



Evaluación de Impacto del Gas Natural sobre Consumidores Residenciales

Documento de Trabajo N° 27

Luis Bendezú Medina

Oficina de Estudios Económicos

Lima, Noviembre 2009

OSINERGMIN

Evaluación de Impacto del Gas Natural sobre Consumidores Residenciales

Documento de Trabajo N° 27, preparado por la Oficina de Estudios Económicos (OEE).

Está permitida la reproducción total o parcial de este documento por cualquier medio, siempre y cuando se cite la fuente.

Elaborado por: Luis Bendezú Medina.

Asistentes: Manuel Álvarez, Paulo Chahuara.

Primera versión: Julio 2008

Esta versión: Noviembre 2009

Para comentarios o sugerencias dirigirse a:

OSINERGMIN

Bernardo Monteagudo 222, Magdalena del Mar

Lima, Perú

Tel. (511) 219-3400, anexo 1057

Fax (511) 219-3413

http://www.osinerg.gob.pe/newweb/pages/Estudios_Economicos/79.htm

Correo electrónico: lbendezu@osinerg.gob.pe.

Organismo Supervisor de la Inversión en Energía y Minería-OSINERGMIN
Oficina de Estudios Económicos
Documento de Trabajo N° 25

Evaluación de Impacto del Gas Natural sobre Consumidores Residenciales¹

Resumen

En este documento se analiza el impacto sobre el nivel de gasto y consumo de los nuevos consumidores con acceso al servicio de gas natural residencial. Utilizando una encuesta diseñada específicamente con este propósito y técnicas de evaluación de impacto, se encuentra que el ahorro es ligeramente mayor a lo pronosticado por las empresas y el regulador.

¹ Documento elaborado por la Oficina de Estudios Económicos del OSINERGMIN. Comentarios y sugerencias a lbendezu@osienrg.gob.pe. Se agradece la asistencia de Manuel Álvarez y Paulo Chahuara en la preparación y edición del documento.

TABLA DE CONTENIDOS

1. Introducción.....	5
2. Breve Panorama del Gas Natural en Consumidores Residenciales.....	6
2.1. Cobertura del Servicio.....	7
2.2. Evolución del Consumo.....	8
3. Metodología.....	9
3.1. Evaluación del Impacto.....	9
3.2. Propensity Score Matching.....	15
4. Implementación Empírica.....	17
4.1. Información Utilizada.....	17
4.2. Resultados.....	19
4.2.1. Impacto sobre el Gasto en Energía.....	20
4.2.2. Sobre el Consumo de Energía.....	21
4.2.3. Evaluación de los Resultados.....	22
5. Conclusiones.....	24
6. Referencias.....	25
Anexo.....	26

1. Introducción

La evaluación de impacto de políticas en el sector energético es un aspecto que no ha sido muy abordado en la literatura especializada. Los estudios que existen al respecto se refieren principalmente a estudios donde se realizan evaluaciones de medidas de políticas de conservación de energía y han sido realizados principalmente en países desarrollados. Por ejemplo, en el caso peruano se han realizado numerosos programas de fiscalización conducentes a la mejora de los servicios que brindan las empresas del sector energético.

Dentro de este contexto, el arribo del gas natural ha representado una oportunidad importante de ahorro de energía, tanto a nivel de hogares como de empresas. En particular, los hogares que se encuentran dentro de la red de distribución podrían beneficiarse de menores precios si es que optan por conectarse a la red, especialmente en un contexto donde el precio internacional de los combustibles ha venido incrementándose de forma notoria. Las primeras evaluaciones del ahorro potencial que podrían tener los consumidores residenciales fueron hechas asumiendo que el nivel de consumo una vez realizada la conversión se mantendría constante, dando como resultado porcentajes de ahorro cercanos al 50% en comparación con el gasto que se tenía antes de conectarse a la red.

Sin embargo, a la fecha no se ha realizado una evaluación que determine el verdadero impacto sobre el consumo y gasto de los hogares derivado de la disponibilidad de gas natural. En este sentido, el presente trabajo utilizará la información proveniente de la Encuesta Residencial de Consumo y Usos de Energía del año 2007 para medir estos efectos. Para ello, el presente trabajo utilizará los métodos empleados en la medición de programas de ahorro energético a la introducción del gas natural domiciliario.

El documento posee cuatro secciones. En la primera de ellas se hace una breve descripción del segmento residencial de gas natural, analizando las tendencias agregadas del consumo, número de usuarios y precios. La segunda sección presenta la metodología a ser empleada en el trabajo, para posteriormente en la tercera sección describir brevemente la información empleada y los resultados de la evaluación. Finalmente, la cuarta parte del documento presenta las conclusiones.

2. Breve Panorama del Gas Natural en Consumidores Residenciales

El servicio de distribución del gas natural consiste en dar suministro a los usuarios de una determinada región o localidad, a través de las redes de ductos o tuberías subterráneas. En el caso peruano, Lima Metropolitana concentra la mayor cantidad de usuarios de este servicio, cuyo gas proviene exclusivamente del yacimiento de Camisea. La red de distribución comprende 2 tipos de redes: la Red Principal (o Troncal), con 61 kilómetros de extensión que atraviesa catorce distritos de Lima y Callao empezando por el City Gate de Lurín, pasando por Pachacamac, Villa El Salvador, Villa María del Triunfo, San Juan de Miraflores, Santiago de Surco, Santa Anita, El Agustino, San Juan de Lurigancho, Cercado de Lima, San Martín de Porres, Carmen de la Legua, el Cercado del Callao y terminando en la Central Termoeléctrica de Ventanilla, y *las* otras Redes (o ramales), que llevan el gas desde la Red Principal hasta el domicilio de los usuarios.

En este mercado podríamos distinguir cuatro segmentos según las características de los usuarios: residencial, comercial, industrial y vehicular. No obstante, según la ley, se han establecido cuatro categorías de consumidores, dependiendo de su nivel de consumo. El Cuadro 2.1 muestra los rangos de consumo existentes que definen las categorías de consumidor.

Cuadro 2.1: Categorías de Consumidor

Categoría	Rango de Consumo (m³ / mes)
A	Hasta 300
B	301 - 17500
C	17501 - 300000
D	Más de 300000

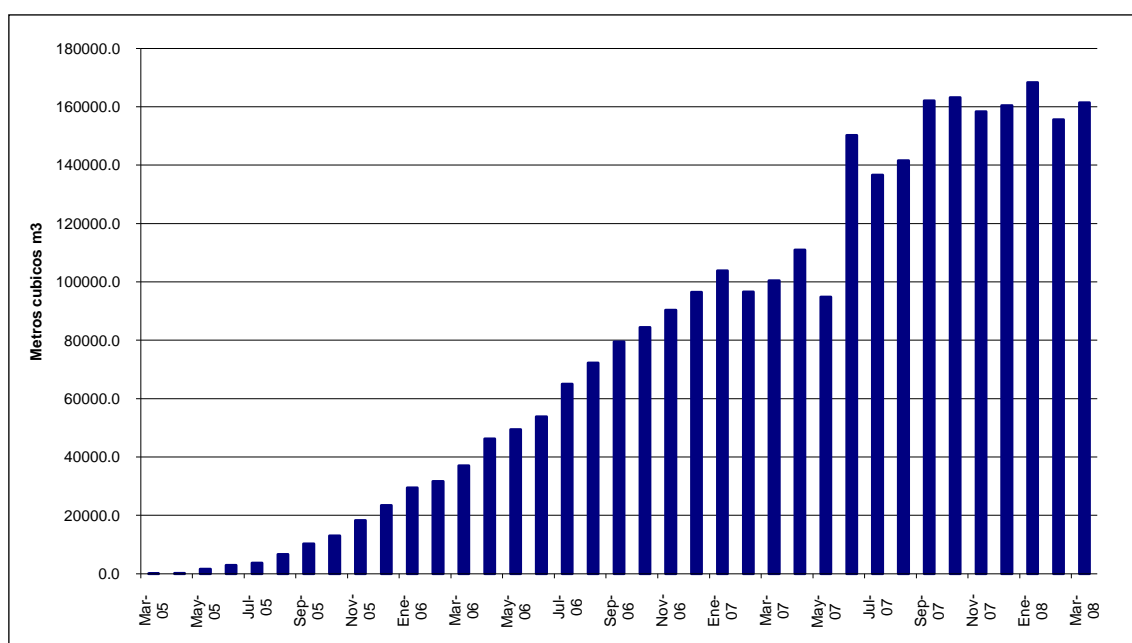
Fuente: resolución OSINERG N° 097-2004-OS/CD

Cabe resaltar que la categoría A cuyo máximo nivel de consumo es de 300 metros cúbicos de gas natural mensual está formada esencialmente por clientes residenciales.

2.2. Evolución del Consumo

El consumo de gas natural de consumidores residenciales muestra una clara tendencia creciente, inicialmente en marzo del 2005 solo se registraba un consumo mensual de 56.6 metros cúbicos de GN registrando 37 meses después en marzo del 2008 un consumo mensual de 161434 metros cúbicos de GN, es decir ha registrado una tasa de crecimiento promedio mensual de 24%. En Enero del 2008 se registró el máximo nivel histórico de consumo mensual alcanzando 168372 metros cúbicos de GN. En el Gráfico 2.2 se muestra esta evolución del consumo de GN para los consumidores residenciales.

Gráfico 2.2: Evolución del consumo residencial de Gas Natural



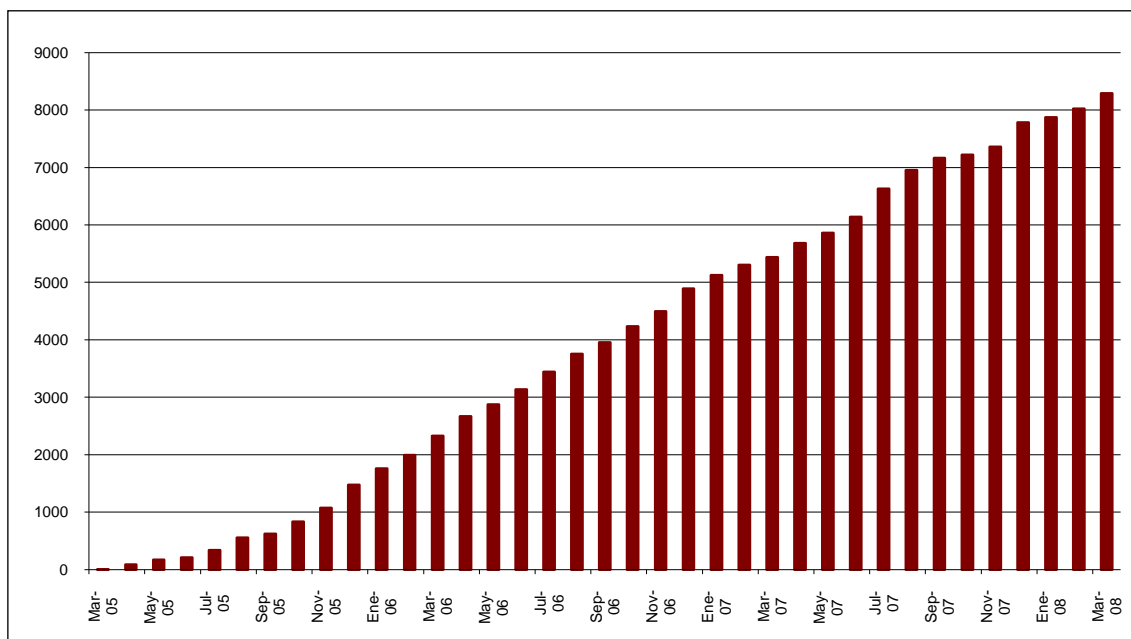
Fuente: Informes mensuales - Ministerio de Energía y Minas (MINEM)
 Elaboración: Oficina de Estudios Económicos – OSINERGMIN.

El consumo de gas natural para los consumidores residenciales muestra una clara tendencia creciente, inicialmente en marzo del 2005 solo se registraba un consumo mensual de 56.6 metros cúbicos de GN registrando 37 meses después en marzo del 2008 un consumo mensual de 161434 metros cúbicos de GN, es decir ha registrado una tasa de crecimiento promedio mensual de 24%. En Enero del 2008 se registró el máximo nivel histórico de consumo mensual alcanzando 168372 metros cúbicos de GN. En el siguiente grafico se muestra esta evolución del consumo de GN para los consumidores residenciales.

Inicialmente en marzo del 2005 solo se registraban cuatro clientes o consumidores residenciales mostrando un crecimiento a marzo del 2008 existen 8293 usuarios, en estos 32 meses de presenta una tendencia claramente creciente, habiéndose incrementado a una tasa de % mensual

promedio.

Gráfico 2.3: Evolución del número de consumidores



Fuente: Informes mensuales - Ministerio de Energía y Minas MINEM
 Elaboración: Oficina de Estudios Economicos – OSINERGMIN.

El consumo promedio de gas natural de un usuario residencial en marzo del 2005 era de 14.2 metros cúbicos mensuales. Hasta el mes de enero del 2008 el consumo promedio se ha elevado en 37%, alcanzando un nivel de 19.5 metros cúbicos mensuales.

3. Metodología

3.1. Evaluación de Impacto³

Según Baker (2002), para realizar la evaluación global de algún programa social, proyecto o política pública es necesario elaborar una adecuada supervisión, de tal forma que permita evaluar si el programa se está implementando de acuerdo a lo planeado, también se debe hacer una buena evaluación de los procesos de tal manera que permita examinar la forma en que funciona el programa y cómo entrega sus servicios, además también se debe realizar una evaluación costo-beneficio que permita comparar los costos del programa con los usos alternativos que se puedan dar a estos recursos y los beneficios que se producen de ellos, y

³ Basado en Cameron y Trivedi (2008).

finalmente, también es necesario hacer una adecuada evaluación de impacto que permita medir los efectos causados por el programa.

En este sentido la evaluación de impacto busca determinar en forma general si el programa, proyecto o política produjo los efectos deseados en las personas, hogares y/o instituciones a las que se dirige y si esos efectos son atribuibles a su intervención. Además, permite examinar las consecuencias imprevistas en sus participantes, ya sean positivas o negativas. De esta manera algunas de las preguntas que se abordan en este tipo de evaluación incluyen por ejemplo las siguientes: ¿Cómo afectó el programa a sus participantes? ¿Algún mejoramiento fue el resultado directo del programa o se habría producido de todas formas? (Baker 2002)

Sin embargo, estas preguntas no se pueden responder simplemente midiendo los resultados del programa ya que puede haber otros factores o sucesos que estén correlacionados con los resultados, sin ser causados por éste. Para asegurar un rigor metodológico, toda evaluación de impacto debe estimar un escenario alternativo llamado “escenario contrafactual”, en cual se simula lo que habría ocurrido si el programa nunca se hubiera realizado (Baker 2002). Por ejemplo, en el caso de un hogar con acceso reciente a la red de gas natural en el que se disminuyó el gasto energético, ¿es esto un efecto directo del programa o dicho hogar de todas formas habría reducido su gasto energético si no hubiera tenido acceso a la red de gas natural?

Por lo tanto, para determinar el escenario contrafactual, es necesario separar el efecto producto de la intervención del programa de otros factores ajenos a él. Esto se logra con la ayuda del grupo de control o de comparación el cual se forman con aquellas personas que no participan en el programa y que posteriormente se comparan con el grupo de tratamiento, el cual está formado con las personas que sí participan en el programa. El grupo de control debe ser lo más semejante posible en todo aspecto al grupo de tratamiento y la única diferencia entre ellos debe ser la participación en el programa.

La evaluación impacto forma parte del análisis causal, porque en ella se trata de responder a las preguntas empíricas más difíciles del análisis económico, las que involucran relaciones de causa y efecto, por ello es necesario tener en cuenta algunos conceptos del análisis causal. Para ello considérese el siguiente modelo⁴ en el cual se define a una población por la letra U , donde cada una de sus unidades es denotada por la letra u . Para cada unidad que pertenece a esta población, $u \in U$, hay asociado un valor Y de la variable que interesa analizar, la cual llamamos variable de respuesta. Cada unidad u puede ser expuesta o no a la acción de una

⁴ Basado en The Neyman-Rubin potential outcome model (RCM) (Neyman, 1923; Rubin, 1974)

causa, la cual podría ser el tratamiento de algún programa, la implementación de algún proyecto social o la aplicación de alguna política pública, para simplificar el análisis se asume que sólo hay 2 causas posibles, las cuales se pueden representar con una variable D que indica la causa a la cual cada unidad es expuesta:⁵

$$D = \begin{cases} t, & \text{si la unidad es expuesta al grupo de tratamiento.} \\ c, & \text{si la unidad es expuesta al grupo de control.} \end{cases}$$

De esta manera, D toma el valor de t si la unidad u es expuesta al grupo de tratamiento, es decir la unidad forma parte del tratamiento del programa, y toma el valor de c si la unidad fue expuesta al grupo de control, es decir, que esta unidad no forma parte del tratamiento del programa

Los valores de la variable Y depende de la causa particular t o c , a la cual la unidad es expuesta. Así, necesitamos dos variables de respuesta para cada unidad u : Y_t y Y_c , donde:

- Y_t : es el valor de la respuesta que sería observada si la unidad fuera expuesta al grupo de tratamiento
- Y_c : es el valor de la respuesta que sería observada si unidad fuera expuesta al grupo de control.

Para hacer operativo nuestro análisis, también se puede expresar al indicador de tratamiento D como una variable *dummy*, donde:

$$D = \begin{cases} D = 1, & \text{si la unidad es expuesta al grupo de tratamiento } (D = c). \\ D = 0, & \text{si la unidad es expuesta al grupo de control } (D = t). \end{cases}$$

El indicador D toma el valor de 1 si la unidad es expuesta al grupo de tratamiento y toma el valor de 0 si esta es expuesta al grupo de control. Por lo tanto, el resultado que toma la variable de respuesta puede expresarse como función de D :

$$Y = DY_1 + (1 - D)Y_0$$

De esta manera: el resultado de cada individuo puede ser reescrito como

$$\text{Si } D = 1 \Rightarrow Y = 1 \times Y_1 + (1 - 1)Y_0 = Y_1$$

$$\text{Si } D = 0 \Rightarrow Y = 0 \times Y_1 + (1 - 0)Y_0 = Y_0$$

Donde las variables de respuesta expresan lo siguiente:

- Y_1 : es el resultado de la unidad si esta fue expuesto al tratamiento.

⁵ En un estudio controlado, D es construido por el experimentador. En un estudio sin control, éste es determinado por factores que están más allá del control del experimentador.

- Y_0 : es el resultado de la unidad en ausencia del tratamiento.

De esta forma la evaluación de impacto busca estimar el efecto causado por el tratamiento del programa para cada una de las unidades, en función a los resultados de la variable de respuesta, mediante la siguiente expresión:⁶

$$\Delta = Y_1 - Y_0$$

Es decir, el efecto del tratamiento se define como la diferencia entre el resultado obtenido por una unidad tratada y el resultado que hubiera obtenido esa misma unidad si no hubiera sido tratada por el programa.

Sin embargo, esta definición plantea el problema fundamental de la inferencia causal y de la evaluación de impacto, el cual indica que es imposible observar al mismo tiempo el valor de Y_1 y Y_0 en la misma unidad u . En otras palabras, no se puede observar al mismo individuo en ambas situaciones participando del programa y dejándolo de hacer, por lo tanto, es imposible observar el efecto del tratamiento individual del programa. Otra forma de expresar este problema es decir que no podemos inferir el efecto del tratamiento por que no tenemos la evidencia contrafactual de cada participante, es decir no sabemos que resultado hubiera tenido cada participante si el programa no se hubiera implementado.

Dado que el efecto de tratamiento para una sola unidad u no puede ser observada, porque se carece de la evidencia contrafactual para cada participante, la solución típica a este problema consiste en identificar el Efecto promedio del tratamiento conocido como ATE (siglas del ingles Average Treatment Effect) para la toda población o para cualquier sub-población de ella. De esta manera el efecto causado por el tratamiento, imposible de observar individualmente, es reemplazado con la posibilidad de estimar el efecto promedio de éste, el cual es un tipo de solución estadística. Formalmente se puede expresar de la siguiente manera:⁷

$$ATE = E[\Delta]$$

$$ATE = E[Y_1 - Y_0]$$

$$ATE = E[Y_1] - E[Y_0]$$

Entonces el ATE queda expresado como la diferencia entre el resultado promedio de todos los individuos tratados por el programa $E[Y_1]$ y el resultado promedio de estos mismos como si no

⁶ Esta definición de efecto causal asume que el tratamiento de un individuo no afecta los posibles resultados de otros individuos. Por ejemplo el acceso de un hogar a la red de gas natural no afecta el gasto energético de los otros hogares que no accedieron al programa.

⁷ Nótese que esta la ecuación está definida para la población completa

hubiesen sido tratados $E[Y_0]$, lamentablemente este resultado no puede ser observado directamente, por lo que la mayoría de los métodos econométricos intentan construir estimaciones consistentes a partir de datos observacionales, mediante el siguiente estimador muestral:⁸

$$ATE = E[\hat{\Delta}]$$

$$ATE = E[\hat{Y}_1 | D = 1] - E[\hat{Y}_0 | D = 0]$$

Una condición suficiente para que este estimador sea consistente con el verdadero ATE, es que el resultado promedio bajo el tratamiento y el resultado promedio bajo control no difieran entre los grupos de tratamiento y de control, lo cual implica que:

$$E[Y_1] = E[Y_1 | D = 1] = E[Y_1 | D = 0]$$

$$E[Y_0] = E[Y_0 | D = 1] = E[Y_0 | D = 0]$$

A fin de satisfacer estas condiciones, es suficiente que las potenciales distribuciones resultantes de Y_1 y Y_0 sean independientes o que no estén correlacionada con la asignación del tratamiento D , esto formalmente se puede representar como:

$$(Y_1, Y_0) \perp D$$

Cabe resaltar, como se verá más adelante, que la principal forma de conseguir esta independencia es a través de la asignación aleatoria del tratamiento.

Sin embargo, en algunas ocasiones la asignación del tratamiento puede no ser aleatoria, o simplemente no hay información disponible sobre cómo aquellos individuos en el grupo de control habrían reaccionado si hubieran recibido el tratamiento por lo que los resultados potenciales del programa no son independientes de la asignación del mismo, lo cual se puede representar formalmente como:

$$(Y_1, Y_0) \not\perp D$$

Esto implica que:

$$E[Y_1] \neq E[Y_1 | D = 1]$$

$$E[Y_0] \neq E[Y_0 | D = 0]$$

El resultado promedio de los tratados es diferente del resultado promedio de únicamente de los tratados y El resultado promedio de los no tratados es diferente del resultado promedio de únicamente de los no tratados. Por consiguiente, al momento de utilizar esta información para obtener el efecto promedio del tratamiento, aparecerá el problema de sesgo de selección:

⁸ Nótese que este estimador evaluado en una muestra obtenida de la población utiliza esperanzas condicionales.

$$E [Y_1 | D = 1] - E [Y_0 | D = 0] \neq E[Y_1] - E[Y_0]$$

Ésta es la base para comprender los posibles sesgos del estimador muestral estándar, los cuales se pueden mostrar de la siguiente manera:

$$ATE = ATE + \underbrace{E[Y_1 | D = 1] - E[Y_0 | D = 0]}_{\text{sesgo de selección}} + (1 - \pi) \underbrace{ATE_{[D=1]} - ATE_{[D=0]}}_{\text{Heterogeneidad del tratamiento}}$$

Donde:

π : Proporción de la población que sería asignada al grupo de tratamiento.

$1 - \pi$: Proporción de la población que sería asignada al grupo de control.

$ATE_{[D=1]}$: Efecto promedio del tratamiento sobre los tratados.

$ATE_{[D=0]}$: Efecto promedio del tratamiento sobre los no tratados.

Esta ecuación especifica las dos fuentes de sesgo que necesitan ser eliminadas de las estimaciones de los efectos causales en los estudios observacionales.

1. Sesgo de Selección: el cual se relaciona con los elementos no observables que pueden sesgar los resultados como por ejemplo las condiciones preexistentes al programa o la capacidad de cada individuo.
2. Heterogeneidad del tratamiento: el efecto del tratamiento del programa no es uniforme para toda la población.

La mayoría de los métodos disponibles solamente se enfocan en el sesgo de selección, asumiendo que el efecto del tratamiento es constante en toda la población.

Dado que el efecto de tratamiento no puede ser observado individualmente, otro tipo de solución alternativo al ATE, consiste en identificar el Efecto promedio del tratamiento para las unidades que han sido tratadas por el programa conocido como ATET (siglas del inglés Average Treatment Effect on the Treated), formalmente se puede expresar de la siguiente manera:

$$ATET = E[\Delta | D = 1]$$

$$ATET = E[(Y_1 - Y_0) | D = 1]$$

$$ATET = E [Y_1 | D = 1] - E [Y_0 | D = 1]$$

Nuevamente, el estimador muestral estándar $ATET = E[\hat{Y}_1 | D = 1] - E[\hat{Y}_0 | D = 0]$ estima consistentemente el ATET solo si se cumple que $E[Y_0 | D = 1] = E[Y_0 | D = 0]$

3.2. Propensity Score Matching

Si bien el supuesto de asignación aleatoria del tratamiento puede parecer poco plausible en la mayoría de aplicaciones empíricas, éste puede lograrse si el investigador cuenta con un conjunto de variables X . La idea detrás de este método es que, la asignación del tratamiento, condicionada a este conjunto de variables, produzca una asignación aleatoria. Formalmente, puede representarse como:

$$Y_1, Y_0 \perp D | X$$

Por lo tanto, cualquier muestra que en principio podría tener sesgo de selección, puede ser transformada en una que carece de él siempre y cuando se cuente con un *set* de variables lo suficientemente amplio como para lograr esta independencia condicional. Este supuesto también es conocido como el de selección en observables (Rubin, 1983). Por otro lado, un supuesto que nos permite realizar comparaciones entre controles y tratados es el siguiente⁹:

$$0 < \Pr D = 1 | X = x < 1$$

No obstante, uno de los principales problemas de los métodos tradicionales de *matching* consiste en el condicionamiento por las variables X . Para efectuar la aleatorización condicional con este vector de características, debemos hacer que las variables asociadas a los individuos de los grupos de control y de tratamiento sean lo más cercanas posibles, empleando algún indicador de distancia (métrica). Si este vector tiene una dimensión muy grande, sería difícil definir en la práctica el grado de cercanía entre dos valores de X . Al respecto, existió un debate bastante importante en cuanto a los criterios para determinar cuándo un individuo perteneciente al grupo de control era lo suficientemente parecido al del grupo de tratamiento¹⁰.

No obstante, Rosenbaum y Rubin (1983) encontraron una forma de resolver este problema. En términos generales, la metodología propuesta por ellos consistía en reducir la dimensionalidad del vector X mediante una función $p(X)$ que generara la aleatorización deseada. Es decir:

$$Y_1, Y_0 \perp D | p(X)$$

A esta función $p(X)$ se le denomina *propensity score* y se define como la probabilidad de recibir tratamiento condicional a las variables que forman parte del vector X . Por lo tanto, si existe un vector X sobre el cual se pueda calcular esta función $p(X)$, la expresión para el efecto de tratamiento sería la siguiente:

⁹ De acuerdo a Heckman, Vytacil y Urzúa (2006), este supuesto es común a todos aquellos modelos que involucren grupos de control y tratamiento.

¹⁰ Los papers de LaLonde (1986) y Dehejia y Wahba (1999) muestran parte de este debate.

$$\Delta X = E Y_1 - Y_0 | p X, D=1 = E Y_1 | p X, D=1 - E Y_0 | p X, D=0$$

Un punto a considerar de especial importancia es la validación del supuesto de aleatorización condicional al *propensity score*. Dado que no existe ningún contraste de hipótesis para determinar la validez del supuesto, generalmente se apela a la calidad de la información utilizada para poder justificar el uso del *matching*. Adicionalmente, Dehejia y Wahba (2002) proponen una serie de criterios para determinar si la función $p X$ es consistente con el criterio de aleatorización.

Por consiguiente, una primera etapa para calcular el efecto de tratamiento consiste en estimar el *propensity score*, para lo cual se puede emplear cualquier modelo de elección binaria. A diferencia de otras aplicaciones, el rol de la estimación del modelo es el de reducir la dimensionalidad, mas no de proveer algún sustento teórico o de comportamiento a la elección del tratamiento (Dehejia y Wahba, 2002). Luego de calcular el modelo, es útil analizar la distribución del *propensity score* entre tratados y no tratados, a fin de determinar si los promedios de las variables que conforman el vector X son similares entre ambos grupos.

Un segundo paso se basa en determinar una métrica que permita encontrar el individuo del grupo de control más parecido al que ha recibido el tratamiento. Esta técnica, conocida como *matching*, es fundamentalmente un esquema de ponderaciones, que determina la magnitud de los pesos que se asignan al grupo de tratamiento (Heckman, Ichimura y Todd, 1998).

$$\hat{\tau}_{T=1} = \frac{1}{|N|} \sum_{i \in N} \left(Y_i - \frac{1}{|J_i|} \sum_{j \in J_i} Y_j \right)$$

Donde N es el grupo de tratamiento, $|N|$ es el número de unidades en el grupo de tratamiento, J_i es el conjunto de unidades de comparación (controles) emparejados con el individuo i tratado, y $|J_i|$ es el número de unidades de comparación en J_i . Dependiendo de la metodología que se utilice para realizar el emparejamiento, éste puede realizarse con o sin reemplazo, o con uno o más vecinos provenientes del grupo de control¹¹. Cada uno de estos métodos tiene sus ventajas y desventajas. En primer lugar, existe un *trade off* entre el sesgo y la varianza. Si se emplea un *matching* sin reemplazo, el sesgo puede incrementarse conforme se vayan eligiendo individuos que se encuentran más lejos del tratado, especialmente cuando el número de controles en la cola derecha de la distribución es bastante reducido. Este procedimiento reduce la precisión del estimador, puesto que cada vez los valores son menos precisos. Por otra parte, el

¹¹ Un resumen de las metodologías de *matching* se muestran en el Anexo 1.

matching con reemplazo permite comparar individuos del grupo de control por más de una vez, lo cual disminuye el sesgo.

4. Implementación Empírica

4.1 Información Utilizada

La información proveniente para esta evaluación de impacto proviene de la Encuesta Residencial de Uso y Consumo de Energía realizada en los meses de septiembre y octubre del 2007, conjuntamente con la encuesta complementaria a usuarios de gas natural realizada en los meses de marzo y abril del 2008. Esta encuesta tiene una representatividad departamental y, dentro de Lima Metropolitana, sobrerrepresenta al grupo de hogares con acceso al servicio de gas natural, siendo también representativa de los distintos niveles socioeconómicos.

Dado que a la fecha el gas natural domiciliario únicamente se encuentra disponible en Lima Metropolitana, nos concentraremos en este ámbito. El grupo de tratamiento se obtuvo a partir de la submuestra realizada a 594 hogares con acceso a gas natural, situados en los distritos de Lima, Pueblo Libre, San Miguel y Santiago de Surco.

La elección del grupo de tratamiento, idealmente, debería comprender a aquellos hogares residentes en el mismo distrito pero que no tenían acceso al gas natural. No obstante, el diseño de la encuesta hace que el número de hogares encuestados sea insuficiente para servir de grupo de control. Por este motivo se decidió elegir, en principio, a todos aquellos hogares residentes en Lima Metropolitana. De este conjunto, se excluyeron a todas las viviendas que no se encontraran en los planes de crecimiento de la red de distribución de gas natural, tal como muestra en la sección 2.1. Como resultado, se eliminaron todos aquellos hogares que residían en distritos periféricos¹².

Hecha esta selección, quedaron 1083 hogares en el grupo de control y 594 en el de tratamiento. Esta muestra representa en la población a 1.395 millones de hogares y 3724 hogares, respectivamente. El Cuadro 4.1 muestra la distribución de ambos grupos.

Cuadro 4.1 Grupos de Control y Tratamiento (Sin Expandir)

¹² Más específicamente, los distritos que fueron excluidos de la muestra fueron Cieneguilla, Carabaylo, Lurigancho, Puente Piedra y Ancón.

Grupos	Observaciones	Porcentaje	Observaciones	Porcentaje
Controles	1,083	64.6%	1,395,562	99.7%
Tratados	594	35.4%	3,724	0.3%
Total	1,677		1,399,286	

Fuente: Encuesta Residencial de Uso y Consumo de Energía 2007.
Elaboración: Oficina de Estudios Económicos – OSINERGMIN.

Por otra parte, las variables sobre las cuales se determinarán los impactos serán el gasto en combustibles (GLP versus gas natural), medido en Nuevos Soles, así como el consumo en unidades físicas (medido en metros cúbicos y kJ/kg). Con respecto a estas variables, el Cuadro 4.2 muestra el promedio de gasto y consumo de cada hogar.

Cuadro 4.2
Gastos y Consumo por Grupos de Control y Tratamiento

	Gasto (S/.)	Diferencia (T-C)	Consumo (m3)	Diferencia (T-C)
Muestra Expandida				
Controles	40.5	-	12.4	8.0
Tratados	20.2	20.3	20.4	
Muestra sin Expandir				
Controles	41.8	-	12.8	
Tratados	20.2	21.6	20.3	7.5

Fuente: Encuesta Residencial de Uso y Consumo de Energía 2007.
Elaboración: Oficina de Estudios Económicos – OSINERGMIN.

En principio, la diferencia de consumo podría ser de aproximadamente 20 soles. Estas cifras son similares a los estimados de obtenidos tanto por el Ministerio de Energía y Minas como por la empresa distribuidora de gas natural en Lima Metropolitana. No obstante, estos trabajos tienen la limitación de considerar únicamente el efecto de la reducción en el precio, sin analizar cuánto varía el consumo. Puede darse el caso que, ante la expectativa de un menor precio por la energía que consumen, los hogares decidan comprar nuevos bienes durables (como calentadores de agua o sistemas de calefacción de ambientes), por lo que el consumo se vea incrementado también. En este sentido, las técnicas presentadas en la sección 3 permiten controlar estos efectos. Con este fin, es preciso definir una serie de variables que, condicionadas a la asignación del tratamiento, la hagan aleatoria.

Con el fin de elegir las variables que serían las más apropiadas para condicionar la asignación al tratamiento, se seleccionaron variables que describieran las características del jefe de hogar, así como las de la vivienda. Las variables que se utilizarán en el procedimiento de *matching* en la siguiente sección se describen en el Cuadro 4.3.

Cuadro 4.3
Descripción de Variables

GRUPO DE CONTROL						
Variable	Observaciones	Promedio	Desv. Estándar	Mínimo	Máximo	
Hogar con equipo de calefacción de agua	1083	0.1247	0.3305	0	1	
Sexo (1=hombre, 0=mujer)	1083	0.7682	0.4222	0	1	
Edad del jefe de hogar	1083	52.23	14.7248	18	99	
Edad del jefe de hogar (al cuadrado)	1083	2944.78	1599.5580	324	9801	
Año en el que llegó a la vivienda	1083	1988.08	14.1361	1932	2007	
Presencia de actividad económica en el hogar	1083	0.1016	0.3022	0	1	
Número de ambientes en el hogar	1083	4.4192	2.1727	1	16	
Vivienda con medidor de agua	1083	0.7488	0.4339	0	1	
Acceso a TV cable	1083	0.3490	0.4769	0	1	
Acceso a Internet	1083	0.1736	0.3789	0	1	
Acceso a Telefonía Móvil	1083	0.7175	0.4504	0	1	
Suministro individual	1083	0.8994	0.3010	0	1	
Vivienda propia	1083	0.8052	0.3963	0	1	
Techo de concreto	1083	0.6833	0.4654	0	1	
GRUPO DE TRATAMIENTO						
Variable	Observaciones	Promedio	Desv. Estándar	Mínimo	Máximo	
Hogar con equipo de calefacción de agua	594	0.4731	0.4997	0	1	
Sexo (1=hombre, 0=mujer)	594	0.7104	0.4539	0	1	
Edad del jefe de hogar	594	63.59	13.2851	19	94	
Edad del jefe de hogar (al cuadrado)	594	4220.22	1613.1750	361	8836	
Año en el que llegó a la vivienda	594	1980.71	12.7350	1940	2007	
Presencia de actividad económica en el hogar	594	0.0623	0.2419	0	1	
Número de ambientes en el hogar	594	5.7374	1.6452	1	13	
Vivienda con medidor de agua	594	0.9916	0.0914	0	1	
Acceso a TV cable	594	0.7643	0.4248	0	1	
Acceso a Internet	594	0.5572	0.4971	0	1	
Acceso a Telefonía Móvil	594	0.8636	0.3435	0	1	
Suministro individual	594	0.9848	0.1223	0	1	
Vivienda propia	594	0.9882	0.1080	0	1	
Techo de concreto	594	0.9781	0.1464	0	1	

Fuente: Encuesta Residencial de Uso y Consumo de Energía 2007.

Elaboración: Oficina de Estudios Económicos – OSINERGMIN.

De este cuadro puede concluirse que, en principio, tanto los grupos de control como de tratamiento presentan ligeras diferencias entre sí. En particular, el grupo de tratamiento pareciera contar con un estatus socioeconómico más alto, dado que cuenta con un mayor porcentaje de equipos de calefacción, una vivienda con mejores materiales, mayor acceso a servicios como cable e Internet y menos hogares donde se realizan actividades económicas. También puede apreciarse que son hogares que llevan relativamente más tiempo de residentes en la misma vivienda. Sin embargo, cabe señalar que los procedimientos de *matching* permiten trabajar con muestras de este tipo, de tal modo que se puedan encontrar los

4.2 Resultados

Para poder realizar el *matching*, es preciso estimar el *propensity score*, que se obtiene a partir de la estimación de un modelo de elección discreta binario (probit o logit). Tal como se ha expuesto en la sección 4, la estimación de este modelo es un requisito para disminuir la dimensionalidad y generar la aleatoriedad del tratamiento condicional al vector de características elegido. En este sentido, se creó una variable dependiente que toma el valor de uno si el hogar posee acceso a gas natural, mientras que es igual a cero si el hogar cuenta con

GLP. Las variables explicativas elegidas fueron algunas características de la vivienda como del jefe de hogar. Los resultados del modelo se muestran en el Cuadro 4.4.

**Cuadro 4.4; Resultados del Modelo Probit
(Variable Dependiente: Acceso a Gas Natural Domiciliario)**

Variable	Coefficiente	Err. Estándar	Estadístico t	p-value
Hogar con equipo de calefacción de agua	0.6411	0.0882	7.27	0.0000
Sexo (1=hombre, 0=mujer)	- 0.0960	0.0875	-1.10	0.2720
Edad del jefe de hogar	- 0.0041	0.0217	-0.19	0.8490
Edad del jefe de hogar (al cuadrado)	0.0002	0.0002	1.19	0.2340
Año en el que llegó a la vivienda	- 0.0045	0.0032	-1.38	0.1680
Presencia de actividad económica en el hogar	- 0.2274	0.1398	-1.63	0.1040
Número de ambientes en el hogar	- 0.0218	0.0216	-1.01	0.3110
Vivienda con medidor de agua	0.9856	0.2366	4.17	0.0000
Acceso a TV cable	0.5043	0.0877	5.75	0.0000
Acceso a Internet	0.5581	0.0908	6.14	0.0000
Acceso a Telefonía Móvil	0.1619	0.1077	1.50	0.1330
Suministro individual	0.1646	0.2634	0.62	0.5320
Vivienda propia	1.4341	0.2266	6.33	0.0000
Techo de concreto	0.9476	0.1783	5.32	0.0000
Constante	4.1298	6.5539	0.63	0.5290
Observaciones	1677			
Wald Chi2(15)	345.17			
p-value	0.0000			
Pseudo R2	0.3702			

Fuente: Encuesta Residencial de Uso y Consumo de Energía 2007.

Elaboración: Oficina de Estudios Económicos – OSINERGMIN.

4.2.1 Impacto Sobre el Gasto en Energía

Los resultados de la evaluación para el gasto en energía se muestran en el Cuadro 5.5. Al respecto, puede verse que, independientemente del método de *matching* empleado, existe un ahorro importante en el gasto en energía. Más específicamente, el gasto promedio en combustible disminuye entre 22 y 24.5 Nuevos Soles, siendo la variabilidad bastante reducida entre los distintos métodos (el coeficiente de variación es del 4%). En términos porcentuales, la disminución sería de aproximadamente 54% en comparación con el escenario en el que un hogar utiliza GLP como fuente de cocción y calefacción. Si se realiza una agregación entre todos los hogares que consumen gas natural al momento en que se efectuó la encuesta, puede verse que el ahorro total promedio al mes sería de 87,704 Nuevos Soles, que a su vez representa 1'052,450 Nuevos Soles al año.

Dado que, según la encuesta residencial del 2007, un elevado porcentaje de hogares optó por financiar la instalación de gas natural en sus hogares, el costo mensual adicional que supone este pago también fue incluido en la evaluación de impacto. Como sería de esperarse, los plazos del financiamiento, así como el monto total involucrado en la instalación domiciliaria hacen que el

costo asociado al gas natural se incrementa de manera sustancial, haciéndolo menos competitivo que el GLP. En particular, mientras dure el financiamiento, el costo promedio sería mayor en alrededor de 48 Nuevos Soles (92 versus 44).

Esta situación podría llevar a un contexto en el que los hogares tengan en consideración lo elevado del financiamiento de la instalación inicial para poder contar con los beneficios futuros. Al respecto, la evidencia muestra que cuando los hogares valoran de forma desigual los beneficios futuros y los costos presentes, éstos pueden tomar decisiones ineficientes y optar, por ejemplo, por no acceder al gas natural. Por este motivo, es necesario proveer facilidades a fin de que los costos presentes de acceso al servicio sean menores y así masificar el servicio.

4.2.2 Sobre el Consumo de Energía

El cuadro 5.5 también muestra los resultados de la disponibilidad del gas natural sobre el consumo de energía. Los resultados muestran que el consumo equivalente se ha incrementado en casi 66.29 MBTU/lb, siendo estos valores significativos y relativamente homogéneos entre los distintos métodos empleados. Más específicamente, se pasa de un consumo de 309.28 MBTU/lb (en el caso del GLP) a 375.58 MBTU/lb con gas natural, lo que representa un aumento de poco más de 20%¹³.

Cuadro 4.5: Resultados de la Evaluación de Impacto (Distintos Métodos de *Matching*)

¹³ Para ello, se asumió que el GLP contiene 100% de propano con un valor calorífico de 21.6 MBTU/lb, mientras que para el gas natural se asumió un valor calorífico de 18.5 MBTU/lb. Si se considera que el GLP puede encontrarse combinado con butano, la diferencia a favor del gas natural sería incluso mayor.

	Nearest Neighbor (Un Vecino) Con Reemplazo	Nearest Neighbor (Un Vecino) Sin Reemplazo	Nearest Neighbor (Cinco Vecinos) Con Reemplazo	Local Linear Regression (Kernel: Epanechnikov)	Matching con Kernel (Kernel: Epanechnikov)
Consumo (MBTU/lb)					
Consumo GLP (MBTU/lb)	306.00	291.09	311.76	312.22	309.28
Consumo Gas Natural (MBTU/lb)	375.58	375.58	375.58	375.58	375.58
Diferencia (MBTU/lb)	69.57	84.49	63.81	63.36	66.29
Error Estándar	22.12	11.23	18.12	17.49	16.94
Estadístico t	3.15	7.52	3.52	3.62	3.91
Aumento en Consumo (%)	22.7%	29.0%	20.5%	20.3%	21.4%
Gasto (Nuevos Soles - No Considera Costo de Instalación)					
Gasto GLP (Nuevos Soles)	44.99	44.21	45.99	46.38	45.75
Gasto Gas Natural (Nuevos Soles)	20.18	20.18	20.18	20.18	20.18
Diferencia (Nuevos Soles)	- 24.81	- 24.03	- 25.81	- 26.20	- 25.57
Error Estándar	3.84	0.64	2.78	2.58	2.42
Estadístico t	- 6.45	- 37.36	- 9.27	- 10.14	- 10.57
Disminución en Gasto (%)	55.1%	54.4%	56.1%	56.5%	55.9%
Gasto (Nuevos Soles - Considera Costo de Instalación y Financiamiento)					
Gasto GLP (Nuevos Soles)	44.99	44.21	45.99	46.38	45.75
Gasto Gas Natural (Nuevos Soles)	92.85	92.85	92.85	92.85	92.85
Diferencia (Nuevos Soles)	47.86	48.64	46.86	46.48	47.11
Error Estándar	4.41	2.25	3.52	3.37	3.24
Estadístico t	10.86	21.60	13.30	13.80	14.53
Aumento en Gasto (%)	106.4%	110.0%	101.9%	100.2%	103.0%

Fuente: Encuesta Residencial de Uso y Consumo de Energía 2007.

Elaboración: Oficina de Estudios Económicos – OSINERGMIN.

Por ello, este ahorro posee una doble dimensión, por cuanto los hogares incrementan su consumo de GN (en comparación con los hogares que utilizan GLP), a un menor costo. Por consiguiente, podríamos afirmar que existe un efecto positivo tanto en precios como en cantidades.

4.2.3 Evaluación de los Resultados

A fin de evaluar la validez del procedimiento de *matching*, se analizó la distribución del *propensity score*, así como los promedios de las variables utilizadas en la estimación del probit. El Gráfico 4.2 muestra la distribución para los grupos de tratamiento (color rojo) y de control (azul). Puede verse que la distribución del *score* en el primer grupo presenta una asimetría hacia la derecha, mientras que ocurre lo contrario para el *score* del segundo grupo. Del gráfico también puede verse que existe una ligera falta de controles para el tramo superior de la distribución del *score*. De ahí la necesidad de adoptar esquemas de ponderación y emparejamientos con reemplazo que permitan solucionar este posible problema.

Una segunda verificación de los supuestos de asignación aleatoria del tratamiento con información observacional consiste en evaluar los promedios de las variables utilizadas para hacer el *matching* tanto antes como después del procedimiento. El Cuadro 5.6 muestra que el sesgo presente en la distribución del *score* entre los grupos de tratamiento y de control habría

logrado ser eliminado en la gran mayoría de las características (se observa una reducción de más del 90% en el sesgo inicial), lo cual respalda la validez del supuesto de *matching*.

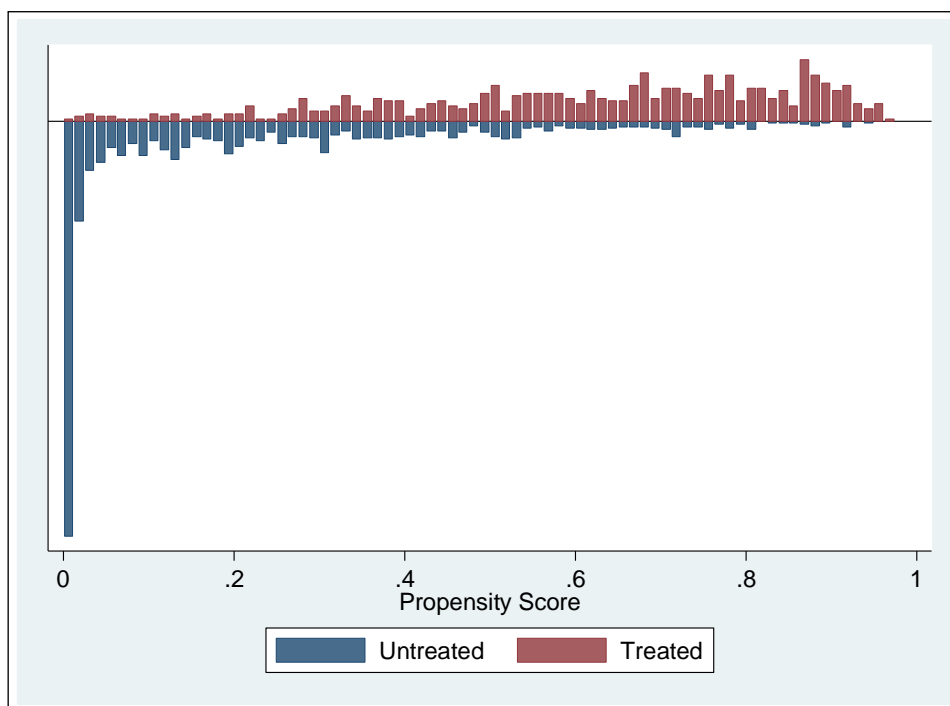
**Cuadro 5.6: Promedio de Variables Empleadas en el Probit
(Tratados y Controles)**

Variable	Muestra	Tratado	Control	% de Sesgo	Reducción en Sesgo (%)	Diferencia de Medias	
						t	p>t
Hogar con equipo de c	Sin matching	0.473	0.125	82.1		17.08	0.0000
	Con matching	0.473	0.473	0.1	99.8	-14.78	0.0000
Sexo (1=hombre, 0=m	Unmatched	0.710	0.769	- 13.3		-2.62	0.0090
	Matched	0.710	0.688	5.2	60.6	4.71	0.0000
Edad del jefe de hoga	Unmatched	63.593	52.228	81.0		15.63	0.0000
	Matched	63.593	64.008	- 3.0	96.3	-15.56	0.0000
Edad del jefe de hoga	Unmatched	4220.2	2944.3	79.4		15.57	0.0000
	Matched	4220.2	4295.0	- 4.7	94.1	-15.64	0.0000
Año en el que llegó a l	Unmatched	1980.7	1988.1	- 54.7		-10.55	0.0000
	Matched	1980.7	1980.3	3.1	94.4	10.19	0.0000
Presencia de activida	Unmatched	0.062	0.102	- 14.4		-2.74	0.0060
	Matched	0.062	0.058	1.6	88.8	3.27	0.0010
Número de ambientes	Unmatched	5.737	4.414	68.7		12.95	0.0000
	Matched	5.737	5.772	- 1.8	97.4	-11.61	0.0000
Vivienda con medidor	Unmatched	0.992	0.748	77.6		13.5	0.0000
	Matched	0.992	0.986	1.9	97.5	-10.96	0.0000
Acceso a TV cable	Unmatched	0.764	0.347	92.4		17.8	0.0000
	Matched	0.764	0.749	3.4	96.3	-15.7	0.0000
Acceso a Internet	Unmatched	0.557	0.172	87.2		17.78	0.0000
	Matched	0.557	0.538	4.3	95.0	-17.14	0.0000
Acceso a Telefonía M	Unmatched	0.864	0.717	36.7		6.92	0.0000
	Matched	0.864	0.862	0.4	99.0	-5.4	0.0000
Suministro individual	Unmatched	0.985	0.899	37.3		6.64	0.0000
	Matched	0.985	0.978	2.8	92.4	-4.3	0.0000
Vivienda propia	Unmatched	0.988	0.805	63.2		11.06	0.0000
	Matched	0.988	0.982	2.0	96.8	-10.34	0.0000
Techo de concreto	Unmatched	0.978	0.682	85.7		15.07	0.0000
	Matched	0.978	0.973	1.5	98.2	-8.97	0.0000

Fuente: Encuesta Residencial de Uso y Consumo de Energía 2007.

Elaboración: Oficina de Estudios Económicos – OSINERGMIN.

Gráfico 5.2: Distribución del *Propensity Score* (Grupos de Control y Tratamiento)



Fuente: Encuesta Residencial de Uso y Consumo de Energía 2007.
Elaboración: Oficina de Estudios Económicos – OSINERGMIN.

5 Conclusiones

Existe evidencia preliminar de que el ahorro derivado de la implementación del gas natural domiciliario ha sido significativo. El efecto producido tendría una doble dimensión: por una parte, los hogares tienen un menor gasto en comparación con la situación anterior, mientras que el consumo (tanto en unidades físicas como en términos caloríficos) se ha incrementado. No obstante, debe considerarse que el costo de acceso podría causar, en una primera instancia, que el beneficio derivado de acceder se vea revertido: dado que el financiamiento de la instalación se paga en cuotas, el ahorro inicial se convierte en un costo (al menos durante los primeros meses). Por ello, si el consumidor valora de manera distinta los beneficios futuros en comparación con los costos presentes, puede estar ocasionándose una menor cobertura que la que realmente debería ocurrir.

6 Referencias

Baker, J. (2000) "Evaluating the Impact of Development Projects on Poverty: A Handbook for Practitioners" Washington: Banco Mundial.

Cameron y Trivedi (2005). *Microeconometrics*.

Heckman, J., E. Vytacil y S. Urzúa (2006) "Understanding Instrumental Variables in Models with Essential Heterogeneity" *The Review of Economics and Statistics* 88(3): 389-432.

Rosenbaum, P. y D. Rubin (1983) "The Central Role of the Propensity Score in Observational Studies for Causal Effects" *Biometrika* 70: 41-55.

Dehejia, R. y S. Wahba (1999) "Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs." *Journal of the American Statistical Association* 94(448): 1053-1062.

Anexo 1

Métodos de Matching

En los métodos de *matching* el estimador del efecto promedio de tratamiento sobre los tratados (ATET) se basa fundamentalmente en un esquema de ponderaciones que se asigna a cada unidad de control:

$$ATE T^m = \frac{1}{N_T} \sum_{i \in D=1} \left[Y_{1,i} - \sum_j w_{i,j} Y_{0,j} \right]$$

Dependiendo de la forma y magnitud que se le de a estas ponderaciones o pesos $w_{i,j}$, se generaran diferentes tipos de estimadores *matching* del ATET, entre las técnicas que se usan para asignar estos pesos, podemos mencionar las que emparejan uno a uno como las que utilizan al o los vecinos más cercanos y las técnicas de promedio suavizado que utilizan funciones del tipo *kernel*, entre otras.

Técnicas de uno a uno

Estas técnicas utilizan únicamente a un conjunto de individuos no tratados para generar el grupo de control para cada individuo tratado, entre ellas tenemos a las que utiliza al vecino mas cercano.

Vecino más cercano

El emparejamiento se realiza con el o los individuos no tratados cuyo vector de características observadas esté a la mínima distancia euclidiana de la del participante, de esta forma, la técnica genera para cada individuo tratado i un grupo de control mediante la siguiente expresión:

$$A_j(x) = \left[j \mid \min_j \|x_i - x_j\| \right]$$

Donde $A_j(x)$ denota al conjunto de individuos no tratados con características observables x que sirve de grupo de control para cada individuo tratado, j representa a un individuo no tratado, el vector x_j representa las características observadas del individuo j , el vector x_i representa las características observadas del individuo i y $\| \cdot \|$ denota la distancia euclidiana entre vectores.

El estimador del ATET bajo esta técnica quedara expresado como el promedio de las diferencias de cada individuo tratado con su respectivo grupo de control. Un ejemplo específico del

estimador ATET es aquel propuesto por Dehejia and Wahba, 2002, en el cual el peso que se le

da a cada matched es el mismo para todos: $w_{i,j} = \frac{1}{N_C}$

$$ATE T^m = \frac{1}{N_T} \sum_{i \in D=1} \left[Y_{1,i} - \frac{1}{N_C} \sum_j Y_{0,j} \right]$$

Donde N_C representa al número de individuos en el conjunto $A_j x$, es decir, es número de individuos no tratados que sirve de grupo de control para cada individuo tratado.

Por ejemplo, si tuviéramos 3 individuos tratados ($i=3$) y utilizáramos como unidades de control a 2 para cada individuo tratado, el estimador matching de esta técnica para el ATET sería:

$$ATE T^m = \frac{1}{3} \left[Y_{1,1} - 0.5Y_{0,1} - 0.5Y_{0,2} + Y_{1,2} - 0.5Y_{0,1} - 0.5Y_{0,2} + Y_{1,3} - 0.5Y_{0,1} - 0.5Y_{0,2} \right]$$

Nótese que si el peso que se le da $w_{i,j} = 1$ en la ecuación general:

$$ATE T^m = \frac{1}{N_T} \sum_{i \in D=1} \left[Y_{1,i} - \sum_j w_{i,j} Y_{0,j} \right]$$

Cuando $j \in A_i x$ y 0 en cualquier otro caso, entonces la técnica utiliza únicamente a un individuo para construir el grupo de control para cada individuo tratado.

Si hay más de un control que cumple la condición, se toma el promedio simple.

A este método se le puede refinar un poco más, de tal forma que se crea un Vecino más próximo restringido: Similar al anterior, pero para formar el conjunto de comparación $A_j X$ se impone la condición adicional de que la distancia debe ser menor que un valor arbitrario r de tal manera que:

$$A_j x = \left[j \mid \min_j \|x_i - x_j\| < r \right]$$

Esto asegura la calidad de los emparejamientos, pero puede restar observaciones.

Otra forma de refinar el Vecino más próximo es dándole una definición aleatoria, de tal manera que si hay más de un control que cumple la condición, escoge uno aleatoriamente.

En general los métodos del vecino más próximo operan con reemplazo, de manera que un mismo control puede ser emparejado con varios participantes.

Matching con funciones de Kernel

Existen alternativas al método paramétrico de estimación de percentiles anteriormente descrito, que no vamos a comentar aquí por su complejidad teórica. Una de ellas se basa en estimar los percentiles aplicando "filtros" a los datos, de tal manera que se aplica mayor peso a los datos próximos al punto del percentil y menor a medida que nos alejamos (*kernel density estimation*).

A diferencia de las técnicas de uno a uno, las técnicas de kernel estiman el resultado contrafactual de cada participante usando el promedio ponderado de los resultados de todos los controles, donde la ponderación es inversamente proporcional a la distancia en las características observadas.

De esta manera el ATET quedara expresado como el promedio de la diferencias entre cada individuo tratado y el promedio ponderado de todo el grupo de control, formalmente:

$$ATE^{m} = \frac{1}{N_T} \sum_{i \in D=1} \left[Y_{1,i} - \sum_j w_{i,j} Y_{0,j} \right]$$

Donde los pesos o ponderaciones de cada unidad de control son inversamente proporcionales a la distancia en las características observadas y tienen la siguiente forma:

$$w(i, j) = \frac{K\left(\frac{x_i - x_j}{h}\right)}{\sum_{j=1}^{N_{c,i}} K\left(\frac{x_i - x_j}{h}\right)}$$

Donde h es el parámetro de suavización llamado banda de suavización o bandwidth, en algunos casos hh puede representar a una ventana de suavización o window width, $K\left(\frac{x_i - x_j}{h}\right)$

representa a la función de kernel, la cual es no negativa, simétrica y unimodal,

Si hacemos $\frac{x_i - x_j}{h} = z$, la forma funcional que adopte la función $K(z)$ determinará el tipo de

Kernel.