)$ also in the presence of heterogeneity. In order to assess whether this is the case, consider the average risk of the insured subpopulation, i.e. individuals who satisfy the restriction $\delta_i z_i + \epsilon_i > 0$:

$$\mathbb{E}(R_i|I_i > 0) = \mathbb{E}(\beta_i z_i|\delta_i z_i > -\epsilon_i) \tag{9}$$

Given the population means μ_β and $\mathbb{E}(z_i) = \mu_z$, the degree of adverse selection may be written as

$$\begin{aligned}\mathbb{E}(\beta_i z_i|\delta_i z_i > -\epsilon_i) - \mathbb{E}(R_i) =& \mu_\beta \times \mathbb{E}(z_i - \mu_z|\delta_i z_i > -\epsilon_i) \\ &+ \mu_z \times \mathbb{E}(\beta_i - \mu_\beta|\delta_i z_i > -\epsilon_i) \\ &+ \mathbb{E}((\beta_i - \mu_\beta)(z_i - \mu_z)|\delta_i z_i > -\epsilon_i).\end{aligned} \tag{10}$$

The first and second term in (10) are positive due to our assumptions. Hence, $\mathbb{E}(R_i) \geq \mathbb{E}(R_i|I_i > 0)$ 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. Concerning a situation with $\mu_\beta < 0$ and $\mu_\delta > 0$ (or vice versa), so that the test suggests advantageous selection, the true selection may be adverse if $\text{Cov}\,(\beta_i, \delta_i) > 0$.

Next, we consider whether parameter heterogeneity is a problem with the approach suggested by Fang et al. (2008). The estimate for $\mathbb{E}\,(\hat{\alpha}_2 \mid R)$ remains unchanged in this scenario. However, an estimate based on equation (6) – which imposes parameter homogeneity on γ_3, the coefficient associated with the unused variable, will deliver a biased estimate of γ_2, the coefficient associated with the risk variable R. This can be shown as follows: specification (6) is equivalent to a model where I_i is regressed on $\tilde{R}_i = R_i - \mu_\beta z_i = (\beta_i - \mu_\beta) z_i + \eta i$ (Angrist and Pischke, 2008). The resulting parameter has expectation $\mathbb{E}(\hat{\gamma}_2 \mid R) = \gamma_2 + \mathbb{E}\left[\text{Var}(z_i) \text{Cov}(\gamma_{3i}, \beta_i) / \text{Var}\left(\tilde{R}\right) \mid \tilde{R}\right]$. The intuition here is that when we regress R on z in order to partial out that variable, we do it without taking heterogeneity into account, and for this reason $\tilde{R}_i$ still includes a component of individual heterogeneity related to z. Hence, whenever the parameters are correlated, the test for selection will deliver biased results.

3 Empirical Implementation

In the previous section we have shown that parameter heterogeneity at the individual level (δ_i, β_i) may lead to incorrect conclusions regarding selection into health insurance. We now turn to an empirical assessment of this claim, first using evidence based on simulated data, and then real-world data regarding the market for private health insurance in England.

3.1 Econometric Model

Throughout the empirical analysis, we compare and contrast results coming out of three different methods. We complement the two standard approaches discussed above (Fang et al., 2008; Finkelstein and McGarry, 2006) with a method that allows for heterogeneity in the parameters associated with the unused observable z. This third approach relies on panel data and is implemented as a multilevel model with random coeffcients (called the 'RC Model' henceforth; cf. Hsiao, 2014). The two estimating equations used in this part are

$$I_{it} = \delta_i z_{it} + X_{it}\phi + \epsilon_{it} \tag{11}$$

$$R_{it} = \beta_i z_{it} + X_{it}\psi + \nu_{it} \tag{12}$$

where I_{it} and R_{it} represent insurance demand and risk of individual i in period t, and z_{it} is the unused observable. The vector X_{it} includes all observable characteristics which may be used for calculating the insurance premium. In order to asses whether conclusions based on the standard approaches can be misleading due to parameter heterogeneity, a critical statistic will be our estimate of $\mathbb{E}\left(\beta_i \delta_i\right) = \mathrm{Cov}(\beta_i, \delta_i) + \mathbb{E}(\beta_i)\mathbb{E}(\delta_i)$. This statistic follows directly from the decomposition of covariance discussed in section 2.2 and measures by how much the risk pool of insured individuals changes when z is increased by one unit – in other words, it measures the degree and direction of selection attributable to z.

In our comparative analysis we apply both 'unused characteristics' approaches by pooling the data in order to compare their results. When applying the approach suggested by Fang et al. (2008), where 2 equations are subsequently estimated, we derive the direction and degree of selection based on the difference in the partial correlation of I and R before and after an 'unused characteristic' is included in the specification.

3.2 Evidence by Simulation

For our simulation analysis, panel data with 1,000 cross-sectional observations and 5 time-series units were generated under different assumptions regarding the distribution of the parameters δ_i and β_i, as well as other variables in the analysis. Assuming that other information X used for calculating the insurance premium is already partialled out before estimation, the simulations are based on the following structural equations, where i refers to an individual and t to the time unit.

$$I_{it} = \delta_i z_{it} + \epsilon_{it} \tag{13}$$

$$R_{it} = \beta_i z_{it} + \eta_{it} \tag{14}$$

z_{it}, ϵ_{it} and η_{it} are standard normal random variables whereas δ_i and β_i are fixed individual-level parameters that vary randomly between individuals and simulations. The population means, standard deviations and the degree of correlation between the parameters δ_i and β_i vary between different scenarios. We consider eight distinct scenarios and run 100 simulations for each one. The main results are presented in Table 1. Since our focus is on the properties of the tests considered, we only present evidence regarding their performance, and do not report the parameter estimates coming out of the regressions.

Table 1: Simulation Results

	Simulation Parameters							Empirical Evaluation								
								FMcG			**FKS**			**RC Model**		
No.	μ_β	μ_δ	σ_β	σ_δ	$\rho(\beta_i, \delta_i)$	$\rho(e_1, e_2)$	$\mathbb{E}[\beta_i, \delta_i]$	TP^n	TP	FP	TP^n	TP	FP	TP^n	TP	FP
1	0.00	0.00	1.00	1.00	-0.30	0.00	-0.30	0.56	0.01	0.00	0.52	0.00	0.00	1.00	1.00	0.00
2	0.00	0.00	1.00	1.00	0.00	0.00	0.00	.	.	0.00	.	.	0.00	.	.	0.06
3	0.00	0.00	1.00	1.00	0.40	0.00	0.40	0.66	0.01	0.00	0.61	0.00	0.00	1.00	1.00	0.00
4	0.00	0.00	2.00	2.00	0.10	0.00	0.40	0.57	0.00	0.00	0.55	0.00	0.00	1.00	0.81	0.00
5	0.30	0.20	1.00	1.00	0.00	0.00	0.06	1.00	1.00	0.00	1.00	0.00	0.00	1.00	0.88	0.00
6	0.30	0.20	1.00	1.00	-0.30	0.00	-0.24	0.00	0.00	1.00	0.00	0.00	0.17	1.00	0.86	0.00
7	0.30	0.20	1.00	1.00	-0.60	0.00	-0.56	0.00	0.00	1.00	0.00	0.00	0.57	1.00	1.00	0.00
8	0.30	0.20	1.00	1.00	-0.60	-0.20	-0.56	0.00	0.00	0.99	0.00	0.00	0.07	1.00	1.00	0.00

Notes: The seven leftmost columns present the parameters used in the simulations, where β is the parameter associating the 'unused observable' z with health risk, and δ associates z with insurance coverage. We denote the correlation coefficient by $\rho(\cdot, \cdot)$. For three test procedures – based on Finkelstein and McGarry (2006), Fang et al. (2008) and a random coefficients model, respectively – three statistics are presented as a fraction of 100 simulations: TP^n is the fraction of cases correctly identifying the sign of the selection related to z, TP is the corresponding fraction that also attains statistical significance, and FP represents the fraction of cases where the suggested direction of selection is wrong and statistically significant.

The left side of the table (the first 7 columns) contains the assumptions underlying each run of simulations, whereas the right side shows the performance of the different tests. For each test, we present three different statistics: TP^n (where TP stands for true positive) counts the fraction of simulations where the conclusion of the test – without considering statistical significance – corresponds to the actual selection imposed by z. It follows that the corresponding error rate (which we will refer to as 'false positives' for convenience) equals $1 - TP^n$, and hence it is not reported in the table. For example, if the scenario imposes adverse selection and the estimates $\hat{\beta} > 0$ and $\hat{\delta} > 0$, then the Finkelstein and McGarry (2006) approach concludes that there is adverse selection, which would be a true positive. The statistic TP in the following column requires in addition that both estimates are significantly different from zero. Finally, FP measures the rate of '*false positives*', representing cases where the tests would reach the wrong conclusion and would do so with statistical significance.

Our first three sets of simulations show what happens if selection is 'hidden' in the sense that it is completely driven by parameter heterogeneity and $\text{Cov}(\beta_i, \delta_i) \neq 0$ – whereas the average effect of z on R and I equals zero. In scenario 1, there is *advantageous* selection: individuals who face a reduced risk due to z ($\beta_i < 0$) are also more likely to purchase insurance ($\delta_i > 0$) on average. In this case, measured by the TP^n criterion, neither the Finkelstein and McGarry (2006) approach nor the Fang et al. (2008) approach is much better than chance, with a success rate of 0.56 and 0.52 respectively. If in addition we require statistical significance, the success rate of both equals zero. A multilevel estimation procedure, which allows for parameter heterogeneity, performs much better and identifies advantageous selection in 100 cases out of 100.

Scenario 2 assumes that there is no selection related to z. In this case there are, of course, no 'true positives' but on the other hand, 'false positives' represent a possibility. The Finkelstein and McGarry (2006) and Fang et al. (2008) procedures do not have any such false positives, and the RC model has a false positive rate of six per cent. In scenario 3 we increase the degree of selection (the number of 'positives') and set $\text{Cov}(\beta_i, \delta_i) > 0$ (*adverse* selection). The approaches of Finkelstein and McGarry (2006) and Fang et al. (2008) are slightly better at identifying selection if measured by counts (TP^n) as compared to scenario 1 (66 per cent as compared to 56 per cent, and 61 per cent as compared to 52 per cent, respectively). However, these 'true positives' do not hold in terms of statistical signficance (as demonstrated by the TP columns) whereas the RC model also detects selection if we consider statistical significance. Scenario 4 changes the variance of parameter heterogeneity, keeping $\text{Cov}(\beta_i, \delta_i)$ identical to scenario 3. These changes do not affect the degree of selection attributable to z, but they decrease the performance of all test procedures.

From scenario 5 onwards, we allow the average effects of z on R and I to be different from zero. This has a great impact on the results of the three test procedures. Scenario 5 represents a case where there is adverse selection related to z ($\mathbb{E}[\beta_i \delta_i] = 0.06$) and no parameter heterogeneity. In this case, the standard 'unused characteristics' procedures correctly identify the direction of selection, but the approach of Finkelstein and McGarry (2006) performs in a superior way when we consider statistical significance as compared to the other approaches. In scenarios 6–8 we combine adverse selection in the average effects and allow for parameter heterogeneity in the opposite direction. For example, in scenario 6, $\text{Cov}\,(\beta_i, \delta_i) = -0.3$ clearly dominates $\mathbb{E}(\beta_i)\mathbb{E}(\delta_i) = 0.06$ in the term $\mathbb{E}\,(\beta_i, \delta_i) = -0.24$ Here, the performance of the standard 'unused characteristics' procedures is strikingly poor as there are no 'true positives'. Moreover, the Finkelstein and McGarry (2006) procedure more often leads to a detection of statistically significant selection in the wrong direction as compared to Fang et al. (2008). The performance of the *RC* model improves if parameter heterogeneity becomes larger as shown in Scenario 7. In these scenarios, both standard test procedures are clearly dominated by the *RC* model which identifies the true direction of selection almost perfectly. The superiority of the *RC* approach is revealed in scenario 8 as well, where the error terms of the equations (13) and (14) are negatively correlated. Such a correlation can reflect the role of additional characteristics in the association of risk and insurance status and it also plays a role in the market for private health insurance in England, as discussed in 3.3. below.

In summary, our simulation analysis shows that both the standard testing approaches considered here will perform poorly when there is parameter heterogeneity which works in the opposite direction to that of the unused observable z on average. We have also seen that the *RC* model dominates the other two approaches under a wide range of assumptions. On the other hand, assuming heterogeneity at the individual level, the RC model requires that either panel data are available or that the parameter heterogeneity can be attributed to some other observable characteristic. Moreover, the RC model tends to have a greater number of false positives than the other two approaches.

3.3 The Market for Private Health Insurance in England

Our simulation analysis above has shown that parameter heterogeneity may invalidate some standard tests used to detect sources of selection. We now turn to an empirical application on real-world data in order to assess whether this theoretical possibility is also of practical relevance. Before conducting the analysis we provide some background information on the market for private health insurance in England.

3.3.1 Institutional Background in England

The population of England 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). The main principle of the NHS is to make health services available to every citizen who is in need.[2] 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.[3] 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 an employer (Boyle, 2011). Private insurance covers services which duplicate those of the NHS (Kiil, 2012) but also provides cover for enhanced services such as faster access and a wider consumer choice, compared to that which is offered by the NHS. Insurers can freely determine the services they offer, but most packages cover surgery as an inpatient or day case, hospital accommodation, nursing care and inpatient tests. Individual contracts tend to be renewable on an annual basis. Insurance companies mainly use medical underwriting and age as information for calculating the risk premium but other variables such as sex, smoking and occupational status may be used. Based on medical underwriting, certain conditions can be excluded from cover; however a policy will typically be offered also in these cases (Boyle, 2011; Association of British

[2] http://www.nhshistory.net/a_guide_to_the_nhs.htm
[3] http://www.londonhealth.co.uk/nhs/index.html

Insurers, 2012).

While there is very little regulation in the English market for private health insurance concerning the pricing of the products, and the products themselves, the UK Financial Conduct Authority (FCA) and the Prudential Regulation Authority (PRA) regulate health insurers in areas such as capital adequacy, consumer welfare and product sales (Karl, 2014). There is a lack of research on the competitiveness of the English market, but there are several reasons to believe it is relatively competitive. First, even though the market is relatively concentrated, no insurer dominates the market; the four largest insurers (BUPA with a market share of 42%; AXA PPP, Norwich Union and Standard Life) had a joint market share of 83.5 per cent in 2006. The Herfindahl-Hirschman Index of 0.25 which these numbers imply, is well below the pre-ACA level of concentration in the U.S. at 0.41 (Boyle, 2011; Dafny et al., 2015). Second, the market power of insurers is limited by the existence of close substitutes such as the NHS and self-funded employer plans (Markar and O'Sullivan, 2012). Third, at around 25 per cent, the premium loading factors were at the lower end of estimated ranges of 25-40 per cent for the individual market in the U.S. (Boyle, 2011; Newhouse, 2004; Brown and Finkelstein, 2007). This is again indicative of a somewhat larger competitive pressure than is experienced in, for example, the US.[4]

There is little evidence of selection in the market for private health insurance in England. Propper et al. (2001) analyse the dynamics in the demand for private health insurance (PHI) between 1978 and 1996 in the UK, using the Family Expenditure Survey. Controlling 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 PHI, are important factors in deciding whether PHI is purchased.

Wallis (2004) also looks for the determinants of demand for PHI in the UK, based on British Household Panel Survey (BHPS) data. The author evaluates the switching behaviour of individuals and focuses both on characteristics influencing the

[4]Finally, confirmation that the UK market is competitive comes from a senior pricing actuary working in the health insurance market (Cheung, 2018).

probability of purchasing insurance and the individual cost of PHI (i.e. the insurance premium). He differentiates between the demand side characteristics of consumers and the supply side aspects that can influence insurance status, e.g. quality of service.

A more recent study, which uses the BHPS data, is done by Olivella and Vera-Hernández (2013), who focus on adverse selection in the market for PHI, using hospitalisation as a measure for being at risk. As Olivella and Vera-Hernández (2013) both theoretically and empirically show there is, overall, a positive correlation between health risk and private health insurance status in England. Assuming that the health status of an individual is independent of receiving PHI as a fringe employment benefit, their results suggest the existence of adverse selection in the PHI market in England. Their findings suggest that this selection may be partly explained by preferences for health (care).

3.3.2 ELSA Data

Our empirical illustration of parameter heterogeneity and selection in insurance plans is based on ELSA, which is a representative individual level dataset for England's age 50+ population. The ELSA dataset contains a broad range of information on each individual's health and financial circumstances, together with the overall demographics, which makes it an ideal source to model both economic decisions and health-related characteristics.

For our analysis we use both the cross-sectional and longitudinal dimension of the ELSA survey so that the period we are using captures the time from year 2002 to 2013. We restrict the data to individuals aged 90 or younger since we cannot identify the actual age of people over age 90. Furthermore, in respect of the longitudinal nature of our dataset, we account for attrition and make the sample we use representative for the first wave by applying the sample weights provided with the ELSA data.

As our main dependent variable, we use self-assessed health (SAH) as a measure for being at risk. This is a widely used indicator, also in the literature on selection in health insurance (Doiron et al., 2008; Bolin et al., 2010), with well-known pros and cons. The main strengths of this indicator for our purposes are that it has exceptional predictive validity regarding some dimensions of health, such as mortality (Idler and Benyamini, 1997; Benyamini and Idler, 1999; DeSalvo et al., 2006) and it captures not only observable information but also information that can affect future demand for health care which cannot be accounted for from using only observable and objective health data. In addition, recent research has shown that the validity of this indicator has increased over time (Schnittker and Bacak, 2014).

On the other hand, there are concerns that SAH is prone to measurement error due to *inter alia* subjective biases in the perception of health. A related issue is that it appears to capture some dimensions of health better than others: alongside mortality, additional aspects such as vitality, mobility and pain are closely related to SAH whereas mental health and social functioning tend to get much less weight (Au and Johnston, 2014). It has also been shown that SAH reflects serious chronic conditions better than less severe conditions (Doiron et al., 2015). Doiron et al. (2015) also show that the relationship between SAH and health care utilisation weakens considerably when conditioned on some objective indicators of health. In order to mitigate the concerns we conduct a robustness check to see if our results are robust to changes in the health indicator used.

ELSA provides the commonly used 5-point scale for Self-Assessed-Health (SAH) which we collapse into binary variables that we call 'High Risk' (HR) and 'Low Risk' (see descriptive statistics in Table 2). The variable HR captures being in 'fair' or 'poor' health status as compared to 'excellent', 'very good' and 'good' (i.e. low risk). The variable HR is used in our analysis to capture information when an individual poses a relatively high health risk from the perspective of an insurance company. The second main dependent variable is a dummy variable (PHI) which equals 1 if someone has private health insurance and 0 otherwise.

We exclude people from the analysis who have PHI cover only as part of the employee benefits package being offered by their employer. This is because the way in which such group cover is purchased by an employer, and the way in which it is

Table 2: Summary statistics

Variable	Mean	Std. Dev.	Min.	Max.	N
PHI	0.126	0.331	0	1	31492
HR	0.266	0.442	0	1	31492
time_avail	0.774	0.418	0	1	31492
female	0.565	0.496	0	1	31492
age	67.633	8.959	50	89	31492
couple	0.690	0.463	0	1	31492
children	0.883	0.322	0	1	31492
retired	0.599	0.49	0	1	31492
unemployed	0.112	0.316	0	1	31492
managerial	0.095	0.293	0	1	31492
intermediate	0.037	0.19	0	1	31492
small employer	0.049	0.217	0	1	31492
lower supervisory	0.024	0.154	0	1	31492
semi-routine	0.083	0.276	0	1	31492
smoke_now	0.129	0.335	0	1	31492
smoke_past	0.5	0.5	0	1	31492
CHH: excellent	0.332	0.471	0	1	27900
CHH: very good	0.334	0.472	0	1	27900
CHH: good	0.208	0.406	0	1	27900
CHH: fair	0.086	0.281	0	1	27900
CHH: poor	0.034	0.181	0	1	27900
CHH: varied a lot	0.006	0.076	0	1	27900

Notes: Information about variable definitions is provided in Appendix A1.

priced, are both very different from the approach which is adopted for individual policies. Therefore, the two groups may not be regarded as being similar in our statistical analysis. As previously mentioned, it is important for our analysis that the econometric model contains all the relevant information used by an insurance company in order to calculate the insurance premium. Our initial selection of variables included all variables used by Olivella and Vera-Hernández (2013) which are also available in the ELSA data, but since our analysis only includes individual policies, some variables could be excluded – in particular variables capturing whether the PHI coverage is provided by an employer. Hence, we condition on the following variables that may used by insurance companies: age, sex, smoking (history), employment status (retired, unemployed or category of National Statistics Socioeconomic Classification), occupation, family status (indicator for being married or cohabiting), and whether the person has children. The last two variables are included since health insurance policies can also include the whole family of a policy holder.[5] We further use dummy variables for the local authority region in which the respondent is living, information that is given to an insurance company when buying a policy. From the perspective of an insurance company this variable is important since health status is known to vary strongly between different English regions (Newton et al., 2015).

Wave 3 of ELSA provides information about the individual's self-assessed health status during childhood. We use this variable in order to partially account for an individual's health history. This is important since insurers may use the past and present health status as factors to take into account when calculating a policy holder's premium (Boyle, 2011). Although self-assessed childhood health is probably not fully able to capture all the health risks of an individual which are available to the insurance company when underwriting, there is a growing literature emphasizing the importance of early child health on health outcomes much later in life (e.g. Blackwell et al., 2001; Delaney and Smith, 2012; Hamad et al., 2016; Bhalotra et al., 2017). Hence, we generate 6 indicator variables in order to capture variation in initial health status – see Appendix A1.

[5]https://boughtbymany.com/news/article/private-health-insurance-cost-uk/#factors

As regards the ‘unused characteristic’, we have to choose a variable from the literature which is likely to drive selection within the private health insurance market but cannot directly be assessed by an insurance company and is therefore not available for calculating a policyholder’s premium. Furthermore, the variable should be available across several waves since we follow a panel-data approach – and, in particular, we should expect it to have a heterogeneous impact on risk and on the propensity to purchase private health insurance. A variable in our data which does meet these criteria is “availability of time”.

From a health economics perspective, restrictions in time availability are of great interest since the decision to take out private health insurance in the UK is known to depend on waiting times for health care (e.g. King and Mossialos, 2005; Johar et al., 2013). As the waiting times in the English market for health care are usually much shorter with private health insurance than with the NHS, a patient’s views on the waiting time and the corresponding opportunity costs will determine whether an individual takes out private insurance or not. If a patient is willing and able to wait longer for treatment then there is less incentive to opt into the PHI market. This is also in line with Buchmueller et al. (2013) who suggest that the high opportunity cost of time may explain a part of the advantageous selection in the Australian private health insurance market.

Concerning health, there is a rich literature documenting a positive relationship between time availability and health. Grossman’s seminal model, where time is an input into the health production process, provides a theoretical rationale (Grossman, 1972): health production requires not only material resources but also time. Subsequent empirical work has confirmed that goods and time are complements in the production of health (Du and Yagihashi, 2017a). Various health measures have also been associated with working hours (Abramowitz, 2016), and in the literature establishing a negative relationship between the business cycle and health, the availability of time is one of the most prominent explanation candidates (Ruhm, 2015). Thus, in general, the economic literature suggests a positive relationship between time availability and health. There is, however, abundant evidence of heterogeneity in the relationship: for example, the effects of retirement on health have been found to be strongly heterogeneous in Germany (Eibich, 2015) and unemployment is sometimes found to have a detrimental effect on individual health (Urbanos-Garrido and Lopez-Valcarcel, 2015; Breuer, 2015).

We use the respondent's answer to the question 'Do you have enough time to do everything?' to capture an individual's availability of time. The ELSA survey offers 6 different choices to the respondent varying between 'strongly agree' and 'strongly disagree'. We collapse it into a binary variable with the aggregate content 'agree' (1) and 'not agree' (0).

We provide estimates of 4 different specifications. The first specification is used for comparison purposes and just accounts for age as an additional covariate. The second specification is our 'baseline'. We condition on variables that are assumed to be used for calculating the insurance premium for the PHI contract. For our third specification, we include time dummies to rule out changes over the time that may affect our estimates. In our fourth specification, we include the indicator variables which capture initial health status.

In Appendix A3 and A4 we provide some robustness checks by varying the cut-off point of our health risk variable and also use a more objective measure for health risk. One concern with the time availability measure is that it is likely to be related to income. Indeed, wages have been shown to have an impact on the time people have available for health-promoting activities (Du and Yagihashi, 2017b). We also take this possibility into account when controlling for personal income.

3.3.3 Results

Each row in Table 3 provides estimates for a different specification. In row 1 we present a minimal specification where we assume that only age is used to calculate the risk premium by health insurance companies. The estimates from our baseline specification are shown in row 2, whereas time effects and childhood health effects are accounted for in rows 3 and 4. The columns allow a comparison of the standard 'unused characteristics' approaches (columns 1 to 5) and the approach that accounts for parameter heterogeneity (columns 6 to 8).

As can be seen in column 3 there is a negative correlation between health risk and insurance across all specifications, which suggests *advantageous selection* in the aggregate. This empirical finding deviates somewhat from that of Olivella and Vera-Hernández (2013, Table 2), where health and insurance takeup are found to be unrelated. This is possibly due to our sample being very different in terms of age and sex. However, as our focus is not on the overall degree of selection, but on selection associated with a specific variable z, we now consider the interpretation of 'unused characteristics' approaches. Turning to the estimates of our baseline specification in row 2, the coefficients of the Linear Probability Model (LPM)[6] based on the approach of Finkelstein and McGarry (2006) (see columns 1 and 2 in Table 3) show that the availability of time is both negatively associated with someone's health risk status (-0.071) and the ownership of private health insurance (-0.018).[7]

Table 3: Estimates ELSA data

Spec.	N	(1) Finkelstein and McGarry	(2)	(3)	(4) Fang et al.	(5)	(6)	(7) RC-Model	(8)
Dependent variable		HR	PHI	PHI	PHI	PHI	HR	PHI	–
Parameter		$\hat{\beta}$	$\hat{\delta}$	$\hat{\alpha}_2$	$\hat{\alpha}_2 - \hat{\gamma}_2$	$\hat{\gamma}_3$	$\mathbb{E}\left(\hat{\beta}_i\right)$	$\mathbb{E}\left(\hat{\delta}_i\right)$	$\mathbb{E}\left(\hat{\beta}_i\hat{\delta}_i\right)$
1–age	31,573	-0.041***	-0.027***	-0.058***	0.001	-0.030***	-0.032***	-0.026***	-0.004***
		(0.008)	(0.007)	(0.005)	[0.845]	(0.007)	(0.008)	(0.006)	(0.001)
2–baseline	31,492	-0.071***	-0.018***	-0.040***	0.001	-0.021***	-0.059***	-0.016***	-0.002***
		(0.008)	(0.006)	(0.005)	[0.810]	(0.006)	(0.008)	(0.006)	(0.001)
3–time	31,492	-0.070***	-0.018***	-0.040***	0.001	-0.021***	-0.056***	-0.018***	-0.002***
		(0.008)	(0.006)	(0.005)	[0.810]	(0.006)	(0.008)	(0.006)	(0.001)
4–CHH	27,900	-0.062***	-0.021***	-0.035***	0.001	-0.024***	-0.049***	-0.020***	-0.002***
		(0.008)	(0.007)	(0.006)	[0.816]	(0.007)	(0.008)	(0.006)	(0.001)

Notes: The columns of each specification show results of unused characteristics approaches. Estimates resulting in columns 1 to 5 are based on Linear Probability Models. Columns 1 and 2 are estimates from the approach of Finkelstein and McGarry (2006) represented by the structural equations (3) and (4), whereas columns 3 to 5 reflect the estimates based on the structural equations (5) and (6) suggested by Fang et al. (2008). Coefficients of other 'used characteristics' are not shown above but detailed regression results for column 2 can be found in Appendix A2. An F test is used to test whether the coefficient of R (based on the specification in column 3) equals the coefficient of R after the inclusion of z into the model. Columns 6 to 7 show the coefficients from a multilevel model where the explanatory variable 'availability of time' is supposed to have a random coefficient and the effect of all other variables is supposed to be fixed. Column 8 contains $\mathbb{E}\left(\hat{\beta}_i\hat{\delta}_i\right)$, the degree of selection associated with z. Standard errors clustered at the individual level; standard errors in (), p-values in []. * p<0.10, ** p<0.05, *** p<0.01

The approach suggested by Fang et al. (2008) also indicates that the unused characteristic z is negatively associated with both risk and insurance. However, the association between R and I hardly changes when z is controlled for (column 4) and

[6]Despite theoretical concerns over the interpretation of the coefficients in a LPM (e.g. Wooldridge, 2003), we find that the LPM fits our data well. Its coefficients are, both in terms of economic relevance and statistical significance, very similar to the marginal effects derived from a probit model.

[7]The results for other control variables can be found in the Appendix A2, but are not of interest for our analysis since we assume that they are used in the calculation of an individual's insurance premium.

the difference is not statistically significant.[8] The interpretation of the positive point estimate in column 4 would be that z is a source of adverse selection, even though its impact is very small and not statistically significant.

When focusing on the estimates from our RC model, we find that the estimates of the fixed part of our model (columns 6 and 7) are similar to those obtained from the approach of Finkelstein and McGarry (2006): z appears to be a source of *adverse* selection according to this measure. However, the multilevel model allows us to account for heterogeneity and calculate $\mathbb{E}\left(\hat{\delta}_i\hat{\beta}_i\right)$; a statistic that measures the overall degree of selection related to z (column 8). This statistic is negative, which implies that on average, z is actually a source of *advantageous* selection – even though it enters negatively in the risk equation and also in the insurance equation.

In summary, our baseline specification has shown that in the scenario considered here, the approach of Finkelstein and McGarry (2006) would suggest that z is a source of adverse selection which is endorsed by the approach of Fang et al. (2008) although here there is not much selection detected that is related to z.

Next, we assess the robustness of the results. The estimates are nearly identical when we account for time fixed effects in specification 3. Following this, we account for someone's health during childhood in specification 4. The number of observations is considerably lower here, since some respondents of the survey did not answer this question. Notwithstanding that point, overall the findings are very similar to the ones in the other specifications 1–3. Here, we also see that the impact of z on private health insurance is decreasing when childhood health is included which suggests that an individual's initial health status is a predictor of the take up of PHI. Nevertheless, the main relationships identified in the other specifications remain robust. Although the standard 'unused characteristics' approaches suggest adverse selection, there is a negative correlation between the coefficients of interest which changes the sign of the selection.

Some additional robustness checks are provided in Appendix A3. First, we vary

[8] However, the literature usually does not test whether this difference is statistically significant (e.g. Buchmueller et al., 2013; Finkelstein and Poterba, 2014).

the cut-off in SAH used to generate our health risk variable, now assigning the categories 'good', 'fair' and 'poor' the value 1. Doing this changes some of the estimates slightly but not qualitatively. In addition, we turn to a more objective measure of health risk by using a variable that captures long-standing illnesses instead. This alternative risk variable takes the value 1 if respondents have a long-standing illness. Again, the results are very similar to our baseline case overall, even though some of the precision in the estimates is lost due to the reduced sample size, as this variable is not available in the first wave of the data. Finally, we check whether our estimates concerning the 'availability of time' are to some extent driven by differences in incomes. The results of these specifications can be found in Appendix A4. As expected, income affects health risk negatively and insurance status positively (coefficients not shown) – but this does not affect our conclusions regarding availability of time.

Our finding that the 'unused characteristic' time availability points into the direction of advantageous rather then adverse selection (after allowing for heterogeneity) is supported by Buchmueller et al. (2013) who argue that advantageous selection into PHI may to some extent be explained by the possibility of avoiding waiting times in the Australian health care sector.

Can our empirical findings concerning correlated coefficients be explained more intuitively? As mentioned above, economic theory suggests a negative relationship between an individual's availability of time and both their demand for health insurance and health risk. Which role can heterogeneity play here? Although there is literature on heterogeneity in time preferences concerning health care decisions (e.g. Grignon, 2009), we do not know the actual mechanism that can explain our empirical finding. We now present one very simple example that would generate the results of our preferred specification above. Suppose the population consists of two equally large groups, denoted 1 and 2 in what follows. In group 1, increased time availability is associated with a reduction in the demand for insurance ($\delta_1 = -0.036$) and a deterioration in health risk ($\beta_1 = 0.11$). We can think of this group as career oriented individuals who perceive a high opportunity cost of time when working hard, and who need to keep in shape in order to meet their job demands. Group 2, on the other hand, does not change their demand for insurance in response to time availability ($\delta_2 = 0$) but they do invest a lot more in their health when more time is available ($\beta_2 = -0.25$). We may think of this group as individuals who perceive time pressure for other reasons than career demands. With the parameter values suggested in this example, we would indeed have a situation where $\mathbb{E}(\beta) = \frac{1}{2} \cdot 0.11 - \frac{1}{2} \cdot 0.25 = -0.07$ and $\mathbb{E}(\delta) = -\frac{1}{2} \cdot 0.036 + \frac{1}{2} \cdot 0 = -0.018$, which leads standard tests to suggest time availability is a source of adverse selection. However, we also have $\mathbb{E}(\beta\delta) = -0.002$ so that the actual influence of z is to *improve* the risk pool of insured individuals.

4 Conclusion

In this paper we provided an overview about commonly used testing procedures for the detection of selection in insurance markets, focusing on their strengths and weaknesses. We argued that, although the classical positive correlation test might lead to incorrect conclusions about selection, it still has the advantage that it uses cross-equation correlations of the residuals acquired at the individual level. Our findings show that, just as two 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 about the nature of selection if the coefficients are heterogeneous concerning their association with both risk and insurance status. This issue arises because the strategy of both Finkelstein and McGarry (2006) and Fang et al. (2008) use two separate equations for detecting selection.

To emphasize the potential problem of parameter heterogeneity we discuss, formally, which circumstances can lead to false conclusions by allowing for the correlation of coefficients across different equations. We demonstrate the relevance of this finding with simulations by imposing different correlation structures between an 'unused characteristic' z, insurance status and risk, and additionally allowing for individual parameter heterogeneity that reflects different directions and degrees of selection. The results show that standard 'unused characteristics' approaches that solely focus on the interpretation of single coefficients do not reveal this kind of heterogeneity, i.e. a detected source of adverse selection may indeed be a source of advantageous selection if the individual coefficients are negatively correlated. The same phenomenon can obviously be found under certain correlation structures of the parameters if advantageous selection or even no selection is detected.

In our empirical implementation, using the market of private health insurance in England, we provided an example with the unused characteristic 'availability of time' which is not directly accessible by insurance companies. This variable may reflect opportunity costs of waiting times in the health care sector, resulting in demand for private health insurance. Our findings show that individual parameter heterogeneity is a relevant issue in real markets as well. Although adverse selection within the insurance market is empirically detected, this finding should not be interpreted as a selection of high risks into the market, since the estimated parameters are strongly negatively correlated due to the heterogeneous outcomes of our explanatory variable. The estimated 'mean' coefficient of such variables can be driven by different parts of the population and does not allow a meaningful interpretation concerning the underlying direction of selection. When interpreting our empirical findings we have to keep in mind that our aim is to focus on specific sources of selection, i.e. the total degree of selection within the insurance market is beyond the scope of this study.

Since the relevance of parameter heterogeneity is an empirical question, and a general conclusion for other markets and characteristics cannot be provided, we suggest that this possibility should be tested if one wants to reveal a specific source of selection in insurance markets. Our findings are important for analysing the efficiency of insurance markets. They are of interest to both the insurance industry and policy makers, and should be accounted for whenever outcomes of structural changes of insurance policies or the insurance market design overall are to be predicted, based on 'unused variables'. We wish to emphasise that our analysis deals with *selection* and that neither of the approaches discussed in this paper are used to reveal an underlying data generating process from a causal perspective. Instead the regression approaches are used solely as descriptive tools. Our contribution is designed in the same spirit, i.e. we do not say that one specific variable is causally related to two different dependent variables. The important thing is that, if there is a correlation between one explanatory variable and both risk and insurance status, the estimated coefficients cannot necessarily be interpreted as an indicator of selection in insurance markets in one way or the other.

Our findings are particularly relevant in the case of unused variables for which, a priori, we cannot assume a specific relationship with either insurance or health status. In this case, it is necessary to be very careful about potential parameter heterogeneity because even random outcomes of such ‘unused characteristics’ may be falsely interpreted as sources of selection in insurance markets and market inefficiencies. Although we do not make any claim about implications for welfare, the economic implications of our ideas should be taken into account in welfare analysis, because they directly affect the interpretation of which kind of selection is being identified in the market of interest.

Needless to say, our study has a number of limitations. With regard to our empirical application, we assume that a subjective health-risk variable is a good indicator for individual health status, but we do not know if it is also a good measure for future 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 (or lower) probability of making a claim in respect of health insurance. Future research should evaluate whether the robustness of this assumption can be supported when using objective data about health care utilization, e.g. number of visits to the doctor or, even better, treatment costs.

We assume a specific correlation structure $\mathrm{Cov}(\delta_\mathrm{i}, \beta_\mathrm{i}) \neq 0$ of the parameters of an ‘unsused characteristic’ on both risk and insurance status, i.e. the parameters are defined for each individual. However, one may also consider a correlation $\mathrm{Cov}(\delta_\mathrm{a}, \beta_\mathrm{a}) \neq 0$ on another level a that confounds the interpretation of the standard ‘unused characteristics’ approaches. If the variable defining groups is in itself an ‘unused observable’, it could be used to solve the issues with standard approaches

– since they will lead to correct conclusions if applied within each subsample separately. We believe that an interesting area for future research may be to allow for heterogeneity in the association between an unused characteristic z and both R and I on another level than the individual one.

References

Abramowitz, J. (2016). The connection between working hours and body mass index in the us: a time use analysis. *Review of Economics of the Household*, 14(1):131–154.

Akerlof, G. (1970). The market for 'lemons': Quality uncertainty and the market mechanism. *The Quarterly Journal of Economics*, 84(3):488–500.

Angrist, J. D. and Pischke, J.-S. (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Association of British Insurers (2012). Are you buying private medical insurance? `http://www.abi.org.uk/globalassets/sitecore/files/documents/publications/public/migrated/health/abi-consumer-guide-on-buying-private-medical-insurance.pdf`. [Online; 5 Jul. 2018].

Au, N. and Johnston, D. W. (2014). Self-assessed health: what does it mean and what does it hide? *Social science & medicine*, 121:21–28.

Bauer, J. M., Schiller, J., Schreckenberger, C., and Trautinger, M. (2017). Selection behavior in the market for private complementary long-term care insurance in Germany.

Benyamini, Y. and Idler, E. L. (1999). Community studies reporting association between self-rated health and mortality: additional studies, 1995 to 1998. *Research on Aging*, 21(3):392–401.

Berniell, M. I. and Bietenbeck, J. (2017). The effect of working hours on health.

Bhalotra, S., Karlsson, M., and Nilsson, T. (2017). Infant health and longevity: Evidence from a historical intervention in Sweden. *Journal of the European Economic Association*, 15(5):1101–1157.

Blackwell, D. L., Hayward, M. D., and Crimmins, E. M. (2001). Does childhood health affect chronic morbidity in later life? *Social science & medicine*, 52(8):1269–1284.

Bolhaar, J., Lindeboom, M., and van der Klaauw, B. (2012). A dynamic analysis of the demand for health insurance and health care. *European Economic Review*, 56(4):669–690.

Bolin, K., Hedblom, D., Lindgren, A., and Lindgren, B. (2010). Asymmetric information and the demand for voluntary health insurance in Europe. Technical report, National Bureau of Economic Research.

Boyle, S. (2011). United Kingdom (England) health system review. *Health systems in transition*, 13(1).

Breuer, C. (2015). Unemployment and suicide mortality: evidence from regional panel data in europe. *Health economics*, 24(8):936–950.

Brown, J. R. and Finkelstein, A. (2007). Why is the market for long-term care insurance so small? *Journal of Public Economics*, 91(10):1967–1991.

Brown, J. R., Goda, G. S., and McGarry, K. (2016). Heterogeneity in state-dependent utility: Evidence from strategic surveys. *Economic Inquiry*, 54(2):847–861.

Browne, M. J. and Zhou-Richter, T. (2014). Lemons or cherries? Asymmetric information in the German private long-term care insurance market. *The Geneva Papers on Risk and Insurance-Issues and Practice*, 39(4):603–624.

Buchmueller, T. C., Fiebig, D. G., Jones, G., and Savage, E. (2013). Preference heterogeneity and selection in private health insurance: The case of Australia. *Journal of Health Economics*, 32(5):757–767.

Cesarini, D., Lindqvist, E., Östling, R., and Wallace, B. (2016). Wealth, health, and child development: Evidence from administrative data on swedish lottery players. *The Quarterly Journal of Economics*, 131(2):687–738.

Cheung, J. (2018). private communication.

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

Chiappori, P.-A. and Salanie, B. (2000). Testing for asymmetric information in insurance markets. *Journal of political Economy*, 108(1):56–78.

Chiappori, P.-A. and Salanié, B. (2013). Asymmetric information in insurance markets: Predictions and tests. In *Handbook of insurance*, pages 397–422. Springer.

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

Cutler, D. M., Finkelstein, A., and McGarry, K. (2008). Preference heterogeneity and insurance markets: Explaining a puzzle of insurance. *The American Economic Review*, 98(2):157–62.

Dafny, L., Gruber, J., and Ody, C. (2015). More insurers lower premiums: Evidence from initial pricing in the health insurance marketplaces. *American Journal of Health Economics*, 1(1):53–81.

De Meza, D. and Webb, D. C. (2001). Advantageous selection in insurance markets. *RAND Journal of Economics*, pages 249–262.

Delaney, L. and Smith, J. P. (2012). Childhood health: trends and consequences over the life-course. *The Future of children/Center for the Future of Children, the David and Lucile Packard Foundation*, 22(1):43.

DeSalvo, K. B., Bloser, N., Reynolds, K., He, J., and Muntner, P. (2006). Mortality prediction with a single general self-rated health question. *Journal of general internal medicine*, 21(3):267–275.

Doiron, D., Fiebig, D. G., Johar, M., and Suziedelyte, A. (2015). Does self-assessed health measure health? *Applied Economics*, 47(2):180–194.

Doiron, D., Jones, G., and 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.

Du, J. and Yagihashi, T. (2017a). Goods–time elasticity of substitution in health production. *Health economics*, 26(11):1474–1478.

Du, J. and Yagihashi, T. (2017b). Health capital investment and time spent on health-related activities. *Review of Economics of the Household*, 15(4):1215–1248.

Eibich, P. (2015). Understanding the effect of retirement on health: mechanisms and heterogeneity. *Journal of health economics*, 43:1–12.

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

Fang, H., Keane, M. P., and Silverman, D. (2008). Sources of advantageous selection: Evidence from the medigap insurance market. *Journal of Political Economy*, 116(2):303–350.

Fang, H. and Wu, Z. (2018). Multidimensional private information, market structure and insurance markets. *Rand Journal of Economics, forthcoming.*

Felfe, C., Lechner, M., and Steinmayr, A. (2016). Sports and child development. *PloS one*, 11(5).

Finkelstein, A. and 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. and Poterba, J. (2014). Testing for asymmetric information using unused observables in insurance markets: Evidence from the UK annuity market. *Journal of Risk and Insurance*, 81(4):709–734.

Gao, F., Powers, M. R., and Wang, J. (2009). Adverse selection or advantageous selection? Risk and underwriting in China's health-insurance market. *Insurance: Mathematics and Economics*, 44(3):505–510.

Grignon, M. (2009). An empirical investigation of heterogeneity in time preferences and smoking behaviors. *The Journal of Socio-Economics*, 38(5):739–751.

Grossman, M. (1972). On the concept of health capital and the demand for health. *Journal of Political economy*, 80(2):223–255.

Hamad, R., Rehkopf, D. H., Kuan, K. Y., and Cullen, M. R. (2016). Predicting later life health status and mortality using state-level socioeconomic characteristics in early life. *SSM-population health*, 2:269–276.

Hendel, I. and Lizzeri, A. (2003). The role of commitment in dynamic contracts: Evidence from life insurance. *The Quarterly Journal of Economics*, 118(1):299–328.

Hsiao, C. (2014). *Analysis of panel data*. Number 54. Cambridge university press.

Idler, E. and Benyamini, Y. (1997). Self-rated health and mortality: a review of twenty-seven community studies. *Journal of Health and Social Behavior*, pages 21–37.

Johar, M., Jones, G., Keane, M., Savage, E., and Stavrunova, O. (2013). The demand for private health insurance: Do waiting lists matter?–revisited. Technical report, Economics Group, Nuffield College, University of Oxford.

Karl, J. B. (2014). A discussion of private health insurance markets in 10 OECD countries. *Journal of Insurance Regulation*, 33:41.

Kesternich, I. and Schumacher, H. (2014). On the use of information in oligopolistic insurance markets. *Journal of Risk and Insurance*, 81(1):159–175.

Kiil, A. (2012). What characterises the privately insured in universal health care systems? A review of the empirical evidence. *Health Policy*, 106(1):60–75.

King, D. and Mossialos, E. (2005). The determinants of private medical insurance prevalence in England, 1997–2000. *Health Services Research*, 40(1):195–212.

Markar, H. and O'Sullivan, G. (2012). *Medical Management: A Practical Guide*. CRC Press.

Newhouse, J. P. (2004). *Pricing the priceless: a health care conundrum*.

Newton, J. N., Briggs, A. D., Murray, C. J., Dicker, D., Foreman, K. J., Wang, H., Naghavi, M., Forouzanfar, M. H., Ohno, S. L., Barber, R. M., et al. (2015). Changes in health in England, with analysis by english regions and areas of deprivation, 1990–2013: a systematic analysis for the global burden of disease study 2013. *The Lancet*, 386(10010):2257–2274.

Olivella, P. and Vera-Hernández, M. (2013). Testing for asymmetric information in private health insurance. *The Economic Journal*, 123(567):96–130.

Olsson, T., Barcellos, L. F., and Alfredsson, L. (2017). Interactions between genetic, lifestyle and environmental risk factors for multiple sclerosis. *Nature Reviews Neurology*, 13(1):25–36.

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

Qi, Q., Chu, A. Y., Kang, J. H., Huang, J., Rose, L. M., Jensen, M. K., Liang, L., Curhan, G. C., Pasquale, L. R., Wiggs, J. L., et al. (2014). Fried food consumption, genetic risk, and body mass index: gene-diet interaction analysis in three US cohort studies. *BMJ*, 348:g1610.

Rothschild, M. and Stiglitz, J. (1976). Equilibrium in competitive insurance markets: An essay on the economics of imperfect information. *The Quarterly Journal of Economics*, 90(4):629–649.

Ruhm, C. J. (2015). Recessions, healthy no more? *Journal of health economics*, 42:17–28.

Schnittker, J. and Bacak, V. (2014). The increasing predictive validity of self-rated health. *PloS one*, 9(1):e84933.

Urbanos-Garrido, R. M. and Lopez-Valcarcel, B. G. (2015). The influence of the economic crisis on the association between unemployment and health: an empirical analysis for spain. *The European Journal of Health Economics*, 16(2):175–184.

Wallis, G. (2004). The determinants of demand for private medical insurance: evidence from the British Household Panel Survey. `http://repec.org/res2004/Wallis.pdf`.

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

A1: Data description

Variable	Description
PHI	owner of private health insurance
HR	2 lowest health categories based on self assessed health
smoke_now	current smoker
smoke_past	past smoker
female	women
age	actual age of respondent
occupation	retired, unemployed and categories from National Statistics Socio-economic Classification
couple	married or cohabit
children	respondent has children
CHH1-6	Childhood health status: Excellent, very good, good, fair, poor, varied a lot

A2: Estimates standard approaches

	(1a)	(1b)	(2a)	(2b)
Approach	Finkelstein and McGarry (2006)		Fang et al. (2008)	
Dep. var.	HR	PHI	PHI	PHI
HR			-0.040***	-0.042***
			(0.005)	(0.005)
time avail	-0.071***	-0.018***		-0.021***
	(0.008)	(0.006)		(0.006)
female	-0.036***	0.014*	0.012*	0.012*
	(0.009)	(0.007)	(0.007)	(0.007)
age	0.003***	-0.002***	-0.002***	-0.001***
	(0.001)	(0.000)	(0.000)	(0.000)
couple	-0.073***	0.042***	0.039***	0.038***
	(0.010)	(0.007)	(0.007)	(0.007)
children	0.031**	-0.031***	-0.031***	-0.030***
	(0.013)	(0.011)	(0.011)	(0.011)
unemployed	0.166***	-0.002	0.006	0.005
	(0.012)	(0.008)	(0.008)	(0.008)
managerial	-0.173***	0.048***	0.045***	0.041***
	(0.012)	(0.014)	(0.014)	(0.014)
intermediate	-0.152***	0.053**	0.050**	0.046**
	(0.015)	(0.020)	(0.021)	(0.021)
small employer	-0.153***	0.054***	0.051***	0.047***
	(0.014)	(0.017)	(0.017)	(0.017)
lower supervisory	-0.093***	-0.056***	-0.057***	-0.060***
	(0.024)	(0.017)	(0.017)	(0.017)
semi-routine	-0.101***	-0.037***	-0.039***	-0.041***
	(0.013)	(0.010)	(0.010)	(0.010)
smoke_now	0.173***	-0.060***	-0.054***	-0.053***
	(0.013)	(0.009)	(0.009)	(0.009)
smoke_past	0.052***	-0.013*	-0.011	-0.011
	(0.009)	(0.008)	(0.008)	(0.008)
_cons	0.238***	0.176***	0.172***	0.186***
	(0.045)	(0.035)	(0.035)	(0.035)
N	31492	31492	31492	31492

Notes: All columns show coefficients from a Linear Probability Model. Columns 1a and b are estimates from the approach of Finkelstein and McGarry (2006), whereas columns 2a/b reflect the coefficients based on Fang et al. (2008). Regional dummies are included and standard errors are clustered at the individual level; standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01

A3: Robustness checks

Spec.	N	(1) Finkelstein and McGarry	(2)	(3) Fang et al.	(4)	(5)	(6) RC-Model	(7)	(8)
Dependent variable		HR	PHI	PHI	PHI	PHI	HR	PHI	–
Parameter		$\hat{\beta}$	$\hat{\delta}$	$\hat{\alpha}_2$	$\hat{\alpha}_2 - \hat{\gamma}_2$	$\hat{\gamma}_3$	$\mathbb{E}\left(\hat{\beta}_i\right)$	$\mathbb{E}\left(\hat{\delta}_i\right)$	$\mathbb{E}\left(\hat{\beta}_i\hat{\delta}_i\right)$
1–time	31,492	-0.070***	-0.018***	-0.040***	0.001	-0.021***	-0.056***	-0.018***	-0.002***
		(0.008)	(0.006)	(0.005)	[0.810]	(0.006)	(0.008)	(0.006)	(0.001)
2–Probit	31,492	-0.070***	-0.015***						
		(0.007)	(0.006)						
3–other cutoff	31,492	-0.058***	-0.018***	-0.031***	0.001	-0.020***	-0.044***	-0.018***	-0.002***
		(0.008)	(0.006)	(0.006)	[0.876]	(0.006)	(0.008)	(0.006)	(0.001)
4–obj. HR	25,104	-0.062***	-0.015**	-0.011*	0.001	-0.015**	-0.049***	-0.014**	-0.001
		(0.009)	(0.007)	(0.006)	[0.908]	(0.007)	(0.009)	(0.006)	(0.001)

Notes: Each row reports results from one specification. The columns of each specification show results of unused characteristics approaches. Estimates resulting in columns 1 to 5 are based on (Linear) Probability Models. Columns 1 and 2 are estimates from the approach of Finkelstein and McGarry (2006) represented by the structural equations (3) and (4), whereas columns 3 to 5 reflect the estimates based on the structural equations (5) and (6) suggested by Fang et al. (2008). Coefficients of other 'used characteristics' are not shown above. An F-test is used to test whether the coefficient of R (based on the specification in column 3) equals the coefficient of R after the inclusion of z into the model. Columns 6 to 7 show the coefficients from a multilevel model where the explanatory variable 'availability of time' is supposed to have a random coefficient and the effect of all other variables is supposed to be fixed. Column 8 contains $\mathbb{E}\left(\hat{\beta}_i\hat{\delta}_i\right)$, the degree of selection associated with z. Specification 2 shows marginal effects of a Probit model, based on specification 1. Standard errors clustered at the individual level; standard errors in (), p-values in []. * p<0.10, ** p<0.05, *** p<0.01

A4: Robustness checks 2

Spec.	N	(1) Finkelstein and McGarry	(2)	(3) Fang et al.	(4)	(5)	(6) RC-Model	(7)	(8)
Dependent variable		HR	PHI	PHI	PHI	PHI	HR	PHI	–
Parameter		$\hat{\beta}$	$\hat{\delta}$	$\hat{\alpha}_2$	$\hat{\alpha}_2 - \hat{\gamma}_2$	$\hat{\gamma}_3$	$\mathbb{E}\left(\hat{\beta}_i\right)$	$\mathbb{E}\left(\hat{\delta}_i\right)$	$\mathbb{E}\left(\hat{\beta}_i\hat{\delta}_i\right)$
1–time	31040	-0.071***	-0.018***	-0.040***	0.001	-0.021***	-0.057***	-0.018***	-0.002***
		(0.008)	(0.006)	(0.006)	[0.813]	(0.006)	(0.008)	(0.006)	(0.001)
2–+income	31040	-0.071***	-0.016***	-0.036***	0.001	-0.019***	-0.057***	-0.017***	-0.002***
		(0.008)	(0.006)	(0.006)	[0.827]	(0.006)	(0.008)	(0.006)	(0.001)

Notes: The columns of each specification show results of unused characteristics approaches. Estimates resulting in columns 1 to 5 are based on Linear Probability Models. Columns 1 and 2 are estimates from the approach of Finkelstein and McGarry (2006) represented by the structural equations (3) and (4), whereas columns 3 to 5 reflect the estimates based on the structural equations (5) and (6) suggested by Fang et al. (2008). Coefficients of variable 'income' are shown in the second row of specification 2. An F-test is used to test whether the coefficient of R (based on the specification in column 3) equals the coefficient of R after the inclusion of z into the model. Columns 6 to 7 show the coefficients from a multilevel model where the explanatory variable 'availability of time' is supposed to have a random coefficient and the effect of all other variables is supposed to be fixed. Column 8 contains $\mathbb{E}\left(\hat{\beta}_i\hat{\delta}_i\right)$, the degree of selection associated with z. Standard errors clustered at the individual level; standard errors in (), p-values in []. * p<0.10, ** p<0.05, *** p<0.01