1 Demand Analysis

Demand analysis is one of the first topics come to in economics.

- Very important especially in the Keynesian paradigm.
- Very important for companies: mainstay of consultancies

As have seen traditional consumer theory is based on the Neoclassical model of consumer choice.

Demand function

$$\begin{aligned} q_i &= q_i(p_1, p_2, ..., p_n, x) : i = 1, ..., n \\ x &= \sum_i p_i q_i \end{aligned}$$

Theory actual gives very little information on functional form, other relevant variables, form of variables. But it does imply restriction which can be useful in reducing number of parameters to be estimated (the degrees of freedom taken up).

Expect homogeneous of degree zero in prices and total expenditure. This means that equal increases in prices and income should leave demand unchanged.

The Slutsky equation suggests own price substitution effects are negative.

$$\frac{\delta q_i}{\delta p_i} + q_i \left(\frac{\delta q_i}{\delta x} \right) < 0$$

This basic theory is used in a relatively 'ad hoc' way in applied work.

1.1 Estimating Demand

Use econometric analysis to estimate demand functions: assume you have basic understanding of regression analysis

- quick summary......

There are alternatives:

- Use consumer interviews: but problems
- Use market experiments
- Consumer clinics: lab experiments
- See Mansfield et al for discussion

1.2 Functional Form

The functional form is chosen for ease of exposition. There are a limited number of explanatory variables: own price, prices of substitutes and complements and possibly the general price level, with a time trend to capture changing tastes.

Popular specification is log linear, which has an added advantage that the coefficients are elasticities.

$$\begin{aligned} q_i &= A(p_i^{\beta_1}, p_j^{\beta_2}, \overline{p}^{\beta_3} x^{\beta_4} \exp(\beta_5 t) \exp(\varepsilon) \\ \log q_i &= \beta_0 + \beta_1 \log p_i + \beta_2 \log p_j + \beta_3 \log \overline{p} + \beta_4 \log \overline{x} + \beta_5 t + \varepsilon \end{aligned}$$

Can impose homogeneity prior to estimation, relative to $\overline{p}$. This implies $\beta_1 + \beta_2 + \beta_3 + \beta_4 = 0$ or $\beta_3 = -\beta_1 - \beta_2 - \beta_4$

$$\log q_i = \beta_0 + \beta_1 \log \left(\frac{p_i}{\overline{p}} \right) + \beta_2 \log \left(\frac{p_j}{\overline{p}} \right) + \beta_4 \log \left(\frac{\overline{x}}{\overline{p}} \right) + ...$$

This restriction could be tested in the normal way by comparing the unrestricted estimates with:

$$\log q_i = \beta_0 + \beta_1 \log p_1 + \beta_2 \log p_j + (-\beta_1 - \beta_2 - \beta_4) \log \overline{p} + ...$$

In some case, where the researcher has needed to save dof this restriction is simply imposed. Using relative prices and real income can also have the advantage of reducing multicollineartiy, something that is often present in time series.

The negativity restriction is an inequality and difficult to impose.

In general studies of demand for single goods or policy orientated studies are often more concerned with estimation of elasticities rather than testing the theory. They just impose the theory.

There are a number of problems, which are discussed in more detail in Thomas.:

1.3 Aggregation Problems

Consumer theory has considerable problems at a conceptual level -family versus individual, but at practical level there are further problems. The data is often presented in broad categories, large groups of individuals, in both time series and cross section.

Just having a theory of the individual doesn't mean it will hold at the aggregate level.

- The implications are that it should.
- Part of N-C/New Keynesian methodology of individual as focus of analysis
- Often cop out -use representative individual

Can aggregate over commodities as long as the groupings mean something, but with individuals require some restrictive assumptions -see Thomas

Problem is worse in the case of non-linear relations. For a linear model

$$\begin{aligned} y_i &= \alpha + \beta x_i \\ \overline{y} &= \alpha + \beta \overline{x} \end{aligned}$$

But if in logs then shouldnt use the arithmetic mean.

1.4 Identification Problem

Simultaneity is possible:

Early studies focused on agricultural products, where the data was available and identification was not a problem. Later manufacturing studies hit the classical identification problems. No reason why supply conditions should be more variable than demand

Simultaneity implies biased and inconsistent estimators of the demand equations.

1.5 Multicollinearity

As noted we might expect multicollinearity between expenditure and prices as both are trended. This will increase standard errors and reduce precision. It can be reduced by imposing homogeneity, but if their is insufficient variance in the explanatory variables, then any remaining multicollinearity, when homogeneity is imposed could make matters worse.

In the past used extraneous estimates of realexpenditure elasticity from cross section studies -assuming absense of price variation, but wide variation in real expenditures. If the extraneous estimates are unbiased then so are the estimates.

$$q_t = \alpha + \beta p_t + \gamma x_t$$

get estimate from cross section $\widehat{\gamma}$

$$q_t - \widehat{\gamma} x_t = \alpha + \beta p_t$$

but problem:

- standard error attached is not considered
- interpretation: cross section and time series are not the same thing, the former can be considered to represent the represents the long run effect and the latter the short run
- So care must be taken

1.6 Autocorrelation

This implies that taking the time series regression

$$Y_t = \alpha + \beta X_t + u_t$$

but in this case there is some relation between the error terms across observations.

$$\begin{aligned} E(u_t) &= 0 \\ E(u_t^2) &= \sigma^2 \\ E(u_s u_p) &\neq 0 \end{aligned}$$

- Thus the error covariances are not zero.
- Means that one of the assumption that makes OLS BLU does not hold.

1.7 Likely causes:

1. Omit variable that ought to be included.
2. Mispecification of the functional form. This is most obvious where a straight line is put through a curve of dots. This would clearly show up in plots of residuals.
3. Errors of measurement in the dependent variable. If the errors are not random then the error term will pick up any sytematic mistakes.

1.8 The Problem

OLS is not the best estimation method.

- It will underestimate the true variance.
- the t values will look too good
- will reject H_0 when it is true

So estimates will be unbiased but inefficient (not least variance)

Focus on simplest form of relation over time: first order autocorrelation which can be written as

$$u_t = \rho u_{t-1} + \varepsilon_t$$

Obviously there could be more complicated forms.

1.8.1 Tests

1. Plot the residuals over time or against a particular variable and see if there is a pattern.

   little change of sign $\Longrightarrow$ positive autocorrelation

2. Durbin Watson Statistic: commonly used

$$DW = \frac{\sum(\hat{u}_t - \hat{u}_{t-1})^2}{\sum \hat{u}_t^2}$$

$$= \frac{\sum \hat{u}_t^2 + \sum \hat{u}_{t-1}^2 - 2\sum \hat{u}_t\hat{u}_{t-1}}{\sum \hat{u}_t^2}$$

   Now when the number of observations is very large $\sum \hat{u}_t^2$ and $\sum u_{t-1}^2$will be almost the same, so

$$\begin{aligned} &= \frac{2\sum \hat{u}_t^2}{\sum \hat{u}_t^2} - \frac{\sum \hat{u}_t\hat{u}_{t-1}}{\sum \hat{u}_t^2} \\ &= 2\left(1 - \frac{\sum \hat{u}_t\hat{u}_{t-1}}{\sum \hat{u}_t^2}\right) \\ &= 2(1-\widehat{\rho}) \end{aligned}$$

   so we have

$$DW \approx 2(1-\widehat{\rho})$$

   - if strong positive autocorrelation then $\widehat{\rho} = 1$ and $DW = 0$
   - if strong negative autocorrelation then $\widehat{\rho} = -1$ and $DW = 4$
   - if no autocorrelation then $\widehat{\rho} = 0$ and $DW = 2$

     So the best can hope for is a $DW$of 2

     But sampling distribution of the DW depends on the values of the explanatory variables and so can only derive upper and lower limits

- – $DW < DW_L$ reject hypothesis no autocorrelation

- $DW > DW_U$ don't reject
- $DW_L < DW < DW_U$ inconclusive

- increasing the observations shrinks the indeterminacy region
- increasing variables increases the indeterminacy region
- Rule of thumb: lower limit for positive autocorrelation = 1.6
- Durbin's h used if LDV
- LM test in Microfit works for higher order
- DW test can also be considered a general mispecification test if there is no autocorrelation

1.8.2 Solutions

- Find cause
- increase number of observations
- find missing values
- specify correctly
- Microfit provides number of procedures eg Cochrane Orcutt -last resort
- Most important: It is easy to confuse mispecified dynamics with serial correlation in the errors.In fact it is best to always start from a general dynamic models and test the restrictions before applying the tests for serial correlation.
- The AR(1) is only one possible dynamic model,

1.9 Heteroscedasticity

In this case

$$Y_i = \alpha + \beta X_i + u_i$$

we assume

$$\begin{aligned} E(u_i) &= 0 \\ E(u_i^2) &= \sigma_i^2 \\ E(u_i u_j) &= 0 \end{aligned}$$

So in this case the errors do not have a common variance.

The effect of this will be to leave the OLS estimator of β unbiased, but the estimated standard error will be biased.

- Number of tests available including that reported in Microfit
- Solution: Correct the standard errors, using the White Heteroscedastic robust errors available in packages.

1.10 Outliers

1.10.1 Problem

Regression parameters can be influenced by a few extreme values or outliers

- Should be able to spot from a careful analysis of the residuals $\widehat{u}_t$
- In the case of a simple bivariate regression you can simply plot the data.
- Outlier is an observation that is very different: usually generated by some unusual factor
- Least squares estimates are very sensitive to outliers, particularly in small samples
- Maddala P89-90 gives examples of data sets that when plotted look very different, but give the same regression results. In two cases this is caused by a single extreme value.

1.10.2 Actions

- Drop the observations with large residuals and reestimate the equation. This should really be a last resort
- The outliers may provide important information. They may not be outliers at all. An example of this is the relation between infant mortality and GDP per capita in Asian countries.
- For cross section should maybe try to get more data rather than drop observations. Also for time series.
- Problem of what is an outlier also relates to leverage: need variation in the data or cant estimate any relationship. Its not always obvious when information on a system becomes an outlier.
- Can treat extreme observations with dummy variables

1.11 Omitted variable bias

If we miss out an important variable it not only means our model is poorly specified it also means that any estimated parameters are likely to be biased.

- Incorrect omission of variables leads to biased estimates of the parameters that are included
- Incorrect inclusion only produces inefficient estimates, so don't have minimum variance
- So better to include the wrong variables rather than exclude the right ones.

1.12 Dynamic Models

1. The problem with specifying the dynamic form of a regression model is that normally the theory provides little information on lag lengths, nature of adjustments etc. So seems better to use the theory to specify the variables to be included, but to allow the data to determine what the dynamic model should look like

   (a) Consider a log linear demand function with only own price as a dependent variable:

   $$q_t = \beta_0 + \beta_1 p_t + \beta_2 p_{t-1} + \beta_3 q_{t-1} + u_t$$

   where q is $\log(q)$ and p is $log(p)$.

   This encompasses a number of different models all with different dynamic structures they are 'nested' in this model meaning the restrictions on the parameters can be written down and they can be tested.

2. Static model: $\beta_2 = \beta_3 = 0$

   $$p_t = \alpha_0 + \alpha_1 p_t + v_{1t}$$

3. AR(1) model: first order autoregression $\beta_1 = \beta_2 = 0$

   $$q_t = \alpha_0 + \alpha_1 q_{t-1} + v_{2t}$$

4. Partial adjustment/habit persistence model $\beta_2 = 0$

   $$q_t = \alpha_0 + \alpha_1 p_t + \alpha_2 q_{t-1} + v_{3t}$$

   This is a habit model if $\alpha_2 > 0$ and a partial adjustment model if $|\alpha_2| < 1$

5. Distributed lag model $\beta_3 = 0$

   $$q_t = \alpha_0 + \alpha_1 p_t + \alpha_2 p_{t-1} + v_{4t}$$

6. First difference $\beta_3 = 1$ and $\beta_2 = -\beta_1$

   $$\Delta q_t = \alpha_0 + \alpha_1 \Delta p_t + v_{5t}$$

   where $\Delta q_t = q_t - q_{t-1}$and $\Delta p_t = p_t - p_{t-1}$

7. Error correction model

   $$\Delta p_t = \alpha_0 + \alpha_1 \Delta q_t + \alpha_2 (p_{t-1} - q_{t-1}) + v_{6t}$$

   where $\beta_3 = 1 + \alpha_2$ and $\beta_2 = -(\alpha_1 + \alpha_2)$ and there is a long run elasticity of demand of unity.

This is now a very commonly used model, because of its use in cointegration analysis

- All of these restrictions can be tested and you will do this in the exercise
- Look for most parsimonious
- Another issue is the long run solutions of these models
- Set $p_t = p_{t-1} = p$ and do the same for q gives the long run solution
- Note that the first difference equation does not have one so if you were to find this was the best model you would only have short run dynamics.
- It is possible that this problem can be dealt with but need to discuss the concept of cointegration

1.13 Early Studies

Engel curves

$$p_i q_i = \alpha + \beta y_i$$

Cross section studies provided a test for Engel's law, that the income elasticity of demand for food was always less than one

- In cross section prices are pretty much fixed, expenditures vary
- But problem that other factors may be important which could give ommited variable bias
  - household characteristics: in particular houselhold size. Early investigators used equivalent adult scales
  - social effects. Can use dummies for social status grouping etc...
  - functional form. Major problem as goods can change from luxuries at low income to necessities at high income. Obvious functional form is sigmoid but is non linear and complex to estimate in practice. Could estimate income ranges separately: lower log linear, upper semi log, middle linear

1.14 Recent Developments

Duality: Use concept of duality to reformulate the consumer problem as choosing quantities so as to minimise the total expenditure necessary to achieve a given utility level .

1.14.1 Demand systems

Rather than focus on individual commodities much recent work has been concerned with complete systems of demand equations.

Advantages are:

- reduce degrees of freedom problem
- can test restrictions to limit number of parameters ratehr than impose ad hoc

Methods:

1. Specify form of utility function and then derive demand curves that satisfy the theoretical restrictions. Advantage is dof saved, but disadvantage is that can't test restrictions and there is a loss of generality.
2. Begin with demand system capable of satisfying restrictions, but that doesn't necessarily do so, and test if they hold. Advantage is can test, disadvantage is dof problem.

Consider forms of demand system

- Linear expenditure system -Stone: explicitly specified utility function
- Rotterdam model -test restrictions, popular until recently
- Indirect addilog and double log -from indirect utility function
- Direct and indirect translog -providing flexible functional form
- Almost Ideal demand system -from duality
- Variants

1.14.2 Linear Expenditure System

First used by Stone (1954) this system has an explicitly specified utility function:

$$\begin{aligned} U &= \beta_1 \log(q_1 - \gamma_1) + ... + \beta_n \log(q_n - \gamma_n) \\ subject\ to \sum p_i q_i &= x \end{aligned}$$

$$p_i q_i = p_i \gamma_i + \beta_i \left[x - \sum_j p_j \gamma_j \right]$$

with $p_i\gamma_i$ representing subsistence expenditure and $x - \sum_j p_j \gamma_j$ supernumerary expenditure.

Advantages:

- expresses q_i as a linear function of real total expenditure x/p_i and of relative prices p_j/p_i

- is the only demand system that satisfies all the theoretical restrictions

But suffers from the fact that the underlying utility function is is additive and hence not general.

1.14.3 Almost Ideal Demand System

Deaton and Muellbauer: start with a general cost function and derive a share equation of the form:

$$\begin{aligned} w_i &= \alpha_i + \sum_j \gamma_{ij} \log p_j + \beta_i \log\left(\frac{x}{P}\right) \\ \log P &= \alpha_0 + \sum_k \alpha_k \log p_k + \frac{1}{2}\sum_j \sum_k \log p_k \log p_j \end{aligned}$$

Usually use approximation:

$$\log P^* = \sum_j w_i \log p_j$$

to estimate by OLS, but to test symmetry need to use the proper version, which requires systems estimation as it implies cross equation restrictions

So estimate

$$\begin{aligned} w_i &= \alpha_i + \sum_j \gamma_{ij} \log p_j + \beta_i \log\left(\frac{x}{P^*}\right) + \varepsilon_i \\ \log P^* &= \sum_j w_i \log p_j \end{aligned}$$

which means a system of equations:

$$\begin{aligned} w_1 &= \alpha_1 + \gamma_{11} \log p_1 + \gamma_{12} \log p_2 + ... + \gamma_{1n} \log p_n + \beta_1 \log\left(\frac{x}{P^*}\right) + \varepsilon_1 \\ w_2 &= \alpha_2 + \gamma_{21} \log p_1 + \gamma_{22} \log p_2 + ... + \gamma_{2n} \log p_n + \beta_2 \log\left(\frac{x}{P^*}\right) + \varepsilon_2 \\ & .... \\ w_n &= \alpha_n + \gamma_{n1} \log p_1 + \gamma_{n2} \log p_2 + ... + \gamma_{nn} \log p_n + \beta n \log\left(\frac{x}{P^*}\right) + \varepsilon n \end{aligned}$$

This is a singular system as the dependent variables are shares and so add up to one across all the commodities. This means that adding up is automatically satisfied

$$\begin{aligned}\sum_i \alpha_i &= 1\\ \sum_i \gamma_{ij} &= 0\\ \sum_i \beta_i &= 0\end{aligned}$$

Homogeneity is testable and implies

$$\sum_j \gamma_{ij} = 0$$

This can be tested equation by equation using OLS, but symmetry

$$\gamma_{ij} = \gamma_{ji}$$

requires MLE

Consider a 4 commodity system

$$\begin{aligned}
w_1 &= \alpha_1 + \gamma_{11}\log p_1 + \gamma_{12}\log p_2 + \gamma_{13}\log p_3 + \gamma_{14}\log p_4 + \beta_1(\log x - \log P^*) + \varepsilon_1\\
w_2 &= \alpha_1 + \gamma_{21}\log p_1 + \gamma_{22}\log p_2 + \gamma_{23}\log p_3 + \gamma_{24}\log p_4 + \beta_2(\log x - \log P^*) + \varepsilon_2\\
w_3 &= \alpha_1 + \gamma_{31}\log p_1 + \gamma_{32}\log p_2 + \gamma_{33}\log p_3 + \gamma_{34}\log p_3 + \beta_3(\log x - \log P^*) + \varepsilon_3\\
w_4 &= \alpha_1 + \gamma_{41}\log p_1 + \gamma_{42}\log p_2 + \gamma_{43}\log p_3 + \gamma_{44}\log p_4 + \beta_4(\log x - \log P^*) + \varepsilon_4
\end{aligned}$$

Homogeneity is tested by imposing the restrictions on the individual equations. Taking the second equation:

$$\gamma_{21} + \gamma_{22} + \gamma_{23} + \gamma_{24} = 0$$

which implies

$$\gamma_{21} + \gamma_{22} + \gamma_{23} = -\gamma_{24}$$

so

$$w_2 = \alpha_1 + \gamma_{21}(\log p_1 - \log p_4) + \gamma_{22}(\log p_2 - \log p_4) + \gamma_{23}(\log p_3 - \log p_4) + \beta_2(\log x - \log P^*) + \varepsilon_2$$

To test the restrictions estimate each equation by OLS restricted and unrestricted and the do LLR or F test.

As the dependent variable is shares the coefficients are not elasticities. We compute the elasticities as:

Expenditure

$$e_i = 1 + \frac{\beta_i}{w_i}$$

Compensated

$$e^*_{ij} = \frac{1}{w_i}\left[\left(\gamma_{ij} + \beta_i\beta_j \log\left(\frac{x}{P^*}\right)\right) - \delta_{ij} + w_j\right]$$

where $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ otherwise
Uncompensated

$$e_{ij} = e^*_{ij} - e_i w_j$$

So the uncompensated is the uncompensated minus the share weighted compensated.

Deaton and Muellbauer reject homogeneity for 4 commodity groups in the UK (food, clothing, housing, transport) and also not sharp decrease in DW statistic. Suggest rejection of homogeneity may be a problem:

- Omitted Variables: dynamics or conditioning variables that may be important
- Price expectations
- Aggregation problem
- Static model assumptions are inadequate
- Argue premature to reject consumer theory as consumption involves intertemporal choices, might need to consider labour supply, failure in models to take account of dynamic factors.
- Note have ommited durables: earlier studies didnt necessarily do so.

Developments:

- General dynamic model:
- Dynamise theory: Introduce a dynamic adjustment process.

1.14.4 Other Models

Havent time to go through these, but check out Thomas

- Indirect addilog and double log -from indirect utility function
- Direct and indirect translog -providing flexible functional form