

Exhibit No.:

Issue(s):

Decoupling//
Rate Design//
Customer Confidentiality//
Line Extension in Unserved
and Underserved Areas//
Economic Development
Rider/Special Contracts

Witness/Type of Exhibit:

Marke/Rebuttal

Sponsoring Party:

Public Counsel

Case No.:

GR-2017-0215 & GR-2017-0216

REBUTTAL TESTIMONY

OF

GEOFF MARKE

Submitted on Behalf of the Office of the Public Counsel

**LACLEDE GAS COMPANY
MISSOURI GAS ENERGY**

CASE NO. GR-2017-0215

CASE NO. GR-2017-0216

October 20, 2017

**BEFORE THE PUBLIC SERVICE COMMISSION
OF THE STATE OF MISSOURI**

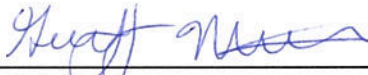
| | |
|---|-----------------------|
| In the Matter of Laclede Gas Company's) Request to Increase Its Revenues for Gas) Service) | Case No. GR-2017-0215 |
| In the Matter of Laclede Gas Company) d/b/a Missouri Gas Energy's Request to) Increase Its Revenues for Gas Service) | Case No. GR-2017-0216 |

AFFIDAVIT OF GEOFF MARKE

STATE OF MISSOURI)
) ss
COUNTY OF COLE)

Geoff Marke, of lawful age and being first duly sworn, deposes and states:

1. My name is Geoff Marke. I am a Regulatory Economist for the Office of the Public Counsel.
2. Attached hereto and made a part hereof for all purposes is my rebuttal testimony.
3. I hereby swear and affirm that my statements contained in the attached testimony are true and correct to the best of my knowledge and belief.




Geoff Marke
Chief Economist

Subscribed and sworn to me this 20th day of October 2017.



JERENE A. BUCKMAN
My Commission Expires
August 23, 2021
Cole County
Commission #13754037



Jerene A. Buckman
Notary Public

My commission expires August 23, 2021.

TABLE OF CONTENTS

| Testimony | Page |
|--|-------------|
| Introduction | 1 |
| Decoupling | 4 |
| Rate Design | 11 |
| Interclass Revenue Shift | 11 |
| Residential Customer Charge | 12 |
| Inclining Block Rates | 15 |
| Customer Confidentiality | 15 |
| Line Extension in Unserved and Underserved Areas | 17 |
| Economic Development Rider and Special Contracts | 21 |

REBUTTAL TESTIMONY

OF

GEOFF MARKE

**LACLEDE GAS COMPANY
CASE NO. GR-2017-0215**

**MISSOURI GAS ENERGY
CASE NO. GR-2017-0216**

1 **I. INTRODUCTION**

2 **Q. Please state your name, title and business address.**

3 A. Geoffrey Marke, PhD, Economist, Office of the Public Counsel (“OPC or “Public Counsel”),
4 P.O. Box 2230, Jefferson City, Missouri 65102.

5 **Q. By whom are you employed and in what capacity?**

6 A. I am employed by the OPC as the Chief Economist.

7 **Q. Please describe your education and employment background.**

8 A. I received a Bachelor of Arts Degree in English from The Citadel, a Masters of Arts Degree
9 from The University of Missouri, St. Louis, and a Doctorate of Philosophy in Public Policy
10 Analysis from Saint Louis University (“SLU”). At SLU, I served as a graduate assistant
11 where I taught undergraduate and graduate course work in urban policy and public finance. I
12 also conducted mixed-method research in transportation policy, economic development and
13 emergency management.

14 I have been in my present position with OPC since April of 2014 where I have been
15 responsible for economic analysis and policy research in electric, gas and water utility
16 operations. Prior to joining OPC, I was employed by the Missouri Public Service
17 Commission as a Utility Policy Analyst II in the Energy Resource Analysis Section, Energy
18 Unit, Utility Operations Department, Regulatory Review Division. My primary duties in that

1 role involved reviewing, analyzing and writing recommendations concerning electric
2 integrated resource planning, renewable energy standards, and demand-side management
3 programs for all investor-owned electric utilities in Missouri. I have also been employed by
4 the Missouri Department of Natural Resources (later transferred to the Department of
5 Economic Development), Energy Division where I served as a Planner III and functioned as
6 the lead policy analyst on electric cases. I have worked in the private sector, most notably
7 serving as the Lead Researcher for Funston Advisory based out of Detroit, Michigan. My
8 experience with Funston involved a variety of specialized consulting engagements with both
9 private and public entities.

10 **Q. Have you been a member of, or participant in, any work groups, committees, or other**
11 **groups that have addressed electric utility regulation and policy issues?**

12 A. Yes. I am currently a member of the National Association of State Consumer Advocates
13 (NASUCA) Distributed Energy Resource Committee which shares information and
14 establishes policies regarding energy efficiency, renewable generation, and distributed
15 generation, and considers best practices for the development of cost-effective programs that
16 promote fairness and value for all consumers. I am also a member of NASUCA's Electricity
17 and Water Committees each tasked with analyzing current issues affecting residential
18 consumers.

19 **Q. Have you testified previously before the Missouri Public Service Commission?**

20 A. Yes. A listing of the cases in which I have previously filed testimony and/or comments
21 before this commission is attached in GM-1.

22 **Q. What is the purpose of your rebuttal testimony?**

23 A. The purpose of this testimony is to respond to direct testimony regarding:

- 24
- Decoupling (“revenue stabilization mechanism” or “RSM”)

- 1 ▪ Laclede Gas Company & Missouri Gas Energy (“Laclede,” “MGE” or the
- 2 “Companies”) witness C. Eric Lobser
- 3 ▪ DE witness Martin R. Hyman
- 4 • Rate Design
- 5 ○ Interclass Revenue Shift
- 6 ▪ Missouri Industrial Energy Consumers (“MIEC”) witness Brian C. Collins
- 7 ○ Customer Charge
- 8 ▪ Laclede & MGE witness C. Eric Lobser
- 9 ▪ National Housing Trust (“NHT”) witness Annika Brink
- 10 ▪ Division of Energy (“DE”) witness Martin R. Hyman
- 11 ▪ Missouri Public Service Commission Staff (“Staff”) witness Robin
- 12 Kliethermes
- 13 ○ Inclining Block Rate (“IBR”)
- 14 ▪ Staff witness Robin Kliethermes
- 15 • Customer Confidentiality
- 16 ▪ Laclede & MGE witness C. Eric Lobser
- 17 • Line Extension in underserved and unserved areas
- 18 ▪ Laclede & MGE witness C. Eric Lobser and Scott A. Weitzel
- 19 • Economic Development Rider and Special Contracts
- 20 ▪ Laclede & MGE witness C. Eric Lobser and Scott A. Weitzel

21 **Q. Please state OPC’s position.**

22 A. OPC recommends that the Commission reject the Company’s decoupling mechanism and
23 adopt Staff’s rate design recommendations with the exception of the proposed residential
24 customer charge(s). In that instance, OPC recommends a \$14.00 customer charge for both
25 Laclede and MGE and correspondingly higher volumetric charge. We do not support a
26 movement to an IBR rate design presently.

1 Additionally, although OPC supports the spirit of the Companies' proposals for customer
2 confidentiality, line extension in underserved areas, and economic development rider/special
3 contracts, OPC cannot support the proposals as presently drafted. OPC is, however, willing to
4 continue to engage the parties in a dialogue on these issues and are hopeful for an amicable
5 resolution.

6 **II. DECOUPLING**

7 **Q. What is decoupling?**

8 A. The term "decoupling" is a blanket phrase that refers to a mechanism or rate design that
9 separates utilities' sales from its profits. A decoupling mechanism is a policy/accounting tool
10 that can be effective in removing the disincentive for utilities to promote energy efficiency in
11 certain situations or could merely transfer weather and economic risk to ratepayers in other
12 situations. Context and details matter otherwise this tool will not work as intended.

13 When "decoupling" is implemented as a mechanism, the "true-up" occurs outside of a rate case
14 and adjusts in isolation of all relevant factors. This is also known as "single-issue ratemaking"
15 and is prohibited in the context of traditional rate of return regulatory settings.

16 When "decoupling" is implemented through rate design, it is generally in the form of a straight-
17 fixed variable ("SFV") design in which distribution costs are recovered entirely in the customer
18 charge.

19 All things being equal, a SFV rate design is a much easier form of decoupling to implement
20 (no true-ups) and effectively avoids the issue of single-issue ratemaking; however, the price
21 signal being sent to ratepayers is one that encourages consumption (i.e., a buffet "all-you-can-

1 consume” style pricing signal) and is largely regressive, favoring consumers with higher
2 incomes who consume large quantities of natural gas (e.g., mansions over rental apartments).¹

3 Only gas companies in Missouri have a statutorily-enabled “opportunity” to make an
4 application to the Commission to approve a decoupling mechanism to reflect variation in
5 revenue increases/decreases due to weather, conservation/energy efficiency or both.² As DE
6 witness Hyman pointed out, it is unclear whether or not Commission rules would need to be
7 promulgated before such a mechanism is approved.

8 **Q. Does Laclede and MGE already have mechanisms in place to stabilize revenues?**

9 A. Yes. Again, as DE witness Mr. Hyman previously stated,³ Laclede and MGE have the
10 Purchased Gas Adjustment (“PGA”) clause and the accompanying Actual Cost Adjustment
11 (“ACA”) mechanism as well as the Infrastructure System Replacement Surcharge (“ISRS”) all
12 of which allow for recovery of costs in a more accelerated manner than is allowed in traditional
13 rate cases.

14 Moreover, both Laclede (current) and MGE (historical and current) have experience with
15 decoupling through rate design in the form of a SFV pricing scheme. On the west-side of the
16 State, a SFV was approved used by MGE from 2007 to 2014. It was a central contested issue
17 in the MGE GR-2009-0355 rate case which was supported by the Company, Staff and
18 ultimately the Commission; however, this rate design was subsequently adjusted to reflect an
19 approximate 17% recovery of revenues through the variable component as a result of a
20 compromised stipulation and agreement in its next rate case: GR-2014-0007. As a result of that
21 case, MGE’s “fixed” customer charge was set at \$23.00 and three years later still represents a
22 statistical pricing outlier compared to other investor-owned gas utilities across the nation.⁴

¹ Livingston, O.V. and K.A. Cort (2011) Analyzing the impact of residential building attributes, demographic and behavioral factors on natural gas usage. http://www.pnl.gov/main/publications/external/technical_reports/PNNL-20235.pdf

² See Section 386.266.3, RSMo.

³ See Direct Testimony of Martin R. Hyman p. 4, 16-20, p. 5, and p. 6, 1-6.

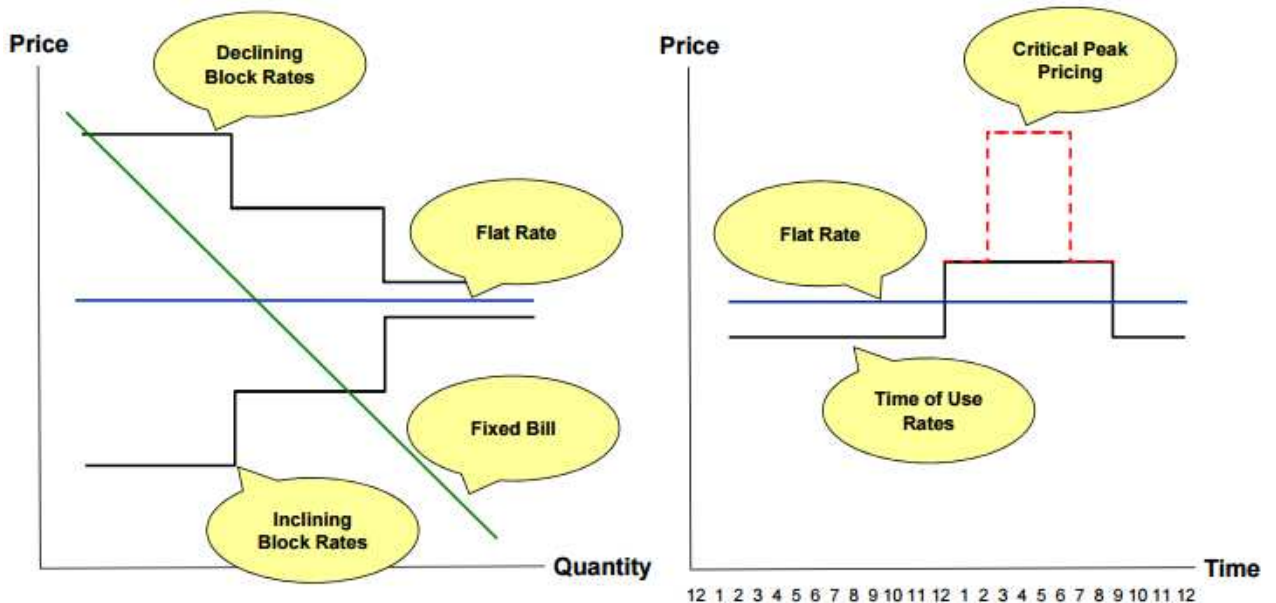
⁴ See the Direct Testimony of Annika Brink p.6.

1 On the east-side of Missouri, Laclede has in place a “Weather Mitigation Rate Design” that
 2 effectively functions as a SFV in the winter season. Laclede’s rate design includes a \$19.50
 3 customer charge but also a no per-therm charge for residential residents after the first 30
 4 therms. To provide an illustrative example of what this means, according to Staff’s residential
 5 cumulative frequency analysis, 95.81% of Laclede residential customers paid a \$47.01 flat fee
 6 to consume as much natural gas as they wanted in January.⁵ This is double the fixed cost
 7 customer charge “outlier” currently in place at MGE.

8 **Q. Does OPC support such rate designs?**

9 A. No. This “buffet style” pricing scheme represents an extreme byproduct of a fully-allocated
 10 embedded cost study that does not accurately characterize the long-run marginal cost of
 11 producing a service. The directional influence of price and usage of different rate designs can
 12 be observed in Figure 1. An SFV rate design increases overall consumption.

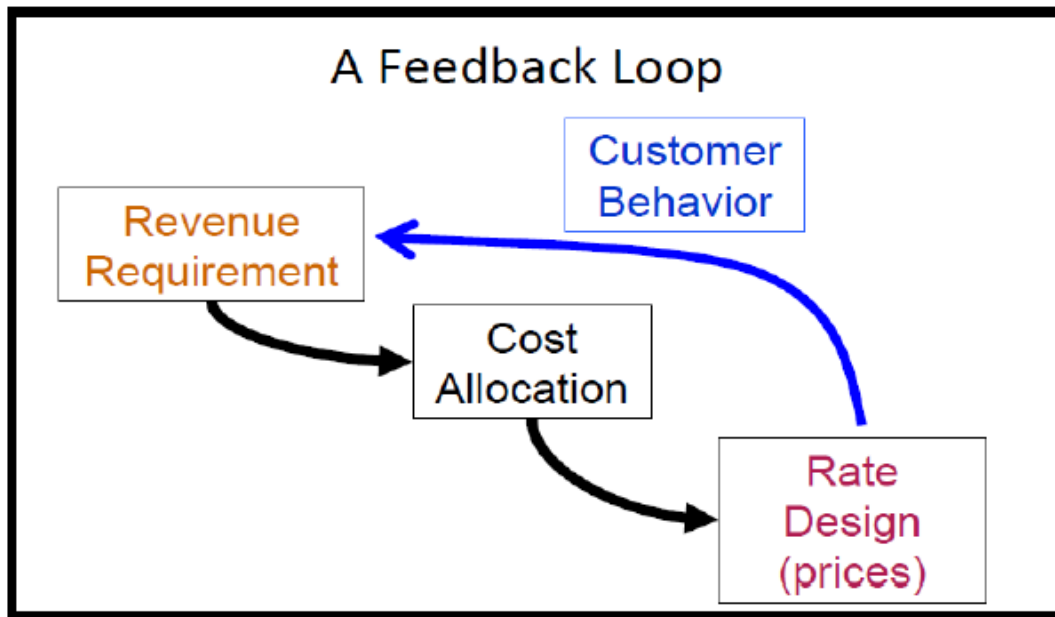
13 Figure 1: Rate Design pricing directions



14 ⁵ Of course \$47.01 would not be the full billed amount, additional costs are also included in Laclede’s customer bill including the ISRS and PGA.

1 Price—both its level and its form—is a powerful determinant of consumer behavior.
2 Accordingly, the setting and design of rates is one of the regulator’s most effective means by
3 which to achieve desired policy objectives. Therefore, how rates are designed will have an
4 impact on ratepayer behavior and future outcomes. For example, we know we can expect a
5 different response to a high customer charge and a low volumetric charge than from a low
6 customer charge and high volumetric charge—even if the two are designed to produce equal
7 revenues in the short run. In the long run, the chosen design will direct future costs because the
8 price signal functions as a feedback loop designed to influence customer behavior. This is
9 illustrated in Figure 2.

10 Figure 2: Feedback loop of rate design price signals



11
12 If consumers were restricted to only eating at all-you-can eat buffets there would no doubt be
13 secondary and tertiary societal impacts that would produce less than optimal outcomes the
14 same is true with pricing natural gas.

1 **Q. What are the combined companies proposing in this case?**

2 A. Laclede and MGE are proposing a decoupling mechanism, as the proposed tariff language
3 states:

4 The purpose of this Revenue Stabilization Mechanism (“RSM”) is to stabilize
5 customers’ utility bills and reduce over and under-recoveries of the base revenues
6 authorized in the Company’s most recent general rate proceeding due to changes in
7 residential and small general service customer usage.⁶

8 The RSM is a rebranded term for a decoupling mechanism. In the Company’s proposal, a “true-
9 up” or “revenue reconciliation” would occur at least once a year but no more than three times
10 annually. It appears as though any prior-period surplus/deficient would be recovered in the
11 RSM on a volumetric basis confined to the residential and general services customer classes
12 separately (not residential and general services together).

13 **Q. From OPC’s perspective, what is the general argument for a decoupling mechanism?**

14 A. A decoupling mechanism can be a regulatory tool to complement policy support for energy
15 efficiency programs.

16 **Q. Again, from OPC’s perspective, what is the general argument against a decoupling
17 mechanism?**

18 A. It further distorts the free market proxy that regulation is supposed to substitute for by shifting
19 risk to captive ratepayers away from shareholders by ensuring recovery of the Company’s
20 profits irrespective of market conditions or inefficient utility behavior. For a gas company, the
21 risk exposure to shareholders profits are, in part, present due to weather volatility, fluctuations
22 in the economy during periods of contraction (recessions) or the loss of customers. A
23 decoupling mechanism effectively eliminates those risks, and if ordered, should be married to
24 an explicit reduction in reward (i.e., a lower return on equity).

⁶ Sheet No. 50.

1 **Q. Is the promotion of gas energy efficiency programs an appropriate trade-off for a**
2 **decoupling mechanism?**

3 A. Not in this case. As the Commission is well aware, both MGE and Laclede already have
4 ratepayer-funded energy efficiency programs. Adding a decoupling mechanism does nothing
5 to enhance the programs or otherwise make them cost effective. Furthermore, the Company
6 has been forthright that they support these programs for purposes of enticing the conversion of
7 future gas customers. The conservation/efficiency argument is, at best, a distant secondary
8 consideration.

9 The reason these programs have not come close to expending their modest budgets (relative to
10 expenditures for electric energy efficiency) in the ten-plus years of existence is both because
11 of how we price natural gas service (SFV rate designs) as well as the low fuel price of natural
12 gas.

13 A customer has little financial incentive to buy an energy efficient furnace if the Commission-
14 approved pricing structure simultaneously encourages that same ratepayer to consume as much
15 as they want after 30 therms in the winter?

16 Additionally, the historic low price of natural gas fuel further complicates the promotion of
17 natural gas energy efficient products. In short, gas is both cheap and abundant while the limited
18 number of natural gas appliances are largely high cost capital items. Of the universe of items
19 to expend discretionary income on, prematurely replacing an inefficient natural gas furnace
20 before its useful life is not a decision that is cost effective or prudent for the vast majority of
21 EE participants in the near-term and is most certainly not a cost effective outcome for the non-
22 participants presently.

23 Based on my professional experience, most of the participants utilizing the rebates to date
24 would be considered “free riders,” that is, these efficient adopters would be purchasing efficient
25 furnaces regardless of whether or not it was subsidized. Because natural gas utilities do not
26 have an energy efficiency performance-based enabling statute in place, like MEEIA, this free

1 rider issue, although unfortunate and suboptimal, is not as disconcerting as it would be for
2 electric utilities who recover a throughput disincentive and earnings opportunity based on
3 verified induced energy and demand savings. Of course it should also be noted that there is no
4 specific future supply side investment deferral either for natural gas utilities as there is for
5 electric utilities.

6 **Q. Is the risk-mitigation from weather and economic volatility an appropriate trade-off for**
7 **a decoupling mechanism?**

8 A. Most assuredly not for ratepayers.

9 **Q. What is OPC's position on the Company proposed decoupling mechanism?**

10 A. The Commission should reject this proposal in total. The appropriate conditions for OPC to
11 support a decoupling mechanism are not currently present for Laclede and MGE. The harm to
12 captive ratepayers outweighs any alleged benefits. The Commission could enable an infinitely
13 more equitable and optimal outcome by adopting the Staff rate design at a minimum and could
14 best serve ratepayers by lowering the customer charge to \$14.00 to be further in line with cost
15 allocation and recovery seen in Missouri's electric and water utilities. This fixed cost recovery
16 would also be in line with peer by natural gas utilities throughout the U.S. as seen in table 1
17 reprinted from the National Housing Trust's witness Annika Brink's direct testimony.

1 Table 1: Residential Fixed Charges of Peer Natural Gas Utilities

| State | Utility | Fixed Charge | Sector | State | Utility | Fixed Charge | Sector |
|-------|------------------------|----------------------|-------------|-----------|---------------------------|----------------------|--------------------|
| SD | MidAmerican Energy | \$5.00* | Other | IA | Alliant Energy | \$13.00 | Residential |
| IN | Citizens Energy Group | \$9.00 [‡] | Residential | NE | Black Hills Energy | \$13.50 | Residential |
| MN | Xcel Energy | \$9.00 | Residential | IL | NICOR, IL | \$13.55 | Residential |
| SD | Mont.-Dakota Util. Co. | \$9.30 | Residential | NE | Metro Util. Dist of Omaha | \$13.72 | Residential |
| MN | CenterPoint Energy | \$9.50 | Residential | MO | Ameren Corporation | \$15.00 | Residential |
| IA | MidAmerican Energy | \$10.00 | Residential | TN | Piedmont Natural Gas Co. | \$15.45 [~] | Residential |
| TN | Memphis LG&W | \$10.00 | Residential | KY | Columbia Gas of KY | \$16.00 | Residential |
| WI | Wisconsin Gas | \$10.23 | Residential | KY | Louisville G&E Co. | \$16.35 | Residential |
| WI | WI Electric & Gas | \$10.23 | Residential | IL | Peoples Gas | \$16.37 [‡] | Residential |
| AR | CenterPoint Energy | \$10.75 | Residential | KS | Kansas Gas Service | \$16.70 | Residential |
| IN | NIPSCO | \$11.00 [^] | Residential | WI | WI Public Srvc Corp. | \$17.00 | Residential |
| IN | Vectren Corporation | \$11.00 | Residential | MO | Laclede proposed | \$17.00 | Residential |
| MI | DTE Energy | \$11.25 | Residential | KS | Black Hills Energy | \$17.25 | Residential |
| AR | Black Hills Energy | \$11.58 | Residential | KY | Atmos Energy Corp. | \$17.50 | Residential |
| MI | Consumers Energy | \$11.75 | Residential | MO | MGE proposed | \$20.00 | Residential |
| IN | Citizens Energy Group | \$12.00 ⁺ | Residential | IL | Peoples Gas | \$30.84 ⁺ | Residential |

*Applies to all customers or all customers may choose this rate, ⁺ Heating customers, [‡]Non-heating customers
[^]For master-metered multifamily housing of 2-5 units the fixed charge is \$12.50 instead. [~]This is the average of two seasonal charges: the April-October charge is \$13.45, November-March charge is \$17.45

2

3 **III. RATE DESIGN**

4 **Interclass Revenue Shift**

5 **Q. Does OPC support any interclass revenue shifts?**

6 A. OPC is still examining the class cost of services studies submitted in this case. Presently, OPC
 7 supports Staff's proposed interclass shifts and consolidation of customer classes. As an
 8 alternative we would support an equal percentage increase to each customer class from any
 9 increase from this case. We reserve the right to amend this recommendation in surrebuttal
 10 based on further inquiry.

1 **Residential Customer Charge**

2 **Q. What is OPC's position on the residential customer charge?**

3 A. OPC applauds the Company for recognizing and reversing an inequitable trend of
4 requesting higher customer charges that Missouri's electric investor-owned utilities have
5 been pursuing for the past few years. Laclede and MGE have proposed to reduce its
6 customer charges to allow ratepayers, and especially low income, low usage ratepayers,
7 greater control over their utility bills. This sentiment was expressed to ratepayers at each
8 public hearing by the Company and was amiably received by those present. If the present
9 customer charge is lowered, this will result in higher bills for above average customers of
10 natural gas but lower bills for below average consumers of natural gas. No doubt, over
11 time, this may result in fewer seasonal shut-offs and even greater gains for the company
12 and customers alike. And although Laclede and MGE's proposal is not in line with OPC's
13 ultimate recommendation, we acknowledge the importance of this departure from past
14 practice and welcome this customer-centric philosophy as the companies move forward as
15 Spire.

16 Presently, OPC recommends a \$14.00 customer charge for both Laclede and MGE. This
17 would place the residential customer charge in line with Missouri electric and water
18 utilities as well as natural gas utilities nationwide.

19 **Q. What is the general argument against a large residential customer charge?**

20 A. When having one or more customers on the system raises the utility's cost regardless of
21 how much the customer uses (billing is an example) then a fixed charge to reflect that
22 additional fixed cost the customer imposes on the system makes perfect economic sense.
23 Utilities can justify a customer charge recovering these basic costs because they are
24 directly related to the number of customers receiving an essential monopoly service. The
25 idea that each household has to cover its customer-specific fixed cost also has obvious

1 appeal on grounds of equity. However, system-wide “fixed” costs such as maintaining the
2 distribution network do not change if one customer were to drop off the system.

3 **Q. Does OPC believe that an increased customer charge would negatively impact low**
4 **income customers?**

5 A. Yes. Low-income and fixed income customers with low usage can all be seen as customer
6 groups with inelastic demands. These groups are subject to paying a higher mark-up above
7 marginal costs than another type of customer under the historic rate SFV rate design and
8 can be seen as a form of price discrimination.

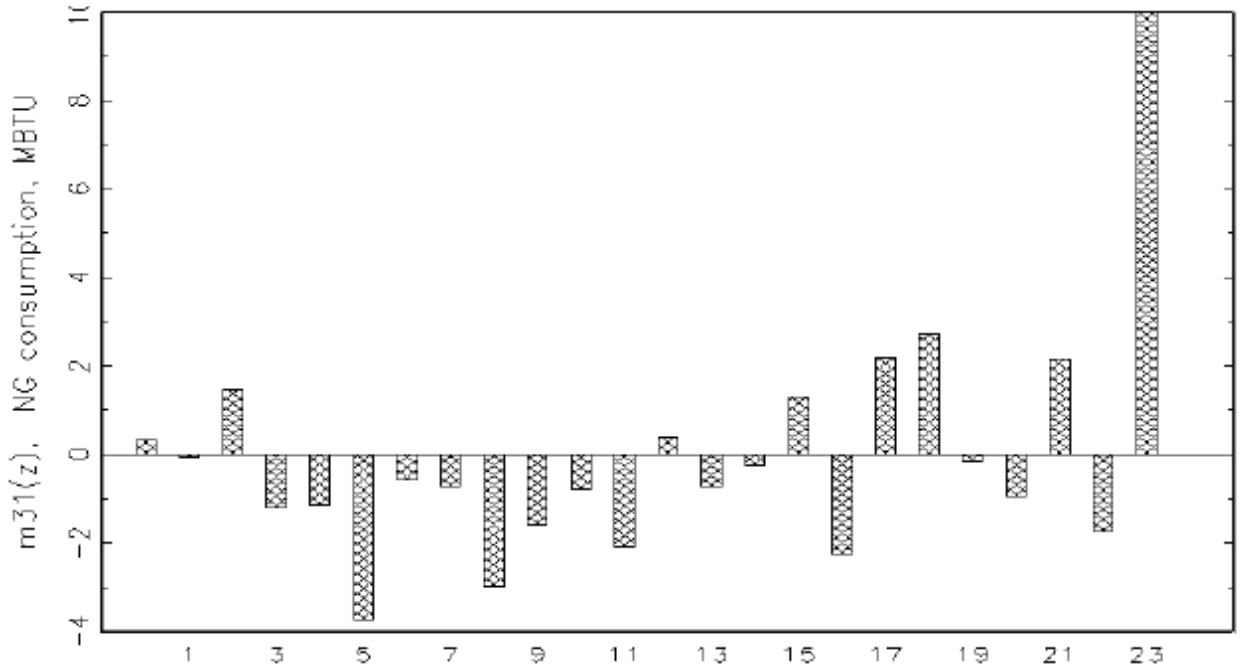
9 **Q. Do you have any empirical data to support the link between income and natural gas**
10 **consumption?**

11 A. Yes. The impact of household income on natural gas usage can be seen in Figure 3 which
12 reprinted from the US Department of Energy’s (“DOE”) analysis of the “Impact of Residential
13 Building Attributes, Demographic and Behavioral Factors on Natural Gas Usage.”

14 The data utilized for this regression analysis is based on the U.S. DOE’s Residential Energy
15 Consumption Survey (“RECS”) microdata and is not Laclede or MGE specific. However,
16 similar high-income/high-usage and low-income/low-usage patterns have been confirmed for
17 Missouri’s investor-owned electric utilities consumers as filed within recent triennial
18 Integrated Resource Plans (“IRP”).

1

Figure 3: Regression for Direction 31: Income and Natural Gas Usage⁷



2

| | | | |
|----|---------------------|----|-----------------------|
| 0 | Less than \$2,500 | 12 | \$50,000 - \$54,999 |
| 1 | \$2,500 - \$4,999 | 13 | \$55,000 - \$59,999 |
| 2 | \$5,000 - \$7,499 | 14 | \$60,000 - \$64,999 |
| 3 | \$7,500 - \$9,999 | 15 | \$65,000 - \$69,999 |
| 4 | \$10,000 - \$14,999 | 16 | \$70,000 - \$74,999 |
| 5 | \$15,000 - \$19,999 | 17 | \$75,000 - \$79,999 |
| 6 | \$20,000 - \$24,999 | 18 | \$80,000 - \$84,999 |
| 7 | \$25,000 - \$29,999 | 19 | \$85,000 - \$89,999 |
| 8 | \$30,000 - \$34,999 | 20 | \$90,000 - \$94,999 |
| 9 | \$35,000 - \$39,999 | 21 | \$95,000 - \$99,999 |
| 10 | \$40,000 - \$44,999 | 22 | \$100,000 - \$119,999 |
| 11 | \$45,000 - \$49,999 | 23 | \$120,000 or more |

3

4

⁷ Livingston, O.V. and K.A. Cort (2011) Analyzing the impact of residential building attributes, demographic and behavioral factors on natural gas usage. http://www.pnl.gov/main/publications/external/technical_reports/PNNL-20235.pdf

1 **Inclining Block Rates**

2 **Q. Please describe Staff’s alternative inclining block rate (“IBR”).**

3 A. OPC does not support an IBR at this time. Given the present “all you can consume” rate design
4 in place for residential customers, OPC would have concerns that such a departure could
5 produce some negative consequences to customers in the winter season. Although we
6 appreciate Staff’s effort to design such an alternative approach, it does not seem appropriate
7 given the present circumstances.

8 **IV. CUSTOMER CONFIDENTIALITY**

9 **Q. What is the proposed customer confidentiality provision?**

10 A. The Company is proposing to incorporate language into the Rules and Regulations section of
11 its tariff specifically related to “Customer Confidentiality.” The proposed language states:

12 All customer specific information will be treated as confidential and will not be
13 released to any other party outside of the Company without specific customer approval.
14 Customer specific information will include all billing statement information, usage
15 data and customer supplier/broker information. Except as provided below, the
16 Company shall notify the customer of any request to disclose such information and
17 shall not disclose such information except upon consent by the customer. This section
18 shall not be construed as precluding the Company from providing information
19 regarding customer status to law enforcement or emergency personnel acting in their
20 official capacity pursuant to procedures established by the Company, in which case the
21 Company shall not be required to notify the customer or obtain the customer’s consent.
22 Nor shall notice to the customer or customer consent be required when customer-
23 specific information is released pursuant to court order, subpoena or other order or
24 requirement issued by a duly constituted authority, or when release of such information
25 is necessary to provide service. In addition, neither notice to the customer nor customer

1 consent shall be required when customer-specific information is released pursuant to
2 request by the Missouri Public Service Commission or the Staff of the Missouri Public
3 Service Commission.

4 **Q. Does OPC support this language?**

5 A. No. Although OPC, in general, supports maintaining customer confidentiality, such language
6 is both unnecessary and may prove to be problematic for Commission-approved interveners in
7 future cases.

8 **Q. Please explain.**

9 A. OPC firmly believes that captive customer information (billing, usage, or otherwise) should
10 not be a channel for non-regulated, “value-added,” revenue-generating services (e.g., third-
11 party security systems—AllConnect, or political polling information); however, the inclusion
12 of the Company’s proposed tariff language may unfairly limit case-specific Commission-
13 approved parties from providing appropriate context and arguments in regards to matters of
14 rate design, economic development riders and special contracts (to cite several instances). In
15 these three examples, historical customer generated usage information may be necessary to
16 ensure that cost causation principles and relevant policy initiatives are upheld.

17 Finally, OPC reminds the Company that §386.480 MO. REV. STAT. (2017) (Information not
18 to be divulged—exception—penalty) suitable protection already exists and also specifically
19 allows OPC access to confidential customer information:⁸

20 No information furnished to the commission by a corporation, person or public utility,
21 except such matters as are specifically required to be open to public inspection by the
22 provisions of this chapter, or chapter 610, shall be open to public inspection or made
23 public except on order of the commission, or by the commission or a commissioner in
24 the course of a hearing or proceeding. The public counsel shall have full and complete

⁸ State of Missouri Revisor of Statutes: <http://revisor.mo.gov/main/Home.aspx>

1 access to public service commission files and records. Any officer or employee of the
2 commission or the public counsel or any employee of the public counsel who, in
3 violation of the provisions of this section, divulges any such information shall be guilty
4 of a misdemeanor (emphasis added).⁹

5 **V. LINE EXTENSION IN UNSERVED AND UNDERSERVED AREAS**

6 **Q. What is the Company proposing?**

7 A. Laclede and MGE are proposing modifications to the existing main extension policies that
8 would permit the Company(s) to extend natural gas service into unserved and underserved
9 areas. The costs associated with these extensions would be borne entirely by customers
10 directly benefitting from these extensions and be recovered through a geographical surcharge
11 for a set period of time. The Company would have the option of offering low interest financing
12 for the high capital costs associated with these extensions for a period of up to fifteen years.

13 **Q. What are unserved and underserved areas?**

14 A. An unserved area is remote from the nearest utility's gas system. A utility may have to make
15 substantial investments to construct a new main line to serve these areas. In contrast, an
16 underserved area may have main lines nearby but many households and businesses are instead
17 using other forms of energy (e.g., electric space heating or propane). For example, certain
18 sections of Jefferson City would be considered underserved areas because they use electricity
19 for space heating from Ameren Missouri electric even though Ameren Missouri gas has nearby
20 natural gas main lines which could be extended cost-effectively.

⁹ Revisor of Statutes, State of Missouri. Title XXV Incorporation and Regulation of Certain Utilities and Carriers:
<http://revisor.mo.gov/main/OneSection.aspx?section=386.480&bid=21737&hl=>

1 **Q. Will existing ratepayers be held harmless?**

2 A. Yes. According to Mr. Lobser there will be no risk transfer to existing customers with new
3 customers bearing the entirety of the incremental costs associated with extended services in
4 unserved or underserved areas.

5 **Q. Could you provide an illustrative example?**

6 A. A developer elects to construct a neighborhood of homes in an “unserved” rural area that
7 requires extending the Company’s distribution system in a manner that would be cost-
8 ineffective.

9 Under the Company’s proposed plan, natural gas service could be extended to this developer
10 and future residents; however, the incremental costs of this natural gas extension service would
11 be borne solely by those consumers and collected through a future geographic-specific
12 surcharge.

13 **Q. How much would the surcharge be?**

14 A. It is not clear. The actual cost incurred would vary based on the extent and conditions of the
15 site the extension line was offered.

16 **Q. In your example of the newly constructed rural neighborhood, what would happen if
17 those homes remained unoccupied?**

18 A. The Company’s shareholders would have to absorb the costs.

19 **Q. What if only one home in a lot of ten was ultimately occupied?**

20 A. As OPC understands the Company’s proposal, that one home would only be allocated one-
21 tenth of the total costs of that extension with the Company absorbing the other nine-tenths.

1 **Q. What if a homeowner entered into a fifteen-year financing arrangement and then**
2 **subsequently had their home foreclosed two-years into the contract?**

3 A. It is not entirely clear. No doubt, this line of questioning can produce multiple “what if”
4 scenarios that have not been discussed in Company testimony to date. Further clarification
5 from the Company is warranted.

6 Ultimately, if the Commission supports the Company’s proposed line extension, OPC would
7 recommend that an FAQ with multiple scenario-based questions be developed and readily
8 accessible on the Company’s website. This would mitigate any potential customer confusion
9 and provide transparency and clarity on billing arrangements as well as accurate price signals
10 for interested customers moving forward.

11 Moreover, the Company should be required to maintain any and all records associated with
12 line extension applications and contractual arrangements as a result of these transactions in the
13 event of future formal complaints.

14 **Q. What factors should consumers consider in regards to fuel switching?**

15 A. Fuel switching to natural gas can make sense for some customers but not others, even when
16 they live in the same neighborhood. Today, customers choosing a natural gas line extension
17 are able to benefit from historically low annual operating costs largely due to the low cost of
18 natural gas fuel. Among the many variables to consider include: the distance from a main line
19 extension, soil conditions (e.g., drilling through rock v. drilling through soil), the relative size
20 of the building unit, the thermal efficiency, and the customer’s preferred indoor ambient
21 temperature. Switching to natural gas may also require special plumbing or retrofit work and
22 purchase of natural gas appliances to justify the large upfront capital costs for line extension
23 and installation. As such, it can be extremely difficult for consumers to make informed
24 decisions without proper knowledge.

1 **Q. Does this market imperfection necessitate regulatory intervention?**

2 A. No. Consumers routinely make decisions in a complex environment where uncertainty,
3 transaction costs, and conflicting information exists. Fuel is no different and will no doubt
4 represent a moving target of future price fluctuations over the life of these line extension
5 investments. Constructing new gas lines to accommodate consumers' desire to switch from
6 one fossil fuel to another is hardly a "public need." Consumer's care about cost first and
7 foremost. Rather than serving a public need, the customer-specific choice decision to switch
8 fossil fuels falls outside the definition of a "necessity," and consequently, should be borne
9 solely by both that individual consumer and the Company to the extent that it elects to offer
10 uneconomic services.

11 **Q. Does OPC support the Company's proposal?**

12 A. Yes, in principle and based on OPC's interpretation of the proposal. If OPC is correct that the
13 line extension policy would hold existing ratepayers harmless and not liable for any non-or
14 under-recovery of incremental costs, OPC would generally be supportive of the spirit of the
15 Company's proposal. However, more information is needed. For example there are still
16 numerous concerns surrounding the specificity of the line extension cost recovery disclosure
17 between developers and future owners and/or tenants, which needs to be articulated in the tariff
18 before OPC can support the proposal in full. As presently drafted, there are entirely too many
19 questions still unanswered. At a minimum, full transparent disclosure of expected costs needs
20 to be communicated to future buyers that elect to enter into this payment arrangement which
21 could extend to multiple actors over many years. Additionally, the Companies need to make a
22 commitment that existing ratepayers will be held harmless, with procedures to make sure the
23 commitment is upheld. OPC is in the process of exploring these conditions and will respond
24 accordingly in surrebuttal testimony as necessary.

1 **VI. ECONOMIC DEVELOPMENT RIDER AND SPECIAL CONTRACTS**

2 **Q. What is the Company’s proposed economic development rider (“EDR”)?**

3 A. According to Company witness Scott Weitzel:

4 The EDR discounts would be available to any existing or potential customer that has
5 or is projected to have (once fully operational) a minimum of 30,000 Dth/yr in annual
6 usage that could be retained or added as a result of the EDR or the customer with
7 expanded usage of 15,000 Dth/yr or more. Commercial as well as industrial customers
8 would qualify for the rider as long as the customer or potential customer is not simply
9 shifting commercial or industrial activity from one portion of the Company’s service
10 territory to another without any significant increase in load or without documentation
11 that the customer might otherwise leave the state without an EDR discount. . . . Under
12 the EDR, eligible customers could receive discounts for a maximum period of 5 years.
13 Over that span, the discounts would be applied to the usage portion of the customer’s
14 base rates and could not exceed an annual average of 20%, nor more than 30% in any
15 one year. . . . As long as these conditions were met and the Company demonstrated in
16 a subsequent rate case proceeding that the EDR discounts were offered as part of an
17 overall effort by state and/or local officials to retain or attract business in the State of
18 Missouri, the actual revenues received or to be received from the customer under the
19 EDR discount arrangement would be used in setting rates.¹⁰

20 **Q. What is the Company’s proposed special contract?**

21 A. The Company is proposing a greater discount and a longer contract term as long as the
22 discounted customer covers the incremental costs to retain/attract its service and “at least
23 some” fixed cost contribution. The terms of the special contract can be extended for up to
24 fifteen years in total.

¹⁰ Direct Testimony of Scott A. Weitzel, p. 15, 20-23 & p. 16. 1-19.

1 **Q. Does OPC support such an incentive rate?**

2 A. Not as presently drafted. OPC does not oppose the inclusion of an EDR and/or Special Contract
3 option to entice new load and economic development but believes that the proposed tariffs
4 need to be more narrowly defined with an overall total annual incentive cap for all economic
5 development-induced discounted rates to minimize overall exposure.

6 It would be OPC's preferred outcome that any discounted rate be offered on a limited basis at
7 a deescalating sliding rate. A five-year discount is a reasonable length. A fifteen-year period
8 special contract rate, however, is five additional years beyond the most recent statutorily-
9 enacted "smelter/steel negotiated rate" terms of possible service from this past summer's
10 special legislative session. OPC has serious reservations committing ratepayers to undefined
11 "economic development" discounts that far into the future.

12 Additionally, because the customer load offers economic advantages for both ratepayers (via
13 a contribution to "some" of the utility's fixed costs) and to utility shareholders (via a utility's
14 earnings) OPC believes that the revenue loss from any discount should be shared equally.

15 Finally, OPC would also prefer any tariff the include language tying the discounted rates to a
16 minimum level of capital investment and/or job creation as well as some assertion that the
17 Special Contracts customer-specific fixed costs associated with the new load are recovered in
18 its rates.

19 To be clear, we believe this initial proposal is better than the existing EDR's which have largely
20 proven to be ineffective in attracting incremental economic development to Missouri and
21 support the spirit of the Company's proposal. OPC is presently working on drafting specific
22 provisions based on discussions with the Company and relevant stakeholders and reserves the
23 right to provide additional feedback in surrebuttal testimony.

24 **Q. Does this conclude your testimony?**

25 A. Yes.

CASE PARTICPATION OF
GEOFF MARKE, PH.D.

| Company Name | Employed Agency | Case Number | Issues |
|---|--------------------------------|-----------------------------|--|
| Missouri Gas Energy | Office of Public Counsel (OPC) | GR-2017-0216 | Rebuttal: Decoupling / Rate Design / Customer Confidentiality / Line Extension in Unserved and Underserved Areas / Economic Development Rider & Special Contracts |
| Laclede Gas Company | OPC | GR-2017-0215 | Rebuttal: Decoupling / Rate Design / Customer Confidentiality / Line Extension in Unserved and Underserved Areas / Economic Development Rider & Special Contracts |
| Indian Hills Utility | OPC | WR-2017-0259 | Direct: Rate Design |
| Rule Making | OPC | EW-2018-0078 | Comments on cogeneration and net metering |
| Empire District Electric Company | OPC | EO-2018-0048 | Integrated Resource Planning: Special Contemporary Topics Comments |
| Kansas City Power & Light | OPC | EO-2018-0046 | Integrated Resource Planning: Special Contemporary Topics Comments |
| KCP&L Greater Missouri Operations Company | OPC | EO-2018-0045 | Integrated Resource Planning: Special Contemporary Topics Comments |
| Missouri American Water | OPC | WU-2017-0296 | Direct: Lead line replacement pilot program Rebuttal: Lead line replacement pilot program Surrebuttal: Lead line replacement pilot program |
| KCP&L Greater Missouri Operations Company | OPC | EO-2017-0230 | Comments on Integrated Resource Plan, preferred plan update |
| Working Case: Emerging Issues in Utility Regulation | OPC | EW-2017-0245 | Comments on Emerging Issues in Utility Regulation / Presentation: Inclining Block Rate Design Considerations |
| Rule Making | OPC | EX-2016-0334 | Comments on Missouri Energy Efficiency Investment Act Rule Revisions |
| Great Plains Energy Incorporated, Kansas City Power & Light Company, KCP&L Greater Missouri Operations Company, | OPC | EE-2017-0113 / EM-2017-0226 | Direct: Employment within Missouri / Independent Third Party Management Audits / Corporate Social Responsibility |

| | | | |
|---|-----|--------------|--|
| and Westar Energy, Inc. | | | |
| Union Electric Company d/b/a Ameren Missouri | OPC | ET-2016-0246 | Rebuttal: EV Charging Station Policy Surrebuttal: EV Charging Station Policy |
| Kansas City Power & Light | | ER-2016-0156 | Direct: Consumer Disclaimer Direct: Response to Commission Directed Questions Rebuttal: Customer Experience / Greenwood Solar Facility / Dues and Donations / Electric Vehicle Charging Stations Rebuttal: Class Cost of Service / Rate Design Surrebuttal: Clean Charge Network / Economic Relief Pilot Program / EEI Dues / EPRI Dues |
| Union Electric Company d/b/a Ameren Missouri | OPC | ER-2016-0179 | Direct: Consumer Disclaimer / Transparent Billing Practices / MEEIA Low-Income Exemption Direct: Rate Design Rebuttal: Low-Income Programs / Advertising / EEI Dues Rebuttal: Grid-Access Charge / Inclining Block Rates /Economic Development Riders |
| KCP&L Greater Missouri Operations Company | OPC | ER-2016-0156 | Direct: Consumer Disclaimer Rebuttal: Regulatory Policy / Customer Experience / Historical & Projected Customer Usage / Rate Design / Low-Income Programs Surrebuttal: Rate Design / MEEIA Annualization / Customer Disclaimer / Greenwood Solar Facility / RESRAM / Low-Income Programs |
| Empire District Electric Company, Empire District Gas Company, Liberty Utilities (Central) Company, Liberty Sub-Corp. | OPC | EM-2016-0213 | Rebuttal: Response to Merger Impact Surrebuttal: Resource Portfolio / Transition Plan |
| Working Case: Polices to Improve Electric Regulation | OPC | EW-2016-0313 | Comments on Performance-Based and Formula Rate Design |
| Working Case: Electric Vehicle Charging Facilities | OPC | EW-2016-0123 | Comments on Policy Considerations of EV stations in rate base |
| Empire District Electric Company | OPC | ER-2016-0023 | Rebuttal: Rate Design, Demand-Side Management, Low-Income Weatherization |

| | | | |
|--|-----|--------------|--|
| | | | Surrebuttal: Demand-Side Management, Low-Income Weatherization, Monthly Bill Average |
| Missouri American Water | OPC | WR-2015-0301 | Direct: Consolidated Tariff Pricing / Rate Design Study Rebuttal: District Consolidation/Rate Design/Residential Usage/Decoupling Rebuttal: Demand-Side Management (DSM)/ Supply-Side Management (SSM) Surrebuttal: District Consolidation/Decoupling Mechanism/Residential Usage/SSM/DSM/Special Contracts |
| Working Case: Decoupling Mechanism | OPC | AW-2015-0282 | Memorandum: Response to Comments |
| Rule Making | OPC | EW-2015-0105 | Missouri Energy Efficiency Investment Act Rule Revisions, Comments |
| Union Electric Company d/b/a Ameren Missouri | OPC | EO-2015-0084 | Triennial Integrated Resource Planning Comments |
| Union Electric Company d/b/a Ameren Missouri | OPC | EO-2015-0055 | Rebuttal: Demand-Side Investment Mechanism / MEEIA Cycle II Application Surrebuttal: Potential Study / Overearnings / Program Design Supplemental Direct: Third-party mediator (Delphi Panel) / Performance Incentive Supplemental Rebuttal: Select Differences between Stipulations |
| The Empire District Electric Company | OPC | EO-2015-0042 | Integrated Resource Planning: Special Contemporary Topics Comments |
| KCP&L Greater Missouri Operations Company | OPC | EO-2015-0041 | Integrated Resource Planning: Special Contemporary Topics Comments |
| Kansas City Power & Light | OPC | EO-2015-0040 | Integrated Resource Planning: Special Contemporary Topics Comments |
| Union Electric Company d/b/a Ameren Missouri | OPC | EO-2015-0039 | Integrated Resource Planning: Special Contemporary Topics Comments |
| Union Electric Company d/b/a Ameren Missouri | OPC | EO-2015-0029 | Ameren MEEIA Cycle I Prudence Review Comments |
| Kansas City Power & Light | OPC | ER-2014-0370 | Direct (Revenue Requirement): Solar Rebates Rebuttal: Rate Design / Low-Income Weatherization / Solar Rebates |

| | | | |
|--|--|--------------|--|
| | | | Surrebuttal: Economic Considerations / Rate Design / Cyber Security Tracker |
| Rule Making | OPC | EX-2014-0352 | Net Metering and Renewable Energy Standard Rule Revisions, Comments |
| The Empire District Electric Company | OPC | ER-2014-0351 | Rebuttal: Rate Design/Energy Efficiency and Low-Income Considerations |
| Rule Making | OPC | AW-2014-0329 | Utility Pay Stations and Loan Companies, Rule Drafting, Comments |
| Union Electric Company d/b/a Ameren Missouri | OPC | ER-2014-0258 | Direct: Rate Design/Cost of Service Study/Economic Development Rider Rebuttal: Rate Design/ Cost of Service/ Low Income Considerations Surrebuttal: Rate Design/ Cost-of-Service/ Economic Development Rider |
| KCP&L Greater Missouri Operations Company | OPC | EO-2014-0189 | Rebuttal: Sufficiency of Filing Surrebuttal: Sufficiency of Filing |
| KCP&L Greater Missouri Operations Company | OPC | EO-2014-0151 | Renewable Energy Standard Rate Adjustment Mechanism (RESRAM) Comments |
| Liberty Natural Gas | OPC | GR-2014-0152 | Surrebuttal: Energy Efficiency |
| Summit Natural Gas | OPC | GR-2014-0086 | Rebuttal: Energy Efficiency Surrebuttal: Energy Efficiency |
| Union Electric Company d/b/a Ameren Missouri | OPC | ER-2012-0142 | Direct: PY2013 EM&V results / Rebound Effect Rebuttal: PY2013 EM&V results Surrebuttal: PY2013 EM&V results Direct: Cycle I Performance Incentive Rebuttal: Cycle I Performance Incentive |
| Kansas City Power & Light | Missouri Public Service Commission Staff | EO-2014-0095 | Rebuttal: MEEIA Cycle I Application testimony adopted |
| KCP&L Greater Missouri Operations Company | Missouri Division of Energy (DE) | EO-2014-0065 | Integrated Resource Planning: Special Contemporary Topics Comments |
| Kansas City Power & Light | DE | EO-2014-0064 | Integrated Resource Planning: Special Contemporary Topics Comments |
| The Empire District Electric Company | DE | EO-2014-0063 | Integrated Resource Planning: Special Contemporary Topics Comments |
| Union Electric Company d/b/a Ameren Missouri | DE | EO-2014-0062 | Integrated Resource Planning: Special Contemporary Topics Comments |
| The Empire District Electric Company | DE | EO-2013-0547 | Triennial Integrated Resource Planning Comments |

| | | | |
|---|-----|-------------------------------|--|
| Working Case: State-Wide Advisory Collaborative | OPC | EW-2013-0519 | Presentation: Does Better Information Lead to Better Choices? Evidence from Energy-Efficiency Labels |
| Independence-Missouri | OPC | Indy Energy Forum 2014 | Presentation: Energy Efficiency |
| Independence-Missouri | OPC | Indy Energy Forum2015 | Presentation: Rate Design |
| NARUC – 2017 Winter | OPC | Committee on Consumer Affairs | NARUC – 2017 Winter Presentation: PAYS Tariff On-Bill Financing |
| NASUCA – 2017 Summer | OPC | Committee on Water Regulation | NASUCA – 2017 Summer Presentation: Regulatory Issues Related to Lead-Line Replacement of Water Systems |



U.S. DEPARTMENT OF
ENERGY

PNNL-20235

Prepared for the U.S. Department of Energy
under Contract DE-AC05-76RL01830

Analyzing the Impact of Residential Building Attributes, Demographic and Behavioral Factors on Natural Gas Usage

OV Livingston
KA Cort

March 2011



Pacific Northwest
NATIONAL LABORATORY

*Proudly Operated by **Battelle** Since 1965*

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor Battelle Memorial Institute, nor any of their employees, makes **any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights.** Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or Battelle Memorial Institute. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

PACIFIC NORTHWEST NATIONAL LABORATORY
operated by
BATTELLE
for the
UNITED STATES DEPARTMENT OF ENERGY
under Contract DE-AC05-76RL01830

Printed in the United States of America

Available to DOE and DOE contractors from the
Office of Scientific and Technical Information,
P.O. Box 62, Oak Ridge, TN 37831-0062;
ph: (865) 576-8401
fax: (865) 576-5728
email: reports@adonis.osti.gov

Available to the public from the National Technical Information Service
5301 Shawnee Rd., Alexandria, VA 22312
ph: (800) 553-NTIS (6847)
email: orders@ntis.gov <<http://www.ntis.gov/about/form.aspx>>
Online ordering: <http://www.ntis.gov>



This document was printed on recycled paper.

(8/2010)

Analyzing the Impact of Residential Building Attributes, Demographic and Behavioral Factors on Natural Gas Usage

OV Livingston
KA Cort

March 2011

Prepared for
the U.S. Department of Energy
under Contract DE-AC05-76RL01830

Pacific Northwest National Laboratory
Richland, Washington 99352

Summary

This analysis examines the relationship between energy demand and residential building attributes, demographic characteristics, and behavioral variables using the U.S. Department of Energy's Residential Energy Consumption Survey 2005 microdata. This study investigates the applicability of the smooth backfitting estimator to statistical analysis of residential energy consumption via nonparametric regression. The methodology utilized in the study extends nonparametric additive regression via local linear smooth backfitting to categorical variables.

The conventional methods used for analyzing residential energy consumption are econometric modeling and engineering simulations. This study suggests an econometric approach that can be utilized in combination with simulation results. A common weakness of previously used econometric models is a very high likelihood that any suggested parametric relationships will be misspecified. Nonparametric modeling does not have this drawback. Its flexibility allows for uncovering more complex relationships between energy use and the explanatory variables than can possibly be achieved by parametric models.

Traditionally, building simulation models overestimated the effects of energy efficiency measures when compared to actual "as-built" observed savings. While focusing on technical efficiency, they do not account for behavioral or market effects. The magnitude of behavioral or market effects may have a substantial influence on the final energy savings resulting from implementation of various energy conservation measures and programs. Moreover, variability in behavioral aspects and user characteristics appears to have a significant impact on total energy consumption. Inaccurate estimates of energy consumption and potential savings also impact investment decisions. The existing modeling literature, whether it relies on parametric specifications or engineering simulation, does not accommodate inclusion of a behavioral component. This study attempts to bridge that gap by analyzing behavioral data and investigate the applicability of additive nonparametric regression to this task.

This study evaluates the impact of 31 regressors on residential natural gas usage. The regressors include weather, economic variables, demographic and behavioral characteristics, and building attributes related to energy use. In general, most of the regression results were in line with previous engineering and economic studies in this area. There were, however, some counterintuitive results, particularly with regard to thermostat controls and behaviors. There are a number of possible reasons for these counterintuitive results including the inability to control for regional climate variability due to the data sanitization (to prevent identification of respondents), inaccurate data caused by self-reporting, and the fact that not all relevant behavioral variables were included in the data set, so we were not able to control for them in the study.

The results of this analysis could be used as an in-sample prediction for approximating energy demand of a residential building whose characteristics are described by the regressors in this analysis, but a certain combination of their particular values does not exist in the real world. In addition, this study has potential applications for benefit-cost analysis of residential upgrades and retrofits under a fixed budget, because the results of this study contain information on how natural gas consumption might change once a particular characteristic or attribute is altered. Finally, the results of this study can help establish a relationship between natural gas consumption and changes in behavior of occupants.

Acronyms and Abbreviations

| | |
|------|---------------------------------------|
| DOE | U.S. Department of Energy |
| HDD | Heating Degree Days |
| CDD | Cooling Degree Days |
| EIA | Energy Information Administration |
| MBtu | million British thermal units |
| NG | natural gas |
| ORC | Opinion Research Corporation |
| R&D | research and development |
| RECS | residential energy consumption survey |
| SBE | smooth backfitting estimation |

Contents

| | |
|---|------|
| Summary | iii |
| Acronyms and Abbreviations | v |
| 1.0 Introduction | 1.1 |
| 2.0 Methodology..... | 2.1 |
| 2.1 Data and Analysis..... | 2.1 |
| 3.0 Results by Attribute | 3.1 |
| 3.1 Weather | 3.2 |
| 3.2 Fuel and Equipment/Appliance Choice..... | 3.2 |
| 3.3 Controls and Thermostat Settings | 3.6 |
| 3.4 Prices and Billing Structure..... | 3.9 |
| 3.5 Home Construction Attributes | 3.10 |
| 3.5.1 Building Shell..... | 3.10 |
| 3.5.2 Size and Design..... | 3.12 |
| 3.6 Vintage | 3.15 |
| 3.7 Home Ownership | 3.16 |
| 3.8 Occupancy..... | 3.17 |
| 3.9 Income..... | 3.18 |
| 4.0 Conclusions | 4.1 |
| 5.0 References | 5.1 |
| Appendix A – Detailed Methodology | A.1 |
| Appendix B – Complete Set of Graphical Results..... | B.1 |

Figures

| | | |
|------|--|------|
| 3.1 | Impact of House Size on NG Use | 3.1 |
| 3.2 | Impact of Heating Degree Days on NG Use | 3.2 |
| 3.3 | Impact of Main Heating Fuel Choice on NG Use | 3.3 |
| 3.4 | Impact of Heating Equipment Choice on NG Use | 3.4 |
| 3.5 | Secondary Heating Equipment Impact on NG | 3.5 |
| 3.6 | Setting During the Winter Day When Someone is Home | 3.6 |
| 3.7 | Setting During the Winter Day When No One is Home | 3.7 |
| 3.8 | Impact of Thermostat and Programmable Thermostat on NG Use | 3.8 |
| 3.9 | Impact of Number of Thermostats on NG Use | 3.9 |
| 3.10 | Impact of How NG is Paid on NG Use | 3.10 |
| 3.11 | Secondary Heating Equipment Impact on NG | 3.11 |
| 3.12 | Impact of Type of Window Glass on NG Use | 3.12 |
| 3.13 | Impact of Heating Garage on NG Use | 3.13 |
| 3.14 | Impact of Basement/crawl space on NG Use | 3.14 |
| 3.15 | Impact of Number of Rooms Not Heated on NG Use | 3.15 |
| 3.16 | Impact of House Age on NG Usage | 3.16 |
| 3.17 | Impact of Ownership/Rental Status on NG Use | 3.17 |
| 3.18 | Impact of Number of Occupants on NG Use | 3.18 |
| 3.19 | Impact of Income on NG Use | 3.19 |

Tables

| | | |
|-----|-------------------------|-----|
| 2.1 | List of Variables | 2.3 |
|-----|-------------------------|-----|

1.0 Introduction

There are three main approaches to residential energy demand analysis: engineering, socio-psychological and econometric. The engineering approach relies on simulating different types of building energy use within an engineering modeling framework such as Energy Plus, DOE-2 and the like (Crawley et al. 2004). These building energy simulation tools construct demand projections by performing hourly energy simulations of buildings, air-handling systems, and equipment based on building and weather characteristics and an assumed operation schedule. The second approach evaluates the impact of institutions, beliefs and group influences on the long-term trends in energy use. The econometric approach links energy use to prices of energy products and their substitutes, as well as household income, demographic characteristics and features of the occupied buildings. This study fits into the third category, exploring the behavioral data on energy consumption at the micro level.

Detailed studies of energy use at the household level using microeconomic data were conducted by Baker et al. (1989), Schmalensee and Stoker (1999), Halvorsen and Larsen (2001), Yatchew and No (2001), Nesbakken (2001), Larsen and Nesbakken (2004), Garcia-Cerruti (2000), Hortedahl and Joutz (2004), Kamerschen and Porter (2004) and Narayan and Smyth (2005) to name a few. The reviewed econometric studies all estimate energy demand functions; however, the explanatory variables employed by these studies differ. These studies can generally be categorized into two groups. The first group includes economic variables such as fuel prices and income level, as well as climate information. The second group of studies incorporates additional household and demographic characteristics of the dwelling into the model. An extensive overview of econometric analysis of residential energy demand predating the above-listed research is included in Madlener (1996).

The focus of this analysis is residential natural gas (NG) demand. Space heating is the single largest end use of energy in residential buildings, and furnaces fueled by natural gas are the primary source of residential heating. Natural gas also provides fuel for residential water heating, cooking, clothes drying, and other miscellaneous uses. In terms of on-site energy use measured in British thermal units (Btu), in 2006 the Energy Information Administration (EIA) estimated that natural gas supplied approximately 65% of 4.4 quadrillion Btu delivered for residential space heating, and approximately 68% of total residential site energy for water heating (DOE/EIA 2009). The primary substitute for natural gas in residential homes is electricity (i.e., electric furnaces, heat pumps, electric water heaters, etc.).

The majority of econometric research on electricity and natural gas consumption relies on a fully specified parametric functional relationship between energy use and its conditioning variables. As a result, there is the potential for severe misspecification of the proposed econometric models. Also, the categorical variables, which are typically present in residential microdata, are usually treated either by including dummy variables or via sub-sample regression. Nonparametric modeling is robust to functional form misspecification. Its flexibility allows for uncovering more complex relationships between energy use and conditioning variables than can be possibly achieved by parametric models.

In this study we adopt additive nonparametric modeling for energy consumption, which would be estimated using the smooth backfitting procedure of Mammen et al. (1999). This procedure achieves convergence rates equal to this of univariate models thus bypassing the curse of dimensionality. In addition, recognizing that both continuous and categorical variables impact energy demand, this

application of backfitting procedure incorporates the kernel smoothing methods of Racine and Li (2003) and Racine et al. (2004) for categorical variables.

The data for this research comes from the Residential Energy Consumption Survey (RECS) designed by the U.S. Department of Energy's Energy Information Administration. The microdata obtained from the 2005 survey covers energy consumption for several major fuel types and includes information on household characteristics, standard demographics, dwelling characteristics, as well as information about televisions and other media devices, personal computers and peripherals, Energy Star labeling, energy efficient lighting, window glazing, window replacement, and thermostat usage. The 2005 survey also incorporates questions on behavioral aspects of energy use. This analysis contributes to existing literature by analyzing and quantifying behavioral impacts on residential energy consumption.

The study is organized into three sections. A brief description of the smooth backfitting approach is presented in Section 2. Section 3 describes the results of the empirical analysis. Section 4 provides the conclusions of this analysis. The local linear smooth backfitting estimator (SBE) for continuous and mixed variables is described in more detail in Appendix A. Appendix B contains a complete set of result charts.

2.0 Methodology

This study investigates the applicability of the smooth backfitting estimator to statistical analysis of residential energy consumption via nonparametric regression. The nonparametric modeling does not require an analyst to assume any particular functional relationship between the energy consumption and analyzed variables. This is one of the advantages that nonparametric approach has over traditionally used parametric models. The quality of any parametric results directly depends on how close the assumed functional form is to the true relationship. Household energy usage depends on a complicated set of variables whose impact is not fully understood or separated.

The model used here is a special case of a very broad class of generalized additive models, which are gaining significant attention in the current econometric literature. The utilized methodology extends nonparametric additive regression via local linear SBE to categorical variables, which are, in this case, attributes of the residential building and demographic characteristics of its occupants.

The smooth backfitting estimator is a projection of the data on the space of additive functions. Projection here is taken with respect to the norm defined by the local polynomial kernel estimator. This particular definition of the estimator allows separating effects (i.e., the effect of natural gas prices versus the effect of exterior wall construction, etc.) within complicated multidimensional problems into one-dimensional effects. Also the number of controlled variables that can be meaningfully utilized in the parametric modeling is usually limited. SBE method is capable of successfully accommodating a large number of explanatory variables. Nielsen and Spierlich (2005) demonstrated that the SBE method produces better results in “extreme cases of complexity and data sparseness” by comparing performance in finite samples on a model with 100 correlated variables. The SBE methodology of Mammen et al. (1999) and computational algorithm outlined by Nielsen and Spierlich (2005) are described in detail in Appendix A of this report.

2.1 Data and Analysis

The data for this research comes from the RECS survey designed by DOE-EIA. The microdata obtained from the 2005 survey covers energy consumption for several major fuel types and includes information on household characteristics, standard demographics, dwelling characteristics, as well as information about televisions and other media devices, personal computers and peripherals, Energy Star labeling, energy efficient lighting, window glazing, window replacement, and thermostat usage. The 2005 survey also incorporates questions on behavioral aspects of energy use. This analysis contributes to existing literature by analyzing and quantifying impacts of demographic and behavioral variables on residential NG consumption.

Upon close examination of the RECS questions and microdata for 2005, it became apparent that it would be an extremely complex task to cover all the end fuel uses for all fuel types included in the survey. The decision was made to investigate the applicability of smooth backfitting by isolating natural gas usage and related variables. RECS data was filtered out to include only households using natural gas, resulting in a subset of 1388 observations. For 1053 of these observations natural gas consumption data came directly from the provider company records. The regressand is natural gas usage in millions of

British thermal units (MBtu). There are 31 regressors,¹ which include demographic and behavioral characteristics, as well as building attributes related to energy consumption. The regressors enter the model additively in the following way:

$$E(Y|X_1=x_1, \dots, X_d=x_d) = m_0 + \sum_{j=1}^d m_j(x_j)$$

where $E(Y|X_1=x_1, \dots, X_d=x_d)$ = conditional mean of natural gas energy consumption
 x_j = regional/residential home attributes, behavioral and demographic characteristics,
 m_0 = unknown scalar parameter,
 $m_j(x_j)$ = unknown function of x_j for all $j=1, \dots, d$,

Out of the 31 regressors, 8 are continuous variables, 14 are unordered categorical variables, and the remaining 9 are ordered categorical variables. These are described in Table 2.1.

Individual cross-validated bandwidth values were computed for each regressor. Although unordered categorical regressors have the potential to violate the mean-zero assumption for each direction to meet the identification conditions as part of the smooth backfitting algorithm, the results of these regressions are reasonable. The results of the ordered categorical regressions suggest that at least some of them could have been treated as continuous variables. Several directional regressions show rather smooth change, which may be suggestive of the particular type of a parametric relationship. Specific results are discussed in Section 3 of this report.

¹ Initially the model was to include 44 categorical variables, but cross-validation produced the bandwidth values equal to the upper bound of $(c_i-1)/c_i$ for 13 of the categorical variables. When the bandwidth takes this upper value, it implies that the regressor is irrelevant and, if included, it will effectively be smoothed out.

Table 2.1. List of Variables

| Regressor | Description | Regressor Code |
|---|--|------------------------------|
| Continuous | Unit of Measurement | (Chart Label) |
| Heating degree days | Degrees Fahrenheit (sanitized) | Direction 1 |
| Cooling degree days | Degrees Fahrenheit (sanitized) | Direction 2 |
| Total house area | Square feet | Direction 3 |
| Price of electricity | Cents/kWh | Direction 4 |
| Price of natural gas | Cents/kBtu | Direction 5 |
| Thermostat setting: Occupied | Degrees Fahrenheit | Direction 6 |
| Thermostat setting: Unoccupied | Degrees Fahrenheit | Direction 7 |
| Thermostat setting: Sleeping | Degrees Fahrenheit | Direction 8 |
| Exterior wall construction | Indescribable, brick, wood, siding, stucco, composition, stone, concrete, glass, other | Direction 9 |
| Garage | No garage, garage not heated, yes garage heated | Direction 10 |
| Ownership status | Owned, rented, occupied without payment | Direction 11 |
| Cooking fuel | Natural gas, propane, electricity, some other fuel | Direction 12 |
| Clothes dryer fuel | Natural gas, propane, electricity, no dryer | Direction 13 |
| Secondary heating equipment | No secondary heating, furnace, radiant (water), built-in floor, built-in room heater, cooking stove | Direction 14 |
| Programmable Thermostat | Not programmable, yes programmable, no thermostat | Direction 15 |
| Programmable Thermostat set-back: night | Not set-back at night, yes set-back at night, no thermostat or not programmable | Direction 16 |
| Programmable Thermostat set-back: day | Not set-back during day, yes set-back during day, no thermostat or not programmable | Direction 17 |
| Main heating fuel | Propane, natural gas, fuel oil, kerosene, electricity, wood, solar | Direction 18 |
| Heating equipment | No heating equipment, radiant (water), heat pump, central furnace, built-in electric wall, built-in floor, built-in room heater (gas, oil, kerosene), wood stove, fireplace, portable electric heaters, portable kerosene heaters, cooking stove | Direction 19 |
| Water heating fuel | Yes natural gas, do not use natural gas | Direction 20 |
| Billing | Household pays all, included in rent, some paid and some included in rent, other | Direction 21 |
| Occupancy | Not occupied typically during day/weekday, typically occupied during day/weekday | Direction 22 |
| Ordered Categorical | Categories | Regressor Code (Chart Label) |
| Number of stories | One story, two stories, three stories, four or more stories, split level, other | Direction 23 |
| Basement/crawlspace heat | No basement, not heated, part heated, all heated | Direction 24 |
| Attic heat | No attic, not heated, partially heated, all heated | Direction 25 |
| Home vintage | Before 1940, 1940-49, 1950-59, 1960-69, 1970-79, 1980-89, 1990-99, 2000-02, 2003, 2004, 2005 | Direction 26 |
| Number of thermostats | Actual number (e.g., 0, 1, 2 . . .) | Direction 27 |
| Number of rooms not heated | Actual number (e.g., 0, 1, 2 . . .) | Direction 28 |
| Type of window glass | Single-pane, double-pane, double pane with low-e, triple-pane glass, triple-pane with low-e | Direction 29 |
| Occupants | Actual number (e.g., 0, 1, 2 . . .) | Direction 30 |
| Income | 5k groupings from 0 to \$120,000 or more | Direction 31 |

3.0 Results by Attribute

The results of this study are presented graphically throughout the Section 3 and also in Appendix B of this report. In all cases, the vertical axes on the graph show changes in natural gas consumption in million British thermal units (MBtu). The horizontal axes represent a characteristic, attribute or a variable of interest specified below the graph. Each graph shows an effect of changes in the variable of interest on natural gas consumption, holding all other variables in the model fixed. Throughout the paper the results are referred to as directional regression results (or Direction 1, 2, 3, etc) because SBE assumes additive separability, thus, we are considering impact of changes only in one direction [dimension] at a time. Note, that vertical axes do not show absolute level of consumption, but represent the magnitude of deviation from the mean. For example, Figure 3.1 illustrates relationship between total square footage of the house and natural gas consumption. Zero on the vertical axes stands for the mean NG consumption of 77.5 MBtu, which corresponds to the house size of approximately 2700 square feet (s.f.). If we consider two identical houses (identical in the sense that all factors that we control for in the model are equal), where one is 2,000 s.f. and the other one is 4,000 s.f., the difference between NG consumption of those is almost 9 MBtu.

It should be noted that nonparametric methods produce estimates of a function at every data point instead of a functional form itself as it is done in parametric estimation. Therefore, the results for continuous variables are presented as vectors of the same size as data, while bar charts are used to depict impact of the categorical variables, where each bar corresponds to a distinct category.

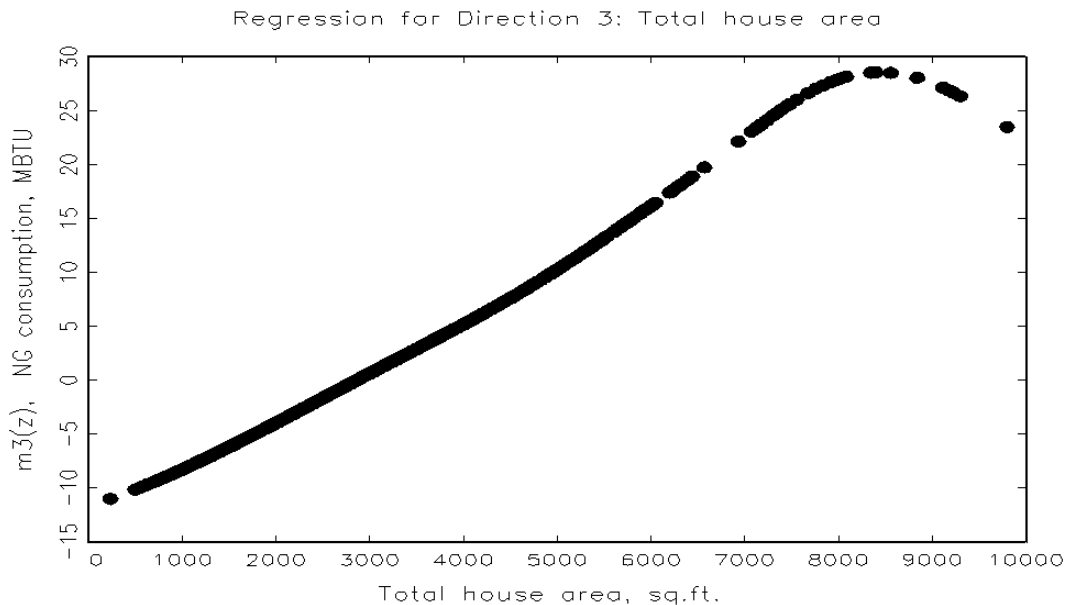


Figure 3.1. Impact of House Size on NG Use

3.1 Weather

Direction 1, heating degree days, seems to correctly represent the increase in natural gas intensity as the number of heating degree days goes up (see Figure 3.2). Heating degree days are a characterization of weather. It is worth noting that RECS microdataset has sanitized data for heating and cooling degree days to prevent identification of survey respondents or specific buildings out of the reported sample. Even with the sanitized data, the overall pattern of dependency is reasonable. Annual heating degree days (HDD) are a measure of how cold a building location is relative to the base temperature. The daily HDD is the numerical difference between a day's average temperature and 65°F, if the average temperature is less than 65°F. Otherwise it is zero. Annual HDD is the sum of the daily HDD for the year. If the thermal integrity (e.g., insulation levels) of the building is known, it is possible to assess heating requirements from this information. The suggested pattern follows the engineering results that building heating requirements are not linear with respect to temperature. Therefore, natural gas use for heating will also have non-linear dependency on temperature. Although this pattern of dependency is well-known from engineering studies, the primary reason for including this variable is to analyze the impact of other factors on energy demand, while controlling for weather. Cooling degree days also contains sanitized data, and their impact is shown graphically as Direction 2 in Appendix B. The cooling degree day pattern of dependency observed is consistent with engineering studies and suggests a non-linear decrease in natural gas usage as the number of cooling degree days goes up.

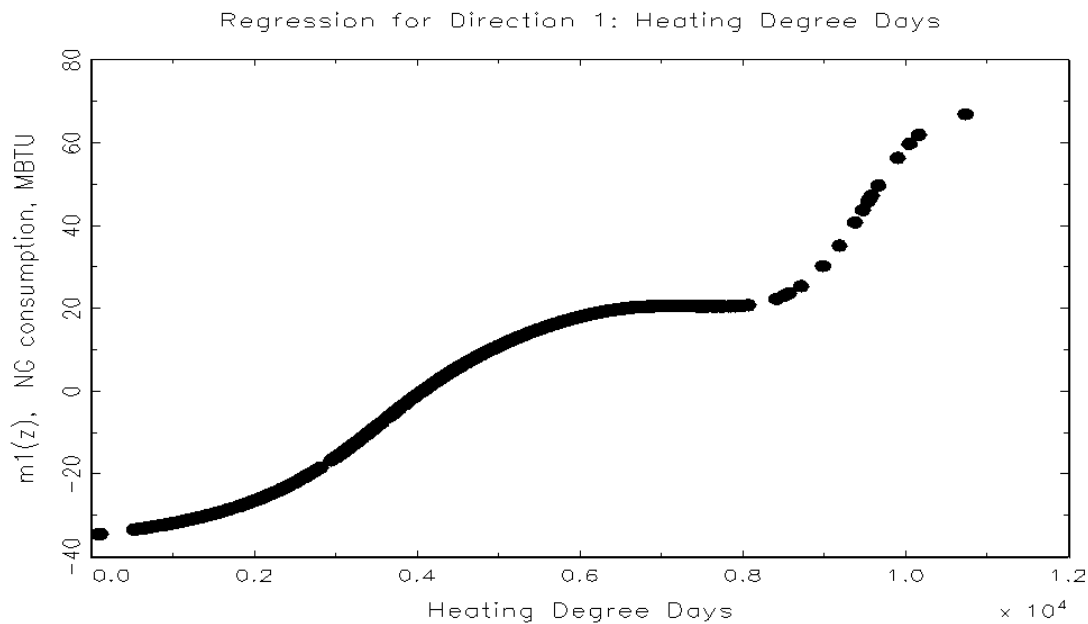


Figure 3.2. Impact of Heating Degree Days on NG Use

3.2 Fuel and Equipment/Appliance Choice

Figure 3.3 presents results related to primary heating systems and the choice of fuels (Direction 18). As expected, NG as primary heating fuel (category 1) would result in the highest NG intensity. If the heating degree days data were not sanitized, it would have been possible to approximately identify the climate zone associated with a particular set of observations. There is a dependency between the climate

zone and choice of fuel for heating that could impact this result. The lowest NG usage is for the houses heated with kerosene or fuel oil. Natural gas consumption for houses that use electricity as primary fuel goes up by 15 MBtu. This could be explained by the fact that some houses with piped natural gas available use electric-source equipment as their primary heating system. The latter use NG for auxiliary heat. Therefore, in this particular case, NG would be used complimentary to electricity. A similar explanation is valid for increase in NG use by 10 MBtu for dwellings using wood and solar energy as a primary heating fuel. Although these are categorical variables, dotted lines connecting results are added on the graphic presented in Figure 3.3 as a visual aid.

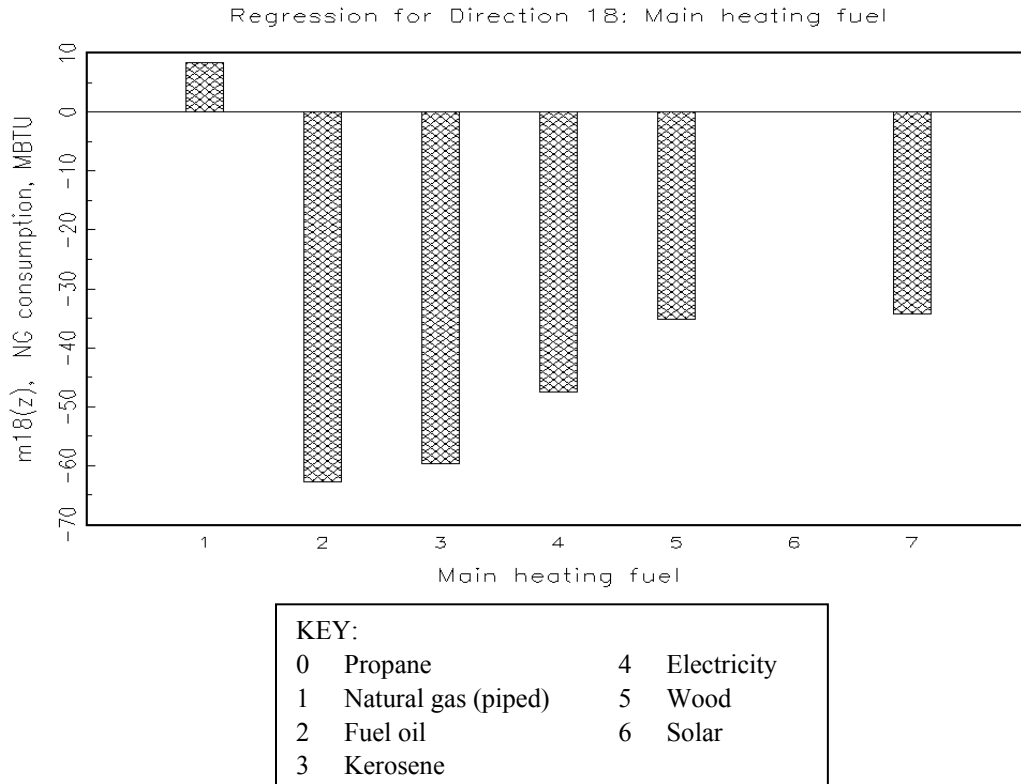
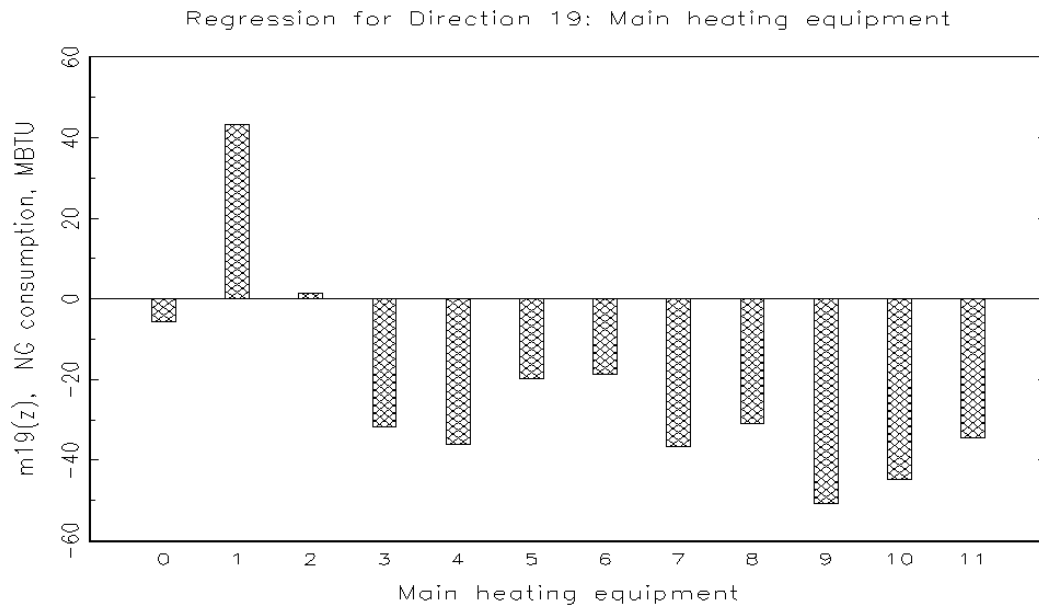


Figure 3.3. Impact of Main Heating Fuel Choice on NG Use

Correlation between the type of heating equipment providing the heat and NG usage is depicted on the graph in Figure 3.4 (Direction 19). The lowest NG usage is suggested where portable electric heaters are used to provide most of the heat (category 9). If the heating load can be met with the portable electric heaters, this would indicate that only very little heating is needed and piped NG is used for water heating and cooking only. Similar explanation is valid for heating stoves burning wood (category 7), portable kerosene heaters (category 10) and cooking stoves used for heating (category 11). The suggestion of highest NG consumption being characteristic of houses with steam/hot water system and radiators/convectors in each room (category 1) is reasonable. High level of NG consumption shown in the graph is expected because this heating system choice impacts natural gas intensity through water-heating requirements, but it is also a manifestation of the climate zone and age/vintage of the house. NG consumption decreases for houses where heat pump is used as a primary equipment, but it is still higher than any other category. This result can also be explained by complimentary use of NG for the auxiliary system that usually turns on as temperatures fall below freezing, as the electric heat pump becomes less

efficient at these colder temperatures. Relatively low NG consumption, according to the regression results, is associated with using central warm-air furnace system with ducts to individual rooms. Considering that this is one of the more efficient heating distribution systems, this is an expected result. Properly designed duct systems have a significant impact on how much heat is lost during delivery. The newest houses have ducts located in the air-conditioned and heated spaces, which results in even more efficient distribution of heat, thus reducing NG intensity. In addition, this is a manifestation of multicollinearity between the house age, quality of construction/insulation and income level of the household.



| KEY: | |
|------|--|
| 0 | No heating equipment used |
| 1 | Steam/hot water system with radiators/convectors in each room or pipes in floors/walls |
| 2 | Heat pump |
| 3 | Central warm-air furnace with ducts |
| 4 | Built-in electric units in room |
| 5 | Built-in floor/wall pipeless furnace |
| 6 | Built-in room heater burning gas/oil |
| 7 | Heating stove burning wood/coal |
| 8 | Fireplace |
| 9 | Portable electric heater |
| 10 | Portable kerosene heater |
| 11 | Cooking stove used for heating |

Figure 3.4. Impact of Heating Equipment Choice on NG Use

Results for Direction 20 represents the type of fuel used to heat water for washing or bathing and are presented in Appendix B. As expected, if the primary water heating fuel is NG, its consumption is higher than for other fuels. The overall difference is 24 MBtu.

Figure 3.5 (Direction 14) shows the dependency between the NG use and the type of secondary heating equipment installed in the house. Typical secondary heating equipment includes central warm-air furnace with ducts (category 1), steam/hot water system with radiators/convectors in each room or pipes in the floor or walls (category 2), built-in floor/wall pipeless furnace (category 3), built-in room heater (category 4) and wood cooking stove used to heat the house (category 5). Cases of no secondary equipment are included as a category with value 0. The result for this category is intuitive because the households with no secondary equipment will have all the heating load provided by the main equipment.

Because the RECS microdataset was filtered to keep only observations with piped natural gas, the result that houses equipped with natural gas intake are more likely to use natural gas as their primary heating fuel is also intuitive. Central warm-air furnace with ducts implies a more efficient heat delivery system; therefore, reduction of the NG consumption for category 1 is also an expected result. The resulting increase in NG consumption that occurs when the secondary heat as built-in room heaters (option 4) is unexpected; however, it is possible that this result is correlated with thermal integrity of the dwelling, because built-in room heaters are more typical for older houses with lower insulation and construction quality.

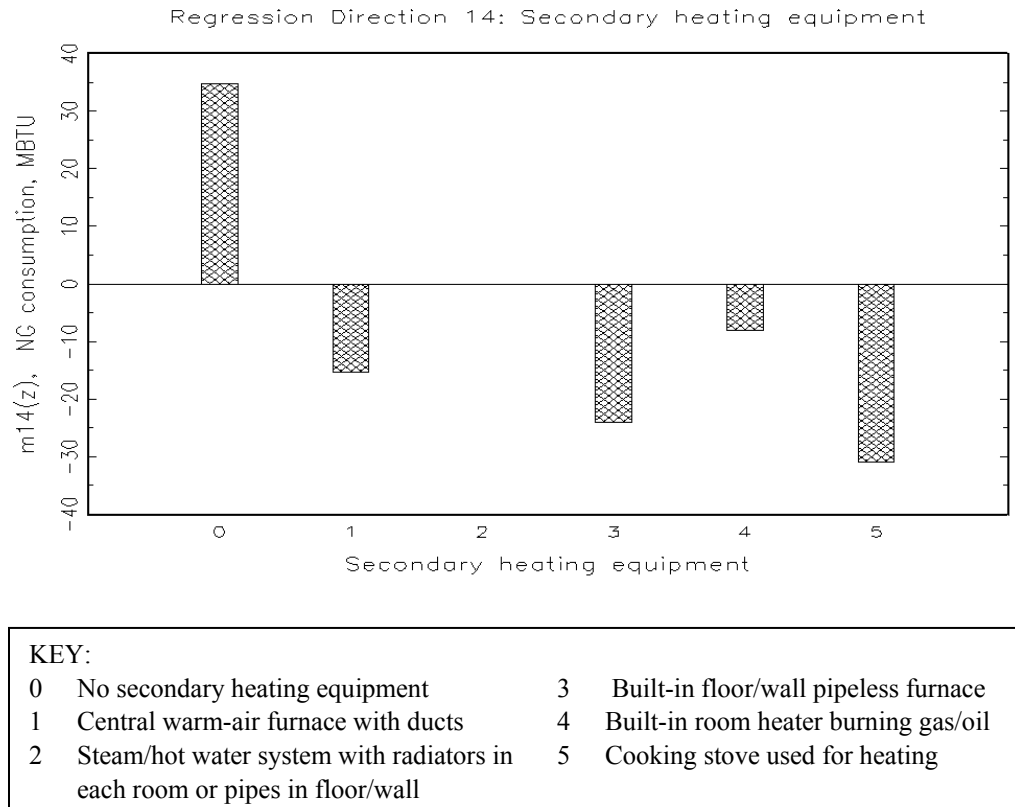


Figure 3.5. Secondary Heating Equipment Impact on NG

The results related to the impact of fuel choice for stovetops (Direction 12) and clothing dryers (Direction 13) are presented in Appendix B. Direction 12 shows the pattern of association between the NG intensity and type of fuel used by burners for cooking on the stove. The peak value is observed for the household equipped with piped natural gas for cooking. There is no difference between using some other fuel (category 0) and bottled propane (category 2). On one hand, these two categories could be combined. However, residents usually refer to both types of fuel (propane and natural gas) generally as gas, so it is worth keeping for clarification. There is a 4 MBtu reduction if the household is using electricity for cooking burners, which is a reasonable result. This result can also be partially attributed to multicollinearity in data, namely if the household has piped natural gas, it is expected that burners would use NG, but so would the water heaters, clothes dryer and potentially other systems. The results related to clothing dryer fuel choice also suggest multicollinearity in the data, where households without dryers are more typical for older neighborhoods with lower construction quality and, therefore, lower thermal integrity.

3.3 Controls and Thermostat Settings

A number of counterintuitive results were observed related to thermostat controls and setting impacts on natural gas usage. Although the reasons for these results are unclear, it is possible that data reporting problems from self-reported data, as well as some unexplained behavioral characteristics, may be the root cause of these results. For example, Direction 6 contains data on the temperature setting during the day in winter when someone is home. Natural gas intensity in this direction seems to misrepresent the direction of dependency. As shown in Figure 3.6, the mean of regressor 6 (option 6 in key) corresponds to the temperature setting of 70°F. While there is a positive correlation between temperature setting and energy consumption for the range between 55°F and 65°F, there is no reasonable explanation why natural gas consumption drops for the ranges from 65°F to 80°F, when the opposite should be observed.

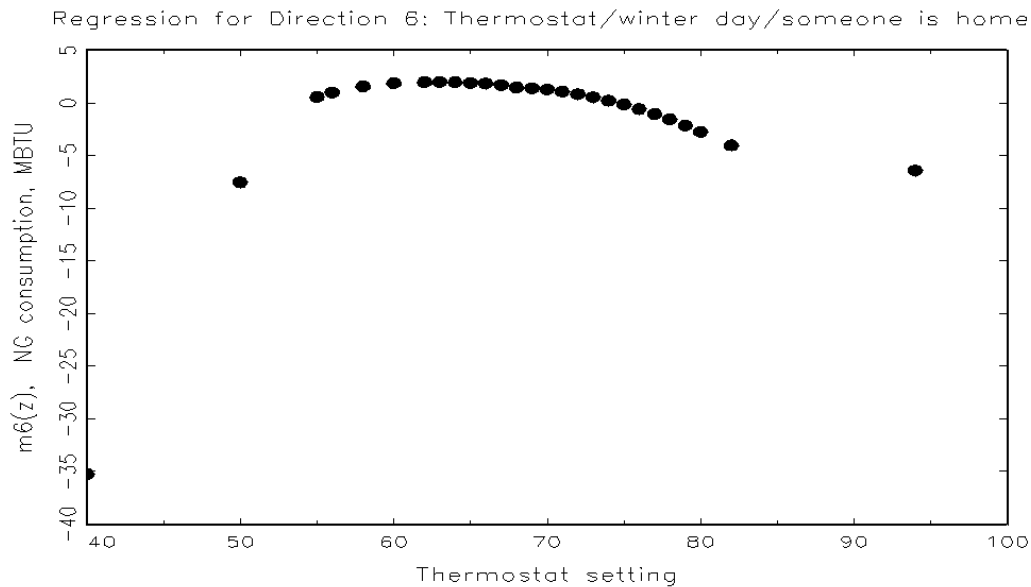


Figure 3.6. Setting During the Winter Day When Someone is Home

The same can be said about the Direction 7, which represents the temperature setting during the day in winter when no one is home and is shown in Figure 3.7. The mean for this regressor is 65°F. The base temperature for heating is 65°F, so thermostats set to the mean temperature would mean no additional heating is required on a 0 HDD. Thus, it is not clear why Direction 7 would indicate a drop in the natural gas consumption while the temperature setting is going up. It might be beneficial to replace these two variables with one that would represent the difference between temperature setting when someone is home and temperature setting when someone is not home. The higher the delta, the less energy is consumed while the building is not occupied. There is also an additional factor that leads to misrepresentation of the relationship for this covariate. All temperature settings are self-reported. In fact, studies have found that persons often report lower-than-actual thermostat settings, even when they know that their settings are being recorded, as shown by Lutzenhiser (1993). No actual readings of the thermostat are taken. As saving energy becomes a more widely-publicized topic, respondents understate heating temperature settings, as well as misreport the way programmable thermostats are used, to fall within the range they perceive as socially acceptable. On the other hand, data on natural gas consumption comes directly from the bill and reflects actual consumption levels. Therefore, even restructuring the variable may not produce a desirable result using existing data.

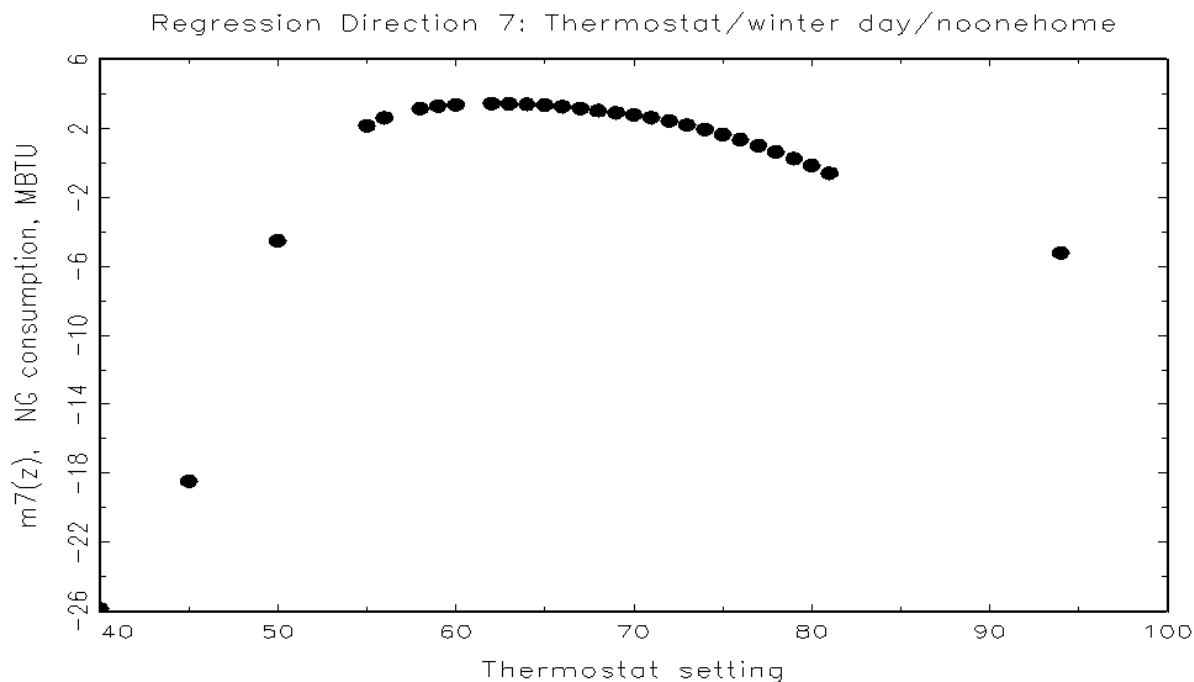


Figure 3.7. Setting During the Winter Day When No One is Home

Direction 8 represents association between the level of natural gas intensity and temperature during the sleeping hours in winter and is presented in Appendix B. As the setting goes up from 50F to 70F so does the NG consumption. The slight drop in the gas usage around that point is unexpected. The concern with temperature setting being self-reported is pertinent here as well, because the owners tend to misreport lowering the thermostat settings. So the houses that are set at much higher temperatures, but underreport to be closer in line with culturally-accepted 65-70°F level, will drive the result for this average level much higher than what it should be. The estimated natural gas consumption will be inflated for the misreported temperature and underestimated for the higher temperature intervals that would otherwise correspond to that actual heating requirement. This makes the results to the right of the anchor level appear lower than at the average setting, thus erroneously suggesting negative correlation over this interval of temperatures.

Figure 3.8 (Direction 15) describes the relationship between NG consumption and the controls installed in the house. There seems to be no difference in NG consumption if there is a programmable (category 1) or non-programmable (category 0) thermostat in the house. These two categories are associated with increased NG demand. The result for category 3 is counterintuitive because it suggests that absence of thermostats is characterized by a significant reduction in NG consumption. Both the direction of change and the magnitude of 16 MBtu are counterintuitive. The explanation might be that absence of thermostat is dictated by a warm climate zone and is an indicator of a non-heated dwelling or very little heating is needed. Although the sample was filtered to retain only the residential buildings that are heated, houses that are in need of very little heating and may not be equipped with thermostats are included in the sample.

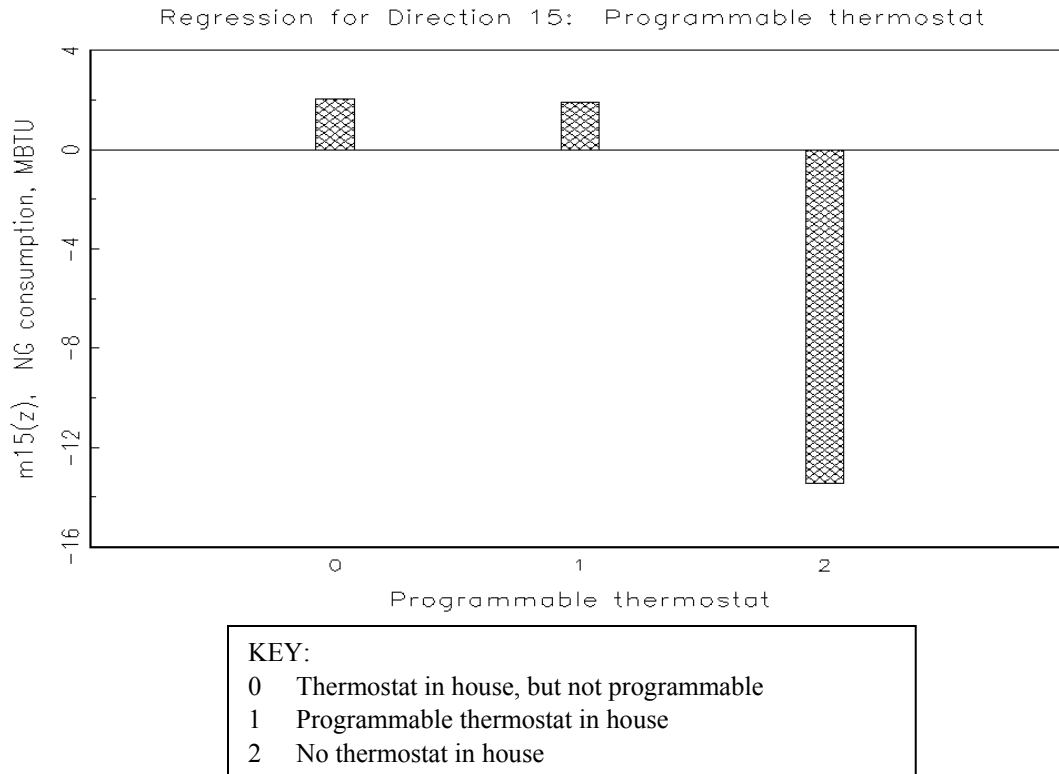


Figure 3.8. Impact of Thermostat and Programmable Thermostat on NG Use

Behavioral information is contained in Directions 16 and 17, which deal with programming thermostats to lower temperature for heat setting at night and, correspondingly, when no one is home. The results of these regressions are found in Appendix B. The result of Direction 16 is counterintuitive because it suggests that programming the thermostat to lower temperature automatically is associated with higher NG use. Neither the direction of change, nor magnitude (3 MBtu) are intuitive. Direction 17 also produced a counterintuitive pattern. It indicates that the highest NG consumption is for the houses with thermostats preprogrammed to lower settings when no one is home during the day. Then it drops by about 1 MBtu for the houses that have no thermostats, and drops down even further for houses where the temperature is not lowered. For detailed analysis of these two variables, more refined data is needed. To separate the behavioral impact, it is necessary to also account for climate. Thermal integrity of the building usually is strongly correlated with the climate. In turn, in more severe climate conditions, where NG intensities are the highest, the inhabitants are more likely to adjust thermostats up or down from the base setting.

Figure 3.9 (Direction 27) shows the impact that the number of thermostats in the house (from zero to six) has on NG use. The drop in the NG consumption between the category with no thermostat and one thermostat by 1 MBtu is reasonable. Then the consumption increases by 17 MBtu for houses with two thermostats. The highest level is registered for three-thermostat houses, leading the previous group by about 2 MBtu. This could be explained by the fact that this variable contains redundant information because number of thermostats is linked to the house size. In addition, the number of thermostats might be a representation of inefficient heating systems with individual dials in each room in older houses. For each additional thermostat after three, the consumption drops.

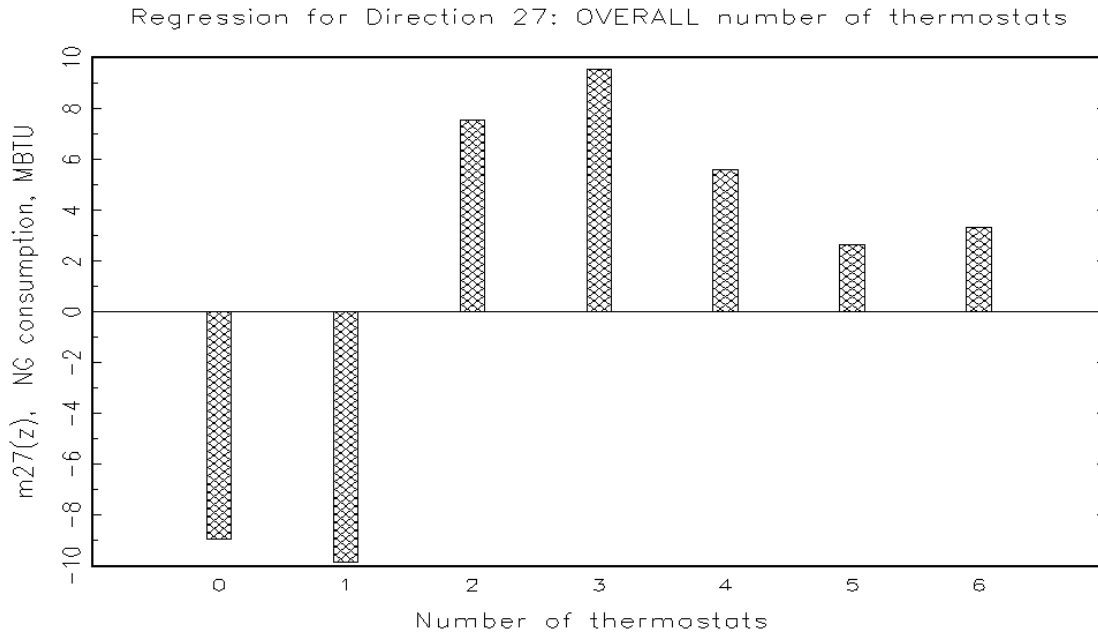


Figure 3.9. Impact of Number of Thermostats on NG Use

3.4 Prices and Billing Structure

The modeling results suggest a positive relationship between electricity price and NG use (Direction 4), which is expected considering that electricity is the primary NG substitute in residential buildings. Increases in electricity prices encourage switching to NG as the primary fuel for the household. The results for own price effect (i.e., price of NG) on NG (Direction 5) is negative, as expected. Increased NG price results in reductions of NG consumption. Both of these price effects are shown in Appendix B.

Direction 21 is of particular interest because it provides some insight on the relationship between the method of how NG is billed and its consumption level. As shown in Figure 3.10, if the household sees the full bill and pays it all, it seems to suggest the lowest result among all categories. Paying the utility bill in full corresponds to category 0. The consumption increases significantly, on the order of 16 MBtu, if all of the payment gets included in rent (category 1) or the household faces only a portion of the total bill for rented dwelling (category 2). This increase could be attributed to differences in willingness to pay for various technology options or invest in energy efficiency between the renters and the owners residing in the house. The result also suggests the difference in NG consumption resulting from the signal of NG prices not reaching the consumer, or a behavioral difference resulting from the "paid for" attitude of the consumer that pays a lump sum irrespective of the actual usage. Such a result is consistent with currently ongoing research on residential energy efficiency.

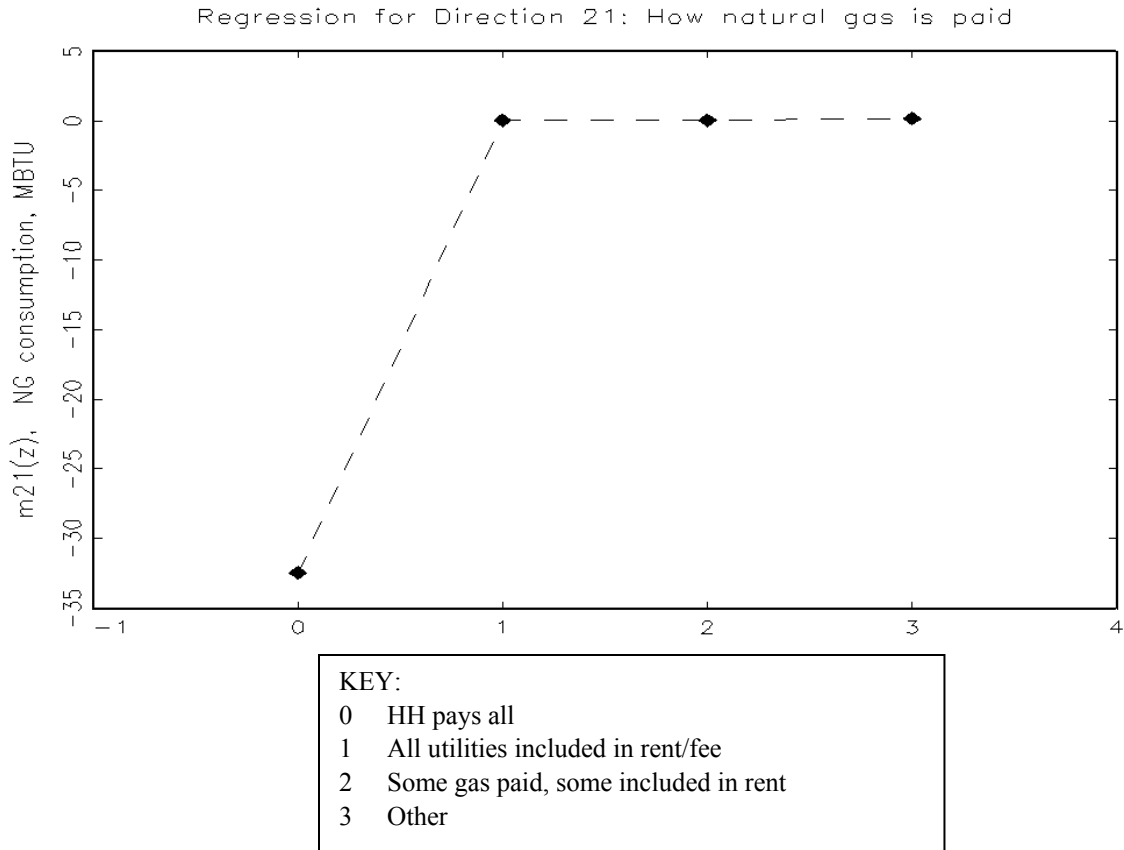
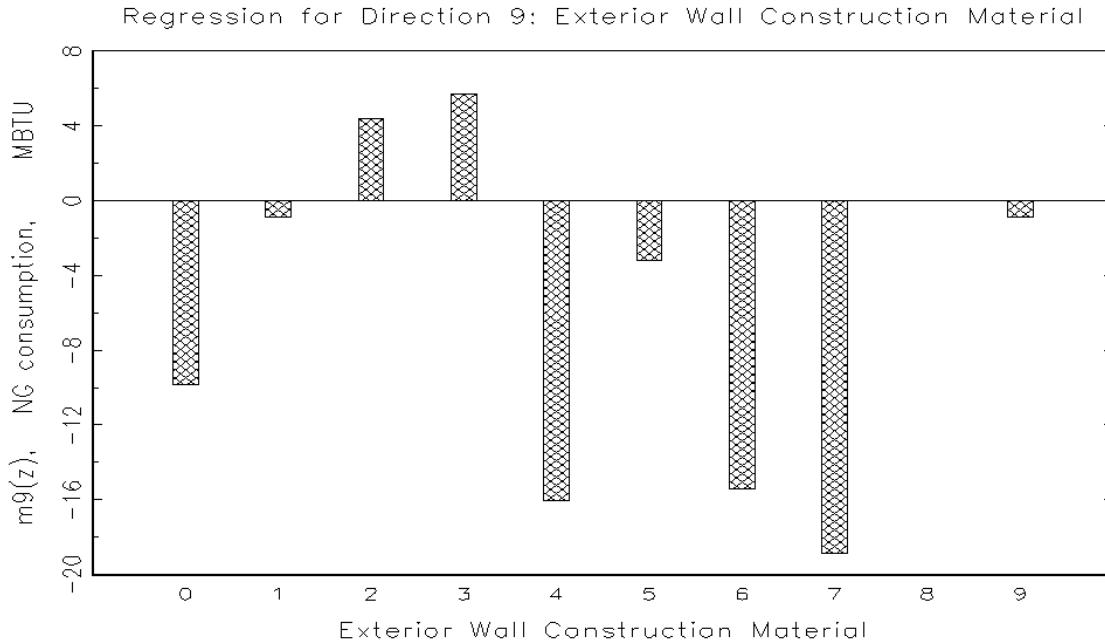


Figure 3.10. Impact of How NG is Paid on NG Use

3.5 Home Construction Attributes

3.5.1 Building Shell

Figure 3.11 (Direction 9) shows the impact of the exterior wall construction material on NG use. All other things held equal, the change of the wall type variable leads to the expected change in the NG intensity. The lowest NG consumption is shown for stucco, concrete block and stone. By stucco, residents usually refer to either the synthetic cladding that is applied over polystyrene panels, which provide extra insulation, or to cement plaster (lime sand and Portland cement). If installed properly, the latter seals the house, but not as thoroughly as synthetic systems. Concrete block and stone will serve as thermal mass storage, slowing down heat loss. The highest NG consumption is shown for houses with aluminium/vinyl/steel siding or wood shingles. This is consistent not only with the properties of each material and construction methods associated with it, but also with the vintage of the homes that would have these materials installed. In turn, there is a strong correlation between house vintage and quality of wall insulation.



| KEY: | |
|------|------------------------------------|
| 0 | Indescribable |
| 1 | Brick |
| 2 | Wood |
| 3 | Siding (Aluminum, vinyl, or steel) |
| 4 | Stucco |
| 5 | Composition (Shingle) |
| 6 | Stone |
| 7 | Concrete or concrete block |
| 8 | Glass |
| 9 | Other |

Figure 3.11. Secondary Heating Equipment Impact on NG

Direction 29 analyzes the building shell component heat load contributions by looking at the windows with various glazing and insulating characteristics. In Figure 3.12, the left side of the chart shows the increase in the natural gas consumption across the first three categories (single-paned glass, double-paned glass and double-paned glass with low-E coating). This result is somewhat counterintuitive because it would be expected that number of window panes (e.g., single-paned versus double-paned) should be negatively correlated with energy demand, because improved windows have higher energy efficiency. One possible explanation might be the size difference between older single-paned windows and newer double-paned. There is a trend to increase size of windows or incorporating additional windows when retrofits are implemented. Also, newer houses tend to have a higher number of windows, which would also increase heat loss and result in the higher NG consumption. In addition, this can also be affected by the climate. Unfortunately the information on window quantity and sizes is not available to test either one of the assertions. Climate information is not included either. NG consumption goes down for categories with triple-pane glass (category 3) and triple-pane glass with low-E coatings (category 4 and 5), which is expected.

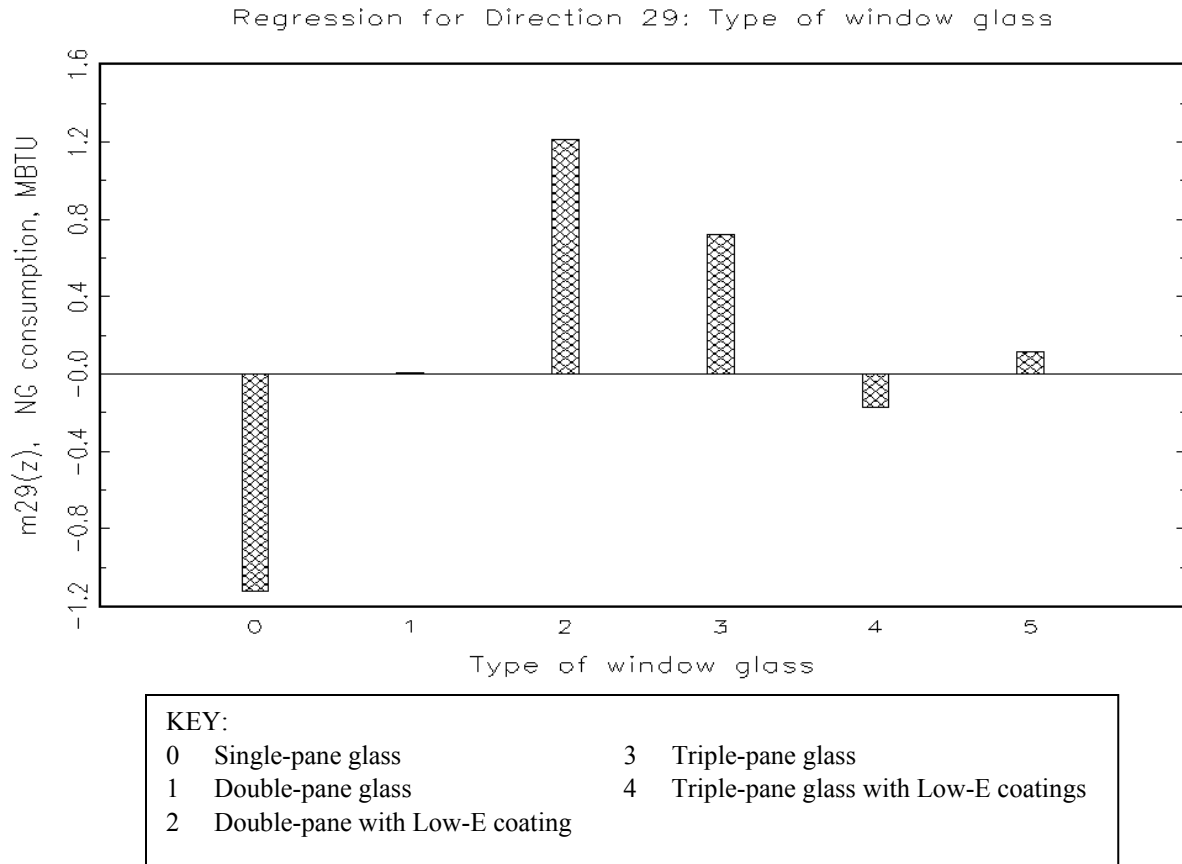


Figure 3.12. Impact of Type of Window Glass on NG Use

3.5.2 Size and Design

Figure 3.1 (Direction 3), which is described in detail at the beginning of the results section, shows the dependency between NG intensity and the total square footage of the house. The suggested relationship is linear over the range of square footage where the most observations are concentrated. So the natural gas demand grows linearly for households between 900 and 6000 s.f. Consumption plateaus after 8000 s.f.; however, this occurrence should not be given much emphasis because there are very few points in this range.

Direction 10 examines the impact of whether or not a home has a garage or heated garage on NG use. The results of the Direction 10 regression were reasonable and are presented in Figure 3.13. Category 0 corresponds to the house with no garage. Category 1 represents the houses where there is a garage, but it is not heated. Attached garage provides additional buffer between the heated part of the house and the environment, thus slowing down heat loss. The results suggest that heating the garage will increase natural gas consumption by up to 14 MBtu. Complete interpretation of this increase also depends on whether garage space is included in the total square footage of the house. Also, this regressor is picking up additional effects impacting NG use. Absence of a garage is more typical of older neighborhoods with lower housing prices. They often share similar quality of construction, amount of insulation and level of equipment. Therefore, fairly high NG intensity for houses with no garage is not an unexpected result.

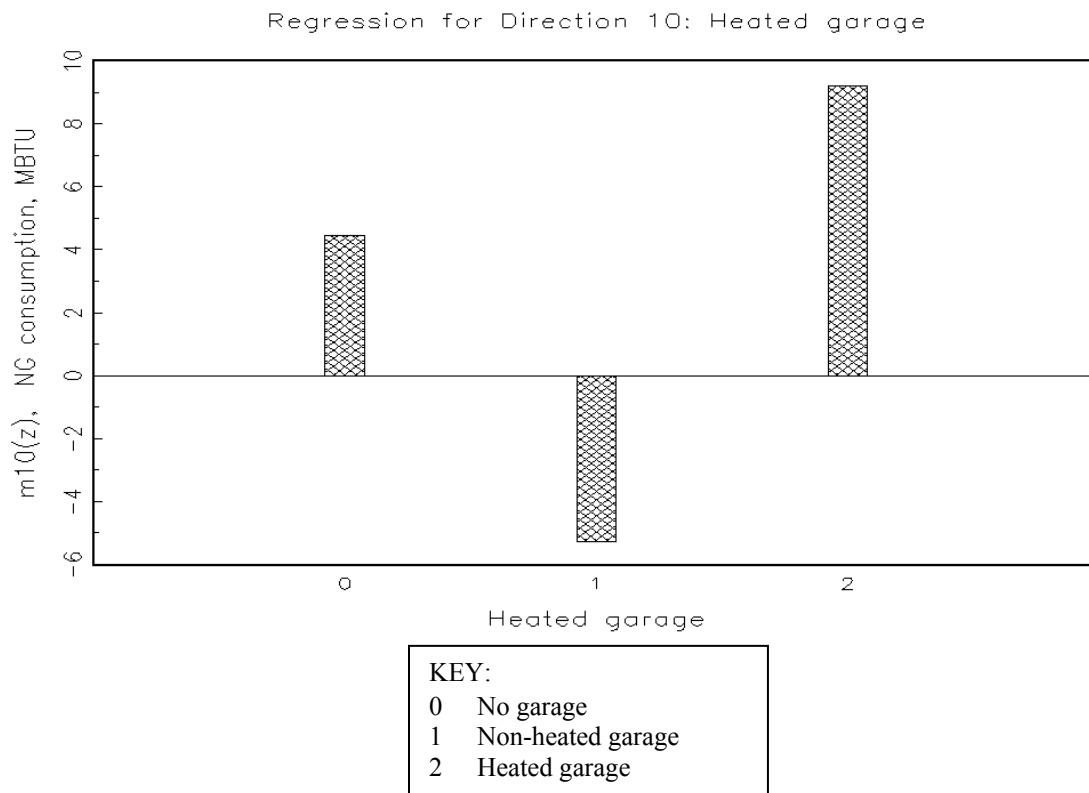


Figure 3.13. Impact of Heating Garage on NG Use

Direction 23 characterizes the impact from the number of stories in the building, and the results are presented in Appendix B. The lowest NG consumption is for the one-story building, followed by the split level house and two-story structure. The highest level is for the three-story dwellings. As the number of stories increases, the structure design tends to change towards narrower buildings. This leads to a much higher exchange surface, which explains higher NG intensity for buildings in this category. It is necessary to note that all apartment complexes were excluded from the sample. The results cover only single-family detached housing units.

Direction 24 produced rather interesting results that are shown in Figure 3.14. Category 3, where the entire basement is heated during winter shows highest NG consumption. The second highest demand for NG is shown for the houses that have a basement but do not heat any portion of it (category 1). It is followed by the houses where there is a basement and portion of it is heated. This result appears counterintuitive, but may have reasonable explanation. Unheated basements are typical for older houses with unfinished basements. If a portion of it is heated, it is likely that the thermal integrity of the basement has been improved. The difference between these two categories is 2 MBtu. This directional result could be different if the regressor is restructured as a binary versus ordered categorical variable, such that it does not attempt to account for a particular portion of the basement which measurement is not defined. Also, if the retrofit information were available, it would be possible to analyze its correlation with the vintage of the house.

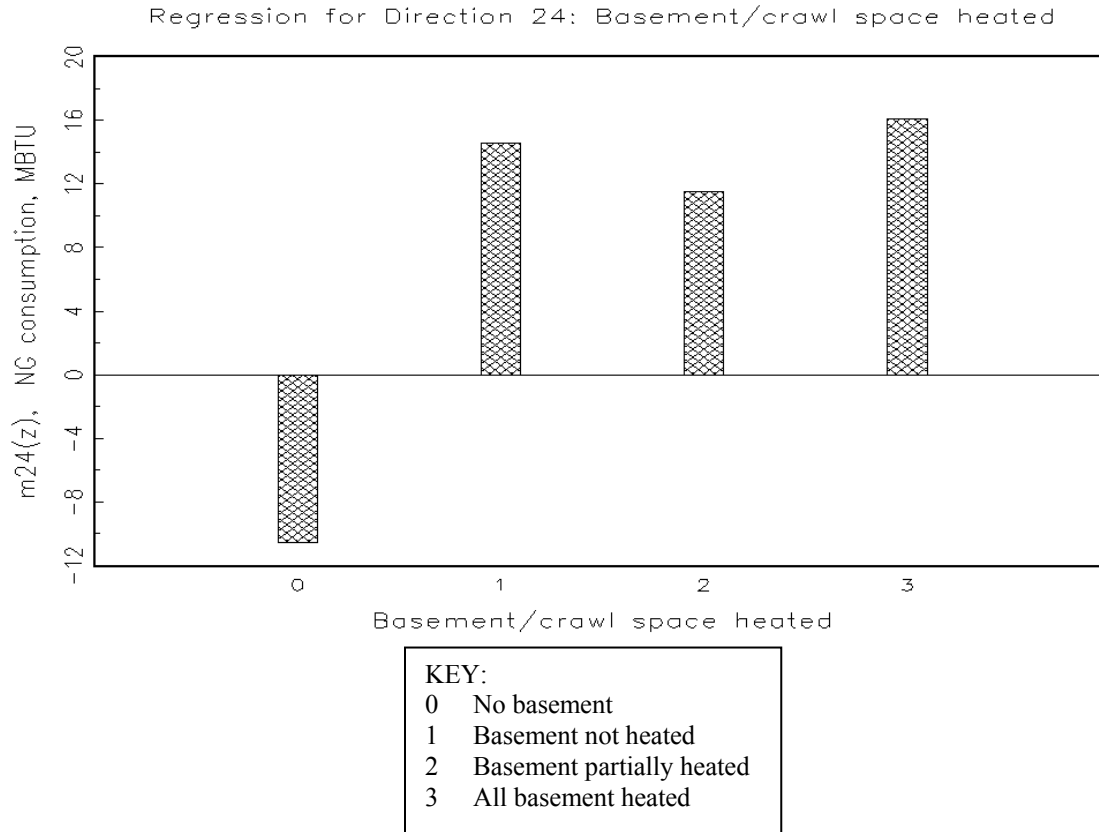


Figure 3.14. Impact of Basement/crawl space on NG Use

Direction 25 describes the portion of the attic that is warm, and the results are reasonable. It suggests a linear relationship between the fraction of attic that is heated and NG consumption, and the results are shown in Appendix B. The difference between a house with no attic versus a house with an unheated attic is approximately 4 MBtu. Usually no attic implies a flat roof with not much room for insulation. Just the presence of an attic has a favorable effect, because it provides a buffer zone slowing down the heat loss in addition to allowing better insulation. This is followed by the partially heated attic with increase in NG demand by about 8 MBtu. The highest NG consumption is shown for fully heated attic, which would be expected.

No particular pattern of dependency between number of rooms not heated during the winter and the NG demand can be derived from the results of Direction 28 (see Figure 3.15). On the surface it would seem likely that this variable should have an inverse impact on NG consumption, because more rooms that are unheated in winter would imply that less NG should be consumed. However, any unheated space that is not zoned appropriately can contribute to the heating load of a house.

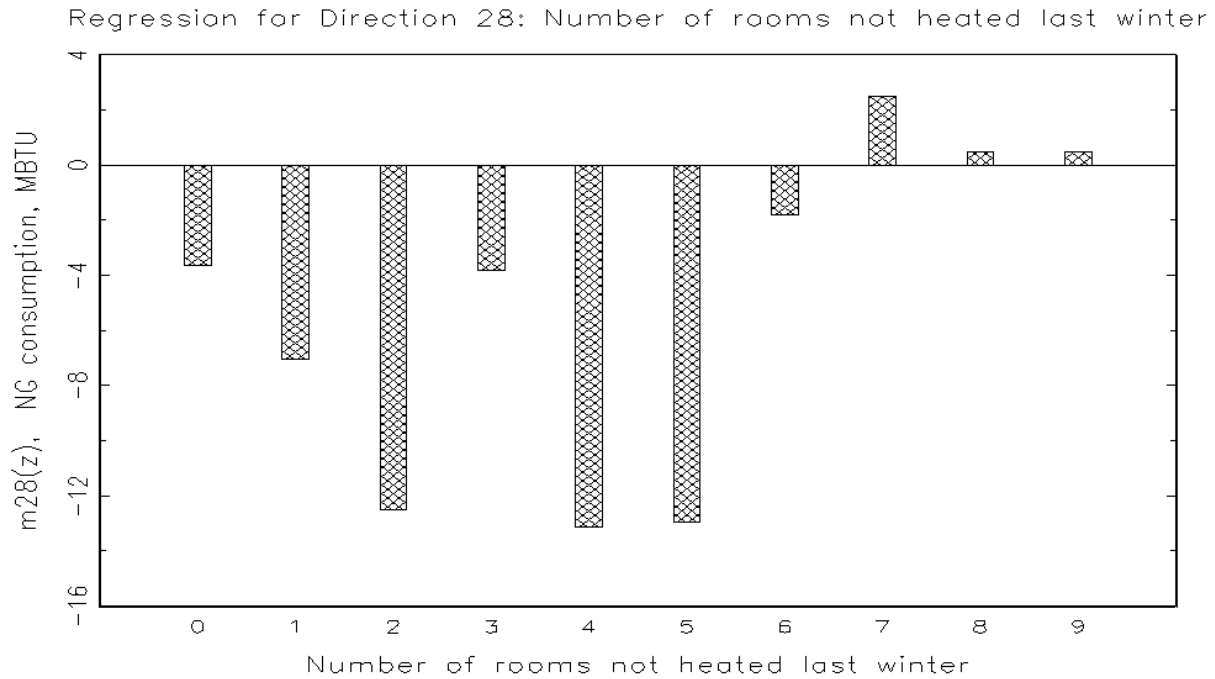


Figure 3.15. Impact of Number of Rooms Not Heated on NG Use

3.6 Vintage

Regression results for house vintage and NG use (Direction 26) are reasonable, and are presented in Figure 3.16. The highest NG consumption is shown for category 0 that represents houses built before 1940. NG demand decreases for the houses built in the 1940s by about 10 MBtu, which is followed by the 1950s vintage. There is an increase in the NG consumption of housing built between 1960 and 1969, up from the level shown for 1950 vintage by 5 MBtu, which may be attributable to changes in construction practices. For houses built between 1970 and 1989, the NG consumption decreases by 8 MBtu, which corresponds to improvements in thermal integrity. This trend reverses for dwellings built after 1990, which can be attributed to several factors. First and foremost, this is the period when houses with high ceilings gained popularity. In addition, this market trend was accompanied by a shift in the design away from standard rectangular houses to designs with less conventional angles and additional coves. The latter contributes to lower overall energy efficiency of the house, and the effect is reinforced by the ceiling height, leading to even more drastic efficiency loss.

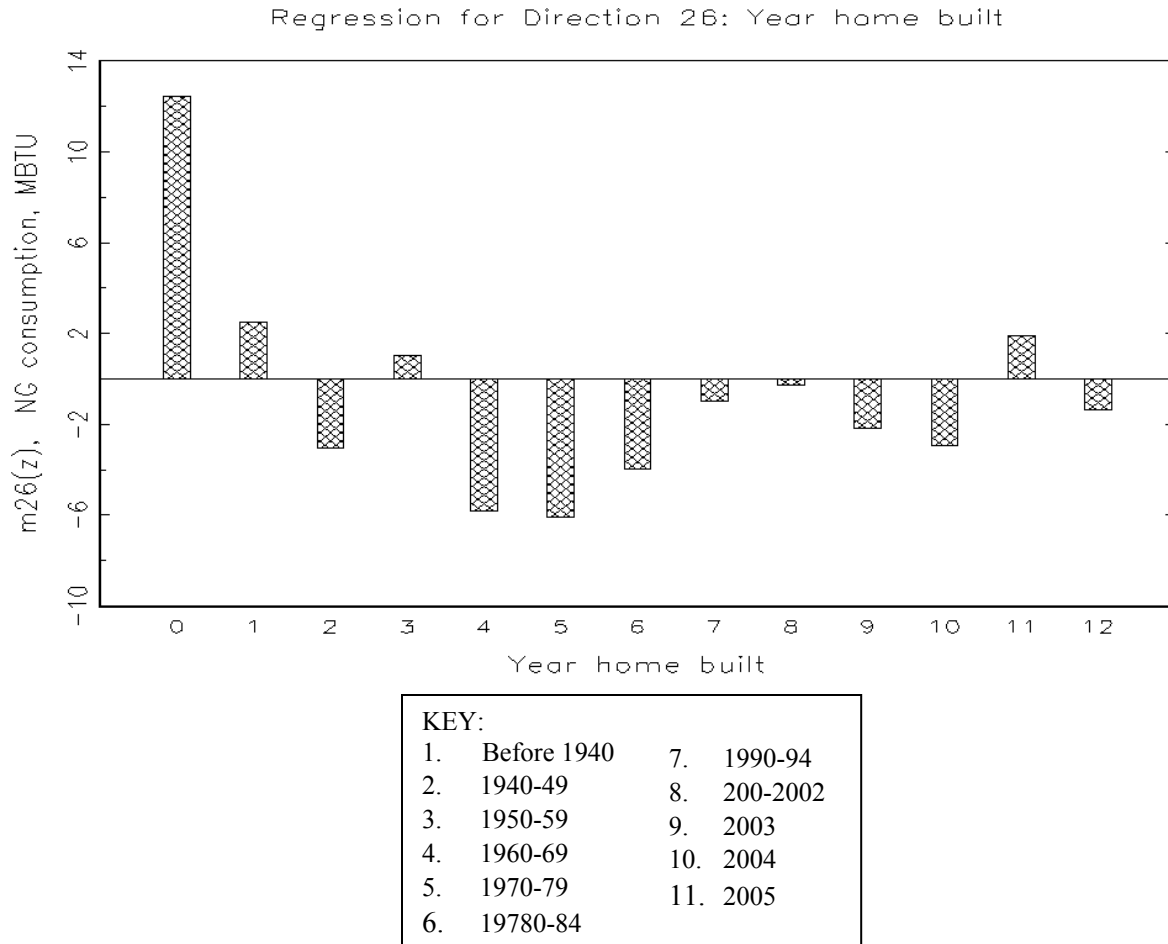


Figure 3.16. Impact of House Age on NG Usage

3.7 Home Ownership

Figure 3.17 (Direction 11) identifies the relationship between the NG intensity and ownership of the house. The result is reasonable because owned houses have lower energy consumption compared to rented (the middle) and occupied without payment (the highest). The difference between three categories is around 4 MBtu, with delta between the second and the third category being over 1 MBtu. This is consistent with previously documented results of the Caravan Opinion Research Corporation (ORC) 2007 surveys. These surveys showed a higher willingness to invest in the energy-saving solutions and high overall concern about the energy efficiency of the residential structure being more typical for the landlords than the renters. There is also a difference in investment decisions associated with primary dwellings versus rentals or additional houses used by relatives or friends without rent payment.

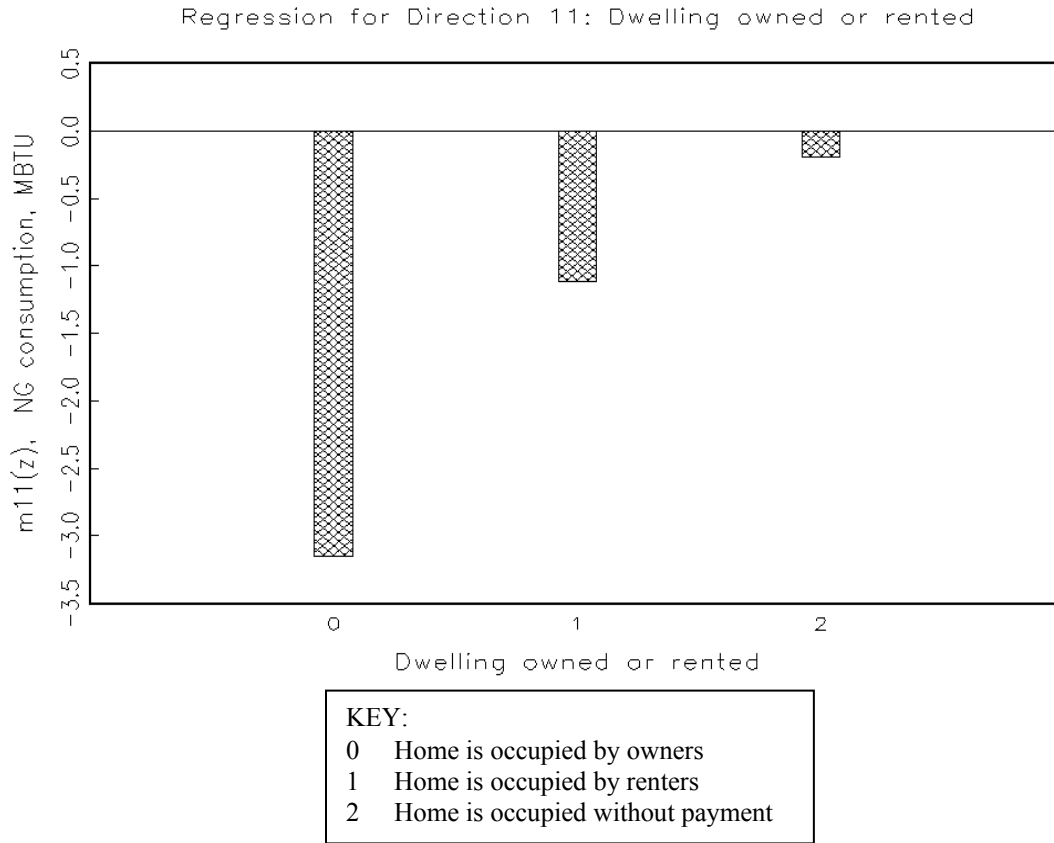


Figure 3.17. Impact of Ownership/Rental Status on NG Use

3.8 Occupancy

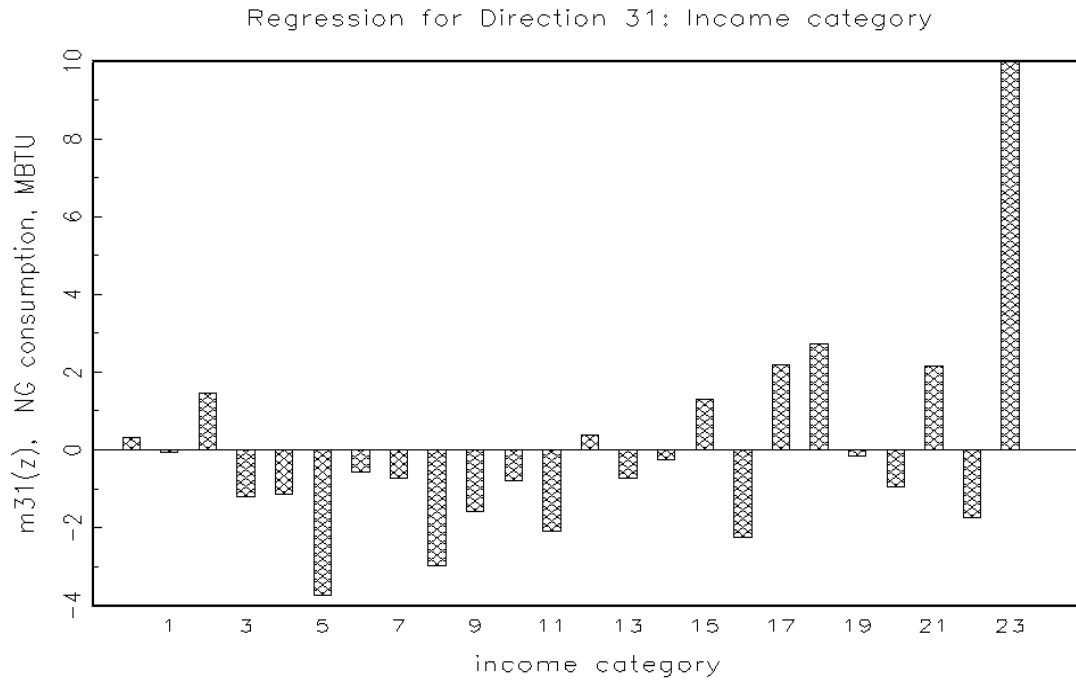
Direction 22 picks up the difference in the natural gas intensity because of someone staying at home the whole day versus the house being unoccupied during working hours. There is approximately a 1.5 MBtu delta resulting from someone reportedly occupying the house during the day. The results are shown in Appendix B. Figure 3.18 (Direction 30) describes the relationship between NG consumption and number of people living in the house. The result is reasonable considering that NG demand would likely increase with each consecutive inhabitant. The magnitude of change is also reasonable because marginal change decreases with each consecutive occupant. Gas consumption drops by 3 MBtu as the number of inhabitants grows from 5 to 7, suggesting that results could plateau after a certain number of residents representing economies of scale in NG usage -- a reasonable result considering heating requirements would not change with each consecutive inhabitant and natural gas consumption associated with water heating, cooking and dryer use would go up at a smaller rate.



Figure 3.18. Impact of Number of Occupants on NG Use (0 = none, up to 10 occupants)

3.9 Income

Direction 31 links the income level with the natural gas consumption of the household, and the results are shown in Figure 3.19. It can be concluded that based on the number of categories, this variable should be treated as continuous. Initially there is a slight drop in NG intensity as the income grows from less than \$2500 to approximately \$25,000. As income grows, an increase in NG consumption is observed. Categories 11 through 18 correspond to the income interval from \$45,000 to \$85,000. Income at these levels would at least be partially linked to the type of the house, quality of construction, level of insulation and types of equipment serving the household, and this would likely be another representation of the multicollinearity in the data. This increase is followed by a drop in NG consumption for income categories in excess of \$85,000. It can be attributed not only to the direct effect caused by a change in willingness to invest in the energy-efficient solutions, but also a change in level of education and environmental considerations, as well as the shift in the initial quality of occupied homes.



| KEY: | | | | | |
|------|----------------------|----|----------------------|----|----------------------|
| 0 | Less than \$2,500 | 6 | \$20,000 to \$24,999 | 12 | \$50,000 to \$54,999 |
| 1 | \$2,500 to \$4,999 | 7 | \$25,000 to \$29,999 | 13 | \$55,000 to \$59,999 |
| 2 | \$5,000 to \$7,499 | 8 | \$30,000 to \$34,999 | 14 | \$60,000 to \$64,999 |
| 3 | \$7,500 to \$9,999 | 9 | \$35,000 to \$39,999 | 15 | \$65,000 to \$69,999 |
| 4 | \$10,000 to \$14,999 | 10 | \$40,000 to \$44,999 | 16 | \$70,000 to \$74,999 |
| 5 | \$15,000 to \$19,999 | 11 | \$45,000 to \$49,999 | 17 | \$75,000 to \$79,999 |

Figure 3.19. Impact of Income on NG Use

4.0 Conclusions

This study employs an econometric approach to analyzing natural gas consumption intensity of residential buildings that can be used in combination with simulations for describing the impact of various household and structure attributes on energy demand. The econometric approach employed uses a local linear smooth backfitting estimator, which is extended to include categorical variables. Satisfactory results were obtained for the majority of the covariates, and the estimation technique was able to accommodate a correlated set of mixed data.

Nonparametric regression estimation revealed patterns of dependency that could not have been achieved by parametric analysis. Some of the results were suggestive of particular parametric relationships. However, these relationships were only sustained over a portion of the regressor range, because the overall result has the appearance of several superpositioned parametric associations depending on what interval of the regressor support is considered.

This analysis could be extended by combining smooth backfitting regression with stochastic frontier estimation via the method suggested by Fan, Li and Weersink (1996) and, more importantly, by using the generalized profile likelihood framework of Severini and Wong (1992). The comparison can be done across residential buildings or groups of residential buildings based on the ranked efficiency score. The regression portion of the analysis would provide the ability to interpret the efficiency scores from the energy management view point because a combination of efficiency scores along with each directional regression result allows further investigation of possible causes. This approach could also provide information on the selection of building technologies and engineering and behavioral solutions that could potentially improve the level of energy intensity of residential buildings. One of the issues with using the suggested approach is to clearly understand how a production frontier can be defined within the context of natural gas usage by residential buildings. If it was possible to isolate only the information that is related to heating, then the thermostat setting could be used as a proxy for the output. The efficiency of maintaining the dwelling at that temperature while all other inputs, attributes and characteristics vary could be compared through ranking. Clusters of houses with similar ranking would provide an insight into what primary features, behavioral characteristics, and house attributes impact the ability to maintain residential buildings at a set temperature.

The benefit of the current analysis is three-fold. The main result, which is the directional impact of each covariate, can be utilized for in-sample prediction to approximate energy demand of a residential building whose characteristics are described by the regressors used in this analysis, but a certain combination of their particular values does not exist in the real world. The only caution is that the best estimates are for the interior of the intervals, where the regressors take values. The closer the values are to the end-points of the regressor range, the less accurate the results.

The second benefit is the information on how natural gas demand might change once a particular characteristic or attribute is altered. For continuous variables, the local linear framework applied in this study produces the values of the slope at each observation as part of the estimation procedure. As far as the categorical variables are concerned, the slope estimates are not calculated as part of the procedure, but they can be easily computed by comparing change in the natural gas usage while moving from one category to another for each of the regressors. For example, results on wall construction material suggest that the natural gas consumption goes down by about 8 MBtu for houses with composite (shingle) siding

versus houses with vinyl siding. Properly installed stucco siding may reduce the gas consumption even further (by about 10 MBtu). Jointly with the cost estimates of such improvements, these results can be used as a quick tool for benefit-cost analysis of residential upgrades and retrofits under a fixed budget.

The third and the most obvious result follows along the lines of the previously discussed benefit, but with a very particular implication. It shapes the message that changing, for example, the thermostat temperature setting several degrees up or down while holding everything else fixed has a very tangible effect on natural gas usage and related household energy expenditures. Another behavioral result is the relationship between natural gas consumption and billing method. Seeing the full bill and paying it in full corresponds to the lowest energy consumption level. The consumption increases significantly if a household faces only portion of the bill, or if the full payment is included in rent and the actual consumer never sees either the amount of natural gas consumed, or associated monthly expenditures. The link is obvious, the link is measurable, and the result is produced by a nonparametric estimation procedure without imposing a particular specification on the shape of that relationship.

The primary objective of this analysis was to investigate the applicability of a particular nonparametric methodology to quantifying the impact of behavioral variables using econometric methods. Behavioral aspects of energy usage are largely treated by traditional parametric models as an unobservable effect. If good-quality microdata is available on behavioral aspects of energy usage, it is possible to extend this nonparametric analysis to a larger number of regressors and encompass the relationship between behavioral changes and energy usage at a more refined level.

General Conclusion

This study investigated the relationship between natural gas demand and characteristics of the dwelling, demographic characteristics of occupants and behavioral variables. The existing modeling literature, whether it relies on parametric specifications or engineering simulation, does not accommodate inclusion of a behavioral component. This study attempts to bridge that gap and investigate the applicability of additive nonparametric regression to this task. The results of this analysis can be used for three primary purposes. The first one is an in-sample prediction for approximating energy demand of a residential building whose characteristics are described by the regressors in this analysis, but a certain combination of their particular values does not exist in the real world. The second potential application is for benefit-cost analysis of residential upgrades and retrofits under a fixed budget, because the results of this study contain information on how natural gas consumption might change once a particular characteristic or attribute is altered. The third purpose is to establish a relationship between natural gas consumption and changes in behavior of occupants. Although information on behavioral variables is generally limited, results of the analysis identify what information would be helpful to further research.

5.0 References

- Baker, P., R. W. Blundell and J. Micklewright. 1989. Modelling household energy expenditures using micro-data. *Economic Journal* 99, 720-738.
- Caravan Opinion Research Corporation, 2007. Study #716287.
- Crawley, Drury B, Linda K Lawrie, Curtis O Pedersen, Frederick C Winkelmann, Michael J Witte, Richard K Strand, Richard J Liesen, Walter F Buhl, Yu Joe Huang, Robert H Henninger, Jason Glazer, Daniel E Fisher, Don B Shirey III, Brent T Griffith, Peter G Ellis and Lixing Gu. 2004. Energy Plus: New, Capable, and Linked. *Journal of Architectural and Planning Research* 21, 4 (Winter 2004).
- DOE-2, 1993. BDL Summary Version 2.1E. LBNL, 34946, Lawrence Berkeley National Laboratory, Berkeley, CA
- Fan, Y., Q. Li, and A. Weersink. 1996. Semiparametric estimation of stochastic production frontier models. *Journal of Business and Economic Statistics* 14, 460-468.
- García-Cerruti, L. 2000. Estimating elasticities of residential energy demand from panel county data using dynamic random variables models with heteroskedastic and correlated error terms. *Resource and Energy Economics* 22, 355-366.
- Halvorsen, B. and B. Larsen. 2001. The flexibility of household electricity demand over time. *Resource and Energy Economics* 23, 1-18.
- Holtedahl, P. and F. Joutz. 2004. Residential electricity demand in Taiwan. *Energy Economics* 26, 201-224.
- Kamerschen, D. and D. Porter. 2004. The demand for residential, industrial and total electricity, 1973-1998. *Energy Economics* 26, 87-100.
- Larsen, B. and R. Nesbakken. 2004. Household electricity end-use consumption: results from econometric and engineering models. *Energy Economics* 26, 179-200.
- Lutzenhiser, L. 1993. Social and behavioral aspects of energy use. *Annual Review of Energy Economics* 18, 247-89.
- Madlener, R. 1996. Econometric analysis of residential energy demand: a survey. *Journal of Energy Literature* 2, 3-32.
- Mammen, E., O. Linton, and J. P. Nielsen. 1999. The existence and asymptotic properties of a backfitting projection algorithm under weak conditions. *Annals of Statistics* 27, 1443--1490.
- Narayan, P. and R. Smyth. 2005. The residential demand for electricity in Australia: an application of the bounds testing approach to cointegration. *Energy Policy* 33, 467-474.

- Nesbakken, R. 2001. Energy consumption for space heating: discrete-continuous approach. *Scandinavian Journal of Economics* 103, 165-184.
- Nielsen J.P. and S.Sperlich. 2005. Smooth backfitting in practice. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 67, 43-61.
- Racine, J. S. and Q. Li. 2004. Nonparametric estimation of regression functions with both categorical and continuous data. *Journal of Econometrics* 119, 99-130.
- Racine, J.S., Q. Li., and X. Zhu. 2004. Kernel estimation of multivariate conditional distributions. *Annals of Economics and Finance* 5, 211-235.
- Schmalensee, R. and T. M. Stoker. 1999. Household gasoline demand in the United States. *Econometrica* 67, 645-662.
- Severini T.A., W.H. Wong, Pro.le Likelihood and Conditionally Parametric Models *Annals of Statistics*, Vol. 20, No. 4 (Dec., 1992), pp.1768-1802
- U.S. Department of Energy. 2005. ENERGY PLUS, Input Output Reference, Washington, D.C.
- U.S. Department of Energy. 2009. Energy Information Administration (DOE/EIA). Annual energy outlook 2009 with projections to 2030. DOE/EIA-0383 (2009). Washington, D.C.
- U.S. Department of Energy. 2009. Energy Information Administration (DOE/EIA). Residential Energy Consumption Survey 2005, Washington D.C.
- Yatchew, A. and J. No. 2001. Household gasoline demand in Canada. *Econometrica* 69, 1697-1709.

Appendix A

Detailed Methodology

Appendix A

Detailed Methodology

A.1 Smooth backfitting for continuous data

The regression model considered here is of the following form:

$$E(Y|X_1 = x_1, \dots, X_d = x_d) = m_0 + \sum_{j=1}^d m_j(x_j)$$

where (Y, X_1, \dots, X_d) is a random vector in \mathbb{R}^{d+1} and we assume that there is a random sample $\{y_i, x_{i1}, \dots, x_{id}\}_{i=1}^n$ of (Y, X_1, \dots, X_d) , m_0 is an unknown scalar parameter, $m_j(x_j)$ is a sufficiently smooth function for all j , and θ_j is the first order derivative of $m_j(x_j)$. Also, for identification purposes, $E(m_j(x_j)) = 0$.

Let $K_h(x_{ij} - x_j) = \frac{1}{h} K\left(\frac{x_{ij} - x_j}{h}\right)$ be a kernel function such that $\int K(\phi) d\phi = 1$, $\int \phi K(\phi) d\phi = 0$, $\int \phi^2 K(\phi) d\phi = 1$. Bandwidth is defined as $h = h(n)$ such that $h \rightarrow 0$ and $nh \rightarrow \infty$ as $n \rightarrow \infty$, and conditions B(1), B(2')-B(4') of Mammen et al. (1999) are met. The backfitting estimator is obtained by minimizing the following objective function

$$\int \sum_{i=1}^n \left[y_i - m_0 - \sum_{j=1}^d m_j(x_j) - \sum_{j=1}^d \theta_j(x_j) (x_{ij} - x_j) \right]^2 \times \prod_{j=1}^d K_h(x_{ij} - x_j) dx$$

The minimization is done with respect to m_0, m_1, \dots, m_d and all first derivatives $\theta_j(x_j)$.

Let

$$\begin{aligned} \hat{p}_j(x_j) &= n^{-1} \sum_{i=1}^n K_h(x_{ij} - x_j), \quad \hat{p}_j^j(x_j) = n^{-1} \sum_{i=1}^n K_h(x_{ij} - x_j) (x_{ij} - x_j), \\ \hat{p}_j^{jj}(x_j) &= n^{-1} \sum_{i=1}^n K_h(x_{ij} - x_j) (x_{ij} - x_j) (x_{ij} - x_j), \\ \hat{p}_{jk}(x_j, x_k) &= n^{-1} \sum_{i=1}^n K_h(x_{ij} - x_j) K_h(x_{ik} - x_k), \\ \hat{p}_{jk}^k(x_j, x_k) &= n^{-1} \sum_{i=1}^n K_h(x_{ij} - x_j) K_h(x_{ik} - x_k) (x_{ik} - x_k), \\ \hat{p}_{jk}^{jk}(x_j, x_k) &= n^{-1} \sum_{i=1}^n K_h(x_{ij} - x_j) K_h(x_{ik} - x_k) (x_{ij} - x_j) (x_{ik} - x_k), \end{aligned}$$

Let

$$\begin{aligned}
A &= \frac{n^{-1} \sum_{i=1}^n K_h(x_{ij} - x_j) y_i}{\hat{p}_j(x_j)} - \sum_{k \neq j}^d \int \tilde{m}_k(x_k) \frac{\hat{p}_{jk}(x_j, x_k)}{\hat{p}_j(x_j)} dx_k \\
&\quad - \sum_{k \neq j}^d \int \tilde{\theta}_k(x_k) \frac{\hat{p}_{jk}^k(x_j, x_k)}{\hat{p}_j(x_j)} dx_k - \tilde{m}_0(x_j), \\
B &= \frac{n^{-1} \sum_{i=1}^n K_h(x_{ij} - x_j) (x_j - X_{ij}) y_i}{\hat{p}_j^j(x_j)} - \sum_{k \neq j}^d \int \tilde{m}_k(x_k) \frac{\hat{p}_{jk}^j(x_j, x_k)}{\hat{p}_j^j(x_j)} dx_k \\
&\quad - \sum_{k \neq j}^d \int \tilde{\theta}_k(x_k) \frac{\hat{p}_{jk}^{jk}(x_j, x_k)}{\hat{p}_j^j(x_j)} dx_k - \tilde{m}_0(x) \\
C &= \frac{\hat{p}_j^j(x_j)}{\hat{p}_j(x_j)}, \quad D = \frac{\hat{p}_j^{jj}(x_j)}{\hat{p}_j^j(x_j)}
\end{aligned}$$

The smooth backfitting estimates of \tilde{m}_0 , \tilde{m}_j and $\tilde{\theta}_j$ are obtained by iteratively solving the two equations below for each regressor $j = 1, \dots, d$

$$\tilde{m}_j(x_j) = A - \tilde{\theta}_j(x_j)C, \quad \tilde{\theta}_j(x_j) = \frac{A - B}{C - D}$$

As a consequence of imposing normalization condition, $\tilde{m}_0 = n^{-1} \sum_{i=1}^n y_i$.

A detailed discussion establishing the asymptotic properties of the smooth backfitting estimator for the case of only continuous regressors is presented in Mammen et al. (1999). Their final result is summarized as the convergence in distribution that holds for any x_1, \dots, x_d with compact support:

$$n^{2/5} \begin{pmatrix} \tilde{m}_1(x_1) - m_1(x_1) + v_{n,1} \\ \vdots \\ \tilde{m}_d(x_d) - m_d(x_d) + v_{n,d} \end{pmatrix} \xrightarrow{d} N \left[\begin{pmatrix} c_h^2 \delta_1(x_1) \\ \vdots \\ c_h^2 \delta_d(x_d) \end{pmatrix}, \text{diag} \{v_j(x_j)\}_{j=1}^d \right],$$

$$\delta_j(x_j) = \frac{\int u^2 K(u) du}{2} \left\{ m_j''(x_j) - \int m_j''(x_j) p_j(x_j) dx_j \right\},$$

$$v_{n,j} = \int m_j(x_j) K_h(x_j - u) p_j(u) du dx_j,$$

$$v_j(x_j) = c_h^{-1} c_k \sigma_j^2(x_j) / p_j(x_j),$$

with $c_k = \int K(u)^2 du$, c_h is a constant such that $n^{1/5}h \rightarrow c_h$. Second derivative of $m_j(x_j)$ is

represented by $m_j''(x_j)$, $p_j(u)$ is the marginal density, and $\sigma_j^2(x_j) = \text{var}[Y - m(x)|X_j = x_j]$ can be consistently estimated from the residuals $\tilde{\varepsilon}_i = y_i - \tilde{m}(x_i)$, $i = 1 \dots n$.

$$n^{2/5} (\tilde{m}(x) - m(x)) \xrightarrow{d} N \left\{ c_h^2 \sum_{j=1}^d \delta_j(x_j), \sum_{j=1}^d v_j(x_j) \right\},$$

where $\tilde{m}(x)$ is a smooth backfitting estimator of $m(x) = m_0 + \sum_{j=1}^d m_j(x_j)$ defined as $\tilde{m}(x) = \tilde{m}_0 + \sum_{j=1}^d \tilde{m}_j(x_j)$.

A.2 Smooth backfitting estimator for mixed data

In a wide variety of applications, especially dealing with microdata, one of the essential features of a regression estimator is its capability to accommodate continuous and categorical conditioning variables. Traditional approaches for estimating the categorical components have relied either on introducing these variables parametrically or implementing a frequency-based estimation. The major drawback of the first approach is a loss of flexibility induced by a fully nonparametric framework, as well as high likelihood of misspecification. The weakness of the second method stems from the requirement to divide the data into cells corresponding to the values taken by the discrete variables. This necessitates fairly large sample size in order for each cell to contain a reasonable amount of data as described in Li and Racine (2007).

Alternative procedures, such as smooth estimation of joint distributions and smooth regression for discrete data, are based on kernel estimation proposed by Aitchison and Aitken (1976). This latter method received attention in the recent literature as kernel smoothing methods have been gaining popularity. Li and Racine (2003) proposed a refined nonparametric kernel approach for estimating an unknown distribution defined over mixed discrete and continuous variables. Nonparametric estimation of regression functions was investigated by Racine and Li (2004), where specific smoothing techniques were considered for treatment of ordered and unordered categorical data. Structure of the proposed estimator is similar to that of Nadaraya-Watson local constant estimator, but with a different kernel employed for smoothing discrete variables. Li and Racine (2004) expanded the regression framework further by constructing a local linear nonparametric estimator for mixed data and investigating the theoretical properties of cross-validated bandwidth selection. In addition, they derived the rate of convergence of the cross-validated bandwidths and established asymptotic normality of the resulting nonparametric regression estimator. These results provide a foundation for incorporating categorical regressors into the local linear smooth backfitting estimator (SBE) and using least squares cross-validation to select bandwidth for both continuous and categorical regressors.

Let $x_j, j = 1, \dots, d$, denote continuous regressors and $x_t, t = 1, \dots, T$ denote the categorical variables. Discrete $x_{it}, i = 1, \dots, n$, takes values $\{0, 1, 2, \dots, c_t - 1\}$. For the local linear regression estimator Li and Racine (2004) propose using a variation of the Aitchison and Aitken (1976) kernel defined as

$$L(x_{it}, x_t, \lambda_t) = \begin{cases} 1, & \text{if } x_{it} = x_t \\ \lambda_t, & \text{if } x_{it} \neq x_t \end{cases} \quad t = 1, \dots, T.$$

This weight function does not add up to one, which cannot support the interpretation of marginal density $p_t(x_t)$ estimated by $\hat{p}_t(x_t) = n^{-1} \sum_{i=1}^n L(x_{it}, x_t, \lambda_t)$ as a proper density. It has been shown by Li and Racine (2004) that it is not the kernel shape, but rather the selection of the bandwidth parameter that has critical impact on the quality of resulting estimates. Therefore, to accommodate interpretation of weighting functions in smooth backfitting estimation as densities, another option is to use the kernel shape suggested by Aitchison and Aitken (1976) for the distribution estimation, namely

$$L(x_{it}, x_t, \lambda_t) = \begin{cases} 1 - \lambda_t, & \text{if } x_{it} = x_t \\ \lambda_t / (c_t - 1), & \text{if } x_{it} \neq x_t \end{cases} \quad t = 1, \dots, T$$

for unordered categorical regressors. The range of λ_t is $[0, (c_t - 1) / c_t]$. This weight function adds up to one. When λ_t assumes its upper value of $(c_t - 1) / c_t$, the kernel becomes $L(x_{it}, x_t, \lambda_t) = 1 / c_t$ regardless of whether $X_{it} = x_t$ or not. The resulting density estimator becomes unrelated to x_t thus smoothing it out. Alternatively, it is possible to use the weighting function that does not add up to one along with the normalization $p = p_t(x_t) / \sum p_t(x_t)$. For ordered categorical variable x_t the kernel of Li and Racine (2004)

$$L(x_{it}, x_t, \lambda_t) = \begin{cases} 1, & \text{if } x_{it} = x_t \\ \lambda_t^{|x_{it} - x_t|}, & \text{if } x_{it} \neq x_t \end{cases}$$

is utilized along with the above-mentioned normalization. The range of λ_t for ordered variables is $[0, 1]$. If λ_t takes its upper value the kernel becomes a uniform weight function. If $\lambda_t = 0$, the kernel turns into an indicator function. An alternative is to use the kernel

$$L(x_{it}, x_t, \lambda_t) = \begin{cases} 1 - \lambda_t, & \text{if } |x_{it} - x_t| = 0 \\ \frac{1 - \lambda_t}{2} \lambda_t^{|x_{it} - x_t|}, & \text{if } |x_{it} - x_t| \geq 1 \end{cases},$$

where x_t is a categorical variable and $x_{it}, i = 1, \dots, n$, takes values $\{0, 1, 2, \dots, c_t - 1\}$, as proposed by Wang and van Ryzin (1981).

The multivariate discrete data kernel is defined as $\prod_{t=1}^T L(x_{it}, x_t, \lambda_t)$, with joint density of discrete variables being estimated by $\hat{p}(x_1, \dots, x_T) = n^{-1} \sum_{i=1}^n \prod_{t=1}^T L(x_{it}, x_t, \lambda_t)$. The multivariate kernel for

mixed data is

$$W(x_{ij}, x_j, h, x_{it}, x_t, \lambda_t) = \sum_{i=1}^n \prod_{j=1}^d K_h(x_{ij} - x_j) \prod_{t=1}^T L(x_{it}, x_t, \lambda_t).$$

The local linear estimator for continuous and discrete data suggested by Li and Racine (2004) has the following structure:

$$\begin{aligned} \widehat{s}(x) &= \begin{bmatrix} \widehat{m}(x) \\ \widehat{\theta}(x) \end{bmatrix} = \left[\sum_{i=1}^n W(x_{ij}, x_j, h, x_{it}, x_t, \lambda_t) \begin{pmatrix} 1 & (x_{ij} - x_j) \\ (x_{ij} - x_j) & (x_{ij} - x_j)^2 \end{pmatrix} \right]^{-1} \\ &\quad \times \sum_{i=1}^n W(x_{ij}, x_j, h, x_{it}, x_t, \lambda_t) \begin{pmatrix} 1 \\ (x_{ij} - x_j) \end{pmatrix} y_i, \end{aligned}$$

where $s(x) = (m(x), \theta(x)')'$, $\theta(x) = \nabla \theta(x) = [\partial m(x)/\partial x_1, \dots, \partial m(x)/\partial x_d]'$. The partial derivative is taken only with respect to continuous variables. This estimator has the local constant shape for the discrete variables and local linear shape for the continuous variables.

The local linear smooth backfitting estimator for mixed continuous and categorical data is a projection of the local linear estimator for mixed regressors onto the space of additive functions. The mixed data local linear smooth backfitting estimator $\widehat{m}^*(x)$ is defined as the argument that minimizes the following objective function

$$\begin{aligned} &\int \sum_{i=1}^n \left[y_i - m_0 - \sum_{j=1}^d m_j(x_j) - \sum_{t=1}^T m_t(x_t) - \sum_{j=1}^d \theta_j(x_{ij} - x_j) \right]^2 \\ &\quad \times \prod_{j=1}^d K_h(x_{ij} - x_j) \prod_{t=1}^T L(x_{it}, x_t, \lambda_t) dx, \end{aligned}$$

where the categorical regressors are indexed by t . Derivation of the first order conditions for this setting follows the same logic as for the continuous regressors, where the minimization is performed over m_0 , $m_j(x_j)$ and $m_t(x_t)$ while preserving mean zero restriction, and over $\theta_j(x_j)$ for the continuous components only.

Using similar notation as before

$$\begin{aligned}
\widetilde{m}_j(x_j) &= \frac{n^{-1} \sum_{i=1}^n K_h(x_{ij} - x_j) y_i}{\widehat{p}_j(x_j)} - \sum_{k \neq j}^d \int \widetilde{m}_k(x_k) \frac{\widehat{p}_{jk}(x_j, x_k)}{\widehat{p}_j(x_j)} dx_k \\
&\quad - \sum_{t=1}^T \int \widetilde{m}_t(x_t) \frac{\widehat{p}_{jt}(x_j, x_t)}{\widehat{p}_j(x_j)} dx_t - \sum_{k \neq j}^d \int \widetilde{\theta}_k(x_k) \frac{\widehat{p}_{jk}^k(x_j, x_k)}{\widehat{p}_j(x_j)} dx_k \\
&\quad - \widetilde{m}_0(x) - \widetilde{\theta}_j(x_j) \frac{\widehat{p}_j^j(x_j)}{\widehat{p}_j(x_j)},
\end{aligned}$$

$$\begin{aligned}
\widetilde{m}_j(x_j) &= \frac{n^{-1} \sum_{i=1}^n K_h(x_{ij} - x_j) (x_{ij} - x_j) y_i}{\widehat{p}_j^j(x_j)} - \sum_{k \neq j}^d \int \widetilde{m}_k(x_k) \frac{\widehat{p}_{jk}^j(x_j, x_k)}{\widehat{p}_j^j(x_j)} dx_k \\
&\quad - \sum_{t=1}^T \int \widetilde{m}_t(x_t) \frac{\widehat{p}_{jt}^j(x_j, x_t)}{\widehat{p}_j^j(x_j)} dx_t - \sum_{k \neq j}^d \int \widetilde{\theta}_k(x_k) \frac{\widehat{p}_{jk}^{jk}(x_j, x_k)}{\widehat{p}_j^j(x_j)} dx_k \\
&\quad - \widetilde{m}_0(x) - \widetilde{\theta}_j(x_j) \frac{\widehat{p}_j^{jj}(x_j)}{\widehat{p}_j^j(x_j)},
\end{aligned}$$

where $\widetilde{m}_0(x)$ is the same as in continuous SBE setting. The iterative equations are shown below:

$$\begin{aligned}
\widetilde{m}_j^*(x_j) &= A - \sum_{t \neq j}^T \int \widetilde{m}_t(x_t) \frac{\widehat{p}_{jt}(x_j, x_t)}{\widehat{p}_j(x_j)} dx_t - \widetilde{\theta}_j^*(x_j) C \\
&= A^* - \widetilde{\theta}_j^*(x_j) C
\end{aligned}$$

$$\begin{aligned}
\widetilde{m}_j^*(x_j) &= B - \sum_{t \neq j}^T \int \widetilde{m}_t(x_t) \frac{\widehat{p}_{jt}^j(x_j, x_t)}{\widehat{p}_j^j(x_j)} dx_t - \widetilde{\theta}_j^*(x_j) D \\
&= B^* - \widetilde{\theta}_j^*(x_j) D
\end{aligned}$$

$$\widetilde{\theta}_j^*(x_j) = \frac{A^* - B^*}{C - D}.$$

Iterative equation for discrete regressors $x_t, t = 1, \dots, T$ is

$$\begin{aligned} \widetilde{m}_t^*(x_t) = & \frac{\sum_{i=1}^n L(x_{it}, x_t, \lambda_t) y_i}{\widehat{p}_t(x_t)} - \sum_{j=1}^d \int \widetilde{m}_j(x_j) \frac{\widehat{p}_{jt}(x_j, x_t)}{\widehat{p}_t(x_t)} dx_j - \widetilde{m}_0(x) \\ & - \sum_{\substack{T \\ k \neq t}} \int \widetilde{m}_k(x_k) \frac{\widehat{p}_{kt}(x_k, x_t)}{\widehat{p}_t(x_t)} dx_k - \sum_{j=1}^d \int \widetilde{\theta}_j(x_j) \frac{\widehat{p}_{jt}^j(x_j, x_t)}{\widehat{p}_t(x_t)} dx_j. \end{aligned}$$

The last four equations jointly with the zero-mean condition describe the solution. Analogously to the continuous regressor densities

$$\begin{aligned} \widehat{p}_t(x_t) &= n^{-1} \sum_{i=1}^n L(x_{it}, x_t, \lambda_t), \\ \widehat{p}_{jt}(x_j, x_t) &= n^{-1} \sum_{i=1}^n K_h(x_{ij} - x_j) L(x_{it}, x_t, \lambda_t), \\ \widehat{p}_{jt}^j(x_j, x_t) &= n^{-1} \sum_{i=1}^n K_h(x_{ij} - x_j) L(x_{it}, x_t, \lambda_t) (x_{ij} - x_j). \end{aligned}$$

The algorithm for computation is as follows:

1. Compute the univariate $\widehat{p}_j(x_j), \widehat{p}_t(x_t)$ for all regressors x_j and $x_t, j = 1, \dots, d$, and $t = 1, \dots, T$; compute $\widehat{p}_j^j(x_j), \widehat{p}_j^{jj}(x_j)$ only for continuous components. Compute bivariate densities.
2. Compute univariate unrestricted $\widehat{m}_t(x_t) = (\sum_{i=1}^n L(x_{it}, x_t, \lambda_t) y_i) / \widehat{p}_t(x_t)$ for all discrete variables and pairs $(\widehat{m}_j(x_j), \widehat{\theta}_j(x_j))$ for all continuous data. Save the results as variables m_{old} and θ_{old} .
3. Set the number of smooth backfitting iteration $iter$ to 1.
 - (a) For $j = 1$ compute expressions A*, B*, C, D. Obtain $\widetilde{m}_j^*(x_j)$ and $\widetilde{\theta}_j^*(x_j)$, save as m_{new} and θ_{new} . Repeat this step for the rest of continuous variables $j = 2, \dots, d$. To compute expressions A* and B*, use updated values from m_{new} and θ_{new} for $k < j$. If $k > j$, use corresponding values from m_{old} and θ_{old} .
 - (b) Perform computation for discrete variables in a similar manner, with the conditional mean of categorical x_k in A being taken only over unique categories of x_k .
4. Define a convergence criteria for all j as $\frac{\sum_{i=1}^n [\widetilde{m}_j^{new}(x_j) - \widetilde{m}_j^{old}(x_j)]^2}{\sum_{i=1}^n [\widetilde{m}_j^{old}(x_j)]^2 + \epsilon} < \epsilon$.
5. Set $iter = iter + 1$, Set $m_{old} = m_{new}$ and $\theta_{old} = \theta_{new}$, then go to step 3a. Iterate steps 3a through 5 until the convergence criteria is met.

If $\int \widehat{p}_{j,k}^{j,k}(x_j, x_k) dx_k = \widehat{p}_{j,k}^j(x_j)$ does not hold, it is necessary to include the norming for $\widetilde{m}_j^*(x_j)$ such that $\widetilde{m}_j^{*,n}(x_j) = \widetilde{m}_j^*(x_j) - \int \widetilde{m}_j^*(x_j) \widehat{p}_j(x_j) dx_j$ after every iterative step for each $j = 1, \dots, T$. When the value of overall sum $m_0 + \sum_{j=1}^d m_j(x_j) + \sum_{t=1}^T m_t(x_t)$ is the primary point of interest, this normalization could be omitted as suggested in Mammen et al. (1999).

A.3 Bandwidth selection

Several different methods for selecting bandwidths for SBE estimation were analyzed recently. Mammen and Park (2005) introduced a bandwidth selection method for smooth backfitting based on minimizing penalized sum of squares residuals. They also compared two additional plug-in methods for local linear SBE. It was suggested that the penalized sum of squared residuals was asymptotically equivalent to cross-validation because this holds true for the classical nonparametric regression, as in Hardle et al. (1988).

Leave-one-out least squares cross-validation is recommended for bandwidth selection by Nielsen and Sperlich (2005). It has an implementation advantage for local linear smooth backfitting if the underlying relationship is additive. In this case, the cross-validation procedure can be simplified because the SB estimator has additively separable bias and variance. Bandwidth selection is based on minimizing mean-integrated squared error $MSE(h_1, \dots, h_d, \lambda_1, \dots, \lambda_d) = \int E [\widetilde{m}(x) - m(x)]^2 p(x) dx$. Because of separability of bias and variance, the mean-integrated squared error for overall regression can be defined as

$$MSE(h_1, \dots, h_d, \lambda_1, \dots, \lambda_d) = \sum_{j=1}^{d+T} MSE_j(x_j),$$

where $MSE_j(x_j)$ is mean-integrated squared error for each regression direction $m_j(x_j)$. Thus, the cross-validation problem of minimizing $CV = \sum_{i=1}^n [y_i - \widetilde{m}^{-i}(x)]^2$, where $\widetilde{m}^{-i}(x)$ is the leave-one-out estimator with observation (y_i, x_i) excluded from the computation, can be separated. It reduces to performing an optimal bandwidth search for each directional regression sequentially. Nielsen and Sperlich (2005) suggest taking starting bandwidths h_1, \dots, h_d that undersmooth for each direction and running the initial SBE estimation. Then the cross-validation criteria is minimized with respect to h_j only, where h_j is the bandwidth for direction j , by using a one-dimensional grid search. Bandwidths for all other directions are kept at their starting values. This is repeated for each direction j individually. It is not necessary to use leave-one-out estimators for all other directions $m_k(x_k)$, $k \neq j$, while searching for the optimal bandwidth for the estimation of $m_j(x_j)$. In addition, all $\widetilde{m}_k(x_k)$ do not need to be estimated at their optimal bandwidth. As shown by Mammen and Park (2005), this procedure results in bandwidths that are optimal for the estimation

of the overall regression. If the primary focus of the estimation is accuracy of each single additive component, Mammen and Park (2005) suggest using plug-in bandwidths that minimize average weighted squared error (ASE) for each direction defined as

$$ASE_j(x_j) = n^{-1} \sum_{i=1}^n w_j^{-i}(x_j) \left[\widetilde{m}_j(x_j) - \widetilde{m}_j^{-i}(x_j) \right]^2,$$

where $\widetilde{m}_j^{-i}(x_j)$ is the leave-one-out estimator of $m_j(x_j)$ and w_j is a weight function.

This paper adopts a simpler method for bandwidth selection. Because smooth backfitting requires computing the unrestricted regression estimates, as well as univariate and bivariate densities for continuous and categorical data, we use four different bandwidth selection routines. To estimate densities for categorical variables we use the cross-validation method of Li and Racine (2007), where the bandwidth λ is chosen separately for each regressor to minimize

$$CV_p(\lambda) = \sum_{x_c \in S_c} [\widehat{p}(x_c)]^2 - 2n^{-2} \sum_{i=1}^n \sum_{v \neq i}^n L_{\lambda,iv},$$

where $L_{\lambda,iv}$ is the previously defined kernel with observation $v = i$ excluded from the computation, $S_c = \{0, \dots, c_t - 1\}$ is the support of x_c and c is the category index. For unrestricted regression estimation for categorical variables, the cross-validation of Li and Racine (2007) is employed. Bandwidth is chosen to minimize

$$CV_{reg}(\lambda) = n^{-1} \sum_{i=1}^n [y_i - \widehat{m}_j^{-i}(x_j)]^2$$

for each j , where $\widehat{m}_j^{-i}(x_j)$ is the leave-one-out Nadaraya-Watson estimator of $m_j(x_j)$ defined as $\widehat{m}_j^{-i}(x_j) = \frac{\sum_{v \neq i}^n y_v L_{\lambda,iv}}{\sum_{v \neq i}^n L_{\lambda,iv}}$. For continuous variables the rule-of-thumb bandwidth selection was used both for estimation of unrestricted univariate regression, as well as densities. Namely, the bandwidth for regression estimation was selected as

$$h_j^{reg} = n^{-1/5} \left\{ s^2 2\sqrt{\pi} (\max(x_j) - \min(x_j)) \cdot \left[\frac{1}{n} \sum_{i=1}^n \left(\widehat{b}_3 + \widehat{b}_4 x_j + 0.5\widehat{b}_5 x_j^2 \right)^2 \right]^{-1} \right\}^{1/5},$$

where b_3, b_4 and b_5 are estimates of coefficients in regressing the dependent variable y on $\beta_1 + \beta_2 x_j + \beta_3 (0.5x_j^2) + \beta_4 (\frac{1}{6}x_j^3) + \beta_5 (\frac{1}{24}x_j^4)$, and s^2 is estimated in a usual manner based on the residual estimates of this regression. The bandwidth for density estimation was computed as $hdens_j = (n^{-1/5}) \cdot 1.01a (2\sqrt{\pi})^{-1/5}$, and $a = q_{75}(x_j) - q_{25}(x_j)$, where q_{75} and q_{25} are upper and lower quartiles of x_j , correspondingly.

A.4 References

- Aitchison, J. and C.G.G. Aitken. 1976. Multivariate binary discrimination by the kernel method. *Biometrika* 63, 413-420.
- Baker, P., R. W. Blundell and J. Micklewright. 1989. Modelling household energy expenditures using micro-data. *Economic Journal* 99, 720-738.
- Buja, A., Hastie, T. and R. Tibshirani. 1989. Linear smoothers and additive models. *Annals of Statistics* 17, 453--510.
- Caravan Opinion Research Corporation, 2007. Study #716287.
- Crawley, Drury B, Linda K Lawrie, Curtis O Pedersen, Frederick C Winkelmann, Michael J Witte, Richard K Strand, Richard J Liesen, Walter F Buhl, Yu Joe Huang, Robert H Henninger, Jason Glazer, Daniel E Fisher, Don B Shirey III, Brent T Griffith, Peter G Ellis and Lixing Gu. 2004. Energy Plus: New, Capable, and Linked. *Journal of Architectural and Planning Research* 21, 4 (Winter 2004).
- DOE-2, 1993. BDL Summary Version 2.1E. LBL, 34946, Lawrence Berkeley National Laboratory, Berkeley, CA
- Fan, Y., Q. Li, and A. Weersink. 1996. Semiparametric estimation of stochastic production frontier models. *Journal of Business and Economic Statistics* 14, 460-468.
- García-Cerruti, L. 2000. Estimating elasticities of residential energy demand from panel county data using dynamic random variables models with heteroskedastic and correlated error terms. *Resource and Energy Economics* 22, 355-366.
- Gauss 9.0 User Guide. 2009. Aptech Systems, Inc.
- Hall, P., J.S. Racine, and Q. Li. 2004. Cross-Validation and the Estimation of Conditional Probability Densities. *Journal of the American Statistical Association* 99, 1015-1026.
- Halvorsen, B. and B. Larsen. 2001. The flexibility of household electricity demand over time." *Resource and Energy Economics* 23, 1-18.
- Härdle, W., P. Hall and J. S. Marron. 1988. How far are automatically chosen regression smoothing parameters from their optimum? (with discussion). *Journal of the American Statistical Association* 83, 86--101.
- Holtedahl, P. and F. Joutz. 2004. Residential electricity demand in Taiwan. *Energy Economics* 26, 201-224.
- Kamerschen, D. and D. Porter. 2004. The demand for residential, industrial and total electricity, 1973-1998. *Energy Economics* 26, 87-100.
- Labandeira, X., J. M. Labeaga and M. Rodriguez. 2006. A residential energy demand system for Spain. *The Energy Journal* 27, 87-111.

- Labandeira, X., J.M. Labeaga and M. Rodríguez. 2004. Microsimulating the effects of household energy price changes in Spain. *Fondazione Eni Enrico Mattei, Working Paper # 161.*
- Larsen, B. and R. Nesbakken. 2004. Household electricity end-use consumption: results from econometric and engineering models. *Energy Economics* 26, 179-200.
- Li, Q. and J. Racine. 2003. Nonparametric estimation of distributions with categorical and continuous data. *Journal of Multivariate Analysis* 86, 266-292.
- Li, Q. and J. S. Racine. 2004. Cross-validated local linear nonparametric regression. *Statistica Sinica* 14, 485-512.
- Li, Q. and J. S. Racine. 2007. *Nonparametric econometrics: theory and practice.* Princeton University Press, 768.
- Linton, O. and J.P. Nielsen. 1995. A kernel method of estimating structured nonparametric regression based on marginal integration, *Biometrika* 82, 93-100.
- Lutzenhiser, L. 1993. Social and behavioral aspects of energy use. *Annual Review of Energy Economics* 18, 247-89.
- Madlener, R. 1996. Econometric analysis of residential energy demand: a survey. *Journal of Energy Literature* 2, 3-32.
- Mammen, E. and B. U. Park. 2005. Bandwidth selection for smooth backfitting in additive models. *Annals of Statistics* 33, 1260--1294.
- Mammen, E., Linton, O. and J. P. Nielsen. 1999. The existence and asymptotic properties of a backfitting projection algorithm under weak conditions. *Annals of Statistics* 27, 1443--1490.
- Martins-Filho, C. 2006. *Applied microeconomics: course notes.* Department of Economics, Oregon State University.
- Narayan, P. and R. Smyth. 2005. The residential demand for electricity in Australia: an application of the bounds testing approach to cointegration. *Energy Policy* 33, 467-474.
- Nesbakken, R. 2001. Energy consumption for space heating: discrete-continuous approach. *Scandinavian Journal of Economics* 103, 165-184.
- Newey, W. 1994. Kernel estimation of partial means. *Econometric Theory* 10, 233-253.
- Nielsen J.P. and S. Sperlich. 2005. Smooth backfitting in practice. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 67, 43-61.
- Opsomer, J. D. and D. Ruppert. 1997. Fitting a bivariate additive model by local polynomial regression. *Annals of Statistics* 25, 186--211.
- Opsomer, J. D. 2000. Asymptotic properties of backfitting estimators. *Journal Multivariate Analysis* 73, 166--179.

- Racine, J. S. and Q. Li. 2004. Nonparametric estimation of regression functions with both categorical and continuous data. *Journal of Econometrics* 119, 99-130.
- Racine, J.S., Q. Li., and X. Zhu. 2004. Kernel estimation of multivariate conditional distributions. *Annals of Economics and Finance* 5, 211-235.
- Schmalensee, R. and T. M. Stoker. 1999. Household gasoline demand in the United States. *Econometrica* 67, 645-662.
- Severini T.A., Wong W.H., Profile Likelihood and Conditionally Parametric Models *Annals of Statistics*, Vol. 20, No. 4 (Dec., 1992), pp.1768-1802
- U.S. Department of Energy. 2005. ENERGY PLUS, Input Output Reference.
- U.S. Department of Energy. 2009. Energy Information Administration (DOE/EIA). Annual energy outlook 2009 with projections to 2030. DOE/EIA-0383 (2009).
- U.S. Department of Energy. 2009. Energy Information Administration (DOE/EIA). Residential Energy Consumption Survey 2005.
- Wang, M.C., and J. Van Ryzin. 1981. A class of smooth estimators for discrete distributions. *Biometrika* 68, 301-309.
- Yatchew, A. and J. No. 2001. Household gasoline demand in Canada. *Econometrica* 69, 1697-1709.

Appendix B

Complete Set of Graphical Results

Appendix B

Complete Set of Graphical Results

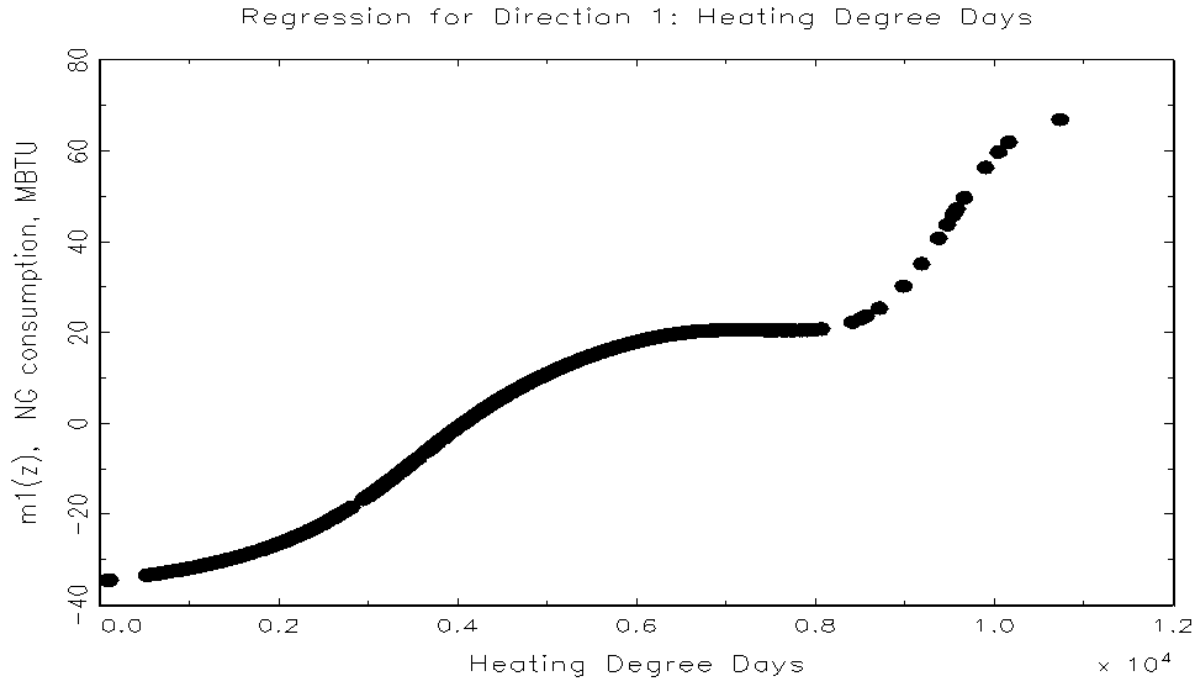


Figure B.1. Heating Degree Days: Base 65, 01 to 12-2005 (Inoculated)

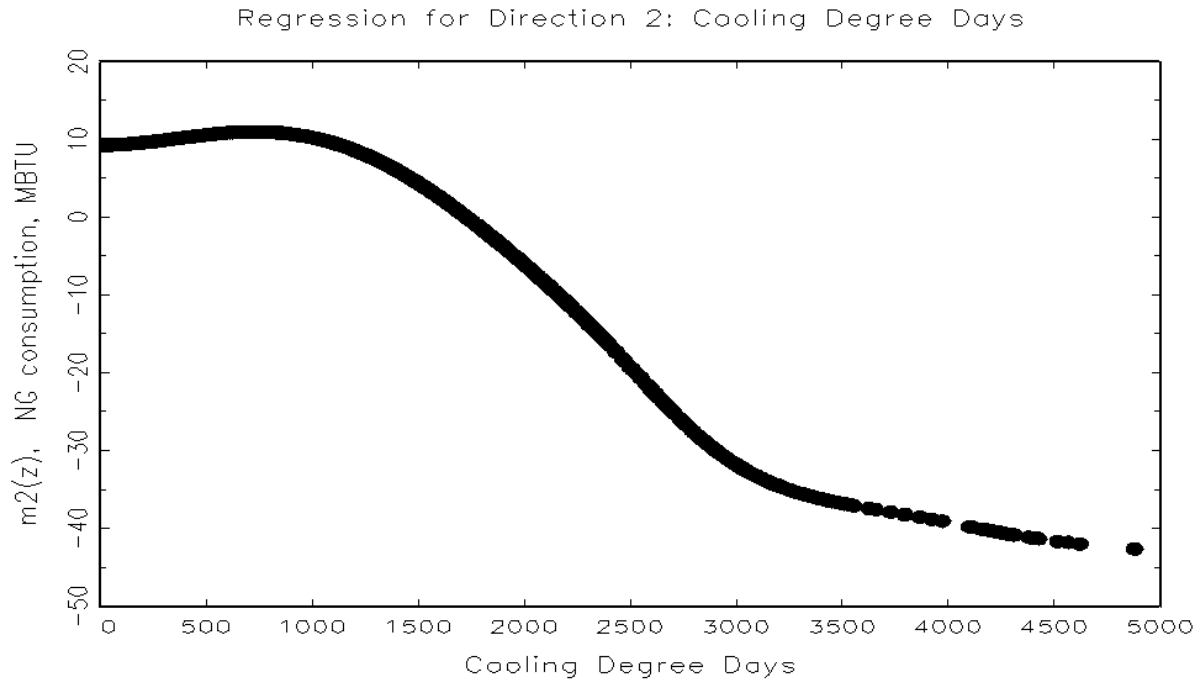


Figure B.2. Cooling DegreeDays: Base 65, 01 to 12-2005 (Inoculated)

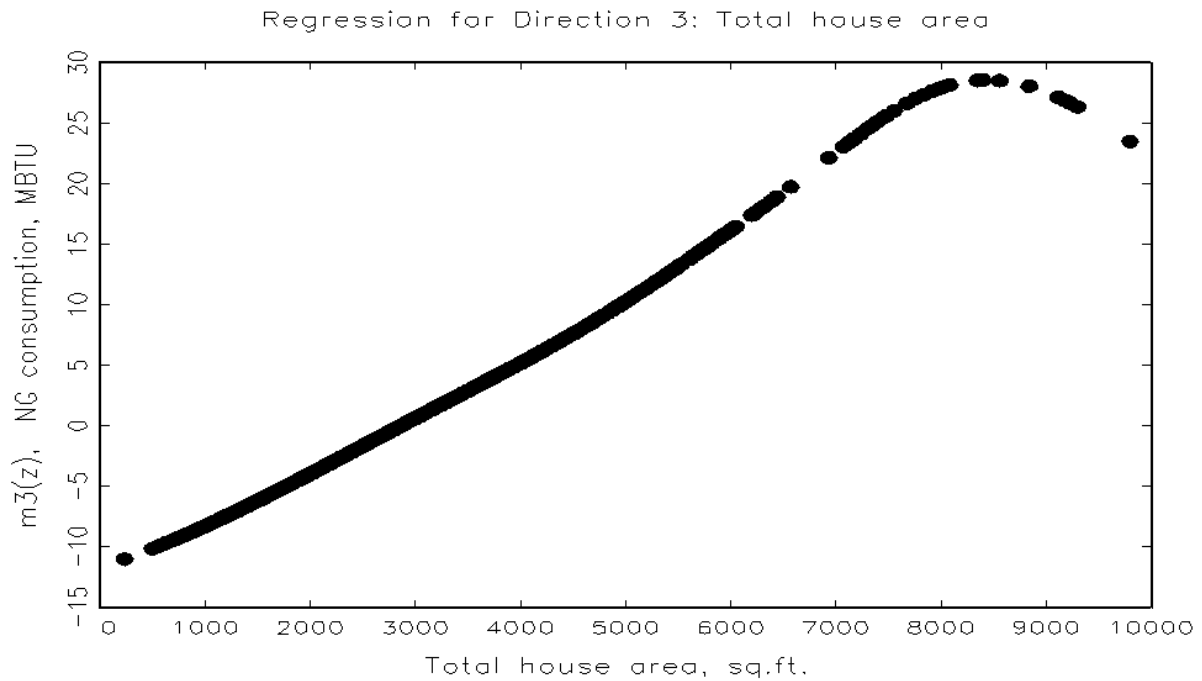


Figure B.3. Total House Area

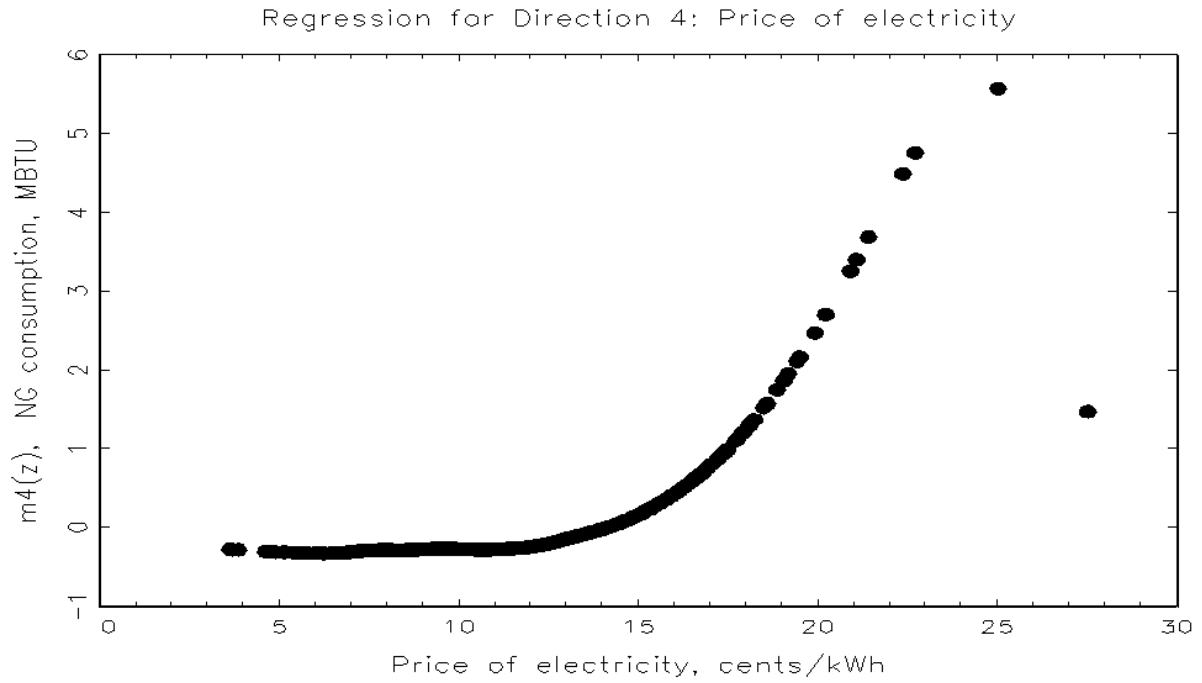


Figure B.4. Price of Electricity

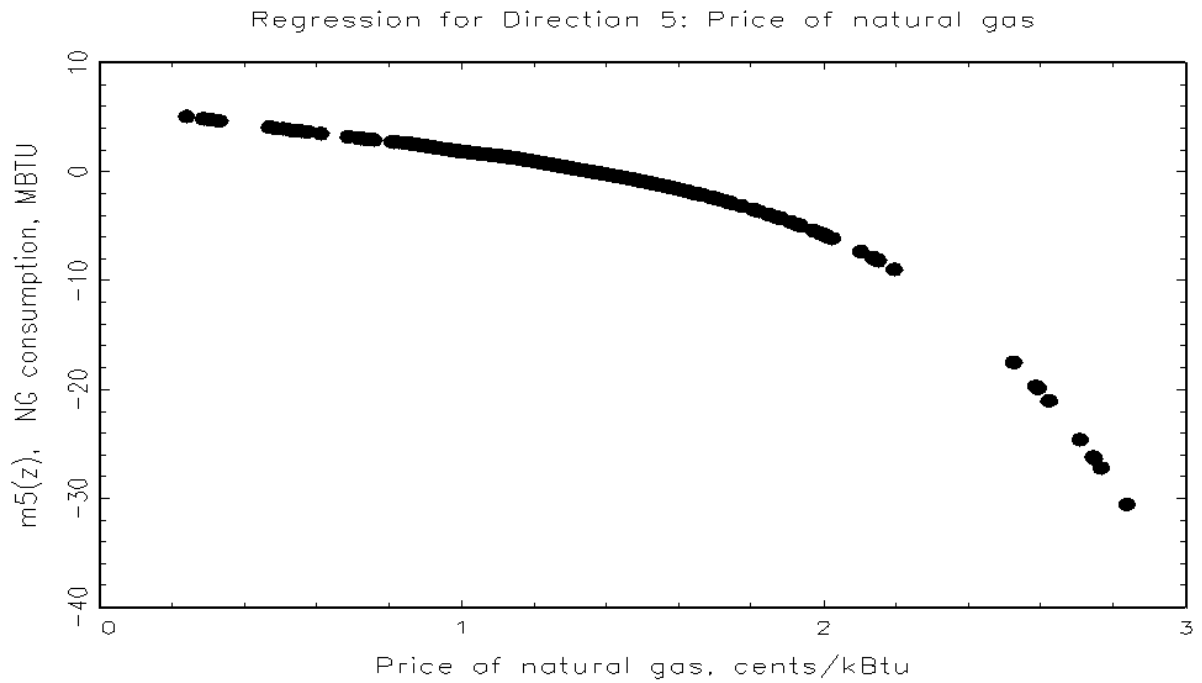


Figure B.5. Price of Natural Gas, Cents/kBtu

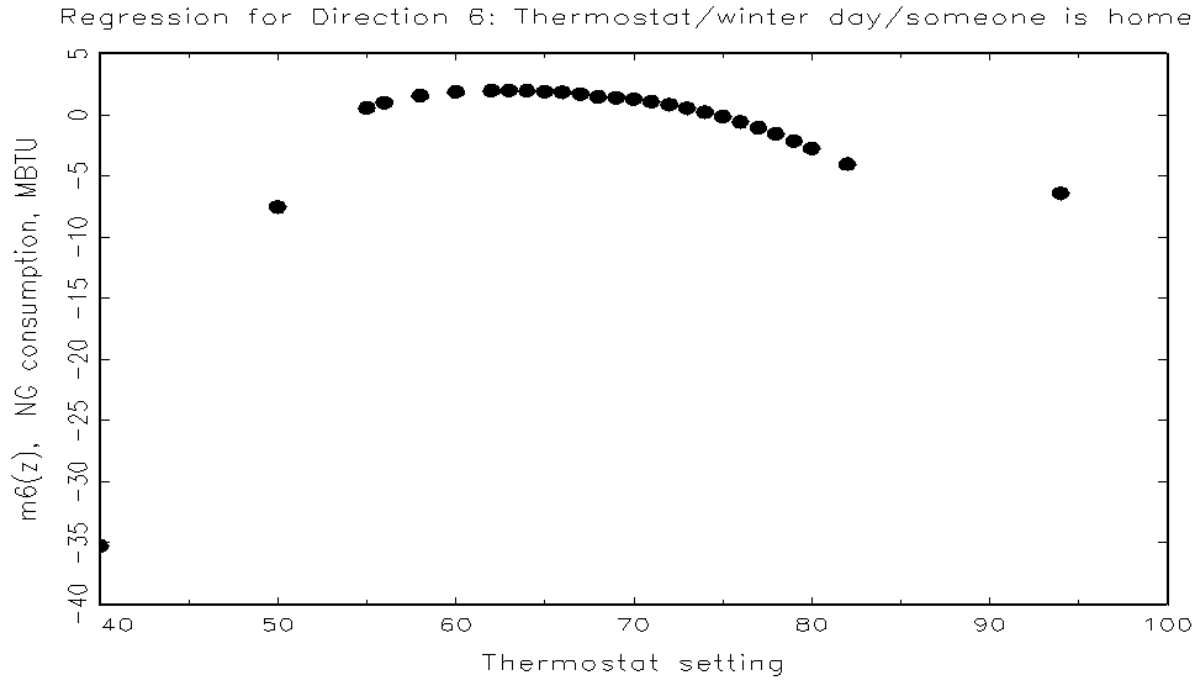


Figure B.6. Setting During the Winter Day When Someone is Home

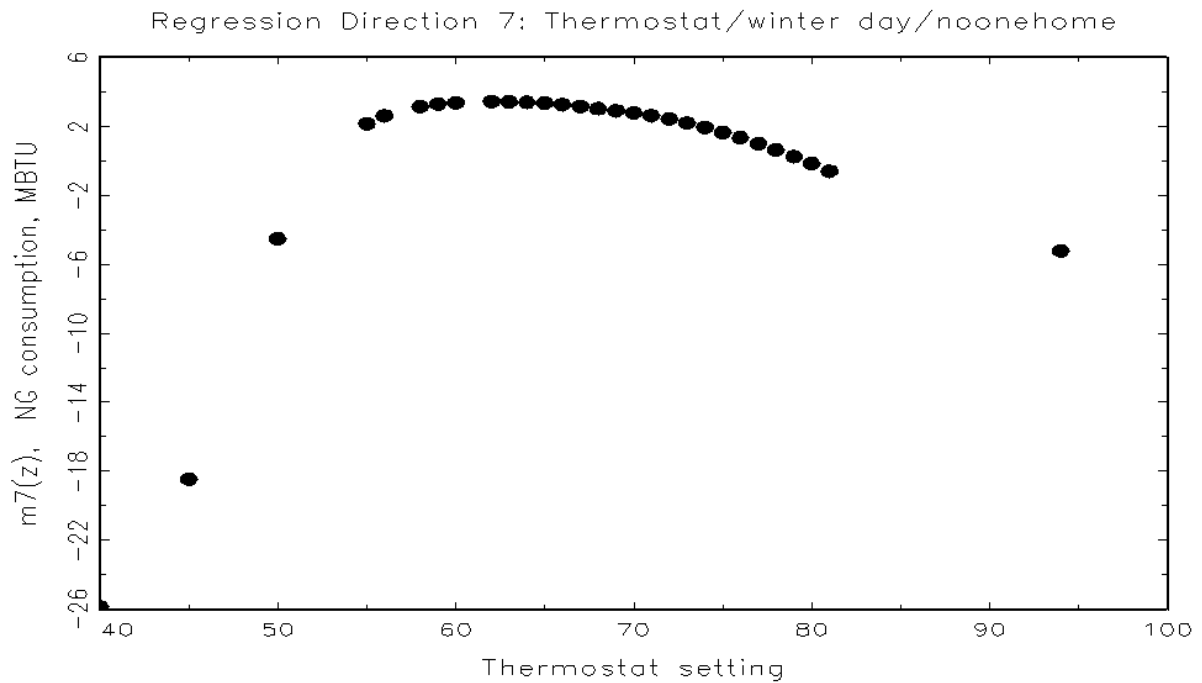


Figure B.7. Setting During the Winter Day When No One is Home

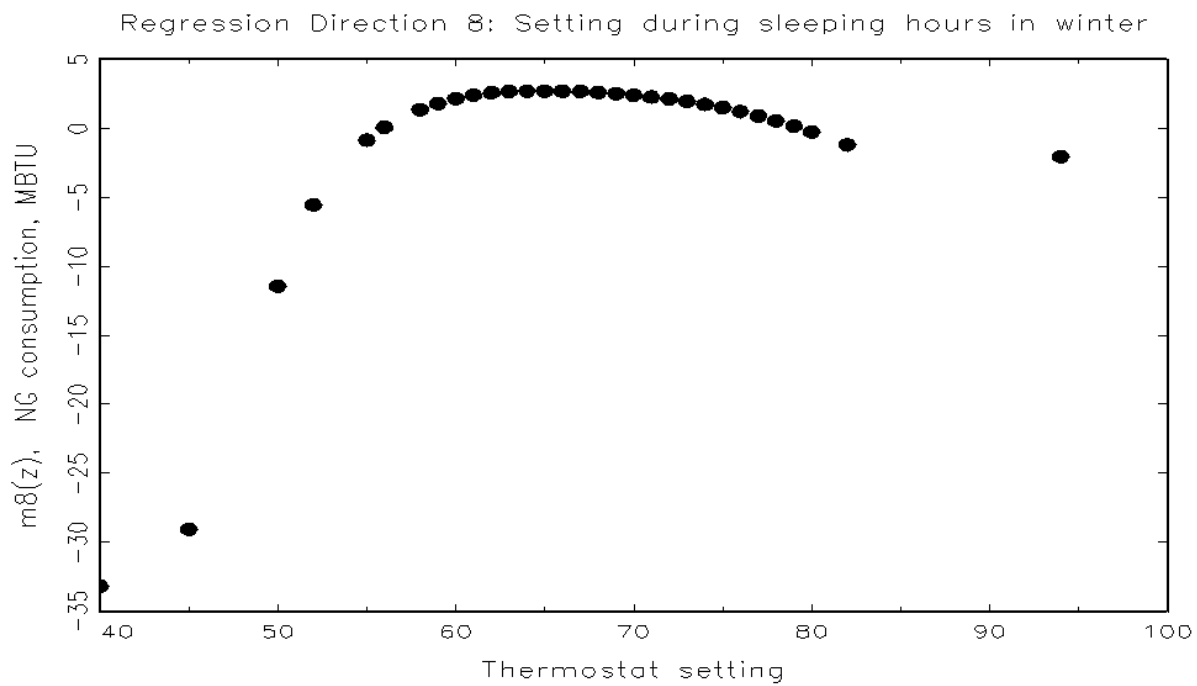


Figure B.8. Setting During Sleeping Hours in Winter



Figure B.9. Exterior Wall Construction Material

- 0 Indescribable
- 1 Brick
- 2 Wood
- 3 Siding (Aluminum, vinyl, or steel)
- 4 Stucco
- 5 Composition (Shingle)
- 6 Stone
- 7 Concrete or concrete block
- 8 Glass
- 9 Other

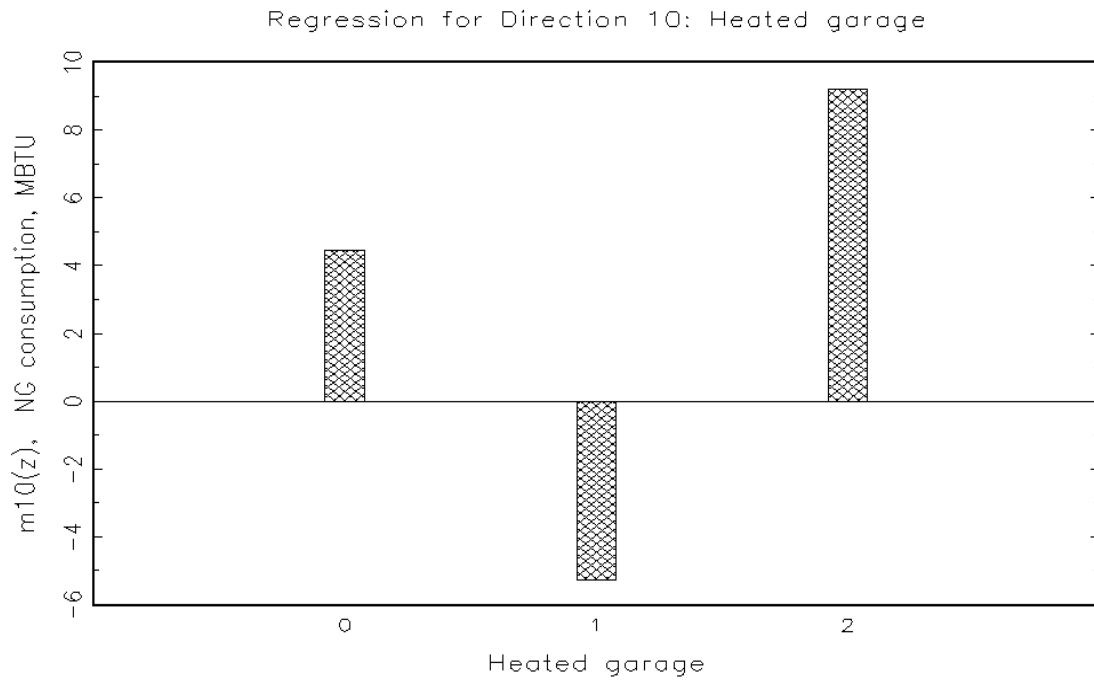


Figure B.10. Is the Garage Heated

- 0 No garage
- 1 Not heated
- 2 Yes

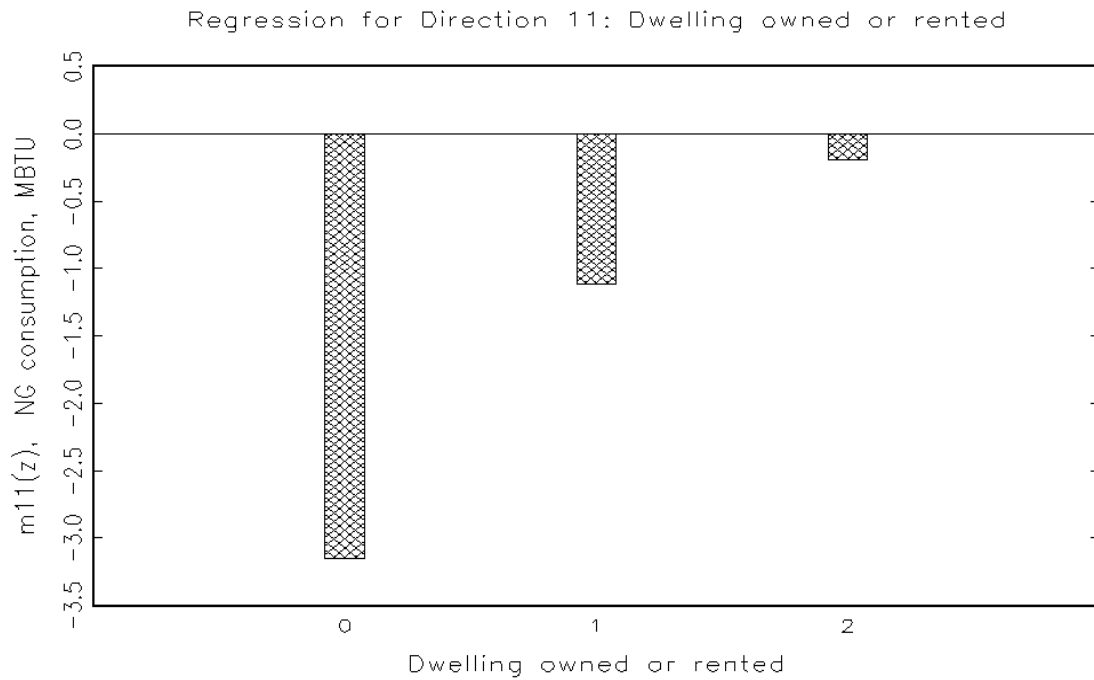


Figure B.11. Dwelling Owned or Rented

- 0 Own
- 1 Rent
- 2 Occupied w/out payment

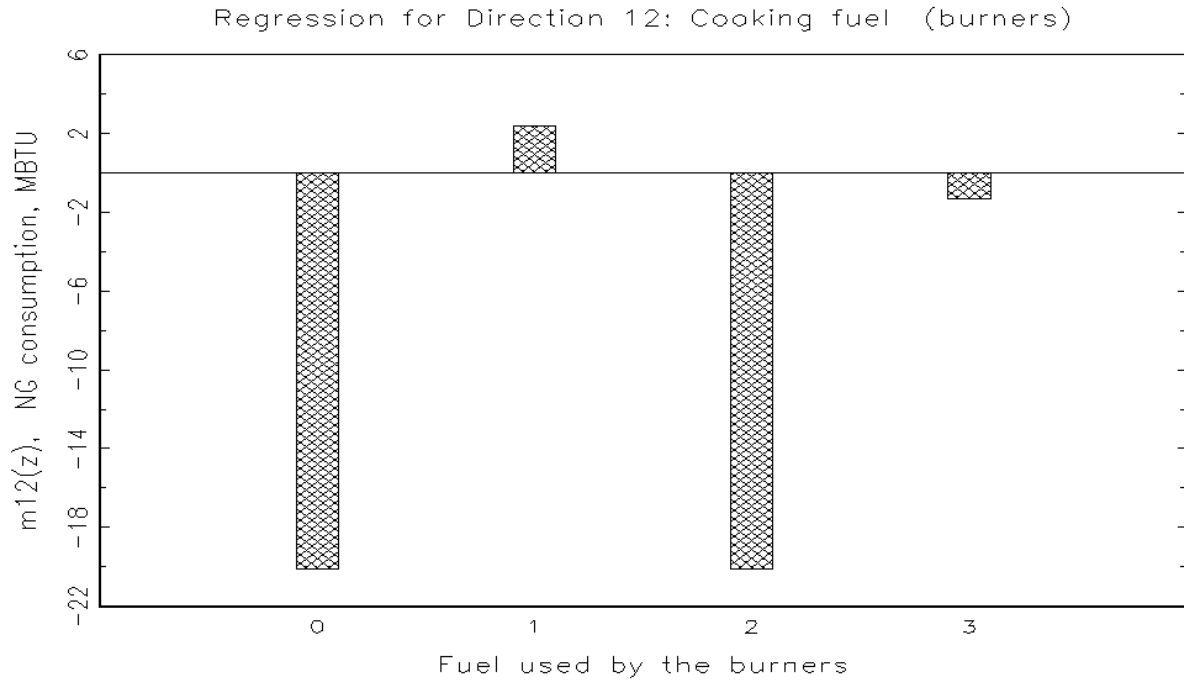


Figure B.12. Fuel Used by the Burners

- 0 Some other fuel
- 1 Natural gas from underground pipes,
- 2 Propane (bottled gas), or
- 3 Electricity

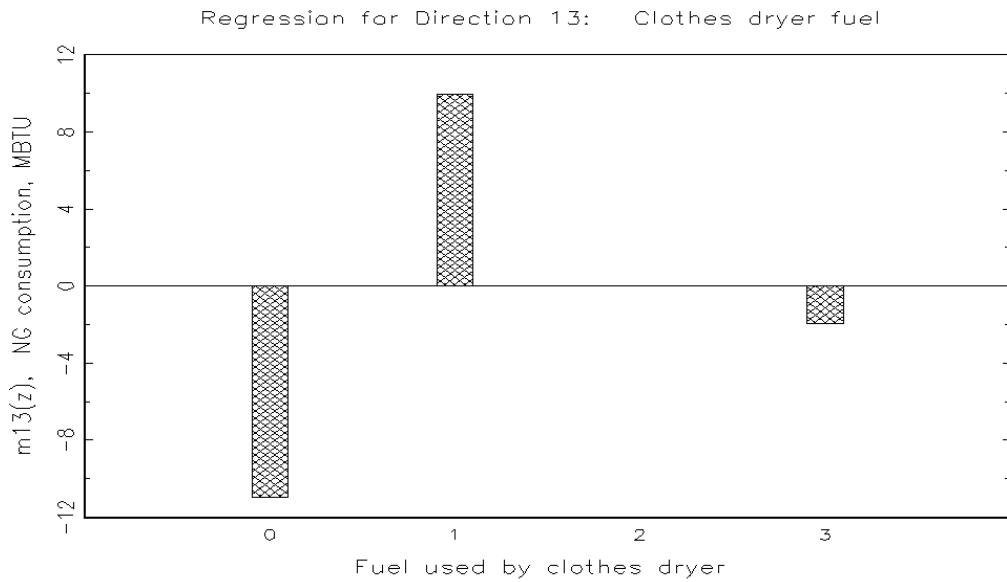


Figure B.13. What Fuel Does Clothes Dryer Use

- 0 No dryer
- 1 Natural gas from underground pipes,
- 2 Propane (bottled gas), or
- 3 Electricity

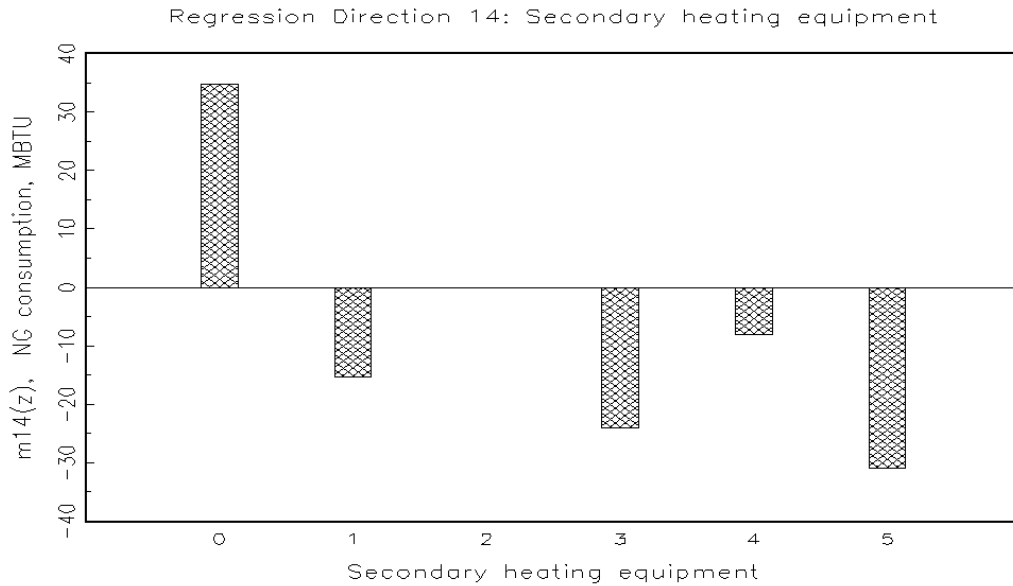


Figure B.14. Combined All Secondary Heating Equipment

- 0 No secondary heating equipment
- 1 Central warm-air furnace with ducts to individual rooms other than a heat pump
- 2 Steam/hot water system with radiators/convectors in each room or pipes in the floor or walls
- 3 Built-in floor/wall pipeless furnace
- 4 Built-in room heater burning gas, oil, or kerosene
- 5 Cooking stove used to heat your home as well as to cook

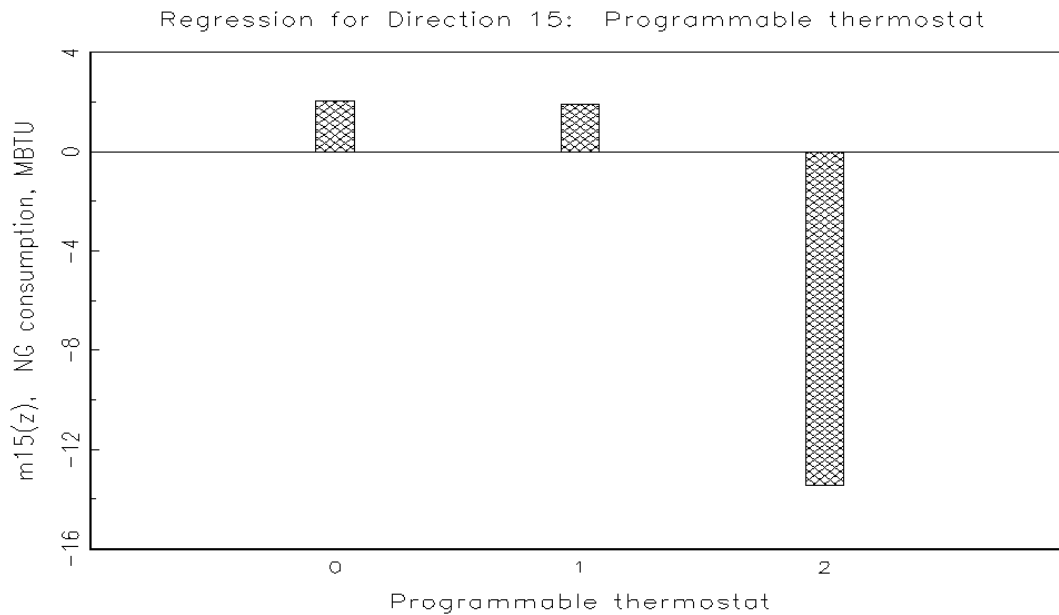


Figure B.15. Is That Thermostat Programmable

- 0 No
- 1 Yes
- 2 No thermostat

Regression for Direction 16: Progr. thermostat lowers heat at night

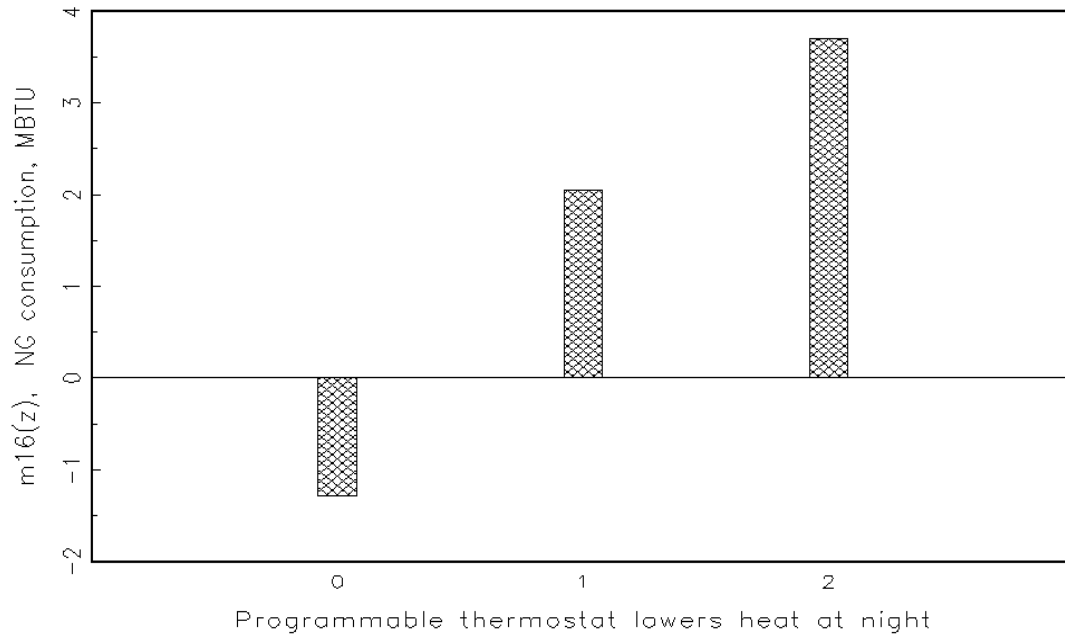


Figure B.16. Programmable Thermostat Lowers Heat at Night

- 0 No
- 1 Yes
- 2 No thermostat or not programmable

Regression for Direction 17: Progr. thermostat lowers heat/ day

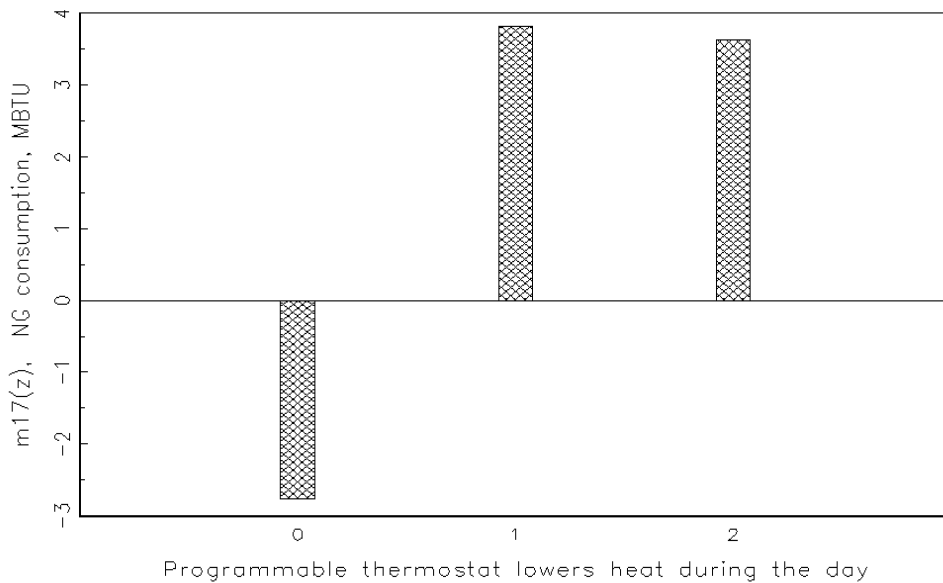


Figure B.17. Programmable Thermostat Lowers Heat During the Day

- 0 No
- 1 Yes
- 2 No thermostat or not programmable

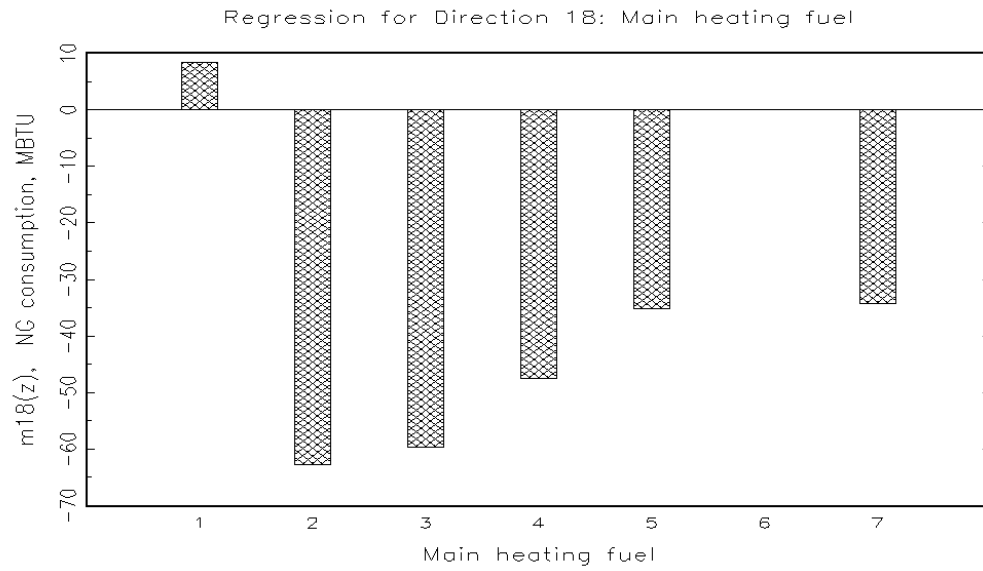


Figure B.18. Main Fuel Used for Heating Home

- 0 Natural gas from underground pipes
- 1 Propane (bottled gas)
- 2 Fuel oil
- 3 Kerosene
- 4 Electricity
- 5 Wood
- 6 Solar

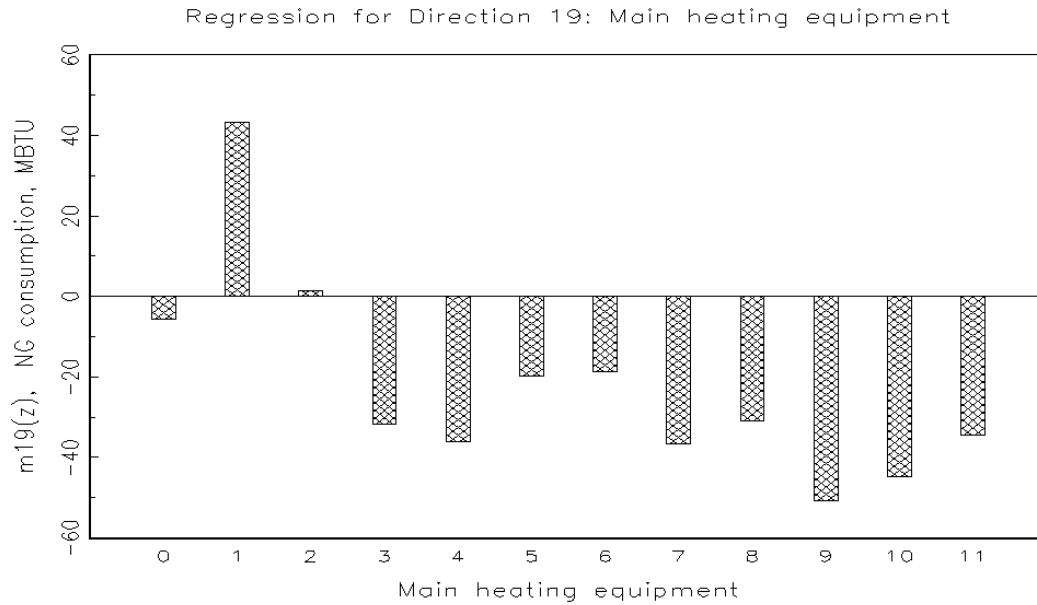


Figure B.19. Type of Heating Equipment Provides the Heat

- 0 No heating equipment used
- 1 Steam/hot water system with radiators/convectors in each room or pipes in the floor or walls
- 2 Central warm-air furnace with ducts to individual rooms other than a heat pump
- 3 Heat pump
- 4 Built-in electric units in each room installed in walls, ceiling, baseboard, or floor
- 5 Built-in floor/wall pipeless furnace
- 6 Built-in room heater burning gas, oil, or kerosene
- 7 Heating stove burning wood, coal, or coke
- 8 Fireplace
- 9 Portable electric heaters
- 10 Portable kerosene heaters
- 11 Cooking stove that is used to heat your home as well as to cook

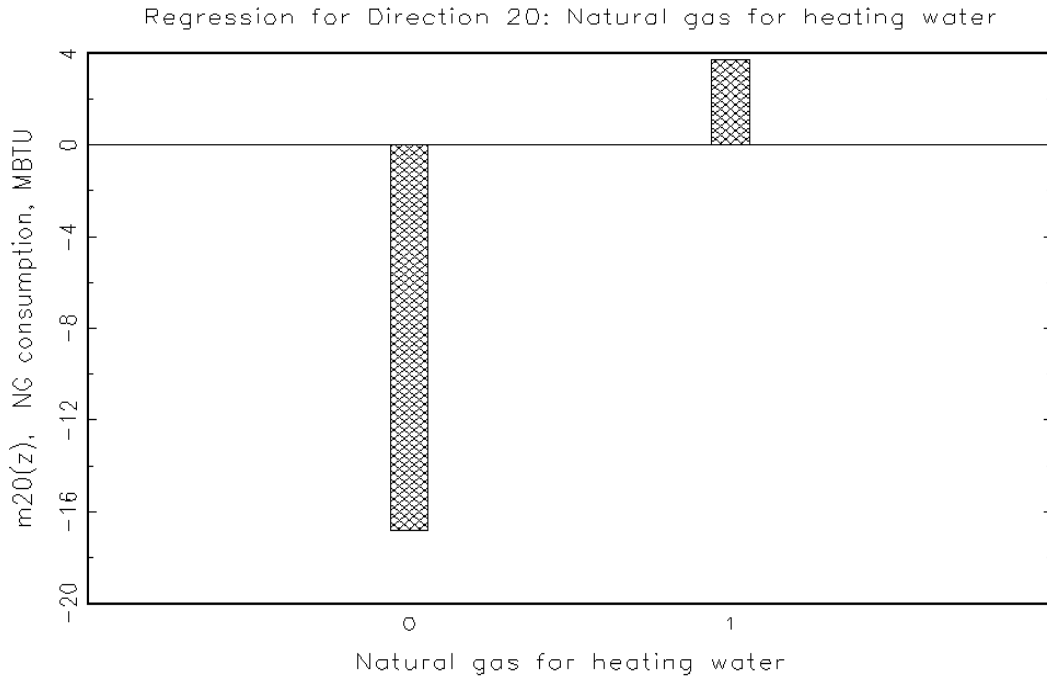


Figure B.20. Natural Gas Used for H2O
 0 No
 1 Yes

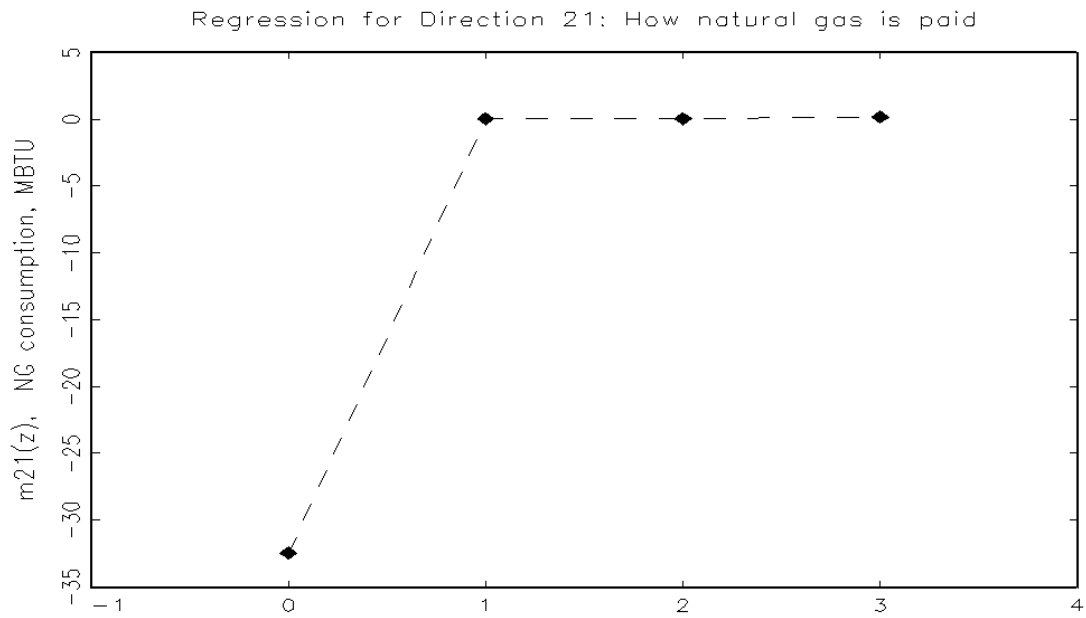


Figure B.21. How Natural Gas is Paid
 0 HH pays all
 1 All in rent/fee
 2 Some paid, some included in rent
 3 Other

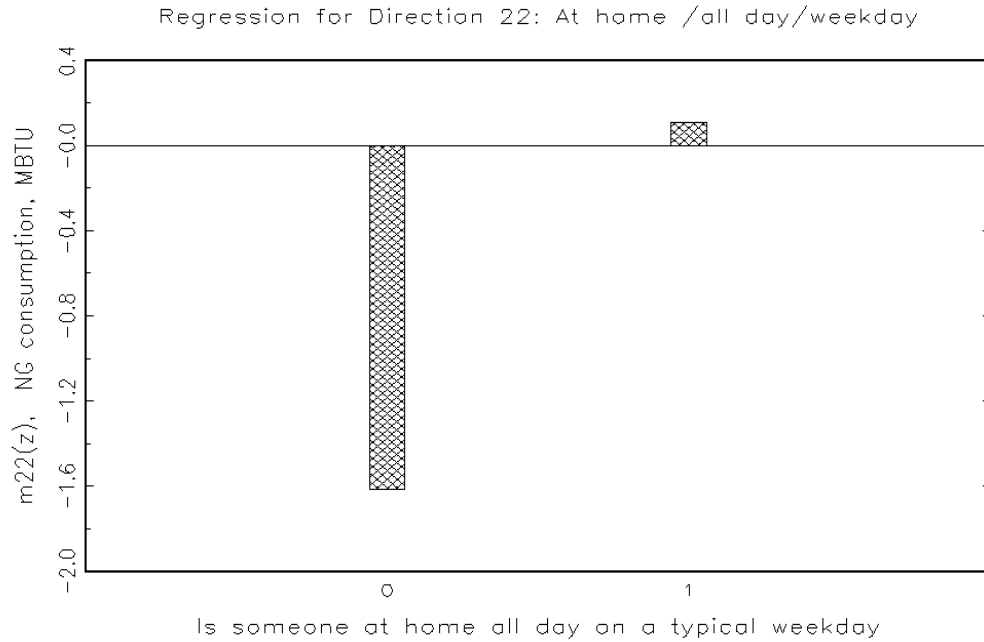


Figure B.22. Is Someone at Home All Day on a Typical Weekday

- 0 No
- 1 Yes

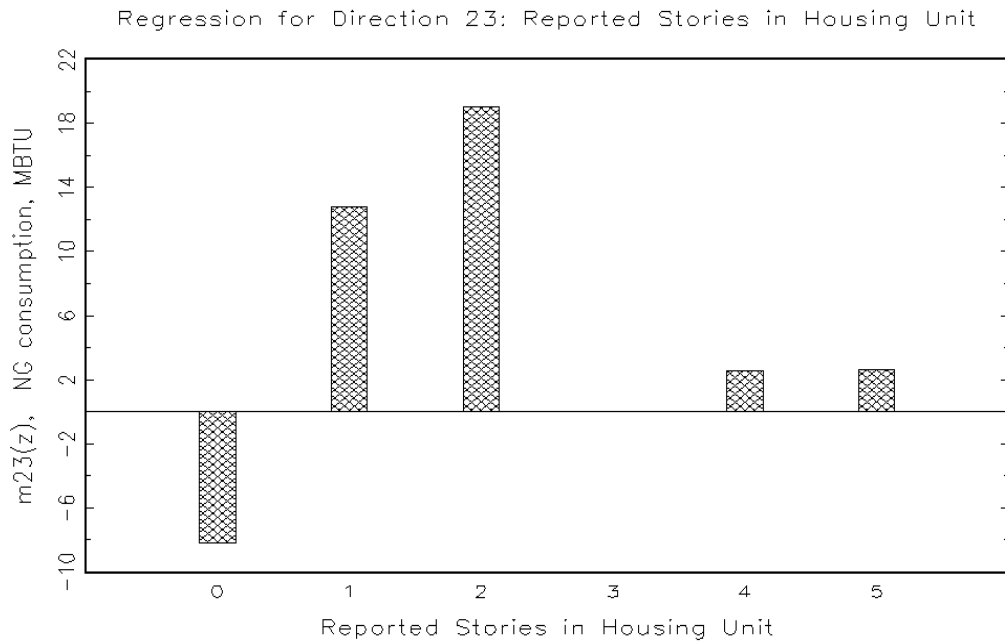


Figure B.23. Reported Stories in Housing Unit

- 0 One story
- 1 Two stories
- 2 Three stories
- 3 Four or more
- 4 Split level
- 5 Other

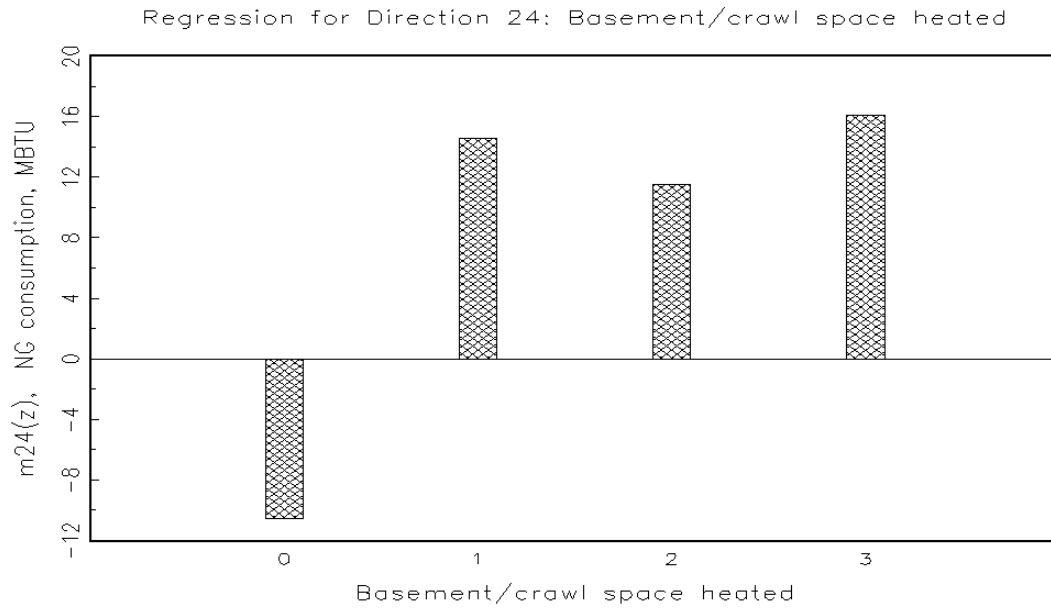


Figure B.24. Basement/Crawl Space Heated

- 0 no basement
- 1 none
- 2 part
- 3 all

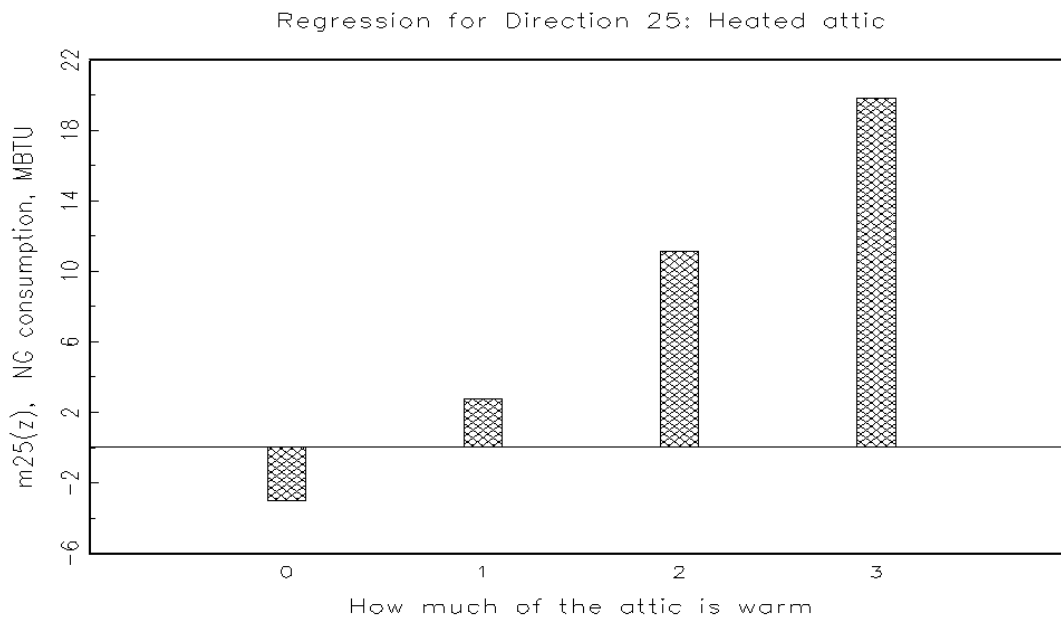


Figure B.25. How Much of the Attic is Warm

- 0 no attick
- 1 none
- 2 part
- 3 all

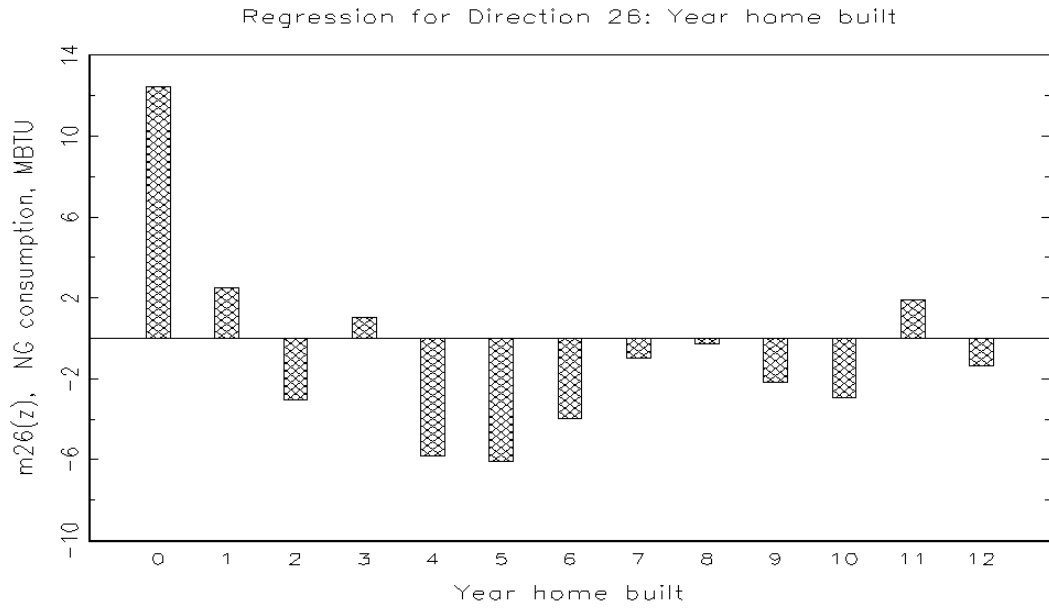


Figure B.26. Year Home Built

- | | |
|---------------|-----------|
| 0 before 1940 | 7 1990-94 |
| 1 1940-49 | 8 1995-99 |
| 2 1950-59 | 9 2000-02 |
| 3 1960-69 | 10 2003 |
| 4 1970-79 | 11 2004 |
| 5 1980-84 | 12 2005 |
| 6 1985-89 | |

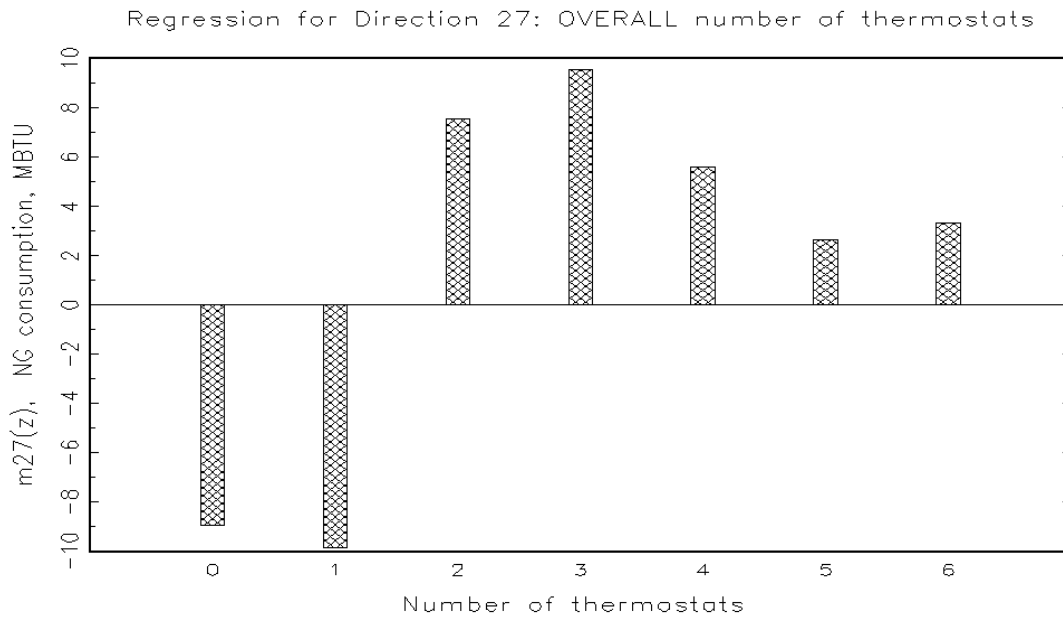


Figure B.27. How Many Thermostats Overall

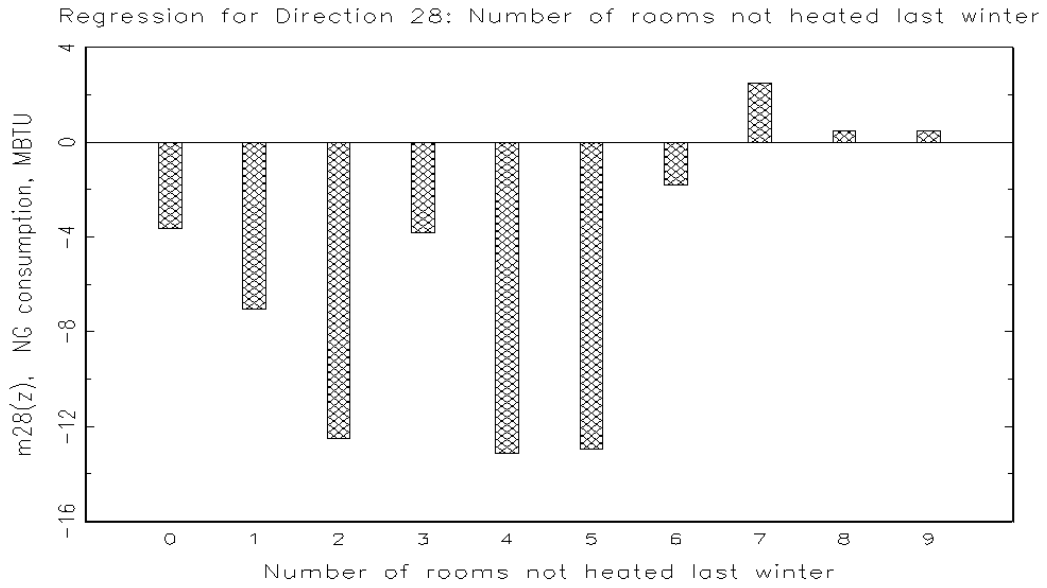


Figure B.28. Number of Rooms Not Heated Last Winter

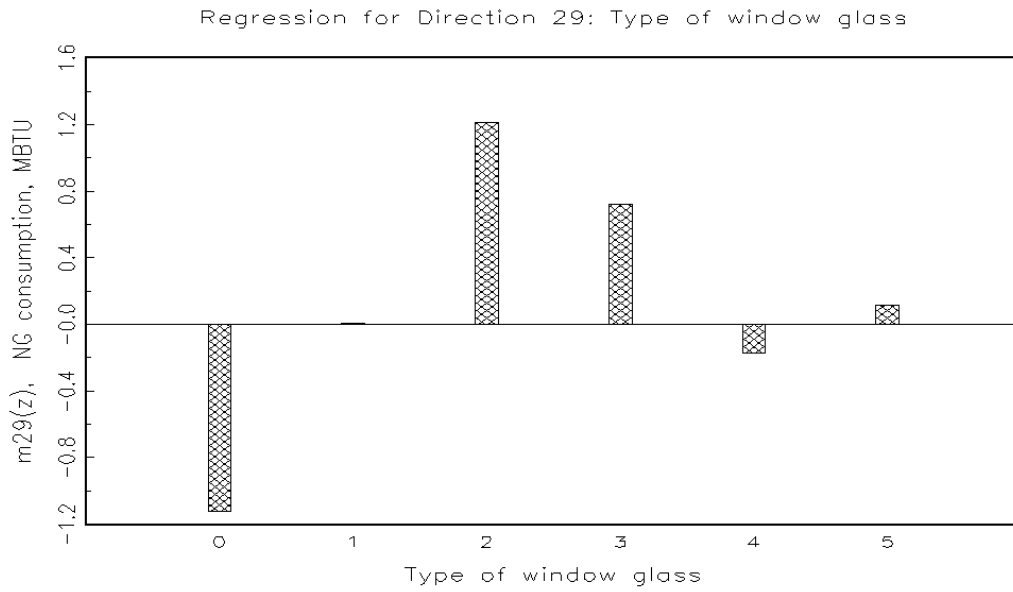


Figure B.29. Type of Window Glass

- 0 Single-pane glass
- 1 Double-pane glass
- 2 Double-pane glass with low-e coating
- 3 Triple-pane glass
- 4 Triple-pane glass with low-e coating

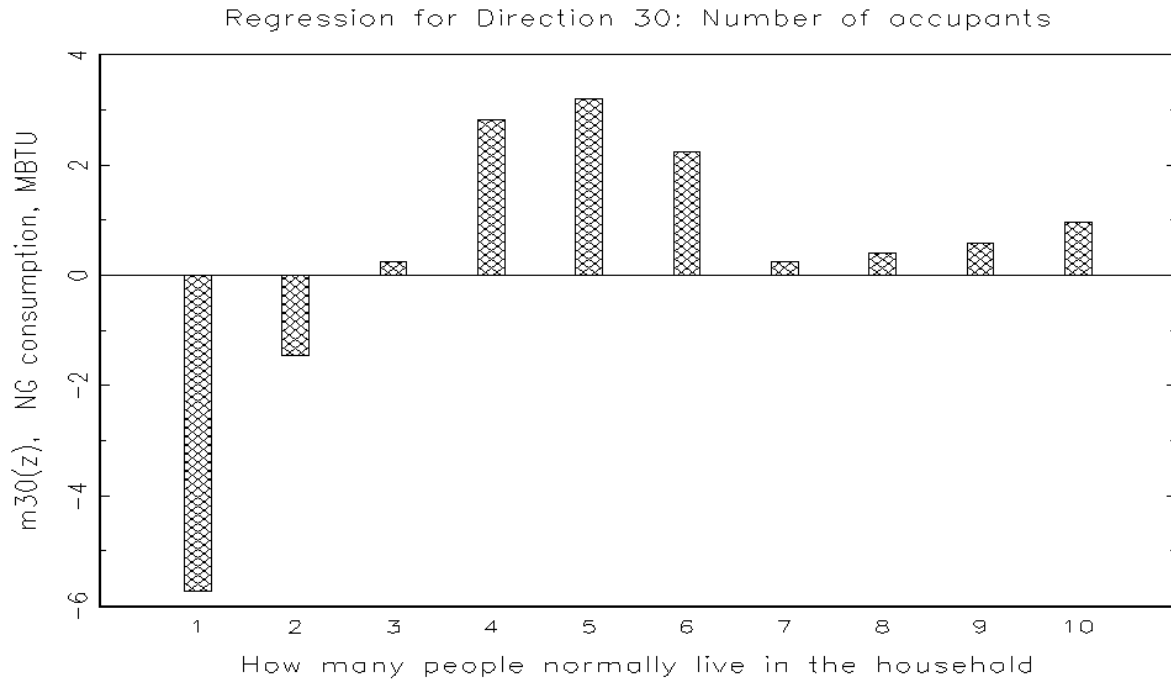


Figure B.30. How Many People Normally Live In This Household
0= none, up to 10

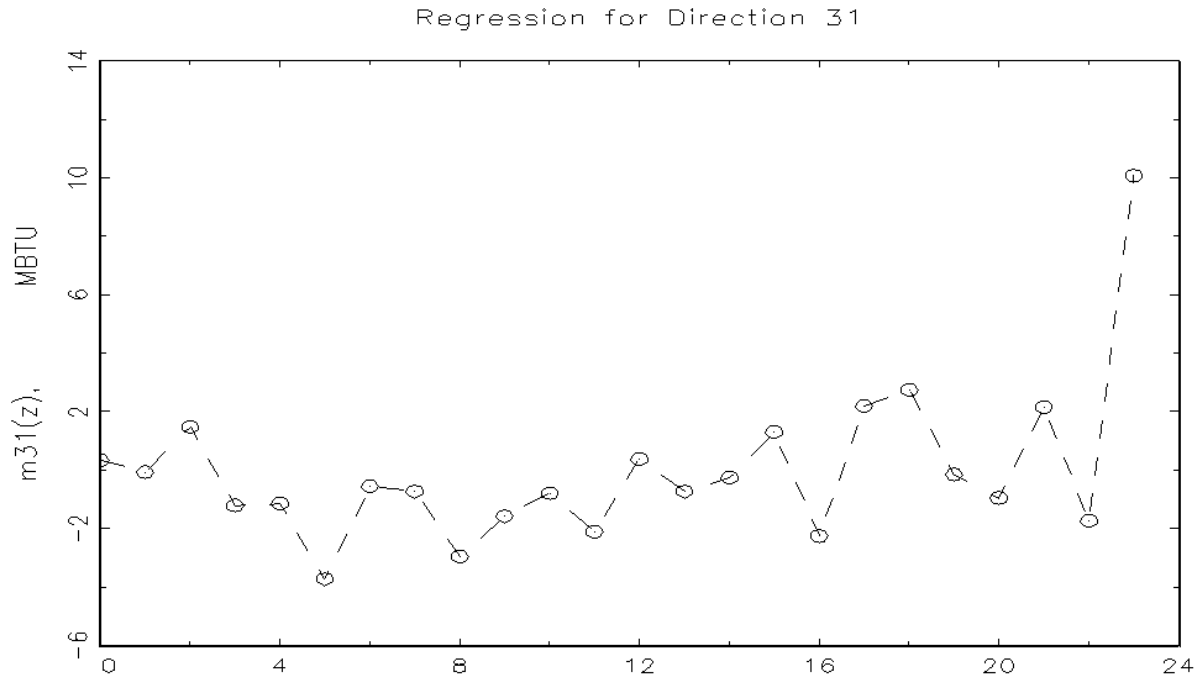


Figure B.31. Total Combined Income in the Past 12 Months

- 0 Less than \$2,500
- 1 \$2,500 to \$4,999
- 2 \$5,000 to \$7,499
- 3 \$7,500 to \$9,999
- 4 \$10,000 to \$14,999
- 5 \$15,000 to \$19,999
- 6 \$20,000 to \$24,999
- 7 \$25,000 to \$29,999
- 8 \$30,000 to \$34,999
- 9 \$35,000 to \$39,999
- 10 \$40,000 to \$44,999
- 11 \$45,000 to \$49,999
- 12 \$50,000 to \$54,999
- 13 \$55,000 to \$59,999
- 14 \$60,000 to \$64,999
- 15 \$65,000 to \$69,999
- 16 \$70,000 to \$74,999
- 17 \$75,000 to \$79,999
- 18 \$80,000 to \$84,999
- 19 \$85,000 to \$89,999
- 20 \$90,000 to \$94,999
- 21 \$95,000 to \$99,999
- 22 \$100,000 to \$119,999
- 23 \$120,000 or more



Pacific Northwest
NATIONAL LABORATORY

*Proudly Operated by **Battelle** Since 1965*

902 Battelle Boulevard
P.O. Box 999
Richland, WA 99352
1-888-375-PNNL (7665)

www.pnl.gov



U.S. DEPARTMENT OF
ENERGY