The Economics of Tobacco and Tobacco Taxation in the Philippines

Stella Luz A. Quimbo, Adele A. Casorla, Marina Miguel-Baquilod, Felipe M. Medalla, Xin Xu, Frank J. Chaloupka

Annex. Modeling the Demand for Cigarettes in the Philippines

In this annex, we provide more details on our estimation of cigarette demand in the Philippines. We focus on cigarettes because these are the predominant tobacco product in the Philippines, accounting for close to 97% of all tobacco-related expenditures recorded by the 2003 Family Income and Expenditure Survey (FIES). Because the price elasticity of demand is a key parameter in understanding changes in consumption given a tax increase, we estimate this for cigarettes using the 2003 FIES.

Data Sources and Limitations

Our primary data source is the 2003 FIES, a nationally representative survey on household incomes and expenditures conducted every three years. Overall, our demand analysis is subject to four important limitations. First, the complete data required for a demand analysis are available only at the household level; thus, our analysis is limited to household demand only. Another important limitation in the household demand analysis is the absence of household-level data on cigarette prices. Instead, we use province-wide average prices of cigarettes estimated by the National Statistics Office in 2003 using data from a nationwide survey of cigarette retail prices. Third, at the household level, tobacco consumption data are available only in expenditure terms. Our basic demand models use the number of units consumed which is estimated using reported expenditure and province-wide average prices. Fourth, our data set does not have direct measures of supply-side variables. Thus, when our econometric model requires supply-side information (i.e., instrumental variables for price), we resort to the use of geographic dummy variables. The justification for our choice of instruments is elaborated in more detail below.

We conduct this analysis using 2003 data, the most recent data available at the time we initiated this study. Moreover, 2003 is also an appropriate base year for the simulations since the relevant tax law was revised in 2004 and the stipulated new excise taxes were applied beginning 2005. In addition, we gain useful insights, especially on parameter estimates and validity of assumptions, by comparing 2005 simulated values with actual figures. We perform all statistical evaluations using STATA version 10.

For the cross-sectional analysis of household demand, we estimate the following basic model:

$$Q_i = \beta_0 + \beta_1 P_i + \beta_2 Y_i + \sum_k \gamma_k X_{ki} + u_i \qquad (1)$$

where:

Q_i : quantity of cigarettes consumed by each household;
P_i : average price of cigarettes in household i's province;
Y_i : annual household income;
X_i : a vector of other household characteristics affecting cigarette consumption; and
$u_i \sim IID(0,\sigma^2)$, a normally distributed error term.

In estimating demand equations, prices are arguably endogenous—prices and quantities consumed (and supplied) are jointly determined. As applied to our model, the price variable will be correlated with the error term in equation (1). In the presence of "endogenous regressors," estimating a demand equation via ordinary least squares (OLS) will result in biased estimates.

At the same time, we emphasize that in our model, our dependent variable is household consumption per capita and the price variable is a provincial average. In this case, each household is too small to influence the provincial price. Thus, each household is a price-taker, which means that from a behavioral point of view, the price is arguably exogenous. Had our dependent variable been provincial consumption, the endogeneity argument would have been applicable.

In addition, the average provincial price in Equation (1) is actually a proxy for the true price faced by the household. The variable P_i is thus subject to measurement errors. To illustrate, a rich household in the Ilocos region faces a relatively low average provincial price but may in fact consume premium-brand cigarettes. On the other hand, an urban poor household in Metro Manila faces relatively higher average provincial prices but in all likelihood would prefer to smoke cheaper brands. Such measurement errors may result in the endogeneity of P_i.

Thus, although economic theory suggests that price endogeneity may not be a problem in Equation (1), measurement errors could cause a correlation between P_i and the error term. We thus formally test for the endogeneity of by performing the Hausman test where P_i is first regressed on all exogenous variables in Equation (1) including the instrumental variables (region fixed effects). We then estimate a second equation where Q_i is regressed on the predicted price and residuals. The significance of the predicted residuals will indicate the endogeneity of P_i.[1]

Once the endogeneity of P_i is ascertained, we proceed by estimating Equation (1) via two-stage least squares (2SLS). In the first stage, P_i is regressed on all exogenous variables in the system, including identifying instruments, namely, variables from the supply equation. We use regional fixed effects (i.e., dummy variables for each region) which proxy for production and marketing cost differences as identifying instruments. In the second stage, Equation (1) is estimated via OLS, using the predicted values of P_i from stage 1 instead of actual values of P_i. This methodology will yield consistent estimates.

We log-transform all continuous variables in Equation (1). Thus, the slope coefficient of P_i can be interpreted as the price elasticity of demand, ε_i, where:

$$\varepsilon_i = \frac{\partial Q_i}{\partial P_i}\frac{P_i}{Q_i} = \frac{\partial \ln Q_i}{\partial \ln P_i} = \beta_1 \qquad (2)$$

To test whether ε_i varies with income, we run sub-sample regressions for four groups consisting of: (i) the first three income deciles, (ii) the fourth to sixth income deciles, (iii) the seventh to ninth income deciles, and (iv) the tenth income decile. Similarly, the coefficient of the income variable, β_2, can be interpreted as the income elasticity of demand:

$$\varepsilon_i^{y} = \frac{\partial Q_i}{\partial Y_i}\frac{Y_i}{Q_i} = \frac{\partial \ln Q_i}{\partial \ln Y_i} = \beta_2 \qquad (3)$$

We choose to examine four income groups so that ε can be estimated for each of the four cigarette brand types for which specific taxes are stipulated. (As described above, cigarettes are classified by price.) In determining the composition of these income groups, we first examine the mean income levels of each income decile. Because income differences across deciles are largest between the last two deciles, the tenth decile is designated as one income group. The remaining nine income deciles are equally divided into three groups. We perform Chow tests to determine the validity of these sub-sample estimates.

A richer demand analysis might generate price elasticities for a particular brand and cross-price elasticities for other brands. The latter is particularly useful in the Philippine context because specific taxes vary by price of tobacco products. Moreover, there is evidence on the intra- and inter-brand switching process as a result of cigarette price increases, taking into account the unique retail practice in the Philippines of selling cigarettes by stick.[2] That is, given an increase in the price of, say, Brand X, smokers may switch to the same number of packs of a cheaper Brand Y or a fewer number of sticks of a more expensive Brand Z. These instances of brand permutations are possible only because cigarettes can be bought by the stick in the Philippines.

Due to the limitations of the price and consumption data, we provide estimates of an average elasticity which measures the responsiveness of demand to changes in the average price of all brands available to households. Thus, if our model yields an estimated price elasticity of –0.8, this means that given a 10% increase in the average price, demand declines by 8%. To illustrate: if there are two brands available (Brands A and B) and prices of both brands change, 10% is the average change in the prices of both brands. Furthermore, if each brand has a 50% market share, a 10% increase in average price could be the result of a 15% increase in Brand A and a 5% increase in Brand B. We attempt, however, to incorporate cross-price elasticities in the simulations by introducing assumptions on intra-brand switching behavior.

The vector "X" contains correlates of household income, particularly characteristics of the household head that influence income. We include these variables in the model to isolate these effects from that of income on cigarette consumption. These variables are age, sex, completion status of college education, and employment status of the household head. Inclusion of these income correlates would thus imply that the slope coefficient of the income variable can be interpreted as the marginal effect of income on cigarette consumption—net or independent of the effects of the household head's age, sex, education, and employment on Q_i. We also include a dummy variable for having expenditure on any form of insurance to control for a household's risk attitude. Households having any insurance expenditure could be more risk-averse than the rest, and therefore, less likely to engage in risky behavior such as smoking.

Data used for this analysis are from the 2003 FIES, a nationwide survey of households conducted by the National Statistics Office. The FIES primarily collects information on income, expenditure, and other variables affecting income and expenditure including household size, ages of household members, and educational attainment and employment status of the household head. The questionnaire lists tobacco products (cigars, cigarettes, and others) as one of the expenditure categories. The FIES uses a multi-stage sampling scheme which involves the selection of sample barangays (villages) in the first stage, the selection of sample enumeration areas in the second stage, and the selection of sample households in the third stage in each stratum for every domain.

Table A1 shows descriptive statistics for the sample of households covered by the 2003 FIES. About 70% of households reported having expenditures on tobacco products, comprising about 1.6% of their total income.

Households without tobacco expenditures differ significantly from those with tobacco expenditures. The latter have household heads who are slightly younger, more likely to be male and employed, but less likely to be college graduates. They are also less likely to have any form of insurance coverage (Table A2).

Table A.1: Mean Household Income and Expenditures on Tobacco Products, 2003

Variable	Mean	N
Annual household income (in pesos*)	137,758.4	42,094
Proportion of households with tobacco expenditures	0.69	42,094
Household expenditures on tobacco products (in pesos*)	1,931.9	28,836
Household expenditures on cigarettes (in pesos*)	1,865.4	28,836
Household expenditures on cigars (in pesos*)	9.5	28,836
Household expenditures on other tobacco products (in pesos*)	57.1	28,836

Source of data: 2003 FIES.
**Exchange rate in 2003: P54 = USD1.*

Table A.2: Socioeconomic Characteristics of the Households, 2003

	Mean*		
Socioeconomic Characteristic	**Households with Tobacco Expenditures**	**Households without Tobacco Expenditures**	**p-value**
Age of household head	45.48	47.65	0.00
Household head is:			
Male	0.89	0.75	0.00
College graduate	0.07	0.14	0.00
Employed	0.89	0.83	0.00
Proportion of households with insurance (any form)	0.28	0.34	0.00

*Note: * To compare means, the t-test was used.*
Source of data: 2003 FIES.

Table A3 presents the average selling (i.e., gross retail) prices of available cigarette brands in retail outlets. These are regional averages computed from prices in 56 provinces. These price data were collected via multi-stage sampling, where municipalities were chosen in the first stage and retail outlets in the second stage. For both stages, sampling was purposive, i.e., sampling units were selected on the basis of criteria such as the size of municipality and the accessibility of stores.

As seen from Table A3, average selling prices vary widely across the country, ranging from 7.4 to 23.1 pesos per pack, despite having a single tax scheme nationwide. Regional averages likely reflect both demand- and supply-side factors. On the demand side, for example, the mean and variability of prices could be a function of household incomes: regions with richer households will prefer to smoke more expensive brands, and regions with an unequal income distribution will have a wider range of brands available in the market. Availability of substitutes is another demand-side variable explaining price variations. For example, cigarette prices could be low in the Ilocos Region where tobacco leaves are primarily sourced, thus explaining the relative popularity of hand-rolled tobacco leaves in this region. On the supply-side, differences in prices could be explained by differences in transportation and distribution-related costs, including regional wages. Based on our interviews, cigarette manufacturers typically pass on their products to regional distributors at a uniform wholesale price. Supply-side differences in retail prices are thus largely explained by distribution costs.

The number of packs consumed by a household is generated by dividing the reported expenditures by the average prices reported in Table A4. Overall, each smoker household consumes an average of 175 packs of cigarettes per year. As evident in Table A4, the average number of packs consumed by households varies substantially across income deciles, with the poorest households consuming under 80 packs per year and the richest households consuming close to 300 packs per year. We note again that because prices are provincial averages rather than what households actually pay, the range of consumption could be overestimated or underestimated based on the brand preferences of rich versus poor households.

We further note that the variable "number of packs consumed," while it may be subject to measurement errors, is on the left-hand side of Equation (3) and will therefore not cause biases in the estimates of the regression coefficients.[1] This is in contrast to the case of P_i, which may also be subject to measurement errors (i.e., differences between actual prices paid by households and the provincial average); however, unlike Q_i, P_i is an explanatory variable in the demand models. Thus, measurement errors in may have implications on the consistency of regression estimates.

Tables A5 and A6 report the results of the Hausman test of endogeneity of P_i. In Stage 1, all region dummy variables are statistically significant ($p<0.001$). More importantly, Stage 2 of the Hausman test indicates that the predicted residuals are statistically significant ($p<0.001$). We therefore conclude that P_i is endogenous and that OLS estimates are biased and inconsistent.

Table A7 presents the results of the Chow tests. We perform pair-wise tests for each of the four sub-samples, and in each test, we compare a sub-sample with the rest of the sample (e.g., first to third deciles vs.

Table A.3: Average Selling Prices of Cigarette, by Province, 2003

Region	Average Selling Price of Cigarettes (per pack, in pesos)
Ilocos Region	8.2
Cagayan Valley	10.6
Central Luzon	12.2
Bicol Region	10.6
Western Visayas	12.4
Central Visayas	7.4
Eastern Visayas	14.6
Western Mindanao	18.2
Northern Mindanao	9.1
Southern Mindanao	23.1
Central Mindanao	17.8
National Capital Region	19.3
Cordillera Administrative Region	11.4
Autonomous Region of Muslim Mindanao	21.3
Caraga	12.5
CALABARZON	10.7
MIMAROPA	7.6
All	13.0

Source: NSO, 2003.

Table A.4: Average Annual Cigarette Consumption of Households, by Income Decile, 2003

Income Decile	Average Number of Packs Consumed By Households Per Year
1	76.5
2	107.5
3	129.6
4	145.4
5	173.1
6	200.0
7	213.7
8	224.3
9	260.6
10	292.6

Source: 2003 FIES

Table A.5: Hausman Test of Endogeneity: Stage 1 (dependent variable: ln(average cigarette price))

Independent Variables	Coefficient	P>t
ln(annual household income)	-0.001	0.874
ln(age of household head)	0.008	0.231
Household head is male	0.010	0.126
Household head is a college graduate	-0.003	0.668
Household head has a job	-0.003	0.669
Household has any form of insurance	0.016	0.002
Ilocos Region	0.205	0.000
Cagayan Valley	0.446	0.000
Central Luzon	0.537	0.000
Bicol Region	0.459	0.000
Western Visayas	0.568	0.000
Central Visayas	0.105	0.000
Eastern Visayas	0.749	0.000
Western Mindanao	0.983	0.000
Northern Mindanao	0.243	0.000
Southern Mindanao	1.223	0.000
Central Mindanao	0.903	0.000
National Capital Region	1.052	0.000
Cordillera Administrative Region	0.526	0.000
Autonomous Region of Muslim Mindanao	1.125	0.000
Caraga	0.584	0.000
CALABARZON	0.446	0.000
Constant	1.868	0.000
R squared	0.599	
Number of observations	21,942	

**Reference for regions: MIMAROPA.*

Table A.6: Hausman Test of Endogeneity: Stage 2 (dependent variable: ln(number of packs smoked by household))

Independent Variables	Coefficient	P>t
Predicted price	–0.881	0.000
Predicted residuals	**–0.874**	**0.000**
Constant	6.593	0.000
R squared	0.061	
Number of observations	21,942	

Table A.7: Chow Tests

Null hypothesis	Income Deciles			
nth income decile is no different from the rest	**1st–3rd**	**4th–6th**	**7th–9th**	**10th**
Computed F	26.22	8.37	5.25	11.64

**Note: The number of parameters estimated is 8. The critical F value for degrees of freedom (8,∞) and α=0.01 is 2.51.*

fourth to tenth deciles). The computed F statistics all exceed the critical F value, and we thus conclude that the four income sub-samples are different.

The full set of two-stage least squares estimates of cigarette demand are presented in Table A8.

All regressors are statistically significant ($p<0.05$) and have the expected signs in the pooled regression. Across sub-samples, regressions results are stable, except for household head's age and employment status as well as household insurance status, which are not statistically significant for some sub-samples. The age and insurance variables also do not have the same signs across sub-samples. Generally speaking, however, the results appear robust across sub-samples. R-squared values are also within the acceptable range for cross-section data.

Household demand for cigarettes is in general price-inelastic ($\varepsilon_i = -0.87$). Price elasticity, in absolute terms, declines with income suggesting that poorer households are more responsive to changes in price than richer ones. In fact, for the first three income deciles, demand appears to be elastic ($\varepsilon_i = -1.09$). Income elasticity is estimated at 0.66 for the full sample and ranges from 0.24 to 1.03 for the four income groups. Income elasticity declines monotonically with income. That is, poorer households' consumption is more responsive to changes in income than richer ones.

We find that other income correlates are significant predictors of cigarette consumption. Male-headed households are more likely to be smoker households ($p<0.001$). A household head's college education predicts lower consumption ($p<0.001$). For all income sub-samples except the 7th to 9th income deciles, employed household heads are more likely to have smokers in their households ($p<0.10$). Younger household heads predict cigarette consumption among the poorest households, but not for the rest of the income groups ($p<0.05$). Finally, we find that having any form of insurance predicts lower cigarette

Table A.8: Two-Stage Least Squares Estimates of Household Consumption of Cigarettes

	Income Deciles									
Independent Variables	ALL		1st - 3rd		4th - 6th		7th-9th		10th	
	Coef.	P-value	Coef.	P-value	Coef.	P-value	Coef.	P-value	Coef.	P-value
ln(average price)	-0.87	0.00	-1.09	0.00	-0.80	0.00	-0.74	0.00	-0.52	0.00
ln(annual household income)	0.66	0.00	1.03	0.00	0.87	0.00	0.40	0.00	0.24	0.03
ln(age of household head)	-0.07	0.04	-0.14	0.01	0.07	0.22	0.06	0.41	-0.09	0.60
Household head is male	0.60	0.00	0.79	0.00	0.53	0.00	0.48	0.00	0.44	0.00
Household head is a college graduate	-0.58	0.00	-0.59	0.01	-0.48	0.00	-0.51	0.00	-0.39	0.00
Household head has a job	0.22	0.00	0.45	0.00	0.18	0.00	0.06	0.33	0.21	0.06
Household has any form of insurance	-0.24	0.00	-0.08	0.40	-0.21	0.00	-0.25	0.00	0.06	0.57
Constant	-1.27	0.00	-4.86	0.00	-4.20	0.00	1.31	0.15	3.00	0.05
R squared	0.15		0.16		0.10		0.09		0.06	
Number of observations	21942		7132		7424		5865		1521	

consumption for the intermediate income deciles (i.e., 4th to 9th), consistent with the hypothesis that risk-averse households have lower consumption of goods that may have harmful effects.

We note that for all 2SLS regressions, we include only those households with positive cigarette expenditures. To address the possible concern that selecting only smoker households in the regressions might introduce a sample selection bias, we estimate an additional model that tests for the presence and corrects for such bias. The results of the model indicate that we can ignore this potential source of bias or that the results presented in Table A8 are likely not subject to a sample selection bias, despite including only smoker households in the analysis.

This model, referred to as a Heckman model with sample selection, is formally defined as follows. Let (Q_i, Z_i, P_i, Y_i, X_i, u_i, v_i) be a random draw from the population, where Z_i is an indicator variable for smoker households, X is a vector of characteristics that measure the utility from smoking, and v is an error term. Recalling that

$$Q_i = \beta_0 + \beta_1 P_i + \beta_3 Y_i + \beta_4 P_i Y_i + u_i$$

we further define

$$Z_i = 1 \text{ if } X_i\delta + v > 0$$

and note that Q_i is observed only when $Z_i = 1$. We estimate

$$E(Q_i \mid P_i, Y_i, Z_i = 1) = \beta_0 + \beta_1 P_i + \beta_2 Y_i + \beta_3 P_i Y_i + \gamma\lambda(X_i\delta)$$

where λ is the inverse Mills ratio. As δ is unknown, we first obtain a probit estimate of the following model:

$$\Pr(Z_i = 1 \mid X_i) = X_i\delta$$

using the full sample of households. We then obtain $\hat{\lambda}$ and proceed to the second stage where we use OLS on the sample of smoker households to estimate:

$$E(Q_i \mid P_i, Y_i, Z_i = 1) = \beta_0 + \beta_1 P_i + \beta_2 Y_i + \beta_3 P_i Y_i + \gamma \lambda(\hat{X_i \delta}) \; .$$

This second-stage estimates are consistent and asymptotically normal. Put differently, estimating Q_i directly using OLS while ignoring λ could result in a bias that is equivalent to an omitted variable bias.[3] This also forms the basis for a test of sample selection bias: A t-test performed on γ will indicate whether we can reject the null hypothesis that a sample selection bias is present.

The second-stage results of our Heckman model are reported in Table A9.

The statistical insignificance of the inverse Mills ratio suggests that we can reject the null hypothesis that a sample selection bias is present.

Finally, we recognize that there are limitations resulting from deflating reported household cigarettes expenditures by the provincial average price and using this as the dependent variable of the demand models. We thus re-estimate all models presented in Table A8 using reported cigarette expenditures as dependent variable. We then obtain the price elasticity of expenditures and then extract from this the price elasticity of demand (measured in number of packs smoked). We find that this indirect method of estimating ε_i yields results that are very similar to those produced by the direct method described and employed above. (For space considerations, we do not report these alternative estimates here, but these are available from the authors upon request.)

Table A.9: Heckman Model of Sample Selection, Stage 2 (dependent variable: In (number of packs smoked by household))

Independent Variables	Coefficient	P>t
In(average price)	-1.02	0.00
In(annual household income)	0.57	0.00
In(age of household head)	0.05	0.82
Household head is male	-0.14	0.92
Household head is a college graduate	0.04	0.97
Household head has a job	0.05	0.85
Household has any form of insurance	-0.05	0.88
Inverse Mills	**-2.28**	**0.57**
Constant	1.63	0.75
R squared	0.15	
Number of observations	21942	

Endnotes for Annex

1 Gujarati, D. Basic Econometrics. Fourth Edition. New York: McGraw-Hill/Irwin. 2003.

2 Chen KJ. The Filipino "tingui" retailing approach and cigarette price increase. Pricing Strategy and Practice. 1997;5(4), 148-155.

3 Wooldridge, J. Econometric Analysis of Cross Section and Panel Data. MIT Press. 2002