TA session

Demand Analysis with the Almost Ideal Demand System (AIDS) Model

Applied IO with Susanna Esteban (IDEA, Spring 2023) TA: Conghan Zheng

Last updated: 2023-01-25

Contents

Latest version of this file:

- pdf;
- html (contains R code chunks).

1 Overview

- In the following example, we will estimate a nonlinear Almost Ideal Demand System (AIDS) model and its linear approximation (LA-AIDS model).

1.1 Model

Recall from microeconomic theory the expenditure function of an n-good system (in most recent applications of the AIDS model, a "good" is an aggregate product category):

$$e(p,u) \equiv \min_{\{q_j\}_{j=1}^n} \sum_{j=1}^n p_j q_j \quad s.t.\ U(q_1,\dots,q_n) \geq u$$

The Hicksian (or compensated) demand function solves the expenditure minimization problem:

$$h(p,u) \equiv \underset{\{q_j\}_{j=1}^n}{\operatorname{argmin}} \sum_{j=1}^n p_j q_j \quad s.t.\ U(q_1,\dots,q_n) \geq u$$

It follows that

$$e(p,u) = p \cdot h(p,u) \tag{1}$$

where the price vector $p = (p_1, \dots, p_n)$. By the Shephard's Lemma, we have

$$\frac{\partial e(p,u)}{\partial p_j} = h_j(p,u) = q_j$$

For a utility-maximizing consumer, her total expenditure X satisfies

$$X = e(p,u) \tag{2}$$

We can then write the expenditure share of product j:

$$w_j \equiv \frac{p_j q_j}{X} \overset{(2)}{=} \frac{p_j q_j}{e(p,u)} = \frac{\partial \ln e(p,u)}{\partial \ln p_j} \tag{3}$$

where the chain rule is used.

Following Deaton and Muellbauer (1980a), we take the *Piglog* class of preferences. The Piglog form was developed to treat aggregate consumer behavior as if it were the outcome of a single rational representative consumer. If this is the case, then a reallocation of a single unit of currency from any one household to another must leave market demand unchanged. In other words, the marginal propensity to spend must be the same for all households, so the quantity consumed is linear in expenditure. Furthermore, we have the following log-expenditure function:

$$\ln e(p,u) = a(p) + b(p) \cdot u \tag{4}$$

Deaton and Muellbauer (1980a) set the functional forms of $a(p)$ and $b(p)$ to ensure that $e(\cdot,\cdot)$ is flexible enough and has enough parameters for its derivatives $\frac{\partial e}{\partial p_j}$, $\frac{\partial e}{\partial u}$, $\frac{\partial e}{\partial p_j \partial p_k}$, $\frac{\partial e}{\partial \partial p_j \partial u}$, $\frac{\partial e}{\partial u^2}$:

$$a(p) = \sum_{j=1}^{n} \alpha_j \ln p_j + \frac{1}{2} \sum_{j=1}^{n} \sum_{k=1}^{n} \gamma_{jk}^* \ln p_j \ln p_k$$
$$b(p) = \prod_{j=1}^{n} p_j^{\beta_j}$$

The translog expression $a(p)$ can be obtained by expanding $\ln e(p)$ in a second-order Taylor series about the point $\ln p = 0$.

The homogeneity of degree of the expenditure function in p requires:

$$\sum_{j=1}^{n} \alpha_j = 1, \ \sum_{j=1}^{n} \gamma_{jk}^* = \sum_{j=1}^{n} \gamma_{kj}^* = \sum_{j=1}^{n} \beta_j = 0 \tag{5}$$

It then follows that

$$\begin{aligned}
(3)\&(4) \Rightarrow w_j &= \frac{\partial \ln e(p,u)}{\partial \ln p_j} \\
&= \alpha_j + \frac{1}{2} \sum_k \gamma_{jk}^* \ln p_k + \frac{1}{2} \sum_k \gamma_{kj}^* \ln p_k + \beta_j \prod_{j=1}^{n} p_j^{\beta_j} \cdot u \\
&\equiv \alpha_j + \sum_k \gamma_{jk} \ln p_k + \beta_j \prod_{j=1}^{n} p_j^{\beta_j} \cdot u \\
&= \alpha_j + \sum_k \gamma_{jk} \ln p_k + \beta_j b(p) u \\
&\overset{(4)}{=} \alpha_j + \sum_k \gamma_{jk} \ln p_k + \beta_j \Big(\ln e(p,u) - a(p) \Big) \\
&\overset{(2)}{\equiv} \alpha_j + \sum_k \gamma_{jk} \ln p_k + \beta_j \ln \left(\frac{X}{P} \right)
\end{aligned} \tag{6}$$

where $\gamma_{jk} = \gamma_{kj} \equiv \frac{1}{2}\left(\gamma^*_{jk} + \gamma^*_{kj}\right)$ and the price index P is defined by

$$\begin{aligned}\ln P \equiv a(p) &= \frac{1}{2}\big(a(p) + a(p)\big) \\ &= \sum_{j=1}^{n} \alpha_j \ln p_j + \frac{1}{2} \sum_{j=1}^{n}\sum_{k=1}^{n} \gamma_{jk} \ln p_j \ln p_k \end{aligned} \tag{7}$$

Expression (6) is the *AIDS demand system* in budget share form.

1.2 Empirical Strategy

Now we have a system of equations that can be econometrically estimated (the AIDS model):

$$w_{jt} = \alpha_j + \sum_k \gamma_{jk} \ln p_{jt} + \beta_j \ln\left(\frac{X_t}{P_t}\right) + \varepsilon_{jt}, \quad \sum_j \varepsilon_{jt} = 0, \forall t \tag{8}$$

Given that the observed shares w_i always sum to one, the constraints on parameters (5) are automatically satisfied.

One problem is that $\sum_j \varepsilon_j = 0$ implies that the covariance matrix is singular, to avoid estimation problems due to this singularity, one of the equations has to be dropped from the system. But no information will be lost by doing this, the coefficients of the dropped equation can be calculated from the restrictions on parameters (5), the estimation results do not depend on which equation is dropped.

The demand system (8) is linear except for the price index P_t. Deaton and Muellbauer (1980a) suggest approximating the price index with the Stone's index to make the entire demand system linear:

$$\ln P^*_t = \sum_j w_{jt} \ln p_{jt} \tag{9}$$

Then (8) can be estimated as (linear approximation of the AIDS, the LA-AIDS model)

$$\begin{aligned} w_{jt} &= \alpha_j - \beta_j \ln \phi + \sum_k \gamma_{jk} \ln p_{jt} + \beta_j \ln\left(\frac{X_t}{P^*_t}\right) + \varepsilon_{jt} \\ &\equiv \alpha^*_j + \sum_k \gamma_{jk} \ln p_{jt} + \beta_j \ln\left(\frac{X_t}{P^*_t}\right) + \varepsilon_{jt} \end{aligned} \tag{10}$$

where $P \simeq \phi P^*$, $\alpha^*_j \equiv \alpha_j - \beta_j \ln \phi$, and $\sum_j \alpha^*_j = 0$ is still required for adding up, since $\sum_j \beta_j = 0$.

Since the number of parameters increases quadratically with the number of products, the estimation of this model requires that T (either time periods or geographic markets for spatial analysis) is substantially larger than the number of products n.

Given data on prices $\{p_{jt}\}_{j,t}$ and quantities $\{q_{jt}\}_{j,t}$, our goal is to estimate the parameters $\{\alpha_j\}_j$, $\{\beta_j\}_j$, and $\{\gamma_{jk}\}_{j,k}$ of the AIDS (8) or the parameters $\{\alpha^*_j\}_j$, $\{\beta_j\}_j$, and $\{\gamma_{jk}\}_{j,k}$ of the LA-AIDS (10).

1.2.1 LA-AIDS model

Estimation Steps Given data on prices $\{p_{jt}\}_{j,t}$ and quantities $\{q_{jt}\}_{j,t}$:

- Calculate the total expenditures $y_{jt} = \sum_j p_{jt} q_{jt}$ and the expenditure shares $w_{jt} = \frac{p_{jt} q_{jt}}{X_t}$ from prices and quantities data.
- Compute the Stone's price index $\ln P^*$ from shares and prices data using expression (9).
- Estimate model (10), obtain an $n \times 1$ vector for α, an $n \times 1$ vector for β, and an $n \times n$ matrix for γ.

Empirical Considerations

1. Simultaneity Bias The Stone's index includes current budget shares (w_{it}), which appear on both the left and right sides of (10). Several scholars (e.g. Blanciforti et al. 1986) suggest using lagged shares in the Stone's price index:

$$\ln P_t^{*L} = \sum_j w_{j,t-1} \ln p_{jt}$$

2. Unit of measurement Obviously, the Stone's price index (9) is not invariant to changes in the units of measurement (e.g. EURO/kg or USD/pound). One solution is to normalize: to choose a "base" period and to use the relative prices (see Moschini, 1995).

1.2.2 AIDS model

Estimation Steps Iterative Linear Least Squares Estimation:

- Given a starting value $\ln P_t^0$ of $\ln P_t$, (8) can be estimated by linear estimation techniques. A good choice for the initial value is the LA-AIDS estimate.
- The price index can be updated with the estimated coefficients using expression (7), after this step, we get $\ln P_t^1$.
- Estimate (8) with the updated value of $\ln P_t$, repeat untill convergence.

At each iteration, (8) is a system of seemingly unrelated regressions (SUR), and we can estimate it using constrained (constrained parameters) feasible GLS (see Greene, 2018, chapter 10).

This iterated SUR estimator generally converges to the FIML estimates (Blanciforti et al. (1986) estimate their AIDS model by FIML), maximum likelihood has no advantages over FGLS in its asymptotic properties (Greene, 2018).

2 Estimation

2.1 Resources[1]

- R (more online resources for now):
  - Package micEconAids (user-written package, still maintained; for Seemingly Unrelated Regression (SUR), it calls the systemfit package).
- Python:
  - No user-written packages just designed for the AIDS model, to my knowledge. For SUR, you can choose the command `SUR` from the `linearmodels` package.
- Stata:
  - `ml` command
  - `nlsur` command (`nlsur` provides similar results but much faster than `ml`)
  - `quaids` command (more than `nlsur`, `quaids` allows the user to fit either the standard AIDS model of Deaton & Muellbauer (1980a), or its quadratic variant of Banks et al., 1997. Demographic variables can be specified, elasticities can be calculated using postestimation commands.)
  - Or, you can estimate the SUR using the commands `sureg` and `suest`.
- SAS:

[1] Stata is certainly not a good tool for structural models, Matlab is not always very fast. For me, R and Python are somewhat similar and they each have their own occasions. Python (`NumPy`) is more general in many cases, e.g., it understands either row-major (C style) or column-major ordered arrays (Fortran style), while R only understands the column-major. Python has the official pystata for Stata integration, while R only has several user-written packages and none of them work well (I remember Wagner wasting a lot of time on these packages in his PS last semester). Katherina told me she found the online resources for structural models mainly in Python (especially of more macro) or Matlab. When it comes to the performance, Python has the clever `@jit` from `Numba` that can speed up the code (Tino recommended it to me). But you can go one step further: integrate R with C++ (`Rcpp`) to improve the performance (Hanna recommended it to me). And more, R also has `shiny` which I guess no labor economist can refuse.

– Procedures `syslin` (LA-AIDS) and `model` (AIDS), an example: https://support.sas.com/rnd/app/ets/examples/aids/index.htm, calculating elasticities: https://support.sas.com/rnd/app/ets/examples/elasticity/index.htm

- Matlab, Fortran, Gauss, C, C++ . . . : You can code it up by yourself, it's not difficult if you are familiar with the language/software.

For TA materials I choose R notebook[2] for illustration purpose.

2.2 Data

Our data set is U.S. food demand data which distinguishes four categories ($j \in \{1, 2, 3, 4\}$) of food:

- meats (`Food1`),
- fruits and vegetables (`Food2`),
- cereal and bakery products (`Food3`), and
- miscellaneous foods (`Food4`).

These data are available for the years 1947 to 1978.

```
## # A tibble: 6 x 13
## # Groups:   Year [2]
##    Year   x_t Food   w_jt  p_jt   w_j   p_j lnp_jt lnx_t lnP_s~1 lnP_s~2 lnx_l~3
##   <int> <dbl> <chr> <dbl> <dbl> <dbl> <dbl>  <dbl> <dbl>   <dbl>   <dbl>   <dbl>
## 1  1947  310. 1     0.298  59.6 0.310  85.9   4.09  5.74    4.05    2.41    1.69
## 2  1947  310. 2     0.172  53.8 0.200  84.7   3.99  5.74    4.05    2.41    1.69
## 3  1947  310. 3     0.134  52.1 0.134  89.8   3.95  5.74    4.05    2.41    1.69
## 4  1947  310. 4     0.397  58   0.355  89.0   4.06  5.74    4.05    2.41    1.69
## 5  1948  326. 1     0.305  67.6 0.310  85.9   4.21  5.79    4.13    2.45    1.66
## 6  1948  326. 2     0.162  55.4 0.200  84.7   4.01  5.79    4.13    2.45    1.66
## # ... with 1 more variable: lnx_lnP_stone_t_1 <dbl>, and abbreviated variable
## #   names 1: lnP_stone_t, 2: lnP_stone_t_1, 3: lnx_lnP_stone_t
```

2.3 Estimation

2.3.1 LA-AIDS

The LA-AIDS model estimates the following equation system:

```
##      [,1]
## [1,] w_2t ~ lnx_lnP_stone_t + lnp_1t + lnp_2t + lnp_3t + lnp_4t
## [2,] w_3t ~ lnx_lnP_stone_t + lnp_1t + lnp_2t + lnp_3t + lnp_4t
## [3,] w_4t ~ lnx_lnP_stone_t + lnp_1t + lnp_2t + lnp_3t + lnp_4t
```

One equation ($j = 4$) is dropped, as explained in section 1.2. The system is subject to the restrictions (5) on the coefficients. Here are the estimation results of the LA-AIDS model.

```
## Demand analysis with the Almost Ideal Demand System (AIDS)
## Estimation Method: LAEstimated Coefficients:
##               Estimate Std. Error    t value   Pr(>|t|)
## alpha_1    0.012755849 0.06120957  0.2083963 0.86920280
## alpha_2    0.204904154 0.03909167  5.2416320 0.12001233
## alpha_3    0.858168488 0.09520606  9.0138013 0.07033959
## alpha_4   -0.075828491 0.08919842 -0.8501103 0.55146440
```

[2] R notebook is just like Jupyter notebook, it allows for independent and interactive execution of the code chunks, and markup languages like markdown and LaTex. R notebook's (it's available only in Rstudio, you can't use it in other IDE) advantage over Jupyter notebook is that it allows inline code. In Jupyter notebook, you need to have the `python-markdown` extension for inline code, which is a non-standard syntax, it doesn't work in JupyterLab for example, so you need to be prepared to take that risk. But for collaboration, the Jupyter project is clearly doing better.

```
## beta_1     0.108276246 0.03591698  3.0146257 0.20390573
## beta_2    -0.042933585 0.02267008 -1.8938433 0.30928060
## beta_3    -0.291859665 0.05581464 -5.2290878 0.12029344
## beta_4     0.226517004 0.05224891  4.3353445 0.14431992
## gamma_11 -0.122665677 0.01703682 -7.2000322 0.08785698
## gamma_12  0.016228192 0.01056249  1.5363982 0.36732238
## gamma_13  0.084008792 0.01978032  4.2470901 0.14721413
## gamma_14  0.022428693 0.01923398  1.1660970 0.45127913
## gamma_21  0.016228192 0.01056249  1.5363982 0.36732238
## gamma_22 -0.053518273 0.02236892 -2.3925285 0.25203733
## gamma_23  0.044077707 0.01834943  2.4021293 0.25113146
## gamma_24 -0.006787626 0.01601504 -0.4238281 0.74479336
## gamma_31  0.084008792 0.01978032  4.2470901 0.14721413
## gamma_32  0.044077707 0.01834943  2.4021293 0.25113146
## gamma_33 -0.053769626 0.03762329 -1.4291580 0.38867746
## gamma_34 -0.074316873 0.03240317 -2.2935065 0.26175396
## gamma_41  0.022428693 0.01923398  1.1660970 0.45127913
## gamma_42 -0.006787626 0.01601504 -0.4238281 0.74479336
## gamma_43 -0.074316873 0.03240317 -2.2935065 0.26175396
## gamma_44  0.058675806 0.04024400  1.4580014 0.38272304
##
##  R-squared:
##       w_1t       w_2t       w_3t       w_4t
##  0.7972795  0.3419233  0.6345755 -6.0232270
```

2.3.2 AIDS

Using the strategy discussed in section 1.2.2, system (8) can be estimated using linear estimation techniques:

```
##      [,1]
## [1,] w_2t ~ lnx_lnP_t + lnp_1t + lnp_2t + lnp_3t + lnp_4t
## [2,] w_3t ~ lnx_lnP_t + lnp_1t + lnp_2t + lnp_3t + lnp_4t
## [3,] w_4t ~ lnx_lnP_t + lnp_1t + lnp_2t + lnp_3t + lnp_4t
```

The initial value $\ln P^0$ (a $T \times 1$ vector) is obtained from the LA-AIDS estimation. Convergence is reached fast:

	LA-AIDS	AIDS		
		No. of iterations = 1	No. of iterations = 5	No. of iterations = 6 (converge)
alpha_1	0.0127558	0.1364995	0.1320134	0.1320135
alpha_2	0.2049042	0.1678119	0.1673457	0.1673455
alpha_3	0.8581685	0.5700479	0.5689800	0.5689796
alpha_4	-0.0758285	0.1256407	0.1316609	0.1316613
beta_1	0.1082762	0.0868199	0.0930121	0.0930119
beta_2	-0.0429336	-0.0512893	-0.0505226	-0.0505224
beta_3	-0.2918597	-0.2989509	-0.2969064	-0.2969059
beta_4	0.2265170	0.2634203	0.2544169	0.2544163
gamma_11	-0.1226657	-0.1507864	-0.1429157	-0.1429155
gamma_12	0.0162282	0.0287326	0.0235330	0.0235330
gamma_13	0.0840088	0.1604749	0.1327353	0.1327347
gamma_14	0.0224287	-0.0384211	-0.0133526	-0.0133522
gamma_21	0.0162282	0.0287326	0.0235330	0.0235330
gamma_22	-0.0535183	-0.0542912	-0.0517749	-0.0517750
gamma_23	0.0440777	0.0142545	0.0282173	0.0282176
gamma_24	-0.0067876	0.0113042	0.0000245	0.0000245
gamma_31	0.0840088	0.1604749	0.1327353	0.1327347
gamma_32	0.0440777	0.0142545	0.0282173	0.0282176
gamma_33	-0.0537696	-0.2579512	-0.1686640	-0.1686629
gamma_34	-0.0743169	0.0832218	0.0077114	0.0077106
gamma_41	0.0224287	-0.0384211	-0.0133526	-0.0133522
gamma_42	-0.0067876	0.0113042	0.0000245	0.0000245
gamma_43	-0.0743169	0.0832218	0.0077114	0.0077106
gamma_44	0.0586758	-0.0561049	0.0056167	0.0056171

2.4 Demand Elasticities

The parameters of most demand models, such as the AIDS, do not have a straightforward interpretation, and the slopes of the demand curves depend on arbitrary units of measurement. We can present the results in terms of elasticities.

I won't give the elasticities expressions here. For the expenditure elasticities of demand and the price elasticities derived from Marshallian demand functions, see Anderson and Blundell (1983, page 400). For the price elasticities derived from Hicksian demand functions, see Deaton and Muellbauer (1980b, page 45).

The demand elasticities are calculated at sample means:

```
## Expenditure Elasticities
##        q_1        q_2        q_3        q_4
##  1.2996758  0.7478217 -1.2136504  1.7161614
##
##  Marshallian Price Elasticities
##            p_1         p_2        p_3         p_4
## q_1 -1.5015562  0.02508309  0.2594339 -0.08263667
## q_2  0.1520449 -1.21573483  0.2824080  0.03346030
## q_3  1.2931987  0.58517488 -1.0148567  0.35013350
## q_4 -0.1357946 -0.12118401 -0.3803176 -1.07886520
##
##  Hicksian (compensated) Price Elasticities
##           p_1        p_2        p_3        p_4
```

```
## q_1 -1.098169  0.2854650  0.4337530  0.3790732
## q_2  0.384150 -1.0659134  0.3827096  0.2991239
## q_3  0.916512  0.3420276 -1.1776376 -0.0810158
## q_4  0.396859  0.2226382 -0.1501374 -0.4691989
```

References

- Deaton, A., & Muellbauer, J. (1980a). An almost ideal demand system. The American economic review, 70(3), 312-326.
- Deaton, A., & Muellbauer, J. (1980b). Economics and consumer behavior. Cambridge university press.
- Anderson, G., & Blundell, R. (1983). Testing restrictions in a flexible dynamic demand system: an application to consumers' expenditure in Canada. The Review of Economic Studies, 50(3), 397-410.
- Blanciforti, L. A., Green, R. D., & King, G. A. (1986). US consumer behavior over the postwar period: an almost ideal demand system analysis. Monographs.
- Moschini, G. (1995). Units of measurement and the Stone index in demand system estimation. American journal of agricultural economics, 77(1), 63-68.
- Banks, J., Blundell, R., & Lewbel, A. (1997). Quadratic Engel curves and consumer demand. Review of Economics and statistics, 79(4), 527-539.
- Henningsen, A. (2017). Demand analysis with the "almost ideal demand system" in r: Package miceconaids. Copenhagen: Department of Food and Resource Economics, University of Copenhagen, 1-36.
- Greene, W. H. (2018). Econometric analysis (eighth edition). Pearson.
- Aguirregabiria, V. (2021). Empirical industrial organization: models, methods, and applications. University of Toronto.
- The Piglog Model