

# Estimativa do limiar de evasão de clientes Sabesp

## ESTIMATION OF THRESHOLD OF SABESP'S ESCAPING CLIENTS

PhD Denisard Cneio de Oliveira Alves <sup>(1)</sup>

Me. Paula Carvalho Pereda <sup>(2)</sup>

Daniel da Silva Grimaldi <sup>(3)</sup>

Alexsandros Fraga <sup>(4)</sup>

<sup>(1)</sup> PhD Yale University, YU, Estados Unidos, professor titular de Econometria e Economia da Universidade de São Paulo, [dcoalves@usp.br](mailto:dcoalves@usp.br)

<sup>(2)</sup> Doutoranda em Teoria Econômica da Universidade de São Paulo, [pereda@usp.br](mailto:pereda@usp.br)

<sup>(3)</sup> Mestrando em Teoria Econômica da Universidade de São Paulo

<sup>(4)</sup> Mestrando em Teoria Econômica da Universidade de São Paulo

**Endereço:** Av. Professor Luciano Gualberto, 908 – FEA1, Sala C113 - Cidade Universitária, São Paulo, SP. CEP: 05508-010.

**Entrada:** 7/4/2009 **Aprovação:** 16/10/2009

### RESUMO

*O objetivo deste estudo é estimar o limiar do fornecimento de água que, dada a atual estrutura tarifária da Sabesp, leva os grandes clientes a buscarem fontes alternativas de fornecimento de água. Foram utilizados modelos de escolha binária para estimar a probabilidade de evasão destes clientes. A principal conclusão que pode ser tirada deste trabalho foi que os níveis de consumo em que a probabilidade chega próxima ao seu máximo, variaram entre 500m<sup>3</sup> e 1200m<sup>3</sup>, indo ao encontro de evidências encontradas pela Sabesp.*

### ABSTRACT

*This work was carried out to estimate the level of water consumption that, given the Sabesp current price structure, leads the clients to search for alternative water suppliers. Binary choice models were used to estimate the clients' probability to leave SABESP system. The main conclusion of this study was that the monthly water consumption level, whose clients' probabilities achieve the maximum, varies from 500m<sup>3</sup> to 1200m<sup>3</sup>.*

**Palavras-chave:** logit com efeitos fixos; modelos de probabilidade; fornecimento de água.

**Key words:** logit and fixed effects models; binary choice models; water supplying.

### 1. INTRODUÇÃO

O objetivo deste estudo é estimar o limiar do fornecimento de água que, dada a atual estrutura tarifária da Sabesp, leva os grandes clientes a buscarem fontes alternativas de fornecimento de água.

Foi identificada pela Sabesp a recorrência de seus maiores clientes na busca por novas fontes de fornecimento de água, sendo as principais fontes alternativas concorrentes: poços (semi) artesianos; contratação de caminhões-pipa; e transporte de água de municípios vizinhos.

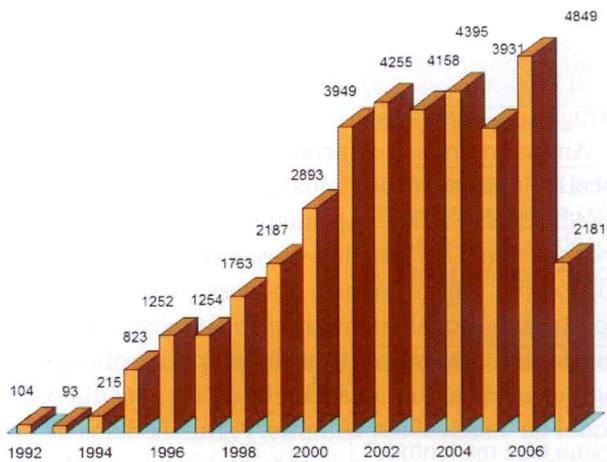
As principais evidências de aumento da concorrência com a Sabesp, ou seja, do crescimento de mercado concorrencial para o fornecimento de água para grandes clientes, são: a perda efetiva de clientes da Sabesp nos últimos anos; e o aumento do número de outorgas concedidas pelo Departamento de Águas e Energia Elétrica (DAEE) para fontes alternativas.

A Sabesp estima que, em decorrência da rigidez tarifária, tenha perdido mais de 100 grandes clientes<sup>1</sup>. Entre os principais clientes perdidos para empresas de soluções locais estão: Saint-Gobain, Volkswagen, Indústrias Anhembí, Shopping Iguatemi, D&D Shopping, Hospital Santa Paula, Hospital Nossa Sra. Da Penha, Liotécnica, Multex Brasil, Osram, Coopers Tools, Jolitec, Cardinal Health.

O Gráfico 1, extraído do site do DAEE<sup>2</sup>, apresenta o crescimento de outorgas expedidas pelo DAEE nos últimos anos até 31/05/2008, evidenciando a migração de clientes para soluções locais em detrimento da utilização da rede pública de abastecimento.

<sup>1</sup> O impacto na receita anual da empresa é estimado em mais de R\$ 40 milhões.

<sup>2</sup> <http://www.dae.sp.gov.br/cgi-bin/Carrega.exe?arq=/outorgafiscalizacao/outorga06.htm>. Consulta feita em 1/1/2008.



Fonte: DAEE.

Acredita-se que os principais motivos que levam à evasão de clientes são o diferencial entre as tarifas da Sabesp e preço da água oferecido pelo mercado de fontes alternativas, a presença de condições necessárias para instalação de poços artesianos e a estrutura tarifária progressiva da Sabesp para o consumo de água.

Com base nas evidências apontadas, este estudo pretende estimar os limiares de consumo que fazem com que os clientes da Sabesp busquem outros tipos de fornecimento de água, que não pelo sistema Sabesp. Além desta breve introdução, este artigo está dividido em quatro seções. A segunda discute a base de dados utilizada e os principais critérios empreendidos na análise. A terceira seção revisa os modelos estimados e métodos de estimação para encontrar o nível de consumo em que os grandes clientes se tornam mais sensíveis a preço, seguida da seção com a apresentação dos resultados principais. A seção final contém as considerações finais mais importantes sobre o estudo.

## 2. Base de Dados

O trabalho se apóia em dados, oferecidos diretamente pela Sabesp, a respeito de seus clientes comerciais e industriais. A população de interesse do estudo é uma fração da população total de clientes industriais e comerciais da Sabesp.

Diante desse quadro, optou-se por contabilizar apenas a parcela da população que representava a população de interesse – os grandes clientes –, subdividindo-a em três tipos de clientes, de acordo com o volume de água mensal médio consumido:

- clientes de pequeno porte, cujo consumo médio mensal esteja acima de  $50\text{m}^3$ ;
- clientes de médio porte, cujo consumo médio esteja acima de  $300\text{m}^3$ ; e
- grandes clientes, cujo consumo médio esteja acima de  $1000\text{m}^3$ .

## 2.1 Organização da Base de Dados

Para cada um dos clientes foram fornecidas pela Sabesp as seguintes informações: registro da ligação de água (denominado RGI); códigos de localidade da ligação; tipo da ligação e de cobrança; endereço; status da ligação; ramo de atividade da empresa; quantidade de economias; consumo médio e faturado mensal dos últimos 61 meses; e status do consumo médio dos últimos 60 meses. O período total de referência da pesquisa, portanto, englobou 61 meses entre julho de 2003 e agosto de 2008.

O número do RGI serviu para definir unicamente cada cliente. O campo denominado *ramo*, que indicava o código do setor de atividade da ligação, foi utilizado para controlar efeitos específicos à atividade fim dos clientes, pois se espera que algumas atividades industriais induzam ao maior ou menor consumo de água.

## 2.2 Critérios para definição da Evasão dos Clientes Sabesp

Na sequência do estudo, pretendeu-se encontrar uma variável que captasse corretamente a evasão de grandes clientes do sistema Sabesp. Técnicos da Sabesp relatam que, na maior parte dos casos, somente por meio de inspeções de equipes de campo é possível identificar consumidores que se retiraram do sistema.

Primeiramente, optou-se por trabalhar sempre com as médias semestrais de consumo, ao invés do consumo mensal, e compará-las com a média de consumo do mesmo período no ano anterior, de maneira a evitar sazonalidades semestrais, como ocorre na indústria de bebidas, por exemplo.

Além disso, foi preciso determinar precisamente o ponto a partir do qual uma queda de consumo faturado deve ser classificada como “incomum”. Em trabalho anterior contratado pela Sabesp<sup>3</sup>, foram elaborados os critérios apresentados no Quadro 1 para classificar as tendências de consumo dos clientes.

Essa classificação, no entanto, se mostrou muito rígida ao definir que somente com consumo zero uma

Quadro 1 – Critério de classificação de cliente

Classificação	Comportamento do consumo
“Normais”	Apresentaram aumento ou redução de consumo a taxas inferiores a 40%
“Esfriados”	Apresentaram uma redução de consumo superior a 40%, mas consumo corrente é superior a zero
“Ex-clientes”	Apresentaram uma redução de consumo superior a 40% e consumo corrente é igual a zero
“Esporádicos”	Clientes não enquadrados em nenhuma das categorias acima

<sup>3</sup> “Perdas de Grandes Clientes” (Agosto de 2008), este trabalho utiliza ferramentas de datamining para definição de critérios.

ligação, ou RGI, poderia ser classificada como um “ex-cliente”. O mais comum entre os clientes que se retiraram do sistema Sabesp é uma queda drástica, mas ainda com a presença de um consumo residual não nulo.

Os clientes industriais e comerciais utilizam um grande volume de água que não requer alta qualidade, sendo essa parcela de consumo substituída. A água utilizada para fins mais nobres permanece sendo demandada junto à Sabesp. Assim, após opiniões de técnicos ligados à companhia, estabeleceu-se como critério para classificação de ex-cliente uma queda de consumo médio semestral igual ou superior a 90%. Tal queda de consumo pode ser fortemente explicada pela retirada do cliente do sistema Sabesp. O Quadro 2 resume os critérios utilizados neste estudo.

Quadro 2 - Critério definido para o estudo

Classificação	Comportamento do consumo
“Normais”	Apresentaram aumento ou redução de consumo a taxas inferiores a 40%
“Esfriados”	Apresentaram uma redução no consumo superior a 40%, mas inferior a 90%
“Ex-clientes”	Apresentaram uma redução no consumo superior a 90%
“Aquecidos”	Apresentaram aumento de consumo superior a 40%

### 3. METODOLOGIA

Para estimar o limiar do fornecimento de água que, dada a atual estrutura tarifária da Sabesp, leva os grandes clientes a buscarem fontes alternativas de fornecimento de água, ferramentas econométricas foram utilizadas.

A econometria é uma área da economia que se baseia no desenvolvimento de métodos estatísticos para estimar relações econômicas em uma população, em um país ou em empresas. Para estimar tais relações econômicas, utilizam-se dados observados (amostras) para inferir resultados sobre a população total, cuja totalidade de informações não é observada. Além de trabalhar com amostras, a econometria enfoca problemas inerentes à coleta e à análise de dados econômicos não experimentais (dados observados), devido à dificuldade de se observar dados experimentais em ciências humanas [Wooldridge, 2006].

A dependência entre as variáveis é medida por meio de uma função matemática que determina o tipo da relação entre as variáveis (linear, quadrática, exponencial, entre outras). Uma vez elaborado o modelo matemático das relações entre este mercado, a análise empírica, baseada na técnica econométrica, estima as relações econômicas com base em métodos estatísticos.

Dois métodos distintos foram utilizados para estimar o volume de fornecimento de água a partir do qual a probabilidade de evasão do cliente Sabesp au-

menta, sendo eles:

- 1) Modelo Probit com dados da comparação de 1 período;
- 2) Modelo Logit com efeitos fixos com dados da comparação de 7 períodos.

Ambos os modelos probit e logit são modelos de escolha binária, ou modelos de probabilidade. Segundo Wooldridge (2001), o princípio básico de modelos de probabilidade é o fato de a variável dependente assumir valor 1 para o caso de interesse, e 0 caso contrário. Em um modelo de regressão clássica, o interesse está no efeito marginal (que é o próprio coeficiente da regressão) de alguma variável  $x$  sobre a média condicional da variável dependente,  $y$ . Em modelos de probabilidade, ou modelos de escolha binária, o interesse se concentra em encontrar o efeito que uma variação em  $x$  provoca na probabilidade condicional de  $y|x^4$ . A seguir, ambos os métodos de estimação são apresentados em detalhes, com base em Wooldridge (2001) e Greene (2003).

#### 3.1 Modelo Probit

Uma vez detectada a evasão do cliente é possível estimar o volume de fornecimento de água a partir do qual a probabilidade de evasão do cliente Sabesp cresce. A amostra utilizada no modelo probit foi representativa para os municípios da Grande São Paulo atendidos pela Sabesp.

O modelo probit permite que se explique o efeito do volume de consumo e de outras variáveis intrínsecas dos clientes na probabilidade de os clientes deixarem o sistema de abastecimento da Sabesp e buscarem alternativas para o fornecimento de água. O volume de água consumido, dada a estrutura tarifária constante, ao longo dos cinco últimos anos, é a variável que permitirá ao cliente se ajustar no seu objetivo de minimizar custos. Assim, o volume médio faturado de água entrou na estimação no nível (em termos de  $m^3$ ) e ao quadrado, de maneira a captar a diferença de sensibilidade dos clientes à estrutura tarifária da Sabesp conforme aumenta o consumo médio dos mesmos.

A constatação de efeito estatisticamente significativo do nível de consumo na probabilidade de evasão do sistema pode ser evidência de elasticidade preço da demanda diferente entre os tipos de clientes da Sabesp (pequeno, médio, ou grande).

Para a construção do modelo, definiu-se:

$$Y = \begin{cases} 1, & \text{se o cliente saiu do sistema SABESP} \\ 0, & \text{caso contrário} \end{cases}$$

A variável  $Y$  assume o valor 1 com base nos critérios definidos na Seção 2.2 deste estudo. Assim, as variáveis de controle são: dummies para ramos de ati-

4 Lê-se “y dado x”.

vidades<sup>5</sup>; volume de água faturado mensalmente na Sabesp; e volume ao quadrado de água faturado mensalmente na Sabesp. Tais variáveis representadas pela matriz X, com número de colunas igual ao total de variáveis (no de dummies + constante + variáveis de volume de água), com N linhas representando as N observações da amostra.

A partir da comparação entre o consumo mensal médio faturado do primeiro semestre de 2008 e do primeiro semestre de 2004, calculou-se a variável Y. A probabilidade de evasão é dada por  $P(Y=1)$ , que é uma curva na forma de S e expressa uma relação não linear entre Y e as variáveis indicadoras.

A relação funcional representando esta curva é a função Probit. A função Probit está relacionada à função de distribuição da normal padronizada [Wooldridge, 2006]<sup>6</sup>. O modelo estatístico expressa a probabilidade do Y ser igual a 1, ou seja, o cliente saiu do sistema Sabesp, é dado por:

$$p = P[Z \leq \Phi(\beta_1 + \beta_2 \text{setor} + \beta_3(m^3) + \beta_4(m^3)^2)]$$

em que os betas são os parâmetros a serem estimados usando-se as N observações da amostra.

É importante ressaltar que, nestes tipos de modelos, o vetor de parâmetros  $\beta$  não tem a interpretação usual dos modelos de regressão tradicionais, não representando diretamente os efeitos marginais das variáveis explicativas sobre Y.

### 3.1.1 Efeitos Marginais

O impacto da variação de uma unidade em qualquer das variáveis explicativas, no caso em estudo: ramo de atividade;  $m^3$ ; e  $(m^3)^2$ , sobre a probabilidade de o cliente sair do sistema Sabesp pode ser escrito da seguinte maneira:

$$\frac{\partial p}{\partial x} = \frac{\partial \Phi(t)}{\partial t} \frac{\partial t}{\partial x} = \phi(t) \beta_k, \text{ com } k = 2, 3, 4 \text{ e}$$

$$t = \beta_1 + \beta_2 \text{setor} + \beta_3 m^3 + \beta_4 (m^3)^2$$

Assim, para se obter a probabilidade marginal de uma variável – ou seja, o efeito de uma variação de uma variável explicativa na probabilidade de evasão – torna-se necessário avaliar a função densidade de probabilidade em algum ponto  $t$  estimado. Uma vez escolhidos os níveis das variáveis explicativas para  $t$ , multiplica-se a função densidade de probabilidade neste ponto pelo coeficiente estimado da variável (beta) que se deseja obter o efeito marginal.

No caso do volume de consumo em  $m^3$ , a presença do termo quadrático (não-linear) altera o cálculo do

efeito marginal do volume médio de consumo na probabilidade do cliente sair. Assim, tem-se:

$$\frac{\partial p}{\partial m^3} = \frac{\partial \Phi(t)}{\partial t} \frac{\partial t}{\partial m^3} = \phi(t) [\beta_3 + 2\beta_4 m^3],$$

$$\text{em que } t = \beta_1 + \beta_2 \text{setor} + \beta_3 m^3 + \beta_4 (m^3)^2$$

Uma vez obtidos os coeficientes estimados do probit, calcula-se os efeitos marginais com base em alguma unidade das variáveis explicativas. É comum que sejam escolhidas medidas de posição, como média e mediana, para descrever os resultados dos efeitos marginais.

## 3.2 Modelo Logit

O modelo logit, assim como o probit, é um modelo de probabilidade, ou escolha binária. Para definir qual dos modelos é o mais adequado, Greene (2003) afirma que as duas distribuições (Normal e Logística) são muito parecidas, exceto nas caudas, onde a distribuição logística apresenta-se mais densa. Entretanto, as duas distribuições tendem a gerar resultados semelhantes.

Ainda de acordo com Greene (2003), embora existam razões práticas para favorecer o uso de um modelo em detrimento de outro por conveniências matemáticas, é muito difícil encontrar justificativas fundamentadas teoricamente para esta escolha, pois ambos os modelos são substitutos. Entretanto, o modelo logit apresenta vantagens sobre o modelo probit quando se tem base de dados em painel e, assim, a presença de efeitos específicos, pois o primeiro não exige que sejam feitas hipóteses sobre a relação do efeito específico não observado e as variáveis independentes do modelo.

Antes de se entrar nos detalhes do modelo logit com efeitos fixos, para facilitar o entendimento do modelo, decidiu-se acrescentar uma introdução sobre o Método de Efeitos Fixos. Os textos a seguir foram escritos com base em Wooldridge (2001).

### 3.2.1 Modelo de Efeitos Fixos

A utilização do Método de Efeitos Fixos, no caso deste estudo, deve-se à possibilidade de transformar a base de dados oferecida em um painel de dados<sup>7</sup>. Essa transformação oferece diversas vantagens na estimação, uma vez que elimina alguns problemas inerentes aos dados em cross-section e o problema da falta de controles na regressão que explica a evasão de clientes da Sabesp. Dados em painel requerem o emprego de modelos econométricos adequados. Um modelo básico de dados em painel pode ser representado por:

$$y_{it} = x'_{it} \beta + z'_i \alpha + \varepsilon_{it}$$

7 Dados em painel envolvem a dimensão de cross-section, neste caso os clientes da Sabesp, identificados pelos RGIs acompanhados ao longo do tempo, neste caso o consumo ao longo dos 61 meses de análise.

5 Dummies são variáveis binárias que têm valor unitário quando ocorre o comportamento que se pretende medir e valor nulo caso contrário.

6 Se Z for uma variável aleatória normal padronizada, a sua função densidade de probabilidade é dada por:

por:  $\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-0.5z^2}$ . A função Probit é, então, dada por:  $\Phi(z) = P(Z \leq z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-0.5z^2} dz$ .

O termo  $x'_{it}$  possui  $K$  regressores (matriz de variáveis explicativas do modelo), incluindo o vetor de 1's (para a constante). Na equação acima, o efeito individual, ou heterogeneidade, é representado por  $z'_i \alpha_i$ , onde  $z_i$  contém um conjunto de variáveis específicas para cada cliente, que podem ser observados ou não, e se mantém constantes ao longo do tempo. O modelo a ser empregado depende do tipo de heterogeneidade assumida.

Se  $z_i$  é não observado, mas correlacionado com  $x_{it}$ , então o estimador de Mínimos Quadrados Ordinários<sup>8</sup> é viesado e inconsistente<sup>9</sup>, como consequência de uma variável relevante omitida. Neste caso, o modelo pode ser formulado por:

$$y_{it} = x'_{it} \beta + \alpha_i + \varepsilon_{it}$$

Nesta equação,  $\alpha_i = z'_i \alpha$  engloba os efeitos não observados, e constantes no tempo, que afetam  $y_{it}$ . Esta abordagem é chamada de efeito fixo exatamente por assumir que  $\alpha_i$  é um termo constante específico de cada indivíduo que não varia ao longo do tempo.

Para cada  $i$ , pode-se tirar a média da equação:

$$\bar{y}_{it} = x'_{it} \beta + \alpha_i + \bar{\varepsilon}_{it}$$

Como o termo  $\alpha_i$  é constante no tempo, ele aparecerá em ambas as equações. Subtraindo a equação de sua média, temos:

$$y_{it} - \bar{y}_{it} = (x_{it} - \bar{x}_{it})' \beta + (\varepsilon_{it} - \bar{\varepsilon}_{it}) \text{ ou } \bar{y}_{it} = \bar{x}_{it} \beta + \bar{\varepsilon}_{it}$$

em que  $\bar{y}_{it} = y_{it} - \bar{y}_{it}$ ;  $\bar{x}_{it} = x_{it} - \bar{x}_{it}$  e  $\bar{\varepsilon}_{it} = \varepsilon_{it} - \bar{\varepsilon}_{it}$ .

É importante notar que o efeito individual,  $\alpha_i$ , desaparece, o que sugere que esta equação pode ser estimada por Mínimos Quadrados Ordinários. Os estimadores obtidos por este procedimento são chamados de estimadores de efeitos fixos ou estimadores *within*. O intuito da formulação por efeitos fixos é eliminar  $\alpha_i$ , ou seja, todos os efeitos não observados que são constantes no tempo<sup>10,11</sup>.

### 3.2.2 Modelo Logit com Efeitos Fixos

O modelo logit apresenta vantagens sobre o modelo probit quando há efeitos fixos, desta forma, utilizou-se tal modelagem na estimação do efeito do tamanho do cliente (volume) na probabilidade de sair do sistema de abastecimento da Sabesp. De acordo com Wooldridge (2002), o modelo pode ser definido por:

8 O Método dos Mínimos Quadrados Ordinários é um dos mais utilizados em análises multivariadas e consiste em uma técnica de otimização matemática que minimiza a soma dos quadrados das diferenças entre a curva ajustada e os dados.

9 Um parâmetro estimado é considerado viesado quando, em média, não acerta o valor verdadeiro.

10 Não se podem incluir variáveis constantes no tempo na estimação de Efeitos Fixos, pois estas seriam perfeitamente colineares à variável  $\alpha_i$  (não atenda hipótese de multicolinearidade perfeita).

11 O modelo de EF perde  $N$  graus de liberdade para estimar, pois é preciso calcular as médias para cada observação no tempo, assim o total de graus de liberdade é  $N.T - N = N(T-1)$ .

$$P(y_{it} = 1 / x_{it}, \alpha_i, 0 \leq n_i \leq 7) = \Lambda(x_{it} \beta + \alpha_i)$$

$$P(y_{it} = 0 / x_{it}, \alpha_i, 0 \leq n_i \leq 7) = 1 - \Lambda(x_{it} \beta + \alpha_i)$$

em que  $\Lambda(\cdot)$  é a função logística,  $x_{it}$  é a matriz de controles,  $\beta$  é o vetor de coeficientes,  $\alpha_i$  o efeito não observado e  $n_i = \sum_{t=1}^7 y_{it}$ .

Similarmente ao modelo probit, o volume médio faturado de água entrou nesta estimação no nível (em termos de  $m^3$ ) e ao quadrado, de maneira a captar a diferença de sensibilidade dos clientes à estrutura tarifária da Sabesp, conforme aumenta o consumo médio dos mesmos. Outros controles como: localização do RGI; ramo de atividade do cliente; entre outras características da ligação, tiveram seus impactos capturados pelo efeito específico não observado, o chamado 'efeito fixo'.

Para a construção deste modelo, definiu-se:

$$Y = \begin{cases} 1, & \text{se o cliente saiu do sistema SABESP} \\ 0, & \text{caso contrário} \end{cases}$$

A variável  $Y$  assume o valor 1 com base nos seguintes critérios estipulados na Seção 2.2. Os dados foram organizados em painel da seguinte maneira: definiu-se, em cada semestre, a variável dependente para o cliente, que mensurou a saída ou permanência do cliente na Sabesp; A variável de consumo médio (variável independente) considerada para cada período é igual ao consumo no semestre de comparação anterior, ou seja, se  $Y$  foi definido a partir da comparação do 1º semestre de 2005 com o 1º semestre de 2004, a média do consumo faturado utilizada como controle foi a do 1º semestre de 2004<sup>12</sup>.

O modelo logit com efeitos fixos depende da hipótese de exogeneidade estrita condicional a para a consistência de seus estimadores, logo:

$$P(y_{it} = 1 / x_i, \alpha_i) = P(y_{it} = 1 / x_{it}, \alpha_i) = \Lambda(x_{it} \beta + \alpha_i).$$

Neste modelo, o que se faz é encontrar a distribuição conjunta de  $y_i = (y_{i1}, \dots, y_{i7})'$  condicional a  $x_i, \alpha_i, n_i$ . Esta distribuição condicional não dependerá de  $\alpha_i$ , sendo a distribuição de  $y_i = (y_{i1}, \dots, y_{i7})'$  condicional a  $x_i, n_i$ <sup>13</sup>.

O método utilizado para estimar os coeficientes ( $\beta$ ) é o de Máxima Verossimilhança Condicional, que gera estimadores logit de efeitos fixos. A maneira de derivar os efeitos marginais é equivalente ao modelo probit, com exceção à suposição que deve ser feita para o efeito fixo.

No logit com efeitos fixos, para que se encontre o  $t$  estimado (combinação linear das variáveis explicativas + efeito fixo), que entra na função densidade de probabilidade para o cálculo dos efeitos marginais, deve-se supor algum valor para o efeito fixo ou considerá-lo zero. Esta segunda alternativa é mais prudente.

12 Para os valores extremos de  $n$ , 0 e 7, a distribuição condicional é dita não informativa para os parâmetros do modelo (betas), sendo estas observações excluídas da modelagem.

13  $P(y_{i1} = y_1, \dots, y_{i7} = y_7 | x_i, \alpha_i, n_i = n) = P(y_{i1} = y_1, \dots, y_{i7} = y_7 | x_i, \alpha_i) / P(n_i = n | x_i, \alpha_i)$

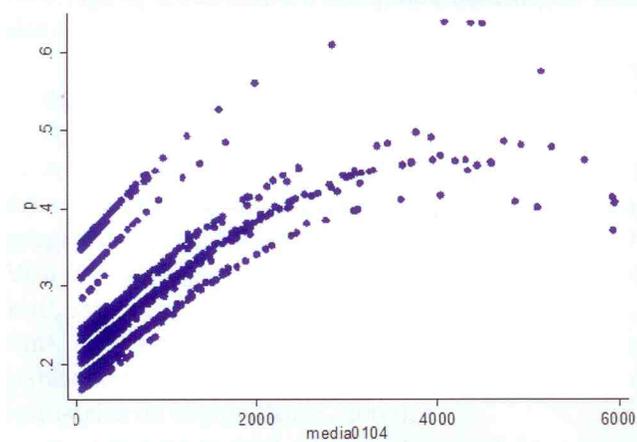
te de acordo com Wooldridge (2002) e menos arbitrária, logo foram considerados nulos os efeitos fixos.

## 4. RESULTADOS DAS ESTIMAÇÕES

### 4.1 Resultados do Probit

Como foi visto na seção anterior, as estimações do modelo probit utilizaram dados em *cross-section* das ligações (clientes) da Sabesp para a Grande São Paulo. A amostra consiste dos clientes cujo consumo faturado médio do 1º semestre de 2004 foi acima de 50m<sup>3</sup>14. O Gráfico 2 apresenta os resultados do modelo, sendo o consumo mensal médio inicial (*media0104*), ou seja, para o primeiro semestre de 2004, representado pelo eixo x e a probabilidade de evasão expressa no eixo y.

Gráfico 2: Probabilidade de evasão por nível de consumo



O Gráfico 2 indica que a probabilidade de evasão da Sabesp é crescente conforme aumenta o nível de consumo dos RGIs15; observou-se uma redução na probabilidade de evasão a partir do consumo médio 3000-4000m<sup>3</sup>, o que corrobora a importância do programa *Sabesp Soluções Ambientais*, que garante contratos com estrutura tarifária diferenciada para clientes cujo consumo seja igual ou superior a 3.000m<sup>3</sup>/mês.

As diferenças de patamar verificadas nos gráficos das probabilidades devem-se à inclusão de *dummies* para os ramos de atividade de cada cliente. As atividades dos clientes que apresentaram maiores probabilidade de evasão, em média, foram: construção civil; P&D; eletricidade e saneamento; extrativismo; telecom; e transporte. Em contrapartida, os setores de atividade com menores probabilidades médias de evasão foram: educação e saúde; atividades imobiliárias; financeiras; e petroquímica.

### 4.2 Resultados do Modelo Logit com Efeitos Fixos

Como foi visto na seção anterior, as estimações do modelo probit utilizaram dados em *cross-section* das ligações (clientes) da Sabesp para a Grande São Paulo. Foram estimados três modelos, utilizando os diversos

critérios de seleção de amostra (cortes no consumo médio)16. Os Gráficos 3 e 4 apresentam os resultados estimados, sendo o consumo mensal médio representado pelo eixo x e a probabilidade de evasão expressa no eixo y. A interpretação dos resultados está resumida na seqüência17.

Gráfico 3: Efeito marginal do consumo

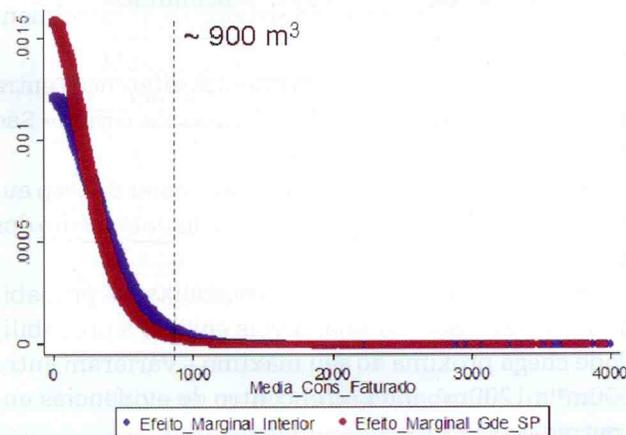
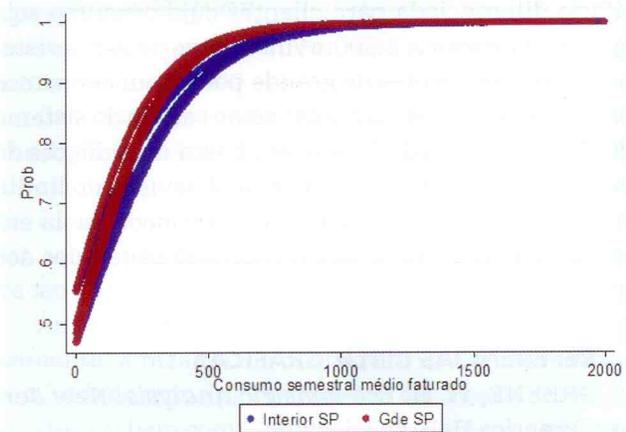


Gráfico 4: Probabilidade de evasão por nível de consumo



Pela análise dos resultados é possível observar que as curvas de efeito marginal são não-crescentes para todas as faixas de consumo estimadas. A contrapartida direta desse fato é que a probabilidade de evasão apresenta comportamento crescente com respeito ao nível de consumo faturado – corroborando a hipótese de que grandes clientes apresentam maior probabilidade de buscar fontes de água alternativas.

Pode-se notar que o efeito marginal converge para valores muito próximos de zero a partir de 900m<sup>3</sup> de consumo médio semestral faturado.

Deve-se ressaltar também que o efeito de localidade não se mostrou muito relevante, pois, independente do critério utilizado, as curvas de probabilidade de evasão estimadas para clientes da grande São Paulo estiveram sempre muito próximas daquelas estimadas para clientes do interior do estado.

## 5. CONSIDERAÇÕES FINAIS

A hipótese que se pretendia testar era a de que a atual estrutura tarifária da Sabesp leva grandes clientes a buscarem fontes alternativas de fornecimento de água. As principais conclusões que podem ser tiradas deste trabalho são:

- há evidências a favor da crença de que empresas de distintos setores de atividade possuem diferentes tendências de sair ou não do sistema de abastecimento da Sabesp;

- há poucas evidências de grandes diferenças entre as sensibilidades de grandes clientes da Grande São Paulo e do Interior de São Paulo;

- a probabilidade de evasão dos clientes Sabesp aumenta conforme cresce o nível de consumo médio dos mesmos;

- os níveis de consumo que estabilizam a probabilidade de evasão – ou seja, níveis em que a probabilidade chega próxima ao seu máximo – variaram entre 500m<sup>3</sup> e 1200m<sup>3</sup>, indo ao encontro de evidências encontradas pela Sabesp.

Com base neste estudo, pode-se concluir que o limite mínimo imposto pelo programa *Sabesp Soluções Ambientais*, que garante contratos com estrutura tarifária diferenciada para clientes cujo consumo seja igual ou superior a 3.000m<sup>3</sup>/mês, deveria ser revisto, uma vez que clientes de grande porte com consumos médios abaixo deste patamar estão saindo do sistema de abastecimento da Sabesp em busca de reduções de custos no mercado concorrencial. A revisão do limite mínimo deste programa acirraria a concorrência entre os fornecedores de água, trazendo benefícios aos utilitários deste insumo.

## REFERÊNCIAS BIBLIOGRÁFICAS

GREENE, W. H. *Econometric Analysis*. New Jersey: Prentice Hall, 5a ed, 2003.

SUPERINTENDÊNCIA DE MARKETING DA SABESP. *Perda de grandes clientes*. São Paulo: Estudo disponibilizado pela SABESP, 2008.

WOOLDRIDGE, J. M. *Econometrics Analysis of Cross Section and Panel Data*. Londres: The Massachusetts Institute of Technology Press, 1a ed., 2001.

WOOLDRIDGE, J. M. *Introdução à Econometria: Uma abordagem moderna*. São Paulo: Thomson Learning, 1a ed., 2006.