The Contemporaneous Effects of Pollution on Subjective Wellbeing*

Tobias Leigh-Wood

18 May, 2020

Abstract

In this paper I try to find the causal contemporaneous effect of air pollution on subjective wellbeing. I construct a 27 year long panel from the Understanding Society Study (USS) with over 500,000 observations of individuals' wellbeing and match this with daily air pollution levels in UK regions and weather variables. I identify the casual effect of air pollution by using a fixed effects model. This controls for region level characteristics such as average road traffic. To account for measurement error and omitted variable bias I make use of atmospheric temperature inversions and regional wind direction as instrumental variables for pollution. I find no significant contemporaneous effects of pollution on wellbeing but do find a small negative statistically significant effect of yesterday's pollution on today's wellbeing and a large effect of average monthly pollution on wellbeing.

Contents

*I am grateful to Dr Johannes Spinnewijn, Dr Matthew Levy, Dr Sefi Roth, Dr Lutz Sager and Miss Leanne Cass for their help and advice.

1 Introduction

The link between long-term air pollution exposure and many health outcomes is well-established and well-studied: a search on Google Scholar for the effects of pollution on human health gives over 2.5 million hits for academic papers. This link is particularly worrying given that so many people are exposed to pollution levels far above the World Health Organisation's (WHO) recommended levels (World Health Organization, 2006). In the Asia Pacific region alone 4 billion people are exposed to dangerous levels of air pollution (Balakrishnan *et al.*, 2019) whilst globally this exposure is estimated to kill 3 million people every year (Lelieveld *et al.*, 2015). Air pollution is not just a problem for the developing world: in Europe and the UK air pollution is in places many times the safe limit, with the London underground having $PM_{2.5}$ (particulate matter less than 2.5 micrometers in diameter) levels 12 times the WHO guidelines (Saunders *et al.*, 2019).

There has been much research into the effects of pollution on economic outcomes. Higher levels of air pollution have been shown to contemporaneously decrease productivity in manufacturing (He, Liu and Salvo, 2019), in farm work (Graff Zivin and Neidell, 2012) and even in professional football (Lichter, Pestel and Sommer, 2017). A short-term rise in pollution also increases crime (Bondy, Roth and Sager, 2018), decreases standardised test scores (Lavy, Ebenstein and Roth, 2014) and increases car accidents (Sager, 2019).

Though there is a large literature looking at the short term-effects of air pollution on other outcomes, there is a lack of research that tries to identify the short-term effects of pollution on wellbeing. Most papers that research the effects of pollution on wellbeing average pollution over a year and aggregate wellbeing data to a high geographic level. For example, Dolan and Laffan (2016) use annual levels of $PM_{2.5}$ and cross-sectional subjective wellbeing (SWB) data matched at the local authority level to show a relationship between annual particulate matter and SWB. Likewise, Knight, Howley and others (2017) use yearly averages of air pollution to show that air pollution decreases life satisfaction.

Subjective wellbeing is a broad measure but can be broken down into three categories: evaluative, experiential and eudaimonic (White *et al.*, 2017). Whilst evaluative and eudaimonic wellbeing encapsulate overall feelings of life satisfaction and meaning, experiential wellbeing captures how people feel on a day-to-day basis. Given that my hypothesis looks at the instantaneous impact of pollution on wellbeing, I choose to focus on experiential wellbeing as my outcome variable of interest.

I use data from the Understanding Society Survey (USS) matched with local pollution and weather data to construct a 30 year panel from 1991 to 2019. I exploit the panel nature of this data and run a fixed effects regression, controlling for daily local weather conditions, to identify the causal impact of air pollution on wellbeing. Using a fixed effects method is key to getting around omitted variable bias that is often not controlled for well enough in cross sectional settings.

To complement this main analysis I use two instrumental variables to address endogeneity and measurement error. These two commonly used instruments are temperature inversions and wind direction. Temperature inversions have been used as instruments in many air pollution settings before and are plausibly exogenous (Sager, 2019). Wind direction has also been used as an instrument in an air pollution setting before (Deryugina *et al.*, 2019).

There are various mechanisms through which air pollution impacts wellbeing. These are important to identify and have been understudied in previous literature. The most obvious mechanism is via a health channel: air pollution has a well documented effect on different health outcomes, see Lippmann and Leikauf (2020) for a review. This could lower wellbeing directly. However, it could also be the case that pollution is affecting wellbeing before noticeable effects on health. This means that individuals may not attribute lower experiential wellbeing to higher levels of air pollution. This distinction is worth making because the welfare effect of pollution could be much greater once we include non-health channel effects.

I focus on the six key pollutants that have been shown to impact health and are found in high concentrations all around the world. These are: Nitrogen dioxide (NO_2), sulphur dioxide (SO_2), carbon monoxide (CO), ozone (O_3), particulate matter less than 10 micrometer (PM_{10}) and particulate matter less than 2.5 micrometer ($PM_{2.5}$). Recently $PM_{2.5}$ has been the most commonly studied pollutant since it is often found in very high concentrations in urban areas but in this paper I widen the scope of study to multiple pollutants.

One limiting factor in my analysis comes from the fact that I was not able to access the data that I had originally planned on using due to the closure of LSE. I had planned on using data from the USS that included the Middle Super Output Area (MSOA) of individuals. However, due to data licensing I am unable to access this data whilst not at LSE. I proceed with the same analysis but only know an individual's region, of which there are 12, as opposed to knowing an individual's MSOA, of which there are 7201. This has meant I measure an individual's exposure to pollution with more error than I would have otherwise.

Using a region fixed effect model I do not get results that are significant at the 5% level. However, I do find that the coefficients on monthly average air pollution are significant at this level. There also seems to be significant heterogeneity with men being more affected. I find no evidence of a health mechanism but also cannot rule one out.

In section 2 I discuss my data and potential identification strategies. In section 3 I present and explain results from the baseline model. I also look into heterogeneous effects across the population and add other extensions to the econometric model. In section 4 I test whether my results are robust to alternative specifications. Section 5 concludes.

2 Data and Empirical strategy

2.1 Data

Wellbeing: The wellbeing data comes from the Understanding Society Study (USS)(University of Essex, 2019). The Understanding Society Study is a longitudinal household study in the UK. It is formed of two key surveys, the UK Household Longitudinal Study (UKHLS) and the British Household Panel Survey (BHPS). The UKHLS currently has nine waves of data and has been running since 2009 whilst the BHPS ran from 1991 until 2008. The main difference between the two studies is in size, the BHPS interviewed 10,000 individuals a year whilst the UKHLS interviews 50,000 individuals a year.

The wellbeing measure is an aggregation of answers to 12 questions about an individual's wellbeing ranging from how well they are sleeping to how confident they feel about day to day life. A low wellbeing score indicates a higher level of wellbeing since these questions are asked on a scale of 0-3 with 0 being the answer indicating the highest wellbeing. An example of one of these questions is "Have you recently been able to concentrate on whatever you're doing?". A full list of questions can be found in the appendix. Generally the questions fall under a mixture of the experiential wellbeing category and the evaluative wellbeing category. I refer to this as the wellbeing measure with a higher wellbeing measure implying lower wellbeing.

I also take a variety of other variables from the USS such as life satisfaction, health satisfaction, monthly income and whether individuals have asthma.

Pollution: The pollution data comes from the Department for Environmental, Food and Rural Affairs (DEFRA)(Department for Environment, Food and Rural Affairs, 2020). DEFRA runs the Automatic Urban and Rural Network (AURN) which currently consists of 150 active monitoring sites. AURN sites measure a range of pollutants with the most commonly measured ones being: PM_{10}; $PM_{2.5}$; ozone pollution (O_3); Nitrogen Dioxide (NO_2); Sulphur Dioxide (SO_2) and Carbon Monoxide (CO). Figure 1 shows a map of the UK pollution sites on top of the NUTS borders.

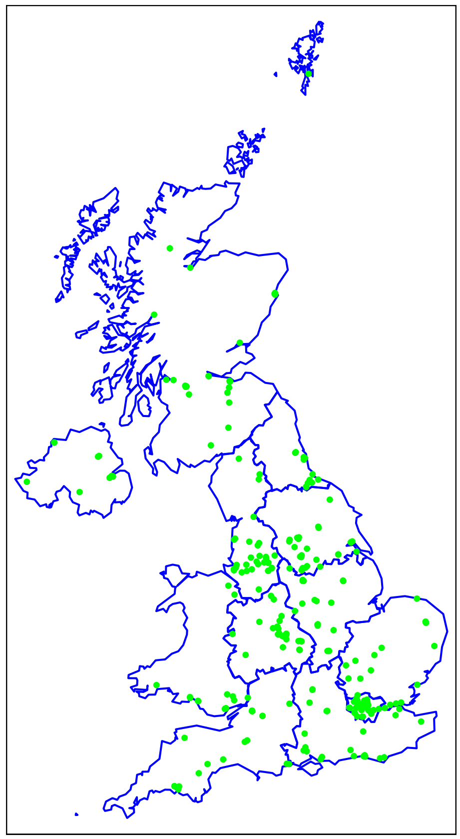

Figure 1: Map of AURN Pollution Sites

To construct my air pollution variable I first average air pollution by day for each site. I then match the sites to a UK NUTS region and average pollution from all sites that belong to that region. I then convert the pollutant levels to an Air Quality Index following US Environmental Protection Agency (EPA) guidelines (US Enviromental Protection Agency, 2018). The AQI is constructed by first finding the AQI score for each individual pollutant using the EPA formula (see equation 7 in the appendix) and taking the maximum of these individual pollutant scores. For example, if on the 1st July 1999 the AQI formula gave CO a score of 30 and NO_2 a score of 23, the AQI for London would be 30. The AQI is the measure of pollution I use throughout.

Weather: Weather data is taken from the UK Met Office Integrated Data Archive System (MIDAS) project which runs a network of sites across the UK (Met Office, 2012). I gather minimum and maximum daily temperatures, daily precipitation, daily wind speed and daily wind direction. I then match the sites to whichever UK NUTS region the site is in and average the weather variables across all sites within a region. Maximum and minimum temperatures are measured in degrees Celsius. Wind speed is measured in knots and rainfall is measured in mm of precipitation.

Atmospheric Data: A temperature inversion is an atmospheric weather event. Generally as you move further from the earth's surface air temperature decreases, and gases emitted at the surface rise up. A temperature inversion is when this process reverses and there is a band of warm air that traps cooler air and pollutants on the surface. I use temperature inversions as an instrumental variable for pollution.

The atmospheric temperature data I use to construct this instrument comes from NASA's MERRA 2 project (Global Modeling and Assimilation Office (GMAO), 2015). To build a measure of temperature inversion strength I take the continuous difference between surface temperature and temperature at an altitude of 600m at 3am. I follow Sager (2019) in choosing to take this difference at 3am, mainly because temperature inversions are occasionally visible during the day which may affect wellbeing. The data comes at a spatial resolution of 0.625 longitude and 0.5 latitude - I take the subset of these points that are within a UK NUTS region. I then average the atmospheric temperature difference inside each region. Hence, the temperature inversion is the continuous difference in surface temperature and temperature at 600m at 3am UTC.

It is also worth noting here that the panel assembled is not balanced. In figure 2 I show the number of years that individuals stay in the survey for. As you can see this is anywhere from one year to the full 28 years. The peaks at 9 years, 17 years and 28 years can be explained by the start of the UKHLS in 2009, BHPS running from 1991 to 2008 and then some people staying in the survey the whole 28 years. I will discuss the identification issues associated with this below.

I present summary statistics for both the pollution and Understanding Society data sets in Table 1.

Table 1: Summary Statistics

Statistic	Max	Min	Mean	St. Dev.	N
AQI	237	1	43.842	19.513	518,910
Wellbeing	36	0	11.118	5.486	518,910
Max Air Temperature	32.773	−5.038	12.973	5.033	518,910
Min Air Temperature	23.190	−12.016	7.561	4.573	518,910
Rain	178.932	0.000	3.649	6.777	518,910
Wind Speed	30.583	0.435	8.728	3.900	518,910
Monthly Income	28,149.670	0.000	1,473.293	1,437.204	518,910
SO_2	298.958	0.000	5.593	9.536	472,935
NO_2	368.514	0.283	32.161	15.499	511,443
PM_{10}	128.854	1.667	20.999	11.195	493,318
$PM_{2.5}$	100.125	0.000	12.236	8.447	350,298
CO	11.872	0.000	0.433	0.388	346,828
O_3	124.727	0.250	41.937	16.380	515,719
Male	1	0	0.449	0.497	518,910
Age	102	15	46.689	18.376	518,910

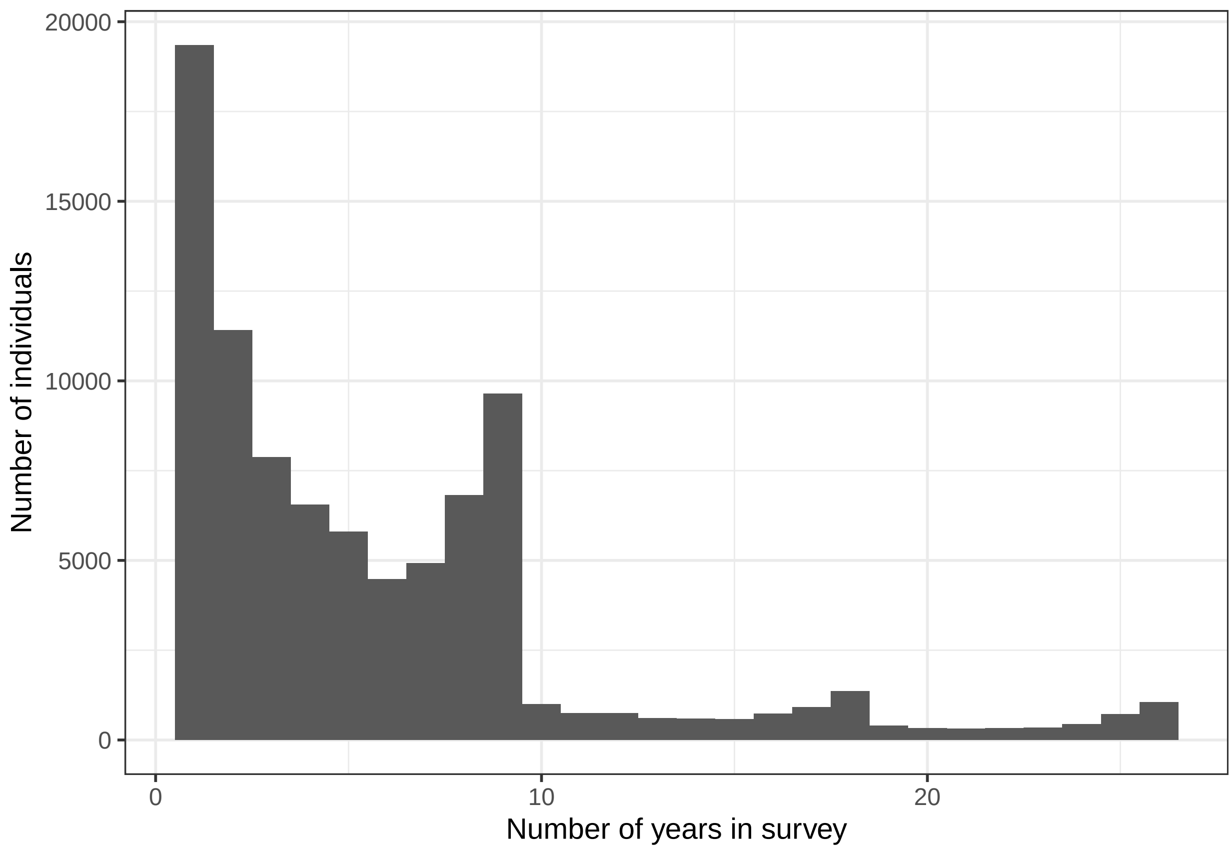

Figure 2: Unbalanced panel Historgram

2.2 Identification

A simple OLS regression of the form $Wellbeing_i = \alpha_i + \beta AQI_i + \epsilon_i$ is biased if there are factors that affect both AQI and wellbeing. One such factor is the location of an individual's house - in urban areas as opposed to rural areas air pollution is higher but so is noise pollution. This means a naive OLS regression will not estimate the effect of air pollution correctly. It has also been shown that air pollution is capitalised into housing prices (Chay and Greenstone (2005)), showing that people are choosing to pay more to avoid air pollution. This selection effect would also bias results.

To account for these unobservable characteristics I can use a fixed effects model. I utilise the fact I have repeated observations of the same person and that I have the date when the interview was carried out. Specifically a fixed effects regression only uses individuals' variation in wellbeing and pollution over time away from their mean rather than using variation between individuals. I also add time fixed effects to account for changes in wellbeing over time that are not to do with pollution. Adding time fixed effects takes cares of factors that change over time that are correlated with air pollution and the same for everyone. One example is seasonal changes in pollution and wellbeing.

However, adding time and individual effects does not control for factors that vary by time and individual. Weather varies by individual by time and affects both pollution and wellbeing and hence is a variable we want to control for. Adding controls for weather is thus very important in ensuring the coefficients are not biased. I add controls for maximum and minimum temperature, precipitation and wind speed. I also collected data on daily sunshine duration but it had over 50% Na so I didn't add it to the regression.

The individual fixed effects equation I estimate is the following.

$$Wellbeing_{it} = \alpha_i + \delta_t + \beta AQI_{it} + \gamma_1 MaxTemp_{it} + \gamma_2 MinTemp_{it} + \gamma_3 Rain_{it} + \gamma_4 WindSpeed_{it} + \epsilon_{it} \tag{1}$$

The coefficient β shows us the effect of a one unit change in the air quality index on the USS wellbeing measure. A positive β indicates that higher pollution causes a lower level of wellbeing. I make a clear distinction between the wellbeing measure which runs from 0 to 36 and 'wellbeing' more generally since a low wellbeing measure implies a high level of wellbeing. Hence, negative coefficients on a regressor indicate higher wellbeing and positive coefficients imply lower wellbeing.

The key identifying assumption for the fixed effect model is strict exogeneity. This means AQI_{it} cannot depend upon current, future or past values of the error term $epsilon_{it}$. After controlling for weather variables I claim that strict exogeneity holds.

The fixed effects model also relies on β being constant for all individuals no matter what an individual's characteristics are. This means there can be no interactions between α_i and AQI_{it}. This second assumption is less plausible, for the individual fixed effects estimator, given that some people will have characteristics that make them more susceptible to pollution, for example a pre existing medical condition. I will also argue that by construction you cannot use individual fixed effects in this setting.

To solve this issue I also run the following equation which includes region fixed effects instead of individual fixed effects. The issue with this regression, although it helps the interaction problem, is that there are still omitted variables which are likely not included in the fixed effects given how large UK regions are. For example, there will be differences in pollution, income and jobs within the South West which brings back some of the omitted variable concerns we discussed earlier.

The region fixed effects equation is the following.

$$Wellbeing_{itr} = \alpha_r + \delta_t + \beta AQI_{itr} + \gamma_1 MaxTemp_{itr} + \gamma_2 MinTemp_{itr} + \gamma_3 Rain_{itr} + \gamma_4 WindSpeed_{itr} + \epsilon_{itr} \tag{2}$$

Where $wellbeing_{itr}$ is the wellbeing of individual i at time t in region r.

There is also a problem of measurement error since we only observe the AQI for a region. UK regions are large and contain urban built up areas as well as rural areas, hence there is likely to be lots of variation of pollution within a region. Moreover, there is large variation in pollution within a day which will impact an individual's exposure. This means we are actually observing $\widetilde{AQI}_{it} = AQI_{it} + u_{it}$ where AQI is an individual's true exposure to pollution and u is the error around AQI. Since we do not observe the true value of the regressor we have measurement error.

Moreover, there is nothing to suggest this measurement error is classical. It may be the case that by averaging sites across a region we are putting in too many of one type of site. For example, the South West AQI estimate may be based off many rural sites and only a few inner city sites yet many individuals could be exposed to the city levels of pollution. We would then be systematically mismeasuring an individual's AQI exposure. With access to data on a finer geographic level this would pose less of a threat to identification.

This problem is exacerbated by the changing number of sites and site types throughout the 27 year period. For example, the number of urban sites increases throughout the period whilst the number of rural sites stays constant. Urban sites have higher pollution readings because there are more sources of pollutants in cities than in the countryside. Hence, my measure of pollution, the AQI, may appear artificially lower or higher than the true exposure for individuals dependent on the geographical distribution of the individuals in the Understanding Society panel.

One way to get around the problem of measurement error is to use an instrumental variable. To be valid an instrument has to satisfy four criteria. Firstly, it needs to be relevant, namely it needs to have a meaningful effect on the endogenous regressor. Secondly, it needs to be independent. This means that assignment of the instrument has to be uncorrelated with the outcome variable. Thirdly, the instrument needs to satisfy the exclusion restriction, that is it can only affect the outcome variable through the regressor of interest. The last criteria for a valid instrument is monotonicity, meaning the instrument affects the endogenous variable in the same direction for any value of the instrument. I use two instruments, temperature inversions and wind direction, which I claim satisfy these criteria.

Temperature inversions are plausibly independent, they are not caused by pollution and occur randomly. They affect pollution in the same direction - there are not places where the effect of a temperature inversion is to lower pollution. One could argue that given temperature inversions are a weather phenomenon then the exclusion restriction is not satisfied because the effects of the temperature inversion could be being felt through this weather channel. However, I control for weather in both stages of the regression so this is not an issue.

Temperature inversions have been widely used as an instrument for air pollution in different settings, for example by Sager (2019) to identify the effects of $PM_{2.5}$ on road accidents and by He, Liu and Salvo (2019) to study the effects of air pollution on productivity.

As explained in the data section the temperature inversion variable is the continuous difference in temperature between the surface and an altitude of 600m at 3am. A temperature inversion equates to this difference being negative. The first stage equation is as follows:

$$AQI_{itr} = \alpha_r + \delta_t + \phi TempDiff_{itr} + \gamma_1 MaxTemp_{tr} + \gamma_2 MinTemp_{tr} + \gamma_3 Rain_{tr} + \gamma_4 WindSpeed_{tr} + \epsilon_{itr} \tag{3}$$

The second stage equation is the following where $\widehat{AQI}_{itr}$ are the fitted values of the first stage regression estimates.

$$Wellbeing_{itr} = \alpha_r + \delta_t + \beta \widehat{AQI}_{itr} + \gamma_1 MaxTemp_{tr} + \gamma_2 MinTemp_{tr} + \gamma_3 Rain_{tr} + \gamma_4 WindSpeed_{tr} + \epsilon_{itr} \tag{4}$$

Where $TempDiff_{it}$ is the temperature inversion for individual i on day t and AQI_{itr} is the pollution level on day t for individual i in region r.

The second instrument I use is wind direction. Wind direction effects pollution through the fact that wind from certain directions increases pollution and wind from other directions decreases pollution. For example if you are north of Heathrow and a southerly is blowing you experience higher levels of pollution than if the wind was blowing from the north. Again I argue that wind direction, after controlling for other weather phenomenon, is a valid instrument.

This identification strategy has been used by Bondy, Roth and Sager (2018) where they look at the effects of pollution on crime in London wards. Likewise Deryugina *et al.* (2019), use wind direction as an instrument when studying the effects of $PM_{2.5}$ on mortality in California.

Wind direction affects pollution differently in different regions based on the sources of pollution and other characteristics of the region. Hence I let the effects of wind direction on pollution differ by region. The wind direction variable is binned into 4 categories of 90 degrees. The estimating equation used are:

$$AQI_{itr} = \alpha_r + \delta_t + \phi_r WindDir_{tr} + \gamma_1 MaxTemp_{tr} + \gamma_2 MinTemp_{tr} + \gamma_3 Rain_{tr} + \gamma_4 WindSpeed_{tr} + \epsilon_{itr} \tag{5}$$

$$Wellbeing_{itr} = \alpha_r + \delta_t + \beta \widehat{AQI}_{itr} + \gamma_1 MaxTemp_{tr} + \gamma_2 MinTemp_{tr} + \gamma_3 Rain_{tr} + \gamma_4 WindSpeed_{tr} + \epsilon_{itr} \tag{6}$$

Where $WindDir_{tr}$ is the wind direction in region r in time period t .

As mentioned in the data section the panel I construct is very unbalanced. This will cause an issue with identification if there is selection in and out of the panel based on the individual or region fixed effects. This means that if people who have lower wellbeing are more likely to leave the panel survey than others, we will not be estimating the effects of pollution on wellbeing consistently. Likewise if individuals who are joining the survey are selecting into it because they have higher wellbeing then we will also have problems.

Since pollution is essentially assigned at the regional level I therefore cluster standard errors at the regional level.

I carry out all my analysis in R and use the lfe package (Gaure *et al.*, 2019) to run my fixed effects regressions.

In summary an OLS regression would be vulnerable to omitted variable bias as well as measurement error. To get around omitted variable bias I use a fixed effect model which accounts for all time invariant individual confounders and time varying confounders that affect all individuals. However, measurement error is still an issue, especially given the data constraints. This means I employ two instruments, wind direction and temperature inversions. This helps overcome the problem of measurement error.

3 Results

3.1 Main Results & Instrumental Variables

Table 2 shows the results from running equations 1 and 2 discussed above alongside other equations. We see that with no controls the coefficient on AQI is positive, which implies a negative effect on wellbeing, but it is insignificant and essentially zero. This is because there are omitted variables biasing the results. An omitted variable is one that is correlated with the explanatory variable, in this case air pollution, and the independent variable, in this case wellbeing. This leads to bias in the estimation of the the true effect. In this case one potential omitted variable would be the location of an individual's house. Urban areas, such as London, may have higher levels of pollution but also have higher levels of income. This would would lead to us underestimating the effects of pollution on wellbeing.

Another omitted variable is rainfall. Rainfall cleans the air through a process called washout (Hemond and Fechner, 2015), meaning on rainy days pollution is lower. However, peoples' wellbeing may also be lower because of the bad weather. This association between rain and pollution would cause us to underestimate the effects of pollution on wellbeing. Essentially the regression with no controls using an OLS specification tells us very little about the effects of pollution on wellbeing.

Adding weather controls can deal with some of the omitted variable bias. We see this in column 2 of Table 2. The coefficient on AQI is now negative and significant although very small at -0.000736. This means that an increase in AQI is associated with a decrease in the wellbeing index and thus an increase in wellbeing. Note that wellbeing in the Understanding Society Survey is measured from 0 to 36 with individuals who have low scores being happier than those with higher scores. This means that negative coefficients imply a positive effect of pollution on wellbeing. The fact we observe a positive effect of pollution on wellbeing in column 2 tells us that there is probably omitted variable bias when using just the OLS regression even with controls.

Another source of omitted variable bias comes from factors that change over time. For example, an economic shock that affects the whole of the country lowers pollution but also increases unemployment which will lower wellbeing. This would lead us to underestimate the effects of pollution or even estimate that lower pollution causes lower wellbeing which is the result we saw in column 2 of Table 2. I now add daily time fixed effects to account for omitted variables that change over time and affect both pollution and wellbeing. The argument that omission of time fixed effects could be causing misestimation is made evident when we run a regression with time fixed effects. We see the coefficient rise to positive (seen in column 3 of Table 2) which means that pollution lowers wellbeing. This is consistent with the argument made above that there was omitted variable bias coming from time shocks.

As discussed in the identification strategy, individual fixed effects can be used to control for individual specific time invariant omitted variables. The results of equation 1, the fixed effects specification, are shown in column 4 of Table 2. We see a coefficient on AQI of -0.000657, this is statistically insignificant. The choice of specification between individual fixed effects and region fixed effects is a difficult one. An understanding of the variation used to estimate the coefficients with each specification is important. In the individual fixed effects specification the coefficients are identified from variation in an individuals wellbeing compared to their own average wellbeing. We are only

Table 2: Main Regressions

	Dependent variable:					
	wellbeing					
	OLS		*felm*			
	No Controls	With controls	Time FE	Individual FE	Region FE	Urban Rural
	(1)	(2)	(3)	(4)	(5)	(6)
AQI	0.000003	−0.000736*	0.000582	−0.000657	0.001427	0.003642**
	(0.000390)	(0.000420)	(0.001219)	(0.000626)	(0.000993)	(0.001282)
MaxTemp		−0.016742***	−0.045118**	−0.011851	−0.009739	−0.003635
		(0.004346)	(0.018372)	(0.011680)	(0.010606)	(0.016929)
MinTemp		0.003294	0.015494	0.012761	0.014793	0.013497
		(0.004705)	(0.018264)	(0.009374)	(0.009284)	(0.013968)
Rain		0.004768***	0.003666	0.002747*	0.000708	0.000348
		(0.001180)	(0.003515)	(0.001493)	(0.001871)	(0.001787)
WindSpeed		−0.008038***	−0.022831**	−0.001257	−0.006556	−0.002321
		(0.002201)	(0.009988)	(0.002644)	(0.004671)	(0.004308)
Constant	11.117480***	11.394920***				
	(0.018730)	(0.046893)				
Individual FE	No	No	No	Yes	No	No
Time FE	No	No	Yes	Yes	Yes	Yes
Region FE	No	No	No	No	Yes	Yes
Observations	518,910	518,910	518,910	518,910	518,910	315,982
R^2	0.000000	0.000197	0.016686	0.550212	0.017659	0.015214
Adjusted R^2	−0.000002	0.000187	0.003137	0.449346	0.004104	0.003728
Note:						*p<0.1; **p<0.05; ***p<0.01

comparing an individual's wellbeing to the same individual's wellbeing in another time period. In the regions fixed effects specification we are comparing individuals across time who are all in the same region.

A fixed effects regression is also biased and inconsistent if there are interaction terms between between α_i and β as discussed in the identification section. In this case individual effects are biased because β is a function of α. This is because there are interactions between the individual characteristics and the effect of pollution on wellbeing. The AQI for a region will be driven up by a rise in pollution in certain areas, this means that if you live in one of those areas where pollution increase is greater than others you will, by construction, be affected more by pollution than those living in other areas. So for the same seen change in AQI there will be a larger β for some people dependent on where they live. Hence, there is an interaction between α_i and AQI so the estimates are biased and we see a negative coefficient. I therefore proceed with the region fixed effects estimator for the remainder of the analysis.

If pollution data was not aggregated at the region level, but we had individual exposure for every individual, an individual fixed effects regression would not be bias. But because pollution assignment is on the regional level, individual fixed effects would be bias.

The results of including region fixed effects and time fixed effects, which control for region specific time invariant characteristics show us that there was bias affecting the results in columns 1, 2, 3 and 4. Omitted variables that are different between regions are taken care of. We now see a coefficient of 0.001427 (column 5 of Table 2) on AQI which indicates a negative effect of pollution on wellbeing.

To get an indication of the magnitude of this result I regress wellbeing on monthly income using the region fixed effects estimator. This gives a coefficient of -0.000293 on monthly income (seen in Table 10 of the appendix). This means that for every £100 of monthly income, wellbeing increases by 0.3 units. Average AQI is 40 with a standard deviation of 20 so a back of the envelope calculation, assuming the model is correct, suggests that a 1 standard deviation decrease in AQI is inequivalent to a £100 increase in monthly income.

I also examine the effects of lags of air pollution on today's wellbeing. Pollution may have a lagged effect on wellbeing especially when the USS wellbeing measure is not an entirely hedonistic measure of wellbeing. By this I mean that due to the nature of the wellbeing questions individuals will be considering a larger scope of their life than just today. This means effects of pollution from the last week or more may be felt. A high pollution day yesterday may get you down today more than yesterday since the effects could be lagged. This could be similar to a bad nights sleep having an effect the day after you have it rather than the day you wake up from it. However, it is also worth noting here that this lagged regression was not in my original analysis plan so any significant results should be treated with scepticism.

The lagged equation looks like equation 2 but includes two lags of pollution. I present the results from the lagged regressions in Figure 3.

The horizontal lines in Figure 3 represent 95% confidence intervals around the point estimates, so if one of the horizontal lines does not touch the 0 vertical line the regressor is statistically different from 0 at the 5% level. This means that under the hypothesis that the true coefficient is 0 there is less than a 5% chance of observing data more adverse to the null than the data we actually observed. Due to the magnitude of difference between the estimates for the weather controls and the estimates for the pollution lags I omit the weather controls from the figure although the equation was estimated with them in. Results tables for two lags and six lags can be found in the appendix. We

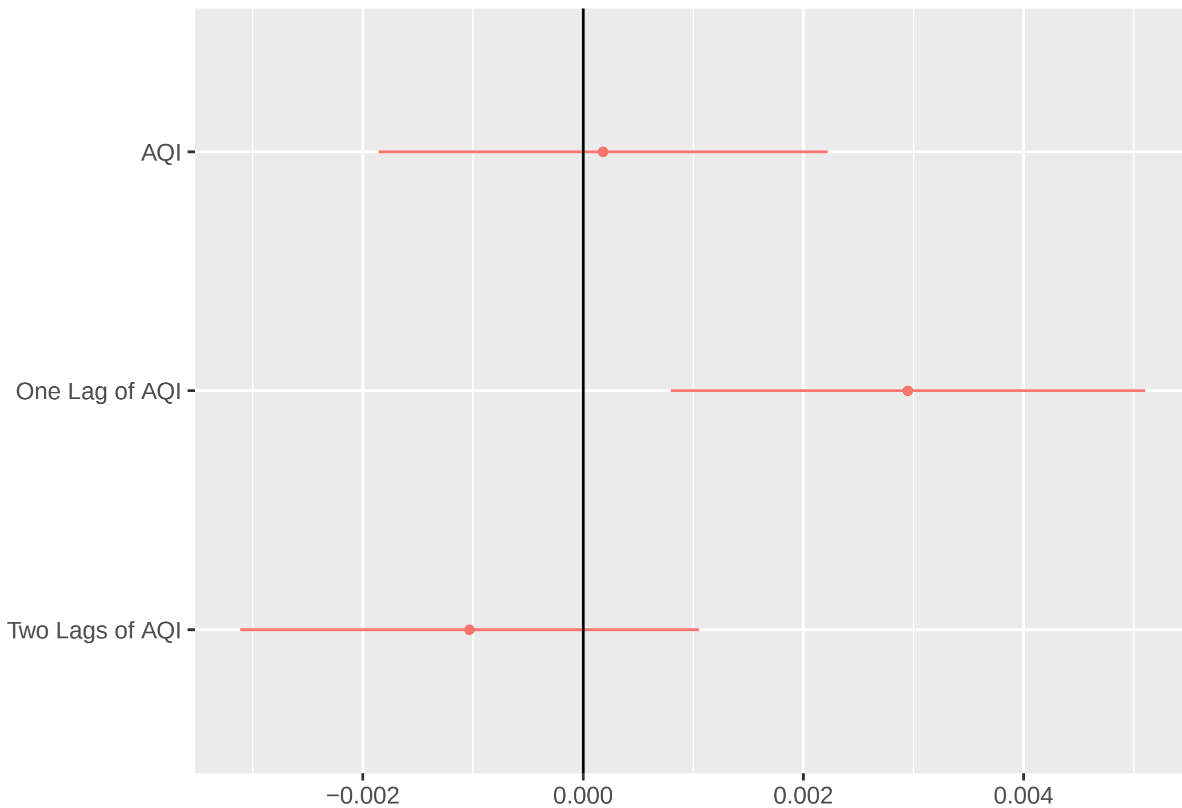

Figure 3: Lagged Regression Estimates

see that pollution with one day of lag has a coefficient of 0.002949. This means pollution yesterday is decreasing wellbeing today.

However, pollution is highly correlated over time so the result may be coming from multicollinearity. Perfect multicollinearity occurs when one explanatory variable can be linearly predicted from the other regressors and this means the effects of the explanatory variable are impossible to identify. Intuitively when two variables are highly correlated it becomes difficult for the model to identify the individual impact of either variable. And this ends up making the estimates highly inaccurate which is what we could be seeing in Figure 3.

As discussed in Section 2 we need to consider measurement error as well as omitted variable bias. The insignificant coefficients reported in columns 4 and 5 of Table 2 are attributable to both measurement error and omitted variable bias. We have measurement error because we only know which region of the UK an individual is in. This means an individual's AQI is calculated from a region's average pollution.

In the case with one regressor with measurement error, the attenuation factor is $\sigma_x^2/(\sigma_x^2 + \sigma_u^2)$ where u is the error term I introduced in the identification strategy (Stock and Watson, 2015). However, measurement error is more of a problem in panel data settings with the attenuation factor becoming $\frac{1}{1+\frac{\sigma_u^2(1-r)}{\sigma_x^2(1-p)}}$ if x_t is stationary where ρ is the autocorrelation coefficient for u and r is the autocorrelation coefficient for x (Griliches and Hausman, 1986). This means that by using fixed effects we may be introducing more bias into the model because of measurement error.

One obvious way around the problem of measurement error is to better assign pollution to an individual using the information we have. In the UKHLS section of the data I know the region of an individual and whether they are in a rural or urban location. The pollution site data also contains information on whether the site is in an urban or rural location. I hence find the average urban pollution level and average rural pollution level on certain days in certain regions and merge this with the information on which region an individual is in and whether they are in an urban or rural location. The results from this regression are presented in column 6 of Table 2. The coefficient on AQI is now 0.003642 and significant at the 1% level. This is more than twice as high as the coefficient seen in column 5 illustrating that measurement error is a significant issue.

As discussed in section 2, one way around the problem of measurement error is to use an instrumental variable. This has to be relevant, independent and monotone and satisfy the exclusion restriction. I use two weather phenomena which I claim cause exogenous variation in the amount of pollution: temperature inversions and wind direction. The results from these regressions are presented in Table 3.

Instrumental variable regression occurs in two stages. First, the explanatory variable is regressed on the instrument. Second, the dependent variable is regressed on the fitted values from the first stage.

The results from the temperature inversion regression are displayed in columns 1 and 2 of Table 3. Column 1 shows the first stage regression of AQI on temperature inversions. The coefficient of -1.57 tells us that for every degree increase in temperature inversion pollution increases by roughly 1.5 AQI points. Weak instruments can cause problems in inference. To test whether I have a weak instrument problem I look at the F statistic from the first stage. Assuming that the effects a temperature inversion are homogeneous, an F statistic of above 20 is a good indication that an instrument is not weak. The first stage F statistic is 260 - this shows temperature inversions are not a particularly weak instrument.

Table 3: IV Regressions

	Dependent variable:		
	AQI	Wellbeing	
	Temperature Inversion	Temperature Inversion	Wind Direction
	1st Stage	2nd Stage	2nd Stage
	(1)	(2)	(3)
MaxTemp	−0.008136	−0.009975	−0.009890
	(0.019809)	(0.010629)	(0.010617)
MinTemp	−0.657588***	0.018965	0.017463
	(0.018226)	(0.011859)	(0.013193)
Rain	−0.020281***	0.000821	0.000780
	(0.002731)	(0.001878)	(0.001887)
WindSpeed	−0.380944***	−0.004510	−0.005246
	(0.007713)	(0.005822)	(0.006557)
Inversion Strength	−1.572846***		
	(0.039440)		
$\widehat{AQI}$		0.006961	0.004969
		(0.009656)	(0.012443)
Individual FE	Yes	Yes	Yes
Time FE	Yes	Yes	Yes
Region FE	No	No	No
Observations	518,910	518,910	518,910
R^2	0.781845	0.017574	0.017624
Adjusted R^2	0.778835	0.004017	0.004068
Note:			*p<0.1; **p<0.05; ***p<0.01

The 2nd stage estimates are presented in column 2 of Table 3. We see that the coefficient on AQI is 0.006961. This is much larger than the coefficient seen in column 5 of Table 2 (0.001427) showing that attenuation bias was an issue and the instrument goes some way to solving it. This coefficient is also higher than the Urban & Rural regression but is not statistically significant.

For the wind direction instrument I interact the wind bins with each region so that the effect of wind direction on pollution differs by region. Without doing this monotonicity could be violated since any one direction of wind may raise pollution in some regions and decrease pollution in other regions. The first stage results are omitted from the table. The F statistic from the first stage is 258 which shows that I do not have a weak instrument problem if the effect of wind is homogeneous. However, wind direction does not have a homogeneous effect within a region, so the F stat test isn't a valid test for weak instruments.

The second stage results has a coefficient of 0.004969 on AQI which is higher than the result seen in the original region fixed effects regression but slightly lower than the result from the temperature inversion regression. This result is also not statistically significant.

The exclusion restriction is harder to argue for given the broadness of wellbeing as an outcome. However, after controlling for weather variables there are few channels that weather inversions and wind direction could affect wellbeing apart from via pollution.

The standard errors of our results increase when we use the instruments rather than the original specification. So we cannot reject that the coefficients on $\widehat{AQI}$ are equal to 0 at any level of significance since the p values are high. To understand what this means in the context of this paper it is worth noting what exactly a p value is. The p value associated with some state of the world θ_0 is the probability of observing data Y at least as extreme as the data that was actually observed, y, when θ_0 is the true state of the world (Stock and Watson, 2015). If the p value is less than 0.05 for the hypothesis that the coefficient is 0, it means that there is less than a 5% chance of observing that data or data more adverse if the coefficient was actually 0. The high p values that are associated with the regressions we have run so far are all associated with the null hypothesis that the effect of pollution is in fact zero.

In this section we see evidence that there is potentially a contemporaneous effect of pollution on wellbeing. Although the region fixed effects estimator is small, the coefficients from the Urban & Rural regression and the instrumental variable regressions were significantly higher than the coefficient seen on the basic region fixed effects estimator. This indicated that there is a degree of measurement error. With access to the data at a finer geographic level we might see more significant results.

3.2 Mechanisms & Heterogeneity Analysis

I now move on to discussing potential channels that air pollution affects wellbeing through. Of course the obvious candidate is through a health effect. People could be exposed to air pollution which then causes a health effect, lowering an individual's wellbeing. However, there could also be effects of pollution through a non-health effect. That is, pollution affects wellbeing independently of the health effect. There is also the question of whether people realise that pollution is affecting their wellbeing and how this may impact their behaviours.

To examine channels which pollution acts through I introduce new variables and look at the interaction of pollution with them. I first examine the effects of pollution on individuals with asthma. Air pollution impacts asthmatics' health more than other individuals (Guarnieri and Balmes, 2014), so I hypothesise that if wellbeing is acted upon through the health effect then asthmatics should have a higher coefficient on AQI. That is they feel the effects more for the same level of pollution. I construct the asthma variable from the UKHLS data since it is included in the BHPS part of the data.

The results from this are shown in column 1 of Table 4. Although we see a significant effect of asthma on wellbeing, we do not see any significant difference in the effect of pollution for those with asthma. If there was a significant difference it would show up in the AQI:Asthma interaction term. The fact there is no significant difference indicates that pollution may not be acting through a respiratory health channel.

When adding interactions and controls we need to be careful that we are not adding any bad controls. A bad control is one that is caused by the independent variable and should really be on the left hand side of the regression equation. In this case asthma could be seen as a bad control because long term exposure to pollution could be causing asthma. However, the evidence for this is low (Anderson, Favarato and Atkinson, 2013). I claim that even if pollution does cause asthma, since we are looking at very short term exposure, the increase in asthma rates from this exposure is negligible. Hence we do not have a bad control issue.

To analyse the health channel further I look at the effects of pollution on health satisfaction. I hypothesise that if pollution is acting through a health channel then we would see effects on health satisfaction as well as effects on wellbeing. The health satisfaction measure I use ranges from 1 to 7 with 7 being the most satisfied. The results from this regression are presented in column 2 of Table 4. We do not see any significant results of short term pollution on health satisfaction and the coefficient is not statistically different from 0. This adds to the evidence that if pollution affects wellbeing in the short term then it does not affect it through a health channel.

The health satisfaction variable is only available in the BHPS section of the data so unfortunately I cannot regress health satisfaction on the interaction between the asthma variable and AQI.

To examine the channels further I now look at which components of the wellbeing measure are affected in which directions. The wellbeing measure used throughout is the aggregation of 12 questions listed in the appendix. Briefly they relate to the following areas: concentration, sleep, decision capability, playing a useful role, overcoming difficulties, enjoy day to day, ability to face problems, losing confidence, self worth and general happiness. I now see whether pollution is impacting one of these more than the others. If pollution were impacting certain components of wellbeing more than others it could potentially give us an indication of the mechanism through which pollution is acting.

I run equation 2 twelve times with each component of wellbeing as a dependent variable. In Figure 4 I report the coefficient on AQI for each of the models.

We see that the largest results are coming from concentration, being under strain and losing confidence. Concentration being affected ties into previous studies that show air pollution impacts exam scores (Lavy, Ebenstein and Roth, 2014) but being under strain and losing confidence are not necessarily expected. However, we also see every standard error is large and none of the estimates are significant, so we cannot conclude much from Figure 4.

I next examine whether the effect of pollution differs by age, gender or race. There have been studies that show the very young and the old are affected more by pollution. If there are significant differences between the effects of pollution on men and women it could help explain inequalities in health outcomes and the such. This is an important question currently given that men seem to be

Table 4: Mechanisms

	Dependent variable:	
	Asthma	Health Satisfaction
	(1)	(2)
AQI	0.001561	0.000357
	(0.000968)	(0.000324)
Asthma	2.338980***	
	(0.152228)	
MaxTemp	−0.002488	0.009124
	(0.016599)	(0.008838)
MinTemp	0.011304	−0.004541
	(0.013943)	(0.005541)
Rain	0.000504	−0.000384
	(0.001794)	(0.000910)
WindSpeed	−0.000096	0.002630
	(0.004127)	(0.001672)
AQI:Asthma	−0.001011	
	(0.002064)	
Individual FE	No	No
Time FE	Yes	Yes
Region FE	Yes	Yes
Observations	320,134	140,308
R^2	0.029371	0.026068
Adjusted R^2	0.018192	0.009298
Note:		*p<0.1; **p<0.05; ***p<0.01

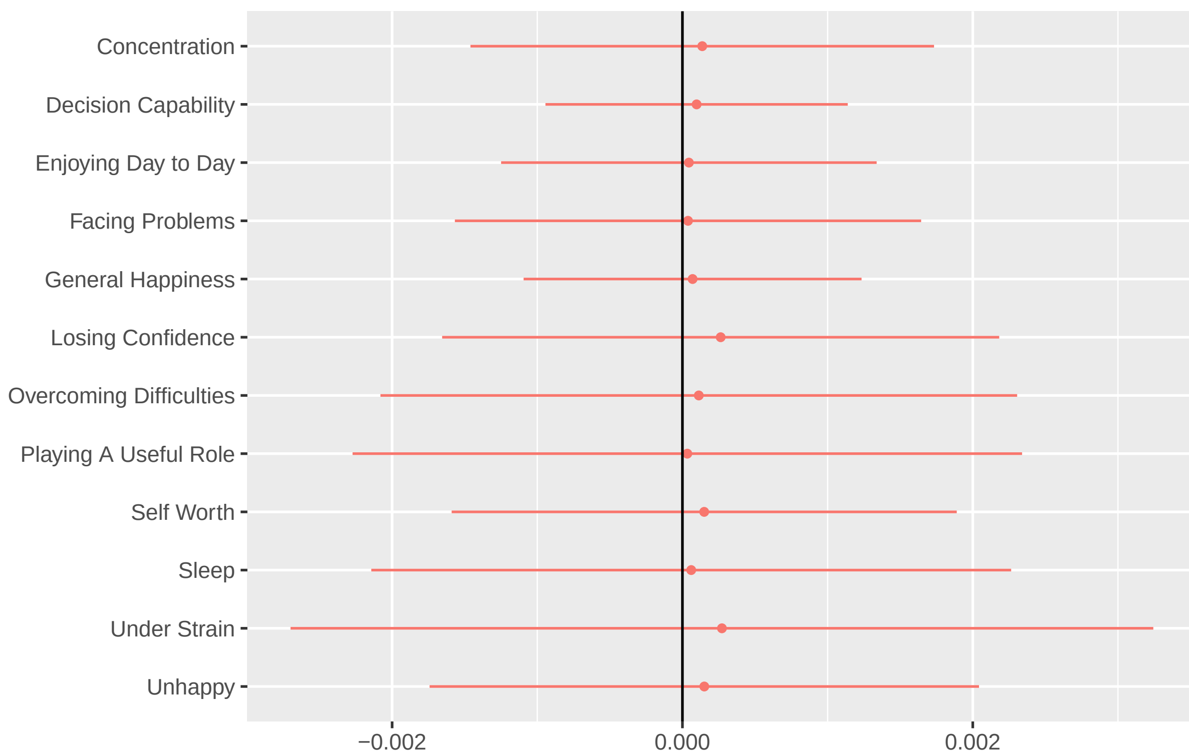

Figure 4: Wellbeing Components

Table 5: Heterogeneity Analysis

	Dependent variable:		
	Wellbeing		
	(1)	(2)	(3)
AQI	0.014635***	−0.001253	−0.000643
	(0.003465)	(0.001595)	(0.000900)
Age	0.011369***		
	(0.002832)		
White		−0.299357**	
		(0.109653)	
Male			−1.448150***
			(0.069156)
MaxTemp	−0.009393	−0.010411	−0.009427
	(0.014302)	(0.014494)	(0.015076)
MinTemp	0.014420	0.014550	0.016218
	(0.010847)	(0.010805)	(0.010623)
Rain	0.000618	0.000769	0.000557
	(0.001414)	(0.001453)	(0.001380)
WindSpeed	−0.006529	−0.006326	−0.006833
	(0.004445)	(0.004528)	(0.004238)
AQI:Age	−0.000287***		
	(0.000071)		
AQI:White		0.003132**	
		(0.001392)	
AQI:Male			0.004463***
			(0.001020)
Individual FE	No	No	No
Time FE	Yes	Yes	Yes
Region FE	Yes	Yes	Yes
Observations	518,910	518,910	518,910
R^2	0.018007	0.017720	0.030486
Adjusted R^2	0.004452	0.004162	0.017103
Note:			*p<0.1; **p<0.05; ***p<0.01

dying from Covid-19 at a faster rate than women and that pollution seems to be associated with a higher Covid death rate. Could it be that the effects of pollution are greater for men than they are for women and that this is a contributing factor to why more men are dying of the disease?

The results from the gendered regression are shown in column 3 of Table 5. We see a large increase in wellbeing overall for men relative to women. There also is a significant difference between the effects of pollution on men and the effects of pollution on women. The coefficient of 0.004463 on the interaction between AQI and the dummy variable for male shows us that the effect is higher for men than it is for women. However, measurement error in pollution exposure could be driving this effect. Men are more likely to work in an occupation that involves lots of time outside during the day. Since air pollution is generally higher outside than inside the effect on men could be just reflecting that exposure is higher for them. But, we cannot rule out that there isn't a greater effect of air pollution for men than for women.

The ethnicity regressions shown in column 2 of Table 5 show us that the effects of pollution are higher for those from white ethnic backgrounds as opposed to those from other ethnic backgrouds. The interaction coefficient is significant at the 5% level and implies a small effect of around 0.002 (-0.0012 + 0.0031) for white ethnic groups. This again may be because of measurement error in pollution exposure. I present results by all ethnic groups in the appendix.

The age regression shown in column 1 of Table 5 shows that the effects of pollution are decreasing with age. For every year increase in age the effect of pollution decreases by 0.000287. This implies that pollution is only negatively affecting younger individuals' wellbeing whilst older individuals' wellbeing is positively associated with air pollution. In the appendix (Table 12) I bin age into 3 groups and run the same regression with an interaction term for each bin. Those results mirror the ones we see in Table 5.

This effect may be explained through the fact that younger people generally live in cities which have higher air pollution. If the day to day variation in pollution is being driven by increases in pollution in cities which only affects younger people then we would expect to see a result like this. It could also be that young people are out and about near roads and industrial sites more than old people are. This would again mean for the same regional average pollution younger people are getting a higher level of pollution exposure which is driving the result.

We have seen in this heterogeneity section that the effect of pollution on an individual differs by age, gender and ethnic group. The differences amongst age and gender illustrate the problems with using non-experimental data when looking at the effects of pollution on wellbeing since the assignment of pollution exposure to an individual is a problem. More precise pollution data would help us get around this problem and better identify the effects of pollution on wellbeing.

3.3 Extensions

Another issue is correlation between pollutants. This makes finding the effects of an individual pollutant difficult because of omitted variable bias between pollutants and multicollinearity. If NO_2 and PM_{10} are correlated and NO_2 has a true effect on wellbeing whilst PM_{10} doesn't, a regression of wellbeing on PM_{10} will suggest that PM_{10} has an effect on wellbeing but it is actually that they are both correlated with NO_2. I show in Table 6 that there is significant correlation between the different pollutants.

Figure 5 shows the differing coefficients when including all pollutants versus just including one. The 'Individual' model shows the coefficients from running the region fixed effect equation with just that

Table 6: Correlation Table

	NO_2	SO_2	CO	PM_{10}	$PM_{2.5}$	O_3
NO_2	1	0	0	0	0	0
SO_2	0.438	1	0	0	0	0
CO	0.608	0.496	1	0	0	0
PM_{10}	0.595	0.463	0.535	1	0	0
$PM_{2.5}$	0.525	0.239	0.374	0.921	1	0
O_3	-0.555	-0.260	-0.384	-0.345	-0.371	1

one pollutant as the explanatory variable. Whereas the 'All Pollutants' model shows the coefficients from including all pollutants in one equation. When including multiple pollutants - namely the individual model - the standard errors rise illustrating the effects of multicollinearity. The point estimates also shift illustrating the effect of omitted variable bias in the individual specification from not including other pollutants.

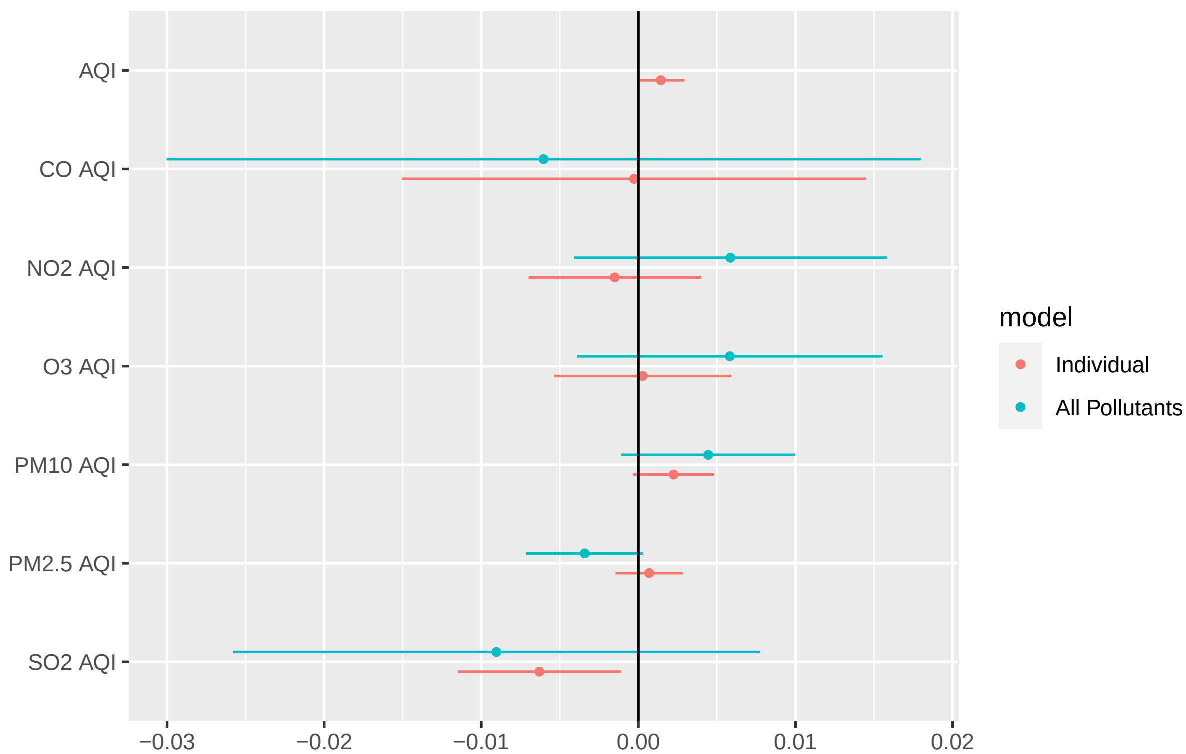

Figure 5: Separate Pollutants

Many previous studies have looked at the effects of long term pollution exposure, typically a year, on wellbeing and life satisfaction. The aim of this paper was to examine the evidence for instantaneous effects of pollution on wellbeing and look at the underlying mechanisms. However, as an extension to my main analysis I now look at whether these longer-term effects occur in my setting. In column 1 of Table 7 I regress wellbeing on yearly averages of the AQI score. In column 2 I regress wellbeing

Table 7: Extensions

	Dependent variable:	
	Wellbeing	
	(1)	(2)
Yearly Pollution	0.006134**	
	(0.002290)	
Monthly Pollution		0.013162**
		(0.004352)
MaxTemp	−0.009569	−0.005989
	(0.014542)	(0.014079)
MinTemp	0.013887	0.014936
	(0.010823)	(0.010685)
Rain	0.000774	0.000702
	(0.001447)	(0.001432)
WindSpeed	−0.007451	−0.007004
	(0.004397)	(0.004333)
Individual FE	No	No
Time FE	Yes	Yes
Region FE	Yes	Yes
Observations	518,910	518,910
R^2	0.017688	0.017704
Adjusted R^2	0.004133	0.004149
Note:	*p<0.1; **p<0.05; ***p<0.01	

on the monthly average AQI.

Although I do not control for variables that would bias these results, such as unequal trends between regions or monthly weather variables, the results in Table 7 are robust to averaging weather variables over a month or year. The large coefficient of 0.013162 seen in column 2 of Table 7 implies that a 40 point decrease in average monthly AQI would increase wellbeing by 0.5 points. The coefficient of 0.006134 seen in column 1 is also significant at the 1% level and shows a negative effect of pollution on wellbeing.

4 Robustness

As discussed in the main results section the region fixed effects model may be biased because it doesn't take into account individual level characteristics such as the area where someone lives. I now test the robustness of the results seen in Table 2. A robustness test involves checking to see if my results are stable after changing the econometric specification.

One route for these results to be biased would be location as an omitted variable. Income is a relatively good proxy for location since wealthier areas often have less pollution. So location affects where someone lives which affects their pollution exposure and location affects wellbeing directly. The results in column 1 of Table 8 show us that income does have an effect on wellbeing as expected. I present the results from the original region fixed effects specification in column 2 for reference. We see that the coefficients on AQI are very similar in both regressions. This tells us that our results are 'robust' to adding a potential omitted variable.

I also run the region fixed effects specification including yearly pollution as well as daily pollution. This tests whether my region fixed effects estimates for the instantaneous effects of pollution are robust to adding in yearly average pollution. Column 3 of Table 8 shows the region fixed effects results are not robust to controlling for yearly pollution levels. However, this could be coming from multicollinearity as opposed to omitted variable bias. More research should be carried out on the effects of different lengths of pollution exposure on wellbeing.

In column 4 of Table 8 I regress wellbeing on a 100 day lead of pollution. Pollution in 100 days should have no effect on wellbeing today. We see a statistically insignificant coefficient of 0. This is reassuring as it shows that pollution in 100 days does not have a significant effect on wellbeing today.

5 Conclusion

There are 3 key findings from this paper. Firstly, that there could be a small and significant instantaneous effect of air pollution on wellbeing. Secondly, that there are potentially large effects of long term exposure. Thirdly, that there are portions of the population that suffer more from pollution either through greater exposure or because particular characteristics make them more susceptible to the negative effects of pollution. I also explored the mechanisms through which an instantaneous effect of pollution may be acting.

In my preferred region effects specification I saw a small negative effect of pollution today on subjective wellbeing as measured in the USS data. I controlled for omitted variables using time fixed effects, regional fixed effects and controls for weather. My results are not significant but there is a large degree of measurement error potentially attenuating them. When wind direction and

Table 8: Robustness Checks

	Dependent variable:			
	Wellbeing			
	(1)	(2)	(3)	(4)
AQI	0.001429*	0.001427	−0.000849	
	(0.000717)	(0.000993)	(0.000634)	
Monthly Income	−0.000293***			
	(0.000024)			
Yearly Pollution			0.006980**	
			(0.002329)	
100 Day Lead				0.001227
				(0.000781)
MaxTemp	−0.010481	−0.009739	−0.009518	−0.009664
	(0.014364)	(0.010606)	(0.014566)	(0.014050)
MinTemp	0.015658	0.014793	0.013270	0.012410
	(0.010923)	(0.009284)	(0.010940)	(0.010889)
Rain	0.000753	0.000708	0.000770	0.000361
	(0.001432)	(0.001871)	(0.001450)	(0.001467)
WindSpeed	−0.006838	−0.006556	−0.007816	−0.008253*
	(0.004481)	(0.004671)	(0.004393)	(0.004533)
Individual FE	No	No	No	No
Time FE	Yes	Yes	Yes	Yes
Region FE	Yes	Yes	Yes	Yes
Observations	518,910	518,910	518,910	509,463
R^2	0.023053	0.017659	0.017689	0.017973
Adjusted R^2	0.009570	0.004104	0.004132	0.004175
Note:				*p<0.1; **p<0.05; ***p<0.01

temperature inversions are used as instruments for pollution the coefficients rise. However, the standard errors also increase meaning that the coefficients remain statistically insignificant.

The extension section briefly looked at the effects of longer term pollution on wellbeing. These results showed a large effect of monthly pollution on wellbeing. Future research could concentrate on finding the length of exposure to pollution that causes large drops in wellbeing.

In the heterogeneity analysis section I show that the effects of pollution differ by individual characteristics. Pollution has a larger effect on men than it does on women. This may be because of physiology but also perhaps because more men work in and travel to areas that have higher pollution levels. We also see that pollution has negative and significant effects on younger age groups but doesn't have an effect for the over 65s. Again I hypothesise that this is because of exposure as opposed to physiology since most studies show that older people are affected more.

I have also shown that the effects of pollution are not acting through a noticeable health channel. Pollution did not have a significant effect on health satisfaction or have larger effects for those with asthma. This points to pollution affecting wellbeing in a more complex way. Further research could be done into this area since it is important for effective policy.

There are also many unanswered questions. What explains the large effects we see for monthly averages of pollution? What would the effects be with a hedonistic wellbeing measure? What would the effects be if we could properly assign pollution to an individual? More research can be done into this area that will impact policy and the way we think about air pollution.

This paper adds to the evidence that pollution, even at the relatively low levels seen in developed countries today, is dangerous and is affecting wellbeing. Beyond this, the results here warn against overlooking the nuanced effects of contemporaneous pollution exposure in favour of an exclusive focus on long-term health implications. This is particularly valuable as the impact of short-term pollution is demonstrably unbalanced across societal groups. A hybrid research approach, combining the study of long-term implications and the contemporaneous effects looked at here, will be crucial to the successful evolution of environmental policy looking forward.

Anderson, H. R., Favarato, G. and Atkinson, R. W. (2013) ‘Long-term exposure to outdoor air pollution and the prevalence of asthma: Meta-analysis of multi-community prevalence studies’, *Air Quality, Atmosphere & Health*. Springer, 6(1), pp. 57–68.

Balakrishnan, K. *et al.* (2019) ‘Air pollution in asia and the pacific: Science-based solutions’.

Bondy, M., Roth, S. and Sager, L. (2018) ‘Crime is in the air: The contemporaneous relationship between air pollution and crime’. IZA Discussion Paper.

Chay, K. Y. and Greenstone, M. (2005) ‘Does air quality matter? Evidence from the housing market’, *Journal of political Economy*. The University of Chicago Press, 113(2), pp. 376–424.

Department for Environment, Food and Rural Affairs (2020) ‘Automatic urban and rural network’. doi: https://uk-air.defra.gov.uk/data/data_selector.

Deryugina, T. *et al.* (2019) ‘The mortality and medical costs of air pollution: Evidence from changes in wind direction’, *American Economic Review*, 109(12), pp. 4178–4219.

Dolan, P. and Laffan, K. (2016) ‘Bad air days: The effects of air quality on different measures of subjective well-being’, *Journal of Benefit-Cost Analysis*. Cambridge University Press, 7(1), pp. 147–195.

Gaure, S. *et al.* (2019) *Linear group fixed effects*. doi: https://github.com/sgaure/lfe.

Global Modeling and Assimilation Office (GMAO) (2015). Greenbelt, MD, USA, Goddard Earth Sciences Data; Information Services Center (GES DISC). doi: 10.5067/SUOQESM06LPK.

Graff Zivin, J. and Neidell, M. (2012) ‘The impact of pollution on worker productivity’, *American Economic Review*, 102(7), pp. 3652–73.

Griliches, Z. and Hausman, J. A. (1986) ‘Errors in variables in panel data’, *Journal of econometrics*. Elsevier, 31(1), pp. 93–118.

Guarnieri, M. and Balmes, J. R. (2014) ‘Outdoor air pollution and asthma’, *The Lancet*. Elsevier, 383(9928), pp. 1581–1592.

He, J., Liu, H. and Salvo, A. (2019) ‘Severe air pollution and labor productivity: Evidence from industrial towns in china’, *American Economic Journal: Applied Economics*, 11(1), pp. 173–201.

Hemond, H. F. and Fechner, E. J. (2015) ‘Chapter 4 - the atmosphere’, in Hemond, H. F. and Fechner, E. J. (eds) *Chemical fate and transport in the environment (third edition)*. Third Edition. Boston: Academic Press, pp. 311–454. doi: https://doi.org/10.1016/B978-0-12-398256-8.00004-9.

Knight, S. J., Howley, P. and others (2017) ‘Can clean air make you happy? Examining the effect of nitrogen dioxide (no2) on life satisfaction’, *Health, Econometrics and Data Group (HEDG) Working Papers*. HEDG, c/o Department of Economics, University of York, 17(08).

Lavy, V., Ebenstein, A. and Roth, S. (2014) *The impact of short term exposure to ambient air pollution on cognitive performance and human capital formation*. National Bureau of Economic Research.

Lelieveld, J. *et al.* (2015) ‘The contribution of outdoor air pollution sources to premature mortality on a global scale’, *Nature*. Nature Publishing Group, 525(7569), pp. 367–371.

Lichter, A., Pestel, N. and Sommer, E. (2017) ‘Productivity effects of air pollution: Evidence from professional soccer’, *Labour Economics*. Elsevier, 48, pp. 54–66.

Lippmann, M. and Leikauf, G. D. (2020) *Environmental toxicants: Human exposures and their health effects.* John Wiley & Sons.

Met Office (2012) 'Met office integrated data archive system (midas) land and marine surface stations data (1853-current)'. NCAS British Atmospheric Data Centre. doi: http://catalogue.ceda.ac.uk/uuid/220a65615218d5c9cc9e4785a3234bd0.

Sager, L. (2019) 'Estimating the effect of air pollution on road safety using atmospheric temperature inversions', *Journal of Environmental Economics and Management.* Elsevier, 98, p. 102250.

Saunders, B. M. *et al.* (2019) 'Spatial variability of fine particulate matter pollution (pm2. 5) on the london underground network', *Urban Climate.* Elsevier, 30, p. 100535.

Stock, J. H. and Watson, M. W. (2015) *Introduction to econometrics.*

University of Essex (2019) 'Understanding society: Waves 1-9, 2009-2018 and harmonised bhps: Waves 1-18, 1991-2009'. UK Data Service. doi: http://doi.org/10.5255/UKDA-SN-6614-13.

US Enviromental Protection Agency (2018) *Technical assistance document for the reporting of daily air quality - the air quality index (aqi).* Office of Air Quality Planning; Standards.

White, M. P. *et al.* (2017) 'Natural environments and subjective wellbeing: Different types of exposure are associated with different aspects of wellbeing', *Health & place.* Elsevier, 45, pp. 77–84.

World Health Organization (2006) 'Air quality guidelines: Global update 2005: Particulate matter, ozone, nitrogen dioxide, and sulfur dioxide'. World Health Organization.

6 Appendix

EPA AQI Formula

$$I = \frac{I_{high} - I_{low}}{C_{high} - C_{low}} (C - C_{low}) + I_{low} \tag{7}$$

Where I is the Air Quality Index, C is the pollutant concentration, C_{low} is the concentration breakpoint that is $\leq C$, C_{high} is the concentration breakpoint that is $\geq C$, I_{low} is the index breakpoint corresponding to C_{low} and I_{high} is the index breakpoint corresponding to C_{high}.

Wellbeing Questions

- Have you recently been able to concentrate on whatever you're doing?
- Have you recently lost much sleep over worry?
- Have you recently felt that you were playing a useful part in things?
- Have you recently felt capable of making decisions about things?
- Have you recently felt constantly under strain?
- Have you recently felt you couldn't overcome your difficulties?
- Have you recently been able to enjoy your normal day-to-day activities?
- Have you recently been able to face up to problems?
- Have you recently been feeling unhappy or depressed?
- Have you recently been losing confidence in yourself?
- Have you recently been thinking of yourself as a worthless person?
- Have you recently been feeling reasonably happy, all things considered?

Table 9: Monthly Income Regression

	Dependent variable:
	wellbeing
monthly_income	-0.0003^{***}
	(0.00002)
MAX_AIR_TEMP	-0.010
	(0.014)
MIN_AIR_TEMP	0.015
	(0.011)
PRCP_AMT	0.001
	(0.001)
MEAN_WIND_SPEED	-0.007
	(0.004)
Observations	518,910
R^2	0.023
Adjusted R^2	0.010
Residual Std. Error	5.460 (df = 511846)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 10: Lagged Regression Estimates

	Dependent variable:	
	wellbeing	
	(1)	(2)
AQI	0.0002	−0.002
	(0.001)	(0.001)
lag_1_aqi	0.003**	0.002**
	(0.001)	(0.001)
lag_2_aqi	−0.001	−0.002*
	(0.001)	(0.001)
MAX_AIR_TEMP	−0.010	
	(0.015)	
MIN_AIR_TEMP	0.014	
	(0.011)	
PRCP_AMT	0.001	
	(0.001)	
MEAN_WIND_SPEED	−0.007	
	(0.004)	
lag_3_aqi		−0.00001
		(0.001)
lag_4_aqi		0.001
		(0.001)
lag_5_aqi		0.001
		(0.001)
lag_6_aqi		−0.0003
		(0.001)
Observations	517,783	515,460
R^2	0.018	0.551
Adjusted R^2	0.004	0.450
Residual Std. Error	5.476 (df = 510723)	4.072 (df = 420626)
Note:		*p<0.1; **p<0.05; ***p<0.01

Table 11: Full Ethnicity Regression

	Dependent variable:
	wellbeing
AQI	0.002
	(0.006)
ethnicityarab	0.263
	(0.603)
ethnicityasian__bangladeshi	0.235
	(0.387)
ethnicityasian__chinese	0.004
	(0.679)
ethnicityasian__indian	−0.047
	(0.633)
ethnicityasian__other	−0.178
	(0.411)
ethnicityasian__pakistani	0.372
	(0.451)
ethnicityblack__african	−1.221*
	(0.623)
ethnicityblack__carribbean	−0.040
	(0.601)
ethnicityblack__other	0.083
	(1.536)
ethnicitymixed__african	−0.108
	(0.847)
ethnicitymixed__asian	1.020
	(0.590)
ethnicitymixed__caribbean	0.388
	(0.442)
ethnicitymixed__other	0.581
	(0.505)
ethnicitywhite__english	−0.255
	(0.456)
ethnicitywhite__gypsy	2.797
	(1.831)

Table 12: Age Binned Results

	Dependent variable:
	wellbeing
AQI	0.004**
	(0.001)
age_bins(44,73]	0.229**
	(0.088)
age_bins(73,102]	0.603***
	(0.167)
MAX_AIR_TEMP	−0.009
	(0.014)
MIN_AIR_TEMP	0.014
	(0.011)
PRCP_AMT	0.001
	(0.001)
MEAN_WIND_SPEED	−0.006
	(0.004)
AQI:age_bins(44,73]	−0.003
	(0.002)
AQI:age_bins(73,102]	−0.018***
	(0.003)
Observations	518,910
R^2	0.018
Adjusted R^2	0.005
Residual Std. Error	5.474 (df = 511842)
Note:	*p<0.1; **p<0.05; ***p<0.01