# Economics Project: Factors affecting survivorship rates from respondents to the English longditudinal survey 2002-16

April 22, 2017

## Contents

# 1 Introduction

question being asked, background and why this is interesting

## 1.1 Background

The demographics of the UK is changing. We have a population that is living longer and the life expectatncy for men and women is predicted to be 93.5 and 96.5 by 2030. [changing population]Exploring the causes of mortality not only allows policy to be formed that more efficiently manages resources for this ageing population but also allows education of individuals to choose lifestyle choices that will reduce their risk of early death. In a developed country such as the UK, mortality rates at younger ages is very low so the most effective improvements in mortality come from focusing on the older population.

The Engish Longditudinal Study of Ageing is a study started in 2002 focusing on following a group of approximately 10,000 individuals over 50 years old. These individuals were in private residences.The questionaire that explores demographics, economic data( House hold income, Wealth, Pension, Employment, Consumption, Expectations (mortality, employment, income adequancy)) and health data. There has been a vast amount of research into the link between smoking and premature death. This survey also wants to see what life was like for individuals before their deaths. So a lot of the health data is concerned with how good the quality of life is for people prior to their death. This includes living with certain conditions, how ability to walk, wash and clothe oneself impacts of mortality. And how mental health affects.

Health sections include, general health, Health behaviours (smoking), mental health, Mobility assement, blood analys, cognitive fucntions

Also included was psychosocial (tv watching etc) included in the study.

The information collected provides data about, household and individual demographics, Health pyhsical and psychosocial, work and pensions, income and assets, housing, cognitive function, social participation, effor and reward(voluntary work and caring), Expectations and walking speed.

## 1.2 Questions being asked

I'd like to see what kind of socio economic effects are heavily linked with mortality. I think things such as income, savings, smoker status, social class, tv

watching, alchol, physical activity are related to mortality. My hypothesis is that percieved unhealthy behaviours such as tv watching, smoking, alochol wil be negatively correlated with mortality and percieved positive behaviours such as physical activity, income, savings and social class will the positively correlated with mortality. Such things as age and sex are already widely regarded as being negatively correlated with mortality and regards sex, women living longer than men so I'm trying to dig deeper than that for this project.

Rference for introduction : http://www.macs.hw.ac.uk/~iain/research/Currie.SM.pdf

### 1.3 Why is this interesting

If we know the causes of mortality policy makers can more efficiently allocate resources and also disseminate information to allow people to make more informed choices of their lifestyle.

## 2 Theory

link to literature, setup model & hypothesis, Q's to answer

heterogenity - unintended variation across groups/participants

The idea of linking mortality to socio economic factors is not new. The ELSA is a very rich form of data and.

Fe analysis is used when interested in the impact of variables over time. FE is interesting when we want to explore the relationship between independent and dependent variables within an entity(group, country etc). We assume that something within the group may bias the independent or dependent variables and we need to control this. We assume there is a correlation between the error terms and dependent variables which is why we use FE. Fe removes the effect so we can assess net effect of indep on dep variables.

Also those group characteristics are not correlated with another groups characteristics. If they are correlated we can detect with the Hausman test. This leads us to model the behaviour using random effects model instead.

Fixed Effects model does not work well for slow changing variables or where there is minimal within group variation.

This may be the model I use

"You could add time effects to the entity effects model to have a time and entity fixed effects regression model: Yit = $\beta 0 + \beta 1X1,it + ... + \beta kXk,it + \gamma 2E2 + ... + \gamma nEn + \delta 2T2 + ... + \delta tTt + uit$ [eq.3] Where –Yit is the dependent variable (DV) where i = entity and t = time. –Xk,it represents independent variables (IV), –$\beta$k is the coefficient for the IVs, –uit is the error term –En is the entity n. Since they are binary (dummies) you have n-1 entities included in the model. –$\gamma$2 is the coefficient for the binary regressors (entities) . –Tt is time as binary variable (dummy), so we have t-1 time periods. –$\delta$ t is the coefficient for the binary time regressors ." - from Panel 101.pdf

We will use least squares dummy variable model (LSDV) -explain

The effect of smokerate is medidated by the differences across couuntries

By adding the dummy for each country we are estimating the pure effect of smokerate by controling for the unobserved heterogeneity

Each dummy is absorbing the effects particular to each country

FE model control for all time-invariant differences between groups so estimated coefficients cannot be biased by time-invariabnt charactertistcs.

Random effects model the variabtion is random and uncorrelated with dep or indep variables in model. If we believe difference across entities have some influence on dep variables the use random effects

### 2.1 Reasons not to use OLS

Taken from a paper so reword"An application I will walk you through many applications of fixed effect models. Here is one to ponder. There have recently been high profile studies of the relationship between staffing in hospitals and patient outcomes. These studies use traditional OLS regression where the unit of observation is the patient, the dependent variable is some outcome measure like mortality (a dummy variable that equals 1 if the patient died in the hospital) and the key predictor is staffing (e.g., nurses per patient). These studies do not use fixed effects, and invariably show that hospitals with more staff have better patient outcomes. These results have had enormous policy implications. Unfortunately, these studies may suffer from omitted variable bias. For example, the key unobservable variable might be the severity of patients' illnesses, which is notoriously difficult to control for with available data. Severity is likely to be correlated with both mortality and staffing, so that the coefficient on staffing will be biased. (What is the direction of the bias?) If you ran a hospital fixed-effects model, you would include hospital dummies in the regression that would control for observable and unobservable differences in severity (and all other factors) across hospitals. This would greatly reduce potential omitted variable bias. None of the current research in this field has done so, perhaps because there is not enough intrahospital variation in staffing to allow for fixed-effects estimation. Even a fixed effects model would not completely eliminate potential omitted variable bias. You must also hope that any changes over time in unobservable patient severity within each hospital are uncorrelated with changes over time in staffing. This might not be such a good assumption. As hospitals experience increases in severity, they may increase staffing. If so, then unobservable severity within a hospital is correlated with staffing and the omitted variable bias is still present. "

## 3 Data

### 3.1 Describe the survey in language

All analaysis done using STATA 13

The English Longitudinal Study of Ageing (ELSA) started in 2002. ELSA is a collaboration between Univeristy College London (UCL), the institute for

Fiscal Sutdies (IFS) and NatCen Social Research. It is a large scale longditudianl panel study of men and women aged 50 and over and their partners. They are living in private households in England. The orginal (Wave 0) participants were drawn from the Health Survey for England (HSE). Approximately 10,000 people are studies in each wave with the sample being refreshed in wave 3,4,6 and 7 due to people dropping out and death. Respondents were intervieved at approximaltey 2 yearly interviews, Wave1- March 2002 to March 2003, Wave2 -June 2004 - July 2005, wave3- May 2006-August 2007, Wave 4 - May 2008-July 2009, Wave 5 - June 2010-July 2011, Wave 6 May 2012 - June 2013, Wave 7 June 2014 - May 2015.

We are concerned with mortality data which has only been updated until Wave 6 as of this time. So we will concern ourselves with individuals who responded in the first wave and every wave up until their death. So there are not gaps in the data set. (Insert number here). When respondent die they can obviously no longer give responses to the questions in the data set. As the sample size is quite large we simply eliminate them from the group. (Maybe talk about possible biases to this)

The areas of interest and questions have been modified slightly throughout the lifecycle of the study. We will be focusing on questions that have been asked in every wave and who's question has not been modified significantly throughout time.

As Wave 1 was the first full questionaaire we can only deal with mortality after Wave 1. So our time periods are mortality prior to Wave 2, 3, 4, 5, 6. Which gives us 5 time series points corresponding to 2005, 2007, 2009, 2011, 2013 approximately.

## 3.2 Describe the data in numbers and graphs

General mortality numbers occuring after wave 1(% ages and counts)

Partnership status, marital status

Splitting up the mortality numbers by quintile on income/savings

Qualitfications (or maybe social status)

men/ women

## 3.3 Basic properties, graph, trends, cycles and other

## 3.4 Cycle of people more likely to die in december, if we have this data?

There tends to be a higher likelihood for people to die in december. This is quite a strong seasonal cycle. Even though the Uk is a temporate climate, poverty and ill health can lead to people not being able to heat their homes adequately in the winter which may lead to this spike. There's some evidence (find some evidence) that the winter fuel allowance should be increased to help alleviate this problem. But this is beyond the scope of this analysis.

Include graph of seasonal deaths since wave 1

# 4 Statistical Model

## 4.1 Estiamte methods to use & why appropriate

## 4.2 Asuumptions about errors

## 4.3 OLS but justify choice of estimator

We are going to use something like a Fixed Estimate model or a Random Effects model.

We will determine which one to use with the Hausman test.

Including dummy variables for each of the waves.

talk about ommitted variable bias

Identifying assumption: Unobservable factors that might simultaneously affect the LHS and RHS of the regression are time-invariant.

The theory behind fixed effects regression

If we believe the regression has the same shape within each of these groups over time, we can estimate the average of the regressions. The way fixed effects regressions handle this is to calculate the changes wrt to the mean of those changes within the group. So we want to know whether variations in mortality are related to variations in our effects.

We are trying to estimate coefficients of regression for a number of factors (listed above) but across a wide range of different groups. To eliminate some of these differences between the groups

Need to include dummy for each wave (ts component) and each group (so they can differ)

need to include trend component

Stata will automatically create dummies

This is known as fixed effects regression becuase it holds constant (fixes) the average effects of each group.

http://www.jblumenstock.com/files/courses/econ174/FEModels.pdf

# 5 Results

## 5.1 Big important results first

## 5.2 Metntion briefly the results that proved insignificant

## 5.3 Include coefficients, std errors, std error regression & diagnostic tests

## 5.4 Explain results

## 5.5 Consider size of coefficients & their significance

# 6 Conclusions

## 6.1 What did we learn?

## 6.2 How were questions answered?

## 6.3 Relevance for practical Q's of forecasting or policy.

## 6.4 & answers consitent with theory

## 6.5 Is model well specified

# 7 Apendicies

How to run FE in stata.
set panels and year with: xtset country year
if country is string: encode country, gen(country1)
xtline y
gen graph (heterogeneity across acountrie)
bysort country: egen y_mean=mean(y) twoway scatter y country, msymbol(circle_hollow) || connected y_mean country, msymbol(diamond) || , xlabel(1 "A" 2 "B" 3 "C" 4 "D" 5 "E" 6 "F" 7 "G")
Heretrogeneity across years
bysort year: egen y_mean1=mean(y) twoway scatter y year, msymbol(circle_hollow) || connected y_mean1 year, msymbol(diamond) || , xlabel(1990(1)1999)

# References


[changing population] http://visual.ons.gov.uk/uk-perspectives-the-changing-population/

[maybeuseful] https://www.cambridge.org/core/services/aop-cambridge-core/content/view/0334A27557D15848549120FE8ECD8D63/S204984701

[1]