2899B

PART IIA PAPER 3 PROJECT

1. Cross-country Variations in Health Outcomes in the European Union 2018

Health care outcomes show wide variation across the Member States of the European Union (OECD 2018, 2019).

Using the OECD Health database and related publications, together with any other data and information that you wish to introduce:

a) Identify and explain the critical factors affecting health status outcomes across Europe.

Health status can be taken to include:

- Life expectancy.
- Avoidable mortality.
- Chronic disease morbidity.
- Self-rated health.

b) On the basis of your analysis, evaluate and examine the contribution to cross-national variations in health status of:

- Risk factors.
- Access to care.
- Quality of care.
- Health resources.

c) In the light of your findings for Parts a) and b), consider critically what policies are appropriate for:

- An overall improvement in health status in the EU.
- Reducing cross-country inequalities in healthcare outcomes in the EU.

(1998 words)

I Introduction

EU health outcomes improved across the last 5 decades. Life expectancy rose from 71 years in 1970 to 81 in 2018. However, these improvements have not been evenly distributed.

Within countries, between countries and in the EU over time there has been a large variation in health outcomes, the causes of which are not clear. Even the impact of the most plausible contributor to health outcomes, health expenditure, has been debated.

Carr-Hill, Hardman and Russell (1987) found a positive contemporaneous relationship between healthcare spending and avoidable mortality. Machenbach (1991) then found no relationship between healthcare spending and avoidable mortality across countries, concluding differences in efficiency were key.

This paper assesses key contributing factors to health outcome trends and variation by building on Heijink et al. (2013) and the working paper by James et al. (2017). I include more countries and cover more time than Heijink et al., whilst avoiding the potential pitfalls of linear data interpolation that James et al. could have run into.

II Theoretical Framework

I use a macroeconomic production function modelling the impact of various factors on health status outcomes. The population model is given by:

$$Y_{i,t} = \alpha_i + \beta_x \boldsymbol{X_{i,t}} + \varepsilon_{i,t}$$

$Y_{i,t}$ health outcome

α_i country-specific effect (time-invariant factors)

$\boldsymbol{X_{i,t}}$ vector of explanatory variables listed in Data section

$\varepsilon_{i,t}$ error

Some variables (including both dependent variables) were transformed logarithmically if it reduced the level of heteroskedasticity. The co-efficients can be interpreted as elasticities or semi-elasticities.

Country-specific effects are difficult to observe, which is problematic if they correlate with explanatory variables. In this case fixed effects or first differencing the population model removes country-specific effects:

$$\Delta \ln (Y_{i,t}) = \beta_x \Delta(\boldsymbol{X_{i,t}}) + \Delta\varepsilon_{i,t}$$

III Data

I assess 35 countries in the OECD, 23 of which are in the EU, from 2000-2016. Non-EU countries are similarly developed to the EU – e.g. Australia, not Rwanda. Observations increased to 499 across 35 cross-sectional units.

Lack of data availability across the period results in missing data. Some previous papers have estimated missing values using basic linear interpolation (and extrapolation). This simple method could lead to flawed conclusions being reached from made-up data, so was avoided. A more sophisticated method would be to use multiple imputation, but due to time constraints and sufficiently good data this was not done.

Variable	Type	Units	Justification	Source
Potential Years of Life Lost (PYLL)	Dep	Years per 100,000 (age standardised)	Potential years of life lost per 100,000. Sums up deaths occuring at each age and multiply by the number of remaining years to live up to Measure weights deaths amongst young people more highly than deaths related to old age, making it a better measure of health status than death rates and life expectancy. More weight should be given to a 25 year old dying in their prime than an elderly person dying mor naturally. Data is age-standardised.	OECD
Avoidable Mortality	Dep	Deaths per 100,000 (age standardised)	Avoidable mortality is the combination of preventable mortality and treatable mortality. The former measures the deaths that could hav been avoided through public health and primary interventions before the cause of death occurred. The latter measures deaths that coul have been avoided through heathcare interventions and treatment. This composite measure shows how many deaths were unnecessary ICD-10 classification used	OECD
Alcohol	Ind	Litres per 15+	Alcohol consumption. Litres of alochol sold domestically per person aged 15+. Alcohol consumption is a health risk factor, e.g. due to live cirrhosis or stroke and then death.	OECD
GDP net health expenditure	Ind	USD per cap, ppp, current prices	GDP minus healthcare expenditure, all divided by the population. Tests to see if being poorer is a risk factor even after controlling for hea spending	OECD
Health Expenditure (HE)	Ind	USD per cap, ppp, current prices	Health expenditure per capita. Low health spending is mesasure of health resources.	OECD
Hospital Beds	Ind	beds	Hospital beds per 100,000 people. Measure of both health resources and healthcare quality. If more beds are available per 100,000 then more patients can be treated at once, reducing waiting times. Specific waiting time data would lead to an unbalanced panel.	OECD
Hours worked	Ind	hours	Hours worked per employed worker. Working longer hours may lead to increased stress and less time for exercise, so could be a risk fact	OECD
Long-tern unemp	Ind	% of the unemployed	Proportion of unemployed who have been out of work for over a year. Impact on mental health could be a factor	OECD
Obesity	Ind	% of 15+ population. Age standardised	Rate of people with BMI over 30. A risk factor for many diseases that can lead to death.	WHO
Out-of-pocket HE	Ind	USD per capita, ppp adjusted, current prices	Out-of-pocket expenditure on health. A form of voluntary health expenditure where the customer pays extra to see a doctor/specialist. H prices could be indicative of poor access to healthcare.	OECD
Pollution	Ind	Nox kg per cap	Measure of pollution in country. More adverse health effects associated with PM2.5 particulate density measure, but due to poor data availability for PM2.5, nitrogen oxide data was used insead. May be a risk factor.	WHO
Alc*HE	Ind	N/A	Alcohol and health expenditure interaction term. Impact of consuming alcohol on mortality may be lower if spending more on health me you are less likely to die.	OECD
Hours*HE	Ind	N/A	Working hours and health expenditure intereaction term. Potential damage of working more on health may be offset by spending more healthcare.	OECD

Excluded Variables	Type	Justification
Smoking	Ind	Smoking is likely a risk factor, as shown by Shaw et al. (2005), but was excluded in this paper to reduce the impact of missing data
Physician Density	Ind	Variables such as physician density could have been included but could have led to misleading results. In this case the co-efficient on health expenditure would be the importa of health expenditure assuming money is not spent on more physicians.
PM2.5	Ind	My environmental measure was chosen because data for it is available for a broad set of countries and it is associated with adverse health effects. The alternative measure, PM2.5, represents a more dangerous environmental concern, but annual data for PM2.5 levels do not extend back far.
Demographics	Ind	Many variables are already age-standardised, so it would be inappropriate to include demographic variables.

Table1: Data

IV Estimation Methods and Diagnostic Tests

Estimation Method	Assumptions and Efficiency
(i) Pooled OLS (POLS)	Under Gauss-Markov assumptions, OLS is BLUE. However, if α_i varies across countries, the strict exogeneity assumption of the error term fails due to unobserved heterogeneity. The BP-LM test likely rejects the null of no difference in unobserved country-specific effects, favouring random effects. POLS biased alcohol, HE and hours worked upwards compared to FE.
(ii) Random Effects (RE)	If unobserved specific effects are uncorrelated with explanatory variables then RE is consistent and efficient, but the Hausman test likely rejects this assumption when comparing with FE, suggesting that the unobserved heterogeneity is causing endogeneity through omitted variables. RE biased GDP net HE and hours worked downwards compared to FE.
(iii) Fixed Effects (FE)	FE removes α_i from the regression so avoids this source of endogeneity. If tests for serial correlation are rejected, FE is likely more efficient that FD.
(iv) First Differences (FD)	FD also removes α_i from the regression, so choice between FE and FD depends on relative efficiency and assumption violations. In presence of high serial correlation of level errors (e.g. random walk) FD is more efficient. Test for serial correlation. FD is also more appropriate if there is a unit root due to potentially spurious regression from I(1) processes in fixed effects.

Table2: Estimation Method of Initial Population Model Overview

VARIABLES	POLS lnyrslost	RE lnyrslost	FE lnyrslost	FD D.lnyrslost	POLS lnavmort	RE lnavmort	FE lnavmort	FD D.lnavmort
alc	**0.291*** **	**0.056** **	0.047		**0.358*** **	0.010	0.008	
	(0.037)	(0.029)	(0.029)		(0.057)	(0.029)	(0.028)	
lnhe	**2.322***	0.948	**1.089***		1.691	0.394	0.219	
	(1.264)	(0.638)	(0.645)		(1.929)	(0.635)	(0.617)	
lnhospbeds	**0.242*** **	0.015	-0.025		**0.149*** **	0.004	-0.034	
	(0.023)	(0.027)	(0.028)		(0.035)	(0.031)	(0.030)	
lngdpnethe	**-0.450*** **	**-0.373*** **	**-0.350*** **		**-0.234*** **	**-0.237*** **	**-0.226*** **	
	(0.044)	(0.031)	(0.031)		(0.066)	(0.030)	(0.029)	
lnhrs	**2.226***	**1.529** **	**1.741*** **		1.136	**1.135***	**1.070***	
	(1.346)	(0.654)	(0.650)		(2.057)	(0.643)	(0.621)	
lru	**-0.004*** **	**-0.002*** **	**-0.002*** **		**0.002***	**-0.001*** **	**-0.001*** **	
	(0.001)	(0.000)	(0.000)		(0.001)	(0.000)	(0.000)	
obese	**0.026*** **	**0.006** **	0.003		**0.021*** **	0.004	0.005	
	(0.002)	(0.003)	(0.003)		(0.002)	(0.003)	(0.003)	
lnoop	**0.063*** **	**0.096*** **	**0.076*** **		0.009	**0.058*** **	**0.048** **	
	(0.022)	(0.022)	(0.023)		(0.033)	(0.022)	(0.021)	
lnpol	0.005	**0.115*** **	**0.140*** **		**-0.041***	**0.143*** **	**0.174*** **	
	(0.015)	(0.020)	(0.023)		(0.023)	(0.021)	(0.022)	
alc_lnhe	**-0.032*** **	-0.003	-0.002		**-0.035*** **	0.003	0.002	
	(0.005)	(0.004)	(0.004)		(0.007)	(0.004)	(0.004)	
lnhe_lnhrs	**-0.300***	**-0.151***	**-0.166***		-0.244	-0.092	-0.066	
	(0.169)	(0.086)	(0.087)		(0.257)	(0.085)	(0.083)	
D.alc				**0.112*** **				**0.064** **
				(0.032)				(0.032)
D.lnhe				1.742				**2.441** **
				(1.070)				(1.059)
D.lnhospbeds				-0.025				-0.024
				(0.035)				(0.034)
D.lngdpnethe				0.035				0.047
				(0.037)				(0.037)
D.lnhrs				1.572				**2.185** **
				(1.089)				(1.076)
D.lru				**-0.001***				-0.000
				(0.000)				(0.000)
D.obese				**0.025** **				**0.036*** **
				(0.010)				(0.010)
D.lnoop				0.004				0.011
				(0.021)				(0.022)
D.lnpol				0.043				0.042
				(0.030)				(0.030)
D.alc_lnhe				**-0.015*** **				**-0.009** **
				(0.004)				(0.004)
D.lnhe_lnhrs				-0.216				**-0.314** **
				(0.143)				(0.142)
Constant	-5.635	0.980	-0.792	**-0.035*** **	-1.206	0.367	0.739	**-0.043*** **
	(9.982)	(4.882)	(4.862)	(0.005)	(15.307)	(4.821)	(4.656)	(0.005)
Observations	499	499	499	455	472	472	472	429
R-squared	0.841		0.887	0.065	0.818		0.903	0.066
Number of c_id		35	35			34	34	

Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Based on initial population model. Not the final results, which are adjusted based on diagnostic tests.

Table3: Estimation Method Regressions

Dep Var	Test	Model(s)	Null Hypothesis	Test Stat, P value	Justification	Action
PYLL	BP LM test	OLS + RE	No random effects	chibar2: 2244; p<0.0001	Tests for country-specific effect. Null rejected.	RE over POLS
PYLL	Hausman test	RE + FE	Same co-efficients	Chi2: 129.8; p<0.0001	If country-specific effects are uncorrelated with the explanatory variables, the co-efficients would be the same. However, the null is rejected, so they are not.	FE over RE
PYLL	Wooldridge test	FE	No first order autocorrelation	F stat: 0.491, p: 0.4882	Tests for serial correlation. Null not rejected. Will proceed without correcting for serial correlation. This suggests FE will be more efficient than FD, as no serial correlation in level errors implies negatively correlated first difference of errors.	FE over FD
PYLL	Modified Wald test	FE	Homoskedasticity	chi2: 4095; p<0.0001	Test for heteroskedasticity. No complete test for heteroskedasticity on panel data, but unlikely for homoskedasticity to exist. Modified Wald test can be used to test for groupwise heteroskedasticity. This tests to see if the variance of residuals is not constant across the cross section. The null is rejected, signalling heteroskedasticity.	robust SEs
PYLL	Breusch Pagan homoskedasticity	FD	Homoskedasticity	chi2: 49.86; p<0.0001	Test for heteroskedasticity. Rejects null of constant variance.	robust SEs
PYLL	Pesaran CD test (2004)	FE + FD	No cross sectional dependence	t: 4.727 ; p<0.0000	Test for cross-sectional dependence. Pesaran CD works for unbalanced data so is appropriate. Test rejects null, so will use Driscoll-Kraay standard errors that are already robust to most forms of heteroskedasticity (Hoechle, 2007) instead of regular robust standard errors. Spatial dependence does not affect consistency, but does lead to biased standard errors which reduce the validity of statistical inference if used. Driscoll-Kray SEs fix this problem.	D-K SEs
PYLL	Fisher-type test	FE	Panels contain unit roots	Inverse chi-sqd: 138.3, inverse normal: - 4.65, inverse logit: - 4.62, mod inverse chi-sqd: 5.77; all p<0.001	Test for unit root and stationarity. Drift added. Variables demeaned to reduce impact of cross-sectional dependence on test result, as recommended by Levin, Lin and Chu (2002). Null of unit root rejected, suggesting processes are stationary. FE analysis is still applicable.	Use FE
PYLL	Skewness Kurtosis test	FE	Normally distributed errors	adj chi2: 12.59 ; p: 0.0018	Test assumption of normality of errors. Null hypothesis rejected, indicating non-normal errors. Under CLT, this assumption is no longer required	Invoke CLT
PYLL	Skewness Kurtosis test	FD	Normally distributed errors	adj chi2: 33.65 ; p<0.0001	Test assumption of normality of errors. Null hypothesis rejected, indicating non-normal errors. Under CLT, this assumption is no longer required	Invoke CLT
PYLL	modified Ramsey RESET	FE	No misspecification	F=5.56, p=0.0041	Ramsey RESET not appropriate for FE or FD analysis due to unimportance of time-constant variables. Modified version (recommended by Wooldridge 2020) used. Null rejected, and F-value reduces when interaction terms are removed. However, given the plausible justification of the interaction terms, I will persist with them in the model. Added squared obesity term	Proceed

Table 4: Potential Years of Life Lost (PYLL) Diagnostic Tests

Dep Var	Test	Model(s)	Null Hypothesis	Test Stat, P value	Justification	Action
AvMort	Breusch Pagan LM test	OLS + RE	No random effects	chibar2: 2386; p<0.0001	Tests for country-specific effect. Null rejected.	RE over POLS
AvMort	Hausman test	RE + FE	No significant difference in co-efficients	Chi2: 103.9; p<0.0001	Tests difference in co-efficients between RE and FE. If country-specific effects are uncorrelated with the explanatory variables, the co-efficients would be the same. However, the null is rejected, so they are not.	FE over RE
AvMort	Wooldridge test	FE	No first order autocorrelation	F stat: 36.57, p<0.0001	Tests for serial correlation. Null rejected so autocorrelation robust standard errors will be used. This suggests FD may be more efficient than FE.	Do FE and FD
AvMort	Modified Wald test	FE	Homoskedasticity	chi2: 1864; p<0.0001	Test for heteroskedasticity. No complete test for heteroskedasticity on panel data, but unlikely for homoskedasticity to exist. Modified Wald test can be used to test for groupwise heteroskedasticity. This tests to see if the variance of residuals is not constant across the cross section. The null is rejected, signalling heteroskedasticity.	robust SEs
AvMort	Breusch Pagan heteroskedasticity	FD	Homoskedasticity	chi2: 44.51; p<0.0001	Test for heteroskedasticity. Rejects null of constant variance. Procede with robust SEs	robust SEs
AvMort	Pesaran CD test (2004)	FE + FD	No cross sectional dependence	t: 3.57 ; p: 0.0004	Test for cross-sectional dependence. Pesaran CD works for unbalanced data so is appropriate. Test rejects null, so will use Driscoll-Kraay standard errors that are already robust to most forms of heteroskedasticity (Hoechle, 2007) instead of regular robust standard errors. Spatial dependence does not affect consistency, but does lead to biased standard errors which reduce the validity of statistical inference if used. Driscoll-Kray SEs fix this problem.	Use D-K SEs
AvMort	Fisher-type test	FE	Panels contain unit roots	Inverse chi-sqd: 142.3, inverse normal: - 5.24, inverse logit: - 5.28, mod inverse chi-sqd: 6.64; all p<0.001	Test for unit root and stationarity. Drift added. Variables demeaned to reduce impact of cross-sectional dependence on test result, as recommended by Levin, Lin and Chu (2002). Null of unit root rejected, suggesting processes are stationary. FE analysis is still applicable.	Use FE and FD
AvMort	Skewness Kurtosis test	FE	Normally distributed errors	adj chi2: 49.52 ; p<0.0001	Test assumption of normality of errors. Null hypothesis rejected, indicating non-normal errors. Under CLT, this assumption is no longer required	Invoke CLT
AvMort	Skewness Kurtosis test	FD	Normally distributed errors	adj chi2: 36.3 ; p<0.0001	Test assumption of normality of errors. Null hypothesis rejected, indicating non-normal errors. Under CLT, this assumption is no longer required	Invoke CLT
AvMort	modified Ramsey RESET	FE	No misspecification	F=1.53, p=0.2177	Ramsey RESET not appropriate for FE or FD analysis due to unimportance of time-constant variables. Modified version (recommended by Wooldridge 2020) used. Null not rejected	Proceed
AvMort	standard Ramsey RESET	FD	No misspecification	F=0.28, p=0.8371	Ramsey RESET appropriate for FD model testing. Null not rejected.	Proceed

Table 5: Avoidable Mortality Diagnostic Tests

V Results Explained

VARIABLES	FE PYLL	FE + lags PYLL	FE+ lags, trend PYLL	FE + lags, trend, D-K errors PYLL	FD + lags PYLL
Alcohol	0.096	**0.151****	**0.171****	**0.171*****	
	(0.057)	(0.072)	(0.064)	(0.028)	
Health Expenditure	**4.971*****	**5.203*****	**5.504*****	**5.504*****	
	(1.380)	(1.575)	(1.482)	(0.463)	
Hospital Beds	-0.009	0.000	0.059	**0.059***	
	(0.068)	(0.081)	(0.051)	(0.031)	
GDP net HE	**-0.248*****	**-0.207*****	**-0.143***	**-0.143*****	
	(0.057)	(0.063)	(0.071)	(0.034)	
Hours Worked	**5.144*****	**5.238*****	**5.172*****	**5.172*****	
	(1.454)	(1.634)	(1.519)	(0.412)	
Long-term unemployment	**-0.001****	**-0.001****	**-0.001****	**-0.001*****	
	(0.001)	(0.001)	(0.001)	(0.000)	
Obesity	**-0.100*****	**-0.095*****	**-0.050***	**-0.050*****	
	(0.018)	(0.023)	(0.025)	(0.009)	
Obesity-sq	**0.002*****	**0.002*****	**0.001*****	**0.001*****	
	(0.000)	(0.000)	(0.000)	(0.000)	
OOP Health Expenditure	**0.093****	**0.075****	**0.060***	**0.060*****	
	(0.040)	(0.033)	(0.030)	(0.014)	
Pollution	**0.102****	**0.103****	0.007	0.007	
	(0.038)	(0.044)	(0.044)	(0.018)	
Alcohol * HE	-0.011	**-0.018***	**-0.021****	**-0.021*****	
	(0.008)	(0.010)	(0.009)	(0.004)	
HE * hours worked	**-0.671*****	**-0.679*****	**-0.703*****	**-0.703*****	
	(0.184)	(0.214)	(0.197)	(0.062)	
L.HE		-0.055	-0.042	-0.042	
		(0.068)	(0.074)	(0.036)	
L2.HE		**-0.079***	-0.043	-0.043	
		(0.039)	(0.048)	(0.038)	
D.Alcohol					**0.172*****
					(0.051)
D.Health Expenditure					1.967
					(1.694)
D.Hospital Beds					-0.038
					(0.036)
D.GDP net HE					0.028
					(0.045)
D.Hours Worked					1.617
					(1.761)
D.Long-term unemployment					-0.001
					(0.000)
D.Obesity					**-0.065****
					(0.027)
D.Obesity-sq					**0.001*****
					(0.000)
D.OOP Health Expenditure					0.012
					(0.020)
D.Pollution					0.044
					(0.041)
D.Alcohol * HE					**-0.022*****
					(0.007)
D.(HE * hours worked)					-0.237
					(0.229)
D.L.HE					-0.030
					(0.043)
D.L2.HE					-0.022
					(0.031)
Constant	**-26.730****	**-28.132****	**-30.179****	**-30.179*****	**-0.022*****
	(10.835)	(12.173)	(11.283)	(2.866)	(0.007)
Observations	499	454	454	454	410
R-squared	0.911	0.906	0.926		0.118
Number of c_id	35	35	35		
Number of groups				35	

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table6: Potential Years of Life Lost Results

Potential Years of Life Lost

Model 3 is most appropriate as it includes a time trend which can be interpreted as exogenous shocks to the production function. It includes lags for health expenditure as intuitively there is a delay between investing in facilities and using them. The lags also reduce potential simultaneity between the outcome and HE. The lack of serial correlation explains more statistically significant findings than first differences due to greater efficiency. In order to work out elasticities, variable means were used for interaction terms and lags.

Alcohol is associated with worse health outcomes. Interacting alcohol with health expenditure shows that as people drink more, HE mitigates the dangers of alcohol consumption. Increasing alcohol consumption by 1 litre p.c. is correlated with an increase in PYLL by 0.72%.

The co-efficient on HE is positive, but its association with PYLL is reduced by interactions with alcohol and working hour. A 10% increase in health expenditure is linked to a 0.69% increase in PYLL. Health expenditure lags reverse this relationship. When jointly tested, the lags are statistically insignificant in Model 3, but in Model 4 with Driscoll-Kraay errors, they are significant at the 5% level. Increasing HE by 10% is associated with a fall in PYLL of 0.12% with lags included.

A 10% increase in GDP net health expenditure is associated with a 1.43% decrease in PYLL but is only significant at the 10% level using Stata's errors. Significant at 1% level using Driscoll-Kraay.

A 10% increase in hours worked is associated with a 3.11% fall in PYLL. The effect is due to the hours worked and HE interaction term, suggesting working longer can be healthier if you then pay for better healthcare.

As the long-term unemployment rate increases by 10 percentage points, PYLL falls by 0.01%. Unemployment is often correlated with improved health outcomes in high-level studies (James et al., 2017). This could be due to fewer traffic accidents when unemployment is

lower, and so fewer deaths. It could also be due to areas with larger unemployment (in the relatively rich countries studied in this paper at least) offering greater benefits to the unemployed and free healthcare. This simultaneously reduces the incentive to work and improves the health outcomes of the unemployed.

The impact of obesity on health outcomes is non-linear. The obesity rate usually has a negative correlation with the rate of underweight people. When obesity increases at low levels, this could be associated with a large number of people who had been malnourished now getting sufficient food to optimise health. There is a tipping point where the downsides of an increasing obesity rate overwhelm the effect of some others now getting enough food. The results suggest that when the obesity rate exceeds 25% the dangers of overfeeding on the health of the population exceed the benefits of less underfeeding.

A 10% increase in out-of-pocket HE is associated with a 0.60% increase in PYLL. This highlights the danger of assuming any statistically significant relationship is causative. OOP spending includes any payments that someone makes for specialist attention in addition to a compulsory or voluntary healthcare scheme. This is clearly not damaging to the patient's health, but it could indicate a low level of universal healthcare.

Pollution is not significant (even with Driscoll-Kraay errors). This is likely because it takes long periods of time for the impact of higher levels of NOx to have an observable impact on health. Hospital beds are only significant using Driscoll-Kraay errors at the 10% level. The time trend reduces the significance of GDP net health expenditure, OOP and pollution, but increased the significance of alcohol.

The first differenced model is likely less efficient as a result of no serial correlation, as it only found alcohol and obesity to have statistically significant results. The results align with the fixed effect model's results, which increases the reliability of the conclusions drawn.

VARIABLES	(1) FE AvMort	(2) FE + lags AvMort	(3) FE+ lags, trend AvMort	(4) FE + lags, trend, D-K SEs AvMort	(5) FD + lags AvMort
Alcohol	0.084	0.103	**0.130****	**0.130*****	
	(0.059)	(0.070)	(0.061)	(0.016)	
Health Expenditure	**5.127*****	**5.263*****	**4.861*****	**4.861*****	
	(1.437)	(1.430)	(1.309)	(0.577)	
Hospital Beds	0.004	0.014	0.035	0.035	
	(0.061)	(0.062)	(0.047)	(0.039)	
GDP net HE	-0.097**	-0.086*	-0.017	-0.017	
	(0.047)	(0.048)	(0.060)	(0.029)	
Hours Worked	**5.426*****	**5.490*****	**4.711*****	**4.711*****	
	(1.490)	(1.527)	(1.369)	(0.625)	
Long-term unemployment	**-0.001***	**-0.001****	-0.001	-0.001	
	(0.001)	(0.000)	(0.001)	(0.000)	
Obesity	**-0.117*****	**-0.105*****	**-0.053****	**-0.053*****	
	(0.017)	(0.018)	(0.020)	(0.012)	
Obesity-sq	**0.002*****	**0.002*****	**0.001*****	**0.001*****	
	(0.000)	(0.000)	(0.000)	(0.000)	
OOP Health Expenditure	**0.070***	0.054	0.046	**0.046*****	
	(0.040)	(0.033)	(0.028)	(0.011)	
Pollution	**0.122*****	**0.124*****	**0.055***	**0.055*****	
	(0.031)	(0.032)	(0.029)	(0.017)	
Alcohol * HE	-0.010	-0.012	**-0.016***	**-0.016*****	
	(0.008)	(0.009)	(0.008)	(0.002)	
HE * hours worked	**-0.704*****	**-0.704*****	**-0.632*****	**-0.632*****	
	(0.191)	(0.197)	(0.175)	(0.081)	
L.HE		-0.044	-0.056	**-0.056***	
		(0.054)	(0.055)	(0.030)	
L2.HE		**-0.095****	-0.055	-0.055	
		(0.042)	(0.042)	(0.035)	
D.Alcohol					**0.093****
					(0.042)
D.Health Expenditure					**3.167*****
					(1.219)
D.Hospital Beds					-0.032
					(0.037)
D.GDP net HE					0.041
					(0.038)
D.Hours Worked					**2.778****
					(1.249)
D.Long-term unemployment					-0.000
					(0.000)
D.Obesity					**-0.066****
					(0.031)
D.Obesity-sq					**0.002*****
					(0.000)
D.OOP Health Expenditure					0.017
					(0.022)
D.Pollution					0.040
					(0.031)
D.Alcohol * HE					**-0.012****
					(0.005)
D.(HE * hours worked)					**-0.407****
					(0.164)
D.L.HE					-0.039
					(0.033)
D.L2.HE					-0.030
					(0.031)
Constant	**-32.626*****	**-33.319*****	**-30.170*****	0.000	**-0.026*****
	(11.093)	(11.096)	(10.207)	(0.000)	(0.007)
Observations	472	435	435	435	392
R-squared	0.932	0.933	0.945		0.112
Number of c_id	34	34	34		
Number of groups				34	

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table7: Avoidable Mortality Results

Avoidable Mortality

The avoidable mortality results are similar to PYLL. To avoid repetition, I focus on the differences. One major difference is that this model displays some serial correlation, leading to FD being more efficient than before. FE still finds more statistically significant findings than FD, suggesting the serial correlation was not very large.

GDP net HE and long-term unemployment become insignificant when lags of HE and time trends are added. OOP is only significant when using Driscoll-Kraay errors. This is a sign of reduced efficiency compared to PYLL. However, pollution becomes significant: a 10% increase in pollution increases avoidable mortality by 0.55%. NOx literarture suggests a long period is needed for effects of to be observable, so I am reluctant to view this as causative.

The FD model worked better with avoidable mortality. The significant co-efficients were in the exact same direction as the FE model predicted for PYLL.

VI Limitations and Sensitivity Analysis

Issue	Explanation
Standard error difference	The most efficient estimator for PYLL (and likely avoidable mortality too due to relatively low levels of serial correlation) was fixed effects. No/Low serial correlation resulted in most variables being statistically significant using either Stata's or Driscoll-Kraay's errors. Given that the latter adjust for cross-sectional dependence, there is reason to believe they are more accurate. This would lead to more intuitive results with respect to health expenditure, as the relationship between PYLL and HE turned negative when lags were included after being found as jointly significant by Driscoll-Kraay estimates.
Optimal Lag Length	This highlights the sensitivity of the findings. Firstly, there is no evidence that 2 lags are optimal. A different number may lead to different conclusions regarding the significance of lags even using Stata's robust errors.
Lack of use of Driscoll-Kraay errors	Secondly, Driscoll-Kraay errors are not used in much of the literature. This may be due to an inherent weakness in their calculation that means they may be inappropriate in some cases. If that applies here, only robust errors should be trusted, which imply a positive relationship between PYLL and HE.
Sensitivity of results to non-EU countries	In order to test the sensitivity of the findings further, I removed each non-EU country one at a time and repeated Model 3 for the PYLL regression. This did not affect the results, increasing the validity of my findings when conducting analysis of EU countries.
Endogeneity from omitted variables	Previous papers from the 1990s were frequently thwarted by endogeneity. Employing fixed effects and first differences removed the impact of country-specific effects. Furthermore, including lags may have reduced simultaneity. Unfortunately, some problems still remain. There are many potential omitted variables that may cause a significant impact that are so far unobservable, e.g. diet. Although good data for vegetable and macronutrient profiles exist for OECD countries, this does take into account important details. For example, processed, hydrogenated food and natural organic food could have the same macronutrients, but have different impacts on health (Bernard et al., 2019). These omitted variables could help explain health outcomes, particularly in rich countries where processed food is ubiquitous.
Endogeneity from reverse causality	Another source of endogeneity is reverse causality. The impact of income (minus HE) may not impact health outcomes but being healthy could allow people to become richer more easily. This problem could be overcome by using instrumental variables but given the lack of good data sources for appropriate instruments, this solution has not been viable.

Table8: Limitations and Sensitivity

VII Explorative Analysis

I undertook explorative analysis using a LSDV version of Model 3 with PYLL. No serial correlation makes LSDV efficient and time-invariant dummies can be added.

Optimal lag length is not obvious. Increasing lags in steps from 2 to 5 showed the t-stat for the third lag was consistently most significant. This could indicate that the most common amount of time taken between health expenditure investment and its delayed effect is 3 years. The lags reduce HE's co-efficient, but it's p-value remained below 0.0001.

Persistent effects from the USSR's collapse may remain. A dummy for countries formerly under the USSR's sphere of influence is significantly positive and is associated with a 20.3% increase in PYLL. Adding an additional EU dummy is statistically insignificant, suggesting EU and non-EU countries that were not in the Soviet sphere of influence have relatively similar country specific effects compared to countries formerly influenced by the USSR.

During the 2007-2009 Great Recession, mental health would have plummeted. These variables are not directly reliably observable, but adding a dummy shows this period is associated with a 2.25% increase in PYYL, which could be caused by omitted variables such as poor mental health due to increased underemployment (which the long-term unemployment variable could not account for).

Country-specific effects are unbiased but inconsistent as N increases. Adding and analysing regional dummies instead avoids this issue as more dummies are not created when N increases. This has been done in healthcare literature before (Cremieux et al., 2005).

VIII Discussion

This paper identifies and explains critical factors affecting health status outcomes across Europe by considering contributors to potential years of life lost and avoidable mortality, and their impact on health inequality in Europe.

Adding time dummies caused the explanatory variables' co-efficients to fall. From 2002-2016 mean ln(YPLL) in EU countries fell by 0.37, with 0.19 of this being explained by the explanatory variables (was 0.34 before addition of time dummies). This could indicate technology driving exogenous shocks in the macroeconomic production function. The size of these exogenous effects was roughly the same across both dependent variables' models. The exogenous shocks are difficult to replicate to improve health outcomes in the future. Countries could foster technological progress by removing red tape strangling innovation or possibly increasing competition in inventive industries.

Roughly half the variation in health outcomes can be explained by variables included in this analysis (other than time). This can be analysed more closely than the exogenous shocks. Ideally, we would calculate the cost-effectiveness of each method, but without this we rely on the contributing factors' impacts over 2000-2016.

Although my model initially suggested a positive relationship between HE and PYLL, this reversed when 3 lags of HE were added (optimised based on explorative analysis and SIC). Switching to a dynamic panel makes intuitive sense as it takes time for healthcare facilities to be built and used to save lives. HE contributes to healthcare resources that can improve quality and access to care.

In the future, obesity will play a significant role in EU health outcomes. The UK, Lithuania and Hungary have already crossed the 25% obesity rate past which dangers linked to rising obesity rates outweigh the benefits of fewer starving people. Countries in Eastern Europe have slightly higher obesity rates and more YPLL/avoidable mortality than Western Europe (OECD 2015). Alcohol consumption is also associated with poor health and is higher in

Eastern Europe. Reducing both obesity and alcohol consumption (risk factors) across Europe would improve health outcomes and reduce health inequalities. This could be done through taxes on sugar and alcohol, and this could be used to fund educational campaigns or subsidise unprocessed, low glycaemic index foods. However, there is a danger that the government could incentivise the wrong behaviour. For decades Western governments recommended reducing dietary saturated fats and replacing them with highly-processed polyunsaturated fats (Lugavere, 2018). It may be better for the government to encourage "better diet and lifestyle" rather than try to push for one particular diet.

The model suggests that the government could offset the impact of higher alcohol consumption with increased health spending. Note that this model views the impact of health spending holding other income fixed – i.e. there must be an increase in overall income to fund it.

IX Conclusion

Exogenous shocks have driven the fall in avoidable mortality and PYLL. Fostering innovation could accelerate this trend also lead to higher GDP, allowing more money to be spent on healthcare. As HE is the second biggest driver (lags included) this improves outcomes further.

Former Soviet-sphere countries have worse outcomes. This could be due to time-invariant effects from the Soviet era (or before). It's also due to time-variant effects i.e. high obesity and alcohol consumption. Reducing these reduces cross-national health inequality in the EU.

Future research should focus on data imputation to reliably increase observations and improve estimates. Better data collection by the OECD since 2010 should be expanded to reduce the need on imputation in the future.

X Bibliography

Baum, C. F. (2006). Stata Tip 38: Testing for Groupwise Heteroskedasticity. The Stata Journal, 6(4), 590–592. https://doi.org/10.1177/1536867X0600600412

Benjamin Born & Jörg Breitung (2016) Testing for Serial Correlation in Fixed-Effects Panel Data Models, Econometric Reviews, 35:7, 1290-1316, DOI: 10.1080/07474938.2014.976524

Benoit (2011): Linear Regression Models with Logarithmic Transformations

Brown, M. B., and A. B. Forsythe. 1974. Robust tests for the equality of variances. Journal of the American Statistical Association 69: 364–367

Crémieux, P.-Y., Meilleur, M.-C., Ouellette, P., Petit, P., Zelder, M. and Potvin, K. (2005), Public and private pharmaceutical spending as determinants of health outcomes in Canada. Health Econ., 14: 107-116. doi:10.1002/hec.922

De Hoyos, R. E., & Sarafidis, V. (2006). Testing for Cross-Sectional Dependence in Panel-Data Models. The Stata Journal, 6(4), 482–496. https://doi.org/10.1177/1536867X0600600403

GOLDBERGER, A. S. (1991). A course in econometrics. Cambridge, Mass, Harvard University Press.

Hall (2002). Testing for Unit Roots in Panel Data

Heijink & Xander Koolman & Gert Westert, 2013.
"Spending more money, saving more lives? The relationship between avoidable mortality and healthcare spending in 14 countries," The European Journal of Health Economics, Springer;Deutsche Gesellschaft für Gesundheitsökonomie (DGGÖ), vol. 14(3), pages 527-538, June.

Hoechle, D. (2007) ‘Robust Standard Errors for Panel Regressions with Cross-Sectional Dependence’, The Stata Journal, 7(3), pp. 281–312. doi: 10.1177/1536867X0700700301.

James, C., M. Devaux and F. Sassi (2017), "Inclusive growth and health", OECD Health Working Papers, No. 103, OECD Publishing, Paris, https://doi.org/10.1787/93d52bcd-en.

Joumard, I., et al. (2008), "Health Status Determinants: Lifestyle, Environment, Health Care Resources and Efficiency", OECD Economics Department Working Papers, No. 627, OECD Publishing, Paris, https://doi.org/10.1787/240858500130.

Mantobaye Moundigbaye, William S. Rea, and W. Robert Reed (2018). Which panel data estimator should I use?: A corrigendum and extension. Economics: The Open-Access, Open-Assessment E-Journal, 12 (2018-4): 1–31. http://dx.doi.org/10.5018/economics-ejournal.ja.2018-4

Moody, Carlisle E., Fixed-Effects Panel Data Models: To Cluster or Not to Cluster (September 17, 2016). Available at
SSRN: https://ssrn.com/abstract=2840273 or http://dx.doi.org/10.2139/ssrn.2840273

Stock, J. and Watson, M. (2007). Introduction to Econometrics, 2nd Ed., Pearson International Edition.

Ryo Okui, 2017.
"Misspecification in Dynamic Panel Data Models and Model-Free Inferences," The Japanese Economic Review, Japanese Economic Association, vol. 68(3), pages 283-304, September.
<https://ideas.repec.org/a/bla/jecrev/v68y2017i3p283-304.html>

W. Robert Reed & Haichun Ye (2011) Which panel data estimator should I use?, Applied Economics, 43:8, 985-1000, DOI: 10.1080/00036840802600087

Wooldridge, J. (2003). Introductory Econometrics, A Modern Approach, 2nd Ed., Thompson

Wooldridge (2007). Advanced Econometrics

https://stats.idre.ucla.edu/stata/webbooks/reg/chapter2/stata-webbooksregressionwith-statachapter-2-regression-diagnostics/

https://www.stata.com/manuals13/xtxtregpostestimation.pdf

https://www.statalist.org/forums/forum/general-stata-discussion/general/1491049-reset-test-for-omitted-variables

https://www.statalist.org/forums/forum/general-stata-discussion/general/1503799-standard-robust-errors-in-the-lsdv-and-fixed-effects-models

https://www.researchgate.net/post/How_can_I_solve_the_cross-sectional_dependence_and_heteroscedasticity_in_the_model_of_fixed_effects

Data Sources

www.data.worldbank.org
www.ons.gov.uk
www.stats.oecd.org
www.who.int/topics/statistics/en
www.ec.europa.eu/eurstat