Health Risk of Heating Fuel Choice: A Simultaneous Causality Analysis

Zheng Liu
Postdoctoral Research Associate
Department of Agricultural Economics
University of Kentucky
435 Charles E. Barnhart Bldg.
Lexington, KY 40546 U.S.

Angelos Pagoulatos
Professor
Department of Agricultural Economics
University of Kentucky
404 Charles E. Barnhart Bldg.
Lexington, KY 40546 U.S.

Wuyang Hu
Associate Professor
Department of Agricultural Economics
University of Kentucky
313 Charles E. Barnhart Bldg.
Lexington, KY 40546 U.S.

Abstract

Combustion-generated pollutants, principally those from solid-fuels including biomass and coal when cooking and heating, bring out a significant public health hazard in both developed and developing countries. Most of the existing studies addressing this issue focus on developing countries, and on exposure when cooking rather than heating. By using Kentucky rural data, this research explores the health risk associated with heating fuel choice. Given the simultaneity between heating fuel choice and prevalence of asthma and allergy, we obtain the instrumental variable (IV) estimate for Logit models through the Generalized Method of Moments (GMM). After correcting for simultaneity bias, we do not find strong evidence supporting the causal relationship between polluting heating use and the prevalence of asthma, allergy, and other respiratory disease. Some demographic and lifestyle factors do have significant effects on the prevalence of these diseases. The results suggest that the current policy regulation and public awareness of the possible hazards associated with different heating fuels may have prompted a strong averting behavior.

Keywords: combustion-generated pollutants, indoor air pollution, heating fuel choice, health risk, GMM-IV Estimation

JEL: Q53 I18

Introduction

Indoor air pollution (IAP) is a public health problem in both developed and developing countries (Ezzati et al. 2003). Since the 1980's, based on comparative risk studies, EPA and its Science Advisory Board (SAB) have consistently ranked indoor air pollution among the top five environmental risks to public health (EPA 1993). Among the four components of indoor pollution (combustion products, chemicals, radon, and biologic agents), combustion-generated pollutants, principally those from solid-fuel such as biomass (wood, dung and crop residues) and coal used in cooking and heating, have been the focus of epidemiologic and physiologic research, especially in developing countries. Biomass or coal smoke contains a large number of pollutants, such as particulate matter (PM), carbon monoxide (CO), nitrogen dioxide, sulfur oxides (mainly from coal), formaldehyde, and polycyclic organic matter, have known health hazards to human body (Ezzati and Kammen 2002). The epidemiological research focus is given to acute (lower) respiratory infections (ALRI), chronic obstructive pulmonary disease (COPD), and lung cancer (due to coal) for which the evidence is the most robust (Bruce et al. 2000; Sharpe et al. 1989).

In a population-based case-control study of lung cancer among white women in Los Angeles County, California, Wu et al. (1985) reported elevated risks for lung cancer when individuals reported that they had exposure to coal burning smoke when heating or cooking during their younger age. In a case-control study of lung cancer carried out in Montreal in 1996–2001, Ramanakumar et al. (2007) calculated the odds ratios by using a few indices of exposure to traditional heating and cooking sources (coal and wood combustion). The results indicated that there were no excess risks for lung cancer among men, while there was higher risk for women. Another case-control study in Eastern and Central Europe and the United Kingdom (Lissowska et al. 2005) showed a moderately higher risk of lung cancer related to solid fuel use when cooking and heating. In developed countries, people that experienced coal and wood for heating and cooking at their younger age are now in the risk age range for respiratory diseases and lung cancer. Although dirty fuels are no longer dominant, in the past decade, there has been some increased consumption of wood for heating and cooking in developed countries, to create mood-setting atmosphere or grill food. In the U.S., from the Residential Energy Consumption Survey (RECS) data (2005), about 2.9 million households (2.6% of the total housing units) use wood as the main heating fuel and about 79% of them live in rural areas.

About 8.9 million households (8% of the total housing units) use wood as their secondary heating fuel. Although the U.S. Environmental Protection Agency (EPA) has promulgated New Source Performance Standards (NSPS), which set up particulate emission standard for wood heaters to be certified, by 1998, only about 11% of the wood stoves in use were EPA certified, and only 4% of the fireplace inserts were EPA certified (Houck et al. 1998).The EPA and the U.S. Consumer Product Safety Commission (2007) also claimed that use of unvented combustion appliances (such as kerosene and oil fueled space heaters) in closed settings may also be associated with health risks because of exposure to polluting emissions. In addition, according to the review report by the U.S. Energy Information Administration, residential coal consumption increased by 9 percent in 2007 and 2008 (Freme 2009). It is of interest to explore the health risk associated with the use of wood, coal burning and other unvented combustion appliances that are fueled by kerosene or fuel oil for heating. In particular, we are interested on whether over time awareness due to information availability in the U.S. regarding the adverse health effects of different heating fuels have prompted a strong averting activity.

Averting activity represents the activities undertaken to avoid or reduce exposure to pollution. Previous economic literature addresses taking account of the averting activity when measuring the marginal value of reduced morbidity (Freeman III 2003). In this study, averting activity includes change of heating choice, such as shifting away from using polluting heating fuel and/or correctly using combustion appliances which are certified by the EPA and following the safety guidelines strictly. The potential effect of averting activity may not allow the observation of the health risk associated with heating fuel choice when exploring cross-sectional data. Using the Kentucky Homeplace Program survey data, this paper explores the relationship between heating fuel choice and the occurrence of disease from a panel set of observations, and tests whether the averting activity exists. If it does, we need to correct the simultaneity bias produced by the endogeneity of heating fuel choice with disease occurrence and obtain the unbiased estimate for the effect of heating choice on the occurrence of disease. The analysis will use the instrumental variable (IV) estimates through the Generalized Method of Moments (GMM).

Model

A choice model was used to identify the exposure–response relationship between the use of polluting heating fuel and the prevalence of some disease. This exposure–response relationship exists conditional on the effects of individual's behavior (averting activity). The individual's averting activity can be captured by the shift from using polluting heating to non-polluting heating fuel or appliance. The basic equation we will consider is:

$$\Pr ob(disease = 1) = \begin{cases} f(ph = 1, X \, / \, AVERT = 0) \\ f(ph = 0, X \, / \, AVERT = 1) \end{cases}$$

Where *disease* is a dummy variable, which is equal to 1 if individual has a disease. The dummy variable *ph* signifies whether or not the individual uses a polluting heating fuel. *X* is a set of other exogenous variables that could influence the occurrence of disease. And *AVERT* is a dummy variable which indicates the existence of some form of averting activity—the use of non-polluting heating---when equal to 1. When *AVERT* does equal to 1*,* the problem of simultaneity associated with having a disease and the heating fuel choice must be carefully considered and corrected for when estimating the exposure-response relationship between the use of polluting heating fuel and the prevalence of disease.

In this paper, we will obtain the instrumental variable (IV) estimate for simultaneous logistic regression by using the Generalized Method of Moments (GMM). The basic principle of GMM is identifying a set of moment functions of known variables and unknown parameters. The moment functions can be simultaneously set to zero based on the orthogonality conditions between the explanatory variable and the error term. A consistent estimate of the parameters can be obtained by solving the moment functions (Hayashi 2000; Greene 2003). Although there is a potential symmetric problem by using GMM-IV estimation due to the averaged definition for error terms in the orthogonality condition functions without specifying the relationship between explanatory variable, dependent variable and the specific error term, some research demonstrates that the GMM estimate can be an approximately consistent estimator even in the symmetric model case (Mullahy 1997; Terza 2006; Johnston et al. 2008).

Data

The data used in this study is the health survey data of the Kentucky Homeplace Program (KHP). The program was to help patients who live in rural areas to use the services they are qualified for, and provide preventive services for some chronic diseases.The surveys are the initial interview with respondents when they enroll into the KHP. The program is open for every resident living in the community. The data were collected in 2005 and 2006 as independent cross-sections. The total number of observations is 9,539. Besides providing the demographic, social-economic and risk factors information, the respondent may choose the type of heating fuel used among electric, gas, coal, wood, fuel oil, kerosene and others. In descending order, the percentage of heating fuel used by the sample housing units are electricity (66.8%), gas (29.9%), wood (7%), kerosene (3.8%),coal (3.4%), fuel oil (0.6 %), and other fuel (0.2%).

Table 1 reports the definitions and statistics of the variables used in the study. The average age of respondents is 53 years old, which is higher than reported median age of 35.9 for Kentucky residents (Kentucky Demographics 2005). 37% of the respondents are male and 95% of the sample are white (not Hispanic or Latino). The average length of education is 11 years of schooling. The average annual income is about \$12,717 which is much lower than the state average of \$40,299 and national average of \$50740 (U.S Census Bureau, 2007). These statistics are relevant to the geographic service area and the service objective of the Kentucky Homeplace Program, which focuses on the rural counties of Kentucky and most of the clients (respondents of the survey) are retirees and lower income receivers. About 45 % of the respondents participate in physical activities, and 53% have used tobacco products.

Two categories were created and used to indicate the heating fuel choice: non-polluting heating (*nph*) and polluting heating (*ph*). The former includes heating fuel choices of electric and gas. The latter includes the choices of coal, wood, fuel oil, and kerosene. For records indicating "others", we consider them each individually. Based on the explanation given by the respondents, we classify these entries into either "*ph*" or "*nph*." In our study sample, about 13.2% of the respondents use polluting heating fuel as their main or secondary heating option.

In this study, condition information related to heating fuel choice is limited to asthma, allergy and other respiratory disease. About, 7.1%, 6.6% and 7.2% of the sample suffers from allergy, asthma, and other respiratory disease respectively. Asthma was separated from the respiratory disease because of the high prevalence of asthma in Kentucky (CDC 2002, 2007). The sample contains 80 counties in Kentucky. These counties are more severely impoverished than the rest of the state (Appalachian Regional Commission, 2010). We created a dummy variable "*eastky*" to indicate whether the respondent lives in eastern Kentucky. A total of 63% of the respondents in the sample lives in this area. Overall, 17.3% of the eastern KY respondents use polluting heating compared to 6.2% of the respondents in other areas. As a result, we need to consider the region factor when exploring the impact of heating fuel choice.

Table 2 reports the sample distribution, cross frequency, and Z test results for rate of disease prevalence in polluting and non-polluting user groups. The prevalence of respiratory disease is higher within polluting heating users than that of within the non-polluting heating users (8.19% versus 7.08%). However, the Z test shows that this difference is not significant. The prevalence of asthma and allergy within polluting heating users is lower than that of within non-polluting heating users and these differences are statistically significant according to the Z test results. Another cross frequency table for polluting heating using rate in different health condition patient groups (Table 3) shows that the polluting heating using rate within people having asthma and allergy is lower than that of within those without these conditions. Moreover, Z test results indicated that these differences are significant.

Z test shows that there is no significant difference in the polluting heating using rates between people suffering with respiratory disease and those who do not. Do above results tell us individuals using polluting heating are less likely to suffer from asthma and allergy or those who have these diseases/symptoms are less likely to use polluting heating? The results in Tables 2 and 3 show the importance of considering the causality between these observations, or testing for the existence of averting activity. The two-way estimation of the relationship between disease prevalence rate and the heating fuel choice should be included in the regression and simultaneity issues should be considered.

Estimation

In this study, first, the Logit model was used to explore the impact of the heating fuel choice together with some demographic and lifestyle characteristics on the occurrence rate of asthma, allergy and other respiratory disease. The three separate equations to be estimated when determining the prevalence of each disease can be generalized as:

$Y_i = f$ (*age, white, male, eduy, income, exercise, smoker, eastky, ph*/ *AVERT*)   (1)

Where Y_i is a dummy variable ($i=1, 2, 3$), which equals to 1 if the individual suffers from one of the three diseases: other respiratory disease, asthma, or allergy respectively. The explanatory variables include polluting heating using (*ph*), and social-economic, demographic and lifestyle variables (the definitions are referring to Table 1). *AVERT* represents whether the averting activity exists. Based on the equation, we can only capture the true exposure- response relationship when *AVERT* = 0.

To test whether *AVERT* = 0, in addition to (1), we estimate another set of models generalized as equation (2).
$ph = f$ (*age, white, male, eduy, income, exercise, smoker, eastky, Y*$_i$)   (2)
Where the variables *ph*, Y_i ($i=1, 2, 3$) and other explanatory variables are defined as above.

Results

Causality between Heating Fuel Choice and Disease

We did a two-way logistic regression to explore whether *AVERT* = 0. Table 4 reports the regression results for the prevalence of the three diseases (where the dependent variables are whether the individual suffers from one of the three diseases), and Table 5 indicates the estimation results for the polluting heating choice (where the dependent variable is whether the individual uses polluting heating fuel). From Table 4, the use of polluting heating does not have significant effect on the prevalence of other respiratory disease. Using polluting heating has a significant negative effect on the prevalence of asthma and allergy. Table 5 shows suffering from other respiratory disease has no significant effect on the choice of using polluting heating fuel while suffering from asthma or allergy has a significant negative effect on people's choice of polluting heating fuel.

These results may be explained by the averting activities of individuals over time. Asthma is a chronic lung disease, and allergy is the 5th leading chronic disease in the U.S (Asthma & Allergy Foundation of America 2008). People having either of these two chronic diseases may shift to non-polluting heating in order to relieve the symptoms of the disease. While most acute respiratory diseases – a very common branch of respiratory disease – are sudden viral infections, there is no strong motivation for people to take some averting behavior (like shifting to non-polluting heating) after the infection passes. The above results indicate that the causal relationship between using polluting heating and the prevalence of asthma and allergy may work in both directions and simultaneous bias is produced if just using standard logistic regression.

There is a concern about the possibility of the simultaneity problem existed between the lifestyle variables and the occurrence of disease. In this case, the two lifestyle variables are "*exercise*" and "*smoking*." Based on literature review (A.D.A.M. Inc. 2009; Allergy Capital of Australia 2006), we do not have sufficient evidence to support that people suffering from these three particular diseases can reduce or relieve the related health risk by changing lifestyle: exercise more or quit smoking. As a result, in this study, the possible simultaneity problems associated with the two lifestyle variables are not explicitly addressed.

GMM-IV Estimation and Results

Due to the existence of averting activity, we use GMM-IV estimation to correct simultaneous bias and explore the true exposure–response relationship between the polluting heating using and the prevalence of asthma and allergy. For the case of asthma, we choose "*age, eduy, eastky, ownhm, citywater*" as IVs for the endogenous variable "*ph*."

We created two dummy variables: "*ownhm*" indicate whether the respondents live in their own home in contrast to those who rent, live in government housing, group home, stay with friend/family, etc; "*citywater*" represent whether the respondent use city water or other water sources, such as well water, coal bank water, or hauled water. We assume people living in their own home may have more freedom to change their choice of heating fuel, thus variable "*ownhm*" may be correlated to polluting heating using (*ph*). People using non-city water may live in remote rural areas and may be limited in their choice of heating sources. We did two-way logistic regressions to test the validity of IVs. Based on the results, all of the five IVs are highly correlated with the polluting heating variable "*ph*" (at the 1% significance level), but do not have significant effect on the prevalence of asthma (*y*). The model of interest is:

$$\Pr(y_t = 1) = \frac{e^{X_t \beta}}{(1 + e^{X_t \beta})} \text{ or } (1 + e^{-X_t \beta})^{-1} \tag{3}$$

y_t equals to 1 if individual *t* suffers from asthma and 0, otherwise. *X* is an N × 5 matrix of regressors which include five explanatory variables (K = 5): "*white, male, income, smoker* and *ph*." Because the model is overidentified, the GMM estimate of the vector $\hat{\beta}$ can be obtained by minimizing the criterion function as $e^T W (W^T \Omega W)^{-1} W^T e$ (4), where *W* is a matrix of instruments, *e* is a vector of sample residuals, Ω is the variance–covariance matrix of the error term. $(W^T \Omega W)$ need to be estimated before minimizing the criterion function. White's (1980) heteroskedasticity consistent covariance matrix estimator can accommodate flexible forms of heterogeneity in $(W^T \Omega W)$.

Model specification for the allergy case is similar as the asthma case. "*age*, *ownhm* ,*citywater*" were chosen to be the IVs for *ph*. Based on the two-way logistic regression results, all of these three variables are highly correlated with the polluting heating variable "*ph*" (at the 1% significance level) and do not influence the prevalence of allergy directly. The model of interest has the same expression as equation (3) except where y_t equal to 1 if individual *t* suffers from allergy and 0, otherwise. *X* represents an N × 7 matrix of regressors. The explanatory variables include "*white, male, income, eduy, eastky, smoker* and *ph*." Under the overidentified context, the GMM estimate of the vector $\hat{\beta}$ can be obtained by minimizing the criterion function (4).

Table 6 presents the estimation results of the standard logistic regression versus GMM- IV Estimation (IVE) for the prevalence of asthma. In the GMM-IVE model, excluding the IVs ("*age, eduy, eastky, ownhm, citywater*"), we kept all regressors in the original model except variable "*exercise*" to make the model converge better. Comparing the results of standard logistic and the GMM-IVE, we can find the coefficients estimates from both methods are almost identical while the standard error and the significance of the estimates (*P* value) have some differences. The standard errors of GMM-IVE are higher than the ones in the standard logistic regression. This is because less information (only a portion of the information in the endogenous variable) is used to produce the slope estimate, and the variance of the IV estimator is larger.

In term of GMM-IVE results, male are less likely to suffer from asthma at the 1% significance level which is same as the results from the standard logistic regression. Asthma is still more prevalent within people with higher income, which may be because the prevalence record here is based on the survey question that "whether you are told by the doctor that you suffer from some certain disease." People with higher income are more likely to go to see doctor and subject to diagnosis for asthma. Smoking is not a significant determinant of whether a person would suffer from asthma, which is different from the results in the standard logistic regression. If the effect of the averting behavior (that is people suffering from asthma may be more likely to choose non-polluting heating to control the symptoms) is excluded, using polluting heating does not have significant effect on the prevalence of asthma, which is different from the result from the standard logistic regression.

Table 7 presents the estimation results of the standard logistic regression and GMM-IVE for the prevalence of allergy. In the GMM-IVE model, we kept all regressors included in the original model except the IVs ("*age, ownhm, citywater*") and variable "*exercise*". The coefficients estimates for both models are very close. The standard errors are higher in GMM-IVE model than in standard Logit regression, especially for "*ph*".

As of GMM-IVE results, same as in the standard regression, white American and female are more likely to be diagnosed with allergy at the 5% and 1% significance level respectively. Income has less significantly positive effect on the prevalence of allergy. Education has same significant positive effect on the prevalence of allergy. Whether people live in eastern Kentucky does have significant effect on the prevalence of allergy. Same as in the asthma case, whether individuals used polluting heating or smoked are not significant determinants of the prevalence of allergy.

Conclusions

By using the standard Logistic regression, the relationship between polluting heating fuel using and prevalence of some diseases (asthma, allergy and other respiratory disease) in rural Kentucky was estimated for the period 2005-2006. The use of polluting heating fuel (including coal, wood, fuel oil, and kerosene) does not have a significant positive effect on the prevalence of respiratory disease (excluding asthma) while have a significant negative effect on the prevalence of asthma and allergy. These results may be explained by the averting behavior of individuals who shifted over time to non-polluting heating fuels such as electric and gas furnaces after they were diagnosed with asthma or allergy. To test whether the averting activity existed, we conducted a two-way logistic analysis.

The results show that people having asthma or allergy are less likely to use polluting heating fuel (at the 5% and 1% significance level respectively). The results suggest that people with asthma or allergy may have changed heating source over time. We used Instrumental Variable Estimation (IVE) to address this simultaneity problem and obtained the unbiased estimates through the Generalized Method of Moments (GMM). After correcting for simultaneity bias resulting from the averting behavior, using polluting heating fuel is not a significant determinant of the prevalence of asthma or allergy. There is no strong evidence to support the positive relationship between polluting heating fuel using and the prevalence of asthma, allergy and other respiratory disease.

There are some possible explanations for these results. First, the lack of detailed data on historical exposure to the pollution and the use of the type of heating fuel as a proxy for the actual exposure could be producing some measurement errors. This indirect approach may not adequately capture the influence of exposure variables (such as type and the location of heater, the actual time exposed to the pollution smoke) and differentiate the effect of different pollution source (fugitive indoor emissions from the stoves themselves, the exposures coming back in from the house chimney, or "neighborhood" levels due to the proximity of other household using the same fuels in the area). To obtain more accountable estimate results, developing methods for exposure assessment should be an important priority for future research in this field. A more plausible explanation related to public policy is that the performance standard promulgated by EPA and the awareness by the consumer of the possible hazards associated with different heating fuels has prompted a strong averting behavior. Information availability on energy source performance standards has allowed better informed decisions by many consumers.

Results from this study show that some demographic and personal lifestyle characteristics do have significant effects on the prevalence of the three diseases. Female are more likely to be diagnosed with asthma and allergy. People who participate in physical activities are less likely to suffer from respiratory disease (excluding asthma) while smokers are more likely to suffer from it, holding other factors constant.

References

A.D.A.M Inc. *Exercise's Effects on the Lungs*. Intercepted online in May 2009 at: http://health.nytimes.com/health/guides/specialtopic/physical-activity/exercise's-effects-on-the-lungs.html

Allergy Capital of Australia. 2006. *Exercise, allergy & anaphylaxis*. Intercepted online at: http://www.allergycapital.com.au/Pages/gpexerc.html#anchor690913

Asthma and Allergy Foundation of America (AAFA). 2008. *Asthma Facts and Figures*. Intercepted online on Dec 10, 2008 at: http://www.aafa.org/display.cfm?id=8&sub=42

Bruce, Nigel, R .Perez-Padilla, and R. Albalak. 2000. Indoor air pollution in developing countries: a major environmental and public health challenge. Bull. WHO 78:1078– 92

Centers for Disease Control and Prevention (CDC). Behavioral Risk Factor Surveillance System (BRFSS) data. Intercepted online in Dec 2008 at:

http://www.cdc.gov/asthma/brfss/02/current/current.pdf (BRFSS 2002 data)

http://www.cdc.gov/asthma/brfss/07/current/tableC1.pdf (BRFSS 2007 data)

Energy Information Administration (EIA). 2005. *Space Heating Characteristics by Urban/Rural Location, Residential Energy Consumption Survey*. Intercepted online in May 2007 at:

http://www.eia.doe.gov/emeu/recs/recs2005/hc2005_tables/hc4spaceheating/pdf/tablehc8.4.pdf

Ezzati, M. and D. M. Kammen. 2002. The health impacts of exposure to indoor air pollution from solid fuels in developing countries: knowledge, gaps, and data needs. *Environmental Health Perspectives* 110(11): 1057-1068.

Ezzati, M, A.D. Rodgers, A.D. Lopez, et al. 2003.*Comparative quantification of health risks: global and regional burden of disease due to selected major risk factors*. Vol 2. Geneva, Switzerland: World Health Organization.

Freme, F. 2009. *U.S. Coal Supply and Demand: 2008 Review*. U.S. Energy Information Administration (EIA). Intercept online in May 2009 at: http://www.eia.doe.gov/cneaf/coal/page/special/feature08.pdf

Freeman, A. M. III. 2003. *The Measurement of Environmental and Resource Values: Theory and Methods*. Resources for the Future, Washington, DC.

Greene, W.H. 2003. *Econometric Analysis*. 5th ed. Upper Saddle River, NJ: Prentice Hall.

Hayashi, F. 2000. *Econometrics*. 1st ed. Princeton, NJ: Princeton University Press.

Houck, J. E., P. E. Tiegs, R. C. McCrillis, C. Keithley and J. Crouch. 1998. Air Emissions from Residential Heating: The Wood Heating Option Put into Environmental Perspective.*The Proceedings of a U.S. EPA and Air Waste Management Association Conference: Emission Inventory: Living in a Global Environment*. v.1, pp 373-384.

Johnston, K. M., P. Gustafson, A. R. Levy and P. Grootendorst. 2008. Use of instrumental variables in the analysis of generalized linear models in the presence of unmeasured confounding with applications to epidemiological research. *Statistics in Medicine*. 27(9):1539 - 1556

Kentucky Demographics. 2005. Intercepted online in Jan 2009 at:

http://www.eredux.com/states/demographics.php?id=1135

Lissowska, J., A. Bardin-Mikolajczak, T. Fletcher, et al. 2005. Lung Cancer and Indoor Pollution from Heating and Cooking with Solid Fuels: the IARC International Multicentre Case-Control Study in Eastern/Central Europe and the United Kingdom. *American Journal of Epidemiology* 162(4):326-333.

Mullahy J. 1997. Instrumental-variable estimation of count data models: applications to models of cigarette smoking behavior. *The Review of Economics and Statistics*.79:586–593.

Ramanakumar,A. V., M. E. Parent, and J. Siemiatyck. 2007. Risk of Lung Cancer from Residential Heating and Cooking Fuels in Montreal, Canada. *American Journal of Epidemiology* 165(6):634-642.

Terza JV.2006. Estimation of policy effects using parametric nonlinear models: a contextual critique of the generalized method of moments. *Health Services and Outcomes Research Methodology*; 6:177–198.

U.S Census Bureau. *State & County Quickfacts*. 2007. Intercepted online in Jan 2009 at:

http://quickfacts.census.gov/qfd/states/21000.html

U.S. Environmental Protection Agency (EPA).1993. *Fact Sheet: Respiratory Health Effects of Passive Smoking*. Intercept online in May2008 at: http://www.epa.gov/smokefree/pubs/etsfs.html

U.S. Environmental Protection Agency (EPA) & U.S. Consumer Product Safety Commission. *A Guide to Indoor Air Quality—Link: What You Should Know About Combustion Appliances and Indoor Air Pollution*. Intercept online in 2007 at: http://www.cpsc.gov/cpscpub/pubs/452.html

White, Halbert. 1980. A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica*.48 (4): 817-838

Wu, A., B. Henderson, M. Pike, et al.1985. Smoking and other risk factors for lung cancer in women. *Journal of the National Cancer Institute*.74:747–51.

Table 1. Descriptive Statistics of Variables Used in the Analysis

Variable	Mean	Median	Std Dev	Definition
age	52.828	53	14.958	continuous; age of the respondent
male	0.373	0	0.484	dummy; = 1 if male
white	0.954	1	0.209	dummy; = 1 if race is" White"
eduy	10.699	12	2.439	continuous; years of education
income	12716.550	11652	8084.590	continuous; household total yearly pre-tax income
eastky	0.630	1	0.483	dummy; = 1 if live in the eastern of Kentucky
smoker	0.526	1	0.499	dummy; = 1 if has ever used tobacco products
exercise	0.445	0	0.497	dummy; = 1 if participate in any physical activity
ele	0.669	1	0.471	dummy; = 1 if use electric as heating type
gas	0.298	0	0.457	dummy; = 1 if use gas as heating type
coal	0.034	0	0.182	dummy; = 1 if use coal as heating type
wood	0.071	0	0.256	dummy; = 1 if use wood as heating type
foil	0.006	0	0.080	dummy; = 1 if use fuel oil as heating type
kero	0.038	0	0.191	dummy; = 1 if use kerosene as heating type
oth	0.002	0	0.047	dummy; = 1 if use "other" heating type
ph	0.132	0	0.339	dummy; = 1 if use polluting heating fuel
resp	0.072	0	0.259	dummy; = 1 if suffer from respiratory disease(except asthma)
asm	0.066	0	0.249	dummy; = 1 if suffer from asthma
alg	0.071	0	0.257	dummy; = 1 if suffer from allergy
N=9539				

Table 2. Cross Frequency and Z Test Results for Rate of Disease Prevalence in Polluting/ Non Polluting Heating User Groups

Characteristic	Sample Distribution (%)	Prevalence Rate of Respiratory Disease (excluding asthma) (%)	Prevalence Rate of Asthma (%)	Prevalence Rate of Allergy (%)
Polluting Heating Users (Ph=1)	13.15	8.19	4.94	4.86
Non-Polluting Heating Users (Ph=0)	86.85	7.08	6.89	7.34
Z test Statistic		-1.426	2.620***	3.252***
P-value		0.154	0.009	0.001

*, **, and *** represent significant at the 10%, 5%, and 1% significance levels respectively.

Table 3. Cross Frequency and Z Test Results for Rate of Polluting Heating Using by Individuals with Different Health Conditions

Characteristic	Sample Distribution (%)	Polluting Heating using rate
resp=1	7.23	14.89
resp=0	92.77	13.01
Z test Statistic		-1.426
P-value		0.154
asm=1	6.63	9.8
asm=0	93.37	13.39
Z test Statistic		2.620
P-value		0.009***
alg=1	7.02	9.12
alg=0	92.98	13.45
Z test Statistic		3.252
P-value		0.001***

*, **, and *** represent significant at the 10%, 5%, and 1% significance levels respectively.

Table 4. Coefficient Estimates to Explain the Prevalence of Disease

Variable	Y_1= resp coeff. (Std.Err.)	Pr > ChiSq	Y_2=asm coeff. (Std.Err.)	Pr > ChiSq	Y_3= alg coeff. (Std.Err.)	Pr > ChiSq
Intercept	-4.405*** (0.403)	<.0001	-3.426*** (0.370)	<.0001	-3.929*** (0.372)	<.0001
age	0.018*** (0.003)	<.0001	0.003 (0.003)	0.292	0.000 (0.003)	0.981
white	0.366 (0.255)	0.151	0.269 (0.222)	0.225	0.522** (0.241)	0.031
male	-0.077 (0.084)	0.360	-0.548*** (0.094)	<.0001	-0.606*** (0.095)	<.0001
eduy	-0.0318* (0.017)	0.067	0.025 (0.018)	0.167	0.0682*** (0.018)	0.000
income	0.000 (0.000)	0.103	0.000011** (0.000)	0.015	0.000012*** (0.000)	0.007
exercise	-0.286*** (0.083)	0.001	-0.152* (0.085)	0.074	0.283*** (0.081)	0.001
smoker	0.953*** (0.091)	<.0001	0.495*** (0.087)	<.0001	-0.141* (0.082)	0.085
ph	0.013 (0.115)	0.912	-0.313** (0.140)	0.026	-0.408*** (0.139)	0.003
eastky	0.463*** (0.092)	<.0001	-0.109 (0.087)	0.208	0.186** (0.087)	0.032
N	9539		9539		9539	
LLR	212.772		83.444		121.612	
P>ChiSq	<.0001		<.0001		<.0001	

*, **, and *** represent significant at the 10%, 5%, and 1% significance levels respectively.

Table 5. Coefficient Estimates to Explain the Choice of Heating Fuel

Variable	Y=ph coeff. (Std.Err.)	Pr > ChiSq	Y=ph coeff. (Std.Err.)	Pr > ChiSq	Y=ph coeff. (Std.Err.)	Pr > ChiSq
Intercept	-1.637*** (0.302)	<.0001	-1.619*** (0.302)	<.0001	-1.636*** (0.302)	<.0001
resp	0.005 (0.115)	0.967				
asm			-0.320** (0.140)	0.023		
alg					-0.397*** (0.140)	0.005
age	-0.007*** (0.002)	0.001	-0.007*** (0.002)	0.0009	-0.007*** (0.002)	0.001
white	0.554*** (0.213)	0.009	0.557*** (0.213)	0.009	0.561*** (0.213)	0.008
male	0.073 (0.064)	0.255	0.064 (0.065)	0.324	0.061 (0.064)	0.344
eduy	-0.106*** (0.013)	<.0001	-0.106*** (0.013)	<.0001	-0.104*** (0.013)	<.0001
income	-0.00002*** (0.000)	<.0001	-0.00002*** (0.000)	<.0001	-0.00002*** (0.000)	<.0001
exercise	0.104* (0.062)	0.095	0.101 (0.062)	0.107	0.110* (0.062)	0.079
smoker	0.272*** (0.064)	<.0001	0.279*** (0.064)	<.0001	0.267*** (0.064)	<.0001
eastky	1.089*** (0.080)	<.0001	1.087*** (0.080)	<.0001	1.093*** (0.080)	<.0001
N	9539		9539		9539	
LLR	421.581		427.148		430.413	
P>ChiSq	<.0001		<.0001		<.0001	

*, **, and *** represent significant at the 10%, 5%, and 1% significance levels respectively.

Table 6. Logit and GMM-IV Estimation Results for Prevalence of Asthma

Variable	LOGIT			GMM-IVE		
	Coeff.	Std. Err.	P[|Z|>z]|	Coeff.	Std. Err.	P[|Z|>z]|
CONSTANT	-3.070***	0.226	0.000	-3.070***	0.459	0.000
WHITE	0.241	0.219	0.272	0.241	0.231	0.297
MALE	-0.555***	0.093	0.000	-0.555***	0.162	0.001
INCOME	0.113D-04**	0.464D-05	0.015	0.103D-04**	0.518D-05	0.047
SMOKER	0.471***	0.085	0.000	0.471	0.793	0.553
PH	-0.356***	0.137	0.009	-0.356	0.652	0.585
N	9539			9539		

*, **, and *** represent significant at the 10%, 5%, and 1% significance levels respectively.

Table 7. Logit and GMM-IV Estimation Results for Prevalence of Allergy

Variable	LOGIT			GMM-IVE		
	Coeff.	Std. Err.	P[|Z|>z]|	Coeff.	Std. Err.	P[|Z|>z]|
CONSTANT	-3.870***	0.316	0.000	-3.888***	0.451	0.000
WHITE	0.546**	0.241	0.023	0.589**	0.267	0.027
MALE	-0.622***	0.094	0.000	-0.655***	0.152	0.000
INCOME	0.121D-04***	0.444D-05	0.006	0.926D-05*	0.493D-05	0.060
EDUY	0.072***	0.017	0.000	0.073***	0.019	0.000
EASTKY	0.207**	0.086	0.016	0.211*	0.112	0.059
SMOKER	-0.141*	0.081	0.083	-0.144	0.710	0.839
PH	-0.387***	0.138	0.005	-0.360	0.847	0.671
N	9539			9539		

*, **, and *** represent significant at the 10%, 5%, and 1% significance levels respectively.