Estimating Demand for Tobacco

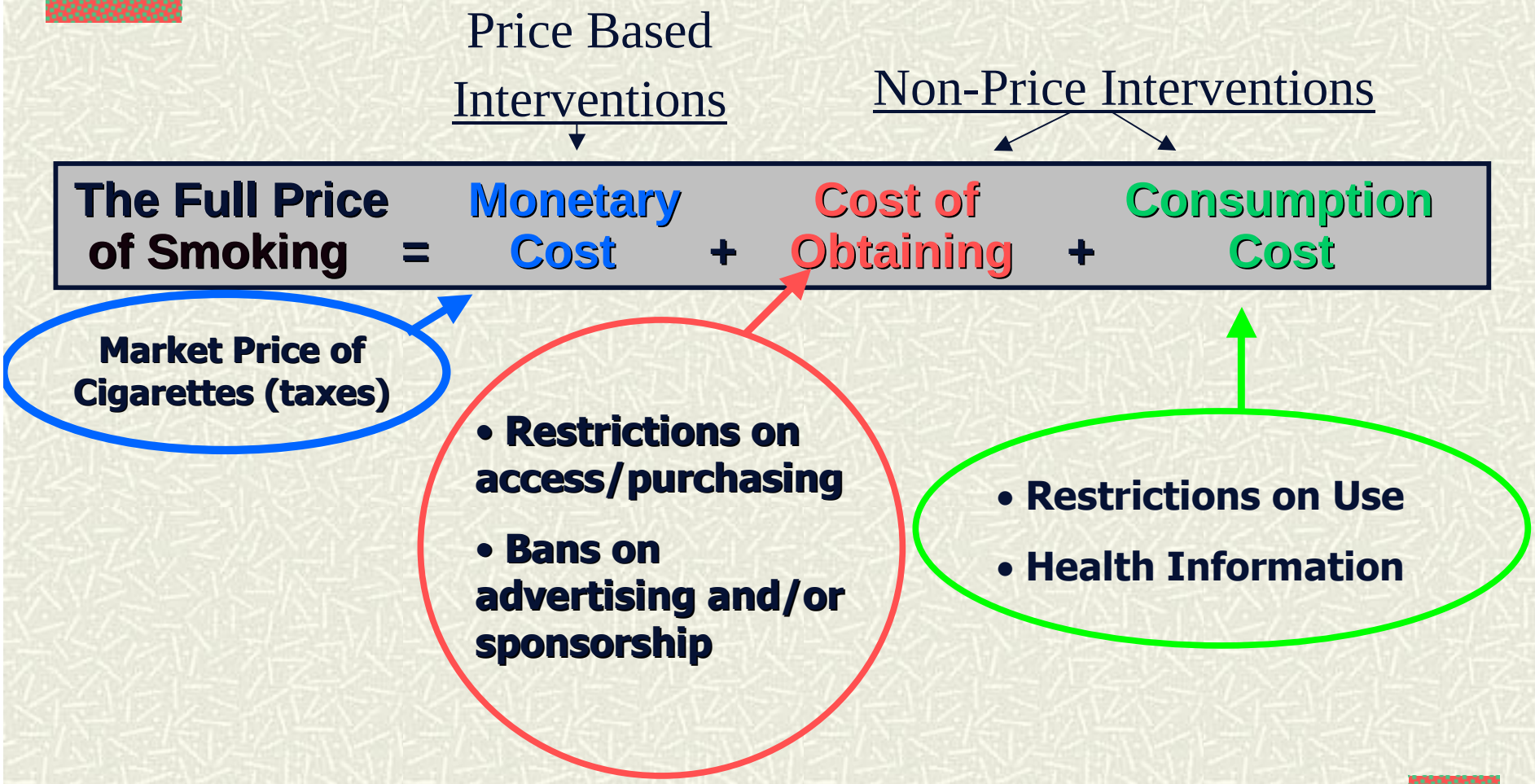
Hana Ross, Ph.D.

Christina Ciecierski, Ph.D.

Analyzing Tobacco Demand

- **Why:** - understand economic determinants of tobacco consumption and how we can manipulate them
  - assess consequences of this manipulation for the economy
  - draft effective public policies
- **Economic theory gives us a framework:**
  - **theory of demand assumes an individual utility function**
  - **utility from smoking = # cigarettes, utility from other goods, and individual tastes**
  - **utility maximization subject to budget constraint (price of cigarettes, income, prices of all other goods)**

Economic Interventions



What we will learn

- Does price/tax affect tobacco consumption, and to what extend? (price/tax elasticities)
- Does smoking depend on personal income? (income elasticity)
- Can public policies change smoking behavior?
- Can advertising encourage tobacco use?
- Will globalization, trade liberalization, and/or privatization change the demand for tobacco products?

Skills we will learn

- How to find and collect data
- How to check data for quality
- How to prepare data for an analysis
- How to conduct an economic analysis
- How to interpret and communicate the results

Collecting Data for Analysis

- Data describes: 1) Consumer market for tobacco
  2) Tobacco industry
  3) Government tobacco legislation
- The coverage and reliability of these data vary between countries
- Data sources: government (Central Statistical Office, Relevant Ministries), NGOs, international organizations (UN, WHO, WB, USDA) commercial agencies, tobacco industry documents, health care organizations, religious groups, etc.
- Type of data: consumption, sale, manufacturing, export, import, smuggling, tax (level, structure, revenue), other regulations, advertising, labeling, cig. content, health knowledge

Aggregate Data Analysis

- **Questions it can answer**
  - overall impact of average retail price on total demand
  - overall impact of disposable income on total demand
  - overall impact of public policies, health knowledge, etc. on total demand

Aggregate Data Analysis

- **Weaknesses**

- simultaneity problem (perfectly elasticity of supply must be assumed)
- cross-border shopping is masked
- smokers' characteristics and preferences (e.g. for low tar/nicotine products) are hidden
- change in # of smokers from change in # of cigarettes smoked by smokers cannot be distinguished
- causal link between tobacco control policies and consumption may be disguised

Individual Level Data Analysis

- **Questions it can answer**

  - Impact of cigarette prices, income, prices of other goods, health knowledge, and socio-demographics on individual cigarette demand

  - Impact of cigarette taxes and other tobacco controls on individual cigarette demand

Individual Level Data Analysis

Strengths over Aggregate Analysis

1. Simultaneity bias is less likely (an individual cannot affect market price)
2. Individual characteristics are less correlated with price and policy measures
3. Can evaluate the impact on both smoking rate and # cigarettes to consumed
4. Different population subgroups (age, gender, education, etc) can be evaluated

Individual Level Data Analysis

Weaknesses

1. Cross-border shopping is masked

   - may lead to overestimating of price elasticity

   Possible solutions: limit the sample, create a control variable (price difference, proximity to border)

2. Ecological bias - omitted variables correlated with those included in a model (e.g. anti-smoking sentiment)

   - may bias estimates

2. Self-reporting bias – assumption: the bias is proportional to true consumption, results not affected

Aggregate Analysis

- **Type of the data**
  - aggregate time-series (most common)
  - aggregate cross-sectional data (small variation in prices)
  - pooled time-series
  - panel cross-sectional data
  - longitudinal data

Aggregate Analysis
Minimum Data Required

- Quantities of tobacco consumed or sold within the particular time periods (year, quarter, month)
- Price of tobacco products (e.g. total expenditures on tobacco products, average price per pack, price index) - can be proxied by tobacco taxes
- Income (e.g. aggregate personal disposable income, national product per capita, etc)

Aggregate Analysis
Other Data of Interest (1)

- Price of other tobacco products (to assess substitution/complimentarity)
- Presence of smuggling
  - prices in neighboring states
  - % of population living near boarders from which smuggling can occur
- Health information
  - warning labels
  - market share of low-tar or filter cigarettes
  - presence and timing of information shocks and counter-advertising campaigns

Aggregate Analysis
Other Data of Interest (2)

- Advertising and promotion expenditures
  - consider all media and forms of promotion
  - preferred are monthly or quarterly data and cross sectional data (higher variation)
  - solution to lack of data:
    - use one media to represent relative ratio between tobacco and other advertising
    - control for timing and range of advertising bans in media (account for offsetting increases of advertising in other media)

Aggregate Analysis
Other Data of Interest (3)

- Smoking restrictions (place and timing)
- Degree of privatization
- Degree of trade liberalization
  - year of entry of foreign companies
  - market share of imported cigarettes
  - trade or total import as a share of GDP
  - existence and level of import duty
- Degree of openness to foreign direct investment
  - # and severity of restrictions
  - exchange rate distortion

Aggregate Analysis
Preparing Data (exercise)

- Evaluate and clean the data (methods, data reliability, # observations, # missing)
- Transport the data to the right format
- Summarize the data and check their quality

  - plot quantity variables against time (outliers)

  - shape of the plot will determine the functional form for the demand equation

Aggregate Analysis Data Coding (1)

- Per capita/household Cigarette Consumption = (Production + Imports – Exports) ÷ (Population 15+)
- Average cigarette price per pack
  - weight prices of product sub-types and types of transaction according to market share, or
  - total expenditures ÷ # packs of cigarettes consumed
  - deflate: nominal price ÷ CPI × 100

Aggregate Analysis Data Coding (2)

- Per capita personal disposable income =
  (aggregate personal disposable income) ÷ (population 15+)
  - deflate: nominal income ÷ CPI × 100
- Dummy for health information shocks and counter-advertising campaigns in each time period
- Expenditure on counter-advertising campaigns
- Dummy for earmarking cigarette tax for health education, etc.

Aggregate Analysis
Data Coding (3)

- Ratio of tobacco related to total advertising & promotional expenditures
  - lagged advertising variable (“stock” effect)
  - separate advertising from other promotional activities (if possible)
- Dummy for introducing advertising ban
- Adverting ban index = the number of media banned

Aggregate Analysis Data Coding (4)

- Index of smoking restrictions
  Scale:
  0–none; 0.25-public places up to 3, not restaurants; 0.5-more than 3 public places, not restaurants; 0.75-in restaurants, not work place; 1-work place
- Dummy for introducing new smoking restrictions
- Enforcement

Aggregate Analysis Data Coding (5)

- Dummy for entry of foreign companies to local market
- Index for degree of openness to foreign direct investment (IMF's Annual Report)

Scale:

0-high; 1-moderate; 2-low or none restrictions

- Degree of openness to foreign direct investment:

Black Market Premium = (Parallel exchange rate/Official exchange rate - 1) * 100

Aggregate Analysis Data Coding (6)

- Degree of trade liberalization:
  - (Exports + Imports of goods and services) ÷ (Gross Domestic Product)
  - (Imports of goods and services) ÷ (Gross Domestic Product)
- Degree of privatization:
  - market share of private versus government companies

Test coded independent variables for multicollinearity

Econometric Models

- Conventional models (static)

  - demand is affected only by current events/characteristics

- Addictive models (dynamic)

  - Myopic model

    - demand is affected by past consumption

  - Rational addition model

    - demand is affected by past and future consumption

Functional Form

- Linear
  - plot against time resembles a line
- Semi-log (log-lin or lin-log)
  - plot against time resembles a logarithmic curve and a line (either dep. and indep. variables)
- Double-log (log-log)
  - plot against time resembles a logarithmic curve
  - coefficients represent elasticities
  - assumes constant price elasticity over time

Aggregate Analysis Demand Equation

- Dependent variable
  - cigarette sale/consumption (per capita or aggregate)
  - tobacco weight (per capita or aggregate)
- Independent variables (min required)
  - deflated price per unit
  - deflated personal disposable income (per capita or aggregate)

Aggregate Analysis Demand Equation (cont.)

- Independent variables (examples)
  - dummy for health information shocks and counter-advertising
  - ratio of tobacco specific to total advertising & promotional expenditures
  - Index of smoking restrictions

Specification error (bias unknown) v. error in variable (bias towards 0)

Examples of Aggregate Econometric Models (1)

- Annual aggregate time-series data
  - usually small number of observations, therefore use limited # of indep. vars
  - usually high collinearity between indep. variables

Example of conventional linear demand:

$$Q_t = b_0 + b_1 P_t + b_2 Y_t + b_3 T_t + b_4 SR_t + b_5 D_m + \varepsilon_t$$

Examples of Aggregate Econometric Models (2)

- Quarterly (or monthly) time-series data
  - control for seasonal variation if sales data is calculated from excise tax revenue
  - suitable for assessing the effect of advertising & promotions

Example of conventional linear demand:

$$Q_t = b_0 + b_1P_t + b_2Y_t + b_3T_t + b_4SR_t + b_5D_m + b_6D_{q2} + b_7D_{q4} + b_8AD_t + \varepsilon_t$$

Performing Aggregate Econometric Analysis

- Document every step (any changes in data or use of a different estimation technique)

1. Ordinary Least Square (OLS)

regress y x

2. Test price for exogeneity (Hausman's test).

regress y p x ; vce

regress p x ; predict p'

regress y p' x ; vce

m = $(b' - b)^2 / [var(b') - var(b)]$

m = $\chi^2(2)$; $\mathbf{H_0}$ = price is exogenous

Performing Aggregate Econometric Analysis (cont.)

3. Find instrumental variable for endogenous price

   - must be correlated with price but cannot affect consumption

   - examples: excise tax, lagged price/tax, real cost of tobacco, average salary cost of cigarette manufacturers

4. Test smoking restrictions for exogeneity (Hausman's test)

Performing Aggregate Econometric Analysis (cont.)

5. Test for non-stationarity /unit root

Time-series data can be:

a) stationary (constant mean and variance over time)

b) trend-stationary (constant variance around a deterministic time trend)

c) non-stationary (stochastic time trend; variance increases with sample size; leads to spurious regression)

Performing Aggregate Econometric Analysis (cont.)

5. Test for non-stationarity/unit root (cont.)

- visually check the data

plot y t ; graph D.y t

- study the autocorrelations: **ac y**
- Dickey-Fuller test

5.1. test for unit root, $\mathbf{y_t}$ in levels, start with $\mathbf{y_{t-1}}$

dfuller y, trend

5.2. continue with $\mathbf{y_{t-2}}$, $\mathbf{y_{t-3}}$, $\mathbf{y_{t-4}}$

dfuller y, trend lags(2, 3, or 4)

Z(t) ; $\mathbf{H_0}$ = unit root exists

Performing Aggregate Econometric Analysis (cont.)

6. Test for first order integration I(1)

6.1. use variable in its first difference

(y_t - y_{t-1}), start with y_{t-1}

dfuller D.y, trend

6.2. continue with y_{t-2} , y_{t-3} , y_{t-4}

dfuller D.y, trend lags(2, 3, or 4)

Z(t) ; H_0 = unit root exists

Performing Aggregate Econometric Analysis (cont.)

- **Test all variables for non-stationarity:**

If all variables are stationary, estimate OLS model with a time-trend.

If some variables are trend-stationary, regress them on time. Use residuals from this regression and apply OLS with time-trend.

If variables are integrated of different order, they are not cointegrated.

If the variables are integrated of same order, test for cointegration.

Performing Aggregate Econometric Analysis (cont.)

7. Test for cointegration (Engle-Granger test)

Cointegration exists if a linear combination of the time series of non-stationary variables is stationary

regress y x ; **predict resid, residual**

Test residuals for stationarity (Dickey-Fuller test up to 4 lags)

dfuller resid, noconstant

Z(t) **;** $\mathbf{H_0}$ = no cointegration

Performing Aggregate Econometric Analysis (cont.)

8. Apply Engle-Granger two-step method for cointegrated time-series

8.1. Estimate a long-run equation without time trend

regress y x ; **test [varname]**

8.2. Estimate a short-term equation without time trend (Error-correction Model - ECM)

regress D.y D.x L.resid

L.resid coefficient measures speed of convergence to the long run equilibrium.

Performing Aggregate Econometric Analysis (cont.)

9. Steps for non-cointegrated time-series

9.1. Make all variables stationary by differencing

9.2. Estimate OLS with a time-trend with these stationary vars.

9.3. Significant time-trend signal an omitted variable. Specify a new model.

9.4. Even if time-trend is not significant, be alert of spurious regression.

Specification and Diagnostic Tests - Aggregate Analysis

1. Test residual for autocorrelation (serial correlation)

1.1. Durbin-Watson (DW) test, no lagged dependent variable in the model. **regress y x**

dwstat

1.2. Durbin-H test, lagged dependent variable in the model.

durbinh

If autocorrelation found, include omitted variable, use different functional form, or correct using robust standard error : **regress y x, robust**

Specification and Diagnostic Tests - Aggregate Analysis

2. Test residual for heteroscedasticity

White's test (also used as model misspecification test)

regress y x

whitetst

$\mathbf{H_0}$ = no heteroscedasticity

Correct for heteroscedasticity by choosing the White robust standard errors

regress y x, robust

Specification and Diagnostic Tests - Aggregate Analysis

3. **Test residual for autoregressive conditional heteroscedasticity (ARCH)**

ARCH Lagrange Multiplier (LM) test

arch y x, arch(1)

H_0 = no autoregressive heteroscedasticity

4. **Test residual for normality**

regress y x, robust

predict e, resid

sktest e

H_0 = residuals are normally distributed

Specification and Diagnostic Tests - Aggregate Analysis

5. Test coefficients' stability and structural change (Chow breakpoint test)

5.1. Split a sample into two sub-samples

gen p1 = 1 if time < t ; **replace p1=0 if time ≥ t**

5.2. Estimate the model for each sub-sample

gen px1 = p1*x1 ; **gen px2 = p1*x2** ; etc.

regress y x1 x2 p1 px1 px2

5.3. Tests coefficients related to split for significance

test p1 px1 px2

$\mathbf{H_0}$ = coefficients are stable

Specification and Diagnostic Tests - Aggregate Analysis

6. Test for model specification

6.1. Ramsey's Regression Specification Error Test (RESET)

regress y x ; ovtest

$\mathbf{H_0}$ = no misspecification

6.2. Check residuals for systematic patterns by graph

rvpplot x, border yline(0) and/or

acprplot, border c(s) bands(10)

Aggregate Analysis
Adjusting the model

If diagnostics suggest a problem you can:

- Apply different regression technique
- Change functional form
- Include/omit different variables
- Substitute proxies where appropriate

Test the new specification.

Continue until tests do not signal any problems.

Individual Level Data Analysis

Required Data

- Tobacco consumption by individuals at given point in time
- Price of tobacco products (tax may serve as a proxy for price) at given point in time; real source of variation in cross-section required

Other Data of Interest

- Public policy interventions (laws, restrictions)
- Socio-demographic information such as:
  - Per capita disposable income
  - Education
  - Age
  - Gender
  - Area of residence

Individual Level Data Analysis

Consumption must be correctly measured.

Specify the type of tobacco being used.

Example of standardized questions from WHO :

Have you ever smoked?

Have you ever smoked at least 100 cigarettes (or tobacco equivalent**) in your lifetime?**

Have you ever smoked daily? Do smoke *daily, occasionally* **or** *not at all*?

On average, what number of the following do you smoke per day? (Manufactured cigarettes; Hand-rolled cigarettes; Bidis; Pipefuls of tobacco; Cigars/cheroots/cigarillos; Goza/hookah)

How many years have you smoked/did you smoke daily? *(ever-daily smokers only)*

How long has it been since you last smoked daily? *(ex-daily smokers only)*

Individual Level Data Analysis

Definition of Smoker

Consumes any tobacco product either daily (***a daily smoker)*** or occasionally (***an occasional smoker)*** at time of survey

Definition of Non-Smoker

Does not smoke at all at time of survey, can belong to 3 categories:

Ex-smoker formerly daily smoker, currently does not smoke at all.

Never-smoker never smoked at all or have never been daily smokers or have smoked less than 100 cigarettes during lifetime.

Ex-occasional smoker former occasional but never daily smoker who smoked 100 or more cigarettes during lifetime

Definition of Lifetime Smoker

Smoked at least 100 cigarettes during lifetime (assumed to be addicted)

Individual Level Data Analysis

Calculating Smoking Prevalence

(# of smokers) : (# of respondents) * 100

Example: A survey of 500 men and 600 women ages 18 and older yielded 250 men and 200 women who smoked.

A) Male smoking prevalence = 250/500 x 100 = 50%

B) Female smoking prevalence=200/600 x 100 =33.3%

Individual Level Data Analysis

Coding Survey Data

Close-ended questions A3:

"How many cigarettes do you smoke each day?"

Answers: **1= none; 2=Less than one cigarette; 3=Less than half a pack; 4=About half a pack; 5=More than half a pack but less than a pack; 6=A pack; 7=More than a pack**

Recode:

```
* DEFINE CIGARETTES PER DAY  *
GEN CIGDAY = 0 IF A3 = = 1
REPLACE CIGDAY =.5 IF A3 = = 2
REPLACE CIGDAY = 5 IF A3 = =3
REPLACE CIGDAY = 10 IF A3 = =4
REPLACE CIGDAY = 15 IF A3 = =5
REPLACE CIGDAY = 20 IF A3 = =6
REPLACE CIGDAY = 30 IF A3 = =7
GEN MCIGDAY = 1 IF A3 = =.
REPLACELSE MCIGSDAY = 0 IF A3 ~=.
```

Individual Level Data Analysis

Coding Survey Data

2. Open-ended questions A1:

"How much do you pay for a single pack (20) of cigarettes"?

Recode:

```
 * DEFINE PRICE PAID PER PACK OF CIGARETTES  *
       GEN PRICE = A1
       REPLACE PRICE = 0 IF A1 = = .
 * DEFINE CIGARETTE PRICE MISSING*
       GEN MPRICE = 1 IF A1 = = .
       REPLACE MPRICE = 0 IF A1~= .
```

Individual Level Data Analysis

Check Data Quality

Check the data for consistency, completeness and accuracy:

- Produce Summary Statistics (Mean, Minimum, Maximum, Sum and Standard Deviation) and look for inconsistencies (e.g. std=0, mean=0, extreme values - “outliers”)
- Generate Frequency Statistics and compare them with values from the questionnaire.
- Computing Response Rates **[R / N-D] x 100**

N = # of respondents in the original random sample

D = # of respondents who could not be surveyed

R = # of respondents who actually replied to the questionnaire

Individual Level Data Analysis

The Econometric Model

Two-Part Model:

Part I: Probit/Logit Model for Smoking Participation

The ***dependent variable*** is a dichotomous measure of smoking (equals1 if the respondent is a "smoker", 0 otherwise)

Part II: Standard Ordinary Least Squares Model (OLS) measuring the Conditional Demand for Cigarettes among Smokers

The ***dependent variable*** is a continuous measure of daily cigarette consumption (# of cigs) among smokers

Individual Level Data Analysis

Two-Part Model (cont'd):

- Part I and Part II use *an identical set* of independent variables.
- Independent variables may include a mix of the following:

1. Demographic Information of each Respondent:

a. Age
b. Gender
c. Area of residency (rural, urban, etc.)
d. Religious background and/or religiosity

2. Socio-Economic Information for Each Respondent:

a. Education
b. Disposable monthly income

3. Price of cigarettes and/or other tobacco products

4. Non-price policy interventions

Individual Level Data Analysis

Modeling Other Non-Price Interventions

Create dichotomous indicator(s)

- **Example: Tobacco Control Environment in Country XYZ**
  - Tar and Nicotine Restrictions required since 1993
  - Health Warning Labels Issued in 1996
  - Smoking Restrictions in Workplaces not yet required
  - Advertising Ban issued in 1998
  - Nicotine Replacement Therapies enter market in 2001

Year	TNContent	Warnings	Workplace	AdBan	NRT
1991	0	0	0	0	0
1992	0	0	0	0	0
1993	1	0	0	0	0
1994	1	0	0	0	0
1995	1	0	0	0	0
1996	1	1	0	0	0
1997	1	1	0	0	0
1998	1	1	0	1	0
1999	1	1	0	1	0
2000	1	1	0	1	0
2001	1	1	0	1	1

Individual Level Data Analysis

Basic STATA Commands

. **set mem 100m** (Assigns on-line memory)

. **set matsize 200** (Provide dimensions for matrices)

. **use "c:\econ\data“, clear** (loads in the data)

. **log using "c:\econ\results.log“**(opens a log file that saves your results)

. **save data1, replace** (creates new dataset, prevents rewriting the original data)

. **use “c:\econ\data\data1, clear”** (reads in temporary data set you can work with)

Individual Level Data Analysis

Basic STATA Commands

How to merge two datasets

. use data2

. sort id

. save data2, replace [the using dataset]

. use data1 [data currently in memory, called the master dataset]

. sort id (variable used for merge must have the same name)

. merge id using data2

. tabulate _merge [1=obs. from master data; 2=obs. from using data;
3=obs.from both master and using data]

Individual Level Data Analysis

Analysis of Individual Level Data

Descriptive Statistics

. Summarize

Drop observations with missing key variables

. Drop if msmoker = = 1 & mprice = = 1

Probit Model (Smoking Participation)

. Probit smoker age mage male mmale educ meduc relig mrelig income mincome city mcity price law1 mlaw1 law2 mlaw2

OLS Model (Smoking Intensity)

. Regress cigday age mage male mmale educ meduc relig mrelig income mincome city mcity price law1 mlaw1 law2 mlaw2 if smoker = = 1

Interpreting results

1. Price results

- Compute price elasticity

1.1. Linear demand function

$$e_{Pt} = b_1 \times P_t / Q_t$$

1.2. Log-lin and Lin-log demand function

$e_{Pt} = b_1 \times P_t$ and $e_{Pt} = b_1 \times (1 / Q_t)$, respectively

1.3. Log-log demand function

Price coefficient = price elasticity (constant)

1.4. Error correction model

$$e_{Pt} = b_1 \times P_t\text{hat} / Q_t\text{hat}$$

Interpreting results (cont.)

2.1. Tax results

- Compute price elasticity

{d(Q) / d(T) * average P / average Q * d(T)/d(P)}

If d(T)/d(P)=1, then tax coefficient is multiplied by the same expression used to compute price elasticity.

Expected values of price elasticities: –0.14 to –1.23.

Higher expected estimates in low- and middle-income countries

Interpreting results (cont.)

2.2. Tax results

- impact of a tax increase on tax revenue

Current tax revenue = tax base(# cigarette packs) * tax rate(tax per pack)

Percentage increase in price = proposed tax/current price*100

Change in demand after a tax increase = current demand(tax base) * price elasticity * percentage increase in price/100

New tax revenue = new demand * new tax rate

Interpreting results (cont.)

3. Results for income

- Compute income elasticity

Apply the same methods as for price elasticity replacing price with income.

If the elasticity is positive, tobacco product is a superior good.

if the elasticity is negative, tobacco product is an inferior good.

Interpreting results (cont.)

4. Results for advertising

- If annual data used, expect statistically not significant coefficient due to diminishing marginal product of advertising expenditures.
- Advertising expenditures can be endogenous, i.e. affected by cigarette sales.
- Effect of advertising ban may be small or insignificant, if other media are still available to the industry.

Interpreting results (cont.)

5. **Results for health information and counter-advertising**
   - Account for diminishing influence over time.
6. **Results for smoking restrictions**
   - Account for possible endogeneity with respect to cigarette consumption (anti-smoking sentiment), and enforcement.
7. **Results for price of other tobacco products**
   - Compute cross-price elasticity. If positive, substitution exists; if negative, complementary exists.