SIEPR Discussion Paper No. 15-020

The Elasticity of Air Quality: Evidence from Millions of Households across the United States

By

Christos Makridis

Stanford Institute for Economic Policy Research
Stanford University
Stanford, CA 94305
(650) 725-1874

The Elasticity of Air Quality: Evidence from Millions of Households Across the United States

Christos Makridis, Stanford University*

April 30, 2015

PRELIMINARY AND INCOMPLETE, do not distribute without permission.

Abstract

This paper estimates the elasticities of substitution between air quality and non-durables consumption, housing services, and leisure in the United States. First, I develop the most comprehensive database to date containing measures of household-level consumption, leisure, and demographics, together with county-level measures of weather, air quality, pollution, and economic development throughout the entire United States between 2005-2010. Second, I formulate and estimate a structural model allowing for nonseparable interactions between air quality and non-durables consumption, housing services, and leisure equal to 1.5, .62, and .32, respectively, and are identified from county-industry-specific deviations in air quality from the county averages after conditioning on shocks common to all counties within a state. Prior literature ignored the ways in which households are able to best respond to changes in environmental amenities through cross-substitution. The multi-dimensionality of the micro-data allows me to characterize heterogeneity in tastes for air quality based on brackets of educational attainment, income, age, and exposure to pollution. Third, applying my elasticity estimates to an analog of the EPA's evaluation of the Clean Air Act Amendments of 1990, I find that the benefits are many orders of magnitude lower because households are able to substitute across different private goods and services.

*Stanford University, Working Paper. Email: cmakridi@stanford.edu; LinkedIn: www.linkedin.com/in/christosmakridis/. PRELIMINARY AND INCOMPLETE, do not cite. I am grateful to detailed discussions with Han Hong, Nicolai Kuminoff, Kerry Smith, and Constantine Yannelis, as well as Lawrence Goulder, Charles Kolstad, Mar Reguant, and seminar participants at Arizona State University, Chapman University, and Stanford University (Applied Microeconomics Lunch and SEEPAC). Partially funded by the NSF Graduate Research Fellowship.

1. Introduction

> *[...] the dynamic interactions created by consumption externalities are of some independent theoretical interest; they may also be of considerable practical importance. (Sandmo, 1980, p. 807).*

Pollution in the United States has been cut in half over the past forty years, displayed below in the left panel of Figure 1, driven by a battery of factors, ranging from increased energy efficiency and productivity (Makridis, 2014) to regulation (Shapiro and Walker, 2014), concurrent with steady rises in both leisure and per capita consumption, displayed in the right panel of Figure 1. While the net benefits of environmental policy hinge crucially on quantitative estimates about household's behavioral cross-substitution elasticities among not only market goods, but also between and non-market goods, all environmental models for cost-benefit evaluation and macroeconomic analysis impose the restrictive assumption that non-market and market goods are perfect substitutes with each other. Mathematically, this takes the form of additive separability between environmental amenities and private goods and services, e.g.,

$$U(c; S) = u(c) + \Phi(S)$$

where c and S denote consumption and air quality, respectively, and where u and Φ are both increasing and weakly concave. While the assumption is unambiguously rejected in the data, current models have deferred to it in the absence of quantitative estimates of behavioral elasticities between market and non-market goods.[1] The purpose of this paper is to estimate these structural elasticities, characterize new sources of omitted variables bias in canonical hedonic regression analysis, and conduct a simple welfare exercise on the benefits of the Clean Air Acts using the estimated elasticities.

While a rich microeconometric literature has emerged to estimate willingness to pay for air quality, all cost-benefit analysis and macroeconomic models with externalities assume that environmental quality is additively separable in household preferences, i.e. changes in the demand for an externality occur independently of changes in the demand for private goods (Davis and Whinston, 1962). The assumption of additive separability took root because Diamond and Mirrlees (1973)

[1] There is an emerging body of empirical evidence, e.g., the effects of pollution on defensive investments (Greenstone et al., 2013), infant mortality (Chay and Greenstone, 2003; Currie and Neidell, 2005), labor productivity (Zivin and Neidell, 2012; Hanna and Oliva, 2015), and health (Moretti and Neidell, 2009; Neidell, 2007; Schlenker and Walker, 2012) and human capital (Bharadwaj et al., 2014; Sanders, 2012; Neidell and Zivin, 2013). See Currie et al. (2014) for a review of the literature on the link between health and pollution. See Guojun (2013) for a regression discontinuity approach from the 2008 Beijing Olympic games (their requirement to meet air quality standards) that found that a 10% decrease in mean PM10 concentrations decreased monthly cardiovascular mortality by 13.6% and Van Hee and Pope (2012) for a randomized medical experiment.

Figure 1: Pollution, Consumption, and Leisure in the United States, 1962-2008
Notes.–Source: EPA, OECD, ATUS, World Bank. The left panel plots the historical trend of particulate matter of 10 micrograms per cubic meter or less (PM10) using data initially collected from Smith (2012). The right panel plots the historical trends of average leisure and per capita consumption (obtained from the World Bank in 2005 constant dollars). Leisure is measured according to the second definition in Aguiar et al. (2007) (activities providing direct utility), including: measure 1 terms and time in sleeping, eating, personal care, and time eating out. Microdata for leisure is collapsed using age, male, race, gender, and number of children frequency weights.

argued it was necessary to rule out unstable equilibria.[2] However, nonseparable interactions between private and non-market goods/services are not anomalies. In fact, there is a robust negative (positive) relationship between air quality and consumption (leisure) (see Figure 2 in Section 4). Interactions between market and non-market goods matter. For example, households with lower tastes for air quality might locate in larger counties with lower air quality because they prefer to have access to less expensive and more diverse consumption goods.[3] Similarly, households with higher tastes for air quality might locate in smaller and less commercial counties because they engage in more outdoor recreational activities. These preferences seem to be fairly stable over

[2]Castle (1965) and Buchanan and Kafoglis (1963) discovered cases of private behavior not being Pareto-inefficient. Diamond and Mirrlees (1973) were concerned with ruling out these "anomalies"—instances where, for example, corrective taxes could induce more of the externality—by deriving some conditions under a narrow class of preferences (linear in income); the slope of the aggregate compensated demand function was not guaranteed to be negative Smith and Carbone (2008). Yet, these were not anomalies—once final product prices are adjusted to include marginal damages and the effects of residuals on production and factor input costs (Kneese et al., 1970), then nonseparabilities can yield these types of results. For example, Sandmo (1980) found that quantity "anomalies" can occur even without income effects (e.g., even when utility is linear in income). Since externalities imply that their effects on demand are not explicitly taken into account by a decentralized equilibrium, instability may well occur and induce nonseparable interactions with private goods and services. While popular applications of carbon taxes in general equilibrium (e.g., tax targeting) use Sandmo (1975) and Kopczuk (2003) as their framework, Sandmo (1975) (p. 92-93) acknowledges that marginal damages are not independent of income and relative price effects under externality targeting in general equilibrium and Kopczuk (2003) (p. 84) says "this conclusion [the generality of the principle of targeting] may break down when issues involving tax avoidance, evasion, and administrative cost are introduced. The approach also ignores general equilibrium considerations." Put differently, within the context of the macroeconomics of labor supply, additive separability implies that households will demand the same level of leisure in the presence of a rise in marginal tax rates. However, because market prices exist for labor supply, additive separability is not as restrictive of an assumption in the case of consumption/leisure since households explicitly choose them within the marketplace. See Klaiber and Smith (2012) for more underlying motivation and an empirical example.

[3]Handbury and Weinstein (2014) assemble the most detailed database to date containing barcode-level data across cities in the U.S. and find that heterogeneity bias is the source of the historical consensus that prices are higher in larger cities.

time.[4]

Unfortunately, there is neither empirical evidence nor theory on the elasticity of substitution between amenities and private goods and services. These elasticities matter greatly because they govern household's dynamic behavioral responses to policy through cross-substitution among private goods. For example, \$1.7 of the \$2 trillion estimated net benefits in the Environmental Protection Agency's (EPA, p. 3, 2011) Second Prospective evaluation of the Clean Air Act Amendments are driven by a failure to model household's behavioral responses and relative preferences for environmental amenities over private goods/services.[5] Given that the 2014 Clean Power Plan reaches similar conclusions as the Second Prospective evaluation—estimated benefits between \$55-93 billion by 2030—establishing credible estimates of parameters governing behavioral responses is a prerequisite to further analysis.[6] Policy interventions that affect the provision of non-market goods will necessarily affect household's demand for market goods since they change the relationship between marginal willingness to pay (MWTP) and marginal cost for other market goods—that is, if costs of environmental compliance rise, and subsequently affect the relative price of consumption, it is now relatively more expensive to consume market goods over non-market goods, like air quality. Households can only choose their desired quantity of non-market goods imperfectly through decisions over where to work and live—that is, by selecting into a geographical location with specified amenities—and the elasticities between market and non-market goods are precisely the parameters the pin down household's behavioral responses to changes in the provision of amenities. Ignoring the interaction between market and non-market goods can lead to gross underestimates of the excess burden of environmental policy in general equilibrium.[7]

Building the most comprehensive database to date containing information on household-level consumption, leisure, demographic attributes, together with county-level information on air quality, pollution, and weather, I exploit spatial and intertemporal variation across counties in the entire United States in order to recover preferences over air quality. Fusing the hedonics literature on equilibrium and labor market sorting literature (Kuminoff et al., 2013) with the quasi-experimental

[4]Although the level of pollution has declined across the United States over the past four decades, the relative rank among counties has stayed about constant. In the Appendix, I show that counties with higher total suspended particulates (micrograms per cubic meter) and ozone (parts per million) in 1990 are also the same counties that tend to have higher pollution in 2010.

[5]Similarly, in their prospective study of "The Benefits and Costs of the Clean Air Act 1990-2010: EPA Report to Congress", 90% of the estimated \$110 billion in benefits were driven by reductions in mortality.

[6]http://www2.epa.gov/carbon-pollution-standards/fact-sheet-clean-power-plan-benefits#benefits

[7]Using Goulder and Williams III (2003) as a benchmark, Carbone and Smith (2008) found that even a small 5% environmental tax could understate the excess burden by nearly 100% when leisure and air quality are complements and overstate it by 50% when they are substitutes. The severity of the bias depends on how far away reality (the nonseparable case) is from the assumption of additive separability. Berry et al. (2014) emphasizes that complementarities may easily arise between a non-market good and unobserved variables, which in this case are other rationed private goods/services.

literature (Greenstone and Gayer, 2009), my structural life-cycle model refines upon Roback (1982) by including nonseparabilities between market and non-market goods, deriving equilibrium conditions over household's locational choice and consumption-labor-housing decisions (Kuminoff et al., 2010). Allowing for interactions between non-market and market goods leads to lower estimates of the willingness to pay for air quality. I characterize three new sources of upwards bias in traditional hedonic models. These models typically use county-level data, ignore consumption/leisure decisions, and regress the aggregate property value or annual earnings on pollution, rather than the price—all of which induce bias. Unlike the Roback-Rosen model, my elasticities are derived from a general equilibrium model containing nonseparable preferences.

The Great Recession provides significant variation in air quality due to the large, but heterogeneous, declines in output across counties.[8] In particular, not only do different counties tend to have different industrial bases, which were impacted heterogeneously by the Great Recession, but also the same county is observed at different points in time with different levels of air quality. Of course, there are two major sources of endogeneity that confront a naive least squares estimator. First, aggregate fluctuations affect both time use (Aguiar et al., 2013) and consumption (Mian et al., 2013) at the household-levels; ignoring these unobserved shocks will generally induce upward bias estimates of willingness to pay (WTP) due to the shock's negative correlation with pollution and housing values. Second, household's locational sorting is driven by their preference for local amenities, which includes air quality (Kuminoff et al., 2013; Banzhaf and Walsh, 2008; Epple, 1987). Motivated by insights from the quasinatural experimental literature (Greenstone and Gayer, 2009), I address these challenges in two ways. First, I introduce detailed fixed effects specifications and nonparametric controls for local demand shocks such that my parameters are identified from the county-industry-specific deviations in air quality from county averages after adjusting for common shocks across all counties within a state. Second, I instrument air quality with wind speeds, leisure with measures of maximum temperature and interactions with fixed effects on the individual's age, housing services with measures of snowfall, and non-durables consumption with the interaction between electricity expenditures and fixed effects over the year in which the household's home was built. I subject these instruments to a battery of robustness checks, introduce additional instruments that leverage household-level variation, and confirm the results through a series of reduced form regressions whose only endogenous regressor is air quality.

[8]While my Census data is in annual frequency (only starting in 2005), this is not a limitation since I am *not* estimating the effects of pollution on health, but rather the nonseparable effects on intratemporal consumption-leisure substitution; the presence of intertemporal effects would accentuate my results and is another topic of my current work. To put my sample period in perspective, Chay and Greenstone (2005, 2003) exploit sharp changes in TSP between 1980-1982; Currie and Neidell (2005) use 1989-2000; Moretti and Neidell (2009) use 1993-2000 in Los Angeles; Neidell (2007) uses 1989-1997 in Los Angeles; Zivin and Neidell (2012) use 2009 and 2010 in Central Valley of California (orange pickers); **?** use 1986-2012.

By providing estimates of the elasticity between air quality and consumption & leisure, my paper enriches three main veins of research at the intersection of public and macro economics. First, building a novel database of household and county -level outcomes, I illustrate that nonseparability is not just a theoretical pursuit. An attractive feature of the EPA air quality index is that it avoids collinearity among different pollutants, allowing me to develop elasticity estimates that are representative for representative air quality. Second, I provide a new identification strategy for recovering causal relationships between market and non-market goods by exploiting quasinatural variation in climate and weather variables, along the lines suggested by Zivin and Neidell (2014) and Neidell and Zivin (2013). I apply my benchmark model across different subsets of the population to test for heterogeneity in preferences over air quality. I find that there is little heterogeneity across age and years of schooling, but a reasonable share when partitioning by income and exposure to pollution. These latter results point macro economists studying heterogeneity (Guvenen, 2012; Guvenen and Smith, 2014; Heathcote et al., 2010b, 2014) to focus on applying these methods for environmental economics and modeling the presence of state-dependent utility (as in the health literature, e.g. (Finkelstein et al., 2013)). Obtaining estimates of these elasticities qualitatively and quantitatively affects a suite of general and partial equilibrium welfare analysis (Carbone and Smith, 2008, 2013; Klaiber and Smith, 2012).[9] Third, I show how these elasticities can be used for partial equilibrium analysis, like those involved in the Prospective Evaluation of the CAAA. Specifically, using my estimated elasticities, I find that the CAAA provided \$70 billion in benefits—much lower than the \$2 trillion that the estimated.

Section 2 discusses relevant literature. Section 3 presents a structural model for disciplining the elasticity estimates and mapping them into macroeconomic models. Section 4 describes the data sources and presents both the methodology and results involving the imputation of non-durables consumption and leisure. Section 5 provides motivating evidence over additional identification problems in the canonical hedonic model. Section 6 estimates the structural model and subjects the results to a battery of robustness checks. Section 7 implements a welfare analysis of the benefits/costs of the rise in air quality over the Great Recession by applying my estimated elasticities in a calibrated model with simulated counterfactual distributions for consumption, housing, and leisure. Section 8 concludes.

[9] For example, the elasticity of substitution between consumption and environmental amenities determines whether environmental degradation is decreasing or increasing in wealth (Shibayama and Fraser, 2014)—a literature on the Environmental Kuznets Curve dating back to Grossman and Krueger (1995).

2. Relevant Literature

The paper complements three streams of literature. The first is an empirical literature on the effects of pollution on individuals' behavior and health. Neidell (2007) exploits random variation in the EPA's deterministic selection rule for issuing smog alerts. To the extent that days just above or below the smog threshold do not vary systematically with time use, their regression discontinuity in Southern California identifies the effect of pollution on leisure time use, in particular a reduction of 6-13%. He also finds that a .01 ppm decrease yields savings of about $417,717 per smog alert season (hospital costs of asthma for children over a 5-day period). Sexton (2011) uses American Time Use Survey (ATUS) data and finds that individuals reduce their time outdoors on smog alert days by 18 percent (21 minutes). Hanna and Oliva (2015) use a triple difference estimator exploiting wind speeds at different altitudes around Mexico City, which experienced the closure of a large refinery, finding that a 20% increase in SO2 led to a 4.2% decline in average hours worked. Currie and Neidell (2005) use individual-level data to study the effect of declines in three pollution criteria on infant mortality, finding that the reductions if carbon monoxide over the 1990s saved approximately 1000 infant lives in California.[10] Greenstone et al. (2013) study the NOx cap-and-trade program and decompose household's willingness to pay for air quality into three components: the effect of pollution on productive work time (valued by the wage rate), the cost of defensive investments (valued by the market price), and the dis-utility of worse health (valued by dollars). Importantly, households undertake a series of defensive expenditures in response to worse air quality, meaning that not only are many WTP studies obtaining too low estimates, but also pollution has important nonseparable contributions on consumption (i.e. defensive medical expenditures).[11]

The second is a more structural and theoretical literature on valuing environmental amenities by observed household's observed behavior introduced by Rosen (1974) and Roback (1982). While Smith and Huang (1995), and more recently Kuminoff et al. (2013), are both excellent surveys of this literature, there are two important contributions to mention. Recognizing that households face mobility costs, Bayer et al. (2009) modify the Roback model and show that typical estimates of willingness to pay are downward bias since they ignore variation in housing prices and wages that is correlated with adjustment costs. Motivated by the concern for omitted variables bias, Chay

[10] See Currie et al. (2014) for a more comprehensive review on the health effects of pollution.

[11] Of course, if households fully internalized the costs of low air quality through defensive expenditures, then additive separability would be a reasonable assumption since households would trade off consumption and air quality perfectly. However, this is not empirically the case. For example, Mansfield et al. (2006) document that many averting behaviors are not observed by the econometrician; even though the American Time Use Survey has been integral in understanding the intertemporal aggregate allocation of activities, disentangling the sources of variation that contribute to an individual spending less time outdoors has not yet been dealt with. Furthermore, there are a suite of issues that would remain, in particular the distorted or unpriced differences in quality among defensive expenditures. (For example, medicine is subsidized, so a simple "market value" is incomplete at best.)

and Greenstone (2005) used quasinatural variation in non-attainment standards of the Clean Air Acts, finding that traditional estimates of willingness to pay were too low. The third is a literature on nonseparability between market and non-market goods. Corlett and Hague (1953) were the first to consider an environment where leisure is complementary to some good and analyzed the competing income and substitution effects that may cause the demand for that good to rise or fall in the presence of a change in prices. Though environmental quality is a non-market good whose prices are not reflected in privately consumed goods, the property of weak complementarity (Maler, 1974)—the point at which there is a cutoff price such that, at zero demand for the amenity, the marginal utility is zero—has been used to identify prices on non-market goods.[12] To the extent that policy intervention affect the reallocation of labor supply and/or consumption, the marginal tradeoffs for market goods will be affected by changes in environmental amenities; see Slesnick (1998) for a more general discussion. Sandmo (1975) took up the broader issue of nonseparability again and showed that equilibria in these economies are not anomalies—as labeled by Diamond and Mirrlees (1973)—but plausible solutions. While analysis of partial demand systems in the tradition of Hanemann and Morey (1992) might be reasonable for purely partial equilibrium exercises, the interaction between market and non-market goods matters much more in general equilibrium.[13] Flores and Graves (2008) illustrate that ignoring the endogeneity of labor-leisure decision-making leads to gross mischaracterizations of the optimal supply of public goods and the costs/benefits of environmental policy intervention. By treating labor supply as inelastic, models with externalities fail to identify the correct level of income since households may fail to think about endogenous changes in their labor-leisure decisions when consuming the public good. Smith et al. (2003) make a related argument in the context of estimates of the value of a statistical life. Carbone and Smith (2008), Carbone and Smith (2013), and Makridis (2014) all show that these nonseparable interactions have big effects in general equilibrium.

3. The Demand for Air Quality

The basic framework for understanding the demand for air quality is based on a simple refinement to the neoclassical growth model incorporating preferences for non-market goods. Households will choose market goods, defined as a triple $X_{j,t} = (C_{j,t}, L_{j,t}, H_{j,t})$, consisting of non-durables

[12] See Bockstael and McConnell (1993).

[13] Amiran and Hagen (2014) also study the wedge between willingness to pay and willingness to accept, arising from an aggregation problem when exploiting cross-sectional variation with consumers with heterogeneous preferences. They clarify an array of complexities that arise in the valuation of non-market goods, for example a potentially infinite willingness to accept when there are non-local restrictions on substitution (e.g., possibly mobility or other adjustment costs to locational sorting).

consumption, leisure, and housing services, indexed by location j in period t. Households are able to choose among j differentiated locations defined by different levels of environmental amenities (e.g., quality); there are no mobility costs.[14] The production side will remain simple since relative prices are exogenous when taken to the data. To keep the model simple, I omit subscripts on location and time. The crucial insight from the setup is the conversion of a spatially and time varying public good into a private good whose price can be identified using geographical-specific variation in local private goods/services, individual-level time allocation, and locational sorting decisions. While my preferences will not allow for an arbitrarily nonlinear price for housing and wages, they are sufficiently flexible to the extent that elasticities are constant across the demand function. Appropriate instruments will allow for consistent identification of preference parameters (Ekeland et al., 2004).

A. Households: Following Rosen (1974), and in particular Roback (1982) who imposes homogeneity in household preferences, suppose that households have preferences over and choose private consumption (C), leisure (L), housing, (H), and environmental quality (S) generated by a constant elasticity of substitution utility function

$$\begin{aligned} U(C,L,H,S) &= \exp(\rho)\Big\{\alpha_C \log\Big[\mu(g_C C)^{\phi} + (1-\mu)(g_S S)^{\phi}\Big]^{\frac{1}{\phi}} \\ &+ \alpha_L \log\Big[\gamma(g_L L)^{\psi} + (1-\gamma)(g_S S)^{\psi}\Big]^{\frac{1}{\psi}} \\ &+ \alpha_H \log\Big[\pi(g_H H)^{\zeta} + (1-\pi)(g_S S)^{\zeta}\Big]^{\frac{1}{\zeta}}\Big\} \end{aligned} \tag{1}$$

where U denotes utility, ρ denotes a preference ("taste") shock, g_C, g_L, g_H, and g_S denote technological trends that affect the value of consumption, leisure, housing, and environmental quality in preferences, α, μ, γ are share parameters, and ϕ, and ψ are elasticity parameters.,[15] Letting $X \in \{C, L, H\}$, $\omega \in \{\phi, \psi, \zeta\}$, and $\theta \in \{\mu, \gamma, \pi\}$, the elasticity of substitution between the private good/service and air quality is given by $\varepsilon_X \in (0, \infty) = -\mathrm{d}\ln(X/S)/\mathrm{d}\ln(U_X/U_S)$ and where $\omega = (\varepsilon_X - 1)/\varepsilon_X$ such that the three extreme cases are given by $\varepsilon_X \equiv 1/(1-\omega) \to 0$ $(\omega \to -\infty)$, $\varepsilon_X \equiv 1/(1-\omega) \to \infty$ $(\omega \to 1)$, and $\varepsilon_X \equiv 1/(1-\omega) \to 1$ $(\omega \to 0)$, which

[14] A reasonable theoretical concern is that individuals working and living in geographical locations sufficiently far away from each other might introduce measurement error air quality since the worker is implicitly consuming it in two locations. While the net effect is only likely to be attenuation since it will merely dull the signal-to-noise relationship between air quality and the market goods, I also conduct a robustness exercise in the empirical section leveraging only the subset of the sample that spends commutes less than half an hour to work, meaning that inference is only over individuals not susceptible to this potential identification problem; the results remain the same.

[15] Although the assumption of additive separability between consumption and leisure is widely applied in the elasticity of labor supply literature (e.g., Altonji, 1986), I address this shortcoming in two ways. First, I estimate a nonseparable version (between consumption and leisure) of preferences, setting the elasticity between the two according to Ziliak and Kniesner (2005). Running GMM on the intratemporal Euler provides similar results, although the expression cannot be neatly decomposed into a linear regression format. Second, the bias is unlikely to alter my results because greater complementarity between consumption and leisure will make consumption an even weaker substitute and leisure a greater complement. In other words, nonseparability seems to further illustrate my argument that air quality is not a perfect substitution for consumption or leisure.

imply perfect complementarity, perfect substitutability, and Cobb Douglas elasticities, respectively (equivalent for leisure). The private good/service and air quality are gross substitutes when $\varepsilon_X > 1$ ($\omega > 0$) and gross complements when $\varepsilon_X < 1$ ($\omega < 0$). The technological parameters (g_C, g_L, g_S) characterize underlying trends that affect household preferences.[16] These preferences are sufficiently parsimonious to capture the nonseparability between consumption-leisure and air quality, but also tractable enough through additive separability between the consumption and leisure aggregates in order to obtain closed form expressions that map into the data.[17] An important abstraction is that the triple $X = (C, L, H)$ is continuous, driven by the homogeneity of preferences after controlling for heterogeneity in individual-level observable tastes. The rationale for allowing the household's decision to take a continuous, rather than discrete, form arises from an indivisibility argument similar to Rogerson (1988).

B. Firms: Suppose that firms use capital and labor to produce a homogeneous output using a constant returns to scale technology

$$Y = F(K, 1-L) = K^{\theta}(1-L)^{1-\theta} \tag{2}$$

and where the wage and price of capital are equal to their marginal products: $w = F_{1-L}$ and $r = F_H$. Wages and housing rents map into the hedonic framework for inferring household's valuation of environmental amenities. Capital produces pollution, which reduces air quality, given by an arbitrarily defined concentration response function, $S = g(K)$, where S is decreasing in K.

C. Equilibrium: Under the assumption that the representative household maximizes their utility subject to a simple budget constraint where consumption is equal to labor income, households solve

$$V(K_t) = \max_{C_t, L_t, H_t, K_{t+1}} \{U(C_t, L_t, H_t; S) + V_{t+1}(K_{t+1})\}$$

subject to his budget constraint

$$C_t + p_t H_t + K_{t+1} = w_t(1 - L_t) + K_t(1 + r_t - \delta)$$

where p is the price of housing and the price of consumption is the numeraire. The production function and relatives prices are given by Equation 2. Dropping time subscripts, optimizing behavior

[16] For example, Hall and Jones (2007) suggest that the rise in healthcare spending can be explained by a saturation of marginal utility for consumption goods; likewise, Aguiar and Hurst (2007) suggest that households have experienced a steady increase in leisure since the 1950s. Importantly, trends in environmental quality (g_S)—governed plausibly by environmental policy (Shapiro and Walker, 2014)—will affect the value of consumption and leisure depending on the share and elasticity parameters.

[17] The preferences avoid the alternative log-linearization approach, which would introduce bias from higher order moments. Carroll (2001) and Ludvigson and Paxson (2001) showed that linearizations of the intertemporal Euler condition for estimating the intertemporal elasticity of substitution led to major sources of bias. Although Attanasio and Low (2004) suggest that linearizations will yield consistent estimates for the elasticity of intertemporal substitution, their results do not apply in this case for two reasons. First, they assume utility is isoelastic, whereas preferences here are non-separable with air quality and leisure. Second, they assume that the sample includes a sufficiently long time period, whereas that is tougher to do when combining demographic and county-level air pollution data.

implies that the following three equilibrium conditions hold

$$U_C(C,S) = \beta \mathbb{E}\left[U_{C'}(C',S')(1-\delta+r')\right] \tag{3}$$

$$U_L(L,S)/U_C(C,S) = w \tag{4}$$

$$U_H(H,S)/U_L(C,S) = p/w \tag{5}$$

together with the resource constraint: $Y = C + p_t H_t + K_{t+1} - (1-\delta)K_t$. When the household is not working, he allocates his time exclusively to leisure.[18] The first condition is the intertemporal Euler, which equates the marginal utilities of consumption over time, together with the externality that is a function of savings decisions since housing is a long-lived asset. The second condition is the intratemporal Euler on consumption-leisure, which equates the ratio of the marginal utilities of leisure to consumption with the wage. Equation 4 will provide a characterization of the equilibrium differential that allocates individuals across locations and compensates those who face lower consumption of environmental amenities. Similarly, the third condition is the intratemporal on housing-leisure, which equates the ratio of marginal utilities of housing and leisure to the price ratio.[19] The nonseparabilities in these equations allow me to identify a virtual price on S—that is, a price associated with air quality as if households could purchase it directly; see Perroni (1992) and Amiran and Hagen (2014) for theoretical justification.

D. Comparison to the Literature: There are a couple of conceptual issues relating to hedonic price theory that need to be contrasted with prior work in canonical hedonic models (Roback, 1982; Rosen, 1974). Equations 4 and 5 are the objects of endogenous sorting processes that undergird the equilibrium in local labor and housing markets. Kuminoff (2012) provides a unified model for households to sort across jobs and housing locations without one implying the other, whereas the canonical models assume that every time a person changes houses, they change jobs, and vice

[18] Marginal utilities are given by

$$U_H(C,L,S,H) = \alpha_H \pi (g_H H)^{\zeta-1} g_H \left[\pi (g_H H)^{\zeta} + (1-\pi)(g_S S)^{\zeta}\right]^{-1}$$

$$U_L(C,L,S,H) = \alpha_L \gamma (g_L L)^{\psi-1} g_L \left[\gamma (g_L L)^{\psi} + (1-\gamma)(g_S S)^{\psi}\right]^{-1}$$

$$U_C(C,L,S,H) = \alpha_C \mu (g_C C)^{\phi-1} g_C \left[\mu (g_C C)^{\phi} + (1-\mu)(g_S S)^{\phi}\right]^{-1}$$

[19] While housing is modeled as a flow—meaning that households are renters—modeling housing as an asset would induce an intertemporal Euler between housing services and consumption, instead of an intratemporal Euler on housing and leisure. Since I use cross-sectional data in this paper, adding another intertemporal arbitrage condition would not allow me to take it to the data. Quantitatively, the only modification that an intertemporal (rather than intratemporal) induces is that housing and air quality appear as the change in logs, rather than purely logs, and the rental rate of housing is included. To the extent that my identification strategy exploits exogenous variation, modeling the change in housing & air quality and/or omitting the the return on housing will not bias the model estimation.

versa. Given that housing and employment opportunities are fundamentally linked—for example, U.S. counties became less stratified by public goods provision and housing demographics as moving costs declined between 1850-1990 (Rhode and Strumpf, 2003)—obtaining unbiased estimates of preferences for local public goods hinges on how these decisions are jointly modeled. Specifically, Equation 5 captures the simultaneity of the labor-housing decision by equating their marginal utilities equal to the price ratio. While the coefficient on air quality in the regression is analogous to the Roback (1982) measure of willingness to pay, the difference in my model is that the coefficient is only an input into the aggregate willingness to pay for air quality, which requires auxiliary information on consumption. Kuminoff (2012) conducts a related exercise by estimating a structural sorting equilibrium model with heterogeneous workers.

E. Empirical Implementation: Applying the Kmenta (1967) approximation around the point $\omega = 0$, Equations 4 and 5 can be estimated using least squares.[20]

$$\log w = (\psi - 1) \log L + (1 - \phi) \log C + (\phi - \psi) \log S + \rho(X) + \epsilon \tag{6}$$

where ϵ includes the constants from T_L, T_C, and other unobserved heterogeneity (including measurement error), and $\rho(X)$ contain a vector of shocks (e.g., household, state, and/or county -level controls). Similarly, taking logs of both sides from Equation 5 and adding an error term yields

$$\log(p/w) = (\zeta - 1) \log H + (1 - \psi) \log L + (\psi - \zeta) \log S + \rho(X) + \nu \tag{7}$$

where ν contains the constant terms from the CES specification. While the simplification makes interpreting elasticities much more convenient, it also may introduce non-classical measurement error since mis-specification in the functional form of the equilibrium conditions is correlated with unobserved shocks to the outcome variable (i.e., by construction of the structural model). Aware of this possibility, I am experimenting with non-linear versions of the model that avoid the approximation, but make the results less stable and intuitive to interpret.[21] The elasticities (dlog$w$/dlog$S$) and (dlog(p/w)/dlog$S$) identify the labor demand and home capitalization effects. The intuition is that variation from Equation 4 is directly informative about household's trade offs between consumption and leisure, and thus the labor supply decision, whereas Equation 5 is directly informative about household's trade offs between housing and leisure, and thus the home ownership decision.

[20] While the nonlinear model could be estimated with method of moments, the database will still contain traces of measurement error from the imputation of consumption and leisure. Although my checks to compare actual (from the CEX and ATUS) with imputed values suggest that they are nearly indistinguishable, I defer to the more conservative strategy of keeping my equations linear for tractability, identification, and ease of interpretation. Another rationale is with respect to the ease of computation. Nonlinear models with millions of observations are very computationally difficult to estimate. Fixed effects specifications in these models are still an active research area.

[21] However, to the extent that these elasticities are identified by an "extensive" margin of sorting, the interactions will matter less and the approximation may suffice. Furthermore, the structural regressions are consistent with my reduced-form regressions.

While hedonic wage regressions have (to my knowledge) only been used to decompose the riskiness of different jobs using partial equilibrium methods (Viscusi and Aldy, 2003; Aldy and Viscusi, 2008; Viscusi, 1979), Equation 4 shows that a relationship between wages and air quality can be derived from a theoretically consistent structural general equilibrium model. The slope of the hedonic wage on air quality has the interpretation as a marginal change in air quality for the individual found at the baseline level (Bockstael and McConnell, 2007). The traditional Roback (1982) framework will take the difference between the estimated coefficient on pollution obtained from a regression of housing values on pollution and labor income on pollution. However, Equations 6 and 7 reveal the potential for bias since these conventional hedonic wage and housing regressions conflate the aggregate price times quantity with simply the price. That is, from the lens of the Roback model, the willingness to pay for air quality is given by the elasticity of air quality in Equation 7. There are other empirical concerns that traditional hedonic regressions face concerning the bundling of unobserved consumption/leisure into the error and the reliance on county, rather than household, -level data; these will be discussed in detail later. The value of a statistical life is the "population"s aggregate willingness to pay for an increase in one expected life saved (Bockstael and McConnell, 2007 p.219). Nevertheless, the interpretation of $\text{dlog}(p/w)/\text{dlog}S$ is: a 1% change in air quality induces a $(\psi-\zeta)$% decline in the ratio of the relative price of housing versus labor. For example, if a policy intervention increases air quality by 1% for N people, then the aggregate WTP would be $N \times (\psi - \zeta)$% and the value of a statistical life would be $(\psi - \zeta)\% \times$ risk of death. Getting the VSL right matters: of the \$2 trillion in estimated net benefits of the CAAA, a staggering \$1.7 trillion are attributed to reductions in mortality.

To my knowledge, there is no empirical evidence on the the elasticity of air quality—only willingness to pay. Most applications focus only on the income elasticity of demand, which is not equal to WTP unless it is also equal to the income elasticity of a public good's virtual price. However, there is little reason to suspect these are the same in the presence of other rationed goods and services, such as consumption and leisure, since the cross-substitution demand elasticities may be non-zero. To put this in perspective, the average household sacrifices over \$5,000 per year to consume non-market amenities in their geographical location (Bieri et al., 2014), bigger cities tend to have differences in availability of consumption goods (Handbury and Weinstein, 2014) and higher wages (Glaeser and Resseger, 2009; Davis and Dingel, 2014).[22] Cross-sectional variation in workers'

[22] Handbury and Weinstein (2014) use detailed barcode data on household-level transactions in 49 U.S. cities to properly measure prices of identical goods sold in comparable stores across these cities. While prior studies introduced an unobserved positive correlation between prices and city size, their paper finds that there are dramatic differences in product availability; this is fully consistent with the theme of nonseparability. Baum-Snow and Pavan (2012) document the city size wage premia and find that most of the variation is explained through returns in experience between small and large locations. Both Glaeser and Resseger (2009); Davis and Dingel (2014) find evidence of higher skills and productivity in larger cities due in part to agglomeration externalities.

equilibrium sorting decisions from Equations 3 and 4 provides a way of estimating preference parameters associated with environmental quality. Because econometricians infer willingness to pay for public goods based on the identifying assumption that the quality of a location-specific amenity is increasing in the price ("value") of its location-specific private goods (e.g., wages or housing rents), then understanding the feedback among the traded and non-market goods is crucial. Individuals choose their consumption of non-market goods only through their consumption of market goods; nonseparability is the only lens to facilitate such an analysis. Put together, these elasticities also imply a measure of WTP.

4. Consumption, Leisure, Climate, and Air Quality in the United States: A New Dataset

4.1. Descriptive Statistics

Through the American Community Survey, the Census Bureau recovers detailed household-level information across the entire United States at disaggregated geographical levels. Unfortunately, the Census does not contain information on aggregate consumption or leisure. To deal with this limitation, the final data set exploits auxiliary data from the Consumption Expenditure Survey and the American Time Use Survey to impute consumption and leisure in the Census, matched with county-level data on air quality and weather, discussed later in the section. The Census' wide coverage is vital since my identification strategy requires variation in regional housing and labor markets. That is, observing observationally equivalent workers in different locations reveals information about their valuation of environmental amenities. The summary statistics below provide a characterization of the cross-sectional and intertemporal variation between 2005-2010.

All of the traditional anecdotal facts about the Great Recession are evident in the established data set: consumption declined by 10%, both the mean and standard deviation of leisure rose by about 3%, housing prices declined by 10%, unemployment doubled, and pollution—whether measured as total suspended particulates or PM10—declined by about 20%. While the rise in consumption inequality (Aguiar and Bils, 2009) and time use (Aguiar et al., 2013) over the Great Recession are not new, the decline in pollution is stark. However, what remains unknown is how counties with higher air quality differ from those with lower air quality. Figure 2 plots the raw correlations in the data between county-level air quality and market "goods", such as consumption, leisure, wages, and housing values. Counties with greater air quality tend to have lower consumption and wages, but higher leisure and housing values, consistent with the hypothesis that environmental amenities are capitalized into both human and physical assets (e.g., time and houses).

Table 1: Summary statistics

	2005-2007		2008-2010	
	Mean	S.D.	Mean	S.D.
Individual				
Nondurable Consumption	10208.69	2560.94	9218.02	2309.84
Leisure (Hours)	6082.56	913.05	6198.69	950.86
Wage (Hourly	16.82	23.50	18.20	29.01
Years of Schooling	13.45	2.96	13.61	3.43
No. of Children	0.56	0.98	0.53	0.95
Age	47.32	15.74	47.74	15.86
Male	0.48	0.50	0.48	0.50
Bedroom	3.13	0.81	3.20	0.96
Housing				
Years Old	1967.38	21.59	1962.59	20.81
1-family Home Detached	0.87	0.34	0.87	0.34
Vacant	0.01	0.12	0.02	0.13
Housing Tenure	16.03	13.02	16.93	12.90
Housing Value	343099.02	491786.41	323360.62	475006.63
County Business				
Unemployment Rate	4.86	0.86	8.68	2.43
No. Employed, 000s	1678.95	2289.91	1626.20	2249.15
Payroll Exp., 000s	73298.88	105827.98	76223.98	112040.24
No. Establishments, 000s	107.04	145.53	106.88	146.57
Population	137628.07	39939.94	148515.87	47389.66
Environment				
Air Quality Index	250.20	14.50	255.09	13.49
Total Suspended Particulates	56.83	18.10	50.76	14.20
PM10	24.73	6.95	21.61	5.67
Fastest 2 Min. Wind	17.34	3.53	17.55	3.36
Fastest 5 Sec. Wind	21.53	4.80	22.81	4.73
Mean Resultant Wind	6.32	1.99	6.35	1.85
Mean Wind	7.87	1.87	7.76	1.77
Precipitation (in. to 100ths)	8.00	60.99	8.01	57.07
Snow (in. to 10ths)	0.13	2.17	0.20	2.55
Max Temp.	69.65	12.82	68.90	13.38
Min Temp	49.11	9.92	48.48	10.36
Observations	710088		715652	

Notes.–Sources: Environmental Protection Agency AirData's air quality index (AQI), the Census Bureau's American Community Survey (ACS), Consumption Expenditure Survey (CEX), and American Time Use Survey (ATUS). The table contains the means and standard deviations of the most relevant variables contained in the econometric specifications. Using the definition of non-durables from Attanasio and Weber (1995), it is the sum of food (home and away), alcoholic beverages, tobacco, services (e.g., repairs), heating, gasoline, transportation, electricity, water, fuel, personal care, clothing, footware, and rents. Using the definition of leisure from Aguiar and Hurst (2007), specifically their "Measure 1", it is the sume of socializing, passive and active leisure, volunteering, pet care, gardening. Housing values are self-reported from the ACS and are upwards biased to the extent homeowners are overly optimistic about the sale price of their property. Air quality, call it S, is transformed from the Environmental Protection Agency's (EPA) "AirNow" air quality index, call it $\tilde{S}$, by taking $S = 300 - \tilde{S}$, where 300 is a hazardous measure of air quality. With the transformation, higher values enter positively into utility. Total suspended particulates and particulate matter are measured in micrograms per cubic meter.

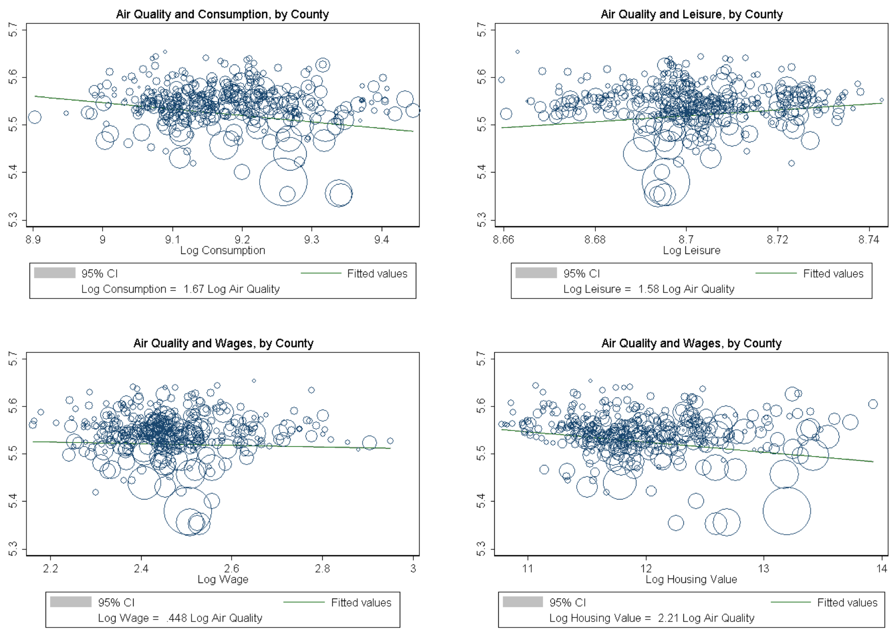


Figure 2: The Relationship between Air Quality and Market Goods, 2005-2010
Notes.–Sources: CEX, ATUS, ACS, EPA. The plots depict air quality with consumption, leisure, wages, and housing values. Each observation is a county-level average between 2005-2010 and the circles are weights based on the population of the county.

While not exogenous variation, the Great Recession is a suitable natural experiment for recovering preferences over environmental amenities because of the vast amount of reallocation and migration dynamics. As counties experienced different magnitudes and durations of labor and housing market shocks, individuals moved locations. For example, many households moved from one county to another for new employment opportunities and/or housing decisions (i.e., to move into a less expensive house). The demographic reallocation and resorting coincides with a massive decline in pollution (rise in air quality) that affected household's locational choices. To the extent that households value environmental amenities, holding all else constant, a homeowner with tastes for non-market goods would choose to locate in an area with better amenities. The density and time series properties of air quality are illustrated below in Figure 3.

4.2. Sources

4.2.1. Demographic Microdata and Economic Development

The Census Bureau began implementing the annual American Community Survey (ACS) starting in 2005 to offer a more frequent measure of demographic data across the United States, relative to its decennial census counterpart. The novelty of the ACS is its comprehensive nature (e.g., relative to the PSID), which is important for estimating an aggregate elasticity for the U.S. and studying heterogeneity in treatment effects. The ACS contains detailed household-level information on years

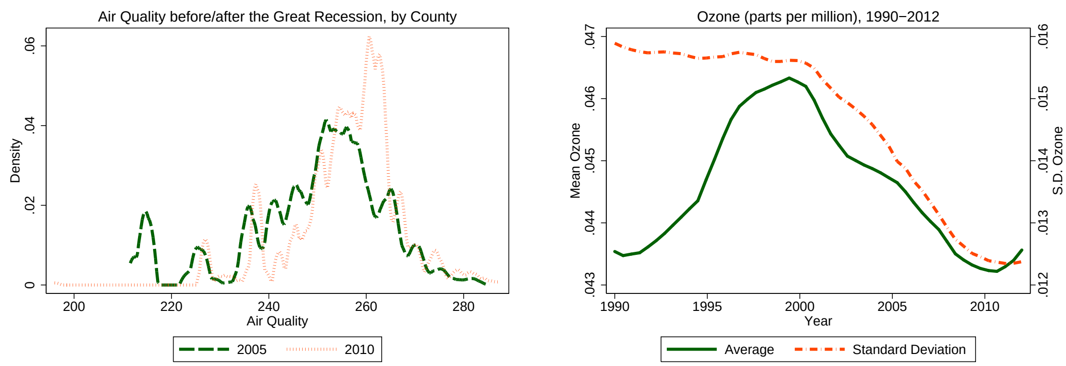


Figure 3: Distribution of Air Quality across Counties (2005/2010) and Average Ozone Pollution (1990-2014)

Notes.–Source: Environmental Protection Agency AirData's air quality index (AQI) and Annual Summary files. Using the same air quality indices as before, the left figure plots the distribution of counties' air quality levels (using the same transformed measure using a smoothed kernel density estimator and, using the annual summary file, the right figure plots the (weighted) average total suspended particulate mean and standard deviation over the past two decades. The averages are weighted by state-level civilian labor forces using Bureau of Labor Statistics state-level data. The TSP measure is computed as the 24-hour average level, but is also robust to the 1-hour average level. See https://aqs.epa.gov/aqsweb/codes/data/SampleDurationCodes.html for details.

of schooling, income, age and other demographics, race, labor force status, etc. Unfortunately, it poses two empirical challenges. First, its consumption and leisure data are not sufficiently detailed to credibly estimate an elasticity—that is, it only contains basic measures of electricity, fuel, and water consumption and hours worked. Second, the ACS identifies people within public use micro-data areas (PUMAs), rather than counties.[23] Because the EPA and NOAA data is disaggregated at the county-level, and the overlap between PUMAs and counties is incomplete, I use a Missouri Data Center sponsored geospatial algorithm to map PUMAs (5% micro-data sample) to counties based on the 2000 Census boundaries (for 2005-2008 in my sample) and 2010 boundaries (for 2009-2010). The mapping is driven by population and demographic data collected at the two decennial censuses at both the PUMA and county -levels.[24]

The Census also provides rich geospatial data that lets me capture the ruggedness of the terrain in any given county. Since topography is correlated with weather shocks, this form of unobserved heterogeneity is useful to control for to make my estimate more precise. Although there are many ways to construct such a measure,[25] I choose the 2D: 3D area ratio. Since the difference between the two is driven by the degree of ruggedness (e.g., mountains), the measure detects whether a county is in a valley or not.

[23] Public use micro-data areas (pumas) tend to be synonymous with FIPS codes. According to the Census Bureau (http://www.census.gov/geo/reference/pdfs/puma/FAQ_version2.pdf): "The puma code must consist of a 5-digit numeric code that is unique within the state. If the five-digit code includes the county FIPS and the number fits the other guidelines for assigning codes to new pumas, then this is an acceptable number to use."

[24] Results are not qualitatively different under slightly different thresholds, but I have been advised by specialists with Census data that this delineation is both reasonable and the best solution.

[25] For example, see: http://gis4geomorphology.com/roughness-topographic-position/.

Using Census County Business Pattern (CPB) data on county-level employment, establishments, and payroll expenditures, together with county-level population controls, I can further control for time varying county-specific productivity shocks. Since state-level employment proxies may be uninformative about the underlying county dynamics, these controls are important otherwise unobserved county-level economic shocks might be driving changes in consumption and leisure preferences, whereas a naive model might attribute it to differences in air quality.[26] The measure has a few important advantages for my application. First, it is scale independent, meaning that the ratio between the two matters—not an absolute level; this is crucial for exploiting the cross-sectional variation *across* counties. Second, it characterizes areas, rather than particular points; many alternative measures are designed for calculating the ruggedness of specific *points*. These county-level data are augmented by Bureau of Labor Statistics data on state-level employment, I can control for state-level productivity shocks proxied by employment outcomes, together with labor force and population controls. State-level changes in economic conditions might interact with county-level shocks in a way that controlling for one, but not the other, could confound estimation. For example, if county-level economic conditions are poor, and state-level conditions are poor, then the county, which may normally rely on additional assistance from the state government during business cycles to provide assistance, may suffer further; these economic shocks are correlated with consumption and leisure and would otherwise bias the elasticity.[27]

4.2.2. Air Quality and Weather: EPA and NOAA

The Environmental Protection Agency (EPA) reports an air quality index (AQI) ("AirNow"), call it $\tilde{S}$, that characterizes air quality within a geographical location, specifically counties. Monitors track the overall air quality with respect to the presence of different pollution criteria and rank the air with a score. An extremely attractive feature of this data is that, although it separately detects different pollutants, the air quality score is unitless and comparable across observations of different pollutants. Air quality values from complete monitor readings are the highest daily reading that they take. To deal with the difference frequency of the data, relative to the Census, I annualize its reported daily format.[28] I introduce a transformation, defining S as the new measure of air quality, taking $S = 300 - \tilde{S}$. The transformed measure implies that counties with higher values of S have better air quality. I use 300 as the normalization factor since any EPA air quality index above 300 indicates that the air is very hazardous for health. To the extent that the transformation introduces

[26] https://www.census.gov/econ/cbp/download/06_data/

[27] http://www.bls.gov/lau/rdscnp16.htm

[28] There are some valid concerns about averaging these values when they are only reported at a daily level. While there is no evidence pointing either way, Sieg et al. (2004) takes the average of the top 30 1-hour daily maximum readings at a given monitor within a year.

measurement error, an added benefit of my instrumental variables approach is that it corrects for the measurement error.[29] These air quality alerts are made publicly known to individuals within a local area and Auffhammer and Kellog (2011) find evidence that their message diffuses across the entire county that it is announced in.

The National Oceanic and Atmospheric Administration (NOAA) reports many measures of weather outcomes, ranging from wind speeds to precipitation. Aside from using wind speeds as part of my instrumental variable strategy, these weather outcomes are vital controls since air quality is highly correlated with air diffusion patterns and precipitation. Absent these controls, two confounders might induce bias. First, changes in consumption or leisure might be attributed to severe weather events, such as a hurricane or major flooding; these are unlikely to induce major bias since these extreme events are rare. Second, and more importantly, time invariant county-level unobservables are correlated with the underlying weather patterns. Even with fixed effects on counties, time varying weather shocks could interact with other unobservables in a way that induces bias.

4.2.3. Consumption and Leisure: CEX and ATUS

The Consumption Expenditure Survey (CEX) is the benchmark source for consumption data in the United States. The Bureau of Labor Statistics maintains and implements the survey to construct the consumer price index at different levels of aggregation. While the diary component is not very reliable, the interview component has been more robustly tested (Bee et al., 2012). Using the Interview portion of the CEX, I extract detailed annual information on household's non-durables expenditures, demographic characteristics, and household-level information on income and assets in order to construct a sample comparable to the ACS data set. Since survey questions are answered about a year that may have already been passed (e.g., households report consumption for overlapping years), I make the following adjustment. Consider a year t_τ with month m_τ where $\tau \in \{Reference, Calandar\}$ denotes whether the time is for the reference (interview) or calendar (actual) period. Since the reference period is always three months (1 quarter) ahead—that is, the reference period $m_{cal} + 3$ contains information about period m_{cal}—I subtract three from the reference months. All variables are deflated with the consumer price index. I follow the definition of non-durables from Attanasio and Weber (1995) as the sum of food (home and away), alcoholic beverages, tobacco, services (e.g., repairs), heating, gasoline, transportation, electricity, water, fuel, personal care, clothing, footware, and rents. Nondurables in the next iteration of the paper will

[29] Aufhammer et al. (2013) caution the use of air quality data because monitors tend to enter/exit, thereby inducing measurement error in monitoring within geographical locations.

also include health care, as in Hai et al. (2015).

Two complexities arise. First, the CEX delineates between a reference person answering the survey versus a spouse. Especially since the ACS has many variables reported at the household-level, I aggregate to the household-level in the CEX. Implicitly I am assuming that households consume equal amounts of the nondurable goods, but the assumption is reasonable for my purposes since I am not studying inequality within a household. Second, the CEX reports on a quarterly basis (also a monthly, which I do not use), whereas the ACS is annual. To annualize the data, I simply aggregate up to a year-level. Since electricity, fuel, gas, and water tend to be truncated at zero—some households have utilities included in rent—I create an adjusted measure for truncated values using propensity score matching.[30] I also impute some of these expenditures using Census averages for the U.S. and an imputed housing consumption measure (see later).[31] Using electricity as a proxy for the demand for non-durables is better than using food since electricity is more highly correlated with transitory changes in non-durables—for example, a negative shock that leads a family to buy fewer household mass-market goods will also involve less electricity use over the period.

The American Time Use Survey (ATUS) is the benchmark source for time allocation data in the United States. Despite the detailed activity breakdown, they create a consistent measure of time allocation through a weighting scheme allowing for representative aggregation. Using the multi year data from 2003-2012, I trim to consider my 2005-2010 sample period. The advantage of the ATUS over the Census Bureau data is in not only the greater reliability in their reporting of time use data (e.g., since participants must fill out an activity log), but also its degree of detail. Defining leisure is much tougher than it is for non-durables consumption; I follow Aguiar and Hurst (2007), but is robust to Ramey and Francis (2009).[32]

Following most of Blundell et al. (2008) for trimming the CEX data set, I cut the Census and both ATUS/CEX such that similar individuals are captured. After adjusting the final measure of

[30] As long as there is overlap in the distribution of covariates in the variables that are correlated between renters and non-renters, and unobservables are not systematically correlated with renter behavior, the truncated-correct values will provide suitable characterizations of the counterfactual as if the household were paying utilities. Including an indicator variable denoting whether the household has zero electricity expenditures or not fails to address the root of the issue.

[31] For the Census averages, see Figures 2 and 3: http://www.census.gov/housing/census/publications/who-can-afford.pdf.

[32] Philosophically, the distinction between leisure and labor is ambiguous. As Isocrates reminds us: "Spend your leisure time in cultivating an ear attentive to discourse." If leisure activities directly enrich our human capital used to increase our earnings, how is it any different from an independent project that a worker decides to begin in order to advance his firm's profitability?) Ramey and Francis (2009) include: sports, fishing, arts/music, dining at restaurants, talking, sleep, movies, church services, reading, walking, meals, TV, hobbies, recreational trips, exercise, meetings, and gardening. Aguiar and Hurst (2007) (Measure 1) include: socializing, passive and active leisure, volunteering, pet care, gardening; their Measure 2 includes Measure 1, and sleeping, eating, and personal activities (excluding medical care). For controls, I use all the ones from the consumption demand model, as well as the number of weeks worked, disability, and industry dummies at the 2 digit level. To correct for measurement error and endogeneity, I use the same instruments as in the case of non-durables.

non-durables with the consumer price index (CPI), I drop households with zero before tax income, those with missing education or state regional records, those with outlier incomes (a level of income below the amount spent on electricity), as well as limiting the age groups to 20-65 years old. I also include the consumer price index and deflate all nondurable consumption goods and energy prices by the appropriate indices.

4.3. Imputing Nondurables Consumption and Leisure

While the Census data contains rich household-level demographic details and comprehensive coverage, it lacks measures of consumption and leisure. A naive way—that fails to provide sufficient variation to identify the parameters of interest—would involve proxying non-durables consumption with electricity expenditures and leisure with non-work hours (e.g., 5100 minus hours worked in a year).[33] There is a detailed theoretical literature on corrections for measurement error and imputation, but a sparse empirical literature.[34] Using a sieve estimator and auxiliary data from the Consumption Expenditure Survey (CEX) and the American Time Use Survey (ATUS), I impute the distributions of non-durables and leisure. The semiparametric estimator is able to capture the rich nonlinearities of the demand for non-durables and leisure based on data observed in all of the datasets and follows thematically in line with Blundell et al. (2008) who estimate a demand system to impute non-durables in the Panel Study of Income Dynamics using data from the CEX.[35] Aside from some flexible semiparametric restrictions, the only limitation is the requirement that the variables used in the imputation are common across all datasets such that only the to-be-imputed variable is missing in the using (primary) data set.

[33] Just as food expenditures were used to proxy for non-durables in the early literature on estimating the intertemporal elasticity of substitution (Hall and Mishkin, 1982), and later found to be weak proxies because of inelasticity and their relative low covariance with the overall consumption bundle (Attanasio and Weber, 1995), electricity expenditures fall into the same trap. Similarly, taking leisure as the difference of total hours and hours worked would attenuate the variation since many allocations of leisure time (e.g., sleeping) are unresponsive to environmental amenities. An entirely separate strategy would involve creating grids of representative households (e.g., households in a certain income bracket, education level, race, etc) and match among the datasets; this approach is guaranteed to provide strictly worse and more inaccurate results since the household-level behavioral responses to county-level air quality is vital for identification—and preferences for environmental amenities differ immensely across the population.

[34] One vein of the literature (e.g., Robins et al. (1994)) focuses on consistent estimation when some of the variables are missing for a subset of the sample series, but not all; these methods have tended to involve inverse probability weights associated with the missing variables. Another vein of the literature (e.g., Chen et al. (2008)) focuses on consistent estimation when variable(s) might be missing for a large subset of the sample series or all of it; these methods have tended to emphasize auxiliary datasets and semiparametric method of moments and propensity score estimators for out-of-sample estimation. See Chen et al. (2011) for a detailed survey.

[35] Although this was the first method that I attempted, it did not succeed in allowing me to impute consumption and leisure, potentially because there is less time series variation. Campos and Reggio (2014) emphasize that instrumenting does not address the asymptotic bias resulting from the covariance between their control variables and the error, which is non-zero especially if demographic and other household-level variables are reported with error differentially in the two datasets. For example, Gibson (2002) finds that household size is correlated with measurement error since a single respondent asked to remember expenditures for an entire household is likely to make more mistakes the greater the number of people he must remember for.

The theory of sieves is is relatively easy to implement and have desirable properties when the approximating functions are unknown (Chen, 2007). Its accuracy depends on the extent to which the target (to-be-imputed) variable is a smooth function of its explanatory variables. The sieve estimator achieves asymptotic consistency by allowing these explanatory variables to enter nonlinearly and with many higher order terms. An alternative approach would be to specify a structural model; see the footnote for details.[36] However, because it relies on the assumption that the model is properly specified, mis-specification in the imputation can give rise to unknown forms of non-classical measurement error in the actual estimation of air quality elasticites. Robustness checks indicate that my sieve estimator yields the most accurate estimates.

Denote $Y_{1i} = (Y_{1i}1, ..., Y_{1iM})' \in \mathcal{R}^M$ as the target M variables (e.g., non-durables and leisure), Y_{2i} as the proxy for Y_{1im}, $X_{1i}^{'} \in \mathcal{X}$ as the "equivalence scale" common across households, with $\dim(X_1) \geq 1$, $\dim(X_2) \geq 1$, and $Z_i = (Y_{1i}^{'}, Y_{2i}, X_i)'$. Denote $\alpha = (\theta, h_1, ..., h_M)$ as all the unknown parameters of interest and $\mathcal{A} \equiv \Theta \times \mathcal{H}_1 \times \cdots \times \mathcal{H}_M$ as the parameter space where $\theta \equiv (\theta_1^{'}, \theta_{2,1}^{'}, ..., \theta_{2,M}^{'})'$ denotes the evector of finite dimensional parameters given by $\theta \in \Theta$, a compact subset of $\mathcal{R}^{d_\theta}$ with $d_\theta \equiv (1 + M) \dim(X_1)$. The terms $h_m \in \mathcal{H}_m$ will denote the unknown demand curves associated with subsets of the target good m, $m = 1, ..., M$, where $\mathcal{H}_m$ is a subset of a space of functions that are square integrable against the probability measure of Y_{2i}. Letting $\rho \equiv (\rho_1, ..., \rho_M)' \in \mathcal{H}^M$ represent

$$\rho_m(Z_i, \alpha) \equiv Y_{1im} - h_m(Y_{2i} - \chi(X_{1i}^{'}, \theta_1)) - X_{1i}^{'}\theta_{2,m} \tag{8}$$

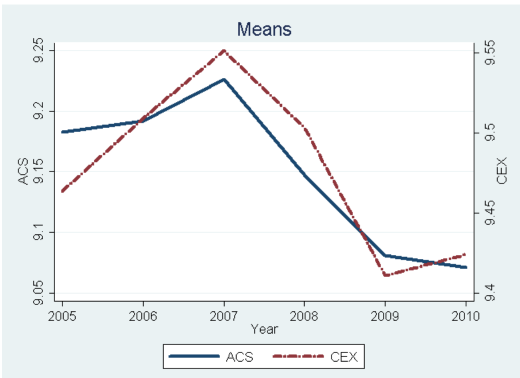
for a specified functional form $\chi(\cdot)$, then individual i facing the same prices for goods $m = 1, ..., M$ will have a demand curve that satisfies $E\left[\rho(Z_i, \alpha_o)|X_i\right] = 0$ where $\alpha_o \equiv (\theta_o, h_{o1}, ..., h_{oM}) \in \mathcal{A}$ is the true (unknown) parameter. The objective is to estimate θ_o and the demand functions h_{om} in order to recover parameters that fully characterize the mapping between the target and input variables of interest. Under very general regularity conditions, Blundell et al. (2007) show that these demand restrictions will be satisfied. Following Ai and Chen (2003) who establish a framework for estimating moment conditions of unknown functional forms, consider approximating functions $h_m \in \mathcal{H}_m$ by $h_{m,n} \in \mathcal{H}_{m,n}$ for $m = 1, ..., M$ where $\mathcal{H}_{m,n}$ is a sieve space for $\mathcal{H}_m$ (e.g., Fourier series, splines, and so on) so that $\mathcal{H}_{m,n}$ becomes dense in $\mathcal{H}_m$ as $n \to \infty$. Arbitrarily fixing a value of $\alpha = (\theta, h_{1,n}, ..., h_{M,n})$ in the sieve parameter space, then the population conditional moment function, characterized by $g(x, \alpha) \equiv (g_1(x, \alpha), ..., g_M(x, \alpha))'$, can be estimated using sieve

[36] For example, let $u(c, l) = [ac^\alpha + (1 - a)l^\alpha]^{1/\alpha}$ denote a constant elasticity of substitution function between consumption and labor. If households maximize utility subject to a budget constraint with no savings (just equal to the wage times labor services), then the intratemporal Euler implies that $(1 - a)l^{\alpha-1}/(aw) = c^{\alpha-1}$. Taking the log of both sides yields $\log c = \log l - (\alpha - 1)^{-1} \log w + \epsilon$, where $\epsilon = (1 - a)/a$. Running some version of least squares with instruments could yield an unbiased measure of $\hat{\alpha}$. Using this estimate, and letting a be the corresponding value share taken from the data, then consumption could be imputed as $\tilde{c} = \left[(1 - a)l^{\hat{\alpha}-1}/(aw)\right]^{1/(\hat{\alpha}-1)}$.

generalized least squares

$$\min_{\alpha \in \mathcal{A}_n} \frac{1}{n} \sum_{i=1}^{N} \rho(Z_i, \alpha) \left[\hat{W}(X_i) \right]^{-1} \rho(Z_i, \alpha)$$

where $\hat{W}$ is a consistent estimator of a positive definite weighting matrix.[37],[38] Using flexible semiparametric functions (e.g., splines and polynomials) over electricity and hours worked for the cases of non-durables and leisure, respectively, I estimate Equation 8 to obtain relationships on demand.[39] While I consider a variety of specifications (see the Appendix), the annual means/variances are plotted below in Figures 4 and 5 for imputed consumption and leisure.



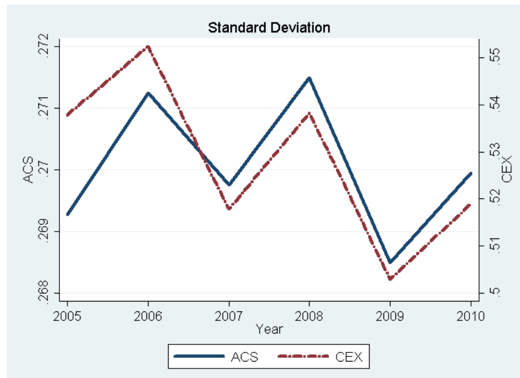


Figure 4: Nondurables Mean and Variance, Actual (CEX) and Imputed (ACS)

Notes.–Source: CEX and ACS. These plot the mean and variance of consumption from the CEX with the mean and variance of the ACS-imputed consumption measure. Using the definition of non-durables from Attanasio and Weber (1995), non-durables is the sum of food (home and away), alcoholic beverages, tobacco, services (e.g., repairs), heating, gasoline, transportation, electricity, water, fuel, personal care, clothing, footware, and rents.

Figures 4 and 5 both characterize the mean log non-durables and leisure in the actual (CEX and ATUS) datasets and the imputed values (in the CEX and ATUS) datasets. To the extent that there is a heavy overlap in the distribution of covariates (see the Appendix for tables), then the coefficients used to impute non-durables and leisure in the CEX and ATUS datasets will be externally valid for application in the Census. For brevity, all my robustness checks are relegated to the Appendix with the summary that the imputed values match the key features of the actual distributions, as defined by the CEX and ATUS. Differences in levels are controlled in my second stage regressions using fixed effects on year. Importantly, remember that the objective here is prediction, not causality, so endogeneity is not germane.[40]

[37]Blundell et al. (2007) provides an excellent methodology for implementing this procedure when the target variables are endogenous and an instrument is needed for Y_2; here, it is taken as exogenous.

[38]Hellerstein and Imbens (1999) show that orthogonality conditions can be created by using moments from auxillary data (e.g., Census and ATUS).

[39]Splines provide a better fit of the data within local regions of the space of points and are well behaved on the endpoints, whereas polynomials can experience unexpected fluctuations ("Runge's phenomenon").

[40]Nonetheless, I also run regressions instrumenting for electricity, water, and gas consumption using interactions between hourly wages (by year, cohort, and years of schooling brackets) and number of children, year, and schooling fixed effects. My results suggest that the IV regressions lead to a much worse fit generally.

Figure 5: Leisure Mean and Variance, Actual (ATUS) and Imputed (ACS)
Notes.–Source: CEX and ACS. These plot the mean and variance of leisure from the ATUS with the mean and variance of the ACS-imputed leisure measure. Using the definition of leisure from Aguiar and Hurst (2007), their "Measure 1" definition includes: socializing, passive and active leisure, volunteering, pet care, gardening.

Even though the aggregate annual measures of consumption and leisure are very close in levels and trends, the individual-specific measure might suffer from (classical) measurement error. Even though my benchmark second-stage regressions are linear in parameters, and include their own instruments that are uncorrelated with measurement error, I experimented with a variety of instrumental variables strategies for this first-stage of imputation. For consumption, I interacted maximum annual temperature and precipitation with log hourly wages and year, schooling, and number of children fixed effects. The identifying assumption is that higher wages only increase non-durables consumption through electricity expenditures on days that are relatively hotter or more humid. Since weather fluctuations are relatively random after controlling for time invariant differences year-to-year, the variation is orthogonal to unobserved shocks to non-durables. For leisure, I interact log commuting time to work with fixed effects on year, number of children, and years of schooling. The identifying assumption is that longer drives to work affect leisure only through its effects on nonwork time within year, family size, and educational bracket. Interestingly, while both these instruments seem to be sufficiently strong, they do not contain enough exogenous variation to affect the prediction quality of consumption and leisure. Because of their worse prediction power—which is what matters in the first-stage—I defer to more flexible semiparametric estimators with splines.[41] The Appendix documents the comparison among a few models (e.g., IV and non-IV) with additional robustness checks.

[41] Commute time is also reported only about 10% of the time in the Census, so it is not pervasive enough to exploit as an IV for a complete imputation—here, I am just using it as a robustness check.

5. Identification Problems in the Hedonic Method

Hedonic methods are used to infer prices on environmental amenities by exploiting individuals' observed choices. The canonical framework that followed from Rosen (1974) considers an environment where households have different tastes and will choose a location that maximizes their utility based on the amenities of the location, including both public and traded goods. Suppose that housing can be decomposed into its various components, differentiating between air quality and all other taste shifters or housing attributes

$$\log H_{icts} = X_{icts}\beta + \log S_{icts}\zeta + \epsilon_{icts} \tag{9}$$

where there exists endogeneity arising from locational sorting based on air quality and omitted variables that covary with air quality. While the standard approach to the omitted variables problem exploit temporal variation in air quality by first differencing counties over time, the problem is that the first differenced regression is not identified in the original hedonics framework absent restrictive functional form assumptions and the time-constancy of the hedonic price function.[42] The naive application of hedonic regressions has encountered decades of identification problems since, on one hand, quasi-experimental methodologies have argued that these regressions ignore important omitted variables, and, on the other hand, hedonic methodologies have argued that some quasi-experimental regressions are not informative for understanding willingness to pay for amenities in the presence of non-marginal changes or heterogeneous treatment effects.[43]

However, there are three additional concerns raised in this article that are new. Even if the parameters of interest in the standard hedonic model are identified, they are neither necessary nor sufficient for cost-benefit evaluation in public policy because the variation that these standard models are exploiting co-moves with other aggregates of interest that structurally pin down willingness to pay. Determining the willingness to pay—the parameter of frequent application in cost-benefit analysis—requires knowledge of the income elasticities of demand for all other rationed goods.[44] Before introducing my identification strategy, I will make three arguments that pertain

[42]Both Ekeland et al. (2004) and Heckman et al. (2010) clarify that this is not a failure of the original setup by Rosen (1974), but rather a convenience that has taken root in much of the literature. Kuminoff and Pope (2014) shows that the first differenced parameter is not necessarily able to identify the capitalization effects of the public good typically used for welfare analysis.

[43]Because many quasi-experimental studies exploit a treatment that primarily varies over time, taking the first differenced housing values removes crucial cross-sectional information that is vital for identifying the full demand curve. These papers implicitly assume that the supply curve is flat even though shocks to the household's budget set may shift the entire distribution of the price function. Bockstael and McConnell (2007) remark that "overcoming or avoiding this type of omitted variable bias is much more difficult... virtually all applied papers ignore this source of bias, and little is known about whether the bias thus generated is of a significant magnitude" (p. 177).

[44]Flores and Carson (1997) proved that there is no a-priori reason to suspect that the income elasticity of demand and income elasticity of willingness to pay are equal; the former might be positive, whereas the latter might be sub-

to the sources of upwards bias in traditional approaches. First, unobserved shocks to housing values are correlated with changes in air quality and other environmental amenities; these shocks are inherently local and affect the county-level demand for goods and services. Second, the variance of these unobserved shocks is time varying and orders of magnitude larger than that in the county-level data. Third, using property values as the dependent variable, rather than the relative price of housing services, creates a correlation between the aggregate expenditures on a resource (e.g., total labor income) and the right hand side variables that are dominated potentially at a unit-level (e.g., leisure). While the purpose of this section is to show that the canonical hedonic model faces major identification problems, a few of these motivating results rest upon an instrumental variables strategy explained in detail in Section 6 exploiting quasinatural variation in climate; for now, take the identifying assumptions as given.

5.1. Unobserved Shocks

Unobserved shocks to housing values, such as local labor market and transitory income shocks, affect housing values and are a source of omitted variables bias, endogeneity, and shifters to the hedonic price function. To explore this possibility, estimate Equation 9 and a supplement to it containing non-durables consumption and leisure. The results highlight the contrasting elasticity estimates of housing to air quality when consumption and leisure are omitted versus when they are included. While they are nearly identical with the naive OLS (columns 1 and 4), and when fixed effects are omitted, the instrumental variables estimates differ by a factor of two. Specifically, a 1% rise in air quality is associated with a 4% rise in housing values when consumption and leisure are omitted, versus 6-7% when they are included. Omitting the two induces a positive correlation between the error and air quality since both are luxury goods and positively correlated with property values; thus, the coefficient on air quality suffers from downwards bias. To the extent that time varying unobservable shocks are ignored, which are negatively correlated with consumption (Mian et al., 2013; Mian and Sufi, 2011) and positively affect leisure (Mian and Sufi, 2014, 2012), then the coefficient on air quality suffers form upwards bias.

5.2. Time Varying Standard Deviations, Counties versus Households

The variance of unobserved shocks to households might vary over time and differ from the variance of unobserved shocks at the county-level. To the extent that this variation is driven by reallocation, or effectively any mechanism correlated with air quality, then the elasticity between air quality and

stantially less than unity (and even negative). Moreover, Flores and Graves (2008) showed that failure to endogenize the labor-leisure choice can lead to an overvaluation of public good amenities when considering an expansion.

Table 2: Hedonic Home Regressions and Consumption/Leisure

	Without OLS	FE	IV-FE	With OLS	FE	IV-FE1	IV-FE2
Main effect:							
Ln(Air Quality)	2.59***	.95***	3.95***	2.72***	.91***	6.19***	7.28***
	.39	.25	1.04	.39	.25	.11	.09
Log Nondurables				.34***	.31***	1.43***	1.29***
				.03	.02	.04	.05
Log Leisure				.04***	.02	1.66***	1.62***
				.01	.01	.06	.08
State FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	No	Yes	Yes	No	Yes	No	Yes
Industry FE	No	Yes	Yes	No	Yes	No	Yes
County FE	No	Yes	No	No	Yes	No	No
Observations	423508	333200	203107	423507	333199	257278	203106
Adjusted R^2	.480	.505	.449	.483	.507	.367	.375

Notes.–Sources: Census, EPA, NOAA. Columns 1-3 estimate Equation 9. Column 1 runs a naive OLS; column 2 includes fixed effects on county, year, and industry; column 3 implements column 2 with two staged least squares and fixed effects on year-by-state and industry. Columns 4-6 implement these same regressions, but now including log non-durables and leisure, together with their respective instruments from Section 5 (wind speeds for air quality, log consumption of electricity interacted with fixed effects on the age of the house, and temperature and its interactions with fixed effects on age for leisure). Specifically, column 4 runs a naive OLS. Column 5 adds fixed effects for year, county, and industry. Column 6 instruments each of the endogenous regressors. Column 7 adds year-by-state and industry fixed effects.

housing prices is biased. Indeed, the data suggests that there are dramatic differences between the standard deviation of household and county -level consumption, leisure, and property values, illustrated below in Figure 6.

The contrasting variances of housing values and earnings at the household and county levels should be viewed analogously to the themes initially articulated by Davis and Haltiwanger (1992). Whereas there is an elasticity of nearly unity between consumption at the household and county -level, and as large as 5.5 for leisure, the elasticity between property values at the household and county -level is both negative and lower in absolute value (-.5). The ratios of consumption and leisure at the county to household-level have a .15 and -.97 correlation with air quality. Only using county-level variation can lead to attenuation at best and bias at worst because of the time varying correlation between the ratio of county-household -level standard deviations to air quality.[45] Given that these unobserved shocks to property values (e.g., consumption and leisure) are large, negatively correlated with property values, and negatively correlated with air quality, then research strategies exploiting only county-level variation incur upwards bias when there is a meaningful difference between the correlation of the sandard deviation of air quality and the difference between the standard deviation of consumption/leisure at the household/county levels.[46]

[45] Auffhammer et al. (2009) study the effects of the CAAA on PM10 across counties and make a related argument: very little effect is observed if the level of aggregation is at a county, rather than monitor-level. Similarly, the theme here is that there is insufficient variation to identify the parameters of interest at purely the county-level—group and individual-level effects are inherently bundled together and dull the signal-to-noise ratio.

[46] Why are these ratios time varying? There are two possibilities. Either the price of unmeasured attributes is changing differentially at an individual versus aggregate level or the size of the shocks to observed covariates is

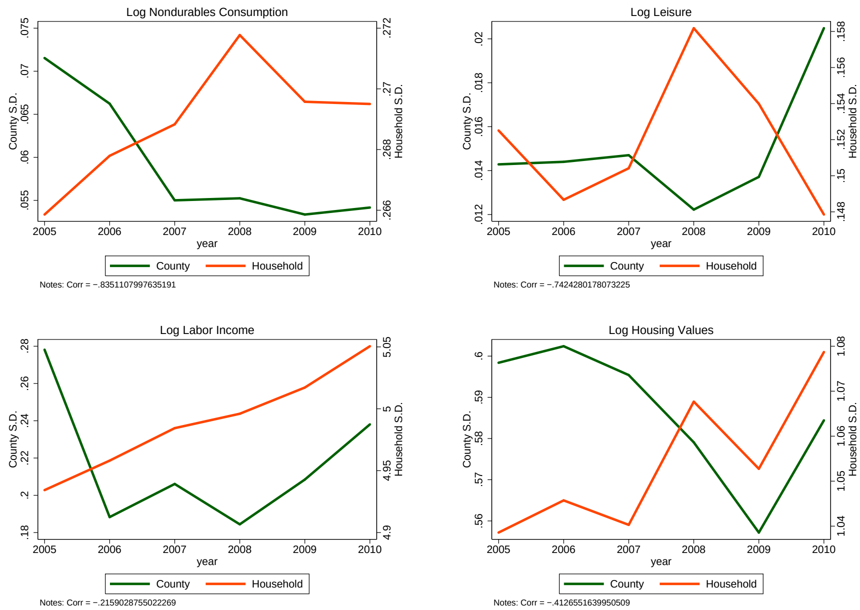


Figure 6: Standard Deviations of Consumption, Leisure, Labor Income, and Property Values, County & Household -Level

Notes.–Source: Census. The plots consist of the standard deviation of log consumption (non-durables expenditures), leisure (hours), labor income (earnings), and housing (property values) at the county and household level. The standard deviations are defined at a county for the base: first, average across all households within a year-county and, second, take the standard deviation within the county.

5.3. Prices and Quantities

Traditional hedonic regressions use property values, rather than the unit price of housing services, in order to estimate willingness to pay. Although the coefficient estimate on air quality may represent some capitalization of an amenity into property values, it does *not* have the interpretation of a marginal willingness to pay for air quality because of unobserved shocks to the gradient of the

changing (Lemieux, 2006). While it is beyond the scope of my analysis to answer the question definitively, consider an exercise similar to Juhn et al. (1993) beginning with the decomposition of the residual variance of property values from Equation 9 into a term consisting of unobserved heterogeneity (u) and measurement error (v); only the former is taken to be correlated with air quality. Letting p denote the price of unobserved heterogeneity, then

$$Var(\epsilon_t) = p_t^2 Var(u_t) + Var(v_t)$$

where ϵ is the residual variance obtained from a regression of property values and earnings on a set of covariates. While the interpretation of p is typically that it measures the price of unobserved tastes, it is indistinguishable from merely a larger variance in shocks over time. The ratio of the mean residual on earnings and housing values aggregated at the county to household level is decreasing for housing and increasing with earnings, suggesting that local labor demand shocks are the dominating channel through which the variance of shocks is heterogeneously transmitting to counties versus households. Recalling the fact that the elasticity of air quality with respect to wages is increasing (decreasing) in the elasticity of consumption (leisure) and with respect to the price ratio of housing to labor is increasing (decreasing) in the elasticity of leisure (housing), I can directly sign the potential for bias that may emerge under canonical hedonic regressions. In the data, the correlations between unobserved earnings variation ratio and air quality and both the consumption and leisure ratios are .12, .65, and -.1, respectively. Even though bias in the coefficient on leisure is going to be smaller, the correlation of the error with consumption and air quality is positve and stronger, inducing upward bias in the elasticity of air quality.

hedonic price function with respect to private goods and services. Derived from a utility-maximizing model containing housing services, Equation 5 shows that the relevant variation must exist over the price of housing services—not the aggregate housing expenditure (e.g., value). To put this in perspective, consider a variant of Equation 4 of the form

$$\log w(1-L) = (\psi - 1)\log L + (1-\phi)\log C + (\phi - \psi)\log S + \rho(x) + \epsilon \tag{10}$$

where $1 - L$ denotes the share of hours allocated to labor services. Aside from an obvious endogeneity problem, this introduces an extra premium to the dependent variable that is negatively correlated with right hand side variables: higher hours worked implies fewer hours in leisure. Similarly, when consumption and leisure are included as right hand side variables, using housing values as the dependent variable induces a positive correlation between housing services and both consumption and leisure. Since housing values are higher in areas with better air quality, there is upwards bias in the coefficient on air quality. To test the severity of this hypothesis, consider additional unrestricted regressions of Equations 4 and 10.

Table 3: Hedonic Wage Regressions with Wage Rates and Labor Income

	Lab. Inc.			Wage			
	OLS	FE	IV-FE	OLS	FE	IV-FE1	IV-FE2
Main effect:							
Log Air Quality	1.88***	-.55***	1.55**	.08***	.04	-.69***	.40*
	.04	.19	.64	.02	.05	.09	.22
Log Nondurables	.20***	.15***	2.04***	.28***	.22***	-.48***	-.87***
	.01	.01	.12	.01	.01	.03	.04
Log Leisure	-25.67***	-19.94***	-19.20***	-2.10***	-1.88***	-6.83***	-6.89***
	.03	.04	.17	.01	.01	.05	.07
State FE	No	No	Yes	Yes	Yes	Yes	Yes
Year FE	No	Yes	Yes	No	Yes	Yes	Yes
Industry FE	No	Yes	Yes	No	Yes	No	Yes
County FE	No	Yes	No	No	Yes	No	Yes
Observations	1403691	1119817	572424	955730	955730	569917	569917
Adjusted R^2	.636	.435	.439	.241	.322	.	.

Notes.–Sources: Census, EPA, NOAA. Columns 1-3 estimate Equation 10, which consists of total labor income as the dependent variable, and columns 4-7 estimate Equation 4, which consists of the hourly wage rate. Column 1 runs a naive OLS; column 2 includes fixed effects on county, year, and industry; column 3 implements column 2 with two staged least squares with year-by-state and industry fixed effects. Columns 4-7 implement these same regressions, but now using the hourly wage rate. Specfically, column 4 runs a naive OLS. Column 5 adds year, county, and industry fixed effects. Column 6 adds year fixed effects together with instrumenting each of the endogenous regressors. Column 7 adds year, industry, and county fixed effects.

The results highlight the importance of disentangling between labor income and the hourly wage. In the naive OLS, using labor income implies a coefficient that is an entire unit higher—many orders of magnitude larger than when the hourly wage is the dependent variable. Introducing fixed effects reverses the point estimate, making it negative. To the extent that air quality is a luxury good and positive attribute, wages should captitalize these amenities, meaning that the negative coefficient is unreasonable. After instrumenting and applying fixed effects over the regression with labor earnings, the result switches back in proximity with the naive OLS. In contrast, under the benchmark specification in column 7, the point estimate is positive, lower than the naive OLS when the dependent variable is earnings, and higher than the naive OLS when the dependent variable is the wage. Interestingly, the coefficients on leisure are implausibly high when the wage is not used as the dependent variable: a 1% rise in leisure is associated with a 20% decline in the hourly wage. These endogenous feedback mechanisms are driven by the fact that unobserved shocks to labor income are larger than those shocks that affect labor supply. Rich fixed effects specifications are important, but no amount of controls will remove the bias if the regression is mis-specified.

5.4. Discussion

These identification problems induce upwards bias in willingness to pay each for their own separate reasons. First, consider mechanisms (1) and (2). Unobserved shocks to wages (e.g., productivity shocks) are positively correlated with wages (Moretti, 2010), but negatively correlated with leisure (higher returns to work) and positively correlated with consumption (higher income). My structural model shows that the causal effect of air quality on wages—a proxy for willingness to pay since it is a more general form of a hedonic wage regression—is decreasing in ψ and increasing in ϕ. Therefore, downwards bias in ψ and upwards bias in ϕ lead to a reinforced upwards bias in the causal effect of air quality on wages. Ontop of all of this, the variance of these shocks differs by orders of magnitude at the household versus county levels. Second, consider mechanism (3). Since the utility of housing or labor services depend on the utility of consumption or leisure in some nonseparable fashion—for example, the marginal utility of housing services depends on amount of leisure time spent in the house—then unobserved shocks to household-level housing values (e.g., local labor demand shocks, denoted u) will drive up leisure ($Corr(u, H) > 0$) and air quality ($Corr(u, S) > 0$), thereby creating upwards bias with air quality. By including the aggregate housing services, rather than just the price of housing services, as the dependent variable, these regressions will simultaneously induce upwards bias since the correlation between aggregate housing services and consumption & leisure is positive and air quality is positive, respectively. While one solution to the latter identification problem involves imposing assumptions about the nature of expectations over unobservables in the hedonic price function (Bajari et al., 2012), these assumptions might be restrictive in periods characterized by volatiltiy and high distortions in the housing and/or labor markets.[47]

[47] There are also other concerns, for example, measurement error in self-reported property values (Pope et al., 2014).

6. Using Intratemporal Variation to Estimate the Demand for Environmental Quality

6.1. Identification

A. Sources of Bias: Unfortunately, a naive application of OLS on Equation 6 will lead to biased estimates for four reasons. The first two are standard empirical challenges and can be resolved easily; the latter three require novel approaches. First, the constant terms—bundled in T_C and T_L—will induce a covariance between the error and (C, L, S) because of time trends. Second, because there is a large mass of individuals who work zero hours within any given year—due to, for example, voluntary or involuntary unemployment—my estimate of $\hat{\psi}$ will be downwards biased.[48]

Third, the right hand side variables of interest household-level consumption-leisure and county-level air quality are endogenous objects that vary with transitory and permanent income shocks. While longitudinal data would enable me to incorporate individual fixed effects and assume that these differences are time invariant, I can only measure shocks at the county-level. Unobserved heterogeneity in workers' productivities can create a downward bias since more productive workers will receive systematically different wage offers, which will also be correlated with leisure and risk preferences (Epple, 1987). Similarly, since people sort into residential communities and local labor markets based on their preferences for public goods (Rhode and Strumpf, 2003), including air quality (Banzhaf and Walsh, 2008), naive applications of least squares will attribute variation in wages to preferences for air quality when it is really driven by unobserved heterogeneity in tastes. Specifically, upward bias emerges because hotter climates tend to be negatively correlated with air quality, and wealthier people tend to favor temperate areas (Albouy et al., 2013).

Fourth, because of non-random sorting across households, the naive least squares estimators bundle combined individual and group effects (Bayer and Ross, 2006).[49] In other words, using individual-level data to estimate Equation 4 bundles the effects of non-random sorting across groups

[48]Consider the regression $w = \gamma l + \sigma s + \epsilon$, where s is an unobserved shock (e.g., non-participation in the labor force). Whether $\hat{\sigma} > 0$ or $\hat{\sigma} < 0$ is unknown ex ante; however, $\hat{\gamma} < 0$ is known: higher leisure reduces wages. Consider the following cases. Case 1: if $\hat{\sigma} < 0$ and $Corr(l, s) < 0$, then $\hat{\gamma} < 0$ implies that $\hat{\gamma}$ will be less negative than the truth. Case 2: if $\hat{\sigma} > 0$ and $Corr(l, s) < 0$, then $\hat{\gamma}$ will be more negative. Case 3: if $\hat{\sigma} < 0$ and $Corr(l, s) > 0$, then $\hat{\gamma}$ will be more negative. Case 4: if $\hat{\sigma} > 0$ and $Corr(l, s) > 0$, then $\hat{\gamma}$ will be less negative. To the extent that censored wages at zero due to non-participation are present in the data set, the regression will overestimate the degree of complementarity for consumption and leisure.

[49]For example, since hours are more volatile than wages over the business cycle (Heathcote et al., 2010a) because of sticky bargaining arrangements (Hall and Milgrom, 2008), controlling for unobserved heterogeneity in match quality is essential. In the absence of such controls, unobserved shocks to hourly wages will load onto the coefficient on leisure and cause it to be downwards biased: job-specific match quality is positively correlated with wages and negatively correlated with leisure since more productive workers and/or matches will face higher returns and rewards to working. Similarly, since both non-durables and housing consumption patterns are relatively clustered based on geographical location (Handbury and Weinstein, 2014), controlling for unobserved heterogeneity in group consumption behavior is equally as important.

and individual behaviors.

$$\log w_{it} = (\psi - 1)\log L_{it} + (1-\phi)\log C_{it} + (\phi - \psi)\log S_{jt} + \beta_1 X_{it} + \beta_2 G_{jt} + \eta_j + \varepsilon_i + \epsilon_{it}$$

where X contains all individual-level preference shifters, G contains all group-level preference shifters, ε_i is the individual-specific idiosyncratic error, and η_j is the j-group idiosyncratic error. Non-random sorting due to unobserved tastes implies that

$$Cov\left[(\beta_1 X_{it} + \varepsilon_{it})(\beta_2 G_{jt} + \eta_j)|X_{it}, \varepsilon_{it}\right] = \mathbb{E}\left[(\beta_1 X_{it}')\eta_j|G_{jt}, \varepsilon_{it}\right] > 0$$

The equation means that there is a positive correlation between unobserved group effects and observed individual-level covariates; the correlation is positive under the assumption that unobserved locational quality enters utility positively and satisfies the single crossing property. To deal with this problem of omitted (local) variables, I apply the insight from Epple and Platt (1998) and Epple and Sieg (1999)—who show that sorting based on a common measure of location quality and its demand implies that each residential location will contain workers within a neighborhood of a specified level of location quality—by using county-level consumption, leisure, and housing values to proxy for unobserved local time varying heterogeneity.

The direction of the bias is very ambiguous ex ante. Endogeneity arising from unobserved shocks to consumption and leisure, like a productivity shock, will tend to cause downwards bias on non-durables and housing, and upwards bias on leisure (since they are associated with cutbacks in disposable income for consumption, declines in home equity, and less time at work either due to a layoff or decline in marginal product). However, endogeneity arising from unobserved tastes for public goods will tend on the correlation between the unobserved taste and the private goods. If public goods are a substitute for consumption and complement with leisure, the bias goes in the same direction as before. Depending on the true parametric relationships, the naive OLS may recover a coefficient close to the truth by accident—that is, the two competing endogeneity problems could cancel each other.

B. Empirical Strategy: I tackle each of these challenges sequentially. The first concern has a simple solution: adding fixed effects on year purges most of the trend and focuses on within-year changes in consumption, leisure, and air quality; results are robust to linear and quadratic time trends. The second concern also has a canonical solution: estimate an instrumental variables treatment effect model to control for selection.[50],[51]

Third, to deal with unobserved shocks to wages that are correlated with consumption, leisure, and air quality, my empirical strategy exploits detailed fixed effects specifications and instrumental variables. The benchmark specifications in all of the regressions that follow will include household/county controls, together with fixed effects on year by state, county, and industry fixed effects. The elasticity of air quality (with respect to consumption and leisure) will be identified from the county-industry-specific deviations in air quality from county averages after adjusting for common shocks across all counties within a state. I instrument for air quality, consumption, and leisure using a quadratic in wind speeds and direction, an interactions between electricity consumption and fixed effects on the age of the homeowner's house, and a quadratic in mean and median maximum temperatures interacted with industry fixed effects, respectively, for each of the endogenous variables. The quadratic terms for the climate variables capture potential nonlinearities (Schlenker and Walker, 2012; Hanna and Oliva, 2015). Since air quality is correlated with temperature and varies over time, county fixed effects alone cannot removes these seasonal shocks (Schlenker and Walker, 2012).[52] Figure 7 characterizes the relevance of these variables by averaging the instrument across counties within year and plotting air quality above and below the median value.

Counties with faster wind speeds have higher air quality; counties with higher temperature have higher leisure; counties with higher electricity expenditures and newer houses have higher non-durables expenditures. The intuition behind the wind speeds instrument is that faster winds blow dirty air out of counties; the intuition behind the temperature instruments is that the cost of labor supply is higher in hotter temperatures (Zivin and Neidell, 2014). To make the latter concrete, suppose that effective leisure is given by $g(\tilde{L}_t, T_t) = \tilde{L}_t T_t^{-\eta}$ where $\partial T/\partial \tilde{L} < 0$ is consistent with the scientific literature on the relationship between fatigue and temperature. Temperature

[50] Cerulli (2014) provides a convenient package that implements probit-2SLS, allowing me to estimate the exact selection process and use it in a second stage to correct for the probability of observing a truncated value. The identifying assumption is that my instrument, years of schooling, is orthogonal to all unobserved shocks affecting selection into non-participation status, after conditioning on observables and fixed effects on state, year, and industry. Controls include: the county-level geographic measure of ruggedness, county-level weather and business patterns, number of children, disability status, age, male, number of bedrooms, age of the house, whether the house has a detached building, vacancy, housing tenure, race dummies, population.

[51] While an alternative strategy is to exclude workers with zero wages/hours, this would underestimate the degree of complementarity since there are two margins through which gains in environmental quality affect private goods and services. In the intensive margin, higher air quality may increase or decrease the marginal utility of consumption or leisure; a reasonable conjecture is that it would increase the marginal utility of leisure, thus its demand. In the extensive margin, higher air quality comes at the cost of greater regulations, which lower wages due to capital-labor and energy complementarity (Hassler et al., 2012), and thus greater leisure; this margin is likely to overwhelm the first, and is ignored by focusing only on workers with positive wages.

[52] Aufhammer et al. (2013) caution the naive application of weather data. To summarize: (1) deviations around the mean temperature might be inaccurate, even if average temperatures are accurately constructed; (2) averaging across nonmissing weather station data induces measurement error; (3) the correlation between weather variables varies across space significantly in sign and magnitude; (3) weather indicators are often spatially correlated because of the extrapolation methods used to create the measures, inducing collinearity.

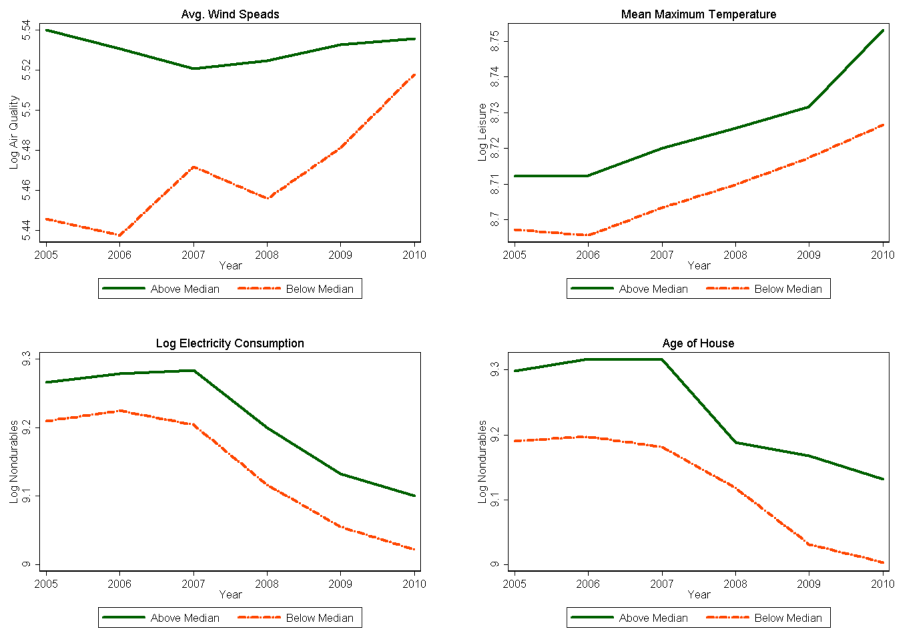


Figure 7: Relevance of Instruments

Notes.–Sources: EPA, NOAA. The figures plot log air quality (by year) by taking the average across counties that are above and below the median of the different instrumental variable groups. The first group is a measure of wind speeds (average within-county); other measures of wind speeds, like the average resultant wind speed (speed and direction) deliver similar relationships. The second group is the mean maximum temperature. The third is the mean log electricity expenditures. The fourth is the mean age of the house.

heterogeneously amplifies the extent of workers' exertion across industries.

Fourth, I follow the control function approach suggested by Bayer and Ross (2006). In the first step, I generate county-level leave-one-out averages of consumption, leisure, and housing to control for unobserved group-level heterogeneity driven by locational sorting, e.g., the unobserved quality of job-matches within a county leading to workers' fluctuations in labor supply. The control function addresses these time varying unobservables since the data is not sufficiently detailed to introduce county-by-year fixed effects.[53] In the second step, I instrument the endogenous consump-

[53] An alternate strategy is to use cohort by years of schooling by industry average hourly wages as a proxy for match quality. This form of aggregation of labor services proxies for common shocks to match quality over the Great Recession since cohort captures labor market experience, years of schooling captures skill content, and industry captures cyclicality and riskiness. However, to the extent that there is selective assortment into industries that are more or less risky based on unobserved tastes, there is an endogeneity problem. Based on an insight from Boualam (2014), I generated a measure of population-wage elasticities by running regressions of the form

$$\log w_{ics} = X_{ics}\beta + \log Pop_{cs}\varepsilon^{w} + \epsilon_{ics}$$

where w denotes the hourly wage, X includes observable controls and public use micro-data area (PUMA) fixed effects, and Pop is the county-level population. The reduced form elasticity, ε^{w}, characterizes the extent to which a rise in population affects the average wage for a given worker. Counties that tend to be more isolated will have a lower (possibly negative) elasticity since geographical remoteness might behave as an impediment to future development. Counties with a high elasticity are more poised for growth and are likely to feature greater labor mobility. Since the generated county-level elasticity is used as an instrument for the quality of matches within cohort, educational attainment, and industry brackets, the exclusion restriction requires that unobserved shocks to wages that are correlated with county-level population elasticities are (a) uncorrelated with cohort, industry, and schooling specific shocks, tion, leisure, and housing individual-level variables using the aforementioned instruments. The IV strategy gives me the separation of group from individual effects for free.

C. Exclusion Restrictions: The exclusion restriction for the air quality instrument requires that unobserved shocks to wind speeds are mean independent of hourly wages—that is, wind speeds affect hourly wages only through the effect on air quality, which induces households to select into one county over another. The exclusion restriction for the consumption instrument requires that unobserved shocks to hourly wages are uncorrelated with electricity consumption within the set of houses built within the same year—that is, electricity consumption within houses of similar arges affects hourly wages only through non-durables consumption. Since these regressions will condition on county-level fixed effects, together with measures of topography, state-by-year, and industry fixed effects, there is no risk of the climate variables—wind speeds and temperature—picking up variation correlated with economic activity through a geography mechanism. Similarly, since previous literature finds that household's demand for electricity is inelastic (Reiss and White, 2005)—often not responding to price signals at all (Shin, 1985; Bushnell and Mansur, 2005)—unobserved shocks to wages are likely to cause households to cut back other forms of consumption, rather than electricity (or, water and gas).

D. Comparison to Other Methods: While the estimated elasticities are entirely novel contributions, another major advantage of my approach is that they can be used either independently or supplementally with other external information (e.g., as sufficient statistics) in order to understand the welfare effects of policy intervention over public goods. In contrast, in the hedonic model, failing to account for adjustment in the hedonic price function biases the coefficients of interest since variation in the relationship between amenities and assets used to infer willingness to pay (e.g., housing markets) is loaded onto air quality (Kuminoff et al., 2010). The identification problem arises from the fact that an exogenous shock to the spatial distribution of a public good, like air quality, changes the gradient of the hedonic price function in order to clear the housing market, introducing a wedge between the average capitalization effect and household's willingness to pay. Rather than conflating willingness to pay for air quality with changes in the shadow price of air quality—represented through the price of consumption or labor—my structural model explicitly controls and instruments for them. Nevertheless, I also implement robustness checks where air quality and wind speeds are interacted with a linear time trend. The hedonic price function is

and (b) uncorrelated with an individual's leisure. While the regression seems to successfully control for unobserved shocks to job-matches, my auxiliary regressions revealed that they also implied $\phi > 1$. Such a value would be unable to generate the observed joint rise in consumption and air quality observed across countries (Makridis, 2014) and seems to be driven by a failure of the exclusion restriction: county-level population-wage elasticities reflect historical patterns of development that are not exogenous to contemporaneous shocks.

linked with the valuation of environmental amenities, requiring some intertemporal link.[54]

E. Interpretation of Treatment Effects: Most quasinatural identification strategies are challenged with the extent to which their estimates can be interpreted as average treatment effects (ATE) over local average treatment effects (LATE). Both Deaton (2009) and Heckman and Urzua (2010) are famous for their critiques of instrumental variables for obtaining measurements of policy relevant ATE parameters.[55] In particular, instrumental variables techniques tend to assume constant treatment effects in that the effect is identical for every observation. In the context of this paper, instruments are used to address the fact that counties with different air qualities are not assigned randomly. To the extent that the instrument introduces exogenous variation into the compliers group—counties (e.g., households residing in them) whose treatment status can be manipulated through the wind speeds instrument—these results may only identify the LATE. However, I provide evidence that these can be interpreted as ATEs for three reasons. First, the distributions of the outcome variables do not differ heavily among compliers, always-takers and never-takers. Second, I implement a reweighting according to Aronow and Carnegie (2013). Third, I explicitly study the presence of heterogeneous treatment effects by income bracket, age, educational attainment, and exposure, and find that they estimates are qualitatively similar, meaning that the average treatment effects are not biased.

6.2. Empirical Estimation

A. Consumption and Leisure Elasticities: To identify the elasticities on consumption and leisure, I exploit variation in hourly wages using Equation 4.[56] Consistency requires that unobserved shocks to hourly wages are mean independent of my instruments. Each endogenous regressor is linked with a separate instrument. For leisure, temperature and its interactions with age captures the way in which temperature fluctuations affect time use (Zivin and Neidell, 2014); for non-durables consumption, interactions between electricity expenditures and the age of the house capture heterogeneity in how households with homes of similar ages use electricity; for air quality, a quadratic in mean and median wind speeds captures the correlation between air diffusion patterns

[54] Including a cubic time trend simply controls for unobserved shocks to hourly earnings—a similar candidate "fix"—it does not control for changes in the slope of the hedonic price function. Interacting air quality and the instruments with year dummies also failed; the estimated coefficients were too imprecise to have any meaning.

[55] Imbens (2010) discusses these concerns through a broader lens of the literature and the constraints inherent in empirical economics—for example, randomization is not always possible. Instrumental variable techniques still have the potential to achieve internal validity, and their external validity will depend on the underlying context.

[56] Although the division bias (Borjas, 1980) is a common problem in labor studies, and is best addressed using another measurement of wages as an instrument, there are no alternative measurements of hourly wages available in the Census. Fortunately, given the sample size, any attenuation resulting from division bias does not have a quantitatively strong effect on my estimates given the sample size and quality of variation. Furthermore, measurement error in the dependent variable (wages are typically an independent variable) will just raise the standard error at worst.

and pollution concentrations. The table below presents the estimated elasticities from Equation 6 under different specifications.

Table 4: Structural Estimates of Consumption/Leisure Elasticities

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ψ	-1.094	-1.177	-3.778	-3.665	-1.857	-2.459	-2.168
se(ψ)	0.033	0.026	0.053	0.054	0.030	0.029	0.028
ϕ	0.756	0.874	0.619	-0.169	-1.230	0.465	0.351
se(ϕ)	0.030	0.011	0.020	0.034	0.045	0.025	0.018
λ	1.850	2.052	4.397	3.495	0.628	2.925	2.519
se(λ)	0.052	0.032	0.061	0.072	0.060	0.042	0.033

Notes.–Sources: Census, CEX, ATUS, EPA, NOAA. These columns present the parameter estimates implied by the structural model from estimating regressions of log hourly wages on the log of leisure, non-durables, air quality, and controls for taste shifters. Column 1 presents a simple OLS. Column 2 adds in county, industry, and year dummies without an instrument. Column 3 instruments only for leisure and consumption with fixed effects on year, industry, and county. Column 4 instruments for each of the three endogenous variables, but only has year, state, and industry fixed effects. Column 5 instruments for air quality, but uses county and year fixed effects. Column 6 instruments for each endogenous variable and contains fixed effects on year, county, and industry. Column 7 uses state by year fixed effects, together with those on industry and county. Leisure is instrumented using cohort by years of schooling interactions; non-durables is instrumented using electricity expenditures; and, air quality is instrumented using wind speeds. Controls on state and county economic conditions include: civilian labor force, population, employment, number of establishments, payroll expenditures; controls on households include: number of children, disability status, age, gender, number of bedrooms, housing tenure, detached family house status, housing tenure, race dummies, and population. Standard errors are heteroskedastic robust and bootstrapped with 50 replications. Because of the length associated with these fixed effects regressions, longer replications have been tested, but are omitted here.

The elasticities of substitution are identified from within county-industry deviations in consumption, leisure, and air quality after adjusting for all the state-year specific shocks that are common to counties within the same state. Under the preferred specifications (columns 5-7), the results suggest that there is complementarity for leisure and air quality, but not between non-durables and air quality. A 1% rise in leisure is associated with a 2.5-3.5% decline in the hourly wage; a 1% rise in non-durables is associated with a .15-.3% decline in the hourly wage; and, a 1% rise in air quality is associated with a 3% rise in the hourly wage. The implied elasticity on labor supply is above the traditional micro-elasticity (around .6 for full time males; Keane (2011)), but is explained by the convexification of labor supply generated through the inclusion of air quality.[57] The fact that air quality is positively associated with the hourly wage reflects the fact that environmental and air quality regulations tend to lead to large reductions in employment (Greenstone, 2002; Walker, 2013)—a result that is consistent with the 2005-2010 period and discussed in a later subsection. Compared to the naive OLS results (columns 1 and 2), the IV results highlight that unobserved shocks to hourly wages are inducing significant upwards bias on the consumption coefficient and

[57] Just as in Imai and Keane (2004) who find an elasticity near four, since local labor markets also include amenities, then the selection of a location to work is tantamount to a choice on public goods. Since air quality is negatively correlated with hours worked, but positively correlated with the hourly wage, then the traditional elasticity is downwards biased.

downwards bias on the leisure coefficient: productivity shocks are positively (negatively) correlated with consumption (leisure) and positively correlated with the hourly wage. Comparing the OLS and FE results (columns 1 and 2), fixed effects do next to nothing to obviate the endogeneity problems. However, the IV results without fixed effects (column 3) has coefficients on leisure and consumption that are twice has high and low, respectively, as the benchmark (column 7), reflecting bias induced by preferences for locational sorting: unobserved tastes are positively (negatively) correlated with consumption (leisure), and negatively correlated with the hourly wage.

Comparing column 3 with columns 6 and 7—where the former instruments for only consumption and leisure, and the latter instruments for all three—the coefficients are not substantially different, suggesting that locational sorting does not matter as much as the unobserved transitory and permanent income shocks. Unobserved shocks to earnings—for example, through higher labor market risk during the Great Recession—are negatively correlated with earnings and consumption (lower disposable income), but positively correlated with leisure. While this reinforces the upwards bias in consumption, it makes leisure downwards biased. The fact that the coefficient on leisure is lower under the naive OLS regressions, relative to the IV regressions, suggests that locational sorting is also causing upwards bias on ψ. Instrumenting for these sources of endogeneity allows me to identify the entire demand curve through an unobserved shock to the supply of the hedonic function under the assumption of homogeneous preferences.[58] Concerns about omitted variables are mitigated, in contrast to the reduced form estimates by evidence by Chay and Greenstone (2005), because the specifications not only include a significant amount of controls, but also impose structure on the relationships in the data.

B. Housing and Leisure Elasticities: The structural model established in Section 3 provides an additional source of intratemporal variation in Equation 5. While I proxy the price of housing services using a household's selected annual owner costs, I impute housing consumption using the definition in Prescott (1997).[59] To address endogeneity, I use a quadratic in county-year specific

[58] As Heckman et al. (2010) discuss, selection inhibits the interpretation of hedonic estimates as measures of willingness to pay. By embedding an instrumental variables strategy to keep the cross-sectional variation, my structural model is able to leverage variation arising from general equilibrium interactions.

[59]

$$H_{icts} = pr_{icts}\left[r_t^f + \tau_{icts}^p - \tau_t^m(m + \tau_{icts}^p) + \delta - \Delta p_{icts} + r_t^e\right]$$

where pr is the self reported property value, r^f is the risk-free rate, τ^p the imputed property tax rate, τ^m the marginal tax rate, m the mortgage rate (10-year average of 30-year fixed rate mortgage rate), δ is the depreciation rate, Δp is the capital gain (change in property value), and r^e the (equity) risk premium. Since the mortgage rate is set according to a 10-year average, I use the 10-year return on a treasury bill from CRSP to obtain r^f. r^e is simply (.02 for now) the difference between the real return on stocks (equity) and the risk free rate.In particular, I use the S&P 500 index composite (monthly close value) as a measure of the real return to equity averaged out over the year. τ^p is imputed by taking the annual property taxes paid by the household divided by the self reported property value. From Harding et al. (2007), I set $m = .055$.While the ACS provides data on the mortgage payments, the actual rate is much more complicated than taking the ratio between the payment and the house. I use TAXSIM's average marginal tax rates to compute τ^m (Feenberg and Coutts, 1993). I set the depreciation rate according to mean and median snowfall measurements—again, robust to conditioning on the housing supply elasticity—and interactions with the housing supply elasticity discussed earlier to instrument for housing services. Differential changes in snowfall cause households to invest more in their houses because they spend more time inside, especially so in counties that tend to have lower housing elasticities (since land is more scarce).

Table 5: Structural Estimates of Housing/Leisure Elasticities

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ζ	-0.624	-0.621	-0.603	-0.618	-1.448	-0.627	-0.628
se(ζ)	0.008	0.007	0.059	0.057	0.011	0.008	0.008
ψ	-1.330	-1.049	-5.014	-4.706	-1.942	-2.608	-2.287
se(ψ)	0.031	0.027	0.065	0.086	0.063	0.042	0.042
λ	-0.127	-0.359	-0.123	-0.715	1.732	-0.729	-0.385
se(λ)	0.189	0.170	0.086	0.153	0.501	0.409	0.396

Notes.–Sources: Census, CEX, ATUS, EPA, NOAA. These columns present the parameter estimates implied by the structural model from estimating regressions of the log of the price ratio of housing to labor on log housing services, leisure, air quality, and controls for taste shifters. Column 1 presents a simple OLS. Column 2 adds in county, industry, and year dummies without an instrument. Column 3 instruments for housing and leisure (not air quality) with fixed effects on state and year. Column 4 also instruments for air quality and adds industry fixed effects. Column 5 adds county fixed effects (in addition to industry and year fixed effects). Column 6 is the same as 5, but uses state by year fixed effects. Column 7 is the same as 6, but also includes industry fixed effects. Leisure is instrumented using a quadratic in mean/median maximum temperature and interactions with age fixed effects; housing consumption is instrumented using a quadratic in mean/median snowfall; and, air quality is instrumented using a quadratic in wind speeds. Controls on state and county economic conditions include: housing supply elasticity, civilian labor force, population, employment, number of establishments, payroll expenditures; controls on households include: number of children, disability status, age, gender, number of bedrooms, housing tenure, detached family house status, housing tenure, race dummies, and population. See the main text for the imputation of housing services, involving information on property values, taxes, and interest rates. Standard errors are heteroskedastic robust and bootstrapped with 50 replications. Because of the length associated with these fixed effects regressions, longer replications have been tested, but are omitted here.

The results are extremely encouraging because the labor supply elasticity is very close to that obtained in the benchmark regression exploiting variation in hourly wages. Since housing and labor markets are separate enough and subject to different shocks, it is unlikely that the similar coefficient is being driven by unobserved heterogeneity. The results also draw a similar contrast between the naive OLS in columns 1-3 and the instrumental variables regressions in columns 4-7. Under the preferred specification (column 6), a 1% rise in housing services, non-durables, and air quality are associated with approximately a 1.6% decline, 3.3% rise, and .4-.8% decline in the price ratio of housing to labor. Since the price ratio is decreasing in air quality, this suggests that labor markets capitalize environmental amenities marginally more than housing markets. While surprising at first, the intuition is that households are able to substitute more effectively through

best practices for annual data: $\delta = .025$. Finally, Δp is simply $(p_t - p_{t-1})/p_{t-1}$ for every household. Although this procedure might sound excessive, environmental quality is correlated with housing services, so measurement error in the housing services will lead to noisy parameter estimates at best. (See Bieri et al. (2014) for spatial variation in the user cost of housing.) This measure of consumption is especially relevant given that air quality is capitalized into property values (i.e. the hedonic property value literature).

the labor, rather than housing, market, potentially because human capital has higher returns on average than physical capital investments in the housing market.

Turning to the comparisons in estimates, notice the very marginal changes in the coefficient on housing services across each of the specifications, whereas the coefficient on leisure changes dramatically throughout. Because local economic shocks and unobserved tastes for air quality work in opposite directions with respect to the bias, the fact that the OLS, FE, and IV specifications are nearly identical for housing suggests that the two effects are exactly cancelling. In contrast, the coefficient on leisure is upwards biased in columns 1 and 2 because unobserved shocks to local labor markets are negatively correlated with housing values and the price ratio of housing to labor.[60] Once detailed fixed effects are introduced in column 3 to control for these labor market shocks, the coefficient becomes much more negative. However, since unobserved tastes for air quality are positively correlated with leisure and negatively correlated with the price ratio of housing to labor, since homeowners earn a relatively higher share of their net income from labor rather than capital income, the coefficient is downwards biased. Correcting for endogeneity through instrumental variables slowly raises the point estimates, but remains robustly negative. To put this in comparison with the Rosen-Roback framework, the standard approach would be to subtract that coefficient on air quality in the regression of earnings on air quality from the coefficient on air quality in the regression of housing values on air quality. Since air quality is a luxury good, the implied WTP from the Rosen-Roback model is not a structural parameter that can be readily applied in different situations.

C. Discussion: To my knowledge, the only other paper that has estimated a potentially comparable elasticity is West and Williams III (2007) between gasoline and leisure, finding evidence of complementarity between the two, although they do not advocate a particular mechanism behind the results. Since gasoline produces emissions, and thus negatively correlated with air quality, the result is evidence of substitutability between air quality and leisure. There are a variety of reasons that motivate the contrasting result. First, they exploit cross-sectional variation in labor supply and cross-state variation in gasoline prices. Because unobserved shocks to gasoline prices are time varying, and they have heterogeneous effects on workers' transitory incomes and local labor demands, they introduce endogeneity. The negative correlation with both consumption and labor supply/earnings induces upwards bias and makes the elasticity between air quality and leisure appear more substitutable. Second, they instrument for earnings using occupation, state, and gender -specific means, real income using an alternative price index, and gasoline prices using national-average gas prices and gasoline refinery outages. However, the variation in these instruments does

[60] The latter assumes that the price on housing is higher than on labor. Since the origin of the labor market shocks largely is from the subprime mortgage crisis, housing is arguably the dominating and precipitating factor.

not match the variation needed to identify household-level valuations of environmental amenities (or gasoline).[61]

While county-level fixed effects are very important, Fu and Ross (2013) caution against inclusion of too finely gridded fixed effects specifications may reduce bias, but also run the risk of attenuating coefficients related to individual-level productivity. An equally, and even more relevant, cautionary note is with respect to the inclusion of county-specific year trends. Interestingly, my robustness exercises involving interactions of linear time trends for each county reversed the elasticity on consumption—turning it negative such that there is complementarity between consumption and air quality. While the the results remain statistically significant, these county-specific trends appear to be removing all of the identifying variation from the instrument. Given that wind speeds at the county-level do not fluctuate massively between 2005-2010, there is simply not enough variation in them in order to identify a meaningful first-stage effect on air quality or even second-stage effect on hourly wages. For example, regressing air quality on these county-specific year trend fixed effects yields an R^2 of .97, showing that little identifying variation remains.

My conclusion is informed by three robustness checks. First, nonparametrically controlling for county-level time varying economic shocks through the inclusion of different bins indicating the number of establishment closures of a particular size (e.g., 1-10 employees, 10-20, and so on) should—if the county-specific trends regressions are valid—make the coefficient on consumption more negative. However, the coefficient remains nearly the same as it was in the setting just with county-level and state-by-year and industry fixed effects. Second, I test for a pre-trend across counties using a flexible estimator operationalized by Mora and Reggio (Forthcoming).[62] Third, and very interestingly, to the extent that employment at the county-level is a broad indicator of a county's economic trajectory, regressing it on the number of establishments and payroll expenditures yields an R^2 of .9927. Evidently, the R^2 cannot rise much further. Adding in county-specific year trends only raises the R^2 to .9999, meaning that these proxies are successfully absorbing the relevant county-specific time varying economic shocks.

6.3. Validity of the Exclusion Restriction

Many economic forces were present during the Great Recession. While it is highly unlikely that counties with different climate variabilities in wind speeds and temperature were on differential trends, such a threat to identification would induce bias. Tests of overidentifying restrictions

[61] They also use a shorter time series (1996-1998) with much lower sample size and fewer controls.

[62] Mora and Reggio (2012) show that recovering causal effects in difference in difference estimators actually requires more than just a parallel path condition, but also similar time trends. They introduce a flexible form for allowing for various types of trends—not just linear—and show that results in many prior papers are not robust to this threat to identifying assumptions.

provide information under the assumption of a null hypothesis that the model is correctly specified and overidentifying restrictions are valid. Because the test is implemented by only excluding some of the instruments, the remainder are assumed to be exogenous to test the subset of excluded instruments. To assure that my elasticities are not driven by unobserved variation in weather outcomes, I examine the plausibility of the exclusion restriction directly and, second, introduce another set of instruments that generate similar results.

While these are explicitly considered in the Appendix, I summarize the basic checks here. First, I show that proxies for economic shocks (establishment closures, employment, pay, and hours at the county-level) are uncorrelated with each of the instruments. Since the only threatening violation to the identifying assumptions of the model is that time varying economic shocks to economic activity are in some way correlated with time varying unobservables. Each of the four proxies of economic factors are uncorrelated with the instruments. Second, estimating Equation 4 and introducing the instruments as additional controls yields statistically insignificant coefficients associated with each of the instruments. The lack of statistical significane implies that—even with a large sample size—the instruments are unlikely to be correlated with omitted variables since they do not have any meaningful relationship with hourly wages (or the price ratio of housing to labor). Third, the distribution of hourly wages is very similar above and below the median level of the given instrument of interest, meaning that the distribution of unobservables is likely to be uncorrelated with the instruments. Taken even further, conditioning on controls and plotting the distribution of residuals implies a nearly identical overlap. Fourth, there is no correlation between two proxies for locational sorting (housing values and commuting time from home to work) with the instruments. To the extent that time varying unobservables are driven by unobserved tastes or locational sorting, these results imply that there is no correlation after conditioning on fixed effects and observable controls.

6.4. Interpretation of Treatment Effects

Quasinatural identification strategies are ultimately susceptible to concerns articulated by Heckman and Urzua (2010) and Deaton (2009) that the estimated parameters are only local average treatment effects. Importantly, there is also a literature specific to hedonic methods that cautions against the interpretation of results when linear approximations to the hedonic price function are used (Ekeland et al., 2004; Heckman et al., 2010). The standard approach in the literature—which they show has severe drawbacks—involves computing linear approximations to the first order conditions implied by a utility-maximizing consumer subject to a budget constraint that nests the hedonic price function. These approaches are typically justified through the rationale of clean instrumental variable or quasinatural experiment strategies. Kuminoff et al. (2010) implement a series of monte carlo experiments and show that many of the conventional results are not robust to functional form, and tend to induce significant bias when the linear approximation to the hedonic price function cannot capture the movement to a new market equilibrium.[63]

These theoretical insights prompt concern about the interpretation of the results thus far and the extent to which heterogeneous treatment effects may bias towards the interpretation of local average treatment effects (LATE). The purpose of this section is to provide evidence of identification that allow results to have an average treatment effects (ATE) interpretation.[64]

First, as Ekeland et al. (2004) emphasize, multimarket data (e.g., different regional housing markets) is essential. To the extent that my results provide an unbiased estimate of the underlying preference parameters, which are common across agents across markets, my model still allows the distribution of individual heterogeneity to vary across markets since the identifying source of variation arises from cross-market differences in prices and locational choice. Furthermore, when interacting the quadratic in wind speeds with a household-level indicator for whether the house uses electricity (rather than gas) for heating, the estimated parameters are nearly the same. The equivalence between these two cases reflects the fact that there is sufficient cross-sectional variation for identifying an ATE.

Second, the typical assumption required to obtain an ATE with instrumental variables is that treatment effects are constant across subpopulations ("identification at infinity").[65] The condition required for extrapolating from subpopulations to achieve an ATE interpretation is that compliers, never-takers, and always-takers are not found to differ substantially in levels with respect to the outcome variable. In the Appendix, I provide plots of the wage distribution for counties above and below the median value of the corresponding instrument (e.g., wind speeds); as expected, the distributions do not meaningfully differ, suggesting that an ATE interpretation suffices.

Third, Section 7 is devoted to understanding the degree to which subgroups have different preferences over environmental amenities across different brackets of income, age, schooling, and

[63] After a policy intervention, the market will obtain a new equilibrium that shifts the hedonic price function (Heckman and Vytlacil, 2007). Kuminoff and Pope (2014) shows that this feature affects implied capitalization effects in hedonic models.

[64] Bayer et al. (2007) introduce an identification strategy for obtaining causal estimates of local amenities through property values when households are heterogeneous by exploiting discontinuities along the boundaries of geographies. To the extent that housing and geographical attributes are more or less continuous throughout metropolitan areas, variation in consumption and leisure within counties with discontinuously different wages or housing values—based on their distance from a specified county or state boundary—would identify the capitalization of air quality.

[65] For an interpretation of LATE, the already discussed identification arguments are sufficient, namely independence (wind speeds is as good as randomly assigned and do not directly affect hourly wages, the outcome), random assignment, the exclusion restriction (wind speeds are uncorrelated with unobserved shocks to wages), and monotonicity ("no defiers"). The monotonicity condition requires that wind speeds affect counties uniformly, which is achieved by conditioning on topography, other weather related variables, and county fixed effects.

exposure to air quality levels. To understand the extent to which these results are informative about aggregate elasticities of air quality, I test for the extent to which heterogeneous treatment effects change my results using a procedure developed by Xie et al. (2012).[66] While certain groups unsurprisingly have more or less tastes for air quality, the qualitative results remain.

6.5. Robustness without Structural Assumptions

A reasonable concern is that the functional forms in the structural model (e.g., Equation 4) impose too much structure on the data, possibly ignoring unobserved sources of heterogeneity. While my instruments and detailed fixed effects assuage concerns of omitted variables and/or unobserved shocks, the following section takes a more reduced-form approach in the tradition of the quasi-natural experimental literature (Chay and Greenstone, 2003, 2005; Greenstone, 2002; Greenstone et al., 2013).[67] The empirical strategy remains largely the same with the exception that the estimating equation is more flexible.

$$\log S_{cst} = \alpha + \theta f(Z_{cst}) + \beta X_{icst} + \nu_{icst}$$

$$\log y_{icts} = \delta + \varepsilon_X \log S_{cst} + \beta X_{icst} + \epsilon_{icst}, \quad \forall y \in \{C, L, H\} \tag{11}$$

As usual, X denotes a vector of controls and $f(Z)$ denotes a quadratic vector of instruments consisting of measures of wind speeds. The objects of interest are the elasticities: $\omega = 1 - (1/\varepsilon_X)$ where $\omega \in \{\phi, \psi, \zeta\}$. To see that this nests the estimating equations in the structural model, note that, for example, if the dependent variable is consumption, then the error grows to incorporate leisure and hourly wages while the coefficient on air quality is scaled by a constant. As long as unobserved shocks to the outcome variable are uncorrelated with wind speeds (and/or other climate variables used as instruments), the two-stage least squares estimator recovers consistent estimates of the reduced-form elasticities

As in the benchmark, the elasticity of air quality is identified from county-industry-specific deviations in weather from the county averages after adjusting for shocks common to all counties in a state. Introducing county-level fixed effects is important to control for sorting based on weather preferences for living in counties, as well as systematic differences contributing to heterogeneous responses to local labor market shocks over the 2005-2010 period. The advantage of abstracting from the endogeneity of consumption/leisure/housing—since they are now the dependent variables—

[66] Crump et al. (2008) develop a fully nonparametric test for treatment effects of heterogeneity.

[67] The caveat is that these specifications will still impose a log-log relationship on the data, which is less flexible than a linear structure. If the elasticity of hours with respect to air quality is a very nonlinear relationship (conditional on observables), the measurement error induced by the log-log model (by ignoring higher order terms) could be correlated with omitted variables, such as adjustment costs associated with moving along the intensive margin. The results are robust to log-linear and linear functional forms, although the interpretation of coefficients lessens.

comes with the plausible concern that the implied elasticities may not map into the structural model due to the integrability problem. Even though the elasticity estimates are not directly comparable, they are remarkably similar and provide informative robustness about household's valuation of environmental amenities without imposing as much structure. The results for each of the different outcomes are displayed below.

Table 6: Reduced-Form Elasticity Results

	OLS	FE		IV		
	(1)	(2)	(3)	(4)	(5)	(6)
Log Air Quality	-.332***	.126***	.00416	-.860***	1.208***	1.576***
	[.0115]	[.0277]	[.0343]	[.0274]	[.106]	[.0347]
Observations	423508	333200	333200	257279	257279	257279
Adjusted R^2	.610	.642	.643	.604	.616	.623
Log Air Quality	-.00756	-.0978***	-.0197	-.0362*	.215***	.145
	[.00685]	[.0167]	[.0193]	[.0196]	[.0550]	[.106]
Observations	423507	333199	333199	257278	203106	257278
Adjusted R^2	.216	.120	.121	.211	.049	.211
Log Air Quality	4.884***	.956***	.323***	5.328***	.319***	.0955*
	[.0381]	[.106]	[.125]	[.105]	[.0674]	[.0556]
State FE	No	No	Yes	No	No	No
Year FE	No	Yes	Yes	No	Yes	Yes
Industry FE	No	Yes	Yes	No	No	Yes
County FE	No	Yes	Yes	No	Yes	Yes
Observations	423508	333200	333200	257279	257279	203107
Adjusted R^2	.461	.510	.511	.431	.472	.470

Notes.–Sources: Census, EPA, NOAA, CEX. Column 1 runs OLS. Column 2 includes year, industry, and county fixed effects, but does not instrument for air quality. Column 3 includes year by industry and county fixed effecs. Column 4 instruments for air quality using wind speeds and includes fixed effects on year, industry, and state. Column 4 instruments for air quality and uses year by state and industry fixed effects. Column 5 instruments for air quality and using year-to-year deviations from average wind speeds within-county, together with year by state, industry, and county fixed effects. Column 6 instruments for air quality and uses year by industry and state fixed effects. All results contain heteroskedastic robust standard errors to address household's heterogeneous willingness to pay for environmental quality, but the results are also robust to clustering at the county-level. Observations are weighted by the ACS person-level weights, and the puma-to-county allocation factor discussed in the Appendix.

Recalling the fact that $\varepsilon = 1/(1-\omega)$, these reduced form elasticities (ε) are very consistent with the structural elasticities for each outcome variable: consumption is a substitute and both leisure and housing are complements with air quality. First, consider the naive least squares estimators in column 1. The consumption and leisure elasticities have non-sensical interpretations—since $\varepsilon \in (0, \infty)$ and cannot be negative—and the housing elasticity implies that it is a substitute, rather than complement, with air quality, which is inconsistent with a large body of equilibrium sorting literature (Smith and Huang, 1995; Kuminoff et al., 2013). Second, consider the case of the fixed effects estimators in column 2. To the extent that local demand shocks are controlled for via the fixed effects and nonparametric proxy for establishment closures, then column 2 represents the reduced form elasticity net of endogeneity arising from the Great Recession. However, the fixed effects estimator does not address unobserved heterogeneity in tastes for air quality and, therefore, leisure/housing will be downward biased (since they are positively correlated with air quality and negatively correlated with demand shocks) and consumption will be upward biased. Third, consider the combination of both the instruments and fixed effects estimators in columns 5 and 6. They show that these two sources of endogeneity may move in the same direction and that controlling only for one is necessary, but not sufficient, for recovering informative elasticities. Under the benchmark specification (column 6), a 1% rise in air quality is associated with a 1.2-1.5%, .145-.215%, and .09-.3% rise in non-durables consumption, leisure, and housing services, respectively. Even though the structural regressions imply leisure is a better complement with air quality than housing, I cannot reject the null that these two elasticities are the same in these reduced form specifications.

6.5.1. Other Issues

A. Alternative Weather Instruments: The main orthogonality condition required to achieve unbiased estimates of air quality in the reduced form regressions is that household's consumption and leisure bundles are influenced by wind speeds only through air quality, conditional on controls and fixed effects. A reasonable alternative to using wind speeds as instruments might be precipitation. However, many alternative climate variables *do* influence household's locational choice and non-pecuniary utility. For example, areas that are subject to a lot of rain tend to be less desireable locations, holding all else constant. Indeed, temperature and snowfall, for example, make good instruments in the structural regressions that include leisure and housing.

B. Measurement of Pollution: The benchmark model has used a transformation over the EPA's recently created air quality index, whereas all of the literature thus far has exploited variation in different pollution criteria (e.g., TSPs). To the extent that households react to air quality and air quality alerts, rather than the mere presence of TSPs, identifying a credible elasticity might be even more of a complex task. To answer this question, replace S^{ω} with $P^{-\omega}$ ($\omega \in \{\phi, \psi, \zeta\}$) in Equation 1, where P denotes pollution (TSP). Because pollution is an economic bad, it is raised to the power of $-\omega$ for $\omega \in \{\phi, \psi, \zeta\}$ in order to characterize the fact that $1/P$ is decreasing in P—that is, higher pollution reduces the quality of consumption, leisure, and housing. The fact that the results are nearly identical is assuring that not only the transformation is reasonable, but also that the instruments are successful at identifying the causal effect of air quality on wages and the price of housing.

C. Capitalization Effects: Kuminoff and Pope (2014) show that capitalization effects may not identify willingness to pay since policy interventions affect the shadow price of the public good. To the extent that the variation during the Great Recession changed the shadow price of air quality—which can be studied through a test they provide for the gradient of the equilibrium price function before and after that defines the capitalization effect—then the WTP still is insufficient to identify welfare consequences. My structural approach provides a way of pinning down willingness to pay through a "consumption equivalent" approach having estimated the demand-side parameters.

D. Willingness to Pay v. Willingness to Accept: Amiran and Hagen (2014) caution that willingness to pay and willingness to accept—which are crucial for identifying the value of a non-market good—may differ, especially when exploiting cross-sectional variation since aggregation problems may arise. Determining the empirical relevance of the concern is likely to vary with the quality of the data (e.g., variation and richness of controls) and is the topic of a companion paper. Nonetheless, this approach provides serious advantages to the traditional hedonic approach. First, since matching site-specific amenities with housing unit and demographic records often occurs at different levels of aggregation, my approach avoids an errors-in-variables problem that would traditionally introduce noise and/or bias depending on the correlation of the errors with other independent variables. Second, while a standard challenge is addressing omitted variables concerns arising from unobserved heterogeneity in housing prices that are correlated with the amenity, my household level data is able to provide a much finer set of controls on local labor market outcomes and housing values.[68]

F. Long versus Short Run Elasticities: The classic controversy in macro-labor economics focuses on the identification of Frisch versus Hicksian elasticities; for example, see Chetty (2012) for the former and Keane and Rogerson (2012); Keane (2011) for the latter. Because my benchmark regressions exploit within county-industry deviations from state-level air quality within the same year, one concern is that my identification strategy only recovers short-run elasticities. In order to recover long-run elasticities, my regressions must contain a sufficient amount of cross-sectional variation. That is, observationally equivalent households must be observed sorting into different counties in order to recover an average treatment effect of air quality on the hourly wage. To let the data speak, I exploit the contrasting sources of variation in my IV specifications with and without fixed effects on state and/or county. In particular, I run regressions on Equation 4 with only year-by-industry fixed effects. I find that $\psi = -3$ and $\phi = .07$. Recognizing that local demand shocks are biasing ψ upwards and ϕ downwards, the similarity with the benchmark specification suggests that my identification strategy is doing quite well at capturing a long-run elasticity recovered through cross-sectional variation among observationally equivalent households living in different counties. Of course, future work should do a better job of modeling investment in *health* capital.

[68] See Bajari et al. (2012) for an alternative approach that exploits rational expectations over the unobserved error.

7. Implications for Macroeconomic Analysis

7.1. Effects on Labor Markets and Search Frictions

The mechanisms underpinning the substitutability between consumption and air quality, and complementarity between leisure & housing and air quality, are driven by the nonseparabilities between private goods and non-market goods. Policy interventions may have large general equilibrium effects that affect the valuation of environmental amenities. For example, larger cities have lower prices on consumption goods and worse air quality; they also have higher productivity in labor services and higher housing values. While the link between environmental policy and employment has been studied before (Greenstone, 2002; Walker, 2013; Curtis, 2014), my data set allows me to answer the question using a different identification strategy (e.g., instrumental variables) over a broader cross-section of the United States using a different proxy for environmental stringency (e.g., air quality).

The 2005-2010 period experienced variation in the stringency of environmental standards in a couple of ways. First, the fraction of counties in non-attainment rose by nearly 12%.[69] Because of the Great Recession, meeting certain attainment status targets may have become harder for certain counties. Second, as Kotchen and Mansur (2014) discuss, even though the American Clean Energy and Security Act did not pass in 2009, it prompted the EPA to issue rules for regulating emissions from fossil fuel generating facilities starting between 2009 and 2010.[70] The identifying source of variation in the following regression is rooted in the heterogeneous response among households and county industrial bases to changes in regulatory stringency. Of course, the fundamental limitation of these regressions, which cannot be completely ruled out, is the inherent local demand shocks that counties are facing over the 2005-2010 period.

A. Search Frictions: I first test for the effect of air quality on search frictions. If counties with better air quality only have it because of more stringent regulation, the regulatory effects could introduce inertia in the labor market to the extent that firms are less willing to hire and/or less willing to fire. Search and match frictions are approximated by studying the degree of labor market turnover (Davis and Haltiwanger, 1992). Denote the job finding rate between t and $t+1$ as $\theta_{c,i,t} = 1 - u_{c,ind,t+1}/u_{c,ind,t}$ where i denotes the group, partitioned by gender, age (decennial) bracket, number of children, and race, to estimate regressions of the form

$$\theta_{c,i,t} = \beta_0 + \beta_1 X_{c,t} + \alpha \log S_{c,t} + \epsilon_{c,t}$$

where X denotes county-state controls on economic activity and demographics. By using de-

[69] The statistic is computed using county-level data on attainment status, kindly provided by Wayne Gray.

[70] Most of the variation is coming subsequent to the period in my sample, but it is likely that plants acted in anticipation of the potential passage of the American Clean Energy and Security Act and/or authorized EPA rules.

viations from trend in the job finding rate, this regression purges all time invariant cross-county-industry unobserved heterogeneity. To the extent that there is time varying heterogeneity in the job finding rate that is correlated with air quality, instrumenting air quality with wind speeds should satisfy the exclusion restriction that unobserved shocks to the job finding rate are uncorrelated with wind speeds. $\hat{\alpha}$ characterizes the elasticity of labor market frictions to environmental policy: how a percent change in air quality affects the job finding rate.

Table 7: Air Quality and the Job Finding Rate

	OLS	FE		IV		
	(1)	(2)	(3)	(4)	(5)	(6)
Log Air Quality	-.532***	.166	1.379*	-1.297**	-2.570	-2.569
	[.117]	[.181]	[.724]	[.515]	[5.847]	[5.846]
Log Employment		.0901**	-1.142*			
		[.0417]	[.627]			
State FE	No	Yes	Yes	No	Yes	Yes
Year FE	No	Yes	Yes	No	Yes	Yes
Industry FE	No	No	No	No	No	Yes
County FE	No	No	Yes	No	Yes	Yes
Observations	233587	233587	233587	140698	140698	140698
Adjusted R^2	.005	.038	.175	.	.199	.198

Notes.–Sources: Census, CEX, ATUS, EPA, NOAA. These columns present the parameter estimates from a regression of the log job finding rate on log air quality and the usual controls, collapsed to the county-industry-year level. The job finding rate is computed by taking one minus the quotient of the number of workers by the population within a county for an industry-year pair. Column 1 runs OLS with the usual controls. Column 2 adds year and state fixed effects. Column 3 includes year-by-state and county fixed effects. Column 4 instruments for air quality using a quadratic in wind speeds interacted. Column 5 introduces year, industry, and county fixed effects. Column 6 is the same as column 5, but uses year-by-state fixed effects. All standard errors are clustered at the county-level and observations are weighted by Census person-weights an aggregated to agender, race, and number of children.

The naive OLS suggests that a 1% rise in air quality is associated with a .5% decline in the job finding rate. However, there are various reasons to think that this estimate is faulty. On one hand, unobserved heterogeneity in tastes for air quality are also correlated with income, and it is possible that more productive workers cluster in areas with greater job finding rates; in this case, the coefficient would be upwards biased. On the other hand, unobserved shocks to the job finding rate are negatively correlated with air quality since the more commerical locations also have more pollution; to the extent that these cases also have a higher job finding rate, the coefficient would be upwards biased. Adding in fixed effects (columns 2 and 3) drives up the estimate, which is consistent with the first hypothesis, especially visible since the inclusion of county-level fixed effects spikes the estimate from .16 to 1.4. However, introducing the instrument reverses the sign and provides evidence of an even more negative effect, relative to the naive OLS. To the extent that year-to-year fluctuations in wind speeds are uncorrelated with unobserved shocks to the job finding rate, then a 1% rise in air quality is associated with a 1.3% decline in the job finding rate. Columns 5 and 6 show that the result is robust to inclusion of more detailed fixed effects, although they lose statistical significance. The intuition behind these results is best absorbed by thinking in context of the Mortensen-Pissarides framework (Mortensen and Pissarides, 1994). Since higher air quality—conditional on county fixed effects—is a proxy for the stringency of environmental regulations, then the negative coefficient on air quality captures the fact that more stringent regulation decreases firms' demand for labor. In the MP model, this decreases the returns to search and, thus, the job-finding rate.

B. Local Demand for Labor: I now test for whether counties with higher air quality have different employment outcomes. Using a similar rationale, more stringent environmental regulation could reduce employment if pollution-intensive inputs are complementary with labor. Exchanging the dependent variable with log employment, there is a similar negative effect on employment.

Table 8: Air Quality and Employment

	OLS	FE		IV	
	(1)	(2)	(3)	(4)	(5)
(mean) laqi	-7.470***	-9.136***	-.0883*	-.326	-1.754
	[1.334]	[1.916]	[.0444]	[2.857]	[3.241]
State FE	No	Yes	No	No	Yes
Year FE	No	Yes	Yes	No	Yes
County FE	No	No	Yes	No	No
Observations	49468	49468	49468	22248	22248
Adjusted R^2	.152	.213	.999	.036	.210

Notes.–Sources: Census, CEX, ATUS, EPA, NOAA. These columns present the parameter estimates from a regression of the log job finding rate on log air quality and the usual controls, collapsed to the county-industry-year level. The job finding rate is computed by taking one minus the quotient of the number of workers by the population within a county for an industry-year pair. Column 1 runs OLS with the usual controls. Column 2 adds year and state fixed effects. Column 3 includes year and county fixed effects. Column 4 instruments for air quality using a quadratic in wind speeds interacted and year fixed effects. Column 5 introduces year and industry fixed effects. All standard errors are clustered at the state-level (since every observation is a county-year-industry) and observations are weighted by Census person weights and subsequently collapsed to the county level by population.

The naive OLS delivers the highly suspect results that a 1% rise in air quality is associated with a 7.4% decline in employment (in thousands). However, the naive OLS is inherently biased since negative shocks to productivity are positively correlated with air quality, thereby inducing a large downwards bias. Even including fixed effects fails to remedy the bias since the shocks are affecting local demand. However, after county-level fixed effects are introduced, the coefficient declines considerable in magnitude and suggests that a 1% rise in air quality is associated with a .08% decline in employment. Furthermore, after instrumenting for air quality using wind speeds, the coefficient becomes even more negative, ranging between -.3 and -1.8, albeit their estimates become less statistically significant since there is no longer any household-level variation.

7.2. A Welfare Evaluation from Changes in Air Quality

Using these estimated elasticities and willingness to pay for air quality, my results allow me to implement a welfare evaluation of changes in air quality, offering a point of comparison with the EPA results on the Clean Air Act Amendments. Higher levels of air quality are valued, but come at the cost of lower household real incomes since firms pass-through the cost of emissions in the form of higher prices (Fabra and Reguant, 2014). A common approach is to assume that preferences are homogeneous and linear with respect to air quality, making the marginal willingness to pay for air quality constant (Freeman, 1974). Unfortunately, the gradient of the hedonic price function provides only an average *marginal* willingness to pay for a one-unit change in air quality or pollution, rather than (total) willingness to pay through the identification of the entire hedonic price function. What's worse is the fact that the hedonic price function may shift over time based on unobserved shocks, undermining the mapping of MWTP estimates into WTP estimates both due to bias (mis-specification) and lack of identification of the new (shifted) hedonic price function (Kuminoff et al., 2010; Kuminoff and Pope, 2014). These purely reduced form methods are not helpful for policy analysis because they do not allow for changes in behavioral responses as a result of shifts in the hedonic price function.

A. Estimating Counterfactual Densities: My alternative strategy asks the following question: "What would households be willing to pay to keep air quality at their 2005 levels in 2010?" Mathematically, this means finding the Δ that solves the following equation.

$$u(C(1+\Delta), L, H, S) = u(\tilde{C}, \tilde{L}, \tilde{H}, S) \tag{12}$$

where $\tilde{D}$ denotes the counterfactual distribution of variable $D \in \{C, L, H\}$ in 2010. While all these variables are observed in the 2005 and 2010 data, the problem with computing Δ directly (without computing counterfactuals) is that Δ will be driven by changes in C and L, rather than just S—the object of the analysis. Answering this, therefore, requires knowledge of the counterfactual consumption, leisure, and air quality densities.[71] Counterfactuals—in this approach—have the interpretation of the density that would have prevailed if attributes (X) stayed the same at their

[71]Dinardo et al. (1996) developed an initial approach for computing these counterfactuals when separating between two groups (e.g., union status). Bound et al. (2010) has recently applied a version of the Dinardo et al. counterfactual simulation in the context of the economics of education. Heckman and Vytlacil (2005) studies it in the context of estimating treatment effects. The counterfactual density is given by $f^{2010}_{2005}(\ln D) =$

$$\int f^{2010}(\ln D|X)h(X|t=2005)dx = \int \left(\rho^{2005}(X)/(1-\rho^{2005}(X))\right)\left(P^{2010}/P^{2005}\right) f^{2010}(\ln D|X)h(X|t=2010)dx$$

where P denotes the proportion of observations in the corresponding years, ρ denotes the predicted probability of a county being a 2005 observation, given the observed distribution of characteristics X. ρ^{2005} can be estimated by running a probit model with X controls on the distribution of attributes within a county, including both household-level attributes and business patterns (e.g., employment and establishments). The counterfactual density is weighted using

2005 level and the relationship between the outcome variable of interest, $D \in \{C, L, H\}$, and economic returns are as observed in 2010.[72] These counterfactuals are displayed below.

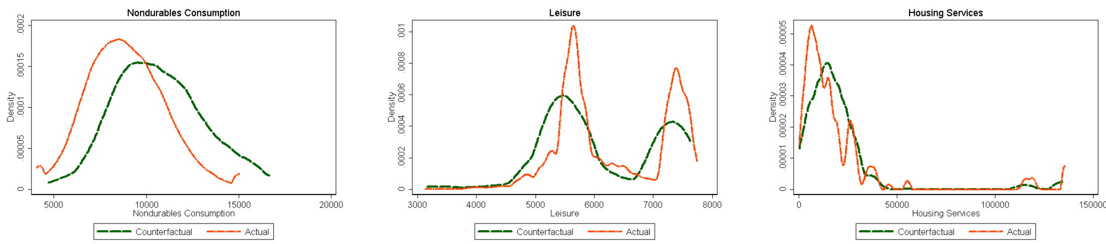


Figure 8: Counterfactual (2010) Consumption, Leisure, and Housing Densities
Notes.–Source: ACS, CEX, ATUS. These plots generate counterfactual 2010 densities for consumption, leisure, and housing. To do this, I generate a dummy variable for 2005 and 2010, together with the share of observations for both years. I subsequently run a probit regression of the dummy on household/demographic characteristics and industry dummies for the consumption weight, and add income and hours worked for the leisure weight. The counterfactual densities are given by: $\Phi(X) = \left(\rho^{2005}(X)/(1-\rho^{2005}(X))\right)\left(P^{2010}/P^{2005}\right)$, where ρ are the weights for simulating the counterfactual 2010 densities holding fixed the distribution in 2005, and P denotes the proportion of observations in the data set.

Much like the treatment effects literature has emphasized (Rosenbaum and Rubin, 1983; Heckman and Robb, 1985), typically these counterfactual exercises imply a causal interpretation only under the assumption of conditional exogeneity, or selection on observables. In order to map these counterfactual distributions into a policy exercise comparable to the EPA's evaluations of the CAAA, I will scale the WTP appropriately.

B. Calibration: Letting $\psi = -4.5$, $\phi = -.6$, and $\zeta = -.5$ from the benchmark specification, the three remaining parameters to calibrate are the α's. Since $\alpha = .48$ typically in log-log preferences of the form $\alpha \log C + (1-\alpha)\log L$ in business cycle models, I keep $\alpha_L = .52$ and set α_C and α_H to match the share of expenditures on non-durables and housing durables, respectively, between 2005-2010. The Bureau of Economic Analysis (BEA) national accounts implies that 60% of private expenditures on non-durables and housing consumption is spent on housing, meaning that $\alpha_C = .48 \times .4 = .192$ and $\alpha_H = .288$.[73] While $\mu = .4608$ and $\gamma = .5136$ are calibrated from Makridis (2014) to match features of the U.S. economy between 1970-2010, setting a reasonable value for π is trickier. However, since the implicit argument in the hedonics literature is that air quality is capitalized into housing values through its effect on land, then I defer to Davis and Heathcote

$$\Phi(X) = \left(\rho^{2005}(X)/(1-\rho^{2005}(X))\right)\left(P^{2010}/P^{2005}\right)$$

[72]General equilibrium changes in the distribution of attributes on the outcome variable are not considered. For example, in the context of air quality, this means that changes in the distribution of attributes did not affect the underlying structure of emissions intensity. I am grateful to David Autor's labor lecture notes for this interpretation; see p. 7: http://economics.mit.edu/files/7714.

[73]I am not including other durables, such as health and financial services. From the BEA national accounts, I include all units in the nondurable goods category (food/beverage, clothing/footwear, gasoline/energy, other) and transportation, recreation, and food services as a composite of non-durables.

(2005) who calibrate land's share in new housing to .106, meaning that $\pi = 1 - .106 = .894$.[74] Under this calibration, then the willingness to pay can be computed closed form.[75]

C. Implied Willingness to Pay for the CAAA: I find that $\Delta = -03$, meaning that households would be willing to forfeit 3% of their consumption in order to be indifferent between their 2005 benchmark level of consumption, relative to that experienced in 2010. How does this map into the EPA's CAAA evaluation? Since annual TSPs were 47.46 in 1990, 44.37 in 2005 and 39.72 in 2010, then the percent declines from 1990-2010 and 2005-2010, respectively, were 17% and 11%. Therefore, the counterfactual value for air quality will be a 17% rise, relative to that in 1990, so Δ should be scaled by 1.06. Using the result from Auffhammer et al. (2009) that approximately 11% of the decline in PM10 can be accounted for by the CAAA, then the rise in air quality needs to be scaled by an additional .11 term.[76] Multiplying this by the aggregate level of consumption expenditures in 2005 yields \$.68 billion—a quantity much lower than the EPA's estimated \$2 trillion.[77]

8. Conclusion

This paper has developed the most comprehensive database to date on household and county consumption, leisure, weather, and air quality outcomes in order to obtain the first elasticities between air quality and consumption-leisure-housing. My structural model provides a refinement upon

[74] Letting H^{adj} denote the air quality adjusted housing services from the CES aggregator with air quality, an alternative strategy is to realize that the share of income (utility) from housing, call it κ, is equal to: $\kappa = (\partial H^{adj}/\partial H)(H/H^{adj})$, or $\kappa = \pi(H/H^{adj})^{\zeta}$. Given data on rental rates on κ, then π can be estimated.

[75] Letting $F = F(\tilde{C}, \tilde{L}, \tilde{H}, \tilde{S}, L, H, S)$ denote a constant consisting of the 2010 counterfactual (and, for air quality, actual) and 2005 leisure/housing/air distributions

$$\Delta = \left\{ \left[\exp(F/\alpha_C)^{\phi} - (1-\mu)S^{\phi} \right] /\mu \right\}^{\frac{1}{\phi}} /C - 1$$

[76] The story for sulfur dioxide is even less optimistic. Greenstone (2004) shows that SO2 nonattainment status is only associated with small reductions in SO2 air pollution rationalized with two explanations. First, many states may not have had the resources to allocate towards developing air diffusion models that demonstrated that all areas within a county adhered to the attainment standards, meaning that many counties may have simply remained in nonattainment. Second, since the EPA was somewhat receptive to lessening the intensity of regulatory oversight in nonattainment counties that did not exceed the National Ambient Air Quality Standard, some counties may have opted simply for more relaxed regulatory oversight over going through the proper procedure to update their attainment status. However, it is possible, as Auffhammer et al. (2009) point out, that the results in Greenstone (2004) might just be averaged out due to too large of a spatial aggregation.

[77] While the same identification strategy could be used to obtain comparable measures of the value of a statistical life, doing so would require additional data on deaths at an industry-county level, which extends far beyond the scope of the analysis. However, doing so is important since nearly 90% of the EPA's estimated benefits of the CAAA are driven by a crude measure of the VSL—in particular, a measure that overestimates the preferred \$2-3 million range in the literature (Mrozek and Taylor, 2002) with approximately \$6 million. On top of the contrast from the best practices in the literature, Smith et al. (2003) emphasize that current measures of the VSL do not take into account endogenous labor supply, which biases the underlying wage/job risk combinations on a few orders of magnitude. I will pursue this in future work.

the canonical Roback (1982) treatment in hedonic theory where preferences are homogeneous and households make locational choice decisions based on preferences for environmental amenities and other considerations. The elasitcities between air quality and consumption, housing, and leisure are 1.35, .6, and .35, implying imperfect substitutability with consumption and complementarity with leisure and housing. Since the parameters are derived from the equilibrium conditions in a structural model, they naturally map into quantitative macroeconomic policy evaluations, like those in Makridis (2014) and Carbone and Smith (2008). My model also highlights the contrast between willingness to pay estimated from a Roback-Rosen framework by showing the ways in which cross-substitution between market and non-market goods in response to policy affects the valuation of non-market goods. My results are subjected to battery of robustness checks examing the plausibility of the identifying assumptions behind the exclusion restrictions of my instrumental variables strategies. I also recover qualitatively similar results from applying entirely separate instrumental variables and reduced form regressions that do not impose the same structural relationships.

My data set also enables me to study new sources that of bias that typically confound hedonic estimation. First, conventional studies do not differentiate between the price and product of price and quantity when regressing on environmental amenities. Even if variation in the amenity is purely exogenous, there is a structural correlation between the quantity of housing or labor services and air quality because of nonseparabilities between market and non-market goods. Second, unobserved shocks to wages and the price of housing services induce upwards bias in the conventional willingness to pay since local labor demand shocks will depress consumption and raise leisure. Third, the variance of these shocks varies dramatically between the county and household levels. Analyses focusing purely on phenomena at the county level effectively aggregate out much of the uncertainty at a household level that is correlated with unobserved tastes and reallocation.

While my benchmark results provide an average treatment effect, further analysis suggests that there is heterogeneity in tastes for air quality. Aside from contrasting with the Roback (1982) model—which my model is able to refine upon through exogenous variation induced by instruments—these results underscore the importance of incorporating different forms of heterogeneity in macroeconomic models. An important task for future research is to more explicitly model household's adaptation to fluctuations in environmental amenities, such as temperature and air quality, that influence the reallocation and cross-substitution patterns among traded goods; see Deschenes (2012) for a survey. I also use my estimated parameters to conduct a counterfactual exercise, asking how much a household would be willing to pay in order to be made indifferent between higher air quality in 2010, relative to 2005, holding fixed the distribution of covariates in non-durables, leisure, and housing services. Scaling the implied WTP appropriately in order to recover a WTP for the EPA's CAAA, I find a significantly lower level of net benefits associated with the legislation than they estimate: $.68 billion, rather than $2 trillion. Because households are able to cross-substitute through locational choices, the value of environmental amenities declines.

Although today's levels of pollution are lower, relative to the 1970s and 1980s, Currie and Neidell (2005) show that these lower levels of pollution are still harmful, which means that my estimates are informative for macroeconomic modeling. My results raise an array of new questions and empirical strategies. First and foremost, disentangling heterogeneity in preferences over private goods/services and amenities matters and cannot be captured purely by examining county-level data. Refinements should incorporate Roy sorting in a more structured way (e.g., as suggested by DeLeire et al. (2013)) in order to confront the inherent endogeneity in standard hedonic wage-risk regressions.[78] Second, there are a variety of distributional issues that can be investigated using accessible software routines (e.g., Frolich and Melly (2010)). In future work, I will use this database and extend it to include up until 2014 in order to study distributional effects of air quality throughout the Great Recession.

References

Aguiar, M. and M. Bils (2009): "Has Consumption Inequality Mirrored Income Inequality?" University of Rochester mimeo.

Aguiar, M. and E. Hurst (2007): "Measuring Trends in Leisure: The Allocation of Time over Five Decades," *Quarterly Journal of Economics*, 122, 969–1006.

Aguiar, M., E. Hurst, and L. Karabarbounis (2013): "Time use during the Great Recession," *American Economic Review*, 103, 1664–1696.

Ai, C. and X. Chen (2003): "Efficient estimation of models with conditional moment restrictions containing unknown functions," *Econometrica*, 71, 1795–1844.

Albouy, D., R. Kellogg, W. Graf, and H. Wolff (2013): "Climate amenities, climate change, and the American quality of life," *Journal of the Association of Environmental and Resource Economists, RR.*

Aldy, J. and W. K. Viscusi (2008): "Adjusting the value of a statistical life for age and cohort effects," *Review of Economics and Statistics*, 90, 573–581.

Altonji, J. G. (1986): "Intertemporal substitution in labor supply: Evidence from the micro data," *Journal of Political Economy*, 94, S176–S215.

Amiran, E. Y. and D. A. Hagen (2014): "Tests of the discrepency between willingness to accept and willingness to pay: A theoretical analysis," *Working paper*.

Aronow, P. M. and A. Carnegie (2013): "Beyond LATE: Estimation of the average treatment effect with an instrumental variable," *Political Analysis*, 1–15.

[78] Kniesner et al. (2012) take an alternate route with panel data and using first-differences to remove time invariant heterogeneity. However, as many have documented (Bound et al., 2001, 1994; Bound and Krueger, 1991), there are massive measurement error problems in the PSID and researchers need to weigh the costs and benefits of using panel data over small samples with cross-sectional data over large samples.

Attanasio, O. P. and H. Low (2004): "Estimating Euler Equations," *Review of Economic Dynamics*, 7, 406–435.

Attanasio, O. P. and G. Weber (1995): "Is Consumption Growth Consistent with Intertemporal Optimization? Evidence from the Consumer Expenditure Survey," *Journal of Political Economy*, 103, 1121–1157.

Auffhammer, M., A. M. Bento, and S. E. Lowe (2009): "Measuring the effects of the Clean Air Act Amendments on ambient PM10 concentrations: The critical importance of disaggregated analysis," *Journal of Environmental Economics and Management*, 58, 15–26.

Auffhammer, M. and R. Kellog (2011): "Clearing the air? The effect of gasoline content regulation on air quality," *American Economic Review*, 101, 2687–2722.

Aufhammer, M., S. M. Hsiang, W. Schlenker, and A. Sobel (2013): "Using weather data and climate model output in economic analyses of climate change," *Review of Environmental Economics and Policy*, 7, 181–198.

Bajari, P., J. C. Fruehwirth, K. I. Kim, and C. Timmins (2012): "A rational expectations approach to hedonic price reressions with time varying unobserved product attributes: The price of pollution," *American Economic Review*, 102, 1898–1926.

Banzhaf, S. H. and R. P. Walsh (2008): "Do people vote with their feet? An empirical test of Tiebouts mechanism," *American Economic Review*, 93, 843–863.

Baum-Snow, N. and R. Pavan (2012): "Understanding the city size wage gap," *Review of Economic Studies*, 79, 88–127.

Bayer, P., F. Ferreira, and R. McMillan (2007): "A unified framework for measuring preferences for schools and neighborhoods," *Journal of Political Economy*, 115, 588–638.

Bayer, P., N. Keohane, and C. Timmins (2009): "Migration and hedonic valuation: The case of air quality," *Journal of Environmental Economics and Management*, 58, 1–14.

Bayer, P. and S. L. Ross (2006): "Identifying individual and group effects in the presence of sorting: A neighborhood effects application," *NBER working paper*.

Bee, A., B. D. Meyer, and J. X. Sullivan (2012): "The validity of consumption data: Are the consumption expenditure interview and diary surveys informative?" *NBER working paper*.

Berry, S., A. Khwaya, A. Musalem, K. C. Wilbur, G. Allenby, B. Anand, P. Chintagunta, W. M. Hanemann, P. Jeziorski, and A. Mele (2014): "Structural models of complementary choices," *Marketing Letters*.

Bharadwaj, P., M. Gibson, C. A. Neilson, and J. G. Zivin (2014): "Grey matters: Fetal pollution exposure and human capital formation," *NBER working paper*.

Bieri, D. S., N. V. Kuminoff, and J. C. Pope (2014): "The role of local amenities in the national economy," *Working paper*.

Blundell, R., X. Chen, and D. Kristensen (2007): "Semi-nonparametric IV estimation of shape-invariant Engel curves," *Econometrica*, 75, 1613–1669.

Blundell, R., L. Pistaferri, and I. Preston (2008): "Consumption inequality and partial insurance," *American Economic Review*, 98, 1887–1921.

Bockstael, N. E. and K. McConnell (1993): "Public goods characteristics of nonmarket commodities," *Economic Journal*, 103, 1244–1257.

Bockstael, N. E. and K. E. McConnell (2007): "Environmental and resource valuation with revealed preferences: A theoretical guide to empirical models," *Springer*.

Boualam, B. (2014): "Getting a first job: Quality of the labor matching in French cities," *University of Geneva, Job Market Paper*.

Bound, J., C. Brown, G. J. Duncan, and W. L. Rodgers (1994): "Evidence on the validity of cross sectional and longitudinal labor market data," *Journal of Labor Economics*, 12, 345–368.

Bound, J., C. Brown, and N. Mathiowetz (2001): "Measurement error in survey data," *Handbook of Econometrics*, 5.

Bound, J. and A. B. Krueger (1991): "The extent of measurement error in longitudinal earnings data: Do two wrongs make a right?" *Journal of Labor Economics*, 9, 1–24.

Bound, J., M. F. Lovenheim, and S. Turner (2010): "Why have college completion rates declined? An analysis of changing student preparation and collegiate resources," *American Economic Journal: Applied Economics*, 2, 129–157.

Bushnell, J. B. and E. T. Mansur (2005): "Consumption under noisy price signals: A study of electricity retail rate deregulation in San Diego," *Journal of Industrial Economics*, 53, 493–513.

Campos, R. G. and I. Reggio (2014): "Measurement error in imputation procedures," *Economic Letters*, 122, 197–202.

Carbone, J. and V. K. Smith (2008): "Evaluating policy interventions with general equilibrium externalities," *Journal of Public Economics*, 92, 1254–1274.

——— (2013): "Valuing nature in a general equilibrium," *Journal of Environmental Economics and Management*, 66, 72–89.

Carroll, C. D. (2001): "Death to the Log-Linearized Consumption Euler Equation! (And Very Poor Health to the Second-Order Approximation)," *Advances in Macroeconomics*, 1, issue 1, Article 6.

Castle, E. N. (1965): "The market mechanism, externalities, and land economics," *Journal of Farm Economics*, 47, 542–556.

Cerulli, G. (2014): "ivtreatreg: A command for fitting binary treatment models with heterogeneous response to treatment and unobservable selection," *Stata Journal*, 14, 453–480.

Chay, K. Y. and M. Greenstone (2003): "The impact of air pollution on infant mortality: Evidence from geographic variation in pollution shocks induced by a recession," *Quarterly Journal of Economics*, 118, 1121–1167.

——— (2005): "Does air quality matter? Evidence from the housing market," *Journal of Political Economy*, 90, 1257–78.

Chen, X. (2007): "Large sample sieve estimation of semi-nonparametric models," *Handbook of Econometrics*, 6, 5550–5588.

Chen, X., G. Hong, and A. Tarozzi (2008): "Semiparametric efficiency in GMM models with auxiliary data," *Annals of Statistics*, 36, 808–843.

Chen, X., H. Hong, and D. Nekipelov (2011): "Nonlinear models of measurement errors," *Journal of Economic Literature*, 49, 901–937.

Chetty, R. (2012): "Bounds on elasticities with optimization frictions: A synthesis of micro and macro evidence on labor supply," *Econometrica*, 80, 969–1018.

Corlett, W. J. and D. C. Hague (1953): "Complementarity and the excess burden of taxation," *Review of Economic Studies*, 21, 21–30.

Crump, R. K., V. J. Hotz, G. W. Imbens, and O. A. Mitnik (2008): "Nonparametric tests for treatment effect heterogeneity," *Review of Economics and Statistics*, XC, 389–405.

Currie, J. and M. Neidell (2005): "Air pollution and infant health: What can we learn from California's recent experience," *Quarterly Journal of Economics*, 120, 1003–1030.

Currie, J., J. G. Zivin, J. Mullins, and M. Neidell (2014): "What do we know about short and long term effects of early exposure to pollution," *Annual Review of Resource Economics*, 6, 1–31.

Curtis, M. (2014): "Who Loses under Power Plant Cap and Trade Programs? Estimating the Impact of the NOx Budget Trading Program on Manufacturing Employment," *Working paper*.

Davis, D. R. and J. I. Dingel (2014): "The comparative advantage of cities," *NBER working paper*.

Davis, M. and J. Heathcote (2005): "Housing and the business cycle," *International Economic Review*, 46, 751–784.

Davis, O. A. and A. Whinston (1962): "Externalities, welfare, and the theory of games," *Journal of Political Economy*, LXX, 241–262.

Davis, S. and J. Haltiwanger (1992): "Gross Job Creation, Gross Job Destruction, and Employment Reallocation," *Quarterly Journal of Economics*, 107, 819–863.

Deaton, A. (2009): "Instruments of development: Randomization in the tropics, and the search for the elusive keys to economic development," *Proceedings of the British Academy, 2008 Lecture*, 162, 123–160.

DeLeire, T., S. Khan, and C. Timmins (2013): "Roy model sorting and nonrandom selection in the valuation of a statistical life," *International Economic Review*, 54, 279–306.

Deschenes, O. (2012): "Temperature, human health, and adaptation: A review of the empirical literature," *NBER working paper*.

Diamond, P. A. and J. A. Mirrlees (1973): "Aggregate production and consumption externalities," *Quarterly Journal of Economics*, 87, 1–24.

Dinardo, J., N. Fortin, and T. Lemieux (1996): "Labor market institutions and the distribution of wages, 1973-1992: A semi Parametric approach," *Econometrica*, LXIV, 1001–1044.

Ekeland, I., J. J. Heckman, and L. Nesheim (2004): "Identification and estimation of hedonic models," *Journal of Political Economy*, 112, S60–S109.

Epple, D. (1987): "Hedonic prices and implicit markets: Estimating demand and supply functions for differentiated products," *Journal of Political Economy*, 95, 59–80.

Epple, D. and G. J. Platt (1998): "Equilibrium and local redistribution in an urban economy hwen households differ in both preferences and incomes," *Journal of Urban Economics*, 43, 23–51.

Epple, D. and H. Sieg (1999): "Estimating equilibrium models of local jurisdictions," *Journal of Political Economy*, 107, 645–681.

Fabra, N. and M. Reguant (2014): "Pass through of emissions costs in electricity markets," *American Economic Review.*

Feenberg, D. and E. Coutts (1993): "An introduction to the Taxsim model," *Journal of Policy Analysis and Management*, 12.

Finkelstein, A., E. F. P. Luttmer, and M. J. Notowidigdo (2013): "What good is wealth without health? The effect of health on the marginal utility of consumption," *Journal of the European Economic Association*, 11, 221–258.

Flores, N. E. and R. T. Carson (1997): "The relationship between the income elasticities of demand and willingness to pay," *Journal of Environmental Economics and Management*, 33, 287–295.

Flores, N. E. and P. E. Graves (2008): "Optimal public goods provision: Implications of endogenizing the labor/leisure choice," *Land Economics*, 84, 701–707.

Freeman, A. M. I. (1974): "On estimating air pollution control benefits from land value studies," *Journal of Environmental Economics and Management*, 1, 74–83.

Frolich, M. and B. Melly (2010): "Estimation of quantile treatment effects with Stata," *State Journal*, 10, 423–457.

Fu, S. and S. L. Ross (2013): "Wage premia in employment clusters: How important is worker heterogeneity?" *Journal of Labor Economics*, 31, 271–304.

Gibson, J. (2002): "Why does the Engel method work? Food demand, economies of size and household survey methods," *Oxford Bulletin of Economics and Statistics*, 64, 341–359.

Glaeser, E. L. and M. G. Resseger (2009): "Complementarity between cities and skills," *NBER working paper.*

Goulder, L. H. and R. C. Williams III (2003): "The substantial bias from ignoring general equilibrium effects in estimating excess burden, and a practical solution," *Journal of Political Economy*, 111, 898–927.

Greenstone, M. (2002): "The impacts of environmental regulations on industrial activity: Evidence from the 1970 and 1977 Clean Air Act Amendments and the Census of Manufacturers," *Journal of Political Economy*, 110, 1175–1219.

——— (2004): "Did the Clean Air Act cause the remarkable decline in sulfur dioxide concentrations?" *Journal of Environmental Economics and Management*, 47, 585–611.

Greenstone, M., O. Deschenes, and J. S. Shapiro (2013): "Defensive investments and the demand for air quality: Evidence from the NOx budget program and ozone reductions," *American Economic Review.*

Greenstone, M. and T. Gayer (2009): "Quasi-experimental and experimental approaches to environmental economics," *Journal of Environmental Economics and Management*, 57, 21–44.

Grossman, G. and A. Krueger (1995): "Economic Growth and the Environment," *Quarterly Journal of Economics*, 110, 353–377.

Guojun, H. (2013): "The effect of air pollution on cardiovascular mortality: Evidence from the 2008 Beijing Olympic games," *Berkeley ARE JMP.*

Guvenen, F. (2012): "Macroeconomics with heterogeneity: A practical guide," *Working paper.*

Guvenen, F. and A. Smith (2014): "Inferring labor income risk and partial insurance from economic choices," *Econometrica*, forthcoming.

Hai, R., D. Krueger, and A. Postlewaite (2015): "On the welfare cost of consumption fluctuations in the presence of memorable goods," *Working paper*.

Hall, R. E. and C. I. Jones (2007): "The Value of Life and the Rise in Health Spending," *Quarterly Journal of Economics*, 122, 39–72.

Hall, R. E. and P. R. Milgrom (2008): "The Limited Influence of Unemployment on the Wage Bargain," *American Economic Review*, 98, 1653–1674.

Hall, R. E. and F. Mishkin (1982): "The Sensitivity of Consumption to Transitory Income: Estimates from Panel Data on Households," *Econometrica*, 50, 461–481.

Handbury, J. and D. E. Weinstein (2014): "Goods prices and availability in cities," *Review of Economic Studies*, forthcoming.

Hanemann, W. M. and E. Morey (1992): "Separability, partial demand systems, and consumer's surplus measures," *Journal of Environmental Economics and Management*, 22, 241–258.

Hanna, R. and P. Oliva (2015): "The effect of pollution on labor supply: Evidence from a natural experiment in Mexico City," *Journal of Public Economics*.

Harding, J. P., S. S. Rosenthal, and C. F. Sirmans (2007): "Depreciation of housing capital, maintenance, and house price inflation: Estimates from a repeat sales model," *Journal of Urban Economics*, 61, 193–217.

Hassler, J., P. Krusell, and C. Olovsson (2012): "Energy saving technical change," *American Economic Review, RR*.

Heathcote, J., F. Perri, and G. L. Violante (2010a): "Unequal we stand: An empirical analysis of economic inequality in the United States, 1967-2006," *Review of Economic Dynamics*, 13, 15–51.

Heathcote, J., K. Storesletten, and G. L. Violante (2010b): "The macroeconomic implications of rising wage inequality in the United States," *Journal of Political Economy*, 118, 681–722.

——— (2014): "Consumption and labor supply with partial insurance: An analytical framework," *American Economic Review*, 104, 1–52.

Heckman, J. J., R. L. Matzkin, and L. Nesheim (2010): "Nonparametric identification and estimation of nonadditive hedonic models," *Econometrica*, 78, 1569–1591.

Heckman, J. J. and R. Robb (1985): "Alternative methods for evaluating the impact of interventions," *Longitudinal Analysis of Labor Market Data*.

Heckman, J. J. and S. Urzua (2010): "Comparing IV with structural models: What simple IV can and cannot identify," *Journal of Econometrics*, 156, 27–37.

Heckman, J. J. and E. Vytlacil (2005): "Structural equations, treatment effects and econometric policy evaluation," *Econometrica*, 73, 669–738.

Heckman, J. J. and E. J. Vytlacil (2007): "Econometric evaluation of social programs, Part I: Causal models, structural models and econometric policy evaluation," *Handbook of Econometrics*, 6, 4779–4874.

Hellerstein, J. K. and G. W. Imbens (1999): "Imposing moment restrictions from auxilary data by weighting," *Review of Economics and Statistics*, LXXXI, 1–14.

Imai, S. and M. P. Keane (2004): "Intertemporal labor supply and human capital accumulation," *International Economic Review*, 45, 601–641.

Imbens, G. (2010): "Better LATE than nothing: Some comments on Deaton (2009) and Heckman and Urzua (2009)," *Journal of Economic Literature*, 48, 399–423.

Juhn, C., K. M. Murphy, and B. Pierce (1993): "Wage inequality and the rise in returns to skill," *Journal of Political Economy*, 101, 410–442.

Keane, M. and R. Rogerson (2012): "Micro and macro labor supply elasticities: A reassessment of conventional wisdom," *Journal of Economic Literature*, 50, 464–476.

Keane, M. P. (2011): "Labor supply and taxes: A survey," *Journal of Economic Literature*, 49, 961–1075.

Klaiber, A. H. and V. K. Smith (2012): "Developing general equilibrium benefit analyses for social programs: An introduction and example," *Journal of Benefit-Cost Analysis*, 3.

Kmenta, J. (1967): "On the estimation of the CES production function," *International Economic Review*, 8, 180–189.

Kneese, A. V., V. Ayres, and R. C. d'Arge (1970): "Economics and the environment: A materials balance approach," *Baltimore: John Hopkins University Press for Resources for the Future*.

Kniesner, T. J., W. K. Viscusi, C. Woock, and J. P. Ziliak (2012): "The value of a statistical life: Evidence from panel data," *Review of Economics and Statistics*, 94, 74–87.

Kopczuk, W. (2003): "A note on optimal taxation in the presence of externalities," *Economic Letters*, 80, 81–86.

Kotchen, M. J. and E. T. Mansur (2014): "How stringent are the US EPA's proposed carbon pollution standards for new plants?" *Review of Environmental Economics and Policy*, 8, 290–306.

Kuminoff, N. (2012): "Partial identification of preferences from a dual-market sorting equilibrium," *Econometrica, RR.*

Kuminoff, N., C. Timmins, and V. K. Smith (2013): "The new economics of equilibrium sorting and policy evaluation using housing markets," *Journal of Economic Literature*, 51, 1007–1063.

Kuminoff, N. V., C. F. Parmeter, and J. C. Pope (2010): "Which hedonic models can we trust to recover the marginal willingness to pay for environmental amenities?" *Journal of Environmental Economics and Management*, 60, 145–160.

Kuminoff, N. V. and J. C. Pope (2014): "Do "capitalization" effects for public goods reveal the public's willingness to pay?" *International Economic Review*, 55, 1227–1250.

Lemieux, T. (2006): "Increasing residual wage inequality: Composition effects, noisy data, or rising demand for skill?" *American Economic Review*, 96, 461–498.

Ludvigson, S. and C. Paxson (2001): "Approximation bias in linearized Euler equations," *Review of Economics and Statistics*, 83, 242–256.

Makridis, C. (2014): "Towards simple environmental policy rules: Fiscal interactions, time inconsistency, and externalities in general equilibrium," *Working paper*.

Maler, K.-G. (1974): "Environmental economics: A theoretical inquiry," *Baltimore: Resources for the Future*.

Mansfield, C., J. F. Reed, and G. Van Houtven (2006): "The missing piece: Valuing averting behavior for children's ozone exposures," *Resource and Energy Economics*, 28, 215–228.

Mian, A., K. Rao, and A. Sufi (2013): "Household balance sheets, consumption, and the economic slump," *Quarterly Journal of Economics*, 128, 1687–1726.

Mian, A. and A. Sufi (2011): "House prices, home equity-based borrowing, and the US household leverage crisis," *American Economic Review*, 101, 2132–56.

——— (2012): "What Explains High Unemployment? The Aggregate Demand Channel," University of Chicago Booth School of Business.

——— (2014): "What explains the 2007-2009 drop in employment?" *Econometrica*, 82, 2197–2223.

Mora, R. and I. Reggio (2012): "Treatment effect identification using alternative parallel assumptions," *Working paper*.

——— (Forthcoming): "dqd: A command for treatment effect estimation under alternative assumptions," *Stata Journal*.

Moretti, E. (2010): "Local labor markets," *Handbook of Labor Economics*, 4b, 1237–1309.

Moretti, E. and M. Neidell (2009): "Pollution, health, and avoidance behavior: Evidence from the ports of Los Angeles," *Journal of Human Resources*, 46, 154–175.

Mortensen, D. T. and C. Pissarides (1994): "Job Creation and Job Destruction in the Theory of Unemployment," *Review of Economic Studies*, 61, 397–415.

Mrozek, J. and L. Taylor (2002): "What determines the value of life? A meta-analysis," *Journal of Policy Analysis and Management*, 21, 253–270.

Neidell, M. (2007): "Information, avoidance behavior, and health: The effects of ozone on asthma hospitalizations," *Journal of Human Resources*, 44, 450–478.

Neidell, M. and J. G. Zivin (2013): "Environment, health, and human capital," *Journal of Economic Literature*, 51, 689–730.

Perroni, C. (1992): "Homothetic representations of regular non-homothetic preferences," *Economics Letters*, 40, 19–22.

Pope, J., C. Makridis, and N. Kuminoff (2014): "Perception versus reality: Implications for hedonic property value models," *Working paper*.

Prescott, E. (1997): "On defining real consumption," *Review of the Federal Reserve Bank of St. Louis*, 47–53.

Ramey, V. A. and N. Francis (2009): "A century of work and leisure," *American Economic Journal: Macroeconomics*, 1, 189–224.

Reiss, P. C. and M. W. White (2005): "Household electricity demand, revisited," *Review of Economic Studies*, 72, 883.

Rhode, P. W. and K. S. Strumpf (2003): "Assessing the importance of Tiebout sorting: Local heterogeneity from 1850 to 1990," *American Economic Review*, 93, 1648–1677.

Roback, J. (1982): "Wages, rents, and the quality of life," *Journal of Political Economy*, 90.

Robins, J. M., A. Rotnitzky, and L. P. Zhao (1994): "Estimation of regression coefficients when some regressors are not always observed," *Journal of the American Statistical Association*, 89, 846–866.

Rogerson, R. (1988): "Indivisible Labor, Lotteries and Equilibrium," *Journal of Monetary Economics*, 21, 3–16.

Rosen, S. (1974): "Hedonic prices and implicit markets: Product differentiation in pure competition," *Journal of Political Economy*, 82, 34–55.

Rosenbaum, P. and D. Rubin (1983): "The central role of the propensity score in observational studies for causal effects," *Biometrika*, 70, 41–55.

Sanders, N. J. (2012): "What doesn't kill you makes you weaker: Prenatal pollution exposure and educational outcomes," *Journal of Human Resources*, 47, 826–850.

Sandmo, A. (1975): "Optimal taxation in the presence of externalities," *Swedish Journal of Economics*, 77, 86–98.

——— (1980): "Anomaly and stability in the theory of externalities," *Quarterly Journal of Economics*, 94, 799–807.

Schlenker, W. and R. Walker (2012): "Airports, air pollution, and contemporaneous health," *Review of Economic Studies, RR.*

Sexton, A. L. (2011): "Responses to air quality alerts: Do Americans spend less time outdoors," *UMN JMP.*

Shapiro, J. S. and R. Walker (2014): "Why is U.S. air quality improving? The roles of trade, regulation, productivity, and preferences," *Working paper.*

Shibayama, K. and I. Fraser (2014): "Nonhomothetic growth models for the environmental Kuznets curve," *International Economic Review*, 55, 919–942.

Shin, J.-S. (1985): "Perception of price when price information is costly: Evidence from residential electricity demand," *Review of Economics and Statistics*, 67, 591–598.

Sieg, H., V. K. Smith, S. Banzhaf, and R. Walsh (2004): "Estimating the general equilibrium benefits of large changes in spatially delineated public goods," *International Economic Review*, 45, 1047–1077.

Slesnick, D. T. (1998): "Empirical approaches to the measurement of welfare," *Journal of Economic Literature*, XXXVI, 2108–2165.

Smith, V. K. and J. C. Carbone (2008): "Environmental economics and the "curse" of the circular flow," *Frontiers in Resource and Rural Economics.*

Smith, V. K. and J.-C. Huang (1995): "Can markets value air quality? A meta analysis of hedonic property value models," *Journal of Political Economy*, 103, 209–227.

Smith, V. K., S. K. Pattanayak, and G. L. Van Houtven (2003): "VSL reconsidered: What do labor supply estimates reveal about risk preferences?" *Economics Letters*, 80, 147–153.

Van Hee, V. C. and C. A. Pope (2012): "From Olympians to mere mortals: The indiscriminate, global challenges of air pollution," *American Journal of Respiratory and Critical Care Medicine.*

Viscusi, W. (1979): "Employment hazards: An investigation of market performance," *Cambridge: Harvard University Press.*

Viscusi, W. K. and J. E. Aldy (2003): "The Value of a Statistical Life: A Critical Review of Market Estimates throughout the World," *Journal of Risk and Uncertainty*, 27, 5–76.

Walker, R. W. (2013): "The transitional costs of sectoral reallocation: Evidence from the Clean Air Act and the workforce," *Quarterly Journal of Economics*, 128, 1787–1835.

West, S. E. and R. C. Williams III (2007): "Optimal taxation and cross-price effects on labor supply: Estimates of the optimal gas tax," *Journal of Public Economics*, 91, 593–617.

Xie, Y., J. E. Brand, and B. Jann (2012): "Estimating heterogeneous treatment effects with observational data," *Sociological Methodology*, 42, 314–347.

Ziliak, J. P. and T. J. Kniesner (2005): "The effect of income taxation on consumption and labor supply," *Journal of Labor Economics*, 23, 769–796.

Zivin, J. G. and M. Neidell (2012): "The impact of pollution on worker productivity," *American Economic Review*, 102, 3652–3673.

——— (2014): "Temperature and the allocation of time: Implications for climate change," *Journal of Labor Economics*, 32, 1–26.

Appendix for *The Elasticity of Air Quality: Evidence from Millions of Households Across the United States*

Christos Makridis, Stanford University*

April 23, 2015

PRELIMINARY AND INCOMPLETE, do not distribute without permission.

1 Supplements for the Introduction

1.1 Motvating Graphs

Interestingly, the micro-data is also consistent with the macro-data: countries with higher consumption as a share of GDP and hours worked tend to have higher levels of pollution.

Since air quality follows an inverse relationship with pollution, and hours follows an inverse relationship with leisure, then 1 suggests that there is imperfect substitutability between consumption and air quality and weak complementarity between leisure and air quality; the regression coefficients on both are much lower than unity (approximately .6 on consumption and .2 for hours.

The counties that polluted more in 1990 are also the counties that tend to pollute more today. To characterize this relationship, Figure 2 depicts a basic scatterplot of these two periods. The positive linear relationship implies that the underlying industrial base contributing to emissions in these counties has not shifted heavily. In fact, 60% and 45% of the variation in 2010 pollution levels can be explained purely from 1990 levels.

1.2 Simple Numerical Example

To highlight the fundamental importance of understanding the relationship between private goods & services and air quality, consider the simplest example possible. Let $T = 2$ and define $u(c_t; S_t) = c_t^\alpha \exp(S_t)^{1-\alpha}$, where $\alpha \in (0,1)$ governs the elasticity and share of consumption versus environmental quality. Initial allocations are $(c^0, S^0) = (1,2)$. Consider a proposed policy intervention that would induce the new bundle $(c^1, s^1) = (.5, 3)$. Depending on the elasticity (α), the welfare implications—measured through lifetime consumption equivalents (e.g., WTP)—are qualitatively different.

Varying the parameter governing preferences over consumption versus air quality, α, Figure 3 shows that households need to be compensated in order to be indifferent between remaining in the "clean" economy with higher air quality and lower consumption when $\alpha \geq .55$. In contrast, when $\alpha \leq .55$, households are willing to forfeit consumption in order to stay in the clean economy than to return to the "dirty" economy with higher consumption and lower air quality. In other words, the nonseparable interactions between non-market goods and private goods have not just quantitative, but qualitative, effects on welfare evaluation. In fact, the

*Stanford University, Working Paper. Email: cmakridi@stanford.edu; LinkedIn: www.linkedin.com/in/christosmakridis/. PRELIMINARY AND INCOMPLETE, do not cite. I am extremely grateful to Kerry Smith and Nicolai Kuminoff for their frequent advice and inspiration of the research topic, and Han Hong for many conversations about econometric theory. Thank you also to Luigi Pistaferri for sponsoring my access to restricted Panel Study of Income Dynamics data, Constantine Yannelis, Robbie Weterings for assistance with the weather and geospatial data, Ryan Pfirrmann-Powell for help with the CEX data, and seminar participants at Arizona State University, Chapman University, and Stanford University. Partially funded by the NSF Graduate Research Fellowship.

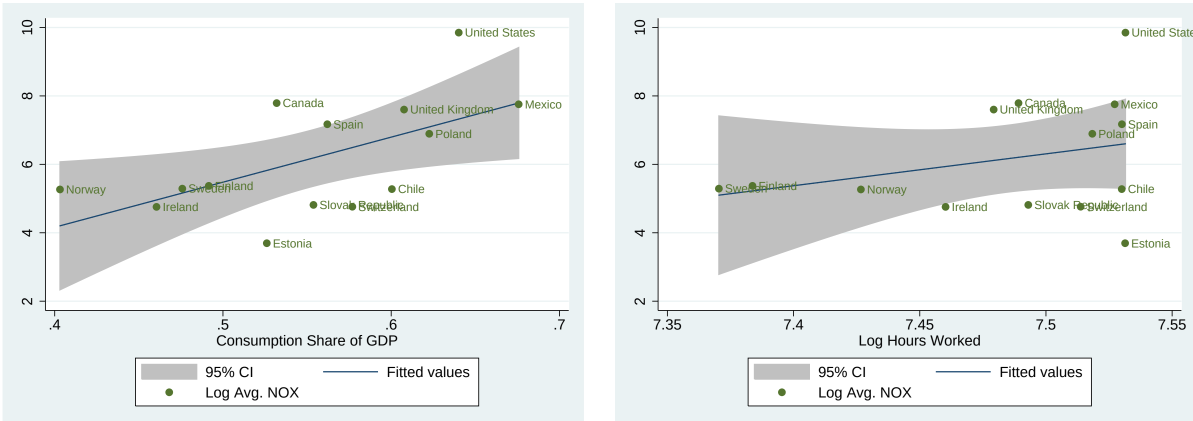


Figure 1: Cross-country Pollution (NOX) and Consumption/Hours Worked

Notes.–Source: OECD environmental and national accounts data. Panel A plots the log of nitrogen oxide with (private) consumption as a share of GDP. Panel B plots the log of nitrogen oxide with hours worked, demeaned of cross-country differences in union density. Hours worked is measured as the average annual hours actually worked per worker. Pollutants are measured in thousands of tons. The relationships for consumption and leisure with pollution hold with other pollutants, including particulate matter and sulfur oxide.

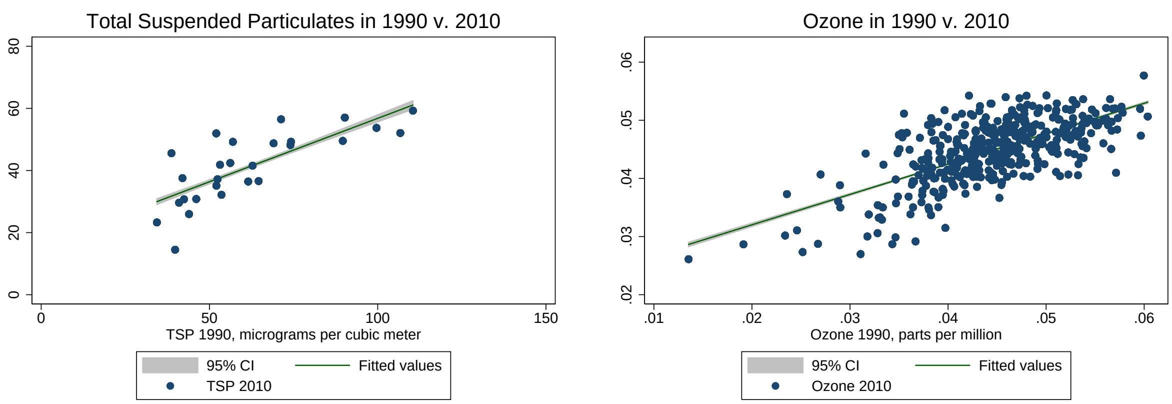


Figure 2: Pollution Criteria over Time

Notes.–Source: EPA Annual Summary Files. The figures plot total suspended particulates (24-hour arithmetic average of micrograms per cubic meter) and ozone (8-hour arithmetic average of parts per million) in 1990 and 2010 at the county-level.

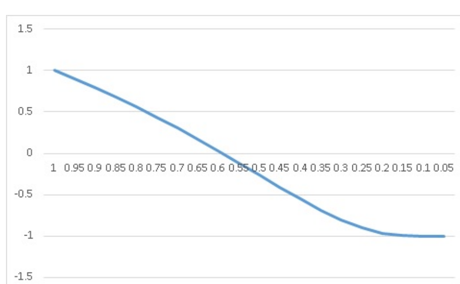


Figure 3: Motivating Willingness to Pay Exercise

Notes.–Letting $\beta = .99$, then compute Δ such that $u(c(1+\Delta); S)^{new} = u(c; S)^{old}$. Denote $\tilde{x}$ as the prospective period-2 allocation for variable $\tilde{x}$ and solve $\Delta = \left[\tilde{c}(\exp(\tilde{S})/\exp(S))^{1-\alpha}\right]^{1/\alpha}/c - 1$ over $0 \leq \alpha \leq 1$. The result has the interpretation of the amount of consumption that it would take to make the representative household's indifferent between the two options.

income elasticity of demand may not equal the income elasticity of WTP for the environmental good under nonseparability, even in the simplest environment possible with preferences linear in consumption (Flores and Carson, 1997; Ebert, 2003; Kristom and Riera, 1996).[1]

2 Extensions in the Analytical Model

2.1 Demand for Air Quality

Since skill biased technical change literature provides a useful lens for thinking about changes in the demand for air quality, I leverage these class of models (Katz and Murphy, 1992; Autor et al., 1998, 2003)to derive useful relationships between prices and quantities with respect to the pollution externality. Assume that $\gamma = \mu \equiv \theta$ and $\phi = \psi \equiv \omega$. Letting $X \in \{C, L\}$, then rewriting the two subutilities yields

$$\tilde{X} = \left[\theta(g_X X)^{\omega} + (1-\theta)(g_S S)^{\omega}\right]^{\frac{\eta}{\omega}}, \quad X \in \{C, L\} \tag{1}$$

where $\tilde{C}$ and $\tilde{L}$ now denote the aggregated consumption-environment and leisure-environment quantities. The challenge is that the subutility functions do not have a straightforward interpretation for an individual household since environmental amenities are not *private* goods and, even if they were, the presence of non-neutral technological shocks affects the valuation of consumption and leisure. In other words, even if technology and policy shocks do not directly affect household preferences over environmental quality directly, these shocks affect relative prices in general equilibrium by changing firms' production possibilities; price changes over private goods affects the degree of substitution with environmental amenities. To be clear, assume that households maximize utility without frictions and take logarithms / first differences over Equation 1

$$\Delta \log\left(\frac{p_X}{p_S}\right) = \Delta \log\left(\frac{\theta}{1-\theta}\right) + \left(\frac{\varepsilon_X - 1}{\varepsilon_X}\right)\Delta \log\left(\frac{g_X}{g_S}\right) - \frac{1}{\varepsilon_X}\Delta \log\left(\frac{X}{S}\right), \quad X \in \{C, L\} \tag{2}$$

If production is exogenously specified, Equation 2 completely characterizes the evolution of consumption and leisure prices over time. Importantly, changes in the relative price of consumption and leisure, which are guaranteed to occur in the presence of policy intervention (Makridis, 2014b), reflect either changes in the relative demand for consumption & leisure or changes in technology. In order to appreciate the effects of policy shocks on the demand for environmental quality-adjusted consumption and leisure, rewrite Equation 2 as follows

$$\log\left(\frac{p_X}{p_S}\right) = \frac{1}{\varepsilon_X}\left[D_t - \log\left(\frac{X}{S}\right)\right]$$

where $D = \varepsilon_X \log\left[(\theta/(1-\theta) + (\varepsilon_X - 1)(g_X/g_S)\right]$ represents shifts in the demand for environmental quality over consumption & leisure. The setup illustrates that changes in environmental quality can affect the demand for consumption and leisure differently by through a bias in technological change, which depends on the extent to which changes in D are driven by non-neutral shocks. For example, if air quality is a substitute for leisure for lower income households, but complement for higher income households, then environmental taxation implicitly induces redistribution to lower income households. These relative demand shifters for environmental quality-adjusted consumption and leisure take the form

$$D = \log\left(\frac{p_X X}{p_S S}\right) + (\varepsilon_X - 1)\log\left(\frac{p_X}{p_S}\right) \tag{3}$$

While a main objective in environmental economics is inferring p_S despite the fact that market prices are not defined for externalities, the more nuanced challenge is that unobserved heterogeneity and/or shocks to environmental quality covary with private goods and services. For example, since households with preferences for environmental quality will locate to areas with higher environmental quality ("locational sorting"), even though household choices reveal information, it is endogenous to the underlying problem.

[1] Hanemann (1991) demonstrated that the income sensitivity for a non-price rationed public good is very different from that of a private good.

2.2 Relationship with the Roback Model

The canonical Roback (1982) model considers households who maximize preferences subject to a budget constraint

$$\max U(C, H; S_i) \text{ s.t. } C + p_i H = W_i$$

where i denotes the spatial location, and W denotes labor income. In equilibrium, households must be indifferent among their locations, meaning that indirect utility is given by

$$V(W_i, p_i; S_i) = \overline{V}$$

Taking the total derivative and exploiting Roy's identity induces the following equilibrium relationship

$$p = H\frac{dp}{dS} - \frac{dW}{dS}$$

where p governs households' willingness to pay for a change in local public goods. Bayer et al. (2009) show that households' WTP is underestimated if mobility costs are abstracted upon. The expression here is similar to the intratemporal Euler on housing and labor, but abstracts for endogenous labor supply. The regression of interest involves the relative price ratio of housing to labor supply. In the absence of endogenous labor supply, willingness to pay will be overestimated since households substitute across labor margins.

3 Data

3.1 Puma to County Crosswalk

The starting point for the data challenge is the unfortunate and irresponsible decision among governmental agencies to create orthogonal identification and measurement schemes; while the Census reports observations based on state-year-puma levels in their 1% ACS datasets, other governmental agencies (EPA, NOAA, etc) report in terms of state-year-county levels. The match between the two is fraught with measurement error, so I try to introduce as many precautions as possible. The goal is to create a crosswalk between pumas (Census) and counties (other) so that different data sources can be merged for consistent estimation. From 2000-2011, the 2000 Census did not change its county identification and, as such, I leverage the 2000 Decennial Census information on counties in order to create the crosswalk for 2005-2010.[2] (In 2012, they changed to the new pumas.) The fundamental challenge is allocating different regions, based on an allocation factor, when some localities contain observations with multiple pumas within a single county, versus other localities with multiple counties within a single puma. Since pumas tend to be larger than counties (at least in urban areas), the problem arises when, for example, a university is in a puma and a specific county, but there is an allocation factor for the remaining counties. A simple application of the allocation factor, ignoring the problem that other observations have multiple pumas for a given county, is that the match would imply educational attainment holds linearly across the matched counties to single puma; in reality, we know that it is much more likely for there to be nonlinearly fewer people in other counties attending the university, relative to people within the county that is located at the university.

One way I overcome the problem that there are multiple pumas in a county for some localities, and multiple counties in a puma for others, is by dropping all observations with less than 65,000 people because the 1 year ACS sample only surveys areas with over 65,000 people. If I did not do this, I would be allocating areas that are not even represented in the ACS 1 year sample. Doing this leaves me with 98.11% of the sample that falls into the category of a direct county to puma match (41.5%) or multiple pumas within a single county; I discard the 64 total observations (only 1.89% of the sample) that still has counties within a puma. Not only am I discarding only a very small amount of data, but also these are the localities that contain the least amount of identifying power since they are rural (smaller) areas that are, therefore, more likely to have better air quality since there are fewer sources of pollution and space for pollution to diffuse.

After matching one to many the crosswalk dataset with the 2005-2010 Census dataset consisting of ACS 1 year estimates, I obtain an 87.53% successful merge consisting of 2,433,952 observations. The remaining

[2]http://mcdc2.missouri.edu/websas/geocorr2k.html

problem is that there is an allocation factor since not every county with multiple pumas should be weighted equally. To address this problem, I weight my regressions not only by person weights, but also by the allocation factor. In addition to the clever weighting, measurement error should balance out given the size of the sample and, at worst, simply increase the standard error of my estimates.

Allocation factors.–Consider an allocation factor of .5 given by the PUMA to county mapping; this implies that about half of the PUMA is associated with the respective county. For an ACS "hweight" or "pweight", say 100, then, if the PUMA is only .5 of the county, then I should reduce the weight of the PUMA by weighting the observation by only 50 instead of 100. (Practically speaking, given the conversion ratio that the geospatial reference provides, multiply this by pweight or hweight.)

Shortcomings of the approach.–These steps assume that the PUMA characteristics are homogeneously replicated in the county portions, which need not hold. For example, just because half of a PUMA's population lives in one county does not mean that's where half of their African American population lives; the same goes for any housing or person level variable. In other words, the mapping from PUMAs to counties may not be homogeneous in the characteristics. Fortunately, this is likely to be flawed only for small counties, namely 20,000 and less, since their allocations are based on estimates from an area that is 5x larger and are just "pro-rated" to the smaller county; however, I do not use these small counties.

Relevant files.–Using the Missouri Geocorr website (http://mcdc2.missouri.edu/websas/geocorr2k.html), I use the 2000 Census as the source and PUMA 5% samples for the target. For years 2009 and 2010, I use the same, but when determining the weighting variable I use the "population 2009 estimate". I thank John Blodgett from the Missouri Census data center for some assistance. I choose not to use their CPR experimental methodology because I am using 1 year ACS sample data, which does not match with their computed ratios that rely on 3 year ACS sample data.[3]

General.–I begin by merging the weather and EPA air quality data together. Starting with the air quality data, I merge the weather data and keep 33% of the observations; about 60% of the data is discarded because I do not have the weather data for all of the same counties that I have air quality for. I subsequently merge the combined Medicare and ACS data using one to many; 84.59% of the data successfully matches. This high number is driven by the fact that many of the mismatches between the previous EPA and NOAA data were for smaller localities. The ACS data highlights that I have rich data for the largest counties. I subsequently add in state-level employment data for controls of economic conditions.

3.2 Weather Cleaning

The data is based on the period 1967 - 2014 and contains 3.2 million data points (each data point represents a monthly average or total for a specific station in a specific year). Not all stations have been recording for the entire period and not all counties had stations, leading to some missing values, particularly in the wind data. There are three measurements: mean/median and counts for records in the source data, which I used to look for outliers. Means/median could be calculated based on the samples in a county, but that would bias towards stations with the most data. As an alternative, stations are weighted equally independent of the data since some stations might be at higher altitudes and therefore have a stronger influence on the county mean if they are sampled thoroughly. The mean of each station is calculated and then used to calculate the mean of all stations within a county.

4 Robustness on Imputation

Before documenting robustness on the imputed aggregates, I begin by characterizing summary statistics for the Census, CEX, and ATUS; these are documented below.

Both tables suggest that there is large overlap in the distribution of covariates between the using and imputing datasets. If this were not the case—for example, considering the extreme example where one dataset consists only of females and the other only of males—then there would not be any variation in relevant covariates that are correlated with leisure/consumption that is needed ot identify the demand parameters of interest. Although one counterargument for imputing is that electricity and the inverse of hours worked are reasonable proxies for nondurables and leisure, electricity tends to be very inelastic and

[3] http://mcdc.missouri.edu/data/acs/Readme.shtml#cpr

Table 1: Imputed v. Actual Summary Statistics

	ACS-imputed		CEX	
	Mean	S.D.	Mean	S.D.
Age	47.53	15.80	45.22	12.38
Number of people in family	3.05	1.28	3.43	1.42
Years of schooling	13.53	3.20	13.40	1.88
Number of children in family	0.54	0.96	1.07	1.22
White	0.79	0.41	0.83	0.37
Log Nondurables	9.15	0.28	9.48	0.53
Years of Household Tenure	16.48	12.97	1.83	1.22
No. Bedrooms	3.17	0.89	3.14	0.96
Male	0.48	0.50	0.50	0.50
Observations	1425740		31714	

Table 2: Imputed v. Actual Summary Statistics

	ACS-imputed		ATUS	
	Mean	S.D.	Mean	S.D.
Age	47.51	17.81	46.44	17.75
Number of people in family	3.05	1.28	2.81	1.55
Years of schooling	13.40	3.22	13.94	3.28
Number of children in family	0.55	0.97	0.91	1.15
White	0.79	0.41	0.82	0.39
Actual Leisure	8.72	0.15	8.71	0.27
Observations	1670955		42441	

hours worked does not contain variation in households time use preferences (e.g., it is only a proxy for the quantity of market supplied time).[4]

Below, I report three additional tests for the accuracy of the imputation. While the results presented in the main text provide the most direct assurance over the validity of the imputation, these reveal additional dimensions of the imputation.

First, consider plotting the predicted versus the actual values from the imputation. That is, generate the imputed values in the dataset containing consumption or leisure, and compare that with the actual values.

Overall, the fit is very good—there's nearly a 1-1 mapping between actual and imputed. Second, while these predicted and actual plots are informative, it is impossible to see the distribution of the imputed and actual. Below, I document the quantiles of these two distributions.

While a perfect match in quantiles would be represented by proximity to the line through the middle, the plot suggests that the imputation is weaker for individuals who have low levels of leisure. One of the reasons is that the Census has wider variance in nonwork hours, which is the main variable used in identifying the distribution of leisure in the imputation. While one alternative is to simply jettison this part of the Census, my interpretation of these results is that these individuals are just not contained in the ATUS and a reasonable mapping exists to those in the Census. Third, I report the regressions used to implement the imputations.

[4]For example, consider a new measure of leisure with the typical approach in macroeconomics of leisure given by "1 - (share of time spent working)". The correlation between my two measures of leisure (which have a .92 correlation with each other) and the aforementioned "macro" definition of leisure is approximately .4. The positive and weak correlation reveals the common theme in time use literature that changes in leisure are driven by changes in not only hours worked, but also other household activities. The challenge is the amount of discretion that classifying leisure and non-leisure time allocations requires. For example, although travel for education might be thought of as a non-work related activity, I classify it as work related because the underlying motivation for it is to supplement labor income and drive greater value for the firm; as a result, I construct four measures of leisure. Because of the discretion required, I construct four measures of leisure. In particular, one question asks specifically the number of market hours that they are paid to work, and the other asks for time spent in a full time and secondary job (if they have one); the latter seems to be less accurately reported.

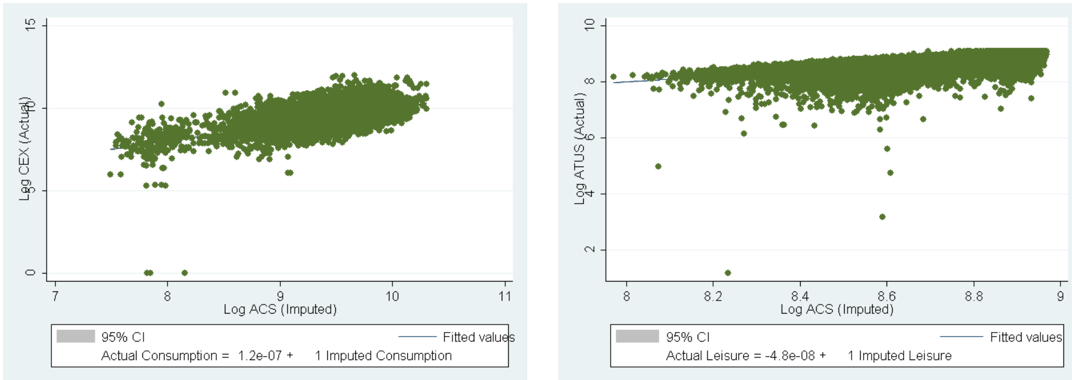


Figure 4: Linear OLS, Actual versus Imputed Log Nondurables Consumption and Log Leisure

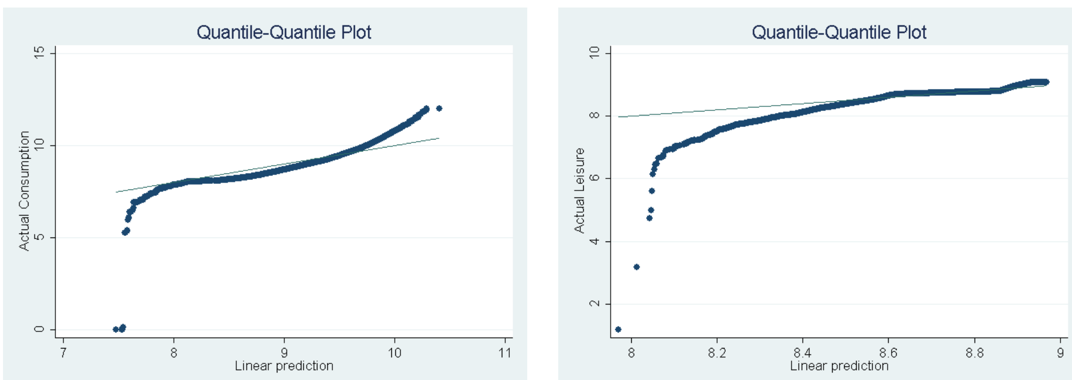


Figure 5: Quantile Plot, Actual versus Imputed Log Nondurables Consumption and Log Leisure

Table 3: Imputation of Nondurables Consumption, with Electricity, Water, Gas Expenditures

	Linear		Log		IV		Log		Spline	
	(1)		(2)		(3)		(4)		(5)	
Elect.	1.36***	[.19]								
Water	2.75***	[.28]								
Gas	1.00***	[.11]								
Log Elect.			.05***	[.00]	.00	[.]	9.52***	[1.62]		
Log Water			.02***	[.00]	.07**	[.03]	7.66	[4.87]	.10***	[.02]
Log Gas			.10***	[.01]	.14***	[.05]	34.05***	[8.68]	.10***	[.01]
Log Elect.-Sq.					-.03	[.02]	-4.69***	[.81]		
Log Water-Sq.							-3.81	[2.43]		
Log Gas-Sq.							-16.94***	[4.34]		
State FE	Yes		Yes		Yes		Yes		Yes	
Year FE	Yes		Yes		Yes		Yes		Yes	
Observations	31485		31485		31485		31485		31485	
Adjusted R^2	.176		.311		.284		.317		.309	

Standard errors in brackets

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4: Imputation of Leisure with Nonwork Time

	Linear		Log		IV		Spline			
	(1)		(2)		(3)		(4)		(5)	
Nonwork Time	.00***	[.00]								
Log Nonwork Time			1.06***	[.01]						
Log Nonwork, Spline 1					1.40**	[.57]	.96***	[.03]	.92***	[.03]
Log Nonwork, Spline 2					-6.31***	[1.43]	.05	[.04]	.06	[.04]
Log Wage, Spline 1							-.00***	[.00]		
Log Wage, Spline 2							.00***	[.00]		
Log Wage									-.00***	[.00]
State FE	Yes		Yes		Yes		Yes		Yes	
Year FE	Yes		Yes		Yes		Yes		Yes	
Observations	27819		27819		27819		42392		42392	
Adjusted R^2	.486		.487		.		.417		.418	

Standard errors in brackets

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

In both of these, column 5 shows the preferred regression results used to generate predicted values. While I test a variety of specifications (many more than reported here), there is little evidence that instrumental variables helps in prediction; in fact, it seems to worsen the fit potentially by introducing other sources of endogeneity. While my initial regressions do not yield parameter values with causal interpretations, the goal is prediction—not inference.

5 Robustness on Structural Elasticity

5.1 Nonlinear Treatment Effects

Prior literature (Heckman et al., 2010; Ekeland et al., 2004) has emphasized the shortcomings of the assumption about linearity. Semiparametric estimates—based on Klein and Spady (1993)—of conditional probabilities are especially important when treatment effects are heterogeneous, when the error might be mis-specified, or when the number of observations in one group is much larger than the other (Lehrer and Kordas, 2013). An alternative way to estimate the causal relationship between air quality and consumption/leisure is replacing $\log S$ in the intratemporal Euler with $g(\log S)$, where g is a polynomial.

The very interesting observation is that consumption becomes more complementary and leisure becomes less—consistent with the time varying demand for air quality—as income and leisure rose, air quality also rose. However, since the estimates are very similar across the specifications, there is little concern of bias relative to the benchmark.

5.2 Validity of the Exclusion Restriction

A. Instruments and Economic Shocks: The first way is to examine the correlation between the instruments and payroll expenditures, establishments, and hours worked.[5] Since the benchmark regressions are exploiting within year-state variation, the following table reports the relevant correlations conditional on controls (not counting the objects of correlation), including fixed effects on county and year.

Although these four variables are only proxies for economic shocks, they are comprehensive proxies since they span the gamut of household and firm measures of economic activity and all unified in their implication that there is effectively zero correlation with the instruments.

B. Instruments and Hourly Wages: The second way is to examine whether there is any statistically different effect of the instruments on hourly wages. To formalize the test, consider regressions of the form

[5]For electricity expenditures, the object of interest is electricity demeaned of other observables and fixed effects. Thus, I work with the residual.

Table 5: Structural Estimates of Elasticity, Semiparametric

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ψ	-1.094	-0.891	-1.571	-2.774	-1.965	-2.240	-1.925
se(ψ)	0.033	0.020	0.019	0.032	0.027	0.028	0.027
ϕ	0.758	0.890	-1.758	-1.573	0.419	0.063	0.074
se(ϕ)	0.029	0.010	0.023	0.025	0.024	0.024	0.023

Notes.–Sources: Census, CEX, ATUS, EPA, NOAA. These columns present the parameter estimates implied by the structural model from estimating regressions of the log of hourly wages on leisure, consumption, a cubic in air quality, and controls for taste shifters. Column 1 presents a simple OLS. Column 2 adds in county, industry, and year dummies without an instrument. Column 3 instruments for housing and leisure (not air quality) with fixed effects on state and year. Column 4 also instruments for air quality and adds industry fixed effects. Column 5 adds county fixed effects (in addition to industry and year fixed effects). Column 6 is the same as 5, but uses state by year fixed effects. Column 7 is the same as 6, but also includes industry fixed effects. Leisure is instrumented using a quadratic in mean/median maximum temperature and interactions with age fixed effects; housing consumption is instrumented using a quadratic in mean/median snowfall; and, air quality is instrumented using a quadratic in wind speeds. Controls on state and county economic conditions include: housing supply elasticity, civilian labor force, population, employment, number of establishments, payroll expenditures; controls on households include: number of children, disability status, age, gender, number of bedrooms, housing tenure, detached family house status, housing tenure, race dummies, and population. See the main text for the imputation of housing services, involving information on property values, taxes, and interest rates. Standard errors are heteroskedastic robust and bootstrapped with 50 replications. Because of the length associated with these fixed effects regressions, longer replications have been tested, but are omitted here.

Table 6: Correlations between Instruments and Economic Indicators

	Wind1	Wind2	Wind3	Wind4	Temp1	Temp2	Elect
Estab	0.001	0.002	0.001	0.001	0.001	0.001	0.001
Emp	0.002	0.002	0.001	0.001	0.003	0.002	0.000
Pay	-0.008	-0.008	-0.008	-0.008	-0.008	-0.008	-0.008
Hours	-0.004	-0.003	-0.005	-0.005	-0.001	-0.001	-0.002

Notes.–Sources: Census, CEX, ATUS, EPA, NOAA. The table reports the correlation of each of the instruments (2 minute wind speeds, 5 second wind speeds, resultant wind speeds, and average wind speeds for air quality; mean and median maximum temperature for leisure; log electricity expenditures for nondurables) and county-level establishments, employment, payroll expenditures, and hours worked. The instruments are demeaned of all variation in controls and county and year fixed effects; fixed effect controls include those used in the benchmark regressions, except for the objects of correlation.

$$\log w_{icts} = T_{cts}\tau_1 + \left[\sum_{k=1}^{K} T_{icts} \times 1(ind_k)\right]\tau_k + W_{cts}\delta + e_{icts}\eta_1 + \left[\sum_{j=1}^{J} e_{icts} \times 1(HouseAge_j)\right]\eta_j + X_{icts}\beta + \epsilon_{icts}$$

If the coefficients $\hat{\vartheta} = (\hat{\tau}_k, \hat{\delta}, \hat{\eta}_j)_{k\in\mathcal{K}, j\in\mathcal{J}}$ are not statistically significant and different from zero, then the results are consistent with the orthogonality between the instruments and the error in the benchmark regressions. Indeed, joint F-statistic tests implies that the null hypothesis that they are jointly equal to zero cannot be rejected (p-value = .44 for temperature × age, .5 for wind, and .11 for electricity × age of house).

C. Distribution of Hourly Wages and Instruments: The third way is to examine the distribution of hourly wage under each of the different instruments. To make these exercises feasible, I consider counties above and below the median of each of the instruments. For example, consider the distribution of hourly wages across counties above and below the mean resultant wind speed; these two distributions are highly overlapping. These graphs are displayed in the next subsection as part of the answer to concerns about having a local average treatment effect.

D. Unobservables and Locational Sorting: The economics of spatial sorting literature (Epple and Platt, 1998; Epple and Sieg, 1999) shows that, if individuals sort across residential locations based on a common measure of locational quality and their demand for it, then each residential location contains workers within a neighborhood of the demand for that location. If my benchmark regressions fail to control for unobserved tastes for market or non-market goods, then there should be a correlation between the unobserved heterogeneity and the location. To obtain a measure of the residual, demean hourly wages of all observable variation

$$\log w_{it} = X_{it}\beta_1 + \epsilon_{it}$$

where X contains a large vector of individual, state, and county covariates, as well as fixed effects. If there is evidence of endogeneity due to unobserved tastes for non-market or market goods, then the residual, $\hat{\epsilon}$, should be correlated with the quality of the individual's residential location. To proxy for locational quality, I use the individual's commute time to work and their self-reported property value, finding that the correlation between the two is -.01 (p-value = .00) and .002 (p-value = .3), respectively. Even though these two variables are imperfect proxies for locational quality, they capture the unobservables inherent local labor and housing markets.

Along the same lines, consider the following thought experiment. Households move into neighborhoods in different periods based on their optimal choice at the time. Two individuals that move into a home at different times will differ a lot in unobservables since different stimuli prompted them to move—for one, it may have been the price of a house on sale and, for the other, it may have been a non-market good. I create ten bins for household tenure and interact them with county fixed effects as in Fu and Ross (2013).

E. "Housing Supply" Instrument: The results are robust to alternative identifying assumptions based on entirely separate instruments. Instead of using weather and electricity as instruments for the three endogenous regressors, consider instrumenting consumption using the housing supply elasticity (discussed below), leisure using interactions between cohort fixed effects and years of schooling and labor market experience, and continuing to use wind speeds as an instrument for air quality.

Before characterizing the results, I first discuss an analog to the traditional housing supply elasticity from Saiz (2010). Suppose that earnings, w, are a function of household-level tastes, $X_1, X_2, ..., X_{J_X}$, local asset values, $A_1, A_2, ..., A_{J_A}$, housing rents (since housing equity is a primary way in which workers buffer against labor market risk (Mian and Sufi, 2011)), w, and non-market amenities, S.

$$w = w(X_1, X_2, ..., X_{J_X}, A_1, A_2, ..., A_{J_A}, h, S)$$

While decomposing housing values with respect to housing attributes is common in environmental economics, a natural refinement is the incorporation of household-level tastes and the opportunity cost of time since they characterize a household's tastes for a location with a certain set of market and non-market amenities. Assuming that individuals maximize their utility when choosing a home, then totally differentiating the home value, h, provides a quantitative link between its value and all of its dependencies.

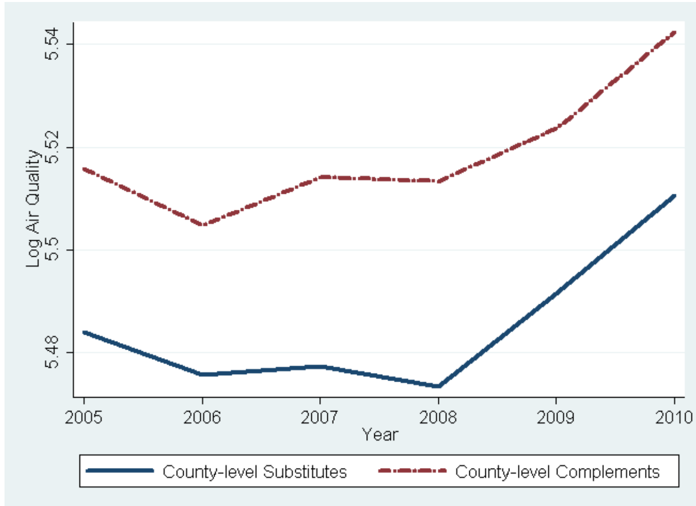


Figure 6: Housing Elasticities and Air Quality, 2005-2010

Notes.–Source: ACS, EPA, NOAA. The figure plots the log of air quality (as previously defined) based on whether within-county elasticities between housing and wages are substitutable or complement, e.g., $\varepsilon^h < 0$ or $\varepsilon^h > 0$, respectively. These elasticities are computed from a regression (Equation 4) of wages on housing prices and the usual controls within-county, but pooling all years that the county is observed, together with a cubic time trend. Counties with substitutability between wages and housing values should exhibit greater air quality since households are being compensated through amenities more, relative to their control counterparts that are compensated through asset values more.

$$dw = \sum_{j_X=1}^{J_X} \frac{dX_{j_X}}{X_{j_X}} + \sum_{j_A=1}^{J_A} \frac{dA_{J_A}}{A_{j_A}} + \frac{dh}{h} + \frac{dS}{S}$$

where all the inputs are exogenous to the housing price.[6] Letting ε^h denote the housing elasticity with respect to the wage rate, then it can be estimated by running regressions of the form

$$\Delta \log w_{ict} = \varepsilon^h_{ct} \Delta \log h_{ict} + \beta_{ct} X_{ict} + \varpi_{ct} A_{ict} + \tau_t + \epsilon_{ict}, \quad \forall c \in \mathcal{C} \tag{4}$$

where Δ denotes the first-difference operator, air quality is omitted because it varies only at the county-level, X contains the usual household-level controls and an industry-year average, ε^h is a county specific elasticity that characterizes the effect of a 1% change in the log housing values. Although the mean elasticity is only $\varepsilon^h = .01$, the range is large: from -.87 to .73. The beautiful feature of this regression is that the capitalization of housing rents into wages reflects the extent to which asset values pass-through to earnings as a market-based form of compensation. Counties with greater housing elasticities with respect to wages will tend to have lower preferences for environmental amenities and air quality since housing values will reflect a stronger correlation with wages. In other words, the identifying assumption is that counties with lower housing elasticities will have greater preferences for air quality, and thus a higher coefficient on air quality in a regression with wages. Denoting counties with $\varepsilon^h > 0$ as "complements" and $\varepsilon^h < 0$ as "substitutes", consider the relationship between air quality and housing elasticities below.

While not causal evidence—since the estimate is more representative of a conditional correlation than an aggregate elasticity—Figure 6 shows that counties that capitalize asset values into wages more heavily ($\varepsilon^h > 0$) have lower levels of air quality. Interestingly, mean nondurables consumption, leisure, and wages are all nearly identical between the two groups, suggesting that the underlying mechanism must be related to tastes for environmental amenities. As a result, this elasticity also behaves as a source of variation for air quality. The exclusion restriction requires that unobserved shocks to wages are uncorrelated with air quality within counties with similar wind speeds and housing elasticities. Since consumption, leisure, and wages are all very similar between counties with and without substitutable housing elasticities, there is little reason

[6] In cross-sectional data, this should be reasonable since the household-specific error is idiosyncratic and not systematically related with any of the inputs over time or space.

to expect the exclusion restriction to be violated.[7] Putting it all together and estimating the intratemporal again yields the following results.

Table 7: Structural Estimates of Elasticity (Alternate Instruments)

	(1)	(2)	(3)	(4)	(5)
ψ	-1.7677079	-.69805484	-.23261382	-.3397118	-.14013804
se(ψ)	.37043278	.28936522	.07457557	.07336621	.06689089
ϕ	-.1567077	.52123166	-1.1394435	-1.9709937	.20090225
se(ϕ)	.14937538	.11237347	.53983212	.57688839	.4840628
λ	1.6110002	1.2192865	-.90682973	-1.6312819	.34104028
se(λ)	.26965019	.2284684	.52428168	.56211107	.48378121

Notes.–Sources: Census, CEX, ATUS, EPA, NOAA. These columns present the same class of regressions as the housing elasticity table, but instead instrument for consumption using an interaction between electricity consumption and an indicator for whether the house uses gasoline for heating, and an interaction between dummies on years of schooling and years of labor market experience.

Although the similar result is only apparent in column 7, the rationale is very important. Since the instruments for leisure are largely coming from skill premia, which are inherently industry-specific, failing to control for time invariant industry unobservables seems to cause systematic bias in the coefficient on consumption since the unobserved variation shows up in these workers' consumption patterns even more than in their time use. However, once state-by-year, county, and industry fixed effects are included, the implied elasticities are qualitatively the same. The elasticity on leisure is lower (more substitutable) since the skill premium is not completely exogenous, but rather correlated with some individual-specific unobservables. Specifically, a positive correlation with earnings based on skill biased technical change and on leisure will induce upwards bias.

5.3 Interpretation of Treatment Effects and Exclusion Restriction

Although a direct test of the exclusion restriction is impossible, this section provides very suggestive evidence that the distribution of wages do not differ between households living in areas above and below the median instrumental variables (e.g., wind speeds). The tests also provide evidence that the parameter estimates have the interpretation of average treatment effects from an empirical perspective. If the distribution of the outcome variable (wages) do not differ substantially among compliers, never-takers, and always-takers, then the LATE and ATE should be similar.

Figure 7 plots the kernel density of hourly wages for observations above and below the median of the group for the corresponding instrument. In other words, consider observations that fall above the median electricity expenditures and plot the density against the density of those that fall below the median; the graph suggests that the distribution of hourly wages is almost identical. The results are very assuring that there are very minor differences in the wage distributions based on the different instruments used to induce exogenous variation in the "treatment" (air quality).

5.3.1 Aggregation and the Willingness to Pay for Air Quality

The fundament challenge with cross-sectional data is removing sources of time varying and invariant heterogeneity that contribute to shifts in the demand for consumption and leisure through channels other than air quality (e.g., locational sorting, unobserved tastes, etc). While my instrumental variables strategy and detailed controls provide assurance that my results are not driven by separate endogenous behavioral responses, Amiran and Hagen (2014) show that a more serious aggregation problem can arise when exploiting

[7] A separate concern is that the elasticity is formed using a regression of housing values on wages, so the generated regressor is correlated with the error. However, that holds only if the unobserved shocks were not controlled in the first stage regression and correlated with air quality in the second stage. For example, if there is a large upwards bias arising from a negative correlation between unobserved housing shocks and wages, then this could induce bias in the second stage if counties with larger shocks are also systematically more likely to have different environmental quality preferences. Generated regressors as instruments have a history in economics, dating back to Murphy and Topel (1985). Dufour and Jasiak (2001) establish finite sample confidence intervals for these estimators and emphasize the importance of the generated regressor having a strong correlation with the instrumented variable of interest.

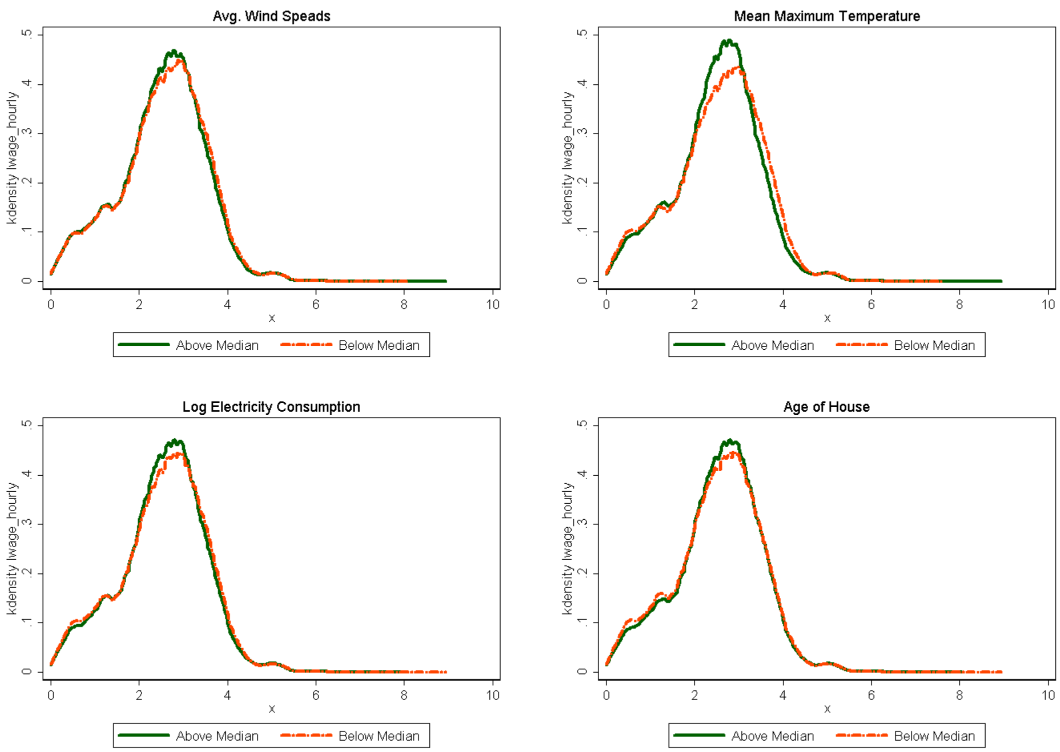


Figure 7: Comparison of Wage Distributions below/above Median (by Instrument)

Notes.–These Figures partition the instruments (wind speeds, maximum temperature, electricity consumption, and age of the house) into two groups: above or below the median. The wind instrument is average wind speed, but is also robust to other measurements, namely the resultant wind speed, or 5 second and 2 minute wind speeds. Similarly, the temperature distribution is robust to other measurements, namely the median maximum temperature.

cross-sectional variation in the presence of heterogeneous preferences. Given their suggestion of exploiting a longitudinal panel with respondents who have experienced significant changes over the sample period, I use detailed longitudinal data from the 1967-2010 Panel Study of Income Dynamics (PSID); see my companion paper (Makridis, 2014a) for a description of the sample selection and data cleaning process.

The dataset has three helpful features. First, it is longitudinal, so I can include person-level fixed effects. By removing all time invariant sources of heterogeneity across individuals, while also conditioning on comprehensive demographic and taste proxy variables, I can provide assurance that unobserved shocks are not driven my elasticity estimates. Second, individuals report whether they moved within or between counties, meaning that I can run the regression only on non-movers. By making the supply of housing—a bundle of consumption and leisure amenities—fixed by construction of the sample, I can guarantee that variation in consumption and leisure is driven only by variation in air quality, thereby eliminating the endogeneity problem of locational sorting. To the extent that there is overlap in the support of the distribution of covariates between these non-movers and movers, these estimates will provide informative elasticities for a general population. Third, because of the time series variation, I can estimate a long-run behavioral response to changes in air quality. Aside from providing a potentially more informative elasticity, a sufficiently long time series will resolve the concern that aggregate shocks bundled in the cross-section are correlated with unobserved heterogeneity in consumption and leisure (Altug and Miller, 1990, 1998).

[waiting for Michigan to send data, update with it]

5.4 Sensitivity Analysis

Imbens (2003) introduced a novel procedure for testing the sensitivity of treatment effects to unobserved heterogeneity by asking the following question: how strongly correlated must an unobserved variable, call it U, need to be with the treatment variable, call it AQI, and the outcome variable, call it Y, in order to make the coefficient on AQI no longer statistically significant? Applying his methodology for sensitivity analysis to my question, consider the following set of equations

$$AQI_{cst} = \beta X_{icst} + \pi U_{icst} + \epsilon_{icst}$$

$$Y_{icst} = \sigma AQI_{cst} + \tilde{\beta} X_{icst} + \tilde{\pi} U_{icst} + \tilde{\epsilon}_{icst}$$

where I am instrumenting for air quality with wind speeds as usual. To generate a proxy for the unobserved variation, modeled through U, I take the residuals from the two equations and a value drawn from a standard normal distribution

$$\mathcal{U}_{icst} = \omega_1 (AQI_{cst} - \beta X_{icst}) + \omega_2 (Y_{icst} - \sigma AQI_{cst} - \tilde{\beta} X_{icst}) + \mathcal{N}(0,1)$$

where ω_1 and ω_2 are weights for each residual. When $\omega_1 > \omega_2$ ($\omega_1 < \omega_2$), then the unobserved variable will be more correlated with air quality (consumption or leisure). Since the unobserved variable needs to be orthogonal to all other covariates in order to generate a valid comparison—otherwise covariances between the controls and unobserved variable would confound the test—estimate the following

$$\tilde{U}_{icst} = \mathcal{U}_{icst} - \hat{\lambda} X_{icst}, \quad \hat{\lambda} = \arg\min_{\lambda} (\mathcal{U}_{icst} - \lambda X_{icst})$$

which is equivalent to regressing U on all the Xs and subtracting the predicted values from U from the predicted values from that regression. In order to operationalize all of this, I draw upon the detailed package created by Harada (2012).[8]

References

Altug, S. and R. Miller (1990): "Household choices in equilibrium," *Econometrica*, 58, 543–570.

[8] See: https://files.nyu.edu/mh166/public/docs/research.html.

——— (1998): "The effects of work experience on female wages and labor supply," *Review of Economic Studies*, 65, 45–67.

Amiran, E. Y. and D. A. Hagen (2014): "Tests of the discrepency between willingness to accept and willingness to pay: A theoretical analysis," *Working paper.*

Autor, D. H., L. F. Katz, and A. B. Krueger (1998): "Computing inequality: Have computers changed the labor market?" *Quarterly Journal of Economics*, 113, 1169–1213.

Autor, D. H., F. Levy, and R. J. Murnane (2003): "The skill content of recent technological change: An empirical exploration," *Quarterly Journal of Economics*, 118, 1279–1333.

Bayer, P., N. Keohane, and C. Timmins (2009): "Migration and hedonic valuation: The case of air quality," *Journal of Environmental Economics and Management*, 58, 1–14.

Dufour, J.-M. and J. Jasiak (2001): "Finite sample limited information inference methods for structural equations and models with generated regressors," *International Economic Review*, 42, 815–843.

Ebert, U. (2003): "Environmental goods and the distribution of income," *Environmental and Resource Economics*, 25, 435–459.

Ekeland, I., J. J. Heckman, and L. Nesheim (2004): "Identification and estimation of hedonic models," *Journal of Political Economy*, 112, S60–S109.

Epple, D. and G. J. Platt (1998): "Equilibrium and local redistribution in an urban economy hwen households differ in both preferences and incomes," *Journal of Urban Economics*, 43, 23–51.

Epple, D. and H. Sieg (1999): "Estimating equilibrium models of local jurisdictions," *Journal of Political Economy*, 107, 645–681.

Flores, N. E. and R. T. Carson (1997): "The relationship between the income elasticities of demand and willingness to pay," *Journal of Environmental Economics and Management*, 33, 287–295.

Fu, S. and S. L. Ross (2013): "Wage premia in employment clusters: How important is worker heterogeneity?" *Journal of Labor Economics*, 31, 271–304.

Hanemann, W. M. (1991): "Willingness to pay and willingness to accept: How much can they differ?" *American Economic Review*, 81, 635–647.

Harada, M. (2012): "Generalized sensitivity analysis," *Working paper.*

Heckman, J. J., R. L. Matzkin, and L. Nesheim (2010): "Nonparametric identification and estimation of nonadditive hedonic models," *Econometrica*, 78, 1569–1591.

Imbens, G. (2003): "Sensitivity to exogeneity assumptions in program evaluation," *American Economic Review*, 93, 126–132.

Katz, L. F. and K. M. Murphy (1992): "Changes in relative wages, 1963-1987: Supply and demand factors," *Quarterly Journal of Economics*, 107, 35–78.

Klein, R. and R. Spady (1993): "An efficient semiparametric estimator for discrete choice models," *Econometrica*, 61, 387–422.

Kristom, B. and P. Riera (1996): "Is the income elasticity of environmental improvements less than one?" *Environmental and Resource Economics*, 7, 45–55.

Lehrer, S. F. and G. Kordas (2013): "Matching using semiparametric propensity scores," *Empirical Economics*, 44, 13–45.

Makridis, C. (2014a): "The Performance Pay Premium, Human Capital and Inequality: Evidence from Over Forty Years of Microdata," *Working paper.*

——— (2014b): "Towards simple environmental policy rules: Fiscal interactions, time inconsistency, and externalities in general equilibrium," *Working paper*.

Mian, A. and A. Sufi (2011): "House prices, home equity-based borrowing, and the US household leverage crisis," *American Economic Review*, 101, 2132–56.

Murphy, K. and R. H. Topel (1985): "Estimation and inference in two-step econometric models," *Journal of Business and Economic Statistics*, 3, 370–379.

Roback, J. (1982): "Wages, rents, and the quality of life," *Journal of Political Economy*, 90.

Saiz, A. (2010): "Geographic determinants of housing supply," *Quarterly Journal of Economics*, 125, 1253–1296.