
by
Braeton J. Smith, Student
Applied Economics Program, Oregon State University
318 Ballard Extension Hall
Corvallis, OR 97331
Phone: 1-503-280-9539/E-mail: braeton@gmail.com

Student Intern/Economist
Sandia National Laboratories
Center for Geoscience, Climate, and Consequence Effects
Albuquerque, NM

Abstract
The Energy Policy Act of 2005 provided incentives for all sectors of the energy industry, particularly to those in the nuclear, coal, and ethanol sectors. With regards to electricity, it provided incentives for the development of renewable and clean energy and exempted the fossil fuel companies from certain environmental standards. This paper uses panel data from the years 2004-2007 to determine the effect, if any, of the policy on residential electricity price while accounting for the fuel composition of the electricity generation mix. It finds that there was a significant increase in the price level after 2005, and determines the price effect of different generation technologies.

1. Introduction
There has been much recent debate surrounding the development of new energy infrastructure for both the more traditional fossil fuels and for renewable sources alike. Mostly under the guise of energy independence, policymakers have long contended that the national energy demand be met by domestic supply. This carries important implications for all sectors of the energy industry, as recent policy has provided incentives for development not only in renewable and “green” technologies, but in the coal, natural gas, and petroleum sectors as well. Specifically, the Energy Policy Act of 2005, signed into law by President George W. Bush in August of that year, provides directives for, energy efficiency, renewable
energy, electricity, and nuclear matters and security. It also provides a suite of incentives for the fossil fuel industries and climate change technology, among other things.\(^1\) Many of the benefits were given directly to the coal, petroleum, ethanol, and nuclear industries.

Along with authorizing subsidies to several of the nation’s energy developers, the act contains goals for reducing energy consumption and pollution. For the years 2007 to 2009 emissions from federal industrial facilities were to be decreased by 3 percent, by 5 percent for 2010 to 2012, and by at least 7.5 percent in 2013 onward.\(^2\) Despite these goals, however, criticisms of the act claim it to be nothing more than a “piñata of perks for the energy industries” which reduce environmental regulation (such as exempting oil and gas companies from some clean-water laws), leading to speculations of fraud.\(^3\) In short, the policy provided for the development of all types of energy without regard for emissions or environmental standards. Some of these benefits were targeted directly at the electricity industry in order to ensure against events such as the blackout that occurred in the eastern part of the country in 2003, encouraging continued development of nuclear, renewable forms such as wind and solar, “clean” coal technologies, hydropower projects, and others forms.

The goal of this paper is to determine the policy’s effects on residential electricity price, if any, while accounting for changes in the type of generation fuel as a result of the policy. It uses some theory from the traditional analysis of electricity demand, but focus is specifically placed on the price variable of the demand equation. The dataset consists of panel data for the years 2004 to 2007, indexed by cross sections for the fifty states. The main equation consists of some of the usual electricity demand determinants, such as income and residential sales, as well as variables describing the output from different fuel types. While it is more theoretically sound to use consumption statistics in determining demand, the generation, consumption, and emissions statistics for each fuel type are almost perfectly correlated. Using any one of the variables, then, merely causes a difference in scale, with the resulting models similar both in the signs of the coefficients and in significant variables. Thus, because only generation data was provided for many of the renewable industries, these data were used in the models. Two models were estimated initially, one including residential sales among the explanatory variables and one weighted by the number of retail customers; however, only one was selected for analysis (residential sales) due to the fact that the two models yield nearly identical results. Several versions of the model were predicted in order to account for heteroskedasticity and serial correlation, including one which incorporates a lagged dependent variable among the explanatory variables. The results indicate that there was indeed a significant increase in the price of residential electricity after 2005 and show some positive and negative effects due to changes in electricity composition. They also show that certain fuel types, such as electricity from wind, have no effect on residential electricity price.


Literature Review

There is no shortage of literature available on the determinants of electricity demand; however, there are few papers which take the choice of fuel into account directly. It is also difficult to find literature using panel data to analyze the effects of a change in specific policies, particularly the Energy Act of 2005, for which there is none to speak of. Also, the majority of papers on the topic focus on electricity demand directly, while developing an equation for prices to aid in their construction of a demand equation.

Halvorsen (1975)\(^4\) provides one of the more important works on the subject. He uses equations for both residential electricity demand and price to estimate the determinants of demand and determine price and income elasticities. Among the variables in Halvorsen’s demand equation are the average residential sales per customer (which serves as his dependent variable), the marginal price of residential electricity, average income per capita, time, average price for different types of gas, average temperature in July, percentage of households in rural areas, household size, and others. He notes that price should have a negative coefficient, while that for income should be positive. He also specifies determinants for the price equation as the nominal marginal price of electricity as a function of average annual sales per customer, cost of labor and fuel, the ratio of total industrial sales to residential sales, time, and others. As expected, he finds that residential sales have a negative effect on price while the cost of fuels has a positive effect. He also finds that the effect of income on demand is positive.

Diabi (1998)\(^5\) estimates total electricity demand for Saudi Arabia using cross-section (regional) data over the period from 1980 to 1992. Diabi notes that, along with price and income, there are at least two other important determinants of electricity consumption, notably changes in the price of complementary electric goods and changes in the price of substitute fuels. He provides a brief literature review over the debate between using marginal versus average prices in determining the amount of consumption. Although he notes that most economists contend that marginal price is theoretically the correct option, he cites several studies which argue that the use of average prices is acceptable, particularly because data on marginal prices is difficult to find. One study (Foster and Beatty, 1979)\(^6\) asserts that average prices are defensible when applied to aggregate data and another (Green, 1987)\(^7\) suggests that it is not unreasonable to assume that consumers respond to average prices since they react to the total amount of their electricity bill as opposed to its individual components.\(^8\) Diabi’s demand equation, then, is the quantity of electricity consumed per region in each time period as a function of real income times the number of customers in each region for each time period, the average electricity price, a one year lag of the dependent variable, and others variables. His results indicate several approaches of the model, among them ordinary least

---


\(^8\) It should also be noted that, in the United States, because electricity prices are largely determined through regulation, marginal and average prices are the same (a fact which challenges the assumption made by this paper of a price determined by the market).
squares, fixed effect, and random effect approaches. He finds that income has a positive effect in all approaches, with price having a positive effect in the OLS model and negative effect in all others, and the lagged variable having a positive effect in each approach.

Nakajima and Hamori (2010)\(^9\) estimate changes to the price elasticity of demand for U.S. residential electricity markets to examine household sensitivity to price changes as a result of utility deregulation. They use an econometric model similar to the other studies, estimating the residential consumption of electricity as a function of real personal income, overall unit price of residential electricity, and two parameters weighing the state population by heating and cooling variables. Their expected coefficients are the same as those in the other studies. They use a quarterly data from the lower 48 states for two periods, 1993-2000 and 2001-2008, forming a panel. Other studies not detailed here, but with similar results and models, are Wilder and Willenborg (1975)\(^10\), Halvorsen and Larsen (2001)\(^11\), Narayan and Smyth (2005)\(^12\), Ziramba (2008)\(^13\), and Tauchman (2006)\(^14\), which takes into account the prices of alternative fuels. There are a plethora of other papers on the topic, which use a range of econometric approaches and data applied to differing parts of the electricity industry, but all with similar basic models for demand.

2. Methods

Conceptual Framework

Although the purpose of this paper is to develop an equation for residential price, the theoretical background begins with a specification of the demand function. Although it is also necessary to specify a supply function so that an equation for price may be derived by solving both functions simultaneously, the demand function specified in this paper is a first step to a more thorough analysis. Most of the studies listed define a similar demand function, one which incorporates, at a minimum, the price of residential electricity, some measure of income, and time. One popular, and indeed important, variable is temperature; however, this variable was not included because data on yearly average temperature (the other data for this analysis are yearly averages) for each state is not particularly illuminating (whereas for


monthly data there exists a large degree of variation). Halvorsen (1975) specifies the general residential electricity demand equation as:

$$Q_d = Q_d(P_m, W, u)$$

where $P_m$ is the marginal price of electricity, $W$ is a vector of all other relevant variables (such as income), and $u$ serves as a disturbance term. Similarly, he defines price as:

$$P_m = P_m(Q, Z, v)$$

where $Z$ is a vector of exogenous variables that define the shape of the rate schedule (such as the supply or composition of electricity) and $v$ is a disturbance term. It is important to note that both equations contain exogenous variables which are not contained in the other, due mainly to the fact that price is not determined by residential demand alone.

The second equation is the equation of interest in this paper, with electricity generation included in $Z$. Clearly, the price of different fuels used to produce electricity is significant in the generation decision; however, this data is not as readily available as actual generation data itself. This paper argues that it is possible to include the aggregate generation from different fuel types to account for their composition of the electricity market, as the price of each fuel is intrinsic in these. As the electricity industry has a somewhat continuous stream of generation, the amount of generation corresponds directly to the amount of electricity consumed (as well as to the amount of emissions produced). Assuming equilibrium (and a competitive market), the price of electricity is determined at some point where consumption and generation are approximately equal, so the composition of electricity is directly related to its price. Thus, while no supply function is specified in this analysis, there are supply-side variables being considered as proxies for consumption.

Included in $Q$ are factors controlling for demand, such as income and consumption data. These variables are necessary for the model to maintain its economic interpretation and value. Residential sales data is suitable to account for consumption and also takes into account the relative population and demand for each state.

In controlling for fuel type and time, it should be possible to estimate the immediate changes which resulted from the Energy Act of 2005, if any. The main goal of the act was to provide sufficient funds and benefits to energy industries in order to detract from the large start up costs associated with certain projects, rather than to make radical immediate changes to the industry (such as cuts in the emissions standard, for example). For that reason, it is entirely possible that no real effects will be observed for the years specified; however, given the large amount of subsidies given to specific energy industries, it is likely that increases in generation and emissions will be observed. The hypothesis is that increased generation from certain fuel types will increase the residential price, while other types, such as hydropower, will have a negative effect.
Empirical Methodology

Most of the empirical demand models follow a similar form in the literature with the variables mentioned in the previous section (they also follow a similar functional form). The common variables are therefore included in the models for price proposed by this paper, as well as new variables incorporating fuel composition and time. There are at least two possible models, which are defined as follows:

\[
\ln P_{i,t} = \alpha_0 + \alpha_1 \ln Q_{i,t} + \alpha_2 \ln I_{i,t} + \sum_{j=1}^{6} \beta_j G_{j,i,t} + \sum_{k=1}^{3} \delta_k T_t + a_{i,t} + u_{i,t}
\]

\[
\ln P_{i,t} = \alpha_0 + \alpha_1 \ln l_{i,t} + \sum_{j=1}^{6} \beta_j G_{j,i,t} + \sum_{k=1}^{3} \delta_k T_t + a_{i,t} + u_{i,t}
\]

Where:
- \( i \) = state
- \( t = 2004, \ldots, 2007 \)
- \( P_{i,t} = \) average price of residential electricity in state \( i \) in year \( t \)
- \( Q_{i,t} = \) total residential sales of electricity in state \( i \) in year \( t \)
- \( I_{i,t} = \) median household income in state \( i \) in year \( t \)
- \( G_{j,i,t} = \) total electricity generation from fuel type \( j \) in state \( i \) in year \( t \), where \( j = 1 \) Coal
  2) Natural Gas
  3) Petroleum
  4) Nuclear
  5) Hydropower
  6) Wind
  and total generation: \( \bar{G} = \sum_{j=1}^{6} G_j \)
- \( T_t = 1 \) for each cross-section from 2005 to 2007 (2004 is base year \( T_0 \))
- \( a_{i,t} = \) fixed, or unobserved, effect (factors not varying over time)
- \( u_{i,t} = \) idiosyncratic error

As pointed out by Diabi (1998), data for the marginal prices as specified by Halvorsen (1975) and others as the theoretically correct price estimate is difficult to collect and not entirely necessary. Since average prices are accepted in the literature as a viable substitute, they were chosen for this analysis. For \( Q \), total residential sales, although not the average residential sales per customer used by Halvorsen (1975), provide a sufficient measure of demand since they essentially account for total residential consumption. In accordance with economic theory, it is hypothesized that this effect will be negative for model (1). The literature provides several different specifications for the income variable \( I \). Median household income by
state is used here because it accounts for the actual individual purchasing unit of residential electricity (the household itself) while eliminating some of the bias found in average incomes due to extreme outliers. Again, in accordance with economic theory it is expected that this effect be positive assuming electricity is a normal good (an increase in I provides further funds to consume more electricity, among other goods). As is common with these three variables, their natural log is included in the model so as to simplify elasticity inferences and meaning. The effect of generation G is ambiguous and depends fully upon its composition.

For the types of fuel \( j \), it is expected that the effects of \( G \) on \( P \) will be: \( \beta_1 < 0 \) for coal, \( \beta_2 > 0 \) for natural gas, \( \beta_3 < 0 \) for petroleum, \( \beta_4 > 0 \) for nuclear, \( \beta_5 < 0 \) for hydropower, and \( \beta_6 > 0 \) for wind. Many of these expectations are made based on the status quo, where, for example, coal is the dominant fuel type based on its relative low price and efficiency. The sign for nuclear is expected to be positive as can be inferred from the fact that the policy in question provided such lofty incentives for it (the nuclear industry received the largest sum of benefits). The same inference is made for wind energy, based on the same logic. Although \( G \) and \( Q \) are likely highly correlated (violating the assumption of no multicolinearity), this problem is partially solved for by decomposing \( G \) in the manner described. \( Q \) may or may not be correlated with any specific \( G_j \).

The time variable \( T_t \) is merely an indicator variable which equals one for each cross-section pertaining to the years 2004 to 2007. This is an important variable because of the stated purpose of this paper. That is, the signs and significance levels of \( T_1 \), \( T_2 \), and \( T_3 \) carry a great deal of weight in the end result as they account for changes in price from year to year. It is expected that the coefficients for each will be positive. The important implication for this paper, however, is the significance of the coefficients for \( T_2 \) and \( T_3 \) (which correspond to 2005 and 2006, respectively). If \( \delta_1 \) is not significantly different from zero, then the log of price for 2005 is the same as for 2004, other things equal. This is an important result if \( \delta_2 \) is positive and different from zero, implying that 2006 prices are higher than those for both 2004 and 2005 (which are essentially equal). If this is not the case, however, then this particular effect is inconclusive.

The difference between the two models is that (2) does not account for \( Q \). Weighting this model by the number of retail customers in each state reintroduces the size component back into the model, essentially imposing on it a weighted least squares (WLS) regression (which reduces the degree of heteroskedasticity). Because \( Q \) and \( C \) are almost perfectly correlated, they cannot be used in the same model, however, it is expected that they both have a nearly identical effect on the models. The difference in these models, then, is assumed to be only a matter of scale and the degree to which they exhibit heteroskedasticity (WLS must be performed on (2) in order for it to be valid). Due to this similarity, only (1) will be discussed for the remainder of this section. Perhaps the most important benefit of (1) is that it is able to be fit using the fixed and random effects models, which do not allow for weights. Its interpretation is also much more straightforward. The two models will be discussed in further detail in the Results section.

Initially, both models are estimated using pooled ordinary least squares (OLS), which includes no tests or corrections for heteroskedasticity or serial correlation. OLS has several assumptions, among the most important being normally distributed standard errors, constant variance (homoskedasticity), and no perfect multicolinearity among the explanatory variables. There are several methods by which to test
these assumptions. First, perfect multicolinearity is easily discovered by creating a correlation matrix of
the explanatory variables. Next, the Breusch-Pagan and White tests both check the assumption of constant
variance (with the null hypothesis of homoskedasticity). If OLS fails this test, the nonconstant variance
may be corrected for by using WLS, feasible generalized least squares (FGLS), or heteroskedastic-robust
standard errors (which do not affect the coefficient estimates).

In time-series estimations, such as this one, there is also the problem of serial correlation, or correlation
between the errors in different time periods. Although there are several tests for serial correlation, almost
none of them may be applied across panels. There are, however, two methods that easily deal the problem
of serial correlation. The first is to incorporate a lagged dependent variable into the model, generally an
autoregressive (AR) component of 1. An AR(1) version of (1) can be shown as:

\[
\ln P_{i,t} = \alpha_0 + \alpha_1 \ln Q_{i,t} + \alpha_2 \ln I_{i,t} + \ln P_{i,t-1} + \sum_{j=1}^{6} \beta_j G_{j,i,t} + \sum_{k=1}^{2} \delta_k T_k + a_{i,t} + u_{i,t}
\]

where \( P_{i,t} \) is the AR(1) component. If this variable is found to be significant, then there is proof of serial
correlation; however, adding the lagged variable to the model helps to resolve the issue. Another method
is to estimate the model using the Newey-West estimator, which also incorporates a lagged component.
This method corrects errors for both heteroskedasticity and serial correlation, while leaving the coefficient
estimates untouched.

Another test for OLS is the joint F-test, which tests the joint significance of explanatory variables
(specifically, that they are all equal to zero). This test is particularly useful for determining the joint
significance of indicator variables and those of a particular group. To the model developed in this paper,
the F-test may be applied to the time and generation variables. An interesting F-test for policy analysis is
the Chow test for structural change across time. This test estimates the residual sum of squares for each
OLS cross-section and calculates an F-statistic and \( p \)-value which may be used to infer structural change.
The test hypothesizes that there is no change across time, and a low enough \( p \)-value would indicate
otherwise.

Finally, there are two useful estimations previously mentioned that are of interest. These are the fixed
effects (FE) and random effects (RE) estimators, which take into account the time-invariant effects
captured by \( a_i \). If \( a_i \) is correlated with the explanatory variables, it is a FE and can be accounted for by
applying the FE transform to the model. If it is uncorrelated with the explanatory variables, it is known as
RE and the proper transformation may be applied to it. FE makes the assumption that there are no time
invariant explanatory variables, while RE does not. While there are a few other differences, this
assumption is the most important.

\textit{Description of the Data}

The data compose a balanced panel dataset, mostly composed of electricity production statistics for the
years 2004 to 2007. The majority of the data in this analysis were downloaded from the Energy
Information Administration’s (EIA) Electric Power Annual Data Tables.\textsuperscript{15} The only non-energy related data is median income by state, which was downloaded from the U.S. Census Bureau.\textsuperscript{16} Data was collected from several of the tables provided by the EIA; however, for the analysis only data from the Average Price by State by Provider, Retail Sales of Electricity by State by Sector by Provider, Net Generation by State by Type of Producer by Energy Source, and Number of Retail Customers by State by Sector datasets were used. On each of the retail data sheets, data were sorted by year and state, filtered for Total Electric Power Industry, and collected for the years 2004 to 2007. For the generation data, after filtering for Total Electric Power Industry, the data was filtered further so that individual data for each fuel type could be gathered for the years of interest. These data were then combined to form a balanced panel data set sorted by state and year. Similar methods were used to incorporate the income data into the dataset.

As there are four years in the dataset for all fifty states, there are a total of 200 observations. The data are sorted such that each state is listed with 4 observations (or panels), one for each year. More observations could easily be added to the dataset by simply adding one more year. The summary statistics for the important vectors (those used in this analysis) are shown in Table 1.

\textbf{Table 1: Summary Statistics of Selected Variables}

<table>
<thead>
<tr>
<th>Variable</th>
<th>Label</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P$</td>
<td>Average Res. Price (cents/KWh)</td>
<td>9.875</td>
<td>3.147</td>
<td>6.1</td>
<td>24.12</td>
</tr>
<tr>
<td>$Q$</td>
<td>Total Res. Sales (TWh)</td>
<td>2.694</td>
<td>2.634</td>
<td>0.206</td>
<td>12.684</td>
</tr>
<tr>
<td>$I$</td>
<td>Mean Household Income (US$)</td>
<td>47,374</td>
<td>7,497</td>
<td>32,875</td>
<td>68,059</td>
</tr>
<tr>
<td>$G_1$</td>
<td>Gen. from Coal (TWh)</td>
<td>3.999</td>
<td>3.833</td>
<td>0</td>
<td>14.888</td>
</tr>
<tr>
<td>$G_2$</td>
<td>Gen. from Natural Gas (TWh)</td>
<td>1.592</td>
<td>3.269</td>
<td>0</td>
<td>19.953</td>
</tr>
<tr>
<td>$G_3$</td>
<td>Gen. from Petroleum (TWh)</td>
<td>0.186</td>
<td>0.496</td>
<td>0</td>
<td>3.726</td>
</tr>
<tr>
<td>$G_4$</td>
<td>Gen. from Nuclear (TWh)</td>
<td>1.582</td>
<td>2.019</td>
<td>0</td>
<td>9.573</td>
</tr>
<tr>
<td>$G_5$</td>
<td>Gen. from Hydropower (TWh)</td>
<td>0.538</td>
<td>1.277</td>
<td>0</td>
<td>8.201</td>
</tr>
<tr>
<td>$G_6$</td>
<td>Gen. from Wind (TWh)</td>
<td>0.046</td>
<td>0.116</td>
<td>0</td>
<td>0.901</td>
</tr>
<tr>
<td>$C$</td>
<td>Res. Customers (x 1000)</td>
<td>242.6</td>
<td>244.5</td>
<td>23.6</td>
<td>1,281.9</td>
</tr>
</tbody>
</table>

3. Results

See the results from all regressions in Table 2. Initially, the model specified in (1) was estimated by OLS, showing interesting results and explaining almost 60 percent of the variation in the price of residential electricity. The coefficients on $Q$ and $I$ are as expected, however, only some of the expectations were correct for $G$. As expected, coal and hydropower have a negative effect and natural gas and nuclear have

\footnotesize
\textsuperscript{16} U.S. Census Bureau. State Median Income. Last revised March 1, 2011. \url{http://www.census.gov/hhes/www/income/data/statemian/index.html}

a positive effect. Somewhat surprising is that generation from wind has a negative effect and petroleum has a positive effect. It is important to note, however, that the coefficient on wind is not significantly different from zero, due mostly to its small magnitude. The results suggest that, of the fuel types, petroleum ($G_1$) has the largest effect, implying that for every 1 terawatt hour (TWh) of electricity produced from petroleum, the average price of residential electricity increases by 10.4 percent. Similarly, the results suggest that the price increases by 3.7 percent for every 1 TWh of natural gas ($G_2$). It should be noted that it is also possible to calculate the price elasticities of demand and income from these results.

---

17 A coefficient is significant at the 95 percent confidence level if its $t$-value is greater than 1.96 (or at the 90 percent level if greater than 1.64), where the $t$-value is the quotient of the coefficient and the standard error shown in the parentheses.
The coefficients of $\ln Q$ and $\ln I$ serve as inverse elasticities, however, calculating the inverse gives the price elasticity. Specifically, the price elasticity of demand is calculated as follows:
\[
\varepsilon = \frac{\partial \ln Q}{\partial \ln P} = \left[ \frac{\partial \ln P}{\partial \ln Q} \right]^{-1} = \alpha_1^{-1}
\]

The time variables, not surprisingly, all have a positive coefficient, although the coefficient on $T_1$ (signifying 2005) is not at all significantly different from zero. This is important because the other two coefficients are positive and significant at the 90% confidence level. This implies that the average residential price of electricity for 2005 was not significantly different from that in 2004 and that there was a significant price increase in 2006 following the Energy Policy Act of 2005. It is also important to note that the coefficients for both 2006 and 2007 are similar (.0753 and .0747, respectively), which suggests that prices for the two years were relatively constant. The interpretation suggests that for the years 2006 and 2007, the average price of residential electricity increases by roughly 7.5 percent. Breusch-Pagan (B-P) and White tests, however, show that the residuals from the OLS estimation exhibit a large degree heteroskedasticity, as the B-P test yields a $p$-value of zero. Therefore, hypothesis of homoskedasticity must be rejected.

To account for this, two estimations were performed: first, a WLS regression on the model specified in (2); and second, a regression with heteroskedasticity-robust standard errors. The first estimation weights (2) by the number of residential customers $C$, yielding very similar results. The results are so similar to the OLS results, in fact, that only the magnitude of the coefficients is different (although the elasticity of demand may not be calculated). The model is no longer heteroskedastic, however, as the B-P test yields a $p$-value of .068 (fail to reject that the standard errors have constant variance). This model, however, may not be used in the FE and RE transformations and the removal of the demand variable from the equation is questionable (it must be removed as $Q$ and $C$ have a correlation of .91). The heteroskedasticity-robust estimation simply strengthens the standard errors without changing the coefficients. Interestingly, after correcting for heteroskedasticity, the coefficient for wind generation $G_6$ becomes significant at the 90% confidence level.

The Chow test for structural change across time was performed on the OLS regression. The test yields a $p$-value of .999 and an $F$-statistic of 1.85. These results do not suggest significant evidence against the null hypothesis of no structural change (i.e. there is no evidence to support structural change).

Next, tests for serial correlation show that the model has errors which are correlated across time. The first test is to regress the model specified in (1b) with an AR(1) component. This variable is extremely significant and has a large effect on the model, which suggests both that the model contains a large degree of serial correlation and that the prices in the current period are greatly affected by the prices from the previous period, more so than by anything else in the model. This estimation only uses 150 observations and renders several of the explanatory variables insignificant while explaining 97 percent of the variation in $P$. Natural gas is now the only fuel type with a significant effect on price and $Q$ is given a positive coefficient, though it is not at all significant. Both 2005 and 2006 are significant at the 90% confidence level; however only 2006 is significant at the 95% level and the magnitude of its coefficient is much larger (2007 was removed because of multicollinearity). While this model suggests interesting results that are quite strong, it is not directly applicable to this dataset because it only analyses changes in price over...
three years. A study with more time periods could easily be specified with this model. The Newey-West estimation, similar to the heteroskedasticity-robust regression but with an AR(1) component, also corrects for serial correlation without changing the coefficients. Using this method causes the generation variable for petroleum to lose significance because of an increase in its standard error. In maintaining the model specified in (1), this regression corrects for both the heteroskedasticity and serial correlation in the model.

Lastly, the FE and RE regressions were performed on (1). The FE estimation suggests once again that natural gas is the only fuel type which has a significant and positive effect on price at the 95% confidence level (at .0046, it is also the largest coefficient for natural gas among the estimations). At the 90% level, however, coal also becomes significant and has a negative effect on price of almost the same magnitude. Also important is the fact that \( I \) is negative, though is no longer significant in the model. The assumptions for this method are well met, as all of the variables vary across time. The RE estimation differs in that the only insignificant variables are the fuel variables for petroleum, hydropower, and wind. The signs for the RE regression are as expected. Calculating the RE estimator \( \lambda \) shows that it is similar to the FE regression. Values close to zero suggest that the RE is more similar to the OLS results, while values close to one suggest that it is more similar to the FE. In this case, \( \lambda = .88 \).

4. Conclusion

The goal of this paper was to estimate the effects of the Energy Act of 2005 on the residential electricity price while accounting for changes in the composition of electricity generation as a result of the policy. In general, the model meets the expectations set forth in the introduction and provides some interesting results. While the model exhibits heteroskedasticity and, to a greater degree, serial correlation, the Newey-West estimation helps to correct these problems while incorporating an AR(1) component. This is perhaps the most robust method, although it does not account for the time-constant factors of the FE method.

The general findings among the different estimation techniques suggest that there is, in fact, a residential electricity price increase after 2005, as well as an effect caused by changes to the composition of electricity generation. The indicator variable for 2005 is insignificant in all but the AR(1), FE, and RE models, while the indicator variables for 2006 and 2007 are significant throughout the estimations. The insignificance of the 2005 variable indicates that prices for 2005 were not statistically different from those in the base year, 2004. Moreover, the fact that 2006 and 2007 are significant (and positive) indicates that there exist price increases after 2005 that are statistically different from previous levels.

Most notable for the generation variables is the effect of natural gas, which is the only generation variable which is significant for each estimation. Other important generation variables are coal, nuclear, and hydropower; however, these variables do not maintain their significance in the both the AR(1) and FE estimations. As anticipated in the methodology section, coal and hydropower are both found to have a negative effect on price and both natural gas and nuclear yield a positive one. Wind has no effect in any of the estimations and petroleum falls out of many of the corrected methods. Interestingly, the two fossil fuel technologies and the two “clean” technologies exhibit opposing signs and potentially offset one another. In fact, in the WLS regression, coal and natural gas are completely offset as the magnitude of the
coefficient is identical (.0022 and -.0022, respectively). Across the estimations, the magnitudes of nuclear and hydropower are also similar. This phenomenon, however, is likely an artifact of the data for states which are dominated by a specific fuel type. Because the dependent variable in the analysis is average price, variations across states are mainly determined by fuel type and, to a large extent, the differing price regimes used by state regulators.

While it may be difficult to directly attribute these effects to the Energy Policy Act of 2005, there is reason to believe that price effects occurred starting in 2006. Since the act was the only major policy affecting the electricity industry in that period, it is not unreasonable to attribute at least some of this effect to the act. Some important implications to recognize from this analysis are the lack of effect on price from wind generation (though the coefficient is, surprisingly, negative) and the offsetting nature of coal/natural gas and nuclear/hydropower. An interesting appendage to this study would be to determine the effects of the policy on both CO₂ emissions and generation from the different fuels using price as a dependent variable. This would essentially be modeling demand for emissions and generation, using the two as instruments for electricity consumption.

The limitations of this study are perhaps best illustrated by the use of the AR(1) model, in which the lagged dependent variable is extremely significant and increases the fit to 97%. This suggests that the price in the previous period has a dramatic impact on the price in the current period, more so than the factors accounted for by this model. This is, however, partially made up for in the Newey-West estimation as it includes an AR(1) component as well. Future work stemming from this research should include a great deal more time periods and placing less focus on the policy itself would give perhaps a more generalized result for the effects of the different fuel types. A more thorough analysis would present a market model which uses two-stage least squares (2SLS) to estimate both supply and demand using several instrumental variables for price. The current analysis, however, is a potential first step in that direction. Using demand alone to derive a price equation gives rise to the identification problem. It is also widely known that there was an increase in the price of natural gas in 2006, which may bias the results of this study. Lastly, because electric utilities are heavily regulated, the assumption that the residential price of electricity is determined by the market is problematic.
References


## Appendix: STATA Code

```stata
use C:\Users\smitbrae\Desktop\ELECTRIC.dta
xtset stateid year

reg lresprice lincome lressale coalgen ngasgen petgen nucgen hydrogen windgen >>> y05 y06 y07
estat hettest,iid
estimates store OLS
reg lresprice lincome coalgen ngasgen petgen nucgen hydrogen windgen y05 y06 >>> y07 [aw=rescust]
estat hettest,iid
estimates store WLS
reg lresprice lincome coalgen ngasgen petgen nucgen hydrogen windgen >>> y05 y06 y07, robust
estimates store Robust
newey lresprice lincome coalgen ngasgen petgen nucgen hydrogen >>>windgen y05 y06 y07, lag(1) force
estimates store Newey
reg lresprice lincome lressale lresprice_1 coalgen ngasgen petgen nucgen hydrogen >>>windgen y05 y06
estimates store AR
xtreg lresprice lincome lressale coalgen ngasgen petgen nucgen hydrogen >>>windgen y05 y06 y07, fe
estimates store FE
xtreg lresprice lincome lressale coalgen ngasgen petgen nucgen hydrogen >>>windgen y05 y06 y07, re
estimates store RE
scalar s_u=e(sigma_u)
scalar s_e=e(sigma_e)
scalar L=1-((s_u^2)/(s_e^2))^2

di L
*lambda=.8824554, RE and FE are basically the same
*Chow Test 1
reg lresprice lincome lressale coalgen ngasgen petgen nucgen hydrogen windgen >>> y05 y06 y07
scalar RSSr=e(rss)
reg lresprice lincome lressale coalgen ngasgen petgen nucgen hydrogen windgen >>> if y04==1
scalar RSS1=e(rss)
reg lresprice lincome lressale coalgen ngasgen petgen nucgen hydrogen windgen >>> if y05==1
scalar RSS2=e(rss)
reg lresprice lincome lressale coalgen ngasgen petgen nucgen hydrogen windgen >>> if y06==1
scalar RSS3=e(rss)
reg lresprice lincome lressale coalgen ngasgen petgen nucgen hydrogen windgen >>> if y07==1
scalar RSS4=e(rss)
scalar RSSur=RSS1+RSS2+RSS3+RSS4
*df numerator = (T-1)*k, k=8
scalar dfn=(4-1)*7
*df denominator = (n-T-Tk), n=200
scalar dfd=((200-4-(4*8)))
scalar Chow=((RSSr-RSSur)/dfn)/(RSSur/dfd)
di Ftail(dfd,dfn,Chow)
*p-value = .999 w n=200
di invFtail(dfd,dfn,.05)
*F-stat = 1.8516 w n=200, Fail to reject H0 (no struct change)
estimates table OLS WLS Robust Newey AR FE RE,b se t p
```