

**Modelo de demanda de energía eléctrica
residencial rural para el Departamento de
Nariño obtenido a través de la información de la
encuesta PERS-Nariño 2013**

**PERS-Nariño¹
Wilson Olmedo Achicanoy Martínez²**

15 de febrero de 2014

¹Plan de Energización Rural Sostenible para el Departamento de Nariño PERS-Nariño.
<http://pers.udenar.edu.co/>.

²Profesor asistente del Departamento de Electrónica, Facultad de Ingeniería, Universidad de Nariño. e-mail: wilachic@udenar.edu.co

Índice

1. Variables	4
2. Regresión	16
2.1. Mod_1	18
2.2. $Mod_{2,1}$, $Mod_{2,2}$ y $Mod_{2,3}$	20
2.3. Mod_3	26
2.4. Mod_4	28
2.5. Mod_5	31
3. Calidad de la regresión	34
3.1. Error Estándar SE (Standar Error)	34
3.2. Estadísticas t y p	35
3.3. Coeficientes de determinación R^2 y R_a^2 y estadística F	37
3.4. Raíz del error cuadrático medio $RMSE$ (Root Mean Squared Error)	38
3.5. Análisis de residuos	39
3.6. Distancia de Cook D_i	46
3.7. Colinealidad	46
3.8. Selección de modelos	50
4. Predicción	51
4.0.1. Modelo Tipo 1 (T1)	51
4.0.2. Modelo Tipo 2 (T2)	53
4.0.3. Modelo Tipo 3 (T3)	54
5. Conclusiones y recomendaciones	55

Introducción

En este documento se resume el proceso de modelación de la demanda de la energía eléctrica residencial para las viviendas rurales del Departamento de Nariño y que utiliza la información suministrada por la encuesta realizada por el PERS-Nariño durante los meses de junio, julio y agosto del año 2013 (la información de la encuesta está disponible en el Sistema de Información del Pers-Nariño: SIPERSN¹). Este proceso de modelación es una metodología de exploración y análisis de modelos econométricos de tipo transversal, estáticos y lineales que se basan en algoritmos de regresión lineal robusta [1][2][3] y que utilizan las variables más representativas de los procesos de consumo de la energía eléctrica en las viviendas rurales del Departamento de Nariño. Esta metodología propone modelos generales y categóricos que hacen uso de variables ficticias (*dummy*).

En la literatura se reporta una tendencia de modelación de la demanda residencial de la energía eléctrica muy particular para cada país. Las fuentes de información que se han utilizado para estudiar esta demanda son encuestas que se han realizado a la población de interés y que en varios países se repiten en periodos de largo plazo. Dos ejemplos representativos de estos estudios son: (i) El que explica la modelación la demanda residencial de la energía eléctrica en Chile [4] y (ii) el análisis micrométrico de la demanda eléctrica residencial de corto plazo en España [5]. Estos dos trabajos se centran en el cálculo de la demanda y las elasticidades con respecto al precio de la energía eléctrica y utilizando modelos transversales lineales y paramétricos no lineales. Aunque la generalización o comparación entre modelos de demanda de diferentes países no es clara y definitiva, en países como Chile ya se cuenta con una base amplia de modelos que permiten hacer evaluaciones de elasticidad sobre las mismas observaciones.

La estructura del modelo que se utiliza para la demanda residencial de la energía eléctrica en las viviendas rurales del Departamento de Nariño es similar a la que se propone en los trabajos antes mencionados. Los modelos que prueban para Nariño son transversales lineales y sirven para proyectar la demanda en el corto y el largo plazo. Aunque no existe un consenso en la forma como determinar el horizonte de la predicción de un modelo de demanda residencial específico, la naturaleza de las variables que participan en el modelo para las viviendas rurales de Nariño lo pueden clasificar como de largo plazo, y en el sentido de que las variables que representan el inventario de equipos eléctricos y electrónicos, la ampliación física de las viviendas, o en general, el uso de la energía eléctrica en los procesos más característicos de la región, no cambia en el largo plazo.

De otro lado, las variables que se utilizan para hacer el estudio de la demanda en Nariño son muy similares a las que se utilizan en los trabajos previamente referenciados; sin embargo, variables significativas en otros países, como la temperatura y que incide directamente en el proceso de calefacción, no son significativas en el modelo de demanda de las viviendas rurales de Nariño.

El modelo de demanda que aquí se presenta parte del análisis de varios modelos y de la selección de los que tienen mejor ajuste y pertinencia con el consumo observado en las viviendas rurales de Nariño. En la metodología se incluye la aplicación de pruebas y criterios de uso común en la regresión lineal, algunos subjetivos y otros más analíticos,

¹Sistema de Información PERS-Nariño (2013). <http://sipersn.udenar.edu.co:90/sipersn/>.

que refuerzan la selección de un modelo en particular. El objetivo de este documento es el de servir de guía para la depuración del modelo de demanda para el sector rural residencial en Nariño y para la exploración de la demanda en otras regiones de Colombia y a través de otras versiones del PERS que encuentran en marcha.

Para la construcción y evaluación del modelo de demanda residencial de energía eléctrica en las viviendas rurales del Departamento de Nariño se siguen los siguientes pasos: (i) Se estudia e identifica las posibles variables que se incluirán en el modelo, (ii) se establece cuáles son las relaciones entre las variables dependiente y explicativas y se computan las respectivas regresiones (estimación de coeficientes) y (iii) se evalúa cada regresión y se selecciona el mejor modelo. A continuación se detalla cada uno de estos pasos.

1. Variables

La construcción de un modelo de demanda de energía eléctrica para las viviendas rurales del Departamento de Nariño exige en primer lugar la identificación de las variables que caracterizan el consumo de dicha energía. Según los resultados de la encuesta PERS-Nariño el consumo de energía eléctrica en el sector rural y residencial del Departamento de Nariño se encuentra caracterizado principalmente por seis procesos:

- Iluminación: Que corresponde al consumo de energía eléctrica por el uso de fuentes de iluminación, principalmente bombillas incandescentes, ahorradoras y fluorescentes.
- Refrigeración: Que corresponde al consumo de energía eléctrica por el uso de sistemas de enfriamiento y refrigeración, principalmente neveras, neverones, cavas y congeladores.
- Ambiente: Que corresponde al consumo de energía eléctrica por el uso de ventiladores.
- Calentamiento de Agua: Que corresponde al consumo de energía eléctrica por el uso de calentadores y duchas eléctricas.
- Cocción: Que corresponde al consumo de energía eléctrica por el uso de estufas eléctricas para la preparación de alimentos y calentamiento de agua.
- Aparatos Eléctricos y Electrónicos: Que corresponde al consumo de energía eléctrica por el uso de aparatos eléctricos y electrónicos, como licuadoras, cafeteras, planchas, lavadoras, aspiradoras, radios, televisores y equipos de audio y video.

En la figura 1 se muestra el porcentaje de participación de cada uno de estos procesos en el consumo de energía eléctrica de las viviendas rurales del Departamento de Nariño. Se observa que la mayor participación está registrada por (en orden descendente): (i) La Iluminación, con el 34 %; (ii) la Refrigeración, con el 33 % y (iii) los Aparatos Eléctricos, con el 27 %. Estos procesos corresponden al 94 % del consumo de energía eléctrica. El

6 % restante del consumo de energía eléctrica está determinado por (en orden descendente): (iv) El Calentamiento de Agua, con el 3 %; (v) la Cocción, con el 2 % y (vi) el Ambiente, con el 1 %. Estos procesos, que caracterizan y cuantifican el comportamiento del consumo de la energía eléctrica en las viviendas rurales del Departamento de Nariño, se tienen en cuenta en la modelación de la demanda de energía eléctrica como variables explicativas².

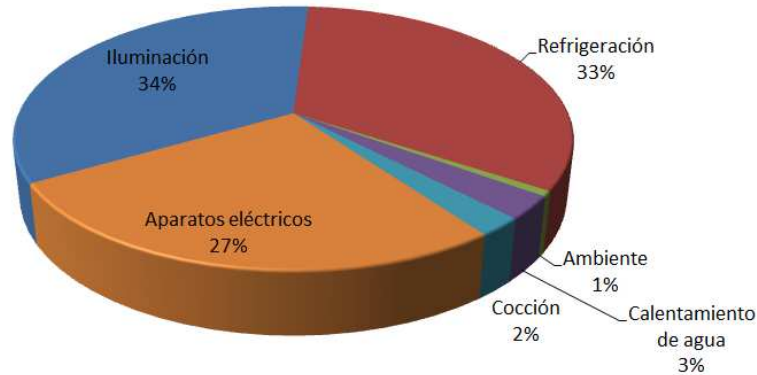


Figura 1: Participación de cada proceso en el consumo de energía eléctrica en el Departamento de Nariño (kWh/Mes). Fuente: SIPERSN <http://sipersn.udenar.edu.co:90/sipersn/>.

De otra parte, la encuesta PERS-Nariño también reporta los valores del consumo de otras fuentes energéticas, como el GLP, el carbón y la leña. En la figura 2 se muestra la participación de cada fuente energética en el consumo de energía del Departamento de Nariño en el sector residencial. Se observa que la fuente energética que mayor participación tiene es la leña, con el 87 %; le sigue el GLP, con el 8 %, y la energía eléctrica, con el 5 %. El uso del carbón en las viviendas rurales nariñenses prácticamente no tiene participación, es 0 %. Las fuentes energéticas distintas a la energía eléctrica, y otras variables relacionadas con la vivienda misma, como el número de personas y cuartos por vivienda, han demostrado ser significativas en el modelo de demanda [4] y se incluyen como variables explicativas en la proposición del modelo de demanda de energía eléctrica.

Es claro que la variable dependiente en el modelo corresponde al consumo de energía eléctrica en las viviendas rurales del Departamento de Nariño. De acuerdo con los datos reportados por la encuesta de energía PERS-Nariño, y que se detallan en la figura 3, el 97 % (2.405) de las viviendas rurales encuestadas cuenta con el servicio de energía eléctrica y solo el 3 % (74) de ellas carecen, o no hacen uso, del servicio. De este 97 % de viviendas rurales que tienen el servicio de electricidad, el 86 % (2.132) acceden a él a través de la red pública, el 7 % (174) lo hacen por medio de una planta municipal, el 3 % (74) por planta compartida y el 1 % (25) cuenta con planta propia.

²Una variable explicativa es una variable que se considera independiente y que causa transformaciones en la variable dependiente, en este caso, que causa transformaciones directas sobre la demanda de energía eléctrica

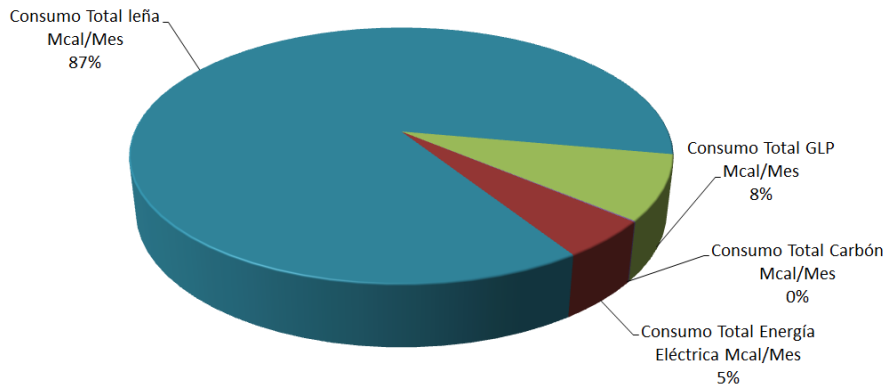


Figura 2: Participación de cada fuente energética en el consumo de energía del Departamento de Nariño (MCal/Mes). Fuente: SIPERSN <http://sipersn.udenar.edu.co:90/sipersn/>.

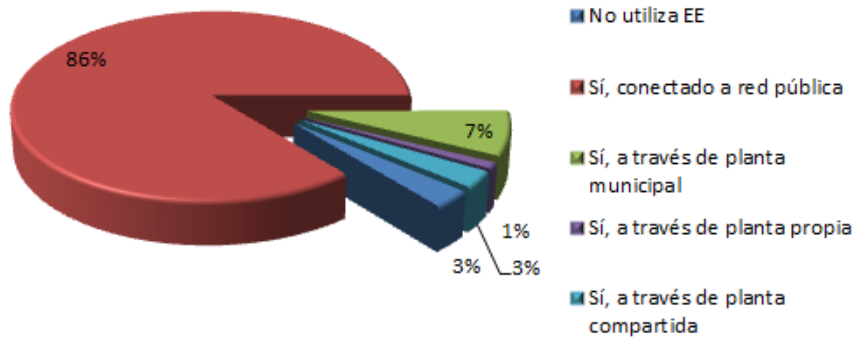


Figura 3: Viviendas con acceso al servicio de energía eléctrica. Fuente: SIPERSN <http://sipersn.udenar.edu.co:90/sipersn/>.

De 2.405 viviendas que reportan el uso de energía eléctrica únicamente se cuenta con $n = 1.397$ observaciones confiables del consumo mensual de este servicio. Estas n observaciones, que corresponden al 56 % del total de viviendas encuestadas (2.479) y que se obtuvieron de las viviendas que sí facilitaron la factura del servicio de electricidad, son la base para el computo del modelo de demanda. La variable denominada `EE_Factura_CkWhMes` es la variable dependiente cuantitativa que almacena estas observaciones. En la figura 4 se muestra el histograma de esta variable.

Este histograma muestra que 617 viviendas registran consumos hasta de 48,5 kWh/mes (es la mayor cantidad de viviendas rurales con servicio de energía eléctrica), le siguen 482 viviendas con consumos en el rango de (48,5 – 95,9] kWh/mes y luego 216 viviendas con consumos en el rango de (95,9 – 143] kWh/mes. En menor cantidad aparecen 51 viviendas con consumos en el rango de (143 – 191] kWh/mes, 18 viviendas con consumos en el rango de (191 – 238] kWh/mes y 13 con valores mayores a éste último rango. Estas últimas viviendas se consideran como *outliers* en la formulación del modelo de demanda.

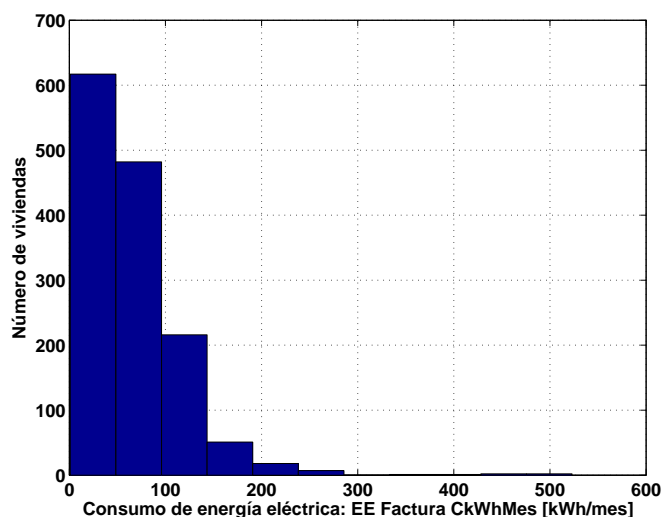


Figura 4: Histograma de la variable dependiente EE_Factura_CkWhMes.

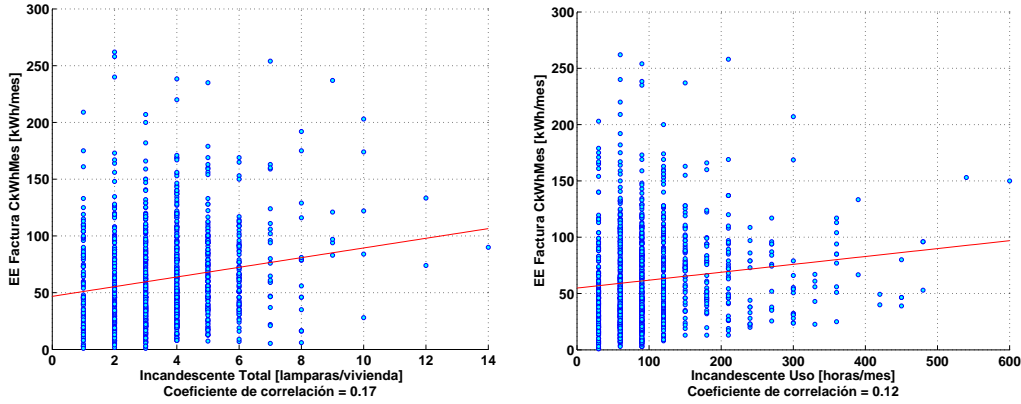
Para explicar el comportamiento de esta variable dependiente, o de la demanda de energía eléctrica en las viviendas rurales de Nariño, se identifica el conjunto de las posibles variables explicativas que lo hacen. Estas variables se extraen de las otras variables que se reportan en la encuesta del PERS-Nariño y que caracterizan el consumo de energía de la población encuestada. Se hace una revisión de la dispersión y de la correlación entre la variable de consumo de energía eléctrica y las variables más significativas reportadas en la encuesta. A continuación se muestra el resumen de estas operaciones de acuerdo a los procesos de consumo de mayor frecuencia identificados por la misma encuesta, otras fuentes de energía utilizadas, el número de personas y cuartos por vivienda y el ingreso de cada vivienda.

- Con respecto al proceso de iluminación, se nota que las variables que mayor número de observaciones tienen son las relacionadas con las lámparas incandescentes. El total de viviendas que utilizan este tipo de lámparas es de 1.137 y corresponde al 81,10% de las viviendas que reportan el consumo de energía eléctrica ($n = 1.397$). Para la proposición del modelo de demanda se utilizan solo las variables que describen el número de lámparas incandescentes por vivienda, `Incandescente_Total`, y su frecuencia de uso mensual, `Incandescente_Uso`. La variable de la potencia de las lámparas incandescentes no se tiene en cuenta en la proposición del modelo debido a su baja correlación y posible colinealidad³ con la variable de frecuencia de uso mensual.

Las figuras 5(a) y 5(b) muestran respectivamente la dispersión y correlación entre los pares de variables `EE_Factura_CkWhMes` vs. `Incandescente_Total` y `EE_Factura_CkWhMes` vs. `Incandescente_Uso`. Se observa que los coeficientes de co-

³La colinealidad, o multicolinealidad, se define como la existencia de relaciones entre las variables explicativas debido a que estas no son ortogonales o independientes entre sí. Su inclusión en el modelo no genera sesgo, pero resultan ser no significativas y disminuyen la eficiencia y precisión del mismo.

relación para estos pares de variables, 0,17 y 0,12 respectivamente, son bajos y no se distingue claramente una relación. Con una línea de color rojo se dibuja una posible tendencia lineal positiva entre los pares de variables y se nota *outliers* en las observaciones.



(a) Consumo de energía eléctrica vs. Número total de lámparas incandescentes (b) Consumo de energía eléctrica vs. Frecuencia de uso de lámparas incandescentes

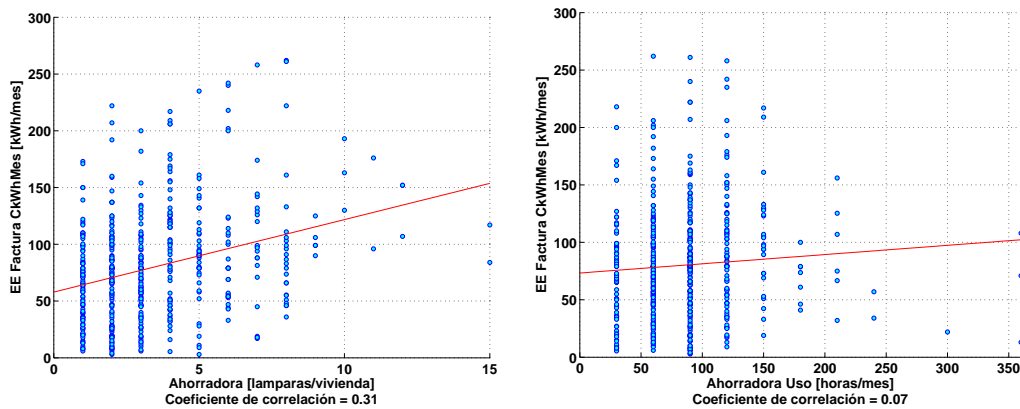
Figura 5: Dispersión y correlación entre el consumo de energía eléctrica y el proceso de iluminación por lámparas incandescentes. (a) EE_Factura_CkWhMes vs. Incandescente_Total. (b) EE_Factura_CkWhMes vs. Incandescente_Uso.

Estos bajos coeficientes de correlación, y la forma como se distribuyen las observaciones en la dispersión, indican que las variables pueden clasificarse no solo como cuantitativas sino también como cualitativas (por la cantidad limitada de valores discretos que asumen). En la propuesta del modelo de demanda se considera la variable Incandescente_Total como explicativa cualitativa y la variable Incandescente_Uso como explicativa cuantitativa.

Por otra parte, el total de viviendas que utilizan lámparas ahorradoras es de 493 y corresponde al 35,23% de las viviendas que reportan el consumo de energía eléctrica. Este porcentaje, aunque relativamente bajo, se tiene en cuenta para la formulación del modelo de demanda. Las variables que se consideran son: El número de lámparas ahorradoras por vivienda, Ahorradora_Total, y su frecuencia de uso mensual, Ahorradora_Uso. La variable de potencia de estas lámparas tiene una correlación un poco más alta que la de la frecuencia de uso de las mismas, sin embargo, por simplicidad, se decide no incluirla en esta versión del modelo.

Las figuras 6(a) y 6(b) muestran respectivamente la dispersión y correlación entre los pares de variables EE_Factura_CkWhMes vs. Ahorradora_Total y EE_Factura_CkWhMes vs. Ahorradora_Uso. También se observa que los coeficientes de correlación para estos pares de variables, 0,31 y 0,07 respectivamente, son bajos y no se distingue claramente una relación. Con una línea de color rojo se dibuja una posible tendencia lineal positiva entre los pares de variables y también se nota *outliers* en las observaciones.

Debido a las mismas razones que se argumentan para las variables de ilumina-



(a) Consumo de energía eléctrica vs. Número total de lámparas ahorradoras (b) Consumo de energía eléctrica vs. Frecuencia de uso de lámparas ahorradoras

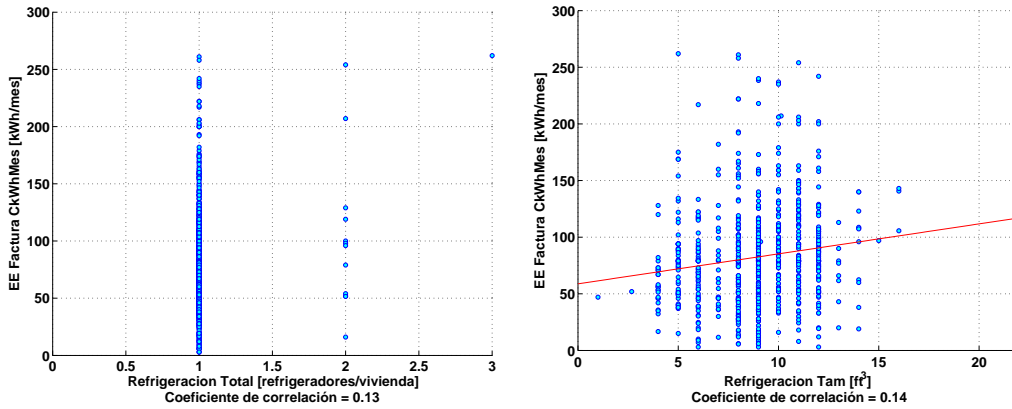
Figura 6: Dispersión y correlación entre el consumo de energía eléctrica y el proceso de iluminación por lámparas ahorradoras. (a) EE_Factura_CkWhMes vs. Ahorradora_Total. (b) EE_Factura_CkWhMes vs. Ahorradora_Uso.

ción con lámparas incandescentes, en la propuesta del modelo de demanda se considera la variable Ahorradora_Total como explicativa cualitativa y la variable Ahorradora_Uso como explicativa cuantitativa.

En el modelo de demanda de energía eléctrica no se tiene en cuenta ninguna de las variables que describen la iluminación con lámparas incandescentes. Su aporte al consumo de electricidad no es significativo porque la cantidad de viviendas que las utilizan es de 14, es decir, están presentes únicamente en el 1% de las viviendas que reportan el consumo de energía eléctrica.

- Para la descripción del proceso de refrigeración se cuenta únicamente con dos variables: La cantidad de neveras por vivienda, Refrigeracion_Total, y el tamaño de las mismas, Refrigeracion_Tam. El número de viviendas que utilizan equipos de refrigeración es de 707, es decir, el 50,61% de las viviendas que reportan el consumo de energía eléctrica. Las figuras 7(a) y 7(b) muestran respectivamente la dispersión y correlación entre los pares de variables EE_Factura_CkWhMes vs. Refrigeracion_Total y EE_Factura_CkWhMes vs. Refrigeracion_Tam. Se observa que los coeficientes de correlación para estos pares de variables, 0,13 y 0,14 respectivamente, son bajos y no se distingue claramente una relación entre ellos. También se nota *outliers* en las observaciones.

En la figura 7(a) se nota claramente que la variable Refrigeracion_Total se debe clasificar como explicativa cualitativa. En la figura 7(b) la línea de color rojo muestra una posible tendencia lineal positiva entre el consumo de energía eléctrica y el tamaño de los refrigeradores. Para el modelo de demanda se decide clasificar la variable Refrigeracion_Tam como explicativa cuantitativa. En la encuesta no se reporta ningún sistema de enfriamiento distinto a la nevera tradicional y que se pueda considerar como significativo para el modelo.



(a) Consumo de energía eléctrica vs. Número total de refrigeradores (b) Consumo de energía eléctrica vs. Tamaño de refrigeradores

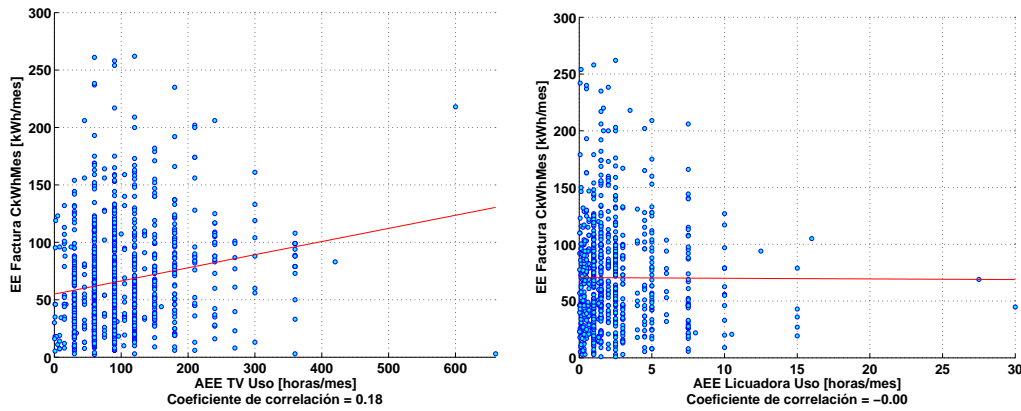
Figura 7: Dispersión y correlación entre el consumo de energía eléctrica y el proceso de refrigeración. (a) EE_Factura_CkWhMes vs. Refrigeracion_Total. (b) EE_Factura_CkWhMes vs. Refrigeracion_Tam.

- El proceso de aparatos eléctricos y electrónicos cuenta con variables que describen la cantidad de equipos (*stock*) presentes en cada vivienda y su frecuencia de uso mensual. Tal como sucede en los anteriores procesos, las variables significativas que representan el número de equipos se clasifican como explicativas cualitativas, mientras que las variables significativas que dan cuenta de la frecuencia de uso de cada aparato se clasifican como explicativas cuantitativas. Para cada clase de equipo eléctrico y electrónico, que se considera significativo en el modelo de demanda, se presenta los resultados de la dispersión y la correlación con la variable de consumo eléctrico.

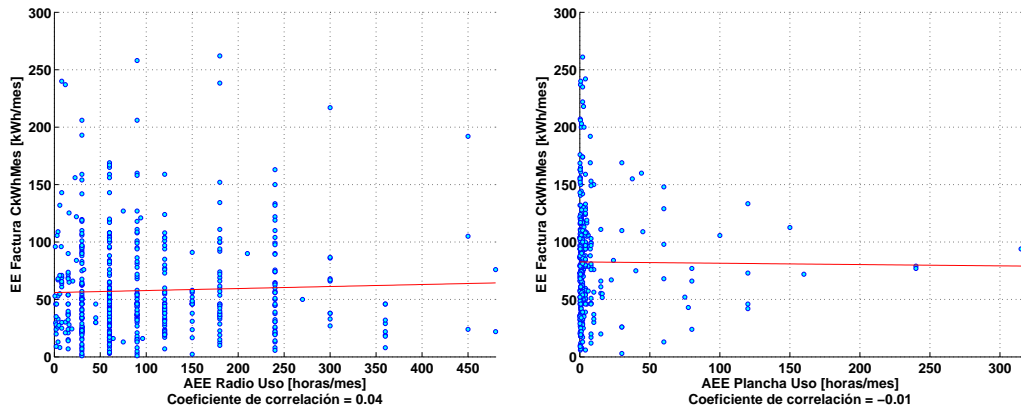
Las viviendas con mayor cantidad de equipos, con su respectivo porcentaje de participación en el total de viviendas que reportan el consumo de energía eléctrica, son (en orden descendente): 1.047 con televisores (74,95%), 863 con licuadoras (61,78%), 495 con radios (35,43%) y 376 con planchas (26,91%). Las variables que describen la cantidad de cada una de estas clases por vivienda son (en el mismo orden): AEE_TV_C, AEE_Licuadora_C, AEE_Radio_C y AEE_Plancha_C. Las figuras 8(a), 8(b), 8(c) y 8(d) muestran respectivamente la dispersión y correlación entre los pares de variables EE_Factura_CkWhMes vs. AEE_TV_Uso, EE_Factura_CkWhMes vs. AEE_Licuadora_Uso, EE_Factura_CkWhMes vs. AEE_Radio_Uso y EE_Factura_CkWhMes vs. AEE_Plancha_Uso.

Las correlaciones de cada uno de estos pares de variables son respectivamente 0,18, $-0,004$, 0,04 y $-0,01$. En este grupo de variables la que mayor correlación tiene con el consumo de energía eléctrica es la frecuencia de uso de televisores, mientras que la frecuencia de uso de las licuadoras es la de menor correlación. Con líneas de color rojo se muestra posibles tendencias lineales entre los pares y también se nota *outliers* en las observaciones.

El número de vivienda con menor cantidad de equipos, con su respectivo porcen-



(a) Consumo de energía eléctrica vs. Frecuencia de uso de televisores (b) Consumo de energía eléctrica vs. Frecuencia de uso de licuadoras

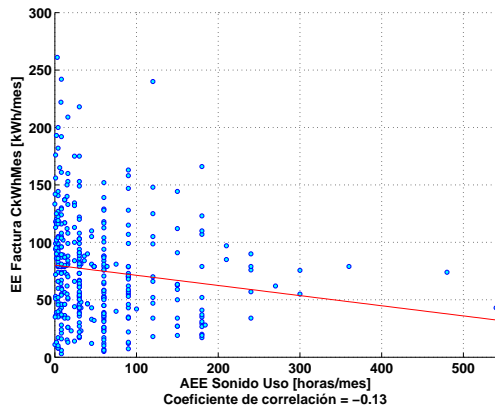


(c) Consumo de energía eléctrica vs. Frecuencia de uso de radios (d) Consumo de energía eléctrica vs. Frecuencia de uso de planchas

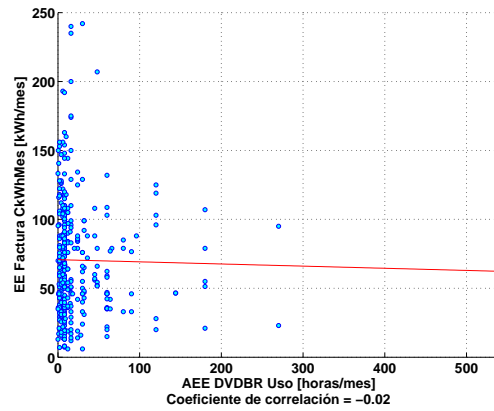
Figura 8: Dispersión y correlación entre el consumo de energía eléctrica y el proceso de aparatos eléctricos y electrónicos (*stock* de mayor cantidad en las viviendas). (a) `EE_Factura_CkWhMes` vs. `AEE_TV_Uso`. (b) `EE_Factura_CkWhMes` vs. `AEE_Licuadora_Uso`. (c) `EE_Factura_CkWhMes` vs. `AEE_Radio_Uso`. (d) `EE_Factura_CkWhMes` vs. `AEE_Plancha_Uso`

taje de participación en el total de viviendas que reportan el consumo de energía eléctrica, son (en orden descendente): 365 con equipos de sonido (26, 13%), 351 con DVDsy y BRs (25, 13%), 207 con ollas arroceras (14, 82%), y 143 con computadores (10, 24%). Las variables que describen la cantidad de cada una de estas clases por vivienda son (en el mismo orden): `AEE_Sonido_C`, `AEE_DVDBR_C`, `AEE_Olla_C` y `AEE_PC_C`. Las figuras 9(a), 9(b), 9(c) y 9(d) muestran respectivamente la dispersión y correlación entre los pares de variables `EE_Factura_CkWhMes` vs. `AEE_Sonido_Uso`, `EE_Factura_CkWhMes` vs. `AEE_DVDBR_Uso`, `EE_Factura_CkWhMes` vs. `AEE_Olla_Uso` y `EE_Factura_CkWhMes` vs. `AEE_PC_Uso`.

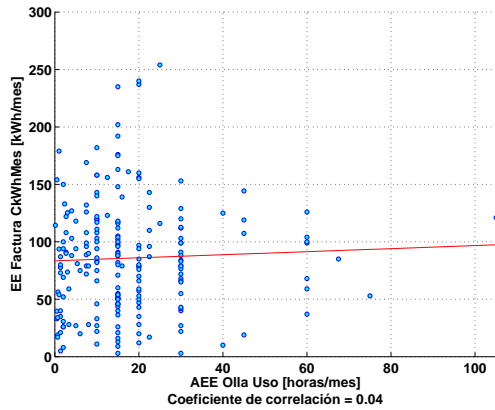
Las correlaciones de cada uno de estos pares de variables son respectivamente $-0,13$, $-0,02$, $0,04$ y $0,03$. En este grupo de variables la que mayor correlación



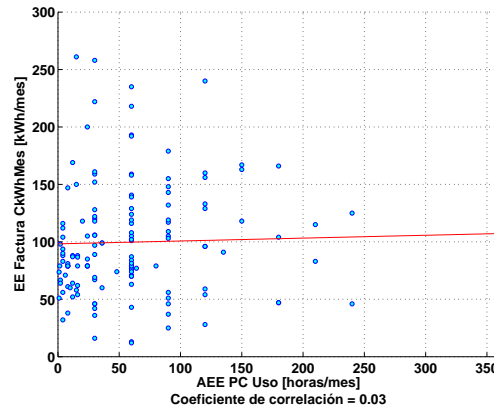
(a) Consumo de energía eléctrica vs. Frecuencia de uso de equipos de sonido



(b) Consumo de energía eléctrica vs. Frecuencia de uso de DVDs y BRs



(c) Consumo de energía eléctrica vs. Frecuencia de uso de ollas arroceras



(d) Consumo de energía eléctrica vs. Frecuencia de uso de computadores

Figura 9: Dispersión y correlación entre el consumo de energía eléctrica y el proceso de aparatos eléctricos y electrónicos (*stock* de menor cantidad en las viviendas). (a) `EE_Factura_CkWhMes` vs. `AEE_Sonido_Uso`. (b) `EE_Factura_CkWhMes` vs. `AEE_DVDBR_Uso`. (c) `EE_Factura_CkWhMes` vs. `AEE_Olla_Uso`. (d) `EE_Factura_CkWhMes` vs. `AEE_PC_Uso`

tiene con el consumo de energía eléctrica es la frecuencia de uso de equipos de sonido, mientras que la variable que menor correlación tiene es la frecuencia de uso de DVDs y BRs. Con líneas de color rojo se muestra posibles tendencias lineales entre los pares, también se nota *outliers* en las observaciones.

- Los otros procesos que tienen participación en el consumo de energía eléctrica están representados por variables que tienen un porcentaje bajo de observaciones. El número de viviendas que reportan información sobre estos procesos es inferior a 140, es decir, inferior al 10% del total de viviendas que reportan el consumo de energía eléctrica. El proceso de cocción, por ejemplo, está descrito por la cantidad de estufas eléctricas que hay en cada vivienda, su potencia y el tiempo dedicado a la

preparación de alimentos; la cantidad de viviendas que reportan este proceso es de 43, es decir, el 3 % del total de viviendas que reportan el consumo de electricidad.

Con respecto al proceso de calentamiento de agua, que está determinado por la frecuencia de uso de duchas eléctricas (no hay calentadores eléctricos en las viviendas), solo se registra en 73 viviendas, es decir, el 5 % de las viviendas que reportan el consumo de electricidad. El proceso de adecuación de ambiente, determinado por la cantidad de ventiladores en la vivienda, su potencia y la frecuencia de uso de los mismos, solo se registra en 75 viviendas, es decir, el 5 % del total de viviendas encuestadas con reporte del consumo de energía eléctrica. No se reporta otro tipo de aparato eléctrico de acondicionamiento de ambiente con uso significativo en las viviendas. Se decide por lo tanto no tener en cuenta ninguna de las variables que describen estos procesos para la proposición del modelo de demanda.

- La encuesta energética PERS-Nariño también reporta la información del consumo de otras fuentes energéticas. El consumo de energéticos como el carbón, y otros derivados del petróleo como la gasolina, no tiene un valor significativo en las viviendas rurales de Nariño; pero sí lo tiene el consumo de energéticos como el gas y la leña. Los porcentajes de viviendas que hacen uso de estas fuentes energéticas, entre las viviendas que reportan el consumo de energía eléctrica, son: 46,46 % (649 viviendas) para el gas y 49,82 % (696 viviendas) para la leña.

Las figuras 10(a) y 10(b) muestran respectivamente las gráficas de la dispersión y correlación entre la variable de consumo de energía eléctrica, *EE_Factura_CkWhMes*, y las variables de consumo de gas, *Gas_CkWhMes*, y leña, *Lena_CkWhMes*. Los coeficientes de correlación para estos pares son respectivamente 0,12 y 0,08. Con una línea de color rojo se dibuja una posible relación lineal entre los pares de variables y se nota la existencia de *outliers* en las observaciones. De acuerdo a estas figuras, las dos variables que describen el consumo de gas y leña presentan una baja correlación y no se distingue claramente una relación entre ellas. Sin embargo, estas dos variables se mantienen para la proposición del modelo de demanda de energía eléctrica y se clasifican como explicativas cuantitativas.

- De acuerdo con [4], variables como el número de personas que habitan la vivienda, el número de cuartos que la componen, el ingreso mensual de los mismos y el valor pagado por el servicio de energía eléctrica, ayudan a explicar en gran porcentaje el consumo de energía eléctrica residencial. Estas variables también se encuentran disponibles en la encuesta y se denominan, respectivamente, *Vivienda_Persona*, *Vivienda_Cuarto*, *Vivienda_Ingreso* y *EE_Factura_Valor*. El número de observaciones registradas para cada variable, y su porcentaje del total de viviendas que reportan el consumo de electricidad, son: 1.391, con el 99,57 %; 1.389, con el 99,43 %, 1.356, con el 97,07 % y 1.375, con el 98,43 %.

Las figuras 11(a), 11(b), 11(c) y 11(d) muestran respectivamente las gráficas de la dispersión y la correlación entre los pares *EE_Factura_CkWhMes* vs. *Vivienda_Persona*, *EE_Factura_CkWhMes* vs. *Vivienda_Cuarto*, *EE_Factura_CkWhMes* vs. *Vivienda_Ingreso* y *EE_Factura_CkWhMes* vs. *EE_Factura_Valor*. Los coeficientes de correlación para estos pares son, respectivamente, 0,13, 0,33, 0,33 y 0,69. Se

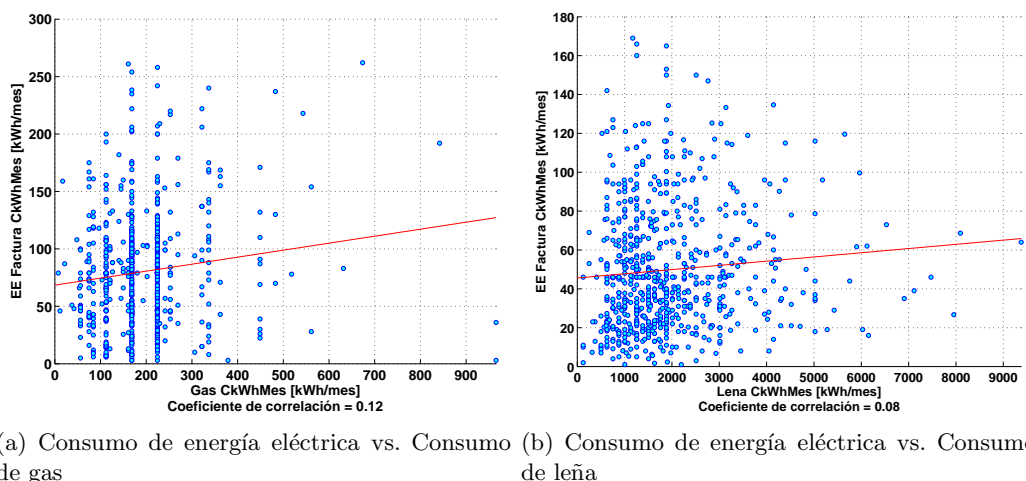


Figura 10: Dispersión y correlación entre el consumo de energía eléctrica y otras fuentes energéticas. (a) EE_Factura_CkWhMes vs. Gas_CkWhMes. (b) EE_Factura_CkWhMes vs. Lena_CkWhMes.

observa que las variables que representan el número de cuartos por vivienda, el ingreso en cada una de ellas y el valor pagado por el servicio, tienen los coeficientes de correlación más altos, incluso mayores que los coeficientes de los otros pares de variables ya analizados. Con líneas de color rojo se dibuja posibles relaciones lineales entre los pares de variables y también se nota la existencia de *outliers* en las observaciones. Estas cuatro variables se tienen en cuenta en la proposición del modelo y se clasifican como explicativas cuantitativas.

Una vez que se ha identificado el conjunto de variables explicativas que pueden contribuir a la explicación de la demanda de energía eléctrica en las viviendas rurales del Departamento de Nariño, o de la variable dependiente, se procede a realizar la regresión de las posibles relaciones entre ellas. El nombre de cada variable (primera columna), su descripción (segunda columna), el tipo (tercera columna), la unidad de medida (cuarta columna) y el porcentaje de observaciones válidas sobre el total de viviendas que reportan el consumo de energía eléctrica, $n = 1.397$, (quinta columna) se muestran en la tabla 1.

Tabla 1: Variables.

Nombre	Descripción	Tipo	Unidad	% de n
Consumo de energía eléctrica				
EE_Factura	Viviendas con servicio de energía eléctrica facturada	Explicativa/ Categorica		100 %
EE_Factura_CkWhMes	Consumo mensual de energía eléctrica de cada vivienda	Dependiente/ Cuantitativa	kWh/mes	100 %

Nombre	Descripción	Tipo	Unidad	% de <i>n</i>
EE_Factura_Valor	Valor facturado del consumo mensual de energía eléctrica de cada vivienda	Explicativa/ Cuantitativa	COP	100 %
Iluminación				
Incandescente_Total	Número de lámparas incandescentes en cada vivienda	Explicativa/ Categoría	lamparas/ vivienda	81 %
Incandescente_Uso	Frecuencia de uso de las lámparas incandescentes	Explicativa/ Cuantitativa	horas/mes	81 %
Ahorradora	Número de lámparas ahorradoras en cada vivienda	Explicativa/ Categoría	lamparas/ vivienda	36 %
Ahorradora_Uso	Frecuencia de uso de las lámparas ahorradoras	Explicativa/ Cuantitativa	horas/mes	36 %
Refrigeración				
Refrigeracion_Total	Número de refrigeradores en cada vivienda	Explicativa/ Categoría	refrigeradores/ vivienda	51 %
Refrigeracion_Tam	Tamaño de los refrigeradores	Explicativa/ Cuantitativa	ft ³	51 %
Aparatos eléctricos y electrónicos				
AEE_TV_C	Número de televisores en cada vivienda	Explicativa/ Categoría	televisores/ vivienda	75 %
AEE_TV_Uso	Frecuencia de uso de televisores	Explicativa/ Cuantitativa	horas/mes	75 %
AEE_Licuadaora_C	Número de licuadoras en cada vivienda	Explicativa/ Categoría	licuadaora/ vivienda	62 %
AEE_Licuadaora_Uso	Frecuencia de uso de televisores	Explicativa/ Cuantitativa	horas/mes	62 %
AEE_Radio_C	Número de radios en cada vivienda	Explicativa/ Categoría	radios/ vivienda	36 %
AEE_Radio_Uso	Frecuencia de uso de radios	Explicativa/ Cuantitativa	horas/mes	36 %
AEE_Plancha_C	Número de plancha en cada vivienda	Explicativa/ Categoría	plancha/ vivienda	27 %
AEE_Plancha_Uso	Frecuencia de uso de plancha	Explicativa/ Cuantitativa	horas/mes	27 %
AEE_Sonido_C	Número de equipos de sonido en cada vivienda	Explicativa/ Categoría	es/vivienda	26 %
AEE_Sonido_Uso	Frecuencia de uso de equipos de sonido	Explicativa/ Cuantitativa	horas/mes	26 %
AEE_DVDBR_C	Número de DVDs y BRs en cada vivienda	Explicativa/ Categoría	DVDBRs/ vivienda	25 %
AEE_DVDBR_Uso	Frecuencia de uso de DVDs y BRs	Explicativa/ Cuantitativa	horas/mes	25 %

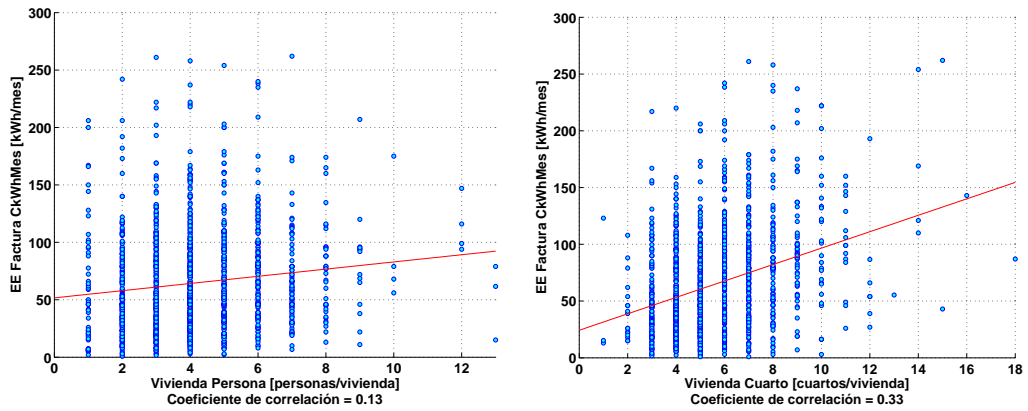
Nombre	Descripción	Tipo	Unidad	% de n
AEE_Olla_C	Número de ollas arroceras en cada vivienda	Explicativa/- Categoría	ollas/vivienda	15 %
AEE_Olla_Uso	Frecuencia de uso de ollas arroceras	Explicativa/- Cuantitativa	horas/mes	15 %
AEE_PC_C	Número de computadores en cada vivienda	Explicativa/- Categoría	PCs/vivienda	10 %
AEE_PC_Uso	Frecuencia de uso de computadores	Explicativa/- Cuantitativa	horas/mes	10 %
Otros fuentes energéticas				
Gas_CkWhMes	Consumo de gas mensual	Explicativa/- Cuantitativa	kWh/Mes	47 %
Lena_CkWhMes	Consumo de leña mensual	Explicativa/- Cuantitativa	kWh/Mes	50 %
Vivienda				
Vivienda_Persona	Número de personas en cada vivienda	Explicativa/- Cuantitativa	personas/- vivienda	100 %
Vivienda_Cuarto	Número de cuartos en cada vivienda	Explicativa/- Cuantitativa	cuartos/- vivienda	99 %
Vivienda_Ingreso_COP	Promedio de ingreso en cada vivienda	Explicativa/- Cuantitativa	COP/- vivienda	97 %

2. Regresión

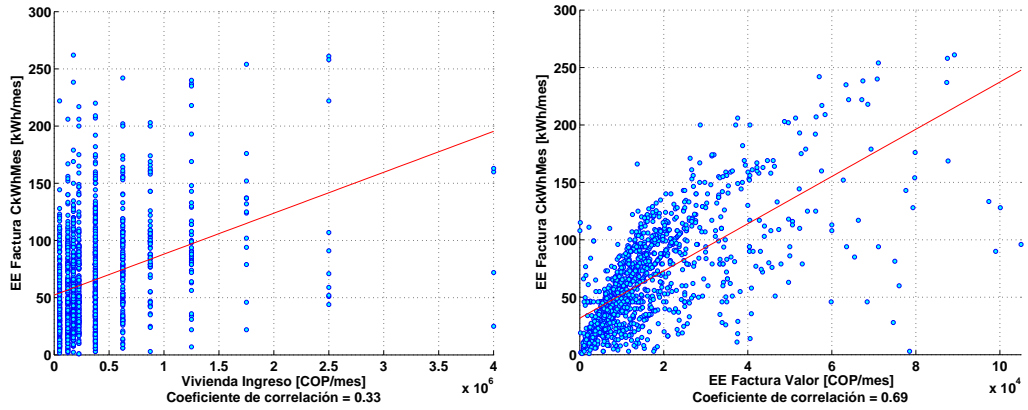
Se realizan varias regresiones entre la variable dependiente y las variables explicativas para identificar posibles modelos de demanda de la energía eléctrica en las viviendas rurales del Departamento de Nariño. De estas regresiones se seleccionan las que tienen los niveles más altos de explicación de la demanda de electricidad y se proponen como ejemplos representativos de una posible metodología de modelación de la demanda de energía eléctrica para las viviendas rurales del Departamento de Nariño y a partir de los datos de la encuesta energética PERS-Nariño.

Los ejemplos que se proponen exploran relaciones lineales, generales y categóricas, entre las variables dependiente y explicativas. Como no existe una regla general para la identificación de estas posibles relaciones, se sigue un proceso de ensayo y error basado en estudios de otros países [4][5] y en el análisis empírico del comportamiento de la demanda de energía eléctrica con respecto a las variables explicativas que se obtienen de la encuesta. Se hace uso de ayudas instrumentales como la revisión de la dispersión y la correlación, el análisis de la distribución de las variables (histograma) y la exclusión manual de *outliers* para lograr las mejores regresiones de estas relaciones.

El objetivo en cada regresión es estimar los coeficientes que hacen parte de las ecuaciones que definen las relaciones entre las variables dependiente y explicativa, con el mayor porcentaje de explicación e incluyendo las variables más significativas. En Teoría



(a) Consumo de energía eléctrica vs. Número de personas por vivienda (b) Consumo de energía eléctrica vs. Número de cuartos por vivienda



(c) Consumo de energía eléctrica vs. Ingreso por vivienda (d) Consumo de energía eléctrica vs. Valor factura

Figura 11: Dispersión y correlación entre el consumo de energía eléctrica y características de las viviendas. (a) $EE_Factura_CkWhMes$ vs. $Vivienda_Persona$. (b) $EE_Factura_CkWhMes$ vs. $Vivienda_Cuarto$. (c) $EE_Factura_CkWhMes$ vs. $Vivienda_Ingreso$. (d) $EE_Factura_CkWhMes$ vs. $EE_Factura_Valor$.

de Regresión Lineal [6], los algoritmos de Mínimos Cuadrados Ordinarios (MCO) y de Regresión Robusta Óptima (RRO) [3] son los de mayor uso en la estimación de estos coeficientes. En este caso, se opta por utilizar un algoritmo RRO por las ventajas que este posee: descarte automático de *outliers* en las observaciones y corrección de la estimación con respecto a la heteroscedasticidad⁴. Este problema es común en modelos de corte transversal y el modelo de demanda de energía eléctrica residencial en el Departamento de Nariño es un posible candidato a tenerla, debido principalmente a la heterogeneidad del consumo de las viviendas que conforman las regiones del departa-

⁴La heteroscedasticidad se define como el cambio en la varianza del error del modelo cuando los valores observados de las variables explicativas cambian también. La heteroscedasticidad no genera problemas de sesgo en la estimación de los coeficientes; pero sí genera un falso valor de significancia en la variable explicativa cuando en realidad esta no la tiene.

mento. A continuación se muestra las mejores regresiones y su significado en el proceso de modelación de la demanda.

2.1. Mod_1

En esta primera regresión se construye un modelo de tipo general que trata de explicar la demanda de la energía eléctrica por medio del comportamiento de la mayor cantidad de variables explicativas que se encuentran disponibles. Dado que el comportamiento de la variable del consumo de energía eléctrica, $EE_Factura_CkWhMes$, presenta una dispersión alta con cada una de las variables explicativas, las observaciones, o viviendas, que se consideran como *outliers* se filtran y se reduce el porcentaje de explicación del modelo. Las observaciones que se descartan en esta regresión son las viviendas que reportan consumos superiores a 250 kWh/mes y que cancelan por el servicio más de 150.000 COP.

En esta regresión se asume que el consumo de energía eléctrica se explica principalmente por la contribución lineal de: (i) La frecuencia de uso de las lámparas incandescentes, $Incandescente_Uso$; (ii) la frecuencia de uso de las lámparas ahorradoras, $Ahorradora_Uso$; (iii) el tamaño de los refrigeradores, $Refrigeracion_Tam$; (iv) el número de personas que habitan en cada vivienda, $Vivienda_Persona$; (v) el número de cuartos, $Vivienda_Cuarto$, (vi) el ingreso mensual en cada una de ellas, $Vivienda_Ingreso$ y (vii) el valor facturado del servicio de electricidad, $EE_Factura_Valor$. En la tabla 2 se explica la asignación de las variables de la encuesta a las variables del modelo.

Tabla 2: Asignación de las variables del modelo Mod_1 .

Variable de Mod_1	Variable de la encuesta
y_i	$EE_Factura_CkWhMes$
$x_{1,i}$	$Incandescente_Uso$
$x_{2,i}$	$Ahorradora_Uso$
$x_{3,i}$	$Refrigeracion_Tam$
$x_{4,i}$	$Vivienda_Persona$
$x_{5,i}$	$Vivienda_Cuarto$
$x_{6,i}$	$Vivienda_Ingreso$
$x_{7,i}$	$EE_Factura_Valor$

La relación entre la variable dependiente y las variables explicativas se describe por medio de la siguiente ecuación:

$$y_i = \alpha_0 + \sum_{k=1}^7 \alpha_k x_{k,i} + \omega_i \quad , \quad (1)$$

donde el subíndice i identifica a cada observación; y_i es la variable dependiente que representa el consumo de energía eléctrica mensual de cada vivienda; $x_{k,i}$, con $k = 1, \dots, 7$ son las variables explicativas; α_0 es el intercepto; α_k son los coeficientes (se consideran constantes) y $\omega_i \stackrel{iid}{\sim} N(0, \sigma_i^2)$ son los errores independientes e idénticamente y normalmente distribuidos, con media cero y varianza σ_i^2 constante. Los coeficientes

de la ecuación 1, incluyendo el intercepto, se estiman con 1.372 observaciones (98,21 % de n). La figura 19 muestra los resultados de esta regresión.

```

Mod_1
-----
Número de observaciones:
1372 (98.21% de n)

Linear regression model (robust fit):
y ~ 1 + x1 + x2 + x3 + x4 + x5 + x6 + x7

Estimated Coefficients:

```

	Estimate	SE	tStat	pValue
(Intercept)	4.4139	1.9925	2.2152	0.026939
x1	0.017412	0.0071865	2.4228	0.015553
x2	0.045293	0.012236	3.7017	0.0002243
x3	0.98287	0.085855	11.448	8.1172e-29
x4	1.0123	0.31054	3.2598	0.0011476
x5	1.6901	0.31829	5.31	1.3143e-07
x6	4.2682e-06	1.6452e-06	2.5943	0.0095979
x7	0.0024559	5.9245e-05	41.453	1.093e-230

```

Number of observations: 1161, Error degrees of freedom: 1153

```

Figura 12: Resultados de la regresión de Mod_1 .

En esta figura se lee que: (i) El número efectivo de observaciones utilizadas por el algoritmo es de 1.161, es decir, se han descartado 211 observaciones que se consideran como *outliers*, y (ii) el número de grados de libertad DoF (*Degrees of Freedom*), que corresponden a estas observaciones menos el número de coeficientes a estimar (incluyendo el intercepto), es $DoF = 1.161 - 8 = 1.153$. Se comprueba que hay suficientes observaciones para realizar una buena estimación. De la columna *Estimate* se puede leer los coeficientes estimados por el algoritmo; luego, la demanda de energía eléctrica residencial mensual para las viviendas rurales del Departamento de Nariño es:

$$\hat{y} = 4,41 + 0,02x_1 + 0,05x_2 + 0,98x_3 + 1,01x_4 + 1,69x_5 + 0,000004x_6 + 0,003x_7 \quad (2)$$

donde \hat{y} es el consumo estimado para cada vivienda (kWh/mes), x_1 es la frecuencia de uso de lámparas incandescentes (horas/mes), x_2 es la frecuencia de uso de lámparas ahorradoras (horas/mes), x_3 es el tamaño de equipos de refrigeración (ft^3), x_4 es el número de personas que habitan la vivienda (personas/vivienda), x_5 es el número de cuartos por vivienda (cuartos/vivienda), x_6 es el ingreso de cada vivienda (COP/mes) y x_7 es el valor pagado por el servicio de electricidad (COP/mes).

Los coeficientes se interpretan de la siguiente manera: $\hat{\alpha}_0 = 4,41$ representa un consumo mínimo de subsistencia en kWh/mes cuando todas las variables explicativas se hacen cero, $\hat{\alpha}_1 = 0,02$ es el incremento en el consumo en kW por cada hora/mes de uso

de lámparas incandescentes, $\hat{\alpha}_2 = 0,05$ es el incremento en el consumo en kW por cada hora/mes de uso de lámparas ahorradoras, $\hat{\alpha}_3 = 0,98$ es el incremento en el consumo en kWh/mes/ft³ por cada ft³ de tamaño de refrigeradores, $\hat{\alpha}_4 = 1,01$ es el incremento en el consumo en kWh/mes/persona/vivienda por cada persona/vivienda, $\hat{\alpha}_5 = 1,69$ es el incremento en el consumo en kWh/mes/cuarto/vivienda por cada cuarto/vivienda, $\hat{\alpha}_6 = 0,000004$ es el incremento en el consumo en kWh/COP por cada COP/mes de ingreso mensual y $\hat{\alpha}_7 = 0,003$ es el incremento en el consumo en kWh/COP por cada COP/mes pagado por el servicio.

2.2. $Mod_{2,1}$, $Mod_{2,2}$ y $Mod_{2,3}$

La encuesta revela que las viviendas rurales del Departamento de Nariño no utilizan en forma significativa la energía eléctrica para el proceso de cocción (no se reporta un uso significativo de estufas eléctricas); en cambio, hacen uso significativo del gas y la leña y de forma exclusiva para este proceso. Las observaciones muestran que el uso del gas y de la leña se hace de manera excluyente en cada una de las viviendas rurales nariñenses, es decir, no hay viviendas que utilizan simultáneamente gas y leña para este proceso.

Debido a que no hay participación representativa de la energía eléctrica en el proceso de cocción, no es posible hacer una regresión que muestre la incidencia de la demanda del gas y de la leña, a través de ese proceso, en la demanda de la energía eléctrica; sin embargo, sí es posible hacer una comparación del consumo de energía eléctrica entre las viviendas que utilizan gas y las viviendas que utilizan leña. Para hacer esto se plantean tres regresiones: $Mod_{2,1}$, $Mod_{2,2}$ y $Mod_{2,3}$. Las dos primeras se hacen respectivamente con observaciones filtradas para las viviendas que usan gas y para las que utilizan leña. La tercera es una regresión categórica que utiliza una variable ficticia para diferenciar las dos clases de viviendas.

- $Mod_{2,1}$ En esta regresión se utiliza solo las observaciones de las viviendas que consumen energía eléctrica y gas. Al igual que para la regresión Mod_1 , las viviendas que reportan consumos superiores a 250 kWh/mes y que cancelan por el servicio más de 150.000 COP se descartan. Aquellas viviendas que registran un consumo de gas equivalente a más de 500 kWh/mes se descartan también.

Se asume que el consumo mensual de energía eléctrica de estas viviendas, $EE_Factura_CkWhMes$, se explica principalmente por la contribución lineal de: (i) El tamaño de los refrigeradores, $Refrigeracion_Tam$; (ii) el número de personas que habitan en cada vivienda, $Vivienda_Persona$; (iii) el número de cuartos, $Vivienda_Cuarto$ y (iv) el valor facturado del servicio de electricidad, $EE_Factura_Valor$. En la tabla 3 se explica la asignación de las variables de la encuesta a las variables del modelo.

La relación entre la variable dependiente y las variables explicativas se describe por medio de la siguiente ecuación:

$$y_i = \alpha_0 + \sum_{k=1}^4 \alpha_k x_{k,i} + \omega_i \quad , \quad (3)$$

Tabla 3: Asignación de las variables del modelo $Mod_{2,1}$.

Variable de $Mod_{2,1}$	Variable de la encuesta
y_i	EE_Factura_CkWhMes
$x_{1,i}$	Refrigeracion_Tam
$x_{2,i}$	Vivienda_Persona
$x_{3,i}$	Vivienda_Cuarto
$x_{4,i}$	EE_Factura_Valor

donde el subíndice i identifica a cada observación; y_i es la variable dependiente que representa el consumo de energía eléctrica mensual de cada vivienda; $x_{k,i}$, con $k = 1, \dots, 4$ son las variables explicativas; α_0 es el intercepto; α_k son los coeficientes (se consideran constantes) y $\omega_i \stackrel{iid}{\sim} N(0, \sigma_i^2)$ son los errores independientes e idénticamente y normalmente distribuidos, con media cero y varianza σ_i^2 constante. Los coeficientes de la ecuación 3, incluyendo el intercepto, se estiman con 627 observaciones (44,88% de n). La figura 20 muestra los resultados de esta regresión.

```

Mod_2_1
-----
Número de observaciones:
627 (44.88% de n)

Linear regression model (robust fit):
y ~ 1 + x1 + x2 + x3 + x4

Estimated Coefficients:
              Estimate      SE      tStat      pValue
(Intercept)    11.455        3.331    3.4389    0.00063253
x1              0.62724       0.1395    4.4964    8.5894e-06
x2              1.3166        0.50642   2.5998    0.0096025
x3              1.1547        0.46133   2.503     0.012631
x4              0.0028184     8.2893e-05 34.001    1.9016e-132

Number of observations: 508, Error degrees of freedom: 503

```

Figura 13: Resultados de la regresión de $Mod_{2,1}$.

En esta figura se lee que: (i) El número efectivo de observaciones utilizadas por el algoritmo es de 508, es decir, se han descartado 119 observaciones que se consideran como *outliers*, y (ii) el número de grados de libertad es $DoF = 508 - 5 = 503$. Se comprueba que hay suficientes observaciones para realizar una buena estimación. De la columna *Estimate* se puede leer los coeficientes estimados por el algoritmo; luego, la demanda de energía eléctrica residencial mensual para las viviendas rurales del Departamento de Nariño, y que consumen gas, es:

$$\hat{y} = 11,46 + 0,63x_1 + 1,32x_2 + 1,15x_3 + 0,003x_4 \quad , \quad (4)$$

donde \hat{y} es el consumo estimado para cada vivienda (kWh/mes), x_1 es el tamaño de equipos de refrigeración (ft³), x_2 es el número de personas que habitan la vivienda (personas/vivienda), x_3 es el número de cuartos por vivienda (cuartos/vivienda) y x_4 es el valor pagado por el servicio de electricidad (COP/mes).

Los coeficientes se interpretan de la siguiente manera: $\hat{\alpha}_0 = 11,46$ representa un consumo mínimo de subsistencia en kWh/mes cuando todas las variables explicativas se hacen cero, $\hat{\alpha}_1 = 0,63$ es el incremento en el consumo en kWh/mes/ft³ por cada ft³ de tamaño de refrigeradores, $\hat{\alpha}_2 = 1,32$ es el incremento en el consumo en kWh/mes/persona/vivienda por cada persona/vivienda, $\hat{\alpha}_3 = 1,15$ es el incremento en el consumo en kWh/mes/cuarto/vivienda por cada cuarto/vivienda y $\hat{\alpha}_4 = 0,003$ es el incremento en el consumo en kWh/COP por cada COP/mes pagado por el servicio.

- *Mod_{2,2}*: En esta regresión se utiliza solo las observaciones de las viviendas que consumen energía eléctrica y leña. Al igual que para la regresión *Mod₁*, las viviendas que reportan consumos superiores a 250 kWh/mes y que cancelan por el servicio más de 150.000 COP se descartan. Las viviendas que registran un consumo de leña equivalente a más de 8000 kWh/mes se descartan también.

Se asume que el consumo mensual de energía eléctrica de estas viviendas, EE_Factura_CkWhMes, se explica principalmente por: i) La frecuencia de uso de las lámparas incandescentes, Incandescente_Uso; (ii) la frecuencia de uso de las lámparas ahorradoras, Ahorradora_Uso; (iii) el tamaño de los refrigeradores, Refrigeracion_Tam; (iv) el número de personas que habitan en cada vivienda, Vivienda_Persona y (v) el valor facturado del servicio de electricidad, EE_Factura_Valor. En la tabla 4 se explica la asignación de las variables de la encuesta a las variables del modelo.

Tabla 4: Asignación de las variables del modelo *Mod_{2,2}*.

Variable de <i>Mod_{2,2}</i>	Variable de la encuesta
y_i	EE_Factura_CkWhMes
$x_{1,i}$	Incandescente_Uso
$x_{2,i}$	Ahorradora_Uso
$x_{3,i}$	Refrigeracion_Tam
$x_{4,i}$	Vivienda_Persona
$x_{5,i}$	EE_Factura_Valor

La relación entre la variable dependiente y las variables explicativas se describe por medio de la siguiente ecuación:

$$y_i = \alpha_0 + \sum_{k=1}^5 \alpha_k x_{k,i} + \omega_i \quad , \quad (5)$$

donde el subíndice i identifica a cada observación; y_i es la variable dependiente que representa el consumo de energía eléctrica mensual de cada vivienda; $x_{k,i}$, con $k = 1, \dots, 5$ son las variables explicativas; α_0 es el intercepto; α_k son los coeficientes (se consideran constantes) y $\omega_i \stackrel{\text{iid}}{\sim} N(0, \sigma_i^2)$ son los errores independientes e idénticamente y normalmente distribuidos, con media cero y varianza σ_i^2 constante. Los coeficientes de la ecuación 5, incluyendo el intercepto, se estiman con 691 observaciones (49,46% de n). La figura 21 muestra los resultados de esta regresión.

```

Mod_2_2
-----
Número de observaciones:
691 (49.46% de n)

Linear regression model (robust fit):
y ~ 1 + x1 + x2 + x3 + x4 + x5

Estimated Coefficients:

```

	Estimate	SE	tStat	pValue
(Intercept)	11.08	1.8377	6.0295	2.8356e-09
x1	0.032344	0.0094634	3.4177	0.00067306
x2	0.073046	0.019739	3.7007	0.0002343
x3	1.8822	0.15423	12.204	7.7931e-31
x4	1.5835	0.37468	4.2262	2.737e-05
x5	0.0017512	8.3868e-05	20.881	1.2817e-73

```

Number of observations: 622, Error degrees of freedom: 616

```

Figura 14: Resultados de la regresión de $Mod_{2,2}$.

En esta figura se lee que: (i) El número efectivo de observaciones utilizadas por el algoritmo es de 622, es decir, se han descartado 69 observaciones que se consideran como *outliers*, y (ii) el número de grados de libertad es $DoF = 622 - 6 = 616$. Se comprueba que hay suficientes observaciones para realizar una buena estimación. De la columna *Estimate* se puede leer los coeficientes estimados por el algoritmo; luego, la demanda de energía eléctrica residencial mensual para las viviendas rurales del Departamento de Nariño, y que consumen leña, es:

$$\hat{y} = 11,08 + 0,03x_1 + 0,07x_2 + 1,88x_3 + 1,58x_4 + 0,002x_5 \quad , \quad (6)$$

donde \hat{y} es el consumo estimado para cada vivienda (kWh/mes), x_1 es la frecuencia de uso de lámparas incandescentes (horas/mes), x_2 es la frecuencia de uso de lámparas ahorradoras (horas/mes), x_3 es el tamaño de equipos de refrigeración (ft^3), x_4 es el número de personas que habitan la vivienda (personas/vivienda) y x_5 es el valor facturado del servicio de electricidad (COP/mes).

Los coeficientes se interpretan de la siguiente manera: $\hat{\alpha}_0 = 11,08$ representa un consumo mínimo de subsistencia en kWh/mes cuando todas las variables explicativas se hacen cero, $\hat{\alpha}_1 = 0,03$ es el incremento en el consumo en kW por cada hora/mes de uso de lámparas incandescentes, $\hat{\alpha}_2 = 0,07$ es el incremento en el consumo en kW por cada hora/mes de uso de lámparas ahorradoras, $\hat{\alpha}_3 = 1,88$ es el incremento en el consumo en kWh/mes/ft³ por cada ft³ de tamaño de refrigeradores, $\hat{\alpha}_4 = 1,58$ es el incremento en el consumo en kWh/mes/persona/vivienda por cada persona/vivienda y $\hat{\alpha}_5 = 0,002$ es el incremento en el consumo en kWh/COP por cada COP/mes pagado por el servicio.

- *Mod_{2,3}*: A diferencia de los dos modelos anteriores, *Mod_{2,1}* y *Mod_{2,2}*, este propone una sola regresión con variable ficticia que facilita el cálculo de la diferencia del consumo de energía eléctrica entre las viviendas que consumen gas y aquellas que consumen leña. Para hacer esta regresión, primero se categorizan las viviendas rurales en dos clases: (i) La clase *A* de viviendas que consumen energía eléctrica y leña y (ii) la clase *B* de viviendas que consumen energía eléctrica y gas; luego se define una clase base, que en este caso se hace corresponder a la clase *A*, y se crea la variable ficticia⁵. Esta variable ficticia toma los siguientes valores: 1, cuando la vivienda pertenece a la clase *B*, y 0, cuando la vivienda no pertenece a esta clase (o pertenece a la clase *A*).

En esta regresión, las viviendas que reportan consumos superiores a 250 kWh/mes y que cancelan por el servicio más de 150.000 COP se descartan. Se asume que el consumo mensual de energía eléctrica de las dos clases de viviendas, *EE_Factura_CkWhMes*, se explica principalmente por: (i) La frecuencia de uso de las lámparas ahorradoras, *Ahorradora_Uso*, y (ii) el valor facturado del servicio de electricidad, *EE_Factura_Valor*. En la tabla 5 se explica la asignación de las variables de la encuesta a las variables del modelo.

Tabla 5: Asignación de las variables del Modelo *Mod_{2,3}*.

Variable de <i>Mod_{2,3}</i>	Variable de la encuesta
y_i	<i>EE_Factura_CkWhMes</i>
$x_{1,i}$	<i>Ahorradora_Uso</i>
$x_{2,i}$	<i>EE_Factura_Valor</i>
$x_{3,i}$	$\begin{cases} 1 & \text{si Gas_CkWhMes} \neq 0 \\ 0 & \text{si Gas_CkWhMes} = 0 \end{cases}$

La relación entre la variable dependiente y las variables explicativas se describe por medio de la siguiente ecuación:

$$y_i = \alpha_{0,A} + \alpha_{1,A}x_{1,i} + \alpha_{2,A}x_{2,i} + (\alpha_{0,B} + \alpha_{1,B}x_{1,i} + \alpha_{2,B}x_{2,i})x_{3,i} + \omega_i \quad , \quad (7)$$

⁵Como regla general siempre se elige una clase base y se crea un número de variables ficticias igual al número de clases menos uno.

donde el subíndice i identifica a cada observación; y_i es la variable dependiente que representa el consumo de energía eléctrica mensual de cada vivienda; $x_{1,i}$ y $x_{2,i}$ son las variables explicativas; $x_{3,i}$ es la variable ficticia; $\alpha_{0,A}$ y $\alpha_{0,B}$ son los interceptos; $\alpha_{1,A}$, $\alpha_{2,A}$, $\alpha_{1,B}$ y $\alpha_{2,B}$ son los coeficientes (se consideran constantes) y $\omega_i \stackrel{\text{iid}}{\sim} N(0, \sigma_i^2)$ son los errores independientes e idénticamente y normalmente distribuidos, con media cero y varianza σ_i^2 constante. Los coeficientes de la ecuación 7, incluyendo los interceptos, se estiman con 1.372 observaciones (98, 21 % de n). La figura 22 muestra los resultados de esta regresión.

```

Mod_2_3
-----
Número de observaciones:
1372 (98.21% de n)

Linear regression model (robust fit):
  Y_Mod_2_3 ~ 1 + X1_Mod_2_3*X3_Mod_2_3 + X2_Mod_2_3*X3_Mod_2_3

Estimated Coefficients:

```

	Estimate	SE	tStat	pValue
(Intercept)	21.02	1.2303	17.085	1.6093e-58
X1_Mod_2_3	0.10183	0.022988	4.4295	1.0342e-05
X2_Mod_2_3	0.0021279	8.8774e-05	23.97	2.1862e-103
X3_Mod_2_3_1	4.0892	1.9947	2.05	0.04059
X1_Mod_2_3:X3_Mod_2_3_1	-0.065265	0.028092	-2.3232	0.020339
X2_Mod_2_3:X3_Mod_2_3_1	0.00079443	0.00011529	6.8905	9.1183e-12

```

Number of observations: 1161, Error degrees of freedom: 1155

```

Figura 15: Resultados de la regresión de $Mod_{2,3}$.

En esta figura se lee que: (i) El número efectivo de observaciones utilizadas por el algoritmo es de 1.161, es decir, se han descartado 211 observaciones que se consideran como *outliers*, y (ii) el número de grados de libertad es $DoF = 1.161 - 6 = 1.155$. Se comprueba que hay suficientes observaciones para realizar una buena estimación. De la columna *Estimate* se puede leer los coeficientes estimados por el algoritmo; luego, la demanda de energía eléctrica residencial mensual para las viviendas rurales del Departamento de Nariño, teniendo en cuenta las clases de viviendas A y B , es:

- Para la clase A ($x_3 = 0$), es decir, viviendas que consumen energía eléctrica y leña:

$$\hat{y}_A = 21,02 + 0,10x_1 + 0,002x_2 \quad , \quad (8)$$

donde \hat{y}_A es el consumo estimado para cada vivienda de la clase A (kWh/mes), x_1 es la frecuencia de uso de lámparas ahorradoras (horas/mes) y x_2 es el valor

facturado del servicio de electricidad (COP/mes). Los coeficientes se interpretan así: (i) $\hat{\alpha}_{0,A} = 21,02$ es un consumo mínimo de subsistencia en kWh/mes cuando todas las variables explicativas se hacen cero, (ii) $\hat{\alpha}_{1,A} = 0,10$ es el incremento en el consumo en kW por cada hora/mes de uso de lámparas ahorradoras y (iii) $\hat{\alpha}_{2,A} = 0,002$ es el incremento en el consumo en kWh/COP por cada COP/mes pagado por el servicio.

- Para la clase B ($x_3 = 1$), es decir, viviendas que consumen energía eléctrica y gas:

$$\hat{y}_B = 25,11 + 0,03x_1 + 0,0028x_2 \quad , \quad (9)$$

donde \hat{y}_B es el consumo estimado para cada vivienda de la clase B (kWh/mes). Los coeficientes se interpretan así: (i) $\hat{\alpha}_{0,A} + \hat{\alpha}_{0,B} = 21,02 + 4,09 = 25,11$ es un consumo mínimo de subsistencia en kWh/mes cuando todas las variables explicativas se hacen cero, (ii) $\hat{\alpha}_{1,A} + \hat{\alpha}_{1,B} = 0,10 - 0,07 = 0,03$ es el incremento en el consumo en kW por cada hora/mes de uso de lámparas ahorradoras y $\hat{\alpha}_{2,A} + \hat{\alpha}_{2,B} = 0,002 + 0,0008 = 0,0028$ es el incremento en el consumo en kWh/COP por cada COP/mes pagado por el servicio.

2.3. *Mod₃*

Una de las variables que reporta la encuesta, y que tiene una de las correlaciones más altas con el consumo de energía eléctrica en las viviendas rurales del Departamento de Nariño, es el número de cuartos por vivienda. Esta variable ha sido reportado como significativa en la explicación de la demanda de energía eléctrica residencial y en esta regresión se propone una categorización del modelo de demanda por el tamaño de la vivienda.

Esta regresión se hace con una variable ficticia que facilita el cálculo de la diferencia de la demanda de energía eléctrica entre las viviendas de las siguientes clases: (i) La clase A de viviendas que tienen más de 4 cuartos y (ii) la clase B de viviendas que tienen 4 cuartos o menos. La clase base de la regresión es la clase A y los valores que toma la variable ficticia son: 1, cuando la vivienda pertenece a la clase B , y 0, cuando la vivienda no pertenece a esta clase (o pertenece a la clase A).

En esta regresión, las viviendas que reportan consumos superiores a 250 kWh/mes, que cancelan por el servicio más de 150.000 COP y con más 15 cuartos se descartan. A diferencia de la regresión *Mod_{2,3}*, en esta regresión no todas las variables explicativas se activan por medio de la variable ficticia. Se asume que el consumo mensual de energía eléctrica de las dos clases de viviendas, *EE_Factura_CkWhMes*, se explica principalmente por: (i) La frecuencia de uso de las lámparas ahorradoras, *Ahorradora_Uso*, (ii) el tamaño de los refrigeradores, *Refrigeracion_Tam*, (iii) el ingreso de cada vivienda, *Vivienda_Ingreso-COP*, y (iv) el valor facturado del servicio de electricidad, *EE_Factura_Valor*. En la tabla 6 se explica la asignación de las variables de la encuesta a las variables del modelo.

La relación entre la variable dependiente y las variables explicativas se describe por medio de la siguiente ecuación:

Tabla 6: Asignación de las variables del Modelo Mod_3 .

Variable de Mod_3	Variable de la encuesta
y_i	EE_Factura_CkWhMes
$x_{1,i}$	Ahorradora_Uso
$x_{2,i}$	Refrigeracion_Tam
$x_{3,i}$	Vivienda_Ingreso_COP
$x_{4,i}$	EE_Factura_Valor
$x_{5,i}$	$\begin{cases} 1 & \text{si Vivienda_Cuarto} \leq 4 \\ 0 & \text{si Vivienda_Cuarto} > 4 \end{cases}$

$$y_i = \alpha_{0,A} + \alpha_{1,A}x_{1,i} + \alpha_{2,A}x_{2,i} + \alpha_{3,A}x_{3,i} + \alpha_{4,A}x_{4,i} + (\alpha_{0,B} + \alpha_{3,B}x_{3,i} + \alpha_{4,B}x_{4,i})x_{5,i} + \omega_i \quad (10)$$

donde el subíndice i identifica a cada observación; y_i es la variable dependiente que representa el consumo de energía eléctrica mensual de cada vivienda; $x_{1,i}$ a $x_{4,i}$ son las variables explicativas; $x_{5,i}$ es la variable ficticia; $\alpha_{0,A}$ y $\alpha_{0,B}$ son los interceptos; $\alpha_{1,A}$ a $\alpha_{4,A}$ y $\alpha_{1,B}$ a $\alpha_{4,B}$ son los coeficientes (se consideran constantes) y $\omega_i \stackrel{iid}{\sim} N(0, \sigma_i^2)$ son los errores independientes e idénticamente y normalmente distribuidos, con media cero y varianza σ_i^2 constante. Los coeficientes de la ecuación 10, incluyendo los interceptos, se estiman con 1.368 observaciones (97, 92% de n). La figura 23 muestra los resultados de esta regresión.

```

Mod_3
-----
Número de observaciones:
1368 (97.92% de n)

Linear regression model (robust fit):
  Y_Mod_3 ~ 1 + X1_Mod_3 + X2_Mod_3 + X3_Mod_3*X5_Mod_3 + X4_Mod_3*X5_Mod_3

Estimated Coefficients:

```

	Estimate	SE	tStat	pValue
(Intercept)	20.009	1.2493	16.016	2.648e-52
X1_Mod_3	0.044053	0.01257	3.5046	0.00047467
X2_Mod_3	1.1582	0.097123	11.925	5.2604e-31
X3_Mod_3	9.6789e-06	1.898e-06	5.0996	3.9724e-07
X4_Mod_3	0.0023015	6.554e-05	35.115	1.4095e-184
X5_Mod_3_1	-7.3618	2.0124	-3.6582	0.00026542
X3_Mod_3:X5_Mod_3_1	-9.3112e-06	3.8059e-06	-2.4465	0.014573
X4_Mod_3:X5_Mod_3_1	0.00058479	0.00012717	4.5986	4.7195e-06

```

Number of observations: 1167, Error degrees of freedom: 1159

```

Figura 16: Resultados de la regresión de Mod_3 .

En esta figura se lee que: (i) El número efectivo de observaciones utilizadas por el algoritmo es de 1.167, es decir, se han descartado 201 observaciones que se consideran como *outliers*, y (ii) el número de grados de libertad es $DoF = 1.167 - 8 = 1.159$. Se comprueba que hay suficientes observaciones para realizar una buena estimación. De la columna *Estimate* se puede leer los coeficientes estimados por el algoritmo; luego, la demanda de energía eléctrica residencial mensual para las viviendas rurales del Departamento de Nariño, teniendo en cuenta las clases de viviendas *A* y *B*, es:

- Para la clase *A* ($x_5 = 0$), es decir, viviendas que consumen energía eléctrica y tienen más de 4 cuartos:

$$\hat{y}_A = 20,00 + 0,04x_1 + 1,16x_2 + 0,00001x_3 + 0,002x_4 \quad , \quad (11)$$

donde \hat{y}_A es el consumo estimado para cada vivienda de la clase *A* (kWh/mes), x_1 es la frecuencia de uso de lámparas ahorradoras (horas/mes), x_2 es el tamaño de equipos de refrigeración (ft³), x_3 es el ingreso por vivienda (COP/mes) y x_4 es el valor facturado del servicio de electricidad (COP/mes). Los coeficientes se interpretan así: (i) $\hat{\alpha}_{0,A} = 20,00$ es un consumo mínimo de subsistencia en kWh/mes cuando todas las variables explicativas se hacen cero, (ii) $\hat{\alpha}_{1,A} = 0,04$ es el incremento en el consumo en kW por cada hora/mes de uso de lámparas ahorradoras, (iii) $\hat{\alpha}_{2,A} = 1,16$ es el incremento en el consumo en kWh/mes/ft³ por cada ft³ de tamaño de refrigeradores, (iv) $\hat{\alpha}_{3,A} = 0,00001$ es el incremento en el consumo en kWh/COP por cada COP/mes de ingreso de la vivienda y (v) $\hat{\alpha}_{4,A} = 0,002$ es el incremento en el consumo en kWh/COP por cada COP/mes pagado por el servicio.

- Para la clase *B* ($x_5 = 1$), es decir, viviendas que consumen energía eléctrica y tienen 4 o menos cuartos:

$$\hat{y}_B = 19,64 + 0,04x_1 + 1,16x_2 + 0,00x_3 + 0,003x_4 \quad , \quad (12)$$

donde \hat{y}_B es el consumo estimado para cada vivienda de la clase *B* (kWh/mes). Los coeficientes se interpretan así: (i) $\hat{\alpha}_{0,A} + \hat{\alpha}_{0,B} = 20,00 - 7,36 = 19,64$ es un consumo mínimo de subsistencia en kWh/mes cuando todas las variables explicativas se hacen cero, (ii) $\hat{\alpha}_{3,A} + \hat{\alpha}_{3,B} = 0,00001 - 0,00001 \approx 0,00$ es el crecimiento en el consumo en kWh/COP por cada COP/mes de ingreso de la vivienda y (iii) $\hat{\alpha}_{4,A} + \hat{\alpha}_{4,B} = 0,0023 + 0,0006 \approx 0,003$ es el incremento en el consumo en kWh/COP por cada COP/mes pagado por el servicio.

2.4. *Mod*₄

Otra de las variables que reporta la encuesta, y que también muestra una de las correlaciones más altas con el consumo de energía eléctrica en las viviendas rurales del

Departamento de Nariño, es el ingreso por vivienda. En esta regresión se explora la diferencia en la demanda de electricidad de las viviendas que se ubican en diferentes categorías de ingreso.

Para esto, las viviendas se categorizan en las siguientes clases: (i) La clase *A* de viviendas con ingresos en el rango de $[0, 350.000]$ COP, (ii) la clase *B* de viviendas con ingresos en el rango de $(350.000, 700.000]$ COP y (iii) la clase *C* de viviendas con ingresos de más de 700.000 COP. La clase base de la regresión se hace corresponder a la clase *A* y se definen dos variables ficticias. La primera variable ficticia toma los siguientes valores: 1, cuando la vivienda pertenece a la clase *B*, y 0, cuando la vivienda no pertenece a esta clase. La segunda variable ficticia toma los siguientes valores: 1, cuando la vivienda pertenece a la clase *C*, y 0, cuando la vivienda no pertenece a esta clase.

En esta regresión, las viviendas que reportan consumos superiores a 250 kWh/mes, que cancelan por el servicio más de 150.000 COP y con más 1.500.000 de ingreso se descartan. Al igual que la regresión *Mod*₃, en esta regresión no todas las variables explicativas se activan por medio de las variables ficticias. Se asume que el consumo mensual de energía eléctrica de las dos clases de viviendas, EE_Factura_CkWhMes, se explica principalmente por: (i) el tamaño de los refrigeradores, Refrigeracion_Tam, (ii) el número de cuartos de cada vivienda, Vivienda_Cuarto, y (iii) el valor facturado del servicio de electricidad, EE_Factura_Valor. En la tabla 7 se explica la asignación de las variables de la encuesta a las variables del modelo.

Tabla 7: Asignación de las variables del Modelo *Mod*₄.

Variable de <i>Mod</i> ₄	Variable de la encuesta
y_i	EE_Factura_CkWhMes
$x_{1,i}$	Refrigeracion_Tam
$x_{2,i}$	Vivienda_Cuarto
$x_{3,i}$	EE_Factura_Valor
$x_{4,i}$	$\begin{cases} 1 & \text{si } 350.000 < \text{Vivienda_Ingreso_COP} \leq 700.000 \\ 0 & \text{en otro caso} \end{cases}$
$x_{5,i}$	$\begin{cases} 1 & \text{si } 700.000 < \text{Vivienda_Ingreso_COP} \\ 0 & \text{en otro caso} \end{cases}$

La relación entre la variable dependiente y las variables explicativas se describe por medio de la siguiente ecuación:

$$\begin{aligned}
 y_i = & \alpha_{0,A} + \alpha_{1,A}x_{1,i} + \alpha_{2,A}x_{2,i} + \alpha_{3,A}x_{3,i} \\
 & + (\alpha_{0,B} + \alpha_{1,B}x_{1,i} + \alpha_{3,B}x_{3,i})x_{4,i} \\
 & + (\alpha_{0,C} + \alpha_{3,C}x_{3,i})x_{5,i} + \omega_i \quad ,
 \end{aligned} \tag{13}$$

donde el subíndice *i* identifica a cada observación; y_i es la variable dependiente que representa el consumo de energía eléctrica mensual de cada vivienda; $x_{1,i}$, $x_{2,i}$ y $x_{3,i}$ son las variables explicativas; $x_{4,i}$ y $x_{5,i}$ son las variables ficticias; $\alpha_{0,A}$, $\alpha_{0,B}$ y $\alpha_{0,C}$ son los

interceptos; $\alpha_{1,A}$, $\alpha_{2,A}$, $\alpha_{3,A}$, $\alpha_{1,B}$, $\alpha_{3,B}$, y $\alpha_{3,C}$ son los coeficientes (se consideran constantes) y $\omega_i \stackrel{iid}{\sim} N(0, \sigma_i^2)$ son los errores independientes e idénticamente y normalmente distribuidos, con media cero y varianza σ_i^2 constante. Los coeficientes de la ecuación 13, incluyendo los interceptos, se estiman con 1.347 observaciones (96,42% de n). La figura 17 muestra los resultados de esta regresión.

```

Mod_4
-----
Número de observaciones:
1347 (96.42% de n)

Linear regression model (robust fit):
  Y_Mod_4 ~ 1 + X2_Mod_4 + X1_Mod_4*X4_Mod_4 + X3_Mod_4*X4_Mod_4 + X3_Mod_4*X5_Mod_4

Estimated Coefficients:

```

	Estimate	SE	tStat	pValue
(Intercept)	4.8796	1.7556	2.7794	0.005535
X1_Mod_4	1.3784	0.15163	9.0907	4.1925e-19
X2_Mod_4	1.9245	0.31064	6.1955	8.0757e-10
X3_Mod_4	0.0030189	8.2689e-05	36.51	1.5811e-194
X4_Mod_4	18.659	2.0532	9.0875	4.3074e-19
X5_Mod_4	13.766	3.6844	3.7363	0.0001959
X1_Mod_4:X4_Mod_4	-1.1549	0.17087	-6.7594	2.1964e-11
X3_Mod_4:X4_Mod_4	-0.0014067	0.00011322	-12.424	2.3715e-33
X3_Mod_4:X5_Mod_4	-0.00042359	0.0001489	-2.8447	0.0045237

```

Number of observations: 1161, Error degrees of freedom: 1152

```

Figura 17: Resultados de la regresión de Mod_4 .

En esta figura se lee que: (i) El número efectivo de observaciones utilizadas por el algoritmo es de 1.161, es decir, se han descartado 186 observaciones que se consideran como *outliers*, y (ii) el número de grados de libertad es $DoF = 1.161 - 9 = 1.152$. Se comprueba que hay suficientes observaciones para realizar una buena estimación. De la columna *Estimate* se puede leer los coeficientes estimados por el algoritmo; luego, la demanda de energía eléctrica residencial mensual para las viviendas rurales del Departamento de Nariño, teniendo en cuenta las clases de viviendas A , B y C , es:

- Para la clase A ($x_4 = 0$ y $x_5 = 0$), es decir, viviendas que consumen energía eléctrica y tienen ingresos en el rango de $[0, 350.000]$ COP:

$$\hat{y}_A = 4,88 + 1,38x_1 + 1,92x_2 + 0,003x_3 \quad , \quad (14)$$

donde \hat{y}_A es el consumo estimado para cada vivienda de la clase A (kWh/mes), x_1 es el tamaño de equipos de refrigeración (ft^3), x_2 es el número de cuartos que tiene la vivienda (cuartos/vivienda) y x_3 es el valor facturado del servicio de electricidad (COP/mes). Los coeficientes se interpretan así: (i) $\hat{\alpha}_{0,A} = 4,88$ es un

consumo mínimo de subsistencia en kWh/mes cuando todas las variables explicativas se hacen cero, $\hat{\alpha}_{1,A} = 1,38$ es el incremento en el consumo en kWh/mes/ft³ por cada ft³ de tamaño de refrigeradores, $\hat{\alpha}_{2,A} = 1,92$ es el incremento en el consumo en kWh/mes/cuarto/vivienda por cada cuarto/vivienda y $\hat{\alpha}_{3,A} = 0,003$ es el incremento en el consumo en kWh/COP por cada COP/mes pagado por el servicio.

- Para la clase B ($x_4 = 1$ y $x_5 = 0$), es decir, viviendas que consumen energía eléctrica y tienen ingresos en el rango de (350.000, 700.000] COP:

$$\hat{y}_B = 23,54 + 0,23x_1 + 1,92x_2 + 0,002x_3 \quad , \quad (15)$$

donde \hat{y}_B es el consumo estimado para cada vivienda de la clase B (kWh/mes). Los coeficientes se interpretan así: (i) $\hat{\alpha}_{0,A} + \hat{\alpha}_{0,B} = 4,88 + 18,66 = 23,54$ es un consumo mínimo de subsistencia en kWh/mes cuando todas las variables explicativas se hacen cero, (ii) $\hat{\alpha}_{1,A} + \hat{\alpha}_{1,B} = 1,38 - 1,15 = 0,23$ es el incremento en el consumo en kWh/mes/ft³ por cada ft³ de tamaño de refrigeradores (iii) y $\hat{\alpha}_{3,A} + \hat{\alpha}_{3,B} = 0,003 - 0,001 \approx 0,002$ es el incremento en el consumo en kWh/COP por cada COP/mes pagado por el servicio.

- Para la clase C ($x_4 = 0$ y $x_5 = 1$), es decir, viviendas que consumen energía eléctrica y tienen ingresos mayores a 700.000 COP:

$$\hat{y}_C = 18,65 + 1,38x_1 + 1,92x_2 + 0,0026x_3 \quad , \quad (16)$$

donde \hat{y}_C es el consumo estimado para cada vivienda de la clase C (kWh/mes). Los coeficientes se interpretan así: (i) $\hat{\alpha}_{0,A} + \hat{\alpha}_{0,C} = 4,88 + 13,77 = 18,65$ es un consumo mínimo de subsistencia en kWh/mes cuando todas las variables explicativas se hacen cero, y (ii) $\hat{\alpha}_{3,A} + \hat{\alpha}_{3,C} = 0,003 - 0,0004 \approx 0,003$ es el incremento en el consumo en kWh/COP por cada COP/mes pagado por el servicio.

2.5. Mod_5

Finalmente, aprovechando que la encuesta reporta el estrato al que pertenecen las viviendas de la muestra, se construye una regresión categórica que permite observar la diferencia del consumo de energía eléctrica de las viviendas que pertenecen a distinto estrato. Para esto, se definen las siguientes clases: (i) La clase A de viviendas que pertenecen al estrato 1 y la clase B de viviendas que pertenecen al estrato 2. La clase base de la regresión se hace corresponder a la clase A y se define una variable ficticia que asume los siguientes valores: 1, cuando la vivienda pertenece a la clase B , y 0, cuando la vivienda no pertenece a esta clase.

En esta regresión, las viviendas que reportan consumos superiores a 250 kWh/mes, que cancelan por el servicio más de 150.000 COP y que pertenecen al estrato 3 se descartan. Al igual que la regresión Mod_4 , en esta regresión no todas las variables explicativas se activan por medio de la variable ficticia. Se asume que el consumo mensual de energía

eléctrica de las dos clases de viviendas, `EE_Factura_CkWhMes`, se explica principalmente por: (i) el tamaño de los refrigeradores, `Refrigeracion_Tam`, (ii) el número de cuartos de cada vivienda, `Vivienda_Cuarto`, y (iii) el valor facturado del servicio de electricidad, `EE_Factura_Valor`. En la tabla 8 se explica la asignación de las variables de la encuesta a las variables del modelo.

Tabla 8: Asignación de las variables del Modelo Mod_5 .

Variable de Mod_5	Variable de la encuesta
y_i	<code>EE_Factura_CkWhMes</code>
$x_{1,i}$	<code>Refrigeracion_Tam</code>
$x_{2,i}$	<code>Vivienda_Cuarto</code>
$x_{3,i}$	<code>EE_Factura_Valor</code>
$x_{4,i}$	$\begin{cases} 1 & \text{si EE_Factura_Estrato} = 2 \\ 0 & \text{en otro caso} \end{cases}$

La relación entre la variable dependiente y las variables explicativas se describe por medio de la siguiente ecuación:

$$y_i = \alpha_{0,A} + \alpha_{1,A}x_{1,i} + \alpha_{2,A}x_{2,i} + \alpha_{3,A}x_{3,i} + (\alpha_{0,B} + \alpha_{1,B}x_{1,i} + \alpha_{3,B}x_{3,i})x_{4,i} + \omega_i \quad , \quad (17)$$

donde el subíndice i identifica a cada observación; y_i es la variable dependiente que representa el consumo de energía eléctrica mensual de cada vivienda; $x_{1,i}$, $x_{2,i}$ y $x_{3,i}$ son las variables explicativas; $x_{4,i}$ es la variable ficticia; $\alpha_{0,A}$ y $\alpha_{0,B}$ son los interceptos; $\alpha_{1,A}$, $\alpha_{2,A}$, $\alpha_{3,A}$, $\alpha_{1,B}$ y $\alpha_{3,B}$ son los coeficientes (se consideran constantes) y $\omega_i \stackrel{\text{iid}}{\sim} N(0, \sigma_i^2)$ son los errores independientes e idénticamente y normalmente distribuidos, con media cero y varianza σ_i^2 constante. Los coeficientes de la ecuación 17, incluyendo los interceptos, se estiman con 1.368 observaciones (97,92% de n). La figura 18 muestra los resultados de esta regresión.

En esta figura se lee que: (i) El número efectivo de observaciones utilizadas por el algoritmo es de 1.167, es decir, se han descartado 201 observaciones que se consideran como *outliers*, y (ii) el número de grados de libertad es $DoF = 1.167 - 7 = 1.160$. Se comprueba que hay suficientes observaciones para realizar una buena estimación. De la columna *Estimate* se puede leer los coeficientes estimados por el algoritmo; luego, la demanda de energía eléctrica residencial mensual para las viviendas rurales del Departamento de Nariño, teniendo en cuenta las clases de viviendas A y B , es:

- Para la clase A ($x_4 = 0$), es decir, viviendas que consumen energía eléctrica y pertenecen al estrato 1:

$$\hat{y}_A = 10,21 + 1,27x_1 + 2,08x_2 + 0,002x_3 \quad , \quad (18)$$

donde \hat{y}_A es el consumo estimado para cada vivienda de la clase A (kWh/mes), x_1 es el tamaño de equipos de refrigeración (ft^3), x_2 es el número de cuartos


```

Mod_5
-----
Número de observaciones:
1368 (97.92% de n)

Linear regression model (robust fit):
  Y_Mod_5 ~ 1 + X2_Mod_5 + X1_Mod_5*X4_Mod_5 + X3_Mod_5*X4_Mod_5

Estimated Coefficients:

```

	Estimate	SE	tStat	pValue
(Intercept)	10.211	1.7309	5.8992	4.787e-09
X1_Mod_5	1.2723	0.098506	12.916	9.3262e-36
X2_Mod_5	2.0801	0.29992	6.9354	6.7125e-12
X3_Mod_5	0.0023735	6.7726e-05	35.046	4.1281e-184
X4_Mod_5	-6.8553	2.2938	-2.9886	0.0028613
X1_Mod_5:X4_Mod_5	-1.216	0.13678	-8.8903	2.2841e-18
X3_Mod_5:X4_Mod_5	0.00080115	0.00011644	6.8807	9.7246e-12

```

Number of observations: 1167, Error degrees of freedom: 1160

```

Figura 18: Resultados de la regresión de Mod_5 .

que tiene la vivienda (cuartos/vivienda) y x_3 es el valor facturado del servicio de electricidad (COP/mes). Los coeficientes se interpretan así: (i) $\hat{\alpha}_{0,A} = 10,21$ es un consumo mínimo de subsistencia en kWh/mes cuando todas las variables explicativas se hacen cero, (ii) $\hat{\alpha}_{1,A} = 1,27$ es el incremento en el consumo en kWh/mes/ft³ por cada ft³ de tamaño de refrigeradores, (iii) $\hat{\alpha}_{2,A} = 2,08$ es el incremento en el consumo en kWh/mes/cuarto/vivienda por cada cuarto/vivienda y (iv) $\hat{\alpha}_{3,A} = 0,002$ es el incremento en el consumo en kWh/COP por cada COP/mes pagado por el servicio.

- Para la clase B ($x_4 = 1$), es decir, viviendas que consumen energía eléctrica y pertenecen al estrato 2:

$$\hat{y}_B = 3,35 + 0,05x_1 + 2,08x_2 + 0,003x_3 \quad , \quad (19)$$

donde \hat{y}_B es el consumo estimado para cada vivienda de la clase B (kWh/mes). Los coeficientes se interpretan así: (i) $\hat{\alpha}_{0,A} + \hat{\alpha}_{0,B} = 10,21 - 6,86 = 3,35$ es un consumo mínimo de subsistencia en kWh/mes cuando todas las variables explicativas se hacen cero, (ii) $\hat{\alpha}_{1,A} + \hat{\alpha}_{1,B} = 1,27 - 1,22 = 0,05$ es el incremento en el consumo en kWh/mes/ft³ por cada ft³ de tamaño de refrigeradores y (iii) $\hat{\alpha}_{3,A} + \hat{\alpha}_{3,B} = 0,002 + 0,0008 \approx 0,003$ es el incremento en el consumo en kWh/COP por cada COP/mes pagado por el servicio.

3. Calidad de la regresión

Las regresiones que se calcularon se evalúan con distintos instrumentos que permiten medir la precisión y la validez de los resultados obtenidos. A continuación se hace la lectura y la aplicación de algunos de estos instrumentos a los ejemplos de regresión ya presentados.

3.1. Error Estándar SE (Standar Error)

En cada regresión se muestra el error estándar SE (columna **SE**) de cada uno de los coeficientes estimados. Este valor mide la variabilidad que tienen el coeficiente estimado debido a que las observaciones se toman de muestra aleatoria de la población, en este caso las observaciones corresponden a una muestra de viviendas del Departamento de Nariño. El error estándar para cada coeficiente se calcula como la relación entre la desviación de los residuos (errores) y la desviación de la variable asociada al coeficiente, es decir,

$$SE(\alpha_k) = \sqrt{\frac{\sum_{i=1}^n e_i^2}{DoF \sum_{i=1}^n (x_{k,i} - \bar{x}_k)^2}} \quad , \quad (20)$$

donde k es el número de coeficientes a estimar, $SE(\alpha_k)$ es el error estándar del coeficiente α_k , $e_i = y_i - \hat{y}_i$ es el error o residuo de la estimación, n es el número de observaciones efectivas utilizadas en la regresión, DoF es el número de grados de libertad, $x_{k,i}$ es cada observación de la variable explicativa asociada al coeficiente α_k y \bar{x}_k es el valor medio de dicha variable.

Los valores del SE (segunda columna), para cada regresión (primera columna) y para cada coeficiente (tercera columna), se muestran en la tabla 3.1. En esta tabla se observa que los coeficientes que miden la misma relación y que se asocian a más de una regresión tienen un SE casi igual en todas ellas⁶. En este caso se concluye que las variables asociadas a cada coeficiente se incluyeron de manera acertada en cada regresión. Por ejemplo, el coeficiente que mide la razón entre el consumo y la frecuencia de uso de lámparas ahorradoras en las regresiones Mod_1 , $Mod_{2,2}$, $Mod_{2,3}$ y Mod_3 , tiene un SE casi igual para todas ellas y por debajo de 0,05; de igual forma, el coeficiente que mide la razón entre el consumo y el valor pagado por el servicio de electricidad tiene un SE prácticamente igual en todas las regresiones y no superior a 0,0001.

Tabla 9: Error estándar SE .

Regresión	$SE(\hat{\alpha}_k)$	Coficiente
Mod_1	1,9925	● $\hat{\alpha}_0$: Intercepto
	0,0072	● $\hat{\alpha}_1$: (consumo mensual)/(frec. uso incandescente)
	0,0122	● $\hat{\alpha}_2$: (consumo mensual)/(frec. uso ahorradora)
	0,0859	● $\hat{\alpha}_3$: (consumo mensual)/(tam. Refrigeradores)
	0,3105	● $\hat{\alpha}_4$: (consumo mensual)/(personas/vivienda)
	0,3183	● $\hat{\alpha}_5$: (consumo mensual)/(cuartos/vivienda)

⁶Las regresiones que reportaron valores de SE más grandes que los que se reportan en esta tabla fueron descartadas.

Regresión	$SE(\hat{\alpha}_k)$	Coficiente
	0,0000 0,0001	<ul style="list-style-type: none"> ● $\hat{\alpha}_6$: (consumo mensual)/(ingresos/vivienda) ● $\hat{\alpha}_7$: (consumo mensual)/(valor pagado)
$Mod_{2,1}$	3,3310 0,1395 0,5064 0,4613 0,0001	<ul style="list-style-type: none"> ● $\hat{\alpha}_0$: Intercepto ● $\hat{\alpha}_1$: (consumo mensual)/(tam. Refrigeradores) ● $\hat{\alpha}_2$: (consumo mensual)/(personas/vivienda) ● $\hat{\alpha}_3$: (consumo mensual)/(cuartos/vivienda) ● $\hat{\alpha}_4$: (consumo mensual)/(valor pagado)
$Mod_{2,2}$	1,8377 0,0095 0,0197 0,1542 0,3747 0,0001	<ul style="list-style-type: none"> ● $\hat{\alpha}_0$: Intercepto ● $\hat{\alpha}_1$: (consumo mensual)/(frec. uso incandescente) ● $\hat{\alpha}_2$: (consumo mensual)/(frec. uso ahorradora) ● $\hat{\alpha}_3$: (consumo mensual)/(tam. Refrigeradores) ● $\hat{\alpha}_4$: (consumo mensual)/(personas/vivienda) ● $\hat{\alpha}_5$: (consumo mensual)/(valor pagado)
$Mod_{2,3}$	1,2303 0,0230 0,0001 1,9947 0,0281 0,0001	<ul style="list-style-type: none"> ● $\hat{\alpha}_{0,A}$: Intercepto ● $\hat{\alpha}_{1,A}$: (consumo mensual)/(frec. uso ahorradora) ● $\hat{\alpha}_{2,A}$: (consumo mensual)/(valor pagado) ● $\hat{\alpha}_{0,B}$: Intercepto ● $\hat{\alpha}_{1,B}$: (consumo mensual)/(frec. uso ahorradora) ● $\hat{\alpha}_{2,B}$: (consumo mensual)/(valor pagado)
Mod_3	1,2493 0,0126 0,0971 0,0000 0,0001 2,0124 0,0000 0,0001	<ul style="list-style-type: none"> ● $\hat{\alpha}_0$: Intercepto ● $\hat{\alpha}_{1,A}$: (consumo mensual)/(frec. uso ahorradora) ● $\hat{\alpha}_{2,A}$: (consumo mensual)/(tam. Refrigeradores) ● $\hat{\alpha}_{3,A}$: (consumo mensual)/(ingresos/vivienda) ● $\hat{\alpha}_{4,A}$: (consumo mensual)/(valor pagado) ● $\hat{\alpha}_{0,B}$: Intercepto ● $\hat{\alpha}_{3,B}$: (consumo mensual)/(ingresos/vivienda) ● $\hat{\alpha}_{4,B}$: (consumo mensual)/(valor pagado)
Mod_4	1,7556 0,1516 0,3106 0,0001 2,0532 3,6844 0,1709 0,0001 0,0001	<ul style="list-style-type: none"> ● $\hat{\alpha}_{0,A}$: Intercepto ● $\hat{\alpha}_{1,A}$: (consumo mensual)/(tam. Refrigeradores) ● $\hat{\alpha}_{2,A}$: (consumo mensual)/(cuartos/vivienda) ● $\hat{\alpha}_{3,A}$: (consumo mensual)/(valor pagado) ● $\hat{\alpha}_{0,B}$: Intercepto ● $\hat{\alpha}_{0,C}$: Intercepto ● $\hat{\alpha}_{1,B}$: (consumo mensual)/(tam. Refrigeradores) ● $\hat{\alpha}_{3,C}$: (consumo mensual)/(valor pagado) ● $\hat{\alpha}_{3,C}$: (consumo mensual)/(valor pagado)
Mod_5	1,7309 0,0985 0,2999 0,0001 2,2938 0,1368 0,0001	<ul style="list-style-type: none"> ● $\hat{\alpha}_{0,A}$: Intercepto ● $\hat{\alpha}_{1,A}$: (consumo mensual)/(tam. Refrigeradores) ● $\hat{\alpha}_{2,A}$: (consumo mensual)/(cuartos/vivienda) ● $\hat{\alpha}_{3,A}$: (consumo mensual)/(valor pagado) ● $\hat{\alpha}_{0,B}$: Intercepto ● $\hat{\alpha}_{1,B}$: (consumo mensual)/(tam. Refrigeradores) ● $\hat{\alpha}_{3,B}$: (consumo mensual)/(valor pagado)

3.2. Estadísticas t y p

Las estadísticas t (columna **tStat**) y p (columna **pValue**) también se reportan en cada regresión. El estadístico t es el resultado de la hipótesis nula de que cada coeficiente,

e individualmente, es igual a cero. Este estadístico tiene una distribución *student* con *DoF* grados de libertad que se define como:

$$\frac{\hat{\alpha}_k}{SE(\hat{\alpha}_k)} \sim t \quad (21)$$

Este resultado, que corresponde a la razón entre las columnas **Estimate** y **SE**, es decir, entre el coeficiente estimado y su error estándar, se compara con el valor de la tabla *t*-student para un determinado nivel de significancia. Para un nivel de significancia del 5%, y para *DoF* > 30, los valores críticos son $-1,96$ y $1,96$; por lo tanto, si el valor de *t* se encuentra por fuera del intervalo $(-1,96, 1,96)$, entonces la hipótesis nula se rechaza y se considera que el coeficiente relacionado al estadístico es diferente de cero.

Por su parte, el estadístico *p* indica cuan probable es que el coeficiente obtenido en la regresión corresponde al coeficiente de la hipótesis nula, es decir, cero. Como regla general, valores más pequeños que 0,05 indican que la probabilidad de observar el coeficiente estimado, con respecto al que debería ser, es cero, y por lo tanto no se debería considerar en la regresión.

Los valores de *t* (segunda columna) y *p* (tercera columna), para cada regresión (primera columna) y para cada coeficiente (cuarta columna), se muestran en la tabla 3.2. En esta tabla se observa que para todos los coeficientes, de cada una de las regresiones calculadas, los estadísticos *t* se encuentran por fuera del intervalo crítico $(-1,96, 1,96)$, es decir, todos los coeficientes son significativos al 5%; de igual modo, las probabilidades *p* son todas inferior a 0,05. Se concluye que todas las variables incluidas en las regresiones son significativas.

Tabla 10: Estadísticas *t* y *p*.

Regresión	<i>t</i>	<i>p</i>	Coefficiente
<i>Mod</i> ₁	2,2152	0,0269	● $\hat{\alpha}_0$: Intercepto
	2,4228	0,0156	● $\hat{\alpha}_1$: (consumo mensual)/(frec. uso incandescente)
	3,7017	0,0002	● $\hat{\alpha}_2$: (consumo mensual)/(frec. uso ahorradora)
	11,4479	0,0000	● $\hat{\alpha}_3$: (consumo mensual)/(tam. Refrigeradores)
	3,2598	0,0011	● $\hat{\alpha}_4$: (consumo mensual)/(personas/vivienda)
	5,3100	0,0000	● $\hat{\alpha}_5$: (consumo mensual)/(cuartos/vivienda)
	2,5943	0,0096	● $\hat{\alpha}_6$: (consumo mensual)/(ingresos/vivienda)
41,4532	0,0000	● $\hat{\alpha}_7$: (consumo mensual)/(valor pagado)	
<i>Mod</i> _{2,1}	3,4389	0,0006	● $\hat{\alpha}_0$: Intercepto
	4,4964	0,0000	● $\hat{\alpha}_1$: (consumo mensual)/(tam. Refrigeradores)
	2,5998	0,0096	● $\hat{\alpha}_2$: (consumo mensual)/(personas/vivienda)
	2,5030	0,0126	● $\hat{\alpha}_3$: (consumo mensual)/(cuartos/vivienda)
	34,0006	0,0000	● $\hat{\alpha}_4$: (consumo mensual)/(valor pagado)
<i>Mod</i> _{2,2}	6,0295	0,0000	● $\hat{\alpha}_0$: Intercepto
	3,4177	0,0006	● $\hat{\alpha}_1$: (consumo mensual)/(frec. uso incandescente)
	3,7007	0,0002	● $\hat{\alpha}_2$: (consumo mensual)/(frec. uso ahorradora)
	12,2044	0,0000	● $\hat{\alpha}_3$: (consumo mensual)/(tam. Refrigeradores)
	4,2262	0,0000	● $\hat{\alpha}_4$: (consumo mensual)/(personas/vivienda)
	20,8805	0,0000	● $\hat{\alpha}_5$: (consumo mensual)/(valor pagado)
<i>Mod</i> _{2,3}	17,0847	0,0000	● $\hat{\alpha}_{0,A}$: Intercepto
	4,4295	0,0000	● $\hat{\alpha}_{1,A}$: (consumo mensual)/(frec. uso ahorradora)

Regresión	t	p	Coefficiente
	23,9704	0,0000	• $\hat{\alpha}_{2,A}$: (consumo mensual)/(valor pagado)
	2,0500	0,0406	• $\hat{\alpha}_{0,B}$: Intercepto
	-2,3232	0,0203	• $\hat{\alpha}_{1,B}$: (consumo mensual)/(frec. uso ahorradora)
	6,8905	0,0000	• $\hat{\alpha}_{2,B}$: (consumo mensual)/(valor pagado)
Mod_3	16,0157	0,0000	• $\hat{\alpha}_0$: Intercepto
	3,5046	0,0005	• $\hat{\alpha}_{1,A}$: (consumo mensual)/(frec. uso ahorradora)
	11,9249	0,0000	• $\hat{\alpha}_{2,A}$: (consumo mensual)/(tam. Refrigeradores)
	5,0996	0,0000	• $\hat{\alpha}_{3,A}$: (consumo mensual)/(ingresos/vivienda)
	35,1153	0,0000	• $\hat{\alpha}_{4,A}$: (consumo mensual)/(valor pagado)
	-3,6582	0,0003	• $\hat{\alpha}_{0,B}$: Intercepto
	-2,4465	0,0146	• $\hat{\alpha}_{3,B}$: (consumo mensual)/(ingresos/vivienda)
	4,5986	0,0000	• $\hat{\alpha}_{4,B}$: (consumo mensual)/(valor pagado)
Mod_4	2,7794	0,0055	• $\hat{\alpha}_{0,A}$: Intercepto
	9,0907	0,0000	• $\hat{\alpha}_{1,A}$: (consumo mensual)/(tam. Refrigeradores)
	6,1955	0,0000	• $\hat{\alpha}_{2,A}$: (consumo mensual)/(cuartos/vivienda)
	36,5097	0,0000	• $\hat{\alpha}_{3,A}$: (consumo mensual)/(valor pagado)
	9,0875	0,0000	• $\hat{\alpha}_{0,B}$: Intercepto
	3,7363	0,0002	• $\hat{\alpha}_{0,C}$: Intercepto
	-6,7594	0,0000	• $\hat{\alpha}_{1,B}$: (consumo mensual)/(tam. Refrigeradores)
	-12,4245	0,0000	• $\hat{\alpha}_{3,C}$: (consumo mensual)/(valor pagado)
	-2,8447	0,0045	• $\hat{\alpha}_{3,C}$: (consumo mensual)/(valor pagado)
Mod_5	5,8992	0,0000	• $\hat{\alpha}_{0,A}$: Intercepto
	12,9163	0,0000	• $\hat{\alpha}_{1,A}$: (consumo mensual)/(tam. Refrigeradores)
	6,9354	0,0000	• $\hat{\alpha}_{2,A}$: (consumo mensual)/(cuartos/vivienda)
	35,0460	0,0000	• $\hat{\alpha}_{3,A}$: (consumo mensual)/(valor pagado)
	-2,9886	0,0029	• $\hat{\alpha}_{0,B}$: Intercepto
	-8,8903	0,0000	• $\hat{\alpha}_{1,B}$: (consumo mensual)/(tam. Refrigeradores)
	6,8807	0,0000	• $\hat{\alpha}_{3,B}$: (consumo mensual)/(valor pagado)

3.3. Coeficientes de determinación R^2 y R_a^2 y estadística F

Los coeficientes de determinación R^2 y R_a^2 , denominados respectivamente **R-squared** y **Adjusted R-Squared** en los resultados de cada regresión, son medidas de la calidad del ajuste del modelo y miden el porcentaje de la varianza de la variable dependiente que se puede explicar con las variables explicativas. Estos coeficientes se definen respectivamente como :

$$R^2 = 1 - \frac{SS_R}{SS_T} \quad (22)$$

y

$$R_a^2 = 1 - \left(\frac{n-1}{n-k} \right) \frac{SS_R}{SS_T} \quad , \quad (23)$$

donde $SS_R = \sum_{j=1}^n e_j^2 = \sum_{j=1}^n (y_j - \hat{y}_j)^2$ es la suma del cuadrado de los residuos, $SS_T = \sum_{j=1}^n (y_j - \bar{y})^2$ es la suma total de cuadrados, \bar{y} es el valor medio de los valores observados de la variable dependiente, n es el número de observaciones efectivas y k es

el número de coeficientes del modelo, incluido el intercepto. El valor de R_a^2 se ajusta con los grados de libertad DoF del modelo y es útil cuando se poseen pocas observaciones para realizar la estimación.

De otra parte, el valor de la estadística F , que se lee en los resultados de la regresión como **F-statistics**, corresponde a la prueba F de la hipótesis nula de que todos los coeficientes del modelo son cero excepto la constante, y permite determinar la significancia del modelo en términos generales. Desde otro punto de vista, esta prueba también permite definir si el valor de los coeficientes de determinación son significativamente diferentes de cero.

Los valores de R^2 (segunda columna), R_a^2 (tercera columna) y F (cuarta columna), para cada regresión (primera columna), se muestran en la tabla 3.3. En esta tabla, los valores de los coeficientes de determinación, R^2 y R_a^2 , indican que las regresiones que tienen mejor explicación de la varianza del modelo son, en orden: Mod_4 , Mod_5 , $Mod_{2,1}$, Mod_1 , Mod_3 , $Mod_{2,3}$ y $Mod_{2,2}$; siendo Mod_4 la regresión con los valores más altos de R^2 y R_a^2 y $Mod_{2,2}$ la regresión con los valores más bajos. Los valores de la prueba F corroboran los valores respectivos de los coeficientes de determinación y se concluye que, en general, todos los modelos son significativos.

Tabla 11: Coeficientes de determinación R^2 y R_a^2 y estadística F .

Regresión	R^2	R_a^2	F (probabilidad)
Mod_1	0,7342	0,7326	455 (0,0000)
$Mod_{2,1}$	0,7429	0,7409	363 (0,0000)
$Mod_{2,2}$	0,6271	0,6241	207 (0,0000)
$Mod_{2,3}$	0,7207	0,7195	596 (0,0000)
Mod_3	0,7218	0,7201	429 (0,0000)
Mod_4	0,7509	0,7492	434 (0,0000)
Mod_5	0,7441	0,7428	562 (0,0000)

3.4. Raíz del error cuadrático medio $RMSE$ (Root Mean Squared Error)

El $RMSE$ es una medida directa del error entre los valores estimados de la variable dependiente y sus observaciones. En este caso, el $RMSE$ se interpreta como una medida de la dispersión general de cada regresión con respecto a los valores observados de consumo de energía eléctrica. El $RMSE$ se expresa en la misma unidad de la variable dependiente y junto con los coeficientes de determinación, R^2 y R_a^2 , y la estadística F se usa para medir la calidad del ajuste de las regresiones. Esta medida se define como:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n e_i^2}{DoF}} \quad (24)$$

Los valores del $RMSE$ (segunda columna) para cada regresión (primera columna) se muestran en la tabla 3.4. Los valores que se leen en esta tabla indican que las regresiones

que tienen mejor desempeño con respecto a los datos observados de consumo son, en orden: $Mod_{2,2}$, Mod_4 , Mod_5 , Mod_1 , Mod_3 , $Mod_{2,3}$ y $Mod_{2,1}$; siendo $Mod_{2,2}$ la regresión con el valor más bajo de $RMSE$ y $Mod_{2,1}$ la regresión con el valor más alto.

Tabla 12: Raíz del error cuadrático medio $RMSE$.

Regresión	$RMSE$
Mod_1	18, 8
$Mod_{2,1}$	21, 2
$Mod_{2,2}$	16, 4
$Mod_{2,3}$	20, 2
Mod_3	19, 3
Mod_4	18, 7
Mod_5	18, 7

3.5. Análisis de residuos

No se requiere que la distribución de los residuos, o de los errores, sea estrictamente normal para que los estimadores MCO y RRO sean los mejores estimadores lineales y sin sesgo; sin embargo, esta condición sí es necesaria para que los estadísticos t y F , y los intervalos de confianza⁷, sean válidos. Únicamente cuando la distribución de los residuos es normal las distribuciones de t y F pueden ser conocidas.

Para realizar el análisis de los residuos se debe partir de una definición de los mismos. Hay varias formas de medir los errores en cada regresión; pero, la más común es la de residuos como tal y que corresponde a la diferencia entre la variable dependiente observada y_i y su valor estimado \hat{y}_i , es decir,

$$e_i = y_i - \hat{y}_i \quad , \quad (25)$$

donde e_i es el residuo e $i = 1, \dots, n$ es el índice que identifica a cada observación de la variable dependiente.

Una vez que se han calculado estos residuos, se procede a realizar el análisis de los mismos mediante el uso de las siguientes técnicas:

- Se procede a dibujar los histogramas de los residuos de cada una de las regresiones para observar la posible correspondencia con una distribución de tipo normal. Las figuras 19(a), 19(b), 19(c), 19(d), 19(e), 19(f) y 19(g) muestran los histogramas de cada regresión. En ellas se observa que las distribuciones de los residuos se pueden aproximar, de manera aceptable, a una distribución normal; pero con un cierto nivel de asimetría (*skewness*) y de desplazamiento del valor

⁷Los rangos de variación de los coeficientes estimados, incluyendo el intercepto, son los intervalos de confianza CI (*Confidence Intervals*). Estos se calculan bajo cierta probabilidad y por ejemplo, para todas las regresiones que se muestran aquí, se calculan como: $CI_k = [\hat{\alpha}_k - 1.96SE_k, \hat{\alpha}_k + 1.96SE_k]$ (95% de confianza).

medio. Estos histogramas también se utilizan para identificar posibles *outliers* en las observaciones⁸.

Una forma más objetiva de analizar si existe relación entre la distribución de los residuos de la regresión y una distribución normal es por medio de los valores de los coeficientes de asimetría (*skewness*) y curtosis. Estos coeficientes deben tener los valores de 0 y 3 respectivamente para que la distribución se pueda considerar como normal. La interpretación de estos valores se acompaña con pruebas de normalidad como las de Jarque-Bera [8] o Kolmogorov-Smirnov [9]; en este caso se hace una prueba de normalidad de Jarque-Bera.

Para cada regresión (primer columna), los valores del coeficiente de asimetría (segunda columna), del coeficiente de curtosis (tercer columna), el indicador de aceptación o rechazo de la hipótesis nula (al 5%) de que los residuos provienen de una distribución normal (cuarta columna) y el valor de la probabilidad (quinta columna) se muestran en la tabla 3.5. Los valores que se leen en esta tabla indican que los residuos que mayor probabilidad tienen de proceder de una distribución normal son los de las regresiones $Mod_{2,2}$ y Mod_4 ; a pesar de que todas las regresiones reportan coeficientes de asimetría y curtosis cercanos a los ideales, solo estas dos regresiones pasan la prueba de Jarque-Bera, es decir, los valores de h son cero (se acepta la hipótesis) y la probabilidad p es superior al 5%.

Tabla 13: Prueba de normalidad de los residuos.

Regresión	Asimetría	Curtosis	h	p
Mod_1	-0,3113	3,5235	1	0,0010
$Mod_{2,1}$	-0,3410	2,6202	1	0,0058
$Mod_{2,2}$	0,1730	3,0362	0	0,1886
$Mod_{2,3}$	-0,1555	2,6755	1	0,0108
Mod_3	-0,3000	3,2575	1	0,0010
Mod_4	-0,1247	2,7910	0	0,0737
Mod_5	-0,3303	3,6395	1	0,0010

El histograma de los residuos, los coeficientes de asimetría y curtosis y la prueba de la normalidad se complementan con la gráfica de la relación entre la distribución de los residuos y una distribución normal. Las figuras 20(a), 20(b), 20(c), 20(d), 20(e), 20(f) y 20(g) muestran las gráficas de la probabilidad de los residuos de cada regresión (los residuos que pueden proceder de distribuciones normales deben situarse sobre la línea punteada). En ellas se observa que la mayoría de los residuos de cada regresión se ubican sobre la línea azul; pero las regresiones $Mod_{2,2}$ y Mod_4 , como ya se ha identificado, son las que mejor hacen la correspondencia

⁸Las regresiones que aquí se presentan descartan residuos con valores superiores a $|40|$ kWh/mes. Porcentajes más altos de explicación de la varianza del modelo se pueden encontrar cuando se disminuye este umbral; sin embargo, se debe evitar que la forma de la distribución de los residuos se distorsione o difiera significativamente de la normal.

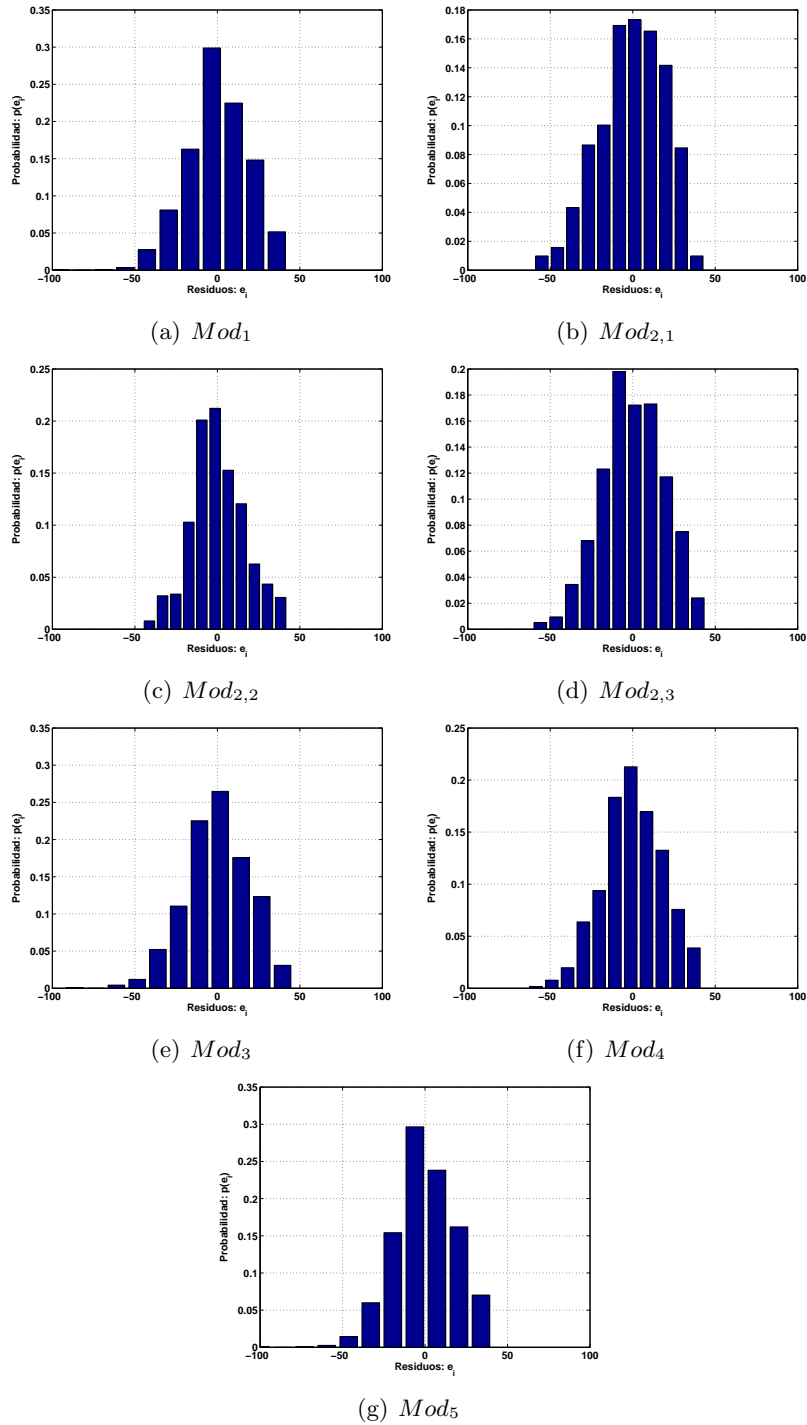


Figura 19: Histogramas normalizados de los residuos de cada regresión. 19(a) Regresión Mod_1 . 19(b) Regresión $Mod_{2,1}$. 19(c) Regresión $Mod_{2,2}$. 19(d) Regresión $Mod_{2,3}$. 19(e) Regresión Mod_3 . 19(f) Regresión Mod_4 . 19(g) Regresión Mod_5 .

con la distribución normal. Posibles outliers también se pueden descartar usando estas gráficas. La asimetría de la distribución también se observa en los residuos que se alejan a la derecha y a la izquierda de la línea recta.

- Para cada regresión también se indaga sobre la correlación entre los residuos. La gráfica de los residuos e_i vs. los residuos *lagged* e_{i-1} y los residuos e_i vs. los valores estimados \hat{y}_i , con $i = 1, \dots, n$, resultan útiles para observar tendencias o correlaciones entre los residuos y posible heteroscedasticidad en las regresiones. Las figuras 21(a), 21(b), 21(c), 21(d), 21(e), 21(f) y 21(g) muestran las gráficas de los residuos e_i vs. los residuos *lagged* e_{i-1} para cada regresión. En estas gráficas no se notan tendencias o comportamientos que sugieran correlación entre los residuos. Las figuras 22(a), 22(b), 22(c), 22(d), 22(e), 22(f) y 22(g) muestran las gráficas de los residuos e_i vs. los valores estimados \hat{y}_i . En estas gráficas no se notan tendencias o comportamientos que sugieran correlación entre los residuos; pero, para valores de la variable estimada (demanda de energía eléctrica) se nota un aumento en el error del modelo; esto puede sugerir una posible heteroscedaticidad, especialmente en las regresiones Mod_3 y Mod_5 .

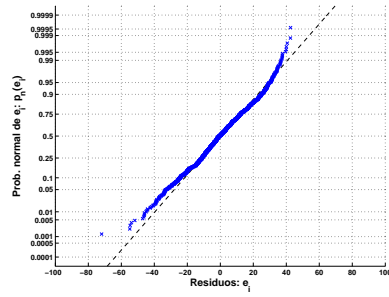
De manera más objetiva, se plantea la prueba de Durbin-Watson [10] para chequear si hay correlación entre los residuos de cada regresión. Esta prueba calcula la estadística DW y la probabilidad p de la hipótesis nula de que los residuos de la regresión lineal no están correlacionados. Valores de p por debajo de 0,05 rechazan la hipótesis. Su definición formal es:

$$DW = \frac{\sum_{i=1}^{n-1} (e_{i+1} - e_i)^2}{\sum_{i=1}^n e_i^2} \quad (26)$$

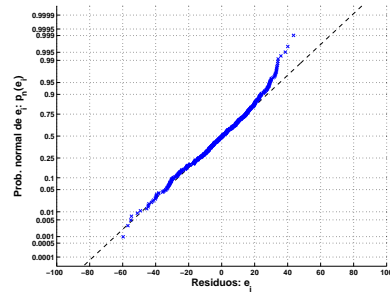
donde e_i es el i -ésimo residuo y n es el número de observaciones. El resultado de este test se utiliza para analizar la hipótesis nula de que los residuos de la regresión lineal están no correlacionados. Los valores del estadístico DW (segunda columna) y su probabilidad p (tercer columna), para cada regresión (primera columna), se muestran en la tabla 3.5. Estos valores aceptan la hipótesis nula para todas las regresiones y refuerzan las anotaciones sobre la no correlación de los residuos que se realizaron a partir de las gráficas anteriores. En orden, las regresiones que presentan los mejores valores de esta prueba son: $Mod_{2,1}$, $Mod_{2,2}$, Mod_4 , Mod_3 , Mod_5 , $Mod_{2,3}$ y $Mod_{2,1}$. Esta última regresión tiene un valor p muy bajo y no se podría decir que la hipótesis es totalmente aceptada.

Tabla 14: Prueba de Durbin-Watson.

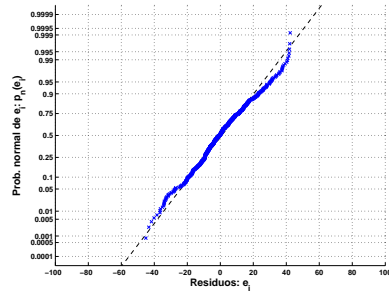
Regresión	DW	p
Mod_1	2,0311	0,8170
$Mod_{2,1}$	2,1731	0,0816
$Mod_{2,2}$	1,9983	0,7426
$Mod_{2,3}$	1,9440	0,2309
Mod_3	1,9635	0,3415
Mod_4	1,9771	0,4468
Mod_5	1,9517	0,2670



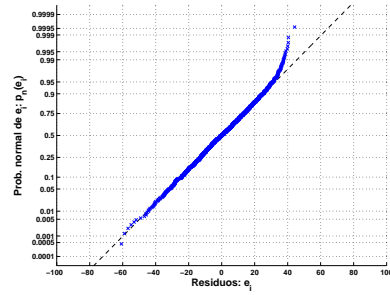
(a) Mod_1



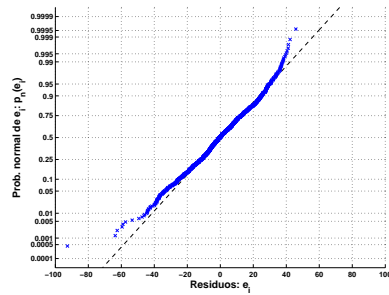
(b) $Mod_{2,1}$



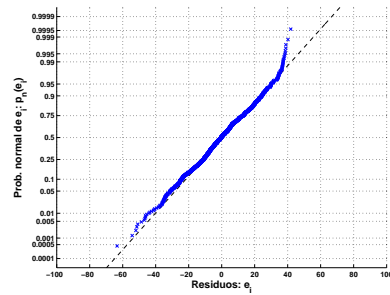
(c) $Mod_{2,2}$



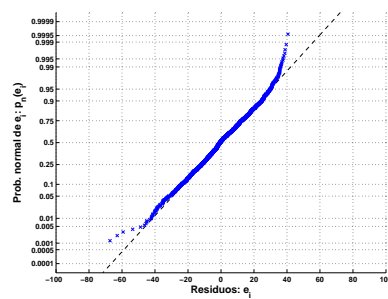
(d) $Mod_{2,3}$



(e) Mod_3



(f) Mod_4



(g) Mod_5

Figura 20: Probabilidad de los residuos de cada regresión. 20(a) Regresión Mod_1 . 20(b) Regresión $Mod_{2,1}$. 20(c) Regresión $Mod_{2,2}$. 20(d) Regresión $Mod_{2,3}$. 20(e) Regresión Mod_3 . 20(f) Regresión Mod_4 . 20(g) Regresión Mod_5 .

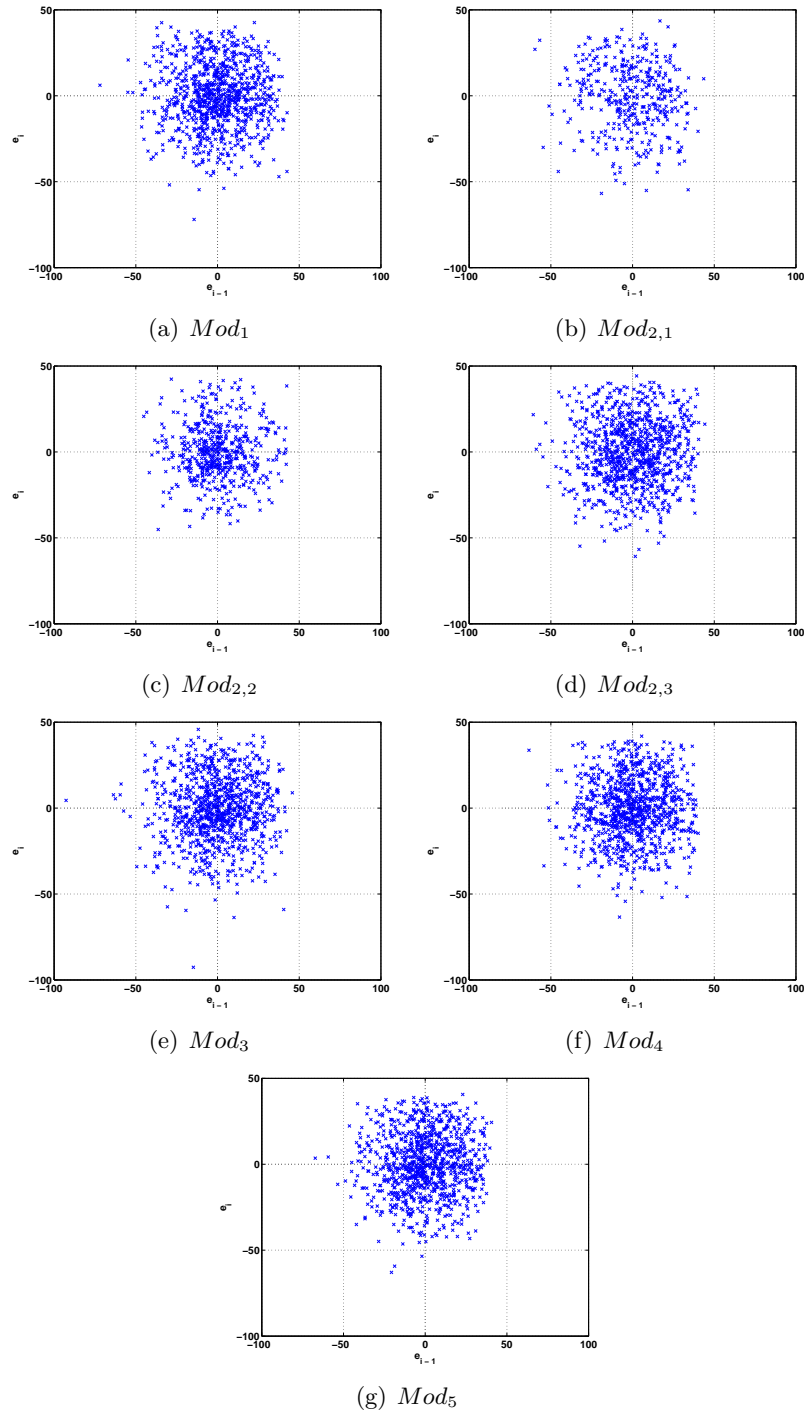


Figura 21: Residuos e_i vs. Residuos *lagged* e_{i-1} . 21(a) Regresión Mod_1 . 21(b) Regresión $Mod_{2,1}$. 21(c) Regresión $Mod_{2,2}$. 21(d) Regresión $Mod_{2,3}$. 21(e) Regresión Mod_3 . 21(f) Regresión Mod_4 . 21(g) Regresión Mod_5 .

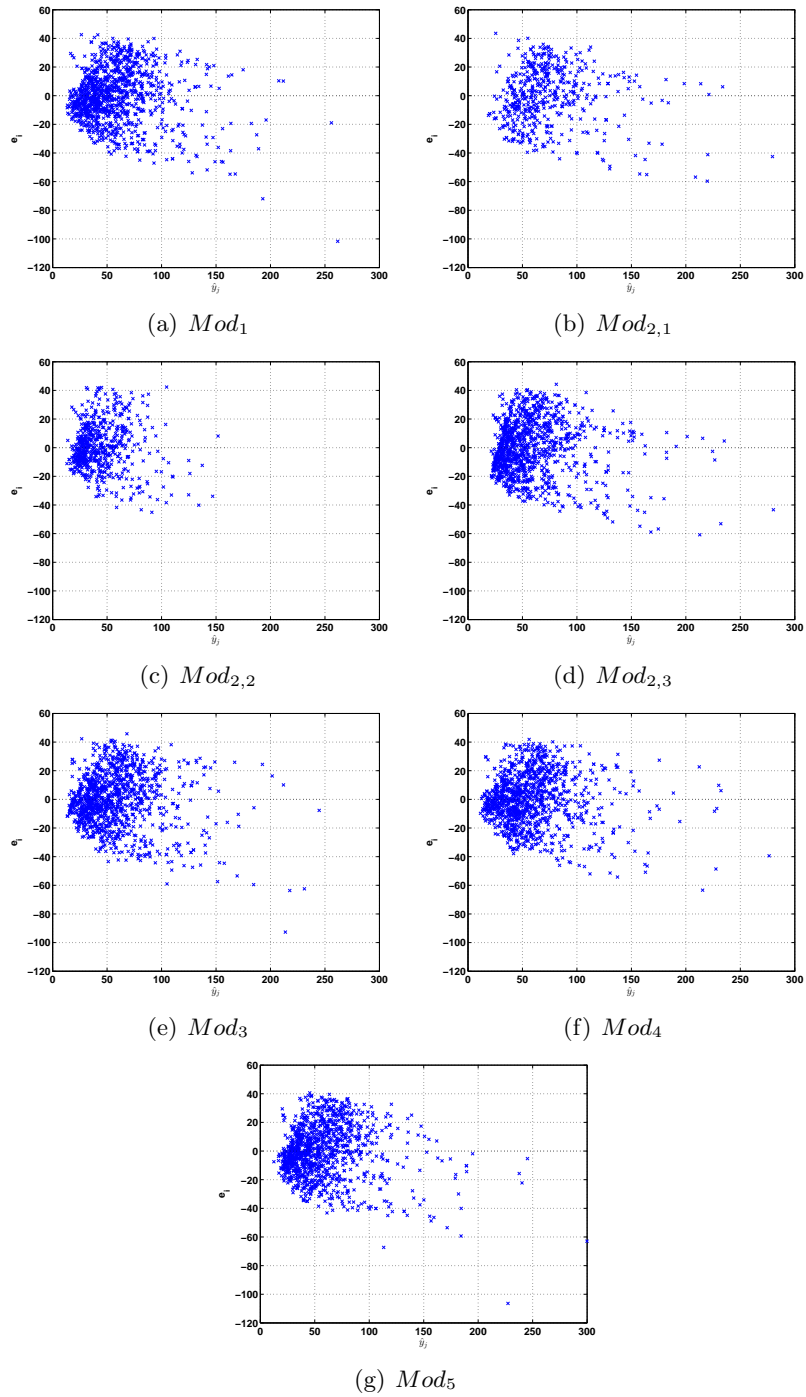


Figura 22: Residuos e_i vs. Valores estimados \hat{y}_i . 22(a) Regresión Mod_1 . 22(b) Regresión $Mod_{2,1}$. 22(c) Regresión $Mod_{2,2}$. 22(d) Regresión $Mod_{2,3}$. 22(e) Regresión Mod_3 . 22(f) Regresión Mod_4 . 22(g) Regresión Mod_5 .

3.6. Distancia de Cook D_i

La distancia de Cook, D_i con $i = 1, \dots, n$, es el cambio normalizado de los coeficientes estimados, vistos como un vector, como consecuencia de la eliminación de la i -ésima observación. Esta distancia es útil para identificar *outliers* en las regresiones. Por definición, se considera una observación como *outlier* si su distancia de Cook es más grande que tres veces la media de las distancias de Cook ($3\bar{D}_i$). D_i se define como:

$$D_i = \frac{\sum_{i=1}^n (\hat{y}_i - \hat{y}_{i \neq r})^2}{(c)(MSE)} \quad (27)$$

donde \hat{y}_i es el i -ésimo valor estimado; $\hat{y}_{i \neq r}$ es el i -ésimo valor ajustado sin incluir la observación r ; c es el número de coeficientes estimados, incluido el intercepto; MSE es el error cuadrático medio y n es el número de observaciones.

Las figuras 23(a), 23(b), 23(c), 23(d), 23(e), 23(f) y 23(g) muestran las gráficas de la distancia de Cook para cada una de las regresiones. La línea de color rojo que se dibuja en cada una de ellas es el límite que se utiliza para clasificar las observaciones como *outliers*. El valor del criterio para descartar observaciones $3\hat{D}_i$ (segunda columna), el número de *outliers* (tercer columna) y su porcentaje con respecto al número total de observaciones (cuarta columna), para cada regresión (primera columna), se muestran en la tabla 15. De acuerdo a los valores que se muestran en esta tabla 15, y como se puede apreciar en cada figura, aún hay *outliers* en cada regresión que se podrían excluir en futuras regresiones. En orden, las regresiones que presentan mayor número de *outliers*, considerando el valor del porcentaje con respecto al número de observaciones efectivas utilizadas en cada regresión, son: $Mod_{2,2}$, $Mod_{2,3}$, Mod_3 , Mod_4 , Mod_1 , Mod_5 y $Mod_{2,1}$.

Tabla 15: Número de *Outliers*.

Regresión	$3\bar{D}_i$	<i>Outliers</i>	% de n
Mod_1	0.0015	47	3.43
$Mod_{2,1}$	0.0038	14	2.23
$Mod_{2,2}$	0.0021	45	6.51
$Mod_{2,3}$	0.0011	70	5.10
Mod_3	0.0014	54	3.95
Mod_4	0.0014	52	3.86
Mod_5	0.0018	34	2.49

3.7. Colinealidad

Por último, aunque este instrumento es fundamental para realizar la depuración de las regresiones, se hace un análisis de colinealidad. La colinealidad se refieren a la posible inclusión de variables explicativas, que no son ortogonales o independientes entre sí, en la regresión. Estas variables se deben descartar de la regresión porque no son relevantes

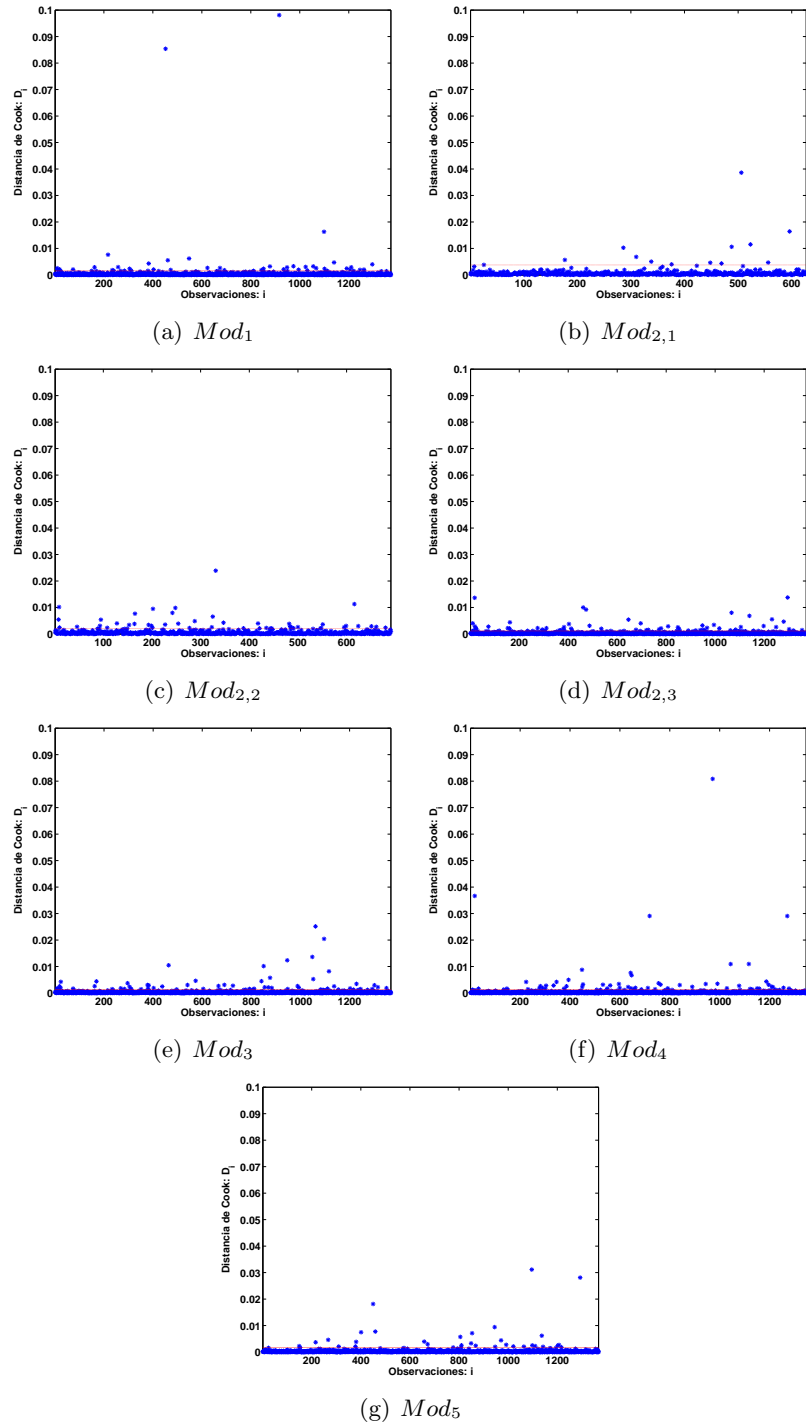


Figura 23: Distancia de Cook D_i . 23(a) Regresión Mod_1 . 23(b) Regresión $Mod_{2,1}$. 23(c) Regresión $Mod_{2,2}$. 23(d) Regresión $Mod_{2,3}$. 23(e) Regresión Mod_3 . 23(f) Regresión Mod_4 . 23(g) Regresión Mod_5 .

para la modelación, es decir, se pueden explicar con las otras variables, y porque reducen la precisión del modelo.

Síntomas característicos de colinealidad son: (i) Valores altos de los factores de determinación, R^2 y R_a^2 , y al mismo tiempo coeficientes no significativos asociados a variables que sí pueden ser relevantes en la regresión; (ii) sensibilidad extrema de los coeficientes cuando variables explicativas se incluyen o remueven de la regresión y (iii) coeficientes con signos y valores contrarios a los esperados. En cada regresión que se reporta aquí, aunque se descartaron las variables que presentaban tendencias no claras (coeficientes de signos no esperados), no se identificaron posibles síntomas de colinealidad entre las variables explicativas utilizadas para la modelación de la demanda de energía eléctrica de las viviendas rurales de Nariño.

De manera analítica, la colinealidad se estima, en este caso, a través del factor de inflación de la varianza VIF (*Variance Inflation Factor*), que mide el grado de inflación de la varianza de cada coeficiente estimado dado que su respectiva variable no es ortogonal o independiente de las otras variables del modelo. La ecuación que describe este factor es:

$$VIF_k = \frac{1}{1 - R_k^2} \quad (28)$$

donde R_k^2 es el coeficiente de determinación de la regresión entre la variable explicativa k y las restantes; k es el número de variables en cada regresión. Los valores del factor VIF_k también corresponden a la diagonal de la inversa de la matriz de correlación R_0 de las variables explicativas del modelo [11]; esta propiedad simplifica un poco el cálculo del factor a través de la matriz R_0 . La matriz de correlación, R_0 , (segunda columna) de las variables explicativas (tercer columna), para cada regresión (primera columna), se muestra en la tabla 3.7. Los valores que se leen en esta tabla no sugieren una fuerte relación o correlación entre las variables que se incluyen en cada regresión.

Tabla 16: Matriz de correlación R_0 .

Regresión	R_0	Variable
Mod_1	1, 0000	Incandescente_Uso
	0, 1023	Ahorradora_Uso
	0, 0362	Refrigeracion_Tam
	0, 2214	Vivienda_Persona
	0, 1982	Vivienda_Cuarto
	0, 1802	Vivienda_Ingreso
	0, 2525	EE_Factura_Valor
	0, 2325	EE_Factura_Valor
Mod_{2_1}	1, 0000	Refrigeracion_Tam
	0, 1348	Vivienda_Persona
	0, 0715	Vivienda_Cuarto
	0, 0911	EE_Factura_Valor
Mod_{2_2}	1, 0000	Incandescente_Uso
	0, 0660	Ahorradora_Uso
	0, 0465	Refrigeracion_Tam
	0, 1130	Vivienda_Persona
	0, 2135	Vivienda_Cuarto
	0, 2554	EE_Factura_Valor

Regresión	R_0	Variable
$Mod_{2,3}$	$\begin{bmatrix} 1,0000 & 0,1802 \\ 0,1802 & 1,0000 \end{bmatrix}$	Ahorradora_Uso EE_Factura_Valor
Mod_3	$\begin{bmatrix} 1,0000 & 0,1007 & 0,1941 & 0,1779 \\ 0,1007 & 1,0000 & 0,2284 & 0,2031 \\ 0,1941 & 0,2284 & 1,0000 & 0,2529 \\ 0,1779 & 0,2031 & 0,2529 & 1,0000 \end{bmatrix}$	Ahorradora_Uso Refrigeracion_Tam Vivienda_Ingreso_COP EE_Factura_Valor
Mod_4	$\begin{bmatrix} 1,0000 & 0,1819 & 0,1946 \\ 0,1819 & 1,0000 & 0,2025 \\ 0,1946 & 0,2025 & 1,0000 \end{bmatrix}$	Refrigeracion_Tam Vivienda_Cuarto EE_Factura_Valor
Mod_5	$\begin{bmatrix} 1,0000 & 0,1976 & 0,2042 \\ 0,1976 & 1,0000 & 0,2281 \\ 0,2042 & 0,2281 & 1,0000 \end{bmatrix}$	Refrigeracion_Tam Vivienda_Cuarto EE_Factura_Valor

El valor del factor de inflación, VIF_k , (segunda columna) de las variables explicativas (tercera columna), para cada regresión (primera columna), se muestra en la tabla 3.7. La regla que sobre el valor VIF_k es que si cada uno de ellos, o en promedio, es mayor a 10, entonces existe la posibilidad de correlación entre las variables explicativas. De acuerdo con los valores que se registran en esta tabla, ninguna regresión reporta un factor de inflación, individual y en promedio, superior a 10; únicamente la regresión Mod_1 podría suponer una débil correlación entre sus variables, y puntualmente entre la variable de ingreso por vivienda, Ingreso_Vivienda, y el valor pagado por el servicio, EE_Factura_Valor. Este factor no es definitivo en el análisis de correlación entre las variables explicativas, se sugiere, en futuras regresiones aplicar otros métodos como el de Belsley [1],

Tabla 17: Factor de inflación de la varianza VIF_k ,

Regresión	VIF_k	Variable
Mod_1	1,1075	Incandescente_Uso
	1,1027	Ahorradora_Uso
	1,0325	Refrigeracion_Tam
	1,2049	Vivienda_Persona
	1,2145	Vivienda_Cuarto
	1,8501	Vivienda_Ingreso
	1,9837	EE_Factura_Valor
$Mod_{2,1}$	1,0226	Refrigeracion_Tam
	1,0701	Vivienda_Persona
	1,8015	Vivienda_Cuarto
	1,8391	EE_Factura_Valor
$Mod_{2,2}$	1,0484	Incandescente_Uso
	1,0773	Ahorradora_Uso
	1,0567	Refrigeracion_Tam
	1,6069	Vivienda_Persona
	1,6905	EE_Factura_Valor
$Mod_{2,3}$	1,0336	Ahorradora_Uso

Regresión	VIF_k	Variable
	1,0336	EE_Factura_Valor
Mod_3	1,0603	Ahorradora_Uso
	1,0824	Refrigeracion_Tam
	1,1315	Vivienda_Ingreso_COP
	1,1127	EE_Factura_Valor
Mod_4	1,0628	Refrigeracion_Tam
	1,0662	Vivienda_Cuarto
	1,0715	EE_Factura_Valor
Mod_5	1,0704	Refrigeracion_Tam
	1,0821	Vivienda_Cuarto
	1,0850	EE_Factura_Valor

3.8. Selección de modelos

Identificar cuál de todas las regresiones ya presentadas es el mejor modelo de demanda de energía eléctrica para las viviendas rurales del Departamento de Nariño no es una decisión simple. Catalogar las mejores regresiones a partir únicamente de los criterios de explicación, R^2 y R_a^2 , de la raíz del error cuadrático medio $RMSE$, o del análisis de residuos, no es conveniente, porque las regresiones desarrolladas no son anidadas y se diferencian entre ellas por las variables explicativas que incluyen y por el tipo de regresión: categórica o no categórica.

Para seleccionar las mejores regresiones no anidadas se utiliza el criterio de información. Dos tipos de criterios de información son de común uso: (i) El criterio de información de Akaike AIC (Akaike Information Criterion) y el criterio de información bayesiano BIC (Bayesian Information Criterion) [7]. El primero se basa en la entropía de la información y el segundo en la teoría de la decisión. Ambos criterios utilizan una función de penalización para la complejidad del modelo que permite observar la bondad del ajuste de las observaciones que se utilizan en la regresión de cada uno de ellos. Se prefiere el modelo que reporte los valores más bajos para estos dos criterios. Los valores de los criterios AIC (segunda columna) y BIC (tercera columna), para cada una de las regresiones, se muestran en la tabla 18.

De acuerdo a los valores que se reportan en esta tabla, y teniendo en cuenta los criterios ya analizados para cada regresión (y especialmente el análisis de los residuos), se llega a la conclusión de que son tres los tipos que se pueden utilizar para representar la demanda de la energía eléctrica en las viviendas rurales del Departamento de Nariño. El modelo tipo 1 (T1), de mayor calidad de ajuste a la regresión y de nivel aceptable de criterio de información, corresponde al modelo que se obtiene a partir de la regresión Mod_4 , es decir, la regresión categorizada por ingresos por vivienda. El modelo tipo 2 (T2), de calidad aceptable de ajuste a la regresión y de mayor nivel de criterio de información, corresponde a las regresiones $Mod_{2,1}$ y $Mod_{2,2}$, es decir, las regresiones filtradas para los grupos de viviendas que consumen gas y leña respectivamente. El modelo tipo 3 (T3), de menor calidad de ajuste a la regresión y menor criterio de información, es la regresión Mod_1 . Este último modelo es relevante porque incluye la

mayor cantidad de variables explicativas y se puede establecer como un modelo base para la comparación con los otros modelos.

Tabla 18: Valores del criterio de información.

Regresión	$AIC (\times 10^3)$	$BIC (\times 10^3)$
Mod_1	10,092	10,132
$Mod_{2,1}$	4,489	4,510
$Mod_{2,2}$	5,259	5,285
$Mod_{2,3}$	10,135	10,165
Mod_3	10,194	10,234
Mod_4	10,043	10,089
Mod_5	10,122	10,158

4. Predicción

Se realiza la predicción de la demanda de energía eléctrica para las viviendas rurales del Departamento de Nariño con los tres tipos de modelos seleccionados.

4.0.1. Modelo Tipo 1 (T1)

El modelo T1 es un modelo categórico que describe la demanda de la energía eléctrica teniendo en cuenta el nivel de ingreso de cada vivienda. El modelo se divide en tres clases: (i) La clase *A* de viviendas que tienen ingresos en el rango $[0, 350.000]$ COP, (ii) la clase *B* de viviendas que tienen ingresos en el rango $(350.000, 700.000]$ COP y (iii) la clase *C* de viviendas que tienen ingresos mayores a 700.000 COP. Las siguientes ecuaciones definen el modelo:

$$\hat{y}_A = 4,88 + 1,38x_1 + 1,92x_2 + 0,003x_3 \quad (29a)$$

$$\hat{y}_B = 23,54 + 0,23x_1 + 1,92x_2 + 0,002x_3 \quad (29b)$$

$$\hat{y}_C = 18,65 + 1,38x_1 + 1,92x_2 + 0,003x_3 \quad , \quad (29c)$$

donde \hat{y}_A es la demanda de la energía eléctrica mensual para las viviendas de la clase *A* (kWh/mes), \hat{y}_B es la demanda de la energía eléctrica mensual para las viviendas de la clase *B* (kWh/mes), \hat{y}_C es la demanda de la energía eléctrica mensual para las viviendas de la clase *C* (kWh/mes), x_1 es el tamaño de equipos de refrigeración (ft^3), x_2 es el número de cuartos que tiene cada vivienda (cuartos/vivienda) y x_3 es el valor facturado del servicio de electricidad (COP/mes).

De las ecuaciones 29a, 29b y 29c se deduce que:

- El intercepto del modelo (kWh/mes), relacionado principalmente con una demanda de subsistencia mínima, es mayor para clase *B*, le sigue la clase *C* y

por último la clase A ; el aporte a la demanda por la capacidad de equipos de refrigeración (kWh/mes/ft³) es igual para las clases A y C y menor para la clase B ; el aporte a la demanda por el número de cuartos que tiene la vivienda (kWh/mes/cuartos/vivienda) es igual para todas las clases y el aporte a la demanda por el valor pagado por el servicio (kWh/COP) es igual para las clases B y C y menor para la clase A .

- Sin incluir el intercepto, la demanda para las clases A y C se explica de igual forma por cada variable independiente; en cambio, la demanda para la clase B se diferencia de las anteriores en el aporte por la capacidad de los equipos de refrigeración y el valor pagado por el servicio (este aporte es menor que el de las otras clases). El comportamiento de estas dos variables es el que modela la demanda de la energía eléctrica de las viviendas rurales del Departamento de Nariño y los interceptos son los que logran distinguir esta demanda por las categorías ya definidas.
- Los rangos de las variables dependiente y explicativas que se utilizaron en la regresión, después de eliminar *outliers*, fueron de: $y \in [1, 240]$ kWh/mes, $x_1 \in [0, 150]$ ft³, $x_2 \in [1, 18]$ cuartos/vivienda y $x_3 \in [170, 87.650]$ COP. Los valores medios de cada variable explicativa (primera columna) para la clase A (segunda columna), para la clase B (tercera columna) y para la clase C (cuarta columna) se muestran en la tabla 19. Los valores en esta tabla están acordes con lo esperado para cada clase.

Tabla 19: Valor medio de las variables explicativas del modelo T1.

Variable	Clase A	Clase B	Clase C
Refrigeracion_Tam (ft ³)	8,8414	8,8073	9,5412
Vivienda_Cuarto (cuartos/vivienda)	5,7556	6,2294	7,6471
EE_Factura_Valor (COP/mes)	16.686	20.513	23.747

- Con los valores medios de cada variable explicativa, que se muestran en la tabla 19, se hace la predicción de la demanda de energía para cada vivienda del Departamento de Nariño. El valor de la demanda, \hat{y} (tercer columna), y los valores límites con el 95% de confianza, \hat{y}^- para el límite inferior de demanda (segunda columna) y \hat{y}^+ para el límite superior de demanda (cuarta columna), para cada clase de vivienda (primera columna), se muestran en la tabla 20. Según los valores registrados en esta tabla, la demanda de una vivienda de clase A es de 78,5164 kWh/mes, la demanda de una vivienda de clase B es de 70,5676 kWh/mes y la demanda de una vivienda de clase C es de 108,1462 kWh/mes; los valores de estas demandas están acordes con la categorización de las viviendas por ingreso, es decir, las viviendas que tienen mayores ingresos tienen una demanda más alta de energía eléctrica que las viviendas que tienen menor ingreso.

Tabla 20: Predicción modelo T1 con los valores de la tabla 19.

Variable	\hat{y}^- kWh/mes	y kWh/mes	y^+ kWh/mes
Clase A	76,4068	78,5164	80,6260
Clase B	68,4031	70,5676	72,7321
Clase C	104,2073	108,1462	112,0851

4.0.2. Modelo Tipo 2 (T2)

El modelo T2 describe la demanda de la energía eléctrica para las dos clases de grupos de vivienda rurales identificadas como : (i) La clase *A* de viviendas que consumen energía eléctrica y utilizan gas y la clase *B* de viviendas que consumen energía eléctrica y utilizan leña. Las siguientes ecuaciones definen el modelo:

$$\hat{y}_A = 11,46 + 0,63x_3 + 1,32x_4 + 0,003x_5 + 1,15x_6 \quad (30a)$$

$$\hat{y}_B = 11,08 + 0,03x_1 + 0,07x_2 + 1,88x_3 + 1,58x_4 + 0,002x_5 \quad , \quad (30b)$$

donde \hat{y}_A es la demanda de la energía eléctrica mensual para las viviendas de la clase *A* (kWh/mes) y \hat{y}_B es la demanda de la energía eléctrica mensual para las viviendas de la clase *B* (kWh/mes), x_1 es la frecuencia de uso de lámparas incandescentes (horas/mes), x_2 es la frecuencia de uso de lámparas ahorradoras (horas/mes), x_3 es el tamaño de equipos de refrigeración (ft³), x_4 es el número de personas que habitan la vivienda (personas/vivienda), x_5 es el valor facturado del servicio de electricidad (COP/mes) y x_6 es el número de cuartos por vivienda (cuartos/vivienda).

De las ecuaciones 30a y 30b se deduce que:

- El intercepto del modelo (kWh/mes), relacionado principalmente con una demanda de subsistencia mínima, es prácticamente igual para las dos clases; el aporte a la demanda por lámparas incandescentes y ahorradoras (kW) es significativo solo para la clase *B*; el aporte a la demanda por la capacidad de equipos de refrigeración (kWh/mes/ft³) es mayor para la clase *B*; el aporte a la demanda por el número de personas que habitan la vivienda (kWh/mes/personas/vivienda) es ligeramente mayor para la clase *B*, el aporte a la demanda por el valor pagado por el servicio (kWh/COP) es ligeramente mayor para la clase *A* y el aporte a la demanda por el número de cuartos que tiene la vivienda (kWh/mes/cuartos/vivienda) es solo significativo para la clase *A*.
- Los dos modelos son prácticamente independientes y la diferencia principal en la modelación de la energía eléctrica está dada por las variables utilizadas para el proceso de iluminación. Los valores medios de cada variable explicativa (primera columna) para la clase *A* (segunda columna) y para la clase *B* (tercera columna) se muestran en la tabla 21.

Tabla 21: Valor medio de las variables explicativas del modelo T2.

Variable	Clase A	Clase B
Incandescentes_Uso (horas/mes)	-	90.2605
Ahorradoras_Uso (horas/mes)	-	16.8017
Refrigeracion_Tam (ft ³)	7,1503	3.1862
Vivienda_Persona (personas/vivienda)	4,1659	4.1129
Vivienda_Cuarto (cuartos/vivienda)	6,2504	-
EE_Factura_Valor (COP/mes)	19.381	11.591

- Con los valores medios de cada variable explicativa, que se muestran en la tabla 21, se hace la predicción de la demanda de energía para cada vivienda del Departamento de Nariño. El valor de la demanda, \hat{y} (tercer columna), y los valores límites con el 95% de confianza, \hat{y}^- para el límite inferior de demanda (segunda columna) y \hat{y}^+ para el límite superior de demanda (cuarta columna), para cada clase de vivienda (primera columna), se muestran en la tabla 22. Según los valores registrados en esta tabla, la demanda de una vivienda de clase A, 83,2658 kWh/mes, es mayor que la demanda de una vivienda de clase B, 48,0349 kWh/mes; es decir, la demanda de energía eléctrica de las viviendas rurales que utilizan gas es mayor que la demanda de las viviendas que utilizan leña.

Tabla 22: Predicción modelo T2 con los valores de la tabla 21.

Variable	\hat{y}^- kWh/mes	y kWh/mes	y^+ kWh/mes
Clase A	81,3480	83,2658	85,1835
Clase B	46,7213	48,0349	49,3484

4.0.3. Modelo Tipo 3 (T3)

El modelo T3 describe la demanda de la energía eléctrica para todas las viviendas rurales del departamento y sin ninguna categorización. La siguiente ecuación describe el modelo:

$$\begin{aligned} \hat{y} = & 4,41 + 0,02x_1 + 0,05x_2 + 0,98x_3 + 1,01x_4 + 1,69x_5 \\ & + 0,00004x_6 + 0,003x_7 \quad , \end{aligned} \quad (31)$$

donde \hat{y} es el consumo estimado para cada vivienda (kWh/mes), x_1 es la frecuencia de uso de lámparas incandescentes (horas/mes), x_2 es la frecuencia de uso de lámparas ahorradoras (horas/mes), x_3 es el tamaño de equipos de refrigeración (ft³), x_4 es el

número de personas que habitan la vivienda (personas/vivienda), x_5 es el número de cuartos por vivienda (cuartos/vivienda), x_6 es el ingreso de cada vivienda (COP/mes) y x_7 es el valor pagado por el servicio de electricidad (COP/mes).

Los valores medios (segunda columna) de cada variable explicativa (primera columna) se muestran en la tabla 23.

Tabla 23: Valor medio de las variables explicativas del modelo T3.

Variable	\bar{x}_k
Incandescentes_Uso (horas/mes)	83.9869
Ahorradoras_Uso (horas/mes)	30.6341
Refrigeracion_Tam (ft ³)	5.2492
Vivienda_Persona (personas/vivienda)	4.1290
Vivienda_Cuarto (cuartos/vivienda)	5.6006
Vivienda_Ingreso (COP/mes)	332.470
EE_Factura_Valor (COP/mes)	15.524

Con lo valores medios de cada variable explicativa, que se muestran en la tabla 23, se hace la predicción de la demanda de energía para cada vivienda del Departamento de Nariño.

El valor de la demanda, \hat{y} (tercer columna), y los valores límites con el 95% de confianza, \hat{y}^- para el límite inferior de demanda (segunda columna) y \hat{y}^+ para el límite superior de demanda (cuarta columna) son , para cada clase de vivienda (primera columna), se muestran en la tabla 24. Es valor de la demanda que muestra esta tabla es inferior al reportado por el modelo T1 y es aproximadamente el promedio del modelo T2.

Tabla 24: Predicción modelo T3 con los valores de la tabla 23.

Modelo	\hat{y}^- kWh/mes	y kWh/mes	\hat{y}^+ kWh/mes
T3	64,4894	65,6133	66,7372

5. Conclusiones y recomendaciones

En el proceso de construcción del modelo de la demanda de la energía eléctrica de las viviendas rurales de Nariño, por medio de las variables reportadas por la encuesta de energía realizada por el PERS-Nariño, se notó en primer lugar que el número de observaciones confiables, de las viviendas que consumen energía eléctrica, corresponde únicamente al 56% del número total de viviendas encuestadas. Estas observaciones corresponden a la información leída directamente de la factura del servicio de energía

eléctrica de cada vivienda, por lo cual se recomienda que para otras encuestas PERS se haga énfasis en la importancia de la recolección de esta información. Otra forma de adquirir y validar esta información es por medio de cruces de información de consumo con bases de datos de los prestadores del servicio, por lo que se debe incluir en la encuesta el nombre del prestador y validar plenamente el número de identificación de la persona a quien se factura el servicio.

En general, todas las variables dependiente y explicativas, presentan observaciones fuera de las tendencias que se deberían esperar (*outliers*). En algunas regresiones estas observaciones tienen un porcentaje alto del total de observaciones disponibles y afectan los resultados para la modelación. Se recomienda para otras encuestas PERS validar la información de cada campo cuando esta se entra directamente al dispositivo de recolección de datos (validar también con la unidad de medida que se hace).

La dispersión y la correlación que presentan las variables explicativas con respecto al consumo de energía eléctrica no son las mejores. Entre el conjunto de posibles variables que se incluyen en el modelo, la variable de valor facturado es la que más correlación presenta (es lógico), le siguen la capacidad de los refrigeradores y el número de cuartos por vivienda. En pruebas realizadas sobre cada regresión, la no inclusión de la variable de valor pagado reduce significativamente la explicación de la demanda de energía eléctrica por medio de las otras variables; sin esta variable el porcentaje de explicación máximo de la varianza del modelo que se obtiene es del 50%. De otra parte, las variables que se espera contribuyan a la explicación de la demanda, como es el consumo por aparatos eléctricos y electrónicos, no logran ser significativas en las regresiones expuestas aquí. Se recomienda a futuro depurar las regresiones a otras categorías donde este tipo de variable si logre ser significativa.

El modelo de demanda de energía eléctrica para las viviendas rurales de Nariño no se debe hacer en una categoría general, es decir, resulta muy difícil calcular un solo modelo que pueda explicar la dispersión alta que reporta la encuesta. Por esta razón, el modelo que aquí se presenta es un conjunto de modelos categorizados por los comportamientos de esta demanda. Entre todos los experimentos que se realizaron, fueron los modelos aquí reportados los que presentaron los mejores resultados en la regresión y una correspondencia lógica y esperada con la demanda. De los tres tipos de modelo de demanda seleccionados, el modelo T1 y el modelo T2 presentan resultados muy interesantes con respecto al nivel de ingreso de las viviendas y al tipo de combustible, diferente a la energía eléctrica, utilizado en el proceso de cocción. Además de los valores de demanda, también son importantes las medidas obtenidas del aporte de cada variable a la demanda (coeficientes o elasticidades).

Las categorías de modelación que se presentan aquí resultaron ser las de mayor impacto, hasta ahora, en la explicación de la demanda; aunque la modelación que se realiza es netamente lineal, la explicación con modelos no lineales debería arrojar mejores resultados; sin embargo, la prueba de modelos no lineales en las categorías definidas no arrojó ningún resultado significativo. Se recomienda a futuro explorar otros tipos de categorías en modelos lineales y no lineales y comparar con los resultados ya obtenidos.

Referencias

- [1] DuMouchel, W. H. y F. L. O'Brien. "Integrating a Robust Option into a Multiple Regression Computing Environment." *ComputerScience and Statistics: Proceedings of the 21st Symposium on the Interface*. Alexandria, VA: AmericanStatistical Association, 1989.
- [2] Holland, P. W. y R. E. Welsch. "RobustRegression Using Iteratively Reweighted Least-Squares." *Communications in Statistics: Theory and Methods*, A6, 1977, pp. 813–827.
- [3] Huber, P. J. *Robust Statistics*. Hoboken, NJ: John Wiley & Sons, Inc., 1981.
- [4] Agostini, C. A., Plottier M. C. y Saavedra, E. H. "La Demanda Residencial por Energía Eléctrica en Chile." *The Selected Works of Claudio A. Agostini*. 2009. Disponible en: http://works.bepress.com/claudio_agostini/35
- [5] Fernández, L. "Análisis Microeconómico de la Demanda Eléctrica Residencial de Corto Plazo España." Universidad de Barcelona, 2006. Disponible en: http://www.tdx.cat/bitstream/handle/10803/2580/02.LFV_2de3.pdf
- [6] Wooldridge, J. *Introductory Econometrics: A Modern Approach*. 4e. Mason, OH: South-Western CENGAGE Learning. 2008
- [7] Box, G. E. P., G. M. Jenkins y G.C. Reinsel. *Time Series Analysis: Forecasting and Control*. 3e. Englewood Cliffs, NJ: Prentice-Hall, 1994.
- [8] Jarque, C. M. y A. K. Bera. "A test for normality of observations and regression residuals." *InternationalStatistical Review*. Vol. 55, No. 2, 1987, pp. 163–172.
- [9] Massey, F. J. "The Kolmogorov-Smirnov Test for Goodness of Fit." *Journal of the American Statistical Association*. Vol. 46, No. 253, 1951, pp. 68–78.
- [10] Atkinson, A. T. *Plots, Transformations, and Regression. An Introduction to Graphical Methods of Diagnostic Regression Analysis*. New York: Oxford Statistical Science Series, Oxford University Press, 1987.
- [11] Belsley, D. A., E. Kuh y R. E. Welsch. *Regression Diagnostics*. Hoboken, NJ: John Wiley & Sons, 1980.