The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search
http://ageconsearch.umn.edu
aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

Maintaining Parameter Invariance in Censored Micro-Level Data

Jay M. Lillywhite
Department of Ag. Economics and Ag. Business
New Mexico State University
Las Cruces, NM 88003-8003
e-mail: lillywhi@nmsu.edu

Paul V. Preckel
Department of Agricultural Economics
Purdue University
West Lafayette, IN 47907-1145
preckel@purdue.edu

James Eales
Department of Agricultural Economics
Purdue University
West Lafayette, IN 47907-1145
eales@purdue.edu

Selected Paper prepared for presentation at the American Agricultural Economics Association Annual Meeting, Providence, Rhode Island, July 24-27, 2005

Copyright 2005 by Jay M. Lillywhite, Paul V. Preckel, and James Eales. All rights reserved. Readers may make verbatim copies of this document for non-commercial purposes by any means, provided that this copyright notice appears on all such copies.

As information technologies develop, researchers have access to unprecedented amounts of individual or micro-level data (e.g., data from surveys and scanner data). The availability of data, combined with interest in understanding demographic differences in consumer behavior, has encouraged researchers to use micro-level data in demand analysis (Capps and Love, Jensen, Cotterill). Micro-level data provide new opportunities as well as challenges in demand estimation. A primary challenge surrounding micro-level demand analysis with commonly used (generalized) least squares estimation procedures is the development of a computationally workable model that ensures non-negativity of predicted quantities and that incorporates constraints implied by economic theory (Dong, Gould, and Kaiser).

A key issue in much of the micro-level demand analysis performed to date is ensurance of parameter invariance while imposing adding up constraints. In the absence of non-consumption (e.g., aggregate data where individual non-consumption is concealed), the imposition of adding-up restrictions in a system of share equations is straightforward. In this case, the conventional method of estimating the system of equations requires that one share equation, along with its corresponding row and column of the error covariance matrix, be dropped. The system of (n-1) equations is estimated, and "adding-up" conditions are used to recover missing parameter values. With aggregate data and a well-defined demand system, parameters estimated with a maximum likelihood approach will be invariant to which equation is eliminated from the system (Barten).

When micro-level data with observations exhibiting non-consumption (censored micro-level data) are used with an Amemiya-Tobin censoring specification in demand analysis the conventional method of imposing adding-up will no longer provide satisfactory results. In this case, demand systems containing censored data will not have identical regressors and parameter estimates will not be invariant. That is, parameter estimates obtained from the econometric estimation will vary depending on which share equation is dropped from the estimated system.

Parameter Invariance in Censored Demand Systems

Researchers have attempted to solve the problem of parameter estimates that vary depending on which share equation is dropped using a number of approaches. Pudney summarizes several alternative approaches for handling the adding-up conditions when estimating a censored demand system in the presence of a budget constraint.

The first, and probably the most popular, approach is to treat one of the expenditure categories as a residual with no specification of its own. In this case, the estimated model would consist of (n-1) equations. The share value for the designated residual category is defined as the difference between one and the sum of the first (n-1) shares. While this approach is simple and in some sense addresses the adding up issue, it fails to account for the parameter invariance problem. That is, the selection of a different "residual" category will typically result in different parameter values. This approach has been used by a number of researchers including: Yen, Lin, and Smallwood and Yen and Huang.

A second approach identified by Pudney is to modify the Tobit model when imposing adding-up. An example of this approach is Heien and Wessells' two-step estimation procedure. The first step of the two-step estimation requires estimation of probit regressions to determine the probability that a given consumer will consume each of the goods in question. The results of the probit regressions are then used to compute inverse Mills ratios for each consumer. The inverse Mills ratios are used as independent variables in the second stage of the demand estimation. To ensure adding-up of the system, this approach implicitly requires the omitted equation to include the negative of the sum of the inverse Mills ratios as an independent variable. While accounting for adding-up within the system, this approach, like to previous one, does not produce parameter estimates that are invariant to which equation is omitted.

An alternative to either of the above approaches is to consider the resource constraint assumed in consumer planning as an *ex ante* rather than an *ex post* concept. That is, planned expenditures will satisfy the adding-up restriction, but actual expenditures (those estimated with econometric techniques) may not add up, in part as the result of "accidents, whims, or mistakes" (Pudney, p. 156). This approach, in a sense, ignores the problem of adding up and the corresponding problem of parameter invariance by avoiding dropping a share equation and omitting the cross equation constraints that would be imposed on parameter estimates in the presence of adding up.

While the alternatives discussed above provide means for imposing adding-up (or describe why adding-up may not hold in actual data), they fail to provide a satisfying solution for obtaining a unique set of parameter estimates when working with censored data. From a strictly econometric standpoint this lack of invariance is not an issue as the parameter values are consistent, but from an applied standpoint the lack of invariance of both parameter estimates and elasticity estimates can be disconcerting. This paper suggests a new method of demand estimation (hereafter referred to as INvariant Seemingly Unrelated Regression or INSUR) that provides an invariant set of parameter estimates. While it does not correct problems associated with ensuring predicted shares sum to one, it can be used with existing procedures (e.g., either the first or second approaches suggested by Pudney and reviewed above) to ensure that parameters estimated by these procedures are at least invariant to which equation is chosen to be dropped or to be modified.

Seemingly Unrelated Regression (SUR) Estimation

Before describing the INSUR approach to demand estimation we provide a brief review of standard SUR estimation commonly used in demand analysis. Zellner's SUR approach is an iterative process that begins with an initial step of estimating equation system parameters using ordinary least squares (OLS). Parameter values obtained from initial OLS regressions are used to estimate residuals between the actual and predicted dependent variables, which are then used to estimate the residual or error covariance matrix. The inverse of the estimated error covariance matrix obtained from this step is then used to weight the errors in order to account for cross-equation correlations. If the procedure is iterated, the process above continues, alternating between coefficient estimation and estimation of the updated error covariance matrix until the parameter estimates and the error covariance matrix converge (Zellner; Judge, et al.).

Consider a linear approximate almost ideal demand system (LA/AIDS) demand system with N goods and T consumers. If the system is defined in share form, where the expenditure share for good i depends on prices and consumer income, the system of share equations can be written as:

$$Y_{1t} = \beta_{10} + \sum_{j=1}^{N} \beta_{1j} X_{jt} + \varepsilon_{1t} \tag{1}$$

$$Y_{2t} = \beta_{20} + \sum_{j=1}^{N} \beta_{2j} X_{jt} + \varepsilon_{2t}$$

$$\vdots$$

$$Y_{Nt} = \beta_{N0} + \sum_{j=1}^{N} \beta_{Nj} X_{jt} + \varepsilon_{Nt}$$

This same demand system can be written in matrix notation as

$$\mathbf{Y}_i = \mathbf{X}_i \boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_i \qquad (2)$$

where the subscript i denotes the i-th equation, and the superscript t refers to the transpose operator, with

$$\mathbf{Y}_i = \begin{bmatrix} Y_{i1} \\ Y_{i2} \\ \vdots \\ Y_{iT} \end{bmatrix} \qquad (3)$$

$$\mathbf{X}_i = \begin{bmatrix} 1 & X_{11} & X_{21} & \cdots & X_{N1} \\ 1 & X_{12} & X_{22} & \cdots & X_{N2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{1T} & X_{2T} & \cdots & X_{NT} \end{bmatrix} \qquad (4)$$

$$\boldsymbol{\beta}_i = \begin{bmatrix} \beta_{i0} \\ \beta_{i1} \\ \beta_{i2} \\ \vdots \\ \beta_{iN} \end{bmatrix} \qquad (5)$$

and

$$\boldsymbol{\varepsilon}_i = \begin{bmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \vdots \\ \varepsilon_{iT} \end{bmatrix} \tag{6}$$

If the demand system is stacked, the complete demand system (representing all consumers and all goods) could be expressed as

$$\begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \vdots \\ \mathbf{Y}_N \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 & & & \\ & \mathbf{X}_2 & & \\ & & \ddots & \\ & & & \mathbf{X}_N \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \vdots \\ \boldsymbol{\beta}_{\mathrm{N}} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \vdots \\ \boldsymbol{\varepsilon}_N \end{bmatrix} \tag{7}$$

Or in more compact notation as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}\ . \tag{8}$$

The SUR approach described above assumes contemporaneous error correlation exists (i.e., errors across equations are correlated) but serial correlation does not exist (i.e., errors across consumers are not correlated). These assumptions imply E[$\varepsilon_{ti}\varepsilon_{sj}$] = σ_{ij} if t = s, but zero if t $\neq$ s. The covariance matrix can be written as

$$\begin{bmatrix} \sigma_{11}I_T & \sigma_{12}I_T & \cdots & \sigma_{1N}I_T \\ \sigma_{21}I_T & \sigma_{22}I_T & \cdots & \sigma_{2N}I_T \\ \vdots & \vdots & \ddots & \\ \sigma_{N1}I_T & \sigma_{N2}I_T & \cdots & \sigma_{NN}I_T \end{bmatrix} = \Sigma \otimes I_T \tag{9}$$

where

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1N} \\ \sigma_{12} & \sigma_{22} & \cdots & \sigma_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{N1} & \sigma_{N2} & \cdots & \sigma_{NN} \end{bmatrix} \tag{10}$$

Incorporating this error structure into the model (substituting $\hat{\Sigma}$ in for the unknown Σ) the generalized least squares estimator for $\boldsymbol{\beta}$ can be written as

$$\hat{\boldsymbol{\beta}} = \left[\mathbf{X}'\left(\hat{\Sigma}^{-1} \otimes I\right)\mathbf{X}\right]^{-1} \mathbf{X}'\left(\hat{\Sigma}^{-1} \otimes I\right)\mathbf{Y} . \tag{11}$$

When linear restrictions in the form $R\boldsymbol{\beta} = \mathbf{r}$ are imposed on the system the estimator above must be modified so that the generalized least squares estimator for $\boldsymbol{\beta}$ is written as (Judge, et. al. 1988):

$$\hat{\hat{\boldsymbol{\beta}}}^{*} = \hat{\hat{\boldsymbol{\beta}}} + \hat{C}R^{t}\left(R\hat{C}R^{t}\right)^{-1}\left(\mathbf{r} - R\hat{\hat{\boldsymbol{\beta}}}\right). \tag{12}$$

where

$$\hat{C} = \left[\mathbf{X}^{t}\left(\hat{\Sigma}^{-1} \otimes I\right)\mathbf{X}\right]^{-1}. \tag{13}$$

and

$$\hat{\hat{\boldsymbol{\beta}}} = \hat{C}\mathbf{X}^{t}\left(\hat{\Sigma}^{-1} \otimes I\right)\mathbf{Y} . \tag{14}$$

The covariance matrix for the coefficients of the restricted SUR estimation procedure is calculated as:

$$\hat{C} - \hat{C}R'\left(R\hat{C}R'\right)R\hat{C} . \tag{15}$$

Invariant Seemingly Unrelated Regression (INSUR) Estimation

As explained earlier, when estimating a micro-level demand system using the traditional approach of dropping one equation and recovering the dropped equation’s parameter estimates from theoretical restrictions, n different sets of parameter estimates will be obtained from each of the n possible systems. The INSUR approach, which provides consistent and invariant parameter estimates, sums the objectives over these n sub-systems of (n-1) equations into one estimation objective function. Thus, using notation similar to Equation 8 above, the objective function to be minimized over the parameters for the INSUR procedure can be written as:

$$\sum_{i=1}^{n} \boldsymbol{\varepsilon}_i{}' \left(\boldsymbol{\Sigma}_i^{-1} \otimes \mathbf{I}\right) \boldsymbol{\varepsilon}_i \qquad (16)$$

where $\boldsymbol{\varepsilon}_\text{i}$ represents the stacked error vector with the errors corresponding to the *i*-th equation omitted and $\left(\boldsymbol{\Sigma}_i^{-1} \otimes \mathbf{I}\right)$ represents elements from the inverted error covariance matrix with the *i*-th row and column removed. The minimization is conducted subject to standard theoretical parameter restrictions used in SUR. Note that objective function in the INSUR approach is simply the summation of the possible n systems that could be estimated in standard SUR demand analysis.

Equivalence of SUR and INSUR in Non-Censored Data: An Example

For demand systems that satisfy the usual invariance property, e.g., systems with non-censored data, the INSUR method produces parameter estimates that are identical to those produced by SUR. In addition, standard errors for each methodology are identical. In this section we illustrate the equivalence of standard SUR estimation and INSUR estimation when non-censored data is used.

Data

Our example uses a LA/AIDS system functional form and aggregated data from a survey conducted by the National Livestock and Meat Board during a five-month period between November 1993 and March 1994. A randomly selected sample of 1,057 households kept a diary of their meat purchases during the five-month period in which they participated.

The sample identified twenty-two different types of meats consumed by one or more households. To simplify our illustration, these twenty-two different meat types were aggregated into three different meats, beef, poultry and pork. After removing non-useable responses and observations in which all three meat types were consumed in positive amounts, the resulting database contained 757 observations.

Functional Form

For purposes of our example a LA/AIDS with a "corrected" Stone's price index is used as the functional form for the demand estimation. The LA/AIDS system, developed by Deaton and Muellbauer is derived from the Almost Ideal Demand System (AIDS) cost function. The AIDS cost function (in log form) is defined as:

$$\ln c(p,u) = \alpha_0 + \sum_{j=1}^{N} \alpha_j \ln p_{tj} + \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \gamma_{ij}^{*} \ln p_{ti} \ln p_{tj} + u\beta_0 \prod_j p_{tj}^{\beta_j} \tag{17}$$

where $\ln c(p,u)$ is the logarithm of the cost function, p_{ti} is the market price faced by consumer t for commodity i, and u is the consumer's utility level.

Using Shephard's lemma the compensated share equations can be derived as

$$w_{ti}(p,u) = \alpha_i + \sum_{j=1}^{N} \gamma_{ij} \ln p_{tj} + \beta_i u \beta_0 \prod_j p_{tj}^{\beta_j} \ . \tag{18}$$

To uncompensate the share equation, the log cost function (17) is inverted to obtain the indirect utility function. The utility level, u, in equation (18) is replaced with the indirect utility obtained from the inversion to provide the uncompensated share equation

$$w_{ti}(p,m) = \alpha_i + \sum_{j=1}^{N} \gamma_{ij} \ln p_{tj} + \beta_i \left(\ln m_t - \ln P_t\right) \tag{19}$$

where m_t is expenditure (or income) for consumer t, and P_t is a price index (a corrected Stone's price index in our case). The "corrected" Stone's price index, the log-linear analogue of the Paasche price index, has the form:

$$\log(P_t) = \sum_{i=1}^{N} w_{ti} \ln\left(\frac{p_{ti}}{p_i^0}\right) \tag{20}$$

Here w_{ti} is the *i*-th share equation for consumer t, p_{ti} is the price associated with the *i*-th good for consumer t, and p_i^0 is a base price (in this case p_i^0 is set equal average price in the sample[1]). With the "corrected" Stone's price index incorporated into the model the share equations are defined by

$$w_{ti}(p,m) = \alpha_i + \sum_{j=1}^{N} \gamma_{ij} \ln p_{tj} + \beta_i\left(\ln m_t - \sum_{j=1}^{N} w_{tj} \ln\left(\frac{p_{tj}}{\overline{p}_j}\right)\right) \tag{21}$$

Finally, in order to empirically estimate the model, a specification for the stochastic nature of the model must be developed. For ease of exposition a homoskedastic, normally distributed error term is added to each share equation so that the final share equation has the form:

$$w_{ti}(p,m) = \alpha_i + \varepsilon_{ti} + \sum_{j=1}^{N} \gamma_{ij} \ln p_{tj} + \beta_i\left(\ln m_t - \sum_{j=1}^{N} w_{tj} \ln\left(\frac{p_{tj}}{\overline{p}_j}\right)\right) \tag{22}$$

Results

As indicated above, the INSUR approach provides parameter estimates and corresponding standard errors that are equal to standard SUR estimates and errors when data does not exhibit censoring. Using the data described above and the LA/AIDS

---

1 Moschini suggests that using the mean value for the base may be more appropriate than other options (e.g., the first period observation). In cross-sectional data sets the mean provides a more appealing base as any type of ordering among survey observations would tend to be subjective.

functional form also reviewed above the following demand results were obtained. Italicized parameter values in the standard SUR section of the table indicate parameter values that were obtained through theoretical restrictions (values not directly estimated econometrically due to singularity of the covariance matrix. Standard errors for these parameters were not estimated and are thus not provided in the table).

Parameter values are equivalent between standard SUR estimation and INSUR estimation. Standard errors for parameter values are also equivalent between the two methods (note slightly different standard error estimates are the result of rounding errors).[2]

Conclusion

Imposing adding-up restraints in demand systems using censored micro-level data to date has been "one of the notorious stigmata in censored demand systems" (Yen, Lin, and Smallwood, p. 460). A key issue in much of the micro-level demand analysis performed to date is insurance of parameter invariance while imposing adding up constraints. This paper has introduced a new methodology for estimating demand systems that ensures parameter invariance when working with censored demand data and SUR estimation.

When applied to situations where parameter invariance holds, e.g., non-censored data in which all individual demands within the system have the same explanatory variables, the INSUR procedure provides parameter estimates and standard errors that are identical to those obtained through the standard approach of dropping one equation. While this procedure is not currently available in pre-programmed econometric software,

[2] INSUR parameters were estimated using GAMS. For ease of calculation standard errors were calculated in GAUSS. The transition between the two programs resulted in slight differences in standard error estimates.

more flexible self-programming packages such as Gauss or GAMS can be used to implement it. Moreover, as the use of micro-level data and censored demand systems becomes more common, perhaps econometric software will incorporate the procedure proposed here as an option.

References

Barten, A.P. "Maximum Likelihood Estimation of a Complete System of Demand Equations." *European Economic Review* 1 (1969): 7-73.

Capps Jr., O. and H.A. Love. "Econometric Considerations in the Use of Electronic Scanner Date to Conduct Consumer Demand Analysis." *American Journal of Agricultural Economics*. 84(August 2002):807–816.

Cotterill, R.W. "Scanner Data: New Opportunities for Demand and Competitive Strategy Analysis." *Journal of Agriculture and Resource Economics* 25(October 1994):125–139.

Deaton, A. and J. Muellbauer. "An Almost Ideal Demand System." *American Economic Review* 70(1980):312–326.

Dong, Diansheng, Brian W. Gould, and Harry M. Kaiser. "Food Demand in Mexico: An Application of the Amemiya-Tobin Approach to the Estimation of a Censored Food System." *American Journal of Agricultural Economics* 86(4) (November 2004):1094–1107.

Heien, D. and C.R. Wessells. Demand Systems Estimation with Microdata: A Censored Regression Approach. *Journal of Business and Economic Statistics* 8(3) (July 1990): 365–371.

Jensen, H.H. "Empirical Analysis Using Scanner Data: Econometric Issues and Policy Implications." *American Journal of Agricultural Economics*. 84(August 2002):824–825.

Judge, George, G., R. Carter Hill, William E. Griffiths, Helmut Lutkepohl, and Tsoung-Chao Lee. *Introduction to the Theory and Practice of Econometrics*, Second Edition. New York: John Wiley and Sons, 1988.

Moschini, Giancarlo. "Units of Measurement and the Stone Index in Demand System Estimation." *American Journal of Agricultural Economics* 77(1995):63–68.

Pudney, S. Modelling Individual Choice: The Econometrics of Corners, Kinks, and Holes. Basil Blackwell, Inc. New York, NY. 1989.

Yen, Steven T. and Chung L. Huang. "Cross-Sectional Estimation of U.S. Demand for Beef Products: A Censored System Approach." Journal of Agricultural and Resource Economics 27(2) (December 2002):320–334.

Yen, S.T., B.H. Lin, and D.M. Smallwood. "Quasi- and Simulated-Likelihood Approaches to Censored Demand Systems: Food Consumption by Food Stamp

Recipients in the United States." *American Journal of Agricultural Economics*. 85(2) (May 2003): 458–478.

Zellner, Arnold. "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias." *Journal of the American Statistical Association* 57(1962):348–368.

Table 1. Parameter Estimates via SUR and INSUR – An Invariant Case

	Standard SUR Equations in estimated system: Equations (1) & (2)		Equations (1) & (3)		Equations (2) & (3)		INSUR	
Parameter	Parameter Estimate	Standard Errors	Parameter Estimate	Standard Errors	Parameter Estimate	Standard Errors	Parameter Estimate	Standard Errors
α_1	0.2952	0.0417	0.2952	0.0417	*0.2952*		0.2952	0.0416
α_2	0.3719	0.0381	*0.3719*		0.3719	0.0381	0.3719	0.0380
α_3	*0.3329*		0.3329	0.0374	0.3329	0.0374	0.3329	0.0373
γ_{11}	0.0020	0.0217	0.0020	0.0217	*0.0020*		0.0020	0.0217
γ_{12}	-0.0121	0.0131	-0.0121	0.0131	*-0.0121*		-0.0121	0.0131
γ_{13}	0.0101	0.0182	0.0101	0.0182	*0.0101*		0.0101	0.0181
γ_{21}	-0.0121	0.0131	*-0.0121*		-0.0121	0.0131	-0.0121	0.0131
γ_{22}	0.0435	0.0136	*0.0435*		0.0435	0.0136	0.0435	0.0136
γ_{23}	-0.0314	0.0119	*-0.0314*		-0.0314	0.0119	-0.0314	0.0119
γ_{31}	*0.0101*		0.0101	0.0182	0.0101	0.0182	0.0101	0.0181
γ_{32}	*-0.0314*		-0.0314	0.0119	-0.0314	0.0119	-0.0314	0.0119
γ_{33}	*0.0213*		0.0213	0.0203	0.0213	0.0203	0.0213	0.0203
β_1	0.0314	0.0086	0.0314	0.0086	*0.0314*		0.0314	0.0085
β_2	-0.0170	0.0078	*-0.0170*		-0.0170	0.0078	-0.0170	0.0078
β_3	*-0.0144*		-0.0144	0.0076	-0.0144	0.0076	-0.0144	0.0076

* Equation definitions (1) beef, (2) poultry, (3) pork.