

Online Appendix: Are Home Buyers Inattentive? Evidence From Capitalization of Energy Costs

Erica Myers

July, 2018

A1 Derivation of Estimation Equation

In what follows, I develop a discrete choice framework where home buyer (i) chooses house (j) in geographic area (a) in year (t) from a choice set with budget constraint w_i . The consumer has an outside option of not buying a house with a utility level normalized to zero. Consumer i 's indirect utility from the purchase of a home is a function of the cost, which has two components: 1) the transaction price, H_{jat} , and 2) the net present value (NPV) of the expected stream of future fuel payments, F_{jat} . Utility is also a function of observable home attributes, \mathbf{X}_{jat} , unobservable home attributes, $\tilde{\xi}_{jat}$, neighborhood-year specific amenities, $\tilde{\lambda}_{at}$, and individual taste ε_{ijat} as follows.

$$U_{ijat} = \eta(w_i - H_{jat} - \gamma F_{jat}) + \mathbf{X}'_{ja}\tilde{\beta} + \tilde{\xi}_{ja} + \tilde{\lambda}_{at} + \varepsilon_{ijat} \quad (\text{A1})$$

The marginal utility of money is represented by η . The implied discount rate is the discount rate that consumers would have to be using for $\gamma = 1$. If the implied discount rate is higher than the borrowing rate for the marginal dollar, then consumers are inattentive energy costs.¹ In other words, demand for homes with high fuel costs is too high relative to what would be optimal.

The choice to sell or buy a home in any given year is driven by exogenous events such as changes in employment or changes in family composition. All potential home buyers in geographic area a in year t have the same income and face the same choice set of homes,

¹The word “inattentive” means that consumers are undervaluing a dollar spent on future energy costs relative to a dollar spent in upfront price. This “mistake” might arise through several potential mechanisms such as imperfect information, biased beliefs, present bias, or bias toward concentration.

where they are trading off the price of a home versus attributes such as square footage or number of bedrooms and bathrooms. Assume a traditional representative consumer logit model where ε_{ijat} is distributed i.i.d. extreme value. Integrating over ε_{ijat} and taking the natural log of both sides gives the following relative choice probability as a function of prices and characteristics.

$$\frac{1}{\eta}(\ln\phi_{jat} - \ln\phi_{0at}) = H_{jat} - \gamma F_{jat} + \mathbf{X}'_{ja}\beta + \xi_{ja} + \lambda_{at} \quad (\text{A2})$$

On the left hand side is the choice probability for a house, ϕ_{jat} , relative to the choice probability of the outside option, ϕ_{0at} . Dividing by η gives the new variables $\mathbf{X}'_{ja}\beta$, ξ_{ja} , and δ_{at} , which can be interpreted as dollar value of the utility represented by $\mathbf{X}'_{ja}\tilde{\beta}$, $\tilde{\xi}_{jt}$, and $\tilde{\delta}_{at}$.

This can be rearranged into an econometric estimating equation of transaction price as a function of fuel costs and a set of fixed effects as follows:

$$H_{jat} = \gamma F_{jat} + \lambda_{at} + \theta_{ja} + \epsilon_{jat} \quad (\text{A3})$$

Variation in the probability of choosing the outside option over time and across space is absorbed by geographic area by year fixed effects λ_{at} , which also control for shocks common to all houses in a given geographic area in a given year. House specific fixed effects (θ_{ja}) control for time invariant observable (X_{ja}) and unobservable characteristics (ξ_{ja}). Since the same house is sold more than once, it is being perceived by different sets of buyers across time periods. The new error term $\epsilon_{jat} = \ln\phi_{jat}$ represents the idiosyncratic changes in the preferences for particular house due to the buyers in a particular period, and is uncorrelated with fuel price. This is a similar theoretical approach to that taken in Allcott and Wozny 2014 in the context of car markets, which uses cross-sectional variation in fuel economy interacted with variation over time in gasoline prices to get plausibly exogenous variation in lifetime fuel costs of cars. In this analysis, I use relative fuel price movements of oil and natural gas as a plausibly exogenous instrument for F_{jat} .

A2 Equivalence of Two Estimation Approaches

In what follows I describe the equivalence of two estimation approaches for estimating the difference-in-differences of interest: the difference in housing price between oil and gas heated homes from one year (e.g. high price difference) to the next (e.g. low price difference).

Let α be the individual effects of each fuel price/fuel type combination as follows:

$$Outcome_{jat} = \beta_0 + \alpha_1(I_{jt}^{oil} \times p_t^{oil}) + \alpha_2(I_{jt}^{oil} \times p_t^{gas}) + \alpha_3(I_{jt}^{gas} \times p_t^{oil}) + \alpha_4(I_{jt}^{gas} \times p_t^{gas}) + \lambda_{at} + \theta_j + \epsilon_{jat} \quad (A4)$$

Following the notation in the paper *Outcome*, is the outcome for unit j in geographic area a in year t . The annual fuel price is p with oil and gas indicated with superscripts, “I” indicates the primary heating fuel with oil and gas as superscripts. λ_{at} are geographic area by year fixed effects, θ_j are individual fixed effects, and ϵ_{jat} is the error term.

Factoring out the fuel indicators in Equation A1 yields the following estimation equation:

$$Outcome_{jat} = \beta_0 + \beta_1 I_{jt}^{oil} \times (p_t^{oil} - p_t^{gas}) + \beta_2 I_{jt}^{gas} \times (p_t^{oil} - p_t^{gas}) + \lambda_{at} + \theta_j + \epsilon_{jat} \quad (A5)$$

The coefficient β_1 in the following equation estimates difference-in-differences of interest. Since one of these terms, e.g. $I_{jt}^{gas} \times (p_t^{oil} - p_t^{gas})$ is collinear with year fixed effects, the estimate of $\beta_1 = (\alpha_1 - \alpha_2) - (\alpha_4 - \alpha_3)$.

The estimation procedure used in the paper yields equivalent results, where this time the terms from Equation A1 are gathered slightly differently:

$$Outcome_{jat} = \beta_0 + \beta_1(I_{jt}^{oil} \times p^{oil} + I_{jt}^{gas} \times p^{gas}) + \beta_2(I_{jt}^{gas} \times p^{oil} + I_{jt}^{oil} \times p^{gas}) + \lambda_{at} + \theta_j + \epsilon_{jat} \quad (A6)$$

Again, one term will be collinear with year fixed effects, e.g. $I_{jt}^{gas} \times p^{oil} + I_{jt}^{oil} \times p^{gas}$. I simplify the first term to p_t since $p_t = I_{jt}^{oil} \times p^{oil} + I_{jt}^{gas} \times p^{gas}$. β_1 yields and equivalent difference-in-differences, i.e. $\beta_1 = (\alpha_1 - \alpha_4) - (\alpha_3 - \alpha_2)$.

A3 Two Stage Approach

A3.1 Data

I use energy expenditure data from the residential energy consumption survey (RECS) for the two-stage procedure detailed below. RECS is an in-home survey, which provides detailed information on housing unit characteristics as well as energy usage and expenditures by fuel

type and end-use. The price and expenditure data are verified with households' residential energy suppliers to ensure their reliability.

The survey is conducted approximately every five years and is designed to be a nationally representative cross-section of U.S. housing units. I use data from 6 surveys performed between 1990 and 2009 in my analysis. I use data from the Northeast Census region to predict energy expenditure as a function of household income controlling for size and other housing characteristics. I limit the sample to owner-occupied, single family houses in the northeast census region—a total of 1545 housing units.

A3.2 Empirical Approach and Results

Since I do not directly observe billing data for each house, estimating the NPV of the stream of future fuel costs, F_{jat} , as a function of house characteristics as a first stage has the potential to introduce bias in the second stage. If the second stage included any variables that affect energy expenditure, and these variables are not included in the first stage, it would introduce mechanical correlation between the first stage residuals and those variables only included in the second stage.² This is the same argument as to why the same exogenous covariates need to be included in both the first and second stage in any two stage least squares (2SLS) estimation (Wooldridge, 2010).

For example, residential energy consumption surveys (RECS) provide a repeated cross-section of energy expenditure, energy consumption, heating fuel type, and housing characteristics. However locational information is limited to large areas such as census regions, which are aggregations of several states. Therefore, it is not possible to estimate the first stage with either house fixed effects or geographic area-by-time fixed effects. If the fixed effects are included in the second stage, but not the first stage, they will likely be correlated with the first stage residuals since unobservable factors that affect housing price likely also affect energy consumption. The inconsistency from the correlation can spillover to all coefficient estimates in the second stage. For completeness, in what follows I provide two-stage estimates. However, since it is not possible to control for location-specific trends or unit fixed effects, the interpretation of the estimates will be affected.

I estimate a two sample two-staged least squares (TS2SLS) model, a variant of the two sample instrumental variables (TSIV) procedure discussed in Angrist and Krueger (1992, 1995).³ This two-sample IV procedure addresses concerns about measurement error in the

²A simulation exercise demonstrating the empirical importance of the bias in this setting can be provided upon request.

³Angrist and Krueger (1992) show that consistent instrumental variables estimation is still possible if one sample contains the outcome, another distinct sample contains the exogenous regressor, and both samples

NPV of fuel expenditure, but the limitation still exists that the only exogenous covariates that can be included as controls must be present in both samples.

In the first stage, I use the RECS data to estimate the effect of heating fuel price movements on energy expenditures as a function of unit characteristics. Then, I estimate a variant of Equation 1 by regressing housing price from the CoreLogic data on the estimated expenditure and unit characteristics. The first stage of the estimation is as follows.

First Stage

$$Exp_{jt} = \beta_0 + \beta_1 p_t + \beta_2 I_{jt}^{oil} + \mathbf{X}_{jt}\beta + \delta_t + \epsilon_{jt} \quad (\text{A7})$$

The dependent variable is expenditure on the primary heating fuel, Exp , for unit j in survey year t . The annual fuel price is p , I_{it}^{oil} indicates oil as the primary heating fuel. As with the main estimation, I use one statewide average price for each of these fuels. δ_t are year fixed effects. X_{jt} is a matrix of covariates and ϵ_{jt} is the error term. The covariates for this estimation include, the number of rooms, bathrooms and stories, flexible controls for square footage, binned by 1000 square foot increments, and indicators for decade built. They were chosen because they were available and in both surveys and are comparable between the two samples. The second stage estimation is as follows, where estimates of expenditure, \hat{Exp}_{jt} , are a function of the house characteristics in the transactions data and the coefficients estimated in the first stage. The coefficient of interest, $\hat{\gamma}$ can be interpreted as the effect of a \$1 increase in the present value of annual fuel expenditure on housing price.

Second Stage

$$H_{jt} = \alpha_0 + \hat{\gamma} \hat{Exp}_{jt} + \alpha_1 I_{jt}^{oil} + \mathbf{X}_{jt}\beta + \delta_t + \epsilon_{jt} \quad (\text{A8})$$

The results from the estimation are displayed in Table A1. The first stage estimate shows that a \$1 increase in the annual MA residential fuel price leads to an increase of \$100 in annual expenditure for houses in the northeast census region. The reduced form estimate and the implied discount rate of 10% are quite close to the basic estimation in the main analysis.

contain the instrumental variable and other exogenous variables included in the model. Their two sample instrumental variables (TSIV) estimator is: $\hat{\beta}_{TSIV} = (Z_2'X_2/n_2)^{-1}(Z_1'Y_1/n_1)$, where Y is the outcome, X contains the endogenous regressor and other exogenous variables, and Z is the matrix of valid instrumental variables, n is the number of observations and subscripts denote the samples 1 and 2. Inoue and Solon (2010) show that $\hat{\beta}_{TS2SLS} = (\hat{X}_1'\hat{X}_1)^{-1}\hat{X}_1'Y_1$ and $\hat{\beta}_{TSIV}$ as proposed by Angrist and Krueger (1992, 1995) have the same probability limit, though TS2SLS is more asymptotically efficient in finite samples due an implicit correction for differences in the distribution of Z between the two samples.

A4 Estimation Using Futures Prices

I examine the sensitivity of the basic estimation to using the discount factor-weighted mean of future fuel prices rather than contemporaneous price. Crude oil and natural gas are traded for as much as 7 to 13 years in advance for later years in the sample. These years will have the largest impact on perceived prices in NPV terms. The discount factor-weighted mean is constructed as follows.

$$\frac{\sum_{i=t}^T \delta^i \cdot p_{ijt}}{\sum_{i=t}^T \delta^i} \quad (\text{A9})$$

Future periods are indexed by i , the discount factor is δ^i , and the fuel price for house j for future year i in year t is p_{ijt} . If the time horizon of the decision were limited to the number of future periods that derivatives are traded T , this would be the perceived price for an agent using the futures market to forecast price. I inflate forward prices according to the trade date and transform crude oil and natural gas prices into residential heating oil and natural gas prices using the average historical relationships between the traded fuel and the residential price. Specifically, I predict the average historical relationship using simple linear regressions of levels of residential retail prices on levels of crude oil or Henry Hub natural gas spot prices. The reason I do this is residential gas prices do not have a separate futures market and residential heating oil is not traded for time horizons of more than 2 to 3 years.

A sensitivity test using the discount factor-weighted average futures prices incorporates all of the information available in the futures market, but implicitly assumes that home buyers will use the discount-weighted average of the prices in traded years for periods beyond the last year for which there are trade data. I replicate Table 2 using the discount factor-weighted average futures price. The results are displayed in Table A2. The magnitude of the point estimates are quite close to those using contemporaneous prices. For the preferred specification (column 5) with geographic area by time and house fixed effects, the estimates using the discount factor-weighted mean futures price actually yield the same implied discount rate of 9.5% as the estimates using contemporaneous price. Futures prices tend to not deviate too far from spot prices, meaning even if consumers were paying attention to them, their decisions would not deviate significantly from a consumer using contemporaneous prices.

A5 Estimation Using Pre-Crisis Data

I examine the sensitivity of the basic estimation to using the sub-sample of data from before the financial crisis (1990-2007). The results are displayed in Table A3. The first column displays results from the preferred estimation with geographic area by time fixed effects and house fixed effects (Equation 1). Column 2 includes results from a model with an oil-specific linear time trend and column 3 includes the results of a model with both an oil-specific linear time trend and age FE. The magnitude of the point estimates and implied discount rates are quite close to those using the full data set. Therefore, it appears the results are not being driven by differential trends in oil and gas homes during the housing crisis.

A6 Estimation of Rates of Conversion from Oil to Gas

A6.1 Data

I use data from the national American Housing Survey (AHS) from 1985 to 2011 to estimate rates of conversion from oil to gas. The AHS is a nationally representative survey, performed every two years. The AHS reports data on many attributes of a housing unit, including primary heating fuel. Importantly, the same housing units are surveyed every two years (with additions to reflect new construction). In the publicly available national survey, the only information on the location of the home is the census region.

Therefore, in order to get a sense of how many housing units are converting from oil to gas in Massachusetts over the period, I focus on the Northeast Census region.⁴

A6.2 Method

I begin by limiting the sample to the 16,896 single family homes in the northeast census region. Next, I drop any homes that have ever had a main heating fuel listed other than oil or gas, or have appeared to switch fuel type more than once, so that I have 10,934 remaining homes. Of those homes, 37% (4007) have oil listed as the primary heating fuel for the first observation of that home. Of the homes that begin with oil as the primary heating fuel, 450, or around 11% list gas as the primary heating fuel in later surveys.

⁴The Northeast Census region, region 1, is comprised of the following states: Maine, New Hampshire, Vermont, New York, Massachusetts, Connecticut, Rhode Island, Pennsylvania, and New Jersey.

Appendix Tables

Table A1: Two Sample 2SLS: Estimation of the Effect of Relative Annual Energy Costs on Relative Transaction Prices Instrumented with Fuel Price

(Dependent Variable)	First Stage (Annual Fuel Expenditure)	Reduced Form (Sales Price)	Second Stage (Sales Price)
Fuel Price	100.4*** (12.92)	-1082.4*** (291.0)	
Estimated Annual Fuel Expenditure			-10.79** (4.89)
Oil Heat Indicator	-157.6*** (39.23)	-6067.9 (4852.7)	-7767.7 (5144.28)
F-stat	20.42		
R ²	0.363		
Attribute Controls	Yes	Yes	Yes
N	1515	909434	909434
Implied Discount Rate			10%
Infinite Horizon			

Notes: First stage regression data are from the Residential Energy Consumption Survey (RECS), northeast census division, survey years 1990, 1993, 1997, 2001, 2005, and 2009. the second stage regressions are based on transaction and unit characteristic data for the state of Massachusetts from CoreLogic, years 1990-2011. Price is the average annual residential retail fuel price for oil or natural gas in dollars per MMBTU. All specifications control flexibly for the house vintage, number of rooms, bedrooms and bathrooms, square footage and year fixed effects. All prices are inflated to 2012 dollars. Standard errors are bootstrapped with 10,000 iterations, clustered at geographic area, and are in parentheses. ***, ** and * denote statistical significance at the 1, 5 and 10 percent levels.

Table A2: Estimation of the Effect of Relative Fuel Prices on Relative Transaction Prices

	Sales Price	Sales Price	Sales Price	Sales Price	Sales Price
Mean Futures Price	-1160.5*** (179.4)	-1028.1*** (202.3)	-1216.3*** (119.0)	-972.9*** (153.2)	-1058.6*** (121.7)
Oil Heat Indicator	-15644.3*** (1055.0)	-8458.9*** (1174.8)	955.0 (1029.9)		
Year FE	Yes	Yes	No	Yes	No
Attribute Controls	No	Yes	Yes	No	No
Geographic Area \times Year FE	No	No	Yes	No	Yes
Unit FE	No	No	No	Yes	Yes
N	909434	870567	870504	529156	529008
Implied Discount Rate Infinite Horizon	8.6%	9.8%	8.2%	10.4%	9.5%

Notes: Regressions are based on transaction and unit characteristic data for the state of Massachusetts from CoreLogic, years 1990-2011. The Mean Futures Price is calculated by weighting all traded futures prices by the discount factor and is measured in dollars per MMBTU. All prices are inflated to 2012 dollars. Standard errors are two-way clustered at the house and geographic unit by year level and are in parentheses. ***, ** and * denote statistical significance at the 1, 5 and 10 percent levels.

Table A3: Estimation of the Effect of Relative Fuel Prices on Relative Transaction Prices: Pre-Crisis Years

	Sales Price	Sales Price	Sales Price
Fuel Price	-1213.7*** (258.0)	-833.8*** (284.4)	-767.1*** (244.0)
Geographic Area \times Year FE	Yes	Yes	Yes
Unit FE	Yes	Yes	Yes
Oil Linear Trend	No	Yes	Yes
Age FE	No	No	Yes
N	412690	412690	412468
Implied Discount Rate Infinite Horizon	8.2%	12.4%	13.6%

Notes: Regressions are based on transaction and unit characteristic data for the state of Massachusetts from CoreLogic, years 1990-2007. Price is the average annual residential retail fuel price for oil or natural gas in dollars per MMBTU. All prices are inflated to 2012 dollars. Standard errors are two-way clustered at the house and geographic unit by year level and are in parentheses. ***, ** and * denote statistical significance at the 1, 5 and 10 percent levels.