Title

---

demandsys — Estimation of flexible demand systems

Description	Quick start	Menu	Syntax
Options	Remarks and examples	Stored results	Methods and formulas
Acknowledgment	References	Also see	

Description

demandsys fits demand systems, sets of equations derived from economic theory that describe consumers' purchases of various goods or services. demandsys allows you to fit eight different demand systems, including the Cobb–Douglas system, the almost ideal demand system (AIDS) of Deaton and Muellbauer (1980b), the translog indirect utility demand system of Christensen, Jorgenson, and Lau (1975), and variants of the latter two. You can also include demographic variables that affect a consumer's or household's demands.

Quick start

Fit an AIDS demand system with four goods with expenditure shares w1, w2, w3, and w4; prices, p1, p2, p3, and p4; and total expenditure, totexp

    demandsys aids w1 w2 w3 w4, prices(p1 p2 p3 p4)   ///
        expenditure(totexp)

Same as above, reporting Marshallian (uncompensated) elasticities rather than coefficients

    demandsys aids w1 w2 w3 w4, prices(p1 p2 p3 p4)   ///
        expenditure(totexp) elasticities(uncompensated)   ///
        nocoeftable

Same as above, labeling goods in output for easier interpretation

    demandsys aids w1 w2 w3 w4, prices(p1 p2 p3 p4)   ///
        expenditure(totexp) elasticities(uncompensated)   ///
        nocoeftable labels("apples bananas carrots dates")

Fit a quadratic AIDS model with four goods, controlling for x1 and x2 using demographic translation

    demandsys quaids w1 w2 w3 w4, prices(p1 p2 p3 p4)   ///
        expenditure(totexp) demographics(x1 x2)

Same as above, but use demographic scaling rather than translating

    demandsys quaids w1 w2 w3 w4, prices(p1 p2 p3 p4)   ///
        expenditure(totexp) demographics(x1 x2, scaling)

Menu

Statistics > Linear models and related > Multiple-equation models > Demand system

Syntax

> demandsys *model* *varlist*$_s$ [*if*] [*in*] [*weight*], prices(*varlist*$_p$)
>
>   expenditure(*varname*) [*options*]

model	Description
cdouglas	Cobb–Douglas demand system
les	linear expenditure system
translog	basic translog demand system
gtranslog	generalized translog demand system
aids	almost ideal demand system (AIDS)
gaids	generalized AIDS
quaids	quadratic AIDS
gquaids	generalized quadratic AIDS

varlist$_s$ indicates the list of G variables containing the expenditure shares of the G goods in the model.

options	Description
Main	
* prices(*varlist*$_p$)	specify the variables containing prices
* expenditure(*varname*)	specify the variable containing total expenditure
demographics(*varlist*$_d$ [, scaling])	specify other variables affecting a consumer's demand for the goods; use demographic scaling instead of translating
piconstant(*#*)	specify constant term in transcendental logarithmic price index
SE/Robust	
vce(*vcetype*)	*vcetype* may be gnr, robust, cluster *clustvar*, bootstrap, or jackknife
Reporting	
elasticities(*e_type*)	report elasticities; *e_type* may be expenditure, compensated, or uncompensated
labels(*string*)	specify labels for goods
level(*#*)	set confidence level; default is level(95)
noheader	suppress the summary header at the top of the output
nocoeftable	do not display the table of estimated coefficients
display_options	control columns and column formats and line width
Optimization	
optimization_options	control the optimization process; seldom used
coeflegend	display legend instead of statistics

*prices() and expenditure() are required.

collect is allowed; see [U] **11.1.10 Prefix commands**.

aweights, fweights, iweights, and pweights are allowed; see [U] **11.1.6 weight**.

coeflegend does not appear in the dialog box.

See [U] **20 Estimation and postestimation commands** for more capabilities of estimation commands.

Options

    ◻ Main

prices(*varlist$_p$*) specifies a list of G variables corresponding to the prices of the G goods faced by each consumer. The number of variables specified here must match the number specified in *varlist$_s$*, and the price variables must be specified in the same order as the share variables. All the price variables must be strictly positive for all the demand systems implemented. prices() is required.

expenditure(*varname*) specifies the variable corresponding to the total expenditure on all goods within the system by each consumer. This variable must be strictly positive. expenditure() is required.

demographics(*varlist$_d$* [, scaling]) specifies one or more demographic variables that affect each consumer's demand for the goods in the system. Suboption scaling, available only with models aids and quaids, requests that demographics be incorporated via demographic scaling rather than demographic translation, the default.

piconstant(*#*) specifies the value of the constant term in the transcendental logarithmic price index; by default, this is set to the logarithm of the minimum expenditure observed in the sample. This option is relevant only for models aids and quaids.

    ◻ SE/Robust

vce(*vcetype*) specifies the type of standard error reported, which includes types that are derived from asymptotic theory (gnr), that are robust to some kinds of misspecification (robust), that allow for intragroup correlation (cluster *clustvar*), and that use bootstrap or jackknife methods (bootstrap, jackknife); see [R] *vce_option*.

vce(gnr), the default, uses the conventionally derived variance estimator for nonlinear models fit using Gauss–Newton regression.

    ◻ Reporting

elasticities(*e_type*) requests that elasticities be reported instead of, or in addition to, the parameter estimates. *e_type* may be expenditure, compensated, or uncompensated. expenditure computes demand elasticity to changes in expenditure. compensated computes demand elasticity to changes in prices, ignoring income effects. These elasticities are also known as Hicksian price elasticities. uncompensated computes demand elasticity to changes in prices. These elasticities are also known as Marshallian price elasticities. The elasticities are computed at the estimation sample means of the prices, expenditures, and any demographic variables specified. For more flexibility in obtaining elasticities, use the postestimation command estat elasticities.

labels(*string*) specifies a set of names with which the goods are to be labeled. By default, if you specify a four-good demand system, then the goods will be labeled generically: "Good 1", "Good 2", "Good 3", and "Good 4".

If you specify `labels("shelter fuel food other")`, then the four goods will be labeled "shelter", "fuel", "food", and "other" in the output. If you specify a demand system with G goods, then you must supply G labels separated by spaces.

`level(#)`; see [R] **Estimation options**.

`noheader` requests that the header summarizing the model, estimation sample, and other statistics not be shown in the output.

`nocoeftable` requests that the table containing the parameter estimates, their standard errors, and so on not be displayed. Typically, you would use this option if you specify the `elasticities()` option.

display_options: `noci`, `nopvalues`, `cformat(%fmt)`, `pformat(%fmt)`, `sformat(%fmt)`, and `nolstretch`; see [R] **Estimation options**.

⌐ Optimization ⌐

optimization_options: <u>iter</u>ate(#), [no]log, <u>tra</u>ce, eps(#), ifgnlsiterate(#), ifgnlseps(#), and delta(#).

`iterate(#)` specifies the maximum number of iterations to use for nonlinear least squares at each round of feasible generalized nonlinear least-squares (FGNLS) estimation. The default is the number set using `set maxiter`, which is 300 by default.

`log` and `nolog` specify whether to display the iteration log. The iteration log is displayed by default unless you used `set iterlog off` to suppress it; see `set iterlog` in [R] *set iter*.

`trace` specifies that the iteration log should include the current parameter vector.

`eps(#)` specifies the convergence criterion for successive parameter estimates and for the residual sum of squares. The default is `eps(1e-5)` (0.00001). `eps()` also specifies the convergence criterion for successive parameter estimates between rounds of iterative FGNLS.

`ifgnlsiterate(#)` specifies the maximum number of FGNLS iterations to perform. The default is the number set using `set maxiter`, which is 300 by default.

`ifgnlseps(#)` specifies the convergence criterion for successive estimates of the error covariance matrix during iterative FGNLS estimation. The default is `ifgnlseps(1e-10)`.

`delta(#)` specifies the relative change in a parameter, δ, to be used in computing the numeric derivatives. The derivative for parameter $beta_i$ is computed as

$$\{f_i(\mathbf{x}_i, \beta_1, \beta_2, \dots, \beta_i + d, \beta_{i+1}, \dots) - f_i(\mathbf{x}_i, \beta_1, \beta_2, \dots, \beta_i, \beta_{i+1}, \dots)\}/d$$

where $d = \delta(|\beta_i| + \delta)$. The default is `delta(4e-7)`.

The following option is available with `demandsys` but is not shown in the dialog box:

`coeflegend`; see [R] **Estimation options**.

Remarks and examples   stata.com

Remarks are presented under the following headings:

Introduction

`demandsys` fits demand systems, sets of equations derived from economic theory that describe consumers' purchases of various goods or services. Typically, you will have a large cross-sectional survey containing consumers' data on their expenditures on various goods and services along with the prices paid for them. We often refer to the items as "goods" for brevity, but of course some of the items may be services. Whether they are physical goods or intangible services, in the context of utility-maximizing consumers, they are both goods in the sense that "more is better". We also use the term "purchase" somewhat loosely; in some cases, what we are interested in is not the actual purchase of goods but rather their consumption over a fixed time period.

To fit a demand system, you must first decide on what is in the consumption basket or bundle and the set of goods whose demands you wish to model. Using the parameters from the model, you may then obtain the elasticities—the effects of changes in prices or changes in expenditure on demand of goods. You may also perform welfare analysis by contrasting changes in demand or utility that occur at different price or expenditure levels. We assume that you have decided which consumption bundle to model already or else that you have several alternative baskets to model and compare.

The left-hand-side variables you specify with `demandsys` are expenditure shares, the shares of total expenditure spent on goods or services. Given p_g (the price of good g), q_g (the quantity of good g purchased), and m (the consumer's total expenditure on all goods within the demand system), the expenditure share for good g is defined as

$$w_g = \frac{p_g \, q_g}{m}$$

`demandsys` requires that you have the prices of all the goods and that you have the total expenditure across all the goods. `demandsys` does not need the quantities of the goods purchased, though you may need them to calculate the expenditure shares.

Consumer theory in microeconomics presents demand models in terms of quantities. In demand system analysis, we model expenditure shares, not quantities. By our definition of total expenditure, $0 \le w_g \le 1$ for all g, and $\sum_g w_g = 1$. For all but the simplest utility or cost functions, the algebra to obtain expenditure shares is arguably easier than that to obtain quantities. `demandsys` checks your expenditure shares for all goods and will exit with an error message if an expenditure share is found outside that range or if the sum is not equal to one (allowing for small rounding errors).

You may also specify additional variables that may affect a consumer's purchase decisions. These are often demographic variables, such as the number of children and adults in a household, or a set of indicator variables to denote the region of a country in which a consumer is located. If you have data that were collected over the course of different months or years, you may wish to include indicators to control for the time at which the consumer's purchases were observed. We call all of these variables demographics, even if some of these variables are not really demographic characteristics.

Based on the theoretical considerations in Lewbel (2001), we strongly encourage you to include demographic variables in your model. He shows that econometrically estimated demand functions will not satisfy rationality conditions unless other variables that affect demand and are correlated with expenditures are included. Conversely, if we include all such variables in our model, then the estimated demands will satisfy rationality even if preferences vary among households. On the other hand, you should also bear in mind that including too many demographic variables can greatly increase the number of parameters in your model, especially if your demand system includes many goods.

demandsys offers just a small sampling of demand systems that have been proposed in the literature. As a practical matter, which one should you use? We have included the Cobb–Douglas model mainly for pedagogical purposes because most students are familiar with Cobb–Douglas utility, though soon thereafter they learn about its severe restrictions on consumer behavior. Stone's (1954) linear expenditure system (LES) is historically important as an early demand system but also places somewhat strong restrictions on consumer behavior. Christensen, Jorgenson, and Lau's (1975) translog model relaxes some of the LES's restrictions. It is not as widely used as some of the newer models, though a very similar translog production function those authors proposed continues to see extensive use. Shortly after the translog model arrived, Deaton and Muellbauer (1980b) proposed their more flexible AIDS, which has been a workhorse model of demand system analysis ever since its arrival. Banks, Blundell, and Lewbel (1997) provided a quadratic extension to AIDS that is also popular.

Banks, Blundell, and Lewbel (1997) suggest using QUAIDS because of its flexibility. Moreover, QUAIDS satisfies certain theoretical properties developed in Gorman (1981); and it allows goods to be either luxuries or necessities depending on a consumer's income, which Banks, Blundell, and Lewbel (1997) show to be important in the data they consider.

With the AIDS and QUAIDS models, you can include demographics via two different methods: demographic translation and demographic scaling; see *Demographic translation* and *Demographic scaling*. If you believe that subsistence or committed quantities are appropriate for the goods you are modeling, you can use what we call the generalized QUAIDS model. If you do that, you should check to see that the subsistence quantity interpretation holds; we do that in example 2 of [R] **demandsys postestimation**. You can also do a simple likelihood-ratio test to see whether the quadratic terms of these QUAIDS models increase their explanatory power over Deaton and Muellbauer's AIDS and its generalized variant; see example 5 below.

Throughout the rest of the discussion, we assume that you are familiar with a few basic microeconomic concepts, including utility maximization, the expenditure function, and the indirect utility function. The presentations in standard texts like Varian (1992, chap. 7–10) and Mas-Colell, Whinston, and Green (1995, chap. 3) provide good introductions to these concepts. The classic monograph by Deaton and Muellbauer (1980a) and the book by Pollak and Wales (1992) describe demand system analysis in much greater detail than space permits us here. More recent survey papers include Holt and Goodwin (2009) and Barnett and Serletis (2008). Fisher, Fleissig, and Serletis (2001) compare many flexible functional forms for demand system analysis, including several that are implemented by demandsys.

Some notation

It will be useful to establish at the outset some standard notation we will use throughout this discussion. We use subscript $i = 1, \ldots, N$ to index observations, $g = 1, \ldots, G$ to index goods, and $d = 1, \ldots, D$ to index demographic variables. In formulas where we must use double summations across goods or else have used subscript g elsewhere, we also use $h = 1, \ldots, G$ and $j = 1, \ldots, G$ to index goods. When no confusion could arise, we omit the observation subscript to reduce the number of subscripts.

Let w_g be the expenditure share for good g defined as $w_g \equiv (p_g \, q_g)/m$, where p_g is the price of good g, q_g is the quantity of good g consumed, and m denotes total expenditure on all the goods in the system being modeled. We use the notation $\mathbf{p}$ to refer to the G-length vector of all G prices. When we refer to w_g, we are referring to the observed expenditure share for good g for consumer i. When we refer to, say, $w_g(\mathbf{p}, m; \boldsymbol{\alpha}, \boldsymbol{\beta})$, we are referring to an expenditure-share equation or function for good g that depends on the prices of all the goods and total expenditure as well as parameter vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. We include the parameter vectors and matrices in the arguments for expenditure shares as well as some other functions to emphasize that they are estimated and will appear in the output from demandsys and that statistics available via predict or estat elasticities after estimation depend on those parameters.

The observed w_g is assumed to be related to the expenditure-share function $w_g(\mathbf{p}, m; \boldsymbol{\alpha}, \boldsymbol{\beta})$ as

$$w_g = w_g(\mathbf{p}, m; \boldsymbol{\alpha}, \boldsymbol{\beta}) + \epsilon_g$$

where ϵ_g is a zero-mean error term that we discuss in more detail in *Methods and formulas*.

Cobb–Douglas

Although widely used in many economic models, the Cobb–Douglas utility function is arguably too simple for serious demand system analysis because of its severe restrictions on the expenditure-share equations and elasticities. For instance, it restricts expenditure elasticities to be identically equal to 1 for all goods. Nevertheless, we include it because it serves as a good starting point for our discussion and allows us to present various aspects of demandsys. For three goods, consumers maximize their Cobb–Douglas utility function subject to the constraint that total expenditure does not exceed the allocated budget,

$$\max_{q_1, q_2, q_3} q_1^{\alpha_1} q_2^{\alpha_2} q_3^{(1-\alpha_1-\alpha_2)}$$
$$\text{subject to } p_1 q_1 + p_2 q_2 + p_3 q_3 \leq m$$

where, without loss of generality, we have made the sum of the exponents of the Cobb–Douglas utility function sum to 1. It is easy to show that the optimal quantities q_1^*, q_2^*, and q_3^* are

$$q_1^* = \alpha_1 \frac{m}{p_1} \qquad q_2^* = \alpha_2 \frac{m}{p_2} \qquad q_3^* = (1 - \alpha_1 - \alpha_2) \frac{m}{p_3}$$

To obtain the expenditure-share function of, say, good 1, we have

$$w_1(\mathbf{p}, m; \boldsymbol{\alpha}) = \frac{p_1 \, q_1}{m} = \alpha_1 \frac{m}{p_1} \frac{p_1}{m} = \alpha_1$$

and likewise for goods 2 and 3. Notice that the expenditure-share function for good g is simply its coefficient α_g in the Cobb–Douglas utility function, or, equivalently, the coefficients in the Cobb–Douglas utility function are simply the shares of total expenditure allocated to each good. Most importantly, the expenditure shares for the Cobb–Douglas model are not functions of prices, income, or any other variables.

We made the sum of the exponents in the Cobb–Douglas utility function sum to 1, and it is clear that $w_1 + w_2 + w_3 = 1$. Without the constraint on the sum, we would have to carry around the divisor $\alpha_1 + \alpha_2 + \alpha_3$ in our expenditure-share equations, but we would gain absolutely no additional insight. Moreover, with the constraint we need estimate only two parameters rather than three.

More generally, for a Cobb–Douglas utility function of the form

$$u(\mathbf{q}; \boldsymbol{\alpha}) = \prod_{g=1}^{G} q_g^{\alpha_g} \qquad \text{with} \qquad \sum_{g=1}^{G} \alpha_g = 1$$

the expenditure-share function for the gth good is simply $w_g(\mathbf{p}, m; \boldsymbol{\alpha}) = \alpha_g$.

▷ Example 1

We first describe the dataset we will use in this example and all others in this manual entry:

```
. use https://www.stata-press.com/data/r18/food_consumption
(Food consumption)

. describe

Contains data from https://www.stata-press.com/data/r18/food_consumption.dta
 Observations:          4,160                  Food consumption
    Variables:             13                  17 Jul 2022 16:03
                                               (_dta has notes)

Variable        Storage   Display    Value
    name           type    format    label     Variable label

w_dairy          float    %10.6f               Expenditure share on dairy
w_proteins       float    %10.6f               Expenditure share on meats and
                                                 proteins
w_fruitveg       float    %10.6f               Expenditure share on fruits and
                                                 vegetables
w_flours         float    %10.6f               Expenditure share on flours,
                                                 breads, pasta, and cereals
w_misc           float    %10.6f               Expenditure share on misc. food
                                                 items
p_dairy          float    %10.6f               Price of dairy
p_proteins       float    %10.6f               Price of meats and proteins
p_fruitveg       float    %10.6f               Price of fruits and vegetables
p_flours         float    %10.6f               Price of flours, breads, pasta,
                                                 and cereals
p_misc           float    %10.6f               Price of misc. food items
expfd            float    %10.6f               Total expenditure on all food
                                                 categories
n_adults         byte     %8.0g                # adults in household
n_kids           byte     %8.0g                # kids in household

Sorted by:
```

The observation level of this dataset is the household, and we have data on 4,160 households. The data include five categories of food: dairy products; proteins, including meats and fish; fruits and vegetables; flour-based products, including breads, pastas, and cereals; and a catchall category. For each of the five categories, we have the expenditure shares and prices, and we have the household's total expenditure on all five categories in the week that the household was surveyed. We also have demographic data for the household, including the numbers of children and adults.

Although we constructed the expenditure shares ourselves and know they sum to one, it is a good idea to verify that you have created your expenditure-share variables properly:

```
. egen wsum = rowtotal(w_dairy w_proteins w_fruitveg w_flours w_misc)

. summarize wsum
```

Variable	Obs	Mean	Std. dev.	Min	Max
wsum	4,160	1	4.40e-08	.9999999	1

demandsys will check that for you and exit with an error message if it is not the case, but doing so during your data management tasks may make debugging easier. demandsys will also check that all your price variables and expenditure variable are strictly positive, though again you might want to check those facts earlier in your data pipeline.

With a dataset in hand, we are ready to fit our first Cobb–Douglas demand system.

```
. demandsys cdouglas w_dairy w_proteins w_fruitveg w_flours w_misc,
> prices(p_dairy p_proteins p_fruitveg p_flours p_misc) expenditure(expfd)

Calculating NLS estimates:
Iteration 0:   Residual SS = 180.2506
Iteration 1:   Residual SS = 180.2506

Calculating FGNLS estimates:
Iteration 0:   Scaled RSS =     16640

FGNLS iteration 2:
Iteration 0:   Scaled RSS =     16640
Parameter change        = 0.00e+00
Covariance matrix change = 0.00e+00
```

```
Cobb–Douglas model                              Number of obs   = 4,160
Expenditure variable: expfd                     Number of goods =     5

Uncentered R2 of model for
  Good 1 = 0.7519
  Good 2 = 0.8851
  Good 3 = 0.8329
  Good 4 = 0.7467
  Good 5 = 0.6836
```

		Estimate	Std. err.	z	P>|z|	[95% conf. interval]
alpha							
	Good						
	1	.1505844	.0013411	112.29	0.000	.147956	.1532129
	2	.3986526	.0022265	179.05	0.000	.3942887	.4030166
	3	.2406533	.0016711	144.01	0.000	.237378	.2439286
	4	.1031129	.0009312	110.73	0.000	.1012877	.104938

Note: **alpha** estimates are expenditure shares.

Normalized parameter

		Estimate	Std. err.	z	P>|z|	[95% conf. interval]
alpha							
	Good						
	5	.1069968	.0011286	94.80	0.000	.1047847	.1092088

Note: **alpha** estimates sum to 1.

```
Shares: w_dairy w_proteins w_fruitveg w_flours w_misc
Prices: p_dairy p_proteins p_fruitveg p_flours p_misc
Expenditure: expfd
```

We specified the command name, demandsys, followed by the model we wish to fit, cdouglas. We then specified the five expenditure-share variables, w_dairy through w_misc. The price variables, p_dairy through p_misc, go into the prices() option, and the expenditure variable, expfd, goes into the expenditure() option. demandsys uses a nonlinear multiple-equation estimator, so the top of the output includes an iteration log showing the model's convergence. In fact, the Cobb–Douglas model is linear, but that is the only model fit by demandsys that is. Stata uses the same nonlinear estimator for cdouglas as we use for the other demand systems. The header of the output includes R^2 values to get some idea of how well the model fits the data.

The table of coefficients is a bit different from most other estimators in Stata, which report coefficients organized by equations. The Cobb–Douglas model is unique in that each equation has its own parameter. However, that is not the case for any other demand system that demandsys fits. For most demand systems, many of the parameters appear in multiple equations, and there is no way to link parameters and equations. Hence, we cannot provide a coefficient table organized by equations as other multiple-equation estimators such as sureg provide. Instead, demandsys organizes estimates by parameter type.

We mentioned having a five-good demand system, so why do we see only four estimated parameters in the main output table? Recall that we used the normalization that the sum of the parameters in the Cobb–Douglas utility function is one. The upshot is that we really have only four free parameters because the fifth parameter must equal one minus the sum of the other four parameters. We provide the estimate of the fifth parameter in a separate table. It is computed separately using the parameter estimates of our fitted demand system and their standard errors. Given how the fifth parameter was computed, it does not share a covariance matrix with the other parameters and cannot be used for testing. We provide it here for reference.

demandsys shows only unconstrained parameters in the main estimation output because, with many goods and more complicated models, the output becomes long. Moreover, the estimated parameters themselves are of less interest than elasticities and other statistics derived from the fitted model.

❑ Technical note

In the header of the output in the previous example, we see that the equation-level R^2 values are labeled as uncentered. Recall that the standard R^2 in regression analysis measures the ability of the model to predict the regressand compared with a model that contains just a constant term. For the Cobb–Douglas demand system, the expenditure shares are simply constant terms to be estimated. Hence, the traditional R^2 is by definition zero. We therefore report the uncentered R^2, which is an alternative that measures the model's explanatory power as a fraction of the sum of squares of an equation's left-hand-side variable. A higher uncentered R^2 is presumably better than a lower uncentered R^2, but you cannot compare the uncentered R^2 of a Cobb–Douglas share equation with the centered R^2 of a share equation from a different demand system.

❑

The command estat elasticities, which is available after demandsys, allows you to calculate expenditure and price elasticities after fitting your demand system, and estat elasticities provides options to specify how those elasticities are calculated and for which observations. The full syntax for that command is listed in [R] **demandsys postestimation**, but the basic command is easy to pick up as we work through examples below.

Here we use `estat elasticities` to obtain the expenditure elasticities for the estimation sample.

```
. estat elasticities, expenditure
Expenditure elasticities                          Number of obs = 4,160
```

Expenditure	Elasticity
Good	
1	1
2	1
3	1
4	1
5	1

```
Note: No standard errors are displayed because all elasticities are
      identically equal to one.
```

The expenditure elasticities for all five goods are identically equal to one, and so the standard errors, test statistics, and confidence interval are not displayed because there is no sampling variance. The Cobb–Douglas function is an example of a utility function with "homothetic preferences", meaning that the ratio of two goods demanded by a consumer with such preferences depends only on the goods' relative prices and not on income. Homothetic preferences also imply that, regardless of changes in income, the shares consumed of each good remain the same; the expansion path of consumption is linear in income. The Cobb–Douglas case is a particularly extreme example, where the expenditure elasticities are one, the uncompensated own-price elasticities are minus one, and uncompensated cross-price elasticities are zero, which you can verify by typing

```
estat elasticities, uncompensated
```

Particularly with more complex demand systems, the coefficients themselves can be more difficult to interpret. Therefore, `demandsys` allows you to request that expenditure, uncompensated price, or compensated price elasticities be displayed in addition to, or in place of, the table of estimated coefficients. The elasticities reported directly by `demandsys` are calculated at the estimation sample means of prices, expenditures, and any demographics that you specify. To obtain, say, the uncompensated price elasticities instead of the coefficient table from a Cobb–Douglas demand system, you would type

```
demandsys cdouglas ..., prices(...) expenditure(...)            ///
                        elasticities(uncompensated) nocoeftable
```

`estat elasticities` gives you much more flexibility in terms of the sample used, and you can even specify particular prices or expenditures at which you want the elasticities to be calculated.

◁

Linear expenditure system (LES)

Nobel laureate Sir Richard Stone's estimation of the LES, developed based on theory by Nobel laureates Paul Samuelson and Lawrence Klein, together with Herman Rubin, represents the genesis of flexible demand system estimation (Stone 1954). LES provides for more flexible consumption patterns than the Cobb–Douglas model. The LES begins with the utility function

$$u(\mathbf{q}; \boldsymbol{\beta}, \boldsymbol{\mu}) = \prod_g (q_g - \mu_g)^{\beta_g}$$

where μ_g is the gth element of $G \times 1$ vector $\boldsymbol{\mu}$, which is to be estimated. This utility function results from translating or shifting the origin of the Cobb–Douglas utility function; utility does not accrue from good g unless its consumption exceeds an amount μ_g. That parameter is known as a subsistence or committed quantity that a consumer must purchase for survival. β_g is the gth element of $G \times 1$ vector $\boldsymbol{\beta}$. β_g measures the share of supernumerary expenditure, namely, the share of remaining expenditure once all subsistence quantities have been purchased, on good g. As in the Cobb–Douglas case, we normalize the $G \times 1$ parameter vector $\boldsymbol{\beta}$ so that $\sum_g \beta_g = 1$. Thus, the model has $2G - 1$ parameters we must estimate.

Solving

$$\max_{\mathbf{q}} \prod_g (q_g - \mu_g)^{\beta_g} \quad \text{s. t.} \quad \mathbf{p}'\mathbf{q} \le m$$

yields the expenditure-share functions

$$w_g(\mathbf{p}, m; \boldsymbol{\beta}, \boldsymbol{\mu}) = \frac{p_g \mu_g}{m} + \beta_g \left(1 - \sum_h \frac{p_h \mu_h}{m}\right)$$

The term to the left of the plus sign is the fraction of expenditure that must be spent on good g to maintain subsistence. To the right, the term bound in parentheses is the fraction of expenditure left after the subsistence level for all other goods has been spent. Of this amount, a share β_g is spent on good g. The normalization that $\sum_g \beta_g = 1$ ensures that $\sum_g w_g(\mathbf{p}, m) = 1$, as must be true of a demand system.

While we have interpreted the μ_g as minimum required amounts of each good, there is no requirement that $\mu_g > 0$ for all g, nor do we impose any such constraints during estimation. Moreover, for this interpretation to hold we must have that for each household i, $\sum_g p_{gi} \mu_{gi} \le m_i$.

▷ Example 2

Here we fit an LES to the same data as in example 1.

```
. use https://www.stata-press.com/data/r18/food_consumption, clear
(Food consumption)
. demandsys les w_dairy w_proteins w_fruitveg w_flours w_misc,
> prices(p_dairy p_proteins p_fruitveg p_flours p_misc)
> expenditure(expfd) nolog
Calculating NLS estimates ...
Calculating FGNLS estimates ...
FGNLS iteration 2 ...
FGNLS iteration 3 ...
FGNLS iteration 4 ...
FGNLS iteration 5 ...
```

Linear expenditure system	Number of obs = 4,160
Expenditure variable: expfd	Number of goods = 5

Centered R2 of model for
  Good 1 = -0.0508
  Good 2 = 0.0494
  Good 3 = 0.0366
  Good 4 = 0.0829
  Good 5 = 0.0603

		Estimate	Std. err.	z	P>|z|	[95% conf. interval]	
beta							
	Good						
	1	.1650894	.0018783	87.89	0.000	.161408	.1687708
	2	.4108231	.0033135	123.98	0.000	.4043288	.4173175
	3	.2262407	.0024899	90.86	0.000	.2213607	.2311208
	4	.0949082	.0013152	72.16	0.000	.0923304	.097486
mu							
	Good						
	1	.1794068	.0383858	4.67	0.000	.1041721	.2546416
	2	.6975436	.0800327	8.72	0.000	.5406824	.8544048
	3	2.368491	.1351367	17.53	0.000	2.103628	2.633354
	4	.4215745	.0202499	20.82	0.000	.3818854	.4612636
	5	.3742182	.0243646	15.36	0.000	.3264644	.421972

Note: **beta** estimates measure how expenditure shares respond to increases in
     supernumerary income.
Note: **mu** estimates are subsistence levels of consumption for each good.

Normalized parameter

		Estimate	Std. err.	z	P>|z|	[95% conf. interval]	
beta							
	Good						
	5	.1029385	.0015833	65.02	0.000	.0998354	.1060416

Note: **beta** estimates sum to 1.

Shares: w_dairy w_proteins w_fruitveg w_flours w_misc
Prices: p_dairy p_proteins p_fruitveg p_flours p_misc
Expenditure: expfd

We specified the nolog option to suppress the detailed iteration log. We first noticed that the R^2 for the model for Good 1 is actually negative. Because the equations here are nonlinear, the value of R^2 is not constrained to be between 0 and 1; that is only true for linear regression. The footer of the table with normalized parameters reemphasizes the fact that we have made $\sum_g \beta_g = 1$.

All the μ parameters are greater than zero. The μ parameter for Good 3, fruits and vegetables, is largest, implying the physical quantity of them required is much higher than the other four goods. However, we should also look at the average prices of the goods:

```
. summarize p_*
    Variable │        Obs        Mean    Std. dev.         Min         Max
─────────────┼─────────────────────────────────────────────────────────────
     p_dairy │      4,160    .4387958    .4238236    .1041262          18
  p_proteins │      4,160    1.729284      .69554    .3333333    9.258823
   p_fruitveg │      4,160    .5472199    .1945325    .1073369    2.672269
    p_flours │      4,160     1.49421    .6816424    .0805687    7.202127
      p_misc │      4,160     1.47115    .8588024    .1428571    15.74629
```

Based on the average price of each good, the dollar amount of protein required for survival is only 9 cents less than the dollar amount of fruits and vegetables required ($2.368 \times 0.547 = \$1.30$ versus $0.698 \times 1.729 = \$1.21$).

◁

As we mentioned in the introductory remarks, if you have demographic variables, you should incorporate them into your model. demandsys incorporates demographics into models by applying Pollak and Wales's (1978) "demographic translation" to the expenditure-share equations. Just as the linear expenditure system is a translated version of the Cobb–Douglas system, Pollak and Wales incorporate demographics by translating the consumer's available level of expenditures.

Suppose for each consumer we have a $D \times 1$ vector of demographic characteristics $\mathbf{d}$. We have G expenditure-share equations, so associated with $\mathbf{d}$ is a $G \times D$ matrix of parameters $\mathbf{N}$ with typical row $\boldsymbol{\nu}_g$. Let $\mathbf{c} = \mathbf{Nd}$ with typical element c_g. In demographic translation, we write each translated demand function as

$$\overline{x}_g(\mathbf{p}, m, \mathbf{c}) = c_g + x_g\left(\mathbf{p}, m - \sum_h p_h c_h\right)$$

For the LES, we have

$$\overline{x}_g(\mathbf{p}, m, \mathbf{d}; \boldsymbol{\beta}, \boldsymbol{\mu}, \mathbf{N}) = (\mu_g + c_g) + \frac{\beta_g}{p_g}\left(m - \sum_h p_h \mu_h - \sum_h p_h c_h\right)$$

so that

$$\overline{w}_g(\mathbf{p}, m, \mathbf{d}; \boldsymbol{\beta}, \boldsymbol{\mu}, \mathbf{N}) = \frac{p_g(\mu_g + \boldsymbol{\nu}_g \mathbf{d})}{m} + \beta_g\left\{1 - \frac{\sum_h p_h(\mu_h + \boldsymbol{\nu}_h \mathbf{d})}{m}\right\}$$

For the LES, the effect of the demographic variables is to adjust the μ parameters, though we must be cautious in interpreting the term $\mu_g + \boldsymbol{\nu}_g \mathbf{d}$ as a minimum required quantity because there is nothing preventing it from being negative.

▷ Example 3

The dataset described in example 1 also includes two demographic characteristics, the numbers of children and adults in each household. Here we refit the LES to these data, controlling for these two demographics.

```
. use https://www.stata-press.com/data/r18/food_consumption
(Food consumption)

. demandsys les w_dairy w_proteins w_fruitveg w_flours w_misc,
> prices(p_dairy p_proteins p_fruitveg p_flours p_misc)
> expenditure(expfd) demographics(n_kids n_adults) nolog

Calculating NLS estimates ...
Calculating FGNLS estimates ...
FGNLS iteration 2 ...
FGNLS iteration 3 ...
FGNLS iteration 4 ...
FGNLS iteration 5 ...
```

Linear expenditure system   Number of obs   =   4,160
Expenditure variable: expfd   Number of goods   =   5
Demographic method: Translating   Number of demographics =   2

Centered R2 of model for
  Good 1 = -0.0444
  Good 2 = 0.0665
  Good 3 = 0.0584
  Good 4 = 0.1576
  Good 5 = 0.0922

	Estimate	Std. err.	z	P>|z|	[95% conf. interval]	
beta						
Good						
1	.1656925	.0023639	70.09	0.000	.1610594	.1703256
2	.4170346	.0042631	97.82	0.000	.408679	.4253901
3	.2434237	.0032072	75.90	0.000	.2371377	.2497098
4	.0800666	.0015889	50.39	0.000	.0769525	.0831808
mu						
Good						
1	-.3932844	.164016	-2.40	0.016	-.71475	-.0718189
2	-.53048	.1798223	-2.95	0.003	-.8829251	-.1780349
3	1.010915	.3046076	3.32	0.001	.4138951	1.607935
4	-.0269827	.0426005	-0.63	0.526	-.1104782	.0565128
5	.0097685	.0526521	0.19	0.853	-.0934277	.1129647
Nu						
Good# c.n_kids						
1	.4332196	.160145	2.71	0.007	.1193413	.747098
2	-.0680289	.1197109	-0.57	0.570	-.302658	.1666002
3	-.7945652	.2003774	-3.97	0.000	-1.187298	-.4018328
4	.3923541	.0273698	14.34	0.000	.3387103	.4459979
5	.3563529	.0368047	9.68	0.000	.284217	.4284888
Good# c.n_adults						
1	.4856673	.1525998	3.18	0.001	.1865771	.7847574
2	.8669319	.1316001	6.59	0.000	.6090005	1.124863
3	.843552	.227083	3.71	0.000	.3984775	1.288626
4	.3576213	.0318802	11.22	0.000	.2951373	.4201054
5	.2639693	.0374888	7.04	0.000	.1904925	.3374461

```
Note: beta estimates measure how expenditure shares respond to increases in
      supernumerary income.
Note: mu estimates are subsistence levels of consumption for each good.
Note: Nu estimates measure the effect of demographic variables on subsistence
      levels of consumption.
```

Normalized parameter

	Estimate	Std. err.	z	P>|z|	[95% conf. interval]
beta					
Good					
5	.0937826	.0019264	48.68	0.000	.090007 .0975582

```
Note: beta estimates sum to 1.
Shares: w_dairy w_proteins w_fruitveg w_flours w_misc
Prices: p_dairy p_proteins p_fruitveg p_flours p_misc
Expenditure: expfd
Demographics: n_kids n_adults
```

The estimated parameter matrix N is shown in the last block of the coefficient table, and demandsys labels the coefficients using factor-variable notation. Yet you should not think of the terms as you would of interactions in a regression. Here the notation means something different. The first five rows of the output for N correspond to the demographic variable n_kids and how it affects the consumption of each good. The remaining five correspond to n_adults. For instance, all the coefficients for n_adults are positive, as we would expect, meaning that having more adults in the family increases consumption of all goods. The coefficient for n_kids on the consumption of Good 3, fruits and vegetables, is negative, meaning consumption decreases as the number of children increases.

Directly interpreting the coefficients on demographic variables is not difficult for the LES, but for more complicated models, it can be. One easy way to see the practical effect of that negative coefficient is to use the margins command to see how our predicted quantity for Good 3 changes as the number of children changes.

```
. margins, predict(quantities equation(#3)) at(n_kids=1 n_kids=2 n_kids=3)
Predictive margins                                   Number of obs = 4,160
Model VCE: GNR

Expression: Predicted quantity of good 3, predict(quantities equation(#3))
1._at: n_kids = 1
2._at: n_kids = 2
3._at: n_kids = 3
```

	Margin	Delta-method std. err.	z	P>|z|	[95% conf. interval]
_at					
1	22.57831	.1734791	130.15	0.000	22.2383 22.91833
2	21.407	.2254155	94.97	0.000	20.96519 21.84881
3	20.23569	.323032	62.64	0.000	19.60256 20.86882

We asked margins to produce predictions of quantities of Good 3, fruits and vegetables, by specifying the equation(#3) option, and we asked it to do so when the number of children is equal to one, again when the number of children is equal to two, and finally when the number of children is equal to three. margins first sets n_kids equal to one for all 4,160 observations in the estimation sample and obtains the predicted quantities; the mean of those predictions is 22.6. When n_kids is equal to two for all observations, the mean is 21.4, and when n_kids is equal to three, the mean is 20.2.

Kids do not always eat their fruits and vegetables, but finding that the quantity of fruits and vegetables actually declines as the household grows is rather surprising. Perhaps we have omitted

other relevant demographic variables from our model, biasing our estimates, or perhaps the model itself is simply too rigid.

Finally, we use `estat elasticities` to obtain uncompensated price elasticities. We will obtain two sets of results, one for households with fewer than three kids and one for households with three or more kids.

```
. estat elasticities if n_kids <= 2, uncompensated
Uncompensated (Marshallian) price elasticities          Number of obs = 3,742
```

Price		Elasticity	Std. err.	z	P>|z|	[95% conf. interval]	
Good 1							
	Good						
	1	-.9537446	.0083361	-114.41	0.000	-.970083	-.9374061
	2	-.00852	.0015032	-5.67	0.000	-.0114662	-.0055738
	3	-.0083598	.0014808	-5.65	0.000	-.0112621	-.0054574
	4	-.0062373	.0011228	-5.56	0.000	-.0084379	-.0040367
	5	-.0071346	.0012983	-5.50	0.000	-.0096793	-.00459
Good 2							
	Good						
	1	-.0621117	.012096	-5.13	0.000	-.0858194	-.038404
	2	-.9289446	.0085912	-108.13	0.000	-.945783	-.9121063
	3	-.0499231	.0062304	-8.01	0.000	-.0621346	-.0377117
	4	-.0390871	.0048967	-7.98	0.000	-.0486844	-.0294897
	5	-.0445883	.0056323	-7.92	0.000	-.0556274	-.0335492
Good 3							
	Good						
	1	-.0503597	.0118715	-4.24	0.000	-.0736274	-.0270921
	2	-.0410463	.0036302	-11.31	0.000	-.0481614	-.0339312
	3	-.886701	.0100534	-88.20	0.000	-.9064052	-.8669968
	4	-.030438	.0027264	-11.16	0.000	-.0357816	-.0250944
	5	-.0348829	.0032109	-10.86	0.000	-.0411761	-.0285897
Good 4							
	Good						
	1	-.0428307	.0038014	-11.27	0.000	-.0502814	-.0353801
	2	-.0373442	.0014421	-25.90	0.000	-.0401707	-.0345178
	3	-.0357288	.0013809	-25.87	0.000	-.0384354	-.0330223
	4	-.7107582	.0101357	-70.12	0.000	-.7306238	-.6908925
	5	-.0306015	.0012718	-24.06	0.000	-.0330941	-.0281089
Good 5							
	Good						
	1	-.0347514	.0047892	-7.26	0.000	-.0441381	-.0253647
	2	-.0296409	.0016063	-18.45	0.000	-.0327893	-.0264926
	3	-.0285217	.0015907	-17.93	0.000	-.0316394	-.025404
	4	-.0212131	.0011859	-17.89	0.000	-.0235375	-.0188887
	5	-.7824229	.0109448	-71.49	0.000	-.8038744	-.7609714

```
. estat elasticities if n_kids >= 3, uncompensated
Uncompensated (Marshallian) price elasticities                    Number of obs = 418
```

Price	Elasticity	Std. err.	z	P>|z|	[95% conf. interval]	
Good 1						
Good						
1	-.9395485	.0146154	-64.28	0.000	-.9681941	-.9109029
2	-.0126359	.0032625	-3.87	0.000	-.0190303	-.0062416
3	-.0146746	.004299	-3.41	0.001	-.0231005	-.0062486
4	-.0082076	.002108	-3.89	0.000	-.0123393	-.0040759
5	-.0093949	.0024408	-3.85	0.000	-.0141788	-.0046109
Good 2						
Good						
1	-.023357	.0100783	-2.32	0.020	-.04311	-.0036039
2	-.9674982	.0135281	-71.52	0.000	-.9940128	-.9409836
3	-.028745	.0152303	-1.89	0.059	-.0585959	.0011058
4	-.0152942	.0064352	-2.38	0.017	-.0279071	-.0026813
5	-.0178107	.007578	-2.35	0.019	-.0326633	-.0029581
Good 3						
Good						
1	-.0000379	.0055043	-0.01	0.995	-.0108261	.0107504
2	.0001398	.0056976	0.02	0.980	-.0110273	.0113068
3	-1.000682	.0201954	-49.55	0.000	-1.040264	-.9610998
4	-.0000842	.0035771	-0.02	0.981	-.0070951	.0069268
5	-.0000569	.0041395	-0.01	0.989	-.0081702	.0080565
Good 4						
Good						
1	-.0451601	.0025582	-17.65	0.000	-.0501741	-.0401461
2	-.0463532	.0024489	-18.93	0.000	-.051153	-.0415534
3	-.052944	.0045994	-11.51	0.000	-.0619587	-.0439293
4	-.6843304	.0131677	-51.97	0.000	-.7101385	-.6585222
5	-.0334225	.0018152	-18.41	0.000	-.0369803	-.0298648
Good 5						
Good						
1	-.0356486	.002797	-12.75	0.000	-.0411305	-.0301666
2	-.0365794	.0027339	-13.38	0.000	-.0419377	-.0312212
3	-.0401475	.0035616	-11.27	0.000	-.0471281	-.033167
4	-.0234098	.0017471	-13.40	0.000	-.0268341	-.0199856
5	-.7542306	.0156676	-48.14	0.000	-.7849385	-.7235226

For example, in the table for households with 3 or more children, the price elasticity for Good 5 with respect to Good 1 is −0.0356486. Among these households, if the price of Good 5 (miscellaneous items) increases by 1%, then the quantity of Good 1 (dairy products) purchased will decrease by an average of 0.036%.

We summarize the own-price elasticities in table 1. We again focus on the demand for fruits and vegetables. Among households with two children or fewer, it is inelastic, while it is (barely) elastic for households with three or more children. Again, we have no explanation for why that would be the case, though we suspect the model is simply too rigid to accurately model the consumption patterns seen in our data.

Table 1: Own-price elasticities based on number of children

	# of Children	
Good	≤ 2	≥ 3
Dairy	-0.954	-0.940
Proteins	-0.939	-0.967
Fruits & vegetables	-0.887	-1.001
Flours	-0.711	-0.684
Miscellany	-0.782	-0.754

◁

Translog

Pollak and Wales (1992, sec. 3.1) discuss an entire class of demand systems that they describe as the transcendental logarithmic or "translog" family. We implement the version they call the "basic translog" function, which is probably the most well known, because it corresponds to the translog indirect utility function proposed by Christensen, Jorgenson, and Lau (1975). We also implement the version Pollak and Wales call the "generalized translog" function, which is a more flexible version of the basic translog that has been translated in a way analogous to how the LES is a shifted version of the Cobb–Douglas utility function.

Basic translog

Christensen, Jorgenson, and Lau (1975) proposed both direct and indirect utility functions with the "translog" functional form. For their direct utility function, they obtain expenditure shares that are a function of expenditure and quantities, which is not the focus of demandsys. We instead implement the demand system based on their translog indirect utility function. The starting point is the indirect utility function

$$\ln v(\mathbf{p}, m; \boldsymbol{\alpha}, \boldsymbol{\Gamma}) = -\sum_g \alpha_g \ln\frac{p_g}{m} - \frac{1}{2}\sum_g \sum_h \gamma_{gh} \ln\frac{p_g}{m} \ln\frac{p_h}{m} \tag{1}$$

Equation (1) is twice differentiable in prices, and based on Young's (1909) theorem, we must have $\gamma_{gh} = \gamma_{hg}$ for all g and h. (Young's theorem states that cross-partial derivatives are equal.)

Roy (1943) introduced what would be known as Roy's identity, which equates a consumer's Marshallian demand function for a good with a function of their indirect utility function. Applying Roy's identity, we have the expenditure-share functions

$$w_g(\mathbf{p}, m; \boldsymbol{\alpha}, \boldsymbol{\Gamma}) = \frac{\alpha_g + \sum_h \gamma_{gh}\ln\frac{p_h}{m}}{1 + \sum_j \sum_h \gamma_{jh}\ln\frac{p_h}{m}}$$

where we have imposed the normalization $\sum_g \alpha_g = 1$. This ensures that the expenditure shares sum to 1, a property known as "additivity".

As we remarked in the introduction, if one has relevant demographic variables, one should include them in the demand system, so we will not present an example of the translog model without demographic variables. Moreover, it is in fact easier to describe how we incorporate demographic variables into the translog system if we first describe the generalized translog model.

Generalized translog

Another way to perform translation and introduce subsistence or committed quantities into a demand system is to translate the indirect utility function, so that consumers' effective expenditures are their expenditures after the subsistence quantities of all goods are purchased. Pollak and Wales (1992, 56) do this to Christensen, Jorgenson, and Lau's (1975) indirect utility function to obtain

$$\ln v(\mathbf{p}, m; \boldsymbol{\alpha}, \boldsymbol{\Gamma}, \boldsymbol{\mu}) = -\sum_g \alpha_g \ln \frac{p_g}{\overline{m}} - \frac{1}{2} \sum_g \sum_h \gamma_{gh} \ln \frac{p_g}{\overline{m}} \ln \frac{p_h}{\overline{m}}$$

$$\overline{m} = m - \sum_h p_h \mu_h \qquad \sum_h \alpha_h = 1 \qquad \gamma_{gh} = \gamma_{hg}$$

where again $\boldsymbol{\mu}$ is a $G \times 1$ vector of subsistence quantities with typical element μ_g. A tedious but straightforward application of Roy's identity gives the expenditure-share functions as

$$w_g(\mathbf{p}, m; \boldsymbol{\alpha}, \boldsymbol{\Gamma}, \boldsymbol{\mu}) = \frac{p_g \mu_g}{m} + \frac{\overline{m}}{m} \left(\frac{\alpha_g + \sum_h \gamma_{gh} \ln \frac{p_h}{\overline{m}}}{1 + \sum_j \sum_h \gamma_{jh} \ln \frac{p_h}{\overline{m}}} \right)$$

$$\overline{m} = m - \sum_h p_h \mu_h \qquad \sum_h \alpha_h = 1 \qquad \gamma_{hj} = \gamma_{jh}$$

It is apparent that the sum of $w_g(\mathbf{p}, m; \boldsymbol{\alpha}, \boldsymbol{\Gamma}, \boldsymbol{\mu})$ across all g is equal to one, as must be true of a demand system. Again, suppose for each consumer that we have a $D \times 1$ vector of demographics $\mathbf{d}$ and an associated $G \times D$ matrix $\mathbf{N}$, and again, let c_g denote the gth element of the vector $\mathbf{c} = \mathbf{Nd}$.

To apply both demographic characteristics and committed quantities, we can translate the translog indirect utility function, this time with the value $(\mu_g + c_g)$ replacing μ_g in the developments above. Thus, the generalized translog system with demographic variables has expenditure-share functions of the form

$$w_g(\mathbf{p}, m, \mathbf{d}; \boldsymbol{\alpha}, \boldsymbol{\Gamma}, \boldsymbol{\mu}, \mathbf{N}) = \frac{p_g(\mu_g + \boldsymbol{\nu}_g \mathbf{d})}{m} + \frac{\overline{m}}{m} \left(\frac{\alpha_g + \sum_h \gamma_{gh} \ln \frac{p_h}{\overline{m}}}{1 + \sum_j \sum_h \gamma_{jh} \ln \frac{p_h}{\overline{m}}} \right)$$

$$\overline{m} = m - \sum_h p_h(\mu_h + \boldsymbol{\nu}_h \mathbf{d}) \qquad \sum_h \alpha_h = 1 \qquad \gamma_{hj} = \gamma_{jh}$$

The basic translog model results if we set $\boldsymbol{\mu} = \mathbf{0}$ and $\mathbf{N} = \mathbf{0}$, and the basic translog model with demographics results if we set just $\boldsymbol{\mu} = \mathbf{0}$. The generalized translog model results if we set just $\mathbf{N} = \mathbf{0}$. The generalized translog model is essentially a translog model with a constant term included among the demographic variables.

▷ Example 4

We first fit a basic translog model to our food consumption data, controlling for the number of children and adults in each household. We will also store these estimation results so we can use them later.

```
. use https://www.stata-press.com/data/r18/food_consumption
(Food consumption)

. demandsys translog w_dairy w_proteins w_fruitveg w_flours w_misc,
> prices(p_dairy p_proteins p_fruitveg p_flours p_misc)
> demographics(n_kids n_adults)
> labels("dairy proteins fruitveg flours misc")
> expenditure(expfd) elasticities(expenditure) nolog nocoeftable

Calculating NLS estimates ...
Calculating FGNLS estimates ...
FGNLS iteration 2 ...
FGNLS iteration 3 ...
FGNLS iteration 4 ...

Basic translog model                      Number of obs          =   4,160
Expenditure variable: expfd               Number of goods        =       5
Demographic method: Translating           Number of demographics =       2

Centered R2 of model for
  dairy    = 0.0448
  proteins = 0.1593
  fruitveg = 0.0976
  flours   = 0.1881
  misc     = 0.1678

Calculating expenditure elasticities ...

Expenditure elasticities
```

Expenditure	Elasticity	Std. err.	z	P>|z|	[95% conf. interval]	
dairy	.90727	.0157948	57.44	0.000	.8763129	.9382272
proteins	1.0912	.0102476	106.48	0.000	1.071115	1.111285
fruitveg	.9629558	.0128444	74.97	0.000	.9377813	.9881303
flours	.8728076	.0152504	57.23	0.000	.8429174	.9026979
misc	1.000333	.0177894	56.23	0.000	.9654664	1.0352

```
Note: Elasticities are calculated at prices', demographic variables', and
      expenditure means.

. estimates store translog
```

We specified the nocoeftable option to suppress the coefficient table and the elastici-
ties(expenditure) option to obtain expenditure elasticities. Finally, we used the labels() option
to name our five goods. Had we not done that, then the output would simply have numbered the
goods from one to five.

For comparison's sake, we also fit the equivalent generalized translog model. We note that despite
the long commands, the following one differs from the previous by just one character:

```
. demandsys gtranslog w_dairy w_proteins w_fruitveg w_flours w_misc,
> prices(p_dairy p_proteins p_fruitveg p_flours p_misc)
> demographics(n_kids n_adults)
> labels("dairy proteins fruitveg flours misc")
> expenditure(expfd) elasticities(expenditure) nolog nocoeftable
Calculating NLS estimates ...
Calculating FGNLS estimates ...
FGNLS iteration 2 ...
FGNLS iteration 3 ...
FGNLS iteration 4 ...
Generalized translog model                  Number of obs         = 4,160
Expenditure variable: expfd                  Number of goods       =     5
Demographic method: Translating              Number of demographics =    2
Centered R2 of model for
  dairy    = 0.0478
  proteins = 0.1615
  fruitveg = 0.0988
  flours   = 0.1903
  misc     = 0.1732
Calculating expenditure elasticities ...
Expenditure elasticities
```

Expenditure	Elasticity	Std. err.	z	P>|z|	[95% conf. interval]	
dairy	.906851	.0174393	52.00	0.000	.8726707	.9410313
proteins	1.074059	.01144	93.89	0.000	1.051637	1.096481
fruitveg	.9716098	.0144251	67.36	0.000	.943337	.9998826
flours	.892461	.0170431	52.36	0.000	.8590571	.9258649
misc	1.023297	.0194942	52.49	0.000	.9850886	1.061504

```
Note: Elasticities are calculated at prices', demographic variables', and
      expenditure means.
. estimates store gtranslog
```

The R^2 statistics from the generalized translog model are trivially higher than for the basic translog model. Because demandsys is based on maximum likelihood estimation and because the translog model is nested within the generalized translog model, we can use a likelihood-ratio test to see whether the inclusion of committed quantities is warranted here.

```
. lrtest gtranslog translog
Likelihood-ratio test
Assumption: translog nested within gtranslog
 LR chi2(5) =  59.00
Prob > chi2 = 0.0000
```

The generalized translog model adds additional G parameters over the basic translog model, so the χ^2 statistic has five degrees of freedom in this example. The result indicates that the use of the generalized version of the translog model is warranted. Of course, if you do this test for your own models, you must use the same demographic specification in both.

◁

As we cautioned with the LES, interpreting the μ parameters of the generalized translog model as subsistence quantities requires care. We do not force the estimates to be positive, nor does demandsys check that $\sum_g p_{gi}(\mu_g + \boldsymbol{\nu}_g' \mathbf{d}_i) \le m_i$ for each observation i in the dataset or that $\mu_g + \boldsymbol{\nu}_g' \mathbf{d}_i \ge 0$ for each observation and each good, both of which must be true if the subsistence argument is to be credible. We have noted that the generalized translog model can produce negative estimates of μ's quite frequently and changing the demographic specification is often enough to produce negative estimates.

For example, if we refit the previous example controlling for just the number of children but not the number of adults, we obtain negative estimates for some of the μ parameters:

```
. demandsys gtranslog w_dairy w_proteins w_fruitveg w_flours w_misc,
> prices(p_dairy p_proteins p_fruitveg p_flours p_misc)
> demographics(n_kids) expenditure(expfd) nolog

Calculating NLS estimates ...
Calculating FGNLS estimates ...
FGNLS iteration 2 ...
FGNLS iteration 3 ...
FGNLS iteration 4 ...

Generalized translog model                    Number of obs           = 4,160
Expenditure variable: expfd                    Number of goods         =     5
Demographic method: Translating                Number of demographics =     1

Centered R2 of model for
  Good 1 = 0.0421
  Good 2 = 0.1600
  Good 3 = 0.0929
  Good 4 = 0.1825
  Good 5 = 0.1706
```

	Estimate	Std. err.	z	P>|z|	[95% conf. interval]	
alpha						
Good						
1	.1745891	.0097353	17.93	0.000	.1555082	.1936699
2	.3416641	.0188744	18.10	0.000	.304671	.3786572
3	.3634044	.0192262	18.90	0.000	.3257218	.401087
4	.0598082	.0066154	9.04	0.000	.0468423	.0727741
Gamma						
Good_g#						
Good_h						
1#1	.0232919	.002496	9.33	0.000	.0183998	.0281841
1#2	-.0146955	.0032932	-4.46	0.000	-.02115	-.0082411
1#3	.0005704	.0018905	0.30	0.763	-.0031349	.0042758
1#4	.001351	.0009723	1.39	0.165	-.0005546	.0032566
1#5	.0026193	.0010776	2.43	0.015	.0005073	.0047312
2#2	.120652	.0085591	14.10	0.000	.1038764	.1374275
2#3	-.0301533	.0055644	-5.42	0.000	-.0410593	-.0192474
2#4	-.0184411	.0027052	-6.82	0.000	-.0237432	-.013139
2#5	-.0228452	.0032034	-7.13	0.000	-.0291238	-.0165667
3#3	.0698867	.0067428	10.36	0.000	.056671	.0831023
3#4	.0006331	.0013394	0.47	0.636	-.0019921	.0032583
3#5	-.0065477	.0017672	-3.71	0.000	-.0100113	-.003084
4#4	.0233853	.0023855	9.80	0.000	.0187097	.0280608
4#5	-.0074747	.0011361	-6.58	0.000	-.0097014	-.005248
5#5	.0323539	.0030878	10.48	0.000	.0263019	.0384059

mu						
Good						
1	−.1560689	.0474664	−3.29	0.001	−.2491014	−.0630364
2	−.811357	.2053695	−3.95	0.000	−1.213874	−.4088403
3	−.7661874	.3576453	−2.14	0.032	−1.467159	−.0652155
4	.2121302	.0367754	5.77	0.000	.1400517	.2842088
5	.0887401	.0446207	1.99	0.047	.0012853	.176195

Nu						
Good#						
c.n_kids						
1	−.5508243	.1957429	−2.81	0.005	−.9344733	−.1671753
2	−1.350384	.1934558	−6.98	0.000	−1.72955	−.9712174
3	−2.749218	.303765	−9.05	0.000	−3.344586	−2.153849
4	.0446037	.0442943	1.01	0.314	−.0422117	.131419
5	−.1036959	.0559887	−1.85	0.064	−.2134318	.0060399

Note: **alpha** estimates are constant expenditures for each good.
Note: **Gamma** estimates measure the effect of price on expenditures shares
     across goods.
Note: **mu** estimates are subsistence levels of consumption for each good.
Note: **Nu** estimates measure the effect of demographic variables on shares and
     subsistence levels of consumption.

Normalized parameter

	Estimate	Std. err.	z	P>|z|	[95% conf. interval]	
alpha						
Good						
5	.0605342	.007785	7.78	0.000	.045276	.0757925

Note: **alpha** estimates sum to 1.

Shares: w_dairy w_proteins w_fruitveg w_flours w_misc
Prices: p_dairy p_proteins p_fruitveg p_flours p_misc
Expenditure: expfd
Demographics: n_kids

Here, again, we see factor-variable notation for the parameter matrix $\mathbf{N}$ but also for Γ. As we mentioned before, you should not interpret the factor-variable notation as you would in a regression output. We see the estimate labeled 1#2 under Good_g#Good_h. This is the estimate of γ_{12}, the coefficient associated with $\ln p_2/\overline{m}$ in the equation for Good 1.

AIDS

One of the most commonly used demand systems is the celebrated AIDS model of Deaton and Muellbauer (1980b). Among other benefits, the AIDS model can be viewed as a first-order approximation to any demand system. Historically, the AIDS model was also favored because it is "almost linear" in the sense that replacing the price index with an approximation results in share equations that are linear. Given the speed with which computers can fit nonlinear systems of equations, we do not implement that approximation.

An additional nice property of the AIDS model is that the functional form for its budget shares is a member of the "price-independent generalized log-linear" and the broader "price-independent generalized linear" families. An advantage to price-independent generalized linear expenditure shares is that they satisfy certain conditions required for the existence of a representative consumer and the related theory of aggregation. In short, expenditure shares of this form can be fit to data aggregated across consumers. The implied market demand curves are consistent with a single utility-maximizing

representative agent and satisfy the same conditions like Slutsky symmetry that hold for demand curves implied by a single consumer's utility-maximizing behavior. For more on price-independent generalized linear demands, see, for example, Muellbauer (1975), Lewbel (1989), and Pollak and Wales (1992, chap. 2).

The starting point for the AIDS model is the expenditure function

$$\ln e(\mathbf{p}, u; \alpha_0, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Gamma}) = (1 - \ln u)\ln a(\mathbf{p}) + \ln u \ \{\ln a(\mathbf{p}) + b(\mathbf{p})\} \tag{2}$$

where u is utility, the price index $\ln a(\mathbf{p})$ is defined as

$$\ln a(\mathbf{p}) \equiv \alpha_0 + \sum_g \alpha_g \ln p_g + \frac{1}{2} \sum_g \sum_h \gamma_{gh} \ln p_g \ln p_h$$

where p_g is the price of good g, and the price aggregator $b(\mathbf{p})$ is defined as

$$b(\mathbf{p}) \equiv \beta_0 \prod_g p_g^{\beta_g}$$

Parameter β_0 cannot be distinguished from u, so we take $\beta_0 = 1$ in the following. In their original formulation, Deaton and Muellbauer (1980b) wrote their expenditure function slightly differently, with u in place of $\ln u$ in (2). We have chosen to use $\ln u$ to draw some comparisons with the QUAIDS model later, and because utility is an ordinal concept, it will not affect the expenditure-share equations.

Suppose that all prices are equal to one. Then (2) with $\beta_0 = 1$ implies that $\ln e(p, u) = \alpha_0$. Thus, $\exp(\alpha_0)$ can be interpreted as the level of expenditure needed for minimal subsistence when all prices are equal to one. In practice, estimating α_0 is difficult, particularly when prices are correlated, as they often are. Therefore, by default, we set α_0 equal to the natural logarithm of the minimum level of expenditure in the estimation sample. You can override the default by specifying the `piconstant()` option. Of course, prices are generally not all equal to one, so you will want to try various values for α_0 to see how sensitive elasticities and other calculations you may perform are to its value.

Expenditure functions consistent with rational consumer choice must be homogeneous of degree one in prices and expenditure, which imply that we must have $\sum_g \alpha_g = 1$, $\sum_g \beta_g = 0$, and $\sum_g \gamma_{gh} = \sum_h \gamma_{gh} = 0$. Slutsky symmetry further requires that $\gamma_{gh} = \gamma_{hg}$ for all g and h. Note that `demandsys` imposes these restrictions on the model being fit. An alternative is to fit the model with and without one or more of those restrictions and then test whether they hold. However, such unrestricted models can often fail to converge because of the increased number of parameters or because the models are "too flexible" and contain unidentifiable parameters.

Shephard's (1970) lemma equates the Hicksian demand functions with the partial derivatives of the expenditure function with respect to the price of the goods. Using Shephard's lemma, we can write the expenditure-share functions as

$$w_g(\mathbf{p}, m; \alpha_0, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Gamma}) = \alpha_g + \sum_h \gamma_{gh} \ln p_h + \beta_g \ln \left\{ \frac{m}{a(\mathbf{p})} \right\} \tag{3}$$

Notice the simplicity of these share equations. Good g's expenditure share depends linearly on the log prices of all the goods as well as the log of an expenditure term. The expenditure term $m/a(\mathbf{p})$ essentially adjusts each household's expenditure to control the cost of living faced by that household. This term is sometimes referred to as deflated expenditure. Equation (3) also makes clear that luxury goods have $\beta_g > 1$ while necessities have $\beta_g < 1$; inferior goods have $\beta_g < 0$.

Setting $e(\mathbf{p}, u) = m$ and manipulating (2) yields the indirect utility function

$$\ln v(\mathbf{p}, m; \alpha_0, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Gamma}) = \frac{\ln m - \ln a(\mathbf{p})}{b(\mathbf{p})} \tag{4}$$

We will come back to this equation below. We show how the indirect utility function plays a role in computing welfare measures in [R] **demandsys postestimation**.

QUAIDS

The Engel curves for the AIDS model have the so-called "Working–Leser" form after Working (1943) and Leser (1963),

$$w_g = \omega_g + \psi_g \ln m$$

so that the share of expenditures spent on good g is a linear function of the logarithm of total expenditures. Banks, Blundell, and Lewbel (1997) provided evidence that in fact linear Engel curves provide a poor fit for many goods. They therefore consider Engel curves with an additional expenditure term,

$$w_g = A_g(\mathbf{p}) + B_g(\mathbf{p}) \ln \left\{ \frac{m}{a(\mathbf{p})} \right\} + C_g(\mathbf{p}) \, f \left\{ \frac{m}{a(\mathbf{p})} \right\} \tag{5}$$

where $A_g(\mathbf{p})$, $B_g(\mathbf{p})$, and $C_g(\mathbf{p})$ are differentiable functions of prices and $f \{m/a(\mathbf{p})\}$ is a differentiable function of real expenditures. $a(\mathbf{p})$ is the same translog price index as in the AIDS model.

Lewbel (1991), building on the work of Gorman (1981), showed that any exactly aggregable demand system must have the form of (5) and that additional terms containing functions of $m/a(\mathbf{p})$ are not theoretically possible if exact aggregability is to hold. Lewbel (1991) termed demand systems of the form (5) as having rank three because there are three linearly independent terms; Gorman (1981) showed that the maximum rank of demand systems is three if exact aggregability is to hold. In contrast, the AIDS model has rank two.

Banks, Blundell, and Lewbel (1997) further showed that the only indirect utility functions consistent with rank-three expenditure-share equations like (5) have the form

$$\ln v(\mathbf{p}, m) = \left[\left\{ \frac{\ln m - \ln a(\mathbf{p})}{b(\mathbf{p})} \right\}^{-1} + \lambda(\mathbf{p}) \right]^{-1}$$

where in their application they use the same definitions for $a(\mathbf{p})$ and $b(\mathbf{p})$ as in the AIDS model and they take $\lambda(\mathbf{p}) = \sum_g \lambda_g \ln p_g$ with $\sum_g \lambda_g = 0$. As with the restrictions implied by economic theory for the AIDS model, demandsys imposes the restriction that the λ's sum to zero. Also notice that if we restrict $\lambda_g = 0$ for all g, we are left with the indirect utility function for the AIDS model shown in (4).

Solving for $\ln m$ yields the cost function

$$\ln e(\mathbf{p}, u) = \frac{b(\mathbf{p})}{\frac{1}{\ln u} - \lambda(\mathbf{p})} + \ln a(\mathbf{p})$$

Applying Shephard's lemma, we have the expenditure-share functions

$$w_g(\mathbf{p}, m; \boldsymbol{\alpha}, \boldsymbol{\Gamma}, \boldsymbol{\Lambda}) = \alpha_g + \sum_h \gamma_{gh} \ln p_h + \beta_g \ln \left\{ \frac{m}{a(\mathbf{p})} \right\} + \frac{\lambda_g}{b(\mathbf{p})} \left[\ln \left\{ \frac{m}{a(\mathbf{p})} \right\} \right]^2$$

This equation also makes clear our earlier claim that if $\lambda_g = 0$ for all g then the QUAIDS model reduces to the AIDS model; see (3).

Controlling for demographic factors

demandsys allows you to choose between two different methods of incorporating demographic characteristics into AIDS and QUAIDS. The first method is the same as we have used with demand systems we have already covered: Pollak and Wales's (1978) demographic translation. An added benefit of translation is that it allows us to introduce constant terms that might be interpreted as subsistence or committed quantities. A possible downside to demographic translation is that the intuition behind how demographic variables affect expenditure shares may not strike all users as entirely lucid. Essentially, demographic variables alter the level of subsistence for each household or observation. However, as we cautioned above in our discussion of the generalized translog model, there are no restrictions to force $\sum_g p_{gi}(\mu_g + \boldsymbol{\nu}'_g \mathbf{d}_i) \leq m_i$, nor are there any restrictions to force $(\mu_g + \boldsymbol{\nu}'_g \mathbf{d}_i) \geq 0$.

The second method of incorporating demographics that demandsys implements is based on Ray (1983) and Poi (2002) and is known as demographic scaling. This method explicitly allows for demographic variables to have "scale" and "composition" effects on expenditures as we explain below. There are two possible downsides to demographic scaling. First, it does not allow one to incorporate subsistence quantities into the demand system. While one may be tempted to include a constant term in the list of demographic variables because that is essentially what demographic translation does to incorporate subsistence quantities, the resulting parameter estimates do not have such a clearcut interpretation. Second, in our experience, models that incorporate demographic scaling sometimes take many iterations to converge. Specifying fewer demographic characteristics may help achieve convergence. We think this method is more intuitive than demographic translation, but either method should prove adequate for most applications.

Demographic translation

We have already discussed demographic translation in the context of the LES and the translog and generalized translog demand systems, so we will not show the details here. In *Methods and formulas*, we show the formula for the expenditure shares for the generalized QUAIDS model, and we discuss the parameter restrictions that give rise to the generalized AIDS model and their nongeneralized variants.

demandsys provides for four models in the AIDS family that can incorporate demographic variables via demographic translation: aids, quaids, gaids, and gquaids. The first two models do not allow for committed quantities, while the latter two represent generalized variants that do. The arguments in Pollak and Wales (1992, 75) suggest that the generalized variants are to be preferred because doing so ensures that a constant term is included in each demand equation (compare with share equation). Whether the constants allow for a subsistence interpretation is another matter, and in [R] **demandsys postestimation**, we provide an example of how to see whether that interpretation holds.

Demographic scaling

demandsys also implements the demographic scaling method used by Ray (1983) for the AIDS model and extended by Poi (2002) to the QUAIDS model. For each consumer, we again have a D-length vector of demographic characteristics $\mathbf{d}$. Suppose $\mathbf{d}$ is the number of children in a household and we are modeling expenditures on household goods, rent, utilities, and food. Then the household's expenditures will presumably increase by virtue of there being more members in the household; call this the scale effect. Second, the consumption pattern of the household may shift as more money is spent on items consumed by children versus adults; think of that as a composition effect. Ray (1983) does this by writing the expenditure function as

$$e(\mathbf{p}, \mathbf{d}, u) = e^R(\mathbf{p}, u) \, \overline{m}_0(\mathbf{d}) \, \phi(\mathbf{p}, \mathbf{d}, u)$$

Here $e^R(\mathbf{p}, u)$ is the expenditure function for a reference household, where $\mathbf{d} = 0$. $\overline{m}_0(\mathbf{d})$ increases total expenditure of the household and accounts for the scale effect. $\phi(\mathbf{p}, \mathbf{d}, u)$ controls for composition effects. For AIDS, Ray (1983) controlled for both of these effects by taking

$$\overline{m}_0(\mathbf{d}) = 1 + \boldsymbol{\rho}'\mathbf{d}$$

for $G \times 1$ parameter vector $\boldsymbol{\rho}$ and

$$\phi(\mathbf{p}, \mathbf{d}, u) = \exp\left\{ u \prod_i p_g^{\beta_g} \left(\prod_g p_g^{\boldsymbol{\eta}_g \mathbf{d}} - 1 \right) \right\}$$

where $\boldsymbol{\eta}_g$ is the gth row of $G \times D$ parameter matrix $\mathbf{H}$. In *Methods and formulas*, we show that the expenditure-share functions for the AIDS model with demographic scaling are

$$w_g(\mathbf{p}, m, \mathbf{d}; \alpha_0, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Gamma}, \boldsymbol{\lambda}, \boldsymbol{\rho}, \mathbf{H}) = \alpha_g + \sum_h \gamma_{gh}\ln p_h + (\beta_g + \boldsymbol{\eta}_i\mathbf{d})\ln\left\{ \frac{m}{\overline{m}_0(\mathbf{d}; \boldsymbol{\rho})\, a(\mathbf{p})} \right\}$$

Basically, the function $\overline{m}_0(\mathbf{d}; \boldsymbol{\rho})$ acts to reduce the effective amount of money available for spending. If $\mathbf{d}$ represents a set of indicator variables for regions of a country, then $\overline{m}_0(\mathbf{d}; \boldsymbol{\rho})$ is controlling for the cost of living in different regions of the country. The function $\phi(\mathbf{p}, \mathbf{d}, u)$ is to make each good's expenditure expansion path a function of the household's demographics. Rather than all households increasing their expenditure share of good g by the same β_g in response to a change in m, now their responses can vary by household composition.

For the QUAIDS model, we show in *Methods and formulas* that an appropriate choice for $\phi(\mathbf{p}, \mathbf{d}, u)$ leads to expenditure share equations

$$\begin{aligned} w_g(\mathbf{p}, m, \mathbf{d}; \alpha_0, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Gamma}, \boldsymbol{\lambda}, \boldsymbol{\rho}, \mathbf{H}) = &\alpha_g + \sum_h \gamma_{gh}\ln p_h + (\beta_g + \boldsymbol{\eta}_i\mathbf{d})\ln\left\{ \frac{m}{\overline{m}_0(\mathbf{d}; \boldsymbol{\rho})\, a(\mathbf{p})} \right\} \\ &+ \frac{\lambda_g}{b(\mathbf{p})\, c(\mathbf{p}, \mathbf{d})} \left[\ln\left\{ \frac{m}{\overline{m}_0(\mathbf{d}; \boldsymbol{\rho})\, a(\mathbf{p})} \right\} \right]^2 \end{aligned}$$

The models `aids` and `quaids` allow for demographic scaling. To request demographic scaling, rather than demographic translation, we specify the `scaling` suboption of the `demographics()` option, as the next example shows.

▷ Example 5

Here we fit a QUAIDS model to our food consumption data, controlling for the numbers of children and adults in each household. We request that demographic scaling be used rather than the default demographic translation.

```
. use https://www.stata-press.com/data/r18/food_consumption
(Food consumption)

. demandsys quaids w_dairy w_proteins w_fruitveg w_flours w_misc,
> prices(p_dairy p_proteins p_fruitveg p_flours p_misc)
> demographics(n_kids n_adults, scaling)
> labels("dairy proteins fruitveg flours misc")
> expenditure(expfd) elasticities(expenditure) nolog nocoeftable

Calculating NLS estimates ...
Calculating FGNLS estimates ...
FGNLS iteration 2 ...
FGNLS iteration 3 ...
FGNLS iteration 4 ...

Quadratic AIDS model                          Number of obs        =  4,160
Expenditure variable: expfd                   Number of goods      =      5
                                              Price index constant =  1.615
Demographic method: Scaling                   Number of demographics =    2

Centered R2 of model for
  dairy    = 0.0453
  proteins = 0.1530
  fruitveg = 0.0949
  flours   = 0.1813
  misc     = 0.1726

Calculating expenditure elasticities ...

Expenditure elasticities
```

Expenditure	Elasticity	Std. err.	z	P>|z|	[95% conf. interval]	
dairy	.8918188	.0176252	50.60	0.000	.857274	.9263636
proteins	1.077283	.0117431	91.74	0.000	1.054267	1.100299
fruitveg	1.043209	.014428	72.30	0.000	1.01493	1.071487
flours	.8409016	.0162384	51.78	0.000	.8090749	.8727284
misc	.9460643	.0189172	50.01	0.000	.9089873	.9831414

```
Note: Elasticities are calculated at prices', demographic variables', and
      expenditure means.

. estimates store quaids_s
```

Because the QUAIDS model with demographic scaling for a 5-good system with 2 demographic variables has 32 estimated parameters, we instructed demandsys to report expenditure elasticities rather than the estimated parameters. We draw your attention to this part of the command:

```
demographics(n_kids n_adults, scaling)
```

Notice that we specified the suboption scaling. Had we not included this suboption, then demandsys would have used demographic translation instead. Finally, we asked Stata to store these estimation results as quaids_s.

The header of the output shows that demandsys fit a QUAIDS model as requested and that our demographic variables were incorporated using demographic scaling. Take note of the line that reads

```
Price index constant = 1.615
```

In our discussion of the basic AIDS model, we discussed the price index $a(\mathbf{p})$ that includes a constant term α_0. We discussed that when all prices are equal to one, then $\exp(\alpha_0)$ can be interpreted as the

level of expenditure needed for minimal subsistence. demandsys looked at our expenditure variable expfd and found that the minimum value within the estimation sample is \$5.03. It therefore set $\alpha_0 = \ln 5.03 = 1.615$. You can specify your own value for α_0 by using the piconstant() option, which stands for price index constant.

The QUAIDS model nests the AIDS model, so we can easily use a likelihood-ratio test to see whether the quadratic terms in the expenditure-share equations contribute to the explanatory power of the model. We include quietly to fit the corresponding aids model without displaying the results and store the estimation results:

```
. quietly demandsys aids w_dairy w_proteins w_fruitveg w_flours w_misc,
> prices(p_dairy p_proteins p_fruitveg p_flours p_misc)
> demographics(n_kids n_adults, scaling)
> labels("dairy proteins fruitveg flours misc") expenditure(expfd)
. estimates store aids_s
```

Because we specified quietly, we did not request expenditure elasticities. Doing so would have required Stata to do more computations that we would not have seen.

We now perform the likelihood-ratio test:

```
. lrtest quaids_s aids_s
Likelihood-ratio test
Assumption: aids_s nested within quaids_s
 LR chi2(4) =  36.58
Prob > chi2 = 0.0000
```

In this example, we reject the null hypothesis that the four free λ parameters are jointly equal to zero. Hence, the use of the QUAIDS model instead of the linear AIDS model is justified here. (Recall that to satisfy the adding-up constraint, we impose $\sum_g \lambda_g = 0$ so that in our example the fifth parameter is determined by the first four and is not estimated.)

◁

When one conducts tests of one model versus another using the likelihood-ratio principle, the more restrictive model must be nested within the more general model. For example, each of the AIDS models implemented here is nested within the corresponding QUAIDS models, and the AIDS model with demographic translation is nested within the generalized AIDS model. But it would make no sense to fit a generalized AIDS model and compare it with a generalized translog model because one is not a restricted variant of the other. In *Methods and formulas*, we describe the most flexible variant within each class of demand system, and we describe the parameter restrictions that give rise to less flexible variants; using lrtest to choose among members within the same class is valid.

Also, for a likelihood-ratio test to be valid, the model specifications—apart from the component we are testing—must be identical. That is, the two models must include the same estimation sample, the same goods, and the same demographic specification, including the method by which demographic variables are introduced for AIDS and QUAIDS models. Moreover, for AIDS and QUAIDS models, you must use the same value for the price index constant α_0. If you do not specify the piconstant() option, demandsys will use the same value, assuming you use the same estimation sample.

▷ Example 6

In this example, we will fit a generalized QUAIDS model, again controlling for the number of children and adults. Because we are fitting a generalized model, the demographics will be incorporated via demographic translation. To demonstrate the use of the piconstant() option, we will specify $\alpha_0 = 3$.

```
. use https://www.stata-press.com/data/r18/food_consumption
(Food consumption)

. demandsys gquaids w_dairy w_proteins w_fruitveg w_flours w_misc,
> prices(p_dairy p_proteins p_fruitveg p_flours p_misc)
> demographics(n_kids n_adults) piconstant(3) expenditure(expfd) nolog
Calculating NLS estimates ...
Calculating FGNLS estimates ...
FGNLS iteration 2 ...
FGNLS iteration 3 ...
FGNLS iteration 4 ...
```

Generalized quadratic AIDS model   Number of obs   = 4,160
Expenditure variable: expfd   Number of goods   =   5
                                          Price index constant   =   3
Demographic method: Translating   Number of demographics =   2

Centered R2 of model for
  Good 1 = 0.0486
  Good 2 = 0.1596
  Good 3 = 0.0996
  Good 4 = 0.1893
  Good 5 = 0.1715

	Estimate	Std. err.	z	P>|z|	[95% conf. interval]	
alpha						
Good						
1	.1911178	.0084914	22.51	0.000	.1744748	.2077607
2	.2916653	.0180615	16.15	0.000	.2562655	.3270651
3	.3606389	.0168958	21.34	0.000	.3275236	.3937541
4	.0716379	.0045291	15.82	0.000	.062761	.0805149
beta						
Good						
1	-.0123296	.0075589	-1.63	0.103	-.0271448	.0024857
2	.0007155	.0191227	0.04	0.970	-.0367642	.0381952
3	-.0083997	.0137386	-0.61	0.541	-.0353268	.0185274
4	.0118892	.0045299	2.62	0.009	.0030106	.0207677
Gamma						
Good_g#						
Good_h						
1#1	.0381355	.0030825	12.37	0.000	.032094	.044177
1#2	-.0326578	.0029964	-10.90	0.000	-.0385305	-.026785
1#3	-.0106876	.0026841	-3.98	0.000	-.0159483	-.0054269
1#4	.0021791	.0012889	1.69	0.091	-.000347	.0047052
2#2	.1579944	.0060979	25.91	0.000	.1460428	.169946
2#3	-.0694334	.0044839	-15.49	0.000	-.0782216	-.0606452
2#4	-.0240681	.0020371	-11.82	0.000	-.0280606	-.0200755
3#3	.0923088	.005649	16.34	0.000	.0812371	.1033806
3#4	-.0001376	.0017618	-0.08	0.938	-.0035908	.0033156
4#4	.0313475	.0023698	13.23	0.000	.0267027	.0359922
lambda						
Good						
1	.0043566	.0028362	1.54	0.125	-.0012022	.0099154
2	.0005707	.0059707	0.10	0.924	-.0111316	.0122729
3	-.006845	.004352	-1.57	0.116	-.0153747	.0016848
4	-.0001594	.0017033	-0.09	0.925	-.0034979	.003179

		Estimate	Std. err.	z	P>|z|	[95% conf. interval]	
mu							
	Good						
	1	1.06405	.2097966	5.07	0.000	.6528563	1.475244
	2	.6058134	.452923	1.34	0.181	-.2818993	1.493526
	3	2.146931	.6621706	3.24	0.001	.8491008	3.444762
	4	.3204353	.0685528	4.67	0.000	.1860742	.4547964
	5	.3588502	.0854053	4.20	0.000	.1914589	.5262414
Nu							
	Good#						
	c.n_kids						
	1	-.2399127	.1953618	-1.23	0.219	-.6228148	.1429893
	2	-1.176653	.1938486	-6.07	0.000	-1.556589	-.7967166
	3	-2.604984	.3121795	-8.34	0.000	-3.216844	-1.993123
	4	.1120979	.0416122	2.69	0.007	.0305396	.1936563
	5	-.0296108	.0541832	-0.55	0.585	-.1358079	.0765863
	Good#						
	c.n_adults						
	1	-1.271932	.2071674	-6.14	0.000	-1.677972	-.8658909
	2	-1.278973	.2249348	-5.69	0.000	-1.719837	-.8381089
	3	-2.698807	.3809636	-7.08	0.000	-3.445482	-1.952132
	4	-.056265	.0483113	-1.16	0.244	-.1509535	.0384234
	5	-.269388	.0569314	-4.73	0.000	-.3809716	-.1578045

Note: **alpha** estimates are constant terms in expenditure-share equations and
also appear in the price index.
Note: **beta** estimates measure sensitivity of expenditure shares to changes in
deflated expenditure and also appear in the price aggregator function.
Note: **Gamma** estimates measure the effect of price on expenditures shares
across goods.
Note: **lambda** estimates measure the sensitivity of expenditure shares to
changes in deflated expenditure.
Note: **Nu** estimates measure the effect of demographic variables on expenditure
shares of each good.

Normalized parameters

		Estimate	Std. err.	z	P>|z|	[95% conf. interval]	
alpha							
	Good						
	5	.0849402	.0049286	17.23	0.000	.0752803	.0946001
beta							
	Good						
	5	.0081246	.0050562	1.61	0.108	-.0017854	.0180345
Gamma							
	Good_g#						
	Good_h						
	1#5	.0030308	.0014661	2.07	0.039	.0001572	.0059044
	2#5	-.0318352	.0023155	-13.75	0.000	-.0363734	-.0272969
	3#5	-.0120502	.0019986	-6.03	0.000	-.0159674	-.008133
	4#5	-.0093209	.0011512	-8.10	0.000	-.0115772	-.0070646
lambda							
	Good						
	5	.0020771	.0019631	1.06	0.290	-.0017705	.0059247

```
Note: alpha estimates sum to 1.
Note: beta estimates sum to 0.
Note: Gamma estimates sum to 0 over goods.
Note: lambda estimates sum to 0 over goods.

Shares: w_dairy w_proteins w_fruitveg w_flours w_misc
Prices: p_dairy p_proteins p_fruitveg p_flours p_misc
Expenditure: expfd
Demographics: n_kids n_adults
```

We see in the header of the output that demandsys did set the price index constant $\alpha_0 = 3$ as per our request.

The μ parameter estimates are all greater than zero, but it is difficult to interpret them in isolation. Because we control for the number of children and the number of adults in each household, the μ parameters themselves would represent subsistence expenditure shares for a hypothetical household with neither any children nor any adults! In a somewhat technical example in [R] **demandsys postestimation**, we continue this example and show how to recover the estimated μ and $\mathbf{N}$ parameters and then use those estimates to calculate the actual subsistence shares for each household. We then compare the calculated subsistence expenditures with actual expenditures to see whether the subsistence interpretation even holds.

◁

Epilogue

demandsys provides easy access to some of the most commonly used demand systems and their extensions. You can control for demographic characteristics. Options and postestimation commands make obtaining elasticities trivial.

Despite what may appear to be a thorough implementation, we have only scratched the surface. demandsys is designed for large cross-sectional datasets. The Rotterdam model of Theil (1965) and Barten (1966) and its extensions are widely used in the analysis of time-series data.

A frequent concern is that datasets often have expenditure shares that are zero for some households. Deaton and Irish (1984) is an early contribution to the literature on how to deal with zero shares. Heien and Wessells (1990) proposed a two-step Heckman estimation procedure, but it has been shown to lead to inconsistent estimates (Vermeulen 2001). Tauchmann (2010) provides an alternative Heckman-type estimator that is consistent. Shonkwiler and Yen (1999) proposed an alternative estimator for zero shares. See also Yen and Lin (2006) for yet another approach and Meyerhoefer, Ranney, and Sahn (2005), who develop an estimator that controls for censoring with panel data. More recently, Caro et al. (2021) have developed the community-contributed Stata command quaidsce, which provides an implementation of Schonkwiler and Yen's estimator.

All of our estimators treat price as an exogenous variable. There are multiple reasons why one may want to allow for endogenous prices. In the classical supply and demand setting, which may occur for example in small villages, there are a small number of buyers and sellers, so the decisions that buyers make will influence prices, causing them to be endogenous in expenditure-share equations. A more pertinent concern is that prices may be measured with error. Consumption is often collected by survey data, and consumers are often not able to recall the exact price paid for an item. Unobserved quality differences also cause prices to be measured with error, especially when the goods being modeled are aggregates of individual items. For example, you may purchase filet mignon every week, while I purchase economy-grade ground beef. Both items are lumped together as "meat" even though, arguably, the two products are not even close substitutes; Nelson (1991) considers this type of problem. Spatial patterns may also make prices endogenous; see Case (1991) as an example.

Lecocq and Robin (2015) provide a community-contributed command called `aidsills` that allows one to fit an AIDS model with endogenous variables. Their command, however, provides the AIDS and QUAIDS models with only one method of including demographic variables that is similar to the demographic translation approach used here.

We have presented the QUAIDS model as somewhat of an endpoint because it has rank 3, which Gorman (1981) and Lewbel (1991) showed is the maximum rank of an exactly aggregable demand system. Lewbel and Pendakur (2009) develop what they call a theory of "implicit Marshallian demands" that are not within the class of demand equations considered by Gorman and Lewbel and hence can have any rank. Lewbel and Pendakur propose a demand system they call the exact affine Stone index that allows for interactions between prices and expenditures and allows for much more flexible Engel curves. Their empirical application shows that observed demands often deviate from those implied by demands that are linear or quadratic in income. Moreover, their model can be fit using the generalized method of moments, so the possibility of controlling for price endogeneity exists.

Stored results

demandsys stores the following in e():

Scalars
e(N)	number of observations
e(k)	number of parameters
e(k_eq)	number of equations in e(b)
e(n_demos)	number of demographic factors
e(n_goods)	number of goods
e(mss_#)	model sum of squares for the #th equation
e(rss_#)	residual sum of squares for the #th equation
e(tss_#)	total sum of squares for the #th equation
e(r2_#)	R^2 for the #th equation
e(ll)	Gaussian log likelihood
e(N_clust)	number of clusters
e(piconstant)	number of constant term
e(rank)	rank of e(V)
e(ic)	number of iterations
e(converged)	1 if converged, 0 otherwise

Macros
e(cmd)	demandsys
e(cmdline)	command as typed
e(wtype)	weight type
e(wexp)	weight expression
e(model)	demand model
e(model_eval)	demand evaluator
e(title)	title in estimation output
e(clustvar)	name of cluster variable
e(vce)	*vcetype* specified in vce()
e(vcetype)	title used to label Std. err.
e(demo_type)	demographic variable specification
e(demos)	demographic variables
e(has_demos)	1 if demographic variables are used, 0 otherwise
e(expenditures)	expenditures variable
e(prices)	price variables
e(shares)	expenditure-share variables
e(r2_type)	R^2 type computed
e(p_index)	whether model has price index
e(properties)	b V
e(estat_cmd)	program used to implement estat
e(predict)	program used to implement predict
e(marginsok)	predictions allowed by margins

Matrices

e(b)	coefficient vector
e(Sigma)	error covariance matrix ($\widehat{\mathbf{\Sigma}}$)
e(V)	variance–covariance matrix of the estimators
e(b_normalized)	coefficient vector of the normalized parameters
e(V_normalized)	variance–covariance matrix of the normalized parameters

Functions

e(sample)	marks estimation sample

In addition to the above, the following is stored in r():

Matrices

r(table)	matrix containing the coefficients with their standard errors, test statistics, p-values, and confidence intervals

Note that results stored in r() are updated when the command is replayed and will be replaced when any r-class command is run after the estimation command.

Methods and formulas

Methods and formulas are presented under the following headings:

> *Introduction*
> *LES*
> *Generalized translog*
> *QUAIDS with demographic translation*
> *QUAIDS with demographic scaling*
> *Estimation*

Introduction

As in the text in *Remarks and examples*, we continue to use subscript $i = 1, \ldots, N$ to index observations, $g = 1, \ldots, G$ to index goods, and $d = 1, \ldots, D$ to index demographic variables. In formulas where we must use double summations across goods, we use $h = 1, \ldots, G$ and $j = 1, \ldots, G$ to index goods. We omit the observation subscript for most of this discussion; all variables and expressions without i subscripts implicitly refer to the ith observation. Only when we sum across observations do we make the i subscript explicit.

Let w_g be the expenditure share for good g defined as $w_g \equiv (p_g \, q_g)/m$, where p_g is the price of good g, q_g is the quantity of good g consumed, and m denotes total expenditure on all the goods in the system being modeled. We use $\mathbf{w}$, $\mathbf{p}$, and $\mathbf{q}$ to represent the $G \times 1$ vectors of expenditure shares, prices, and quantities for observation i. Let $\mathbf{d}$ represent the $D \times 1$ vector of demographic characteristics.

We use the notation w_g to denote the observed expenditure share for good g for observation i. We use the notation $w_g(\mathbf{p}, m; \boldsymbol{\theta})$ to represent an expenditure-share functions where we will replace generic parameter vector $\boldsymbol{\theta}$ with each model's parameter vectors and matrices. We adorn direct and indirect utility functions and expenditure functions analogously.

We use the notation E_h^g to denote the uncompensated (Marshallian) elasticity of the quantity of good g with respect to the price of good h. We use the notation E^g to note the expenditure elasticity of good g. Then, given the definition of w_g, one can easily verify the following facts that are useful in obtaining the formulas for the elasticities:

$$E^g = \frac{1}{w_g} \frac{\partial w_g}{\partial \ln m} + 1 = \frac{m}{w_g} \frac{\partial w_g}{\partial m} + 1$$

$$E_g^g = \frac{1}{w_g} \frac{\partial w_g}{\partial \ln p_g} - 1 = \frac{p_g}{w_g} \frac{\partial w_g}{\partial p_g} - 1$$

$$E_h^g = \frac{1}{w_g} \frac{\partial w_g}{\partial \ln p_h} \qquad = \frac{p_h}{w_g} \frac{\partial w_g}{\partial p_h}$$

The Slutsky equation can be written in elasticity form to obtain the compensated (Hicksian) elasticity of the quantity of good g with respect to the price of good h: $\widetilde{E}_h^g = E^g \widehat{w}_h + E_h^g$, where $\widehat{w}_h$ is the predicted expenditure share for good h. We discuss elasticities in more detail, including why we use $\widehat{w}_h$ rather than w_h, in [R] **demandsys postestimation**.

For each demand system, we provide the equations for the gth expenditure share, the direct utility function (if available), the indirect utility function, the expenditure function (if available), and the elasticities of the quantity of good g with respect to expenditure and the price of good h. Predicted values of the first four items are available via `predict`; the elasticities are available via `estat elasticities`. When one demand system is nested within a more general demand system, we provide the equations for only the more general demand system and note what parameter restrictions would result in the less general demand system. Moreover, we include demographic variables in our exposition; versions of models without demographic variables result when the corresponding parameter vectors and matrices are set to zero. We also note what parameter restrictions are imposed at estimation to ensure adding up, homogeneity, and Slutsky symmetry.

LES

Let $c_g = \mu_g + \boldsymbol{\nu}_g \mathbf{d}$, where $\boldsymbol{\nu}_g$ is the gth row of $G \times D$ parameter matrix $\mathbf{N}$. The LES begins with the utility function

$$u(\mathbf{q}, \mathbf{d}; \boldsymbol{\beta}, \boldsymbol{\mu}, \mathbf{N}) = \prod_g (q_g - c_g)^{\beta_g}$$

Utility maximization yields the expenditure-share equations

$$w_g(\mathbf{p}, m, \mathbf{d}; \boldsymbol{\beta}, \boldsymbol{\mu}, \mathbf{N}) = \frac{p_g c_g}{m} + \beta_g \left(1 - \sum_h \frac{p_h c_h}{m} \right)$$

To enforce adding up, we impose the constraint $\sum_g \beta_g = 1$. Slutsky symmetry and homogeneity are implied by the functional form of the LES model. Straightforward algebra shows the indirect utility function is

$$v(\mathbf{p}, m, \mathbf{d}; \boldsymbol{\beta}, \boldsymbol{\mu}, \mathbf{N}) = \frac{\prod_g \beta_g^{\beta_g}}{\prod_g p_g^{\beta_g}} \left(m - \sum_g p_g c_g \right)$$

and inverting provides the expenditure function

$$e(\mathbf{p}, u, \mathbf{d}; \boldsymbol{\beta}, \boldsymbol{\mu}, \mathbf{N}) = u \frac{\prod_g p_g^{\beta_g}}{\prod_g \beta_g^{\beta_g}} + \sum_g p_g c_g$$

The expenditure and uncompensated price elasticities are given by

$$E^g = \frac{\beta_g m}{p_g c_g + \beta_g \left(m - \sum_h p_h c_h\right)}$$

$$E_h^g = \begin{cases} \dfrac{p_g c_g (1 - \beta_g)}{p_g c_g + \beta_g \left(m - \displaystyle\sum_h p_h c_h\right)} - 1 & g = h \\[2em] \dfrac{-\beta_g p_h c_h}{p_g c_g + \beta_g \left(m - \displaystyle\sum_h p_h c_h\right)} & g \neq h \end{cases}$$

The Cobb–Douglas demand system results if we set $\boldsymbol{\mu} = \mathbf{0}$. Models without demographics result if we set $\mathbf{N} = \mathbf{0}$.

Generalized translog

Continuing to use $c_g = \mu_g + \boldsymbol{\nu}_g \mathbf{d}$, we can obtain the generalized translog expenditure-share equations with demographics by first obtaining the regular translog expenditure-share equations and then applying translation with both a constant term and demographic characteristics to those expenditure-share equations to obtain

$$w_g(\mathbf{p}, m, \mathbf{d}; \boldsymbol{\alpha}, \boldsymbol{\Gamma}, \boldsymbol{\mu}, \mathbf{N}) = \frac{p_g c_g}{m} + \frac{\overline{m}}{m} \frac{\alpha_g + \sum_h \gamma_{gh} \ln\left(\frac{p_h}{\overline{m}}\right)}{1 + \sum_h \sum_j \gamma_{hj} \ln\left(\frac{p_j}{\overline{m}}\right)}$$

$$\overline{m} = m - \sum_j p_j c_j$$

The normalization $\sum_g \alpha_g = 1$ ensures the expenditure shares sum to 1, and Slutsky symmetry requires that $\gamma_{jh} = \gamma_{hj}$. Homogeneity is implied by the functional form of the expenditure-share equation.

Noting that the translated indirect utility function has the same form as the untranslated version except with $\overline{m}$ replacing m, we have the indirect utility function

$$\ln v(\mathbf{p}, m, \mathbf{d}; \boldsymbol{\alpha}, \boldsymbol{\Gamma}, \boldsymbol{\mu}, \mathbf{N}) = -\sum_g \alpha_g \ln\left(\frac{p_g}{\overline{m}}\right) - \frac{1}{2} \sum_g \sum_h \gamma_{gh} \ln\left(\frac{p_g}{\overline{m}}\right) \ln\left(\frac{p_h}{\overline{m}}\right)$$

A limitation of the generalized translog model is that we cannot solve the indirect utility function for a closed-form expression for the cost function, nor does a closed-form expression for the direct utility function exist. (The direct utility function described in Christensen, Jorgenson, and Lau [1975] is not the utility function that is implied by their indirect utility function.)

Demand systems with translated demographics and committed quantities tend to produce elasticity formulas that are rather long. Omitting function arguments for clarity, we begin by writing the expenditure-share equation for good g as

$$w_g = A_g + B_g \times \frac{N_g}{D_g}$$

where

$$A_g = \frac{p_g c_g}{m}$$

$$B_g = \frac{\overline{m}}{m} = 1 - \frac{\sum_h p_h c_h}{m}$$

$$N_g = \alpha_g + \sum_h \gamma_{gh} \ln p_h - \ln\left(m - \sum_h p_h c_h\right) \sum_h \gamma_{gh}$$

$$D_g = 1 + \sum_j \sum_h \gamma_{jh} \ln p_h - \ln\left(m - \sum_h p_h c_h\right) \sum_j \sum_h \gamma_{jh}$$

Applying the chain and quotient rules of elementary calculus, we have

$$\frac{\partial w_g}{\partial m} = \frac{\partial A_g}{\partial m} + B_g \left(\frac{D_g \dfrac{\partial N_g}{\partial m} - N_g \dfrac{\partial D_g}{\partial m}}{D_g^2} \right) + \frac{N_g}{D_g} \frac{\partial B_g}{\partial m}$$

where

$$\frac{\partial A_g}{\partial m} = -\frac{p_g c_g}{m^2} \qquad\qquad \frac{\partial B_g}{\partial m} = \frac{\sum_h p_h c_h}{m^2}$$

$$\frac{\partial N_g}{\partial m} = -\frac{1}{m - \sum_h p_h c_h} \sum_h \gamma_{gh} \qquad \frac{\partial D_g}{\partial m} = -\frac{1}{m - \sum_h p_h c_h} \sum_j \sum_h \gamma_{jh}$$

Given $\partial w_g / \partial m$, the expenditure elasticity of good g is calculated as $E^g = (m/\widehat{w}_g)(\partial w_g/\partial m) + 1$. We also have

$$\frac{\partial w_g}{\partial p_h} = \frac{\partial A_g}{\partial p_h} + B_g \left(\frac{D_g \dfrac{\partial N_g}{\partial p_h} - N_g \dfrac{\partial D_g}{\partial p_h}}{D_g^2} \right) + \frac{N_g}{D_g} \frac{\partial B_g}{\partial p_h}$$

where

$$\frac{\partial A_g}{\partial p_h} = \begin{cases} c_g/m & h = g \\ 0 & h \neq g \end{cases}$$

$$\frac{\partial B_g}{\partial p_h} = -\frac{c_g}{m}$$

$$\frac{\partial N_g}{\partial p_h} = \frac{\gamma_{gh}}{c_h} + \frac{c_h}{m - \sum_j p_j c_j} \sum_j \gamma_{gj}$$

$$\frac{\partial D_g}{\partial p_h} = \frac{1}{p_h} \sum_j \gamma_{jh} + \frac{c_h}{m - \sum_j p_j c_j} \sum_{jk} \gamma_{jk}$$

Given $\partial w_g / \partial p_h$, the uncompensated price elasticity of good g with respect to price h is calculated as $E_h^g = (p_h/w_g)(\partial w_g/\partial p_h) + \mathbb{I}_{h=g}$, where $\mathbb{I}_x$ is the indicator function that takes on the value one if x is true and zero otherwise.

The basic translog results if we set $\boldsymbol{\mu} = \mathbf{0}$, and models without demographics have $\mathbf{N} = \mathbf{0}$.

QUAIDS with demographic translation

Translating Banks, Blundell, and Lewbel's (1997) indirect utility function, we have

$$\ln V(\mathbf{p}, m, \mathbf{d}; \alpha_0, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Gamma}, \boldsymbol{\lambda}, \boldsymbol{\mu}, \mathbf{N}) = \left[\left\{ \frac{\ln \overline{m} - \ln a(\mathbf{p})}{b(\mathbf{p})} \right\}^{-1} + \lambda(\mathbf{p}) \right]^{-1}$$

$$\ln a(\mathbf{p}) = \alpha_0 + \sum_g \alpha_g \ln p_g + \frac{1}{2} \sum_g \sum_h \gamma_{gh} \ln p_g \ln p_h$$

$$b(\mathbf{p}) = \prod_g p_g^{\beta_g}$$

$$\lambda(\mathbf{p}) = \sum_g \lambda_g \ln p_g$$

where α_0 is the value specified in the `piconstant()` option. Solving for $\overline{m}$ and recalling that $m = \overline{m} + \sum_g p_g c_g$ for $c_g = \mu_g + \nu_g d$, we have the expenditure function

$$e(\mathbf{p}, u, \mathbf{d}; \alpha_0, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Gamma}, \boldsymbol{\lambda}, \boldsymbol{\mu}, \mathbf{N}) = \exp \left\{ \frac{b(\mathbf{p}) \ln u}{1 - \lambda(\mathbf{p}) \ln u} + \ln a(\mathbf{p}) \right\} + \sum_g p_g c_g \tag{6}$$

Applying Shepherd's lemma yields the expenditure-share equations

$$w_g(\mathbf{p}, m, \mathbf{d}; \alpha_0, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Gamma}, \boldsymbol{\lambda}, \boldsymbol{\mu}, \mathbf{N}) =$$

$$\frac{p_g c_g}{m} + \frac{\overline{m}}{m} \left(\alpha_g + \sum_h \gamma_{gh} \ln p_h + \beta_g \ln \left\{ \frac{\overline{m}}{a(\mathbf{p})} \right\} + \frac{\lambda_g}{b(\mathbf{p})} \left[\ln \left\{ \frac{\overline{m}}{a(\mathbf{p})} \right\} \right]^2 \right) \tag{7}$$

To obtain the expenditure elasticities, we first write (7) as

$$w_g = R_g + S_g \times T_g \tag{8}$$

where

$$R_g = \frac{p_g c_g}{m} \qquad S_g = \frac{\overline{m}}{m}$$

$$T_g = \alpha_g + \sum_h \gamma_{gh} \ln p_h + \beta_g \ln \left\{ \frac{\overline{m}}{a(\mathbf{p})} \right\} + \frac{\lambda_g}{b(\mathbf{p})} \left[\ln \left\{ \frac{\overline{m}}{a(\mathbf{p})} \right\} \right]^2$$

Taking the derivative with respect to $\ln m$, we have

$$\frac{\partial w_g}{\partial \ln m} = \frac{\partial R_g}{\partial \ln m} + S_g \frac{\partial T_g}{\partial \ln m} + T_g \frac{\partial S_g}{\partial \ln m}$$

where

$$\frac{\partial R_g}{\partial \ln m} = -\frac{p_g c_g}{m} \qquad \frac{\partial S_g}{\partial \ln m} = 1 - S_g$$

$$\frac{\partial T_g}{\partial \ln m} = \frac{1}{S_g} \left[\beta_g + 2 \frac{\lambda_g}{b(\mathbf{p})} \ln \left\{ \frac{\overline{m}}{a(\mathbf{p})} \right\} \right]$$

Then the expenditure elasticity for good g is given by $E^g = (1/w_g)(\partial w_g / \partial \ln m) + 1$.

To obtain the uncompensated price elasticity of good g with respect to price h, we again use (8). Taking the partial derivative with respect to $\ln p_h$, we have

$$\frac{\partial w_g}{\partial \ln p_h} = \frac{\partial R_g}{\partial \ln p_h} + S_g \frac{\partial T_g}{\partial \ln p_h} + T_g \frac{\partial S_g}{\partial \ln p_h}$$

The required partials are

$$\frac{\partial R_g}{\partial \ln p_h} = \begin{cases} \frac{p_h c_h}{m} & h = g \\ 0 & h \neq g \end{cases}$$

$$\frac{\partial S_g}{\partial \ln p_h} = -\frac{p_h c_h}{m}$$

$$\frac{\partial T_g}{\partial \ln p_h} = \gamma_{gh} + \beta_g \frac{\partial}{\partial \ln p_h} \left[\ln \left\{ \frac{\overline{m}}{a(\mathbf{p})} \right\} \right] + \frac{\partial}{\partial \ln p_h} \left(\frac{\lambda_g}{b(\mathbf{p})} \left[\ln \left\{ \frac{\overline{m}}{a(\mathbf{p})} \right\} \right]^2 \right)$$

where

$$\frac{\partial}{\partial \ln p_h} \left[\ln \left\{ \frac{\overline{m}}{a(\mathbf{p})} \right\} \right] = - \left(\frac{p_h c_h}{m - \sum_j p_j c_j} + \alpha_h + \sum_j \gamma_{hj} \ln p_j \right)$$

$$\frac{\partial}{\partial \ln p_h} \left(\frac{\lambda_g}{b(\mathbf{p})} \left[\ln \left\{ \frac{\overline{m}}{a(\mathbf{p})} \right\} \right]^2 \right) = 2 \frac{\lambda_g}{b(\mathbf{p})} \ln \left\{ \frac{\overline{m}}{a(\mathbf{p})} \right\} \frac{\partial}{\partial \ln p_h} \left[\ln \left\{ \frac{\overline{m}}{a(\mathbf{p})} \right\} \right]$$

$$- \frac{\lambda_g \beta_h}{b(\mathbf{p})} \left[\ln \left\{ \frac{\overline{m}}{a(\mathbf{p})} \right\}^2 \right]$$

The uncompensated price elasticity of good g with respect to price h is $E_h^g = (p_h/w_g)(\partial w_g/\partial p_h) + \mathbb{I}_{h=g}$.

The AIDS model results if $\boldsymbol{\lambda} = \mathbf{0}$. The nongeneralized variants result if $\boldsymbol{\mu} = \mathbf{0}$. The variants without demographics result if $\mathbf{N} = \mathbf{0}$.

QUAIDS with demographic scaling

As can be gleaned from (6), the expenditure function for a QUAIDS model without any form of translation or scaling is

$$e(\mathbf{p}, u; \alpha_0, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Gamma}, \boldsymbol{\lambda}) = \exp \left\{ \frac{b(\mathbf{p}) \ln u}{1 - \lambda(\mathbf{p}) \ln u} + \ln a(\mathbf{p}) \right\} = \exp \left\{ \frac{b(\mathbf{p})}{\frac{1}{\ln u} - \lambda(\mathbf{p})} + \ln a(\mathbf{p}) \right\}$$

$$a(\mathbf{p}) = \alpha_0 + \sum_j \alpha_j \ln p_j + \frac{1}{2} \sum_j \sum_k \gamma_{jk} \ln p_j \ln p_k$$

$$b(\mathbf{p}) = \prod_j p_j^{\beta_j}$$

$$\lambda(\mathbf{p}) = \sum_j \lambda_j \ln p_j$$

The AIDS case results when $\boldsymbol{\lambda} = \mathbf{0}$, and we obtain the expenditure function given by (2).

To implement Ray's (1983) demographic scaling as extended to the QUAIDS case in Poi (2002), we require a function that can be split into two parts, one part that depends on prices and utility and one part that does not:

$$m_0(\mathbf{p}, u, \mathbf{d}) = \overline{m}_0(\mathbf{d})\, \phi(\mathbf{p}, \mathbf{d}, u)$$

For $\overline{m}_0(\mathbf{d})$, we use the same function as in Ray,

$$\overline{m}_0(\mathbf{d}; \boldsymbol{\rho}) = 1 + \boldsymbol{\rho}'\mathbf{d}$$

where $\boldsymbol{\rho}$ is a $D \times 1$ parameter vector. For $\phi(\mathbf{p}, \boldsymbol{\delta}, u)$, we follow Poi (2002) and use

$$\phi(\mathbf{p}, \mathbf{d}, u; \boldsymbol{\beta}, \boldsymbol{\lambda}, \mathbf{H}) = \exp\left\{ \frac{b(\mathbf{p})\left(\prod_j p_j^{\boldsymbol{\eta}_j \mathbf{d}} - 1\right)}{\frac{1}{\ln u} - \lambda(\mathbf{p})} \right\}$$

where $\boldsymbol{\eta}_j$ is the jth row of $G \times D$ parameter matrix $\mathbf{H}$. For AIDS when $\boldsymbol{\lambda} = 0$, we are left with the function used by Ray (1983). The expenditure function for the QUAIDS model with demographic scaling is then

$$e(\mathbf{p}, u, \mathbf{d}; \alpha_0, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Gamma}, \boldsymbol{\lambda}, \boldsymbol{\rho}, \mathbf{H}) = e(\mathbf{p}, u; \alpha_0, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Gamma}, \boldsymbol{\lambda}) \times \overline{m}_0(\mathbf{d}; \boldsymbol{\rho}) \times \phi(\mathbf{p}, \mathbf{d}, u; \boldsymbol{\beta}, \mathbf{H})$$

Applying Shepherd's lemma, we have the share equations

$$w_g(\mathbf{p}, m, \mathbf{d}; \alpha_0, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Gamma}, \boldsymbol{\lambda}, \boldsymbol{\rho}, \mathbf{H}) =$$

$$\alpha_g + \sum_h \gamma_{gh} \ln p_h + (\beta_g + \boldsymbol{\eta}_g \mathbf{d}) \ln\left[\frac{m}{\overline{m}_0(\mathbf{d}; \boldsymbol{\rho})\, a(\mathbf{p})} \right] + \frac{\lambda_g}{b(\mathbf{p})\, c(\mathbf{p}, \mathbf{d})} \left[\ln\left\{ \frac{m}{\overline{m}_0(\mathbf{d}; \boldsymbol{\rho})\, a(\mathbf{p})} \right\} \right]^2$$

where the demographic-adjusted price aggregator is defined as $c(\mathbf{p}, \mathbf{d}) \equiv \prod_j p_j^{\boldsymbol{\eta}_j \mathbf{d}}$. The indirect utility function is

$$\ln V(\mathbf{p}, m, \mathbf{d}; \alpha_0, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\lambda}, \boldsymbol{\rho}, \mathbf{H}) = \left[\frac{b(\mathbf{p})\, c(\mathbf{p}, \mathbf{d})}{\ln\{m - \overline{m}_0(\mathbf{d}; \boldsymbol{\rho}) - a(\mathbf{p})\}} + \lambda(\mathbf{p}) \right]^{-1}$$

Because we are not dealing with demographic translation here, the partial derivatives required for elasticities are not as involved as they are for the QUAIDS model with demographic translation or the generalized translog model. We have

$$\frac{\partial w_g}{\partial \ln m} = \beta_g + \boldsymbol{\eta}_g \mathbf{d} + \frac{2\lambda_g}{b(\mathbf{p})\, c(\mathbf{p}, \mathbf{d})} \ln\left\{ \frac{m}{(1 + \boldsymbol{\rho}'\mathbf{d})\, a(\mathbf{p})} \right\}$$

and

$$\frac{\partial w_g}{\partial \ln p_h} = \gamma_{gh} - \left[\beta_g + \boldsymbol{\eta}_g \mathbf{d} + \frac{2\lambda_g}{b(\mathbf{p})\, c(\mathbf{p}, \mathbf{d})} \ln\left\{ \frac{m}{(1 + \boldsymbol{\rho}'\mathbf{d})\, a(\mathbf{p})} \right\} \right] \times \left(\alpha_h + \sum_k \gamma_{hk} \ln p_k \right)$$

$$+ \frac{(\beta_h + \boldsymbol{\eta}_h \mathbf{d})\, \lambda_g}{b(\mathbf{p})\, c(\mathbf{p}, \mathbf{d})} \left[\ln\left\{ \frac{m}{(1 + \boldsymbol{\rho}'\mathbf{d})\, a(\mathbf{p})} \right\} \right]^2$$

The AIDS model with demographic scaling results if $\boldsymbol{\lambda} = 0$, and models without demographic scaling result if $\boldsymbol{\rho} = \mathbf{0}$ and $\mathbf{H} = \mathbf{0}$.

Estimation

The expenditure shares of a demand system represent a set of nonlinear seemingly unrelated regression (SUR) equations. Hence, estimation is performed using nlsur with just one complication. Our system of equations for observation i and generic parameter vector $\boldsymbol{\theta}$ can be written as

$$w_{1i} = w_1(\mathbf{p}_i, m_i, \mathbf{d}_i; \boldsymbol{\theta}) + \epsilon_{1i}$$
$$w_{2i} = w_2(\mathbf{p}_i, m_i, \mathbf{d}_i; \boldsymbol{\theta}) + \epsilon_{2i}$$
$$\vdots \quad = \qquad \vdots$$
$$w_{Gi} = w_G(\mathbf{p}_i, m_i, \mathbf{d}_i; \boldsymbol{\theta}) + \epsilon_{Gi}$$

where $\boldsymbol{\epsilon}_i = (\epsilon_{1i}, \epsilon_{2i}, \ldots, \epsilon_{Gi})$ is a vector of zero-mean disturbances. Because $\sum_g w_{gi} = 1$ and by the construction of our demand systems we also have $\sum_g w_g(\mathbf{p}_i, m_i, \mathbf{d}_i; \boldsymbol{\theta}) = 1$, it must be the case that $\sum_g \epsilon_{gi} = 0$. Therefore, $\det(\boldsymbol{\Sigma}) = \det\{E(\boldsymbol{\epsilon}\boldsymbol{\epsilon}')\} = \mathbf{0}$, and neither the (quasi)maximum likelihood estimator nor the feasible generalized nonlinear SUR estimator is defined.

Barten (1969) showed that to obtain a well-defined likelihood function, we can drop any one of the expenditure-share equations and fit the system containing the remaining $G - 1$ equations. Parameter restrictions that force the demand system's equations to sum to 1 can then be used to recover the parameters of the dropped equation. demandsys drops the final equation from the demand system, but which equation is dropped should not concern you: Barten (1969) showed that you obtain the same likelihood function regardless of which one is dropped.

Moreover, the nonlinear SUR model satisfies the so-called Oberhofer–Kmenta (1974) conditions that allow us to cycle between estimating the parameters $\boldsymbol{\theta}$ and the error covariance matrix $\boldsymbol{\Sigma}$ as nlsur does when we specify option ifgnls. The upshot is that we fit our expenditure-share equations by calling nlsur with option ifgnls. The resulting parameter estimate $\widehat{\boldsymbol{\theta}}$ is the (quasi)maximum likelihood estimate of the parameters of our demand system. The $(G - 1) \times (G - 1)$ estimated error covariance matrix is returned in matrix e(Sigma).

Estimation weights, variance–covariance matrix options, and options to control the optimization process available with demandsys are simply passed to nlsur. See *Methods and formulas* in [R] **nlsur**.

When you type predict ..., shares eq(#), you obtain the predicted shares for the #th equation, based on the expenditure-share equation for the demand system estimated. By construction, the sum of predicted shares for each observation is equal to one.

When you type predict ..., residuals eq(#), you obtain the residuals for the #th equation defined as $r_{\#i} = w_{\#i} - \widehat{w_{\#i}}$, where $\widehat{w_{\#i}}$ is the predicted share for equation # for observation i.

When you type predict ..., quantities eq(#), predict first computes the predicted shares $\widehat{w_{\#i}}$ and then computes $\widehat{q_{\#i}} = m_i \widehat{w_{\#i}} / p_{\#i}$.

When you type predict ..., iuf, predict evaluates the indirect utility function based on the formulas given above.

When you type predict ..., ef utilities(u), predict evaluates the expenditure function based on the formulas given above at the levels of utility specified in variable u.

Acknowledgment

References

Banks, J., R. Blundell, and A. Lewbel. 1997. Quadratic Engle curves and consumer demand. *Review of Economics and Statistics* 79: 527–539. https://doi.org/10.1162/003465397557015.

Barnett, W. A., and A. Serletis. 2008. Consumer preferences and demand systems. *Journal of Econometrics* 147: 210–224. https://doi.org/10.1016/j.jeconom.2008.09.009.

Barten, A. P. 1966. Theorie en empirie van een volledig stelsel van vraagvergelijkingen. PhD thesis, University of Rotterdam.

———. 1969. Maximum likelihood estimation of a complete system of demand equations. *European Economic Review* 1: 7–73. https://doi.org/10.1016/0014-2921(69)90017-8.

Caro, J. C., G. Melo, J. A. Molina, and J. C. Salgado. 2021. Censored demand system estimation with quaidsce. Presented at the Stata Conference 2021, August 6, 2021, virtual. https://www.stata.com/meeting/us21/slides/US21_Melo.pdf.

Case, A. C. 1991. Spatial patterns in household demand. *Econometrica* 59: 953–965. https://doi.org/10.2307/2938168.

Christensen, L. R., D. W. Jorgenson, and L. J. Lau. 1975. Transcendental logarithmic utility functions. *American Economic Review* 65: 367–383. https://doi.org/10.2307/1927992.

Deaton, A. S., and M. Irish. 1984. Statistical models for zero expenditures in household budgets. *Journal of Public Economics* 23: 59–80. https://doi.org/10.1016/0047-2727(84)90067-7.

Deaton, A. S., and J. Muellbauer. 1980a. *Economics and Consumer Behavior*. Cambridge: Cambridge University Press.

———. 1980b. An almost ideal demand system. *American Economic Review* 70: 312–326.

Fisher, D., A. R. Fleissig, and A. Serletis. 2001. An empirical comparison of flexible demand system functional forms. *Journal of Applied Econometrics* 16: 59–80. https://doi.org/10.1002/jae.585.

Gorman, W. M. 1981. Some Engle curves. In *Essays in the Theory and Measurement of Consumer Behavior: In Honour of Sir Richard Stone*, ed. A. Deaton, 7–30. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511984082.003.

Heien, D., and C. R. Wessells. 1990. Demand systems estimation with microdata: A censored regression approach. *Journal of Business and Economic Statistics* 8: 365–371. https://doi.org/10.2307/1391973.

Holt, M. T., and B. K. Goodwin. 2009. The almost ideal and translog demand systems. MPRA Paper 15092, University Library, Ludwig Maximilians University, München, Germany. https://mpra.ub.uni-muenchen.de/15092/.

Lecocq, S., and J.-M. Robin. 2015. Estimating almost-ideal demand systems with endogenous regressors. *Stata Journal* 15: 554–573.

Leser, C. E. V. 1963. Forms of Engel functions. *Econometrica* 31: 694–703. https://doi.org/10.2307/1909167.

Lewbel, A. 1989. Household equivalence scales and welfare comparisons. *Journal of Public Economics* 39: 377–391. https://doi.org/10.1016/0047-2727(89)90035-2.

———. 1991. The rank of demand systems: Theory and nonparametric estimation. *Econometrica* 59: 711–730. https://doi.org/10.2307/2938225.

———. 2001. Demand systems with and without errors. *American Economic Review* 91: 611–618. https://doi.org/10.1257/aer.91.3.611.

Lewbel, A., and K. Pendakur. 2009. Tricks with Hicks: The EASI demand system. *American Economic Review* 99: 827–863. https://doi.org/10.1257/aer.99.3.827.

Mas-Colell, A., M. D. Whinston, and J. R. Green. 1995. *Microeconomic Theory*. New York: Oxford University Press.

Meyerhoefer, C. D., C. K. Ranney, and D. E. Sahn. 2005. Consistent estimation of censored demand systems using panel data. *American Journal of Agricultural Economics* 87: 660–672. https://doi.org/10.1111/j.1467-8276.2005.00754.x.

Muellbauer, J. 1975. Aggregation, income distribution, and consumer demand. *Review of Economic Studies* 42: 525–543. https://doi.org/10.2307/2296792.

Nelson, D. B. 1991. Conditional heteroskedasticity in asset returns: A new approach. *Econometrica* 59: 347–370. https://doi.org/10.2307/2938260.

Oberhofer, W., and J. Kmenta. 1974. A general procedure for obtaining maximum likelihood estimates in generalized regression models. *Econometrica* 42: 579–590. https://doi.org/10.2307/1911792.

Poi, B. P. 2002. From the help desk: Demand system estimation. *Stata Journal* 2: 403–410.

——. 2012. Easy demand-system estimation with quaids. *Stata Journal* 12: 433–446.

Pollak, R. A., and T. J. Wales. 1978. Estimation of complete demand systems from household budget data: The linear and quadratic expenditure systems. *American Economic Review* 68: 348–359.

——. 1992. *Demand System Specification and Estimation.* New York: Oxford University Press.

Ray, R. 1983. Measuring the costs of children: An alternative approach. *Journal of Public Economics* 22: 89–102. https://doi.org/10.1016/0047-2727(83)90058-0.

Roy, P. R. 1943. La hiérarchie des besoins et la notion de groupes dans l'économie de choix. *Econometrica* 11: 13–24. https://doi.org/10.2307/1905715.

Shephard, R. W. 1970. *Theory of Cost and Production Functions.* Princeton, NJ: Princeton University Press.

Shonkwiler, J. S., and S. T. Yen. 1999. Two-step estimation of a censored system of equations. *American Journal of Agricultural Economics* 81: 972–982. https://doi.org/10.2307/1244339.

Stone, R. 1954. Linear expenditure systems and demand analysis: An application to the pattern of British demand. *Economic Journal* 64: 511–527. https://doi.org/10.2307/2227743.

Tauchmann, H. 2010. Consistency of Heckman-type two-step estimators for the multivariate sample-selection model. *Applied Economics* 42: 3895–3902. https://doi.org/10.1080/00036840802360179.

Theil, H. 1965. The information approach to demand analysis. *Econometrica* 33: 67–87. https://doi.org/10.2307/1911889.

Varian, H. R. 1992. *Microeconomic Analysis.* 3rd ed. New York: W. W. Norton.

Vermeulen, F. 2001. A note on Heckman-type corrections in models for zero expenditures. *Applied Economics* 33: 1089–1092. https://doi.org/10.1080/00036840010004004.

Working, H. 1943. Statistical laws of family expenditure. *Journal of the American Statistical Association* 38: 43–56. https://doi.org/10.2307/2279311.

Yen, S. T., and B.-H. Lin. 2006. A sample selection approach to censored demand systems. *American Journal of Agricultural Economics* 88: 742–749. https://doi.org/10.1111/j.1467-8276.2006.00892.x.

Young, W. H. 1909. On the conditions for the reversibility of the order of partial differentiation. *Proceedings of the Royal Society of Edinburgh, Series B: Biological Sciences* 29: 136–164. https://doi.org/10.1017/S0370164600008865.

Also see

[R] **demandsys postestimation** — Postestimation tools for demandsys

[R] **nlsur** — Estimation of nonlinear systems of equations

[U] **20 Estimation and postestimation commands**