The Effect of Expenditure Aggregation on Hypothesis Tests in Consumer Demand Systems

Christopher J. Nicol


Stable URL: http://links.jstor.org/sici?sici=0020-6598%28199105%2932%3A2%3C405%3ATEOEAO%3E2.0.CO%3B2-6

*International Economic Review* is currently published by Economics Department of the University of Pennsylvania.

________________________________________________________________________

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at http://www.jstor.org/about/terms.html. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at http://www.jstor.org/journals/ier_pub.html.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

________________________________________________________________________

The JSTOR Archive is a trusted digital repository providing for long-term preservation and access to leading academic journals and scholarly literature from around the world. The Archive is supported by libraries, scholarly societies, publishers, and foundations. It is an initiative of JSTOR, a not-for-profit organization with a mission to help the scholarly community take advantage of advances in technology. For more information regarding JSTOR, please contact support@jstor.org.
THE EFFECT OF EXPENDITURE AGGREGATION ON HYPOTHESIS TESTS IN CONSUMER DEMAND SYSTEMS*

BY CHRISTOPHER J. NICOL

In this paper, cluster analysis is used to identify "stochastic Hicksian aggregates" (SHA). Such aggregates comprise goods for which relative prices are constant, up to a random element. The need to identify such aggregates arises since inappropriate aggregation could influence subsequent estimation and tests results. Evidence is found that inappropriate aggregation of expenditures influences the outcome of hypothesis tests. Furthermore, cluster analysis appears to identify goods which can be treated as SHA. It would thus appear that these methods should be used to identify the appropriate goods to aggregate, prior to the estimation or testing of demand models.

1. INTRODUCTION

Under certain conditions, aggregation of expenditures into a limited number of categories could affect estimation and test results in applied demand studies. (Some evidence of this is provided in Nicol, 1985.) It would therefore be useful if aggregates could be constructed in a way which minimised such effects. One way to implement this principle would be to rely on the separable structure of preferences and infer appropriate groupings. This has the advantage that the aggregates generated would be economically meaningful. A difficulty, however, is that the source of the rule for determining appropriate separable groups has to be devised.

One technique of determining separable groups was developed by Pudney (1981). In this method, a demand system was estimated and Slutsky cross-elasticities were computed. Groupings were then defined based on these Slutsky elasticities, in terms of separability definitions. The main drawback of this approach is that a demand system must first be specified, to estimate the Slutsky elasticities. Specification error could render the estimated elasticities inconsistent. Aggregates based on these estimates would then be misleading, and subsequently constructed estimates or tests rendered invalid.

As Lewbel (1988) states, the usual justification for aggregation is based on Hicks' (1936) composite commodity theorem. This permits the aggregation of goods with perfectly correlated prices. In practice, however, data on prices are not perfectly correlated. Further, even if prices of certain goods are highly correlated, grouping...
them can still imply that the disturbances of demand models based on these data are systematically affected. Lewbel (1988) therefore proposes “stochastic Hicksian aggregation” (SHA). This is a technique by which models with aggregates composed of imperfectly correlated prices can be estimated. The resulting models explicitly account for the errors induced through aggregation, and contain constants attributable to the aggregation process. Furthermore, the SHA need not be separable from other groups of goods, and in general would not be.

Under the assumptions of Lewbel’s (1988) polynomial SHA (SHA in models which are functions of, at most, polynomials in price variables), although constants appear in the models, they have no effect on the validity of restrictions such as homogeneity and symmetry. If the disaggregated model parameters obey homogeneity and symmetry, the grouped model parameters obey an analogous form of homogeneity and symmetry. (See Lewbel, 1988, Theorem 3.)

It would therefore be useful if goods could be grouped so the requirements for SHA could be met. There is no guarantee that any choice of grouping would ensure this result. Situations under which homogeneity and symmetry restrictions are supported by one proposed grouping, but are rejected by another, support this contention.

An attractive feature of SHA is that the goods to be included in broad aggregates can be determined before estimation of any demand system. All possible groupings will not generally reflect the conditions required for SHA. It would thus be helpful to have procedures for finding stochastic Hicksian aggregates. This is one of the main purposes of this paper. Cluster analysis is used to find the grouping of goods which “best” reflects SHA, in a statistical sense. The cluster analysis groupings and conventional groupings (durables, nondurables and services) are then analysed to determine which better meet the conditions for SHA, using test procedures suggested by Lewbel (1988). Finally, demand systems are estimated using aggregates implied by: conventional groupings, clustering methods, and disaggregated expenditure categories. Homogeneity and symmetry restrictions are tested and compared in the context of each aggregation regime.

Results show that the hypothesis tests are extremely sensitive to the goods which are aggregated. Rejections seem to occur significantly less frequently when aggregates are constructed based on clustering. An interesting avenue for future research would be to conduct a simulation study to examine the generality of these findings.

The remainder of the paper is structured as follows. Section 2 highlights the effects of inappropriate aggregation. Grouping goods by cluster analysis is discussed in Section 3. The distance measure used is also related and compared to Lewbel’s (1988) tests for SHA. In Section 4, the results of the cluster analysis and the implied aggregates are presented. These aggregates are analysed using Lewbel’s (1988) test procedure, discussed previously. Section 5 contains the outcome of estimating the demand systems with the various aggregates. Section 6 contains a summary and concluding comments.
2. EMPIRICAL EVIDENCE OF AGGREGATION EFFECTS

Suppose households’ budget-shares, at a disaggregate level of expenditure categories, can be represented by Deaton and Muellbauer’s (1980) “Almost Ideal Demand System” (AIDS),

\[ w_m = \alpha_m + \sum_{j}^{M} \beta_{mj} \ln p_j + \gamma_m \ln (y/\Pi), \quad \forall \ m = 1, \ldots, M \]

where \( w_m = \frac{p_m \cdot q_m}{y} \) is the \( m \)'th budget-share; \( q_m \) is the quantity demanded of the \( m \)'th good; \( p_m \) is the \( m \)'th price; and \( y \) is total expenditure on the \( M \) goods in the complete system. For convenience, it will be assumed that the price index, \( \Pi \), can be approximated by \( \sum_m^{M} \bar{w}_m \cdot p_m \), where \( \bar{w}_m \) are mean budget-shares.

In practice, \( M \) will be large, requiring estimation of many parameters. The number of equations in (1) can therefore be reduced by aggregating the \( M \) categories into a smaller number of \( R \) groupings. If these groupings satisfy the conditions for SHA, the outcome of tests such as homogeneity and symmetry will be unaffected by the aggregation procedure (Lewbel 1988). Lewbel (1988) suggests aggregates of the form

\[ \ln p_j = \ln P_r + k_j + \mu_j, \quad j \in r, \quad r = 1, \ldots, R \]

where \( \ln P_r = \sum_{j \in r} \omega_j \ln p_j \) is the price index for the \( r \)'th aggregate, with \( \sum_{j \in r} \omega_j = 1 \). Under the assumptions that \( k_j = E(\ln p_j - \ln P_r) \) and \( \mu_j \sim N(0, \Omega) \), the aggregated budget-share system is

\[ w_r = A_r + \sum_{j}^{R} B_{rj} \ln p_j + G_r \ln (y/\Pi) + \epsilon_r, \quad \forall \ r = 1, \ldots, R \]

where \( w_r = \sum_{j \in r} w_j, A_r = \sum_{j \in r}[\alpha_j + \sum_{k}^{M} \beta_{jk} k_j], B_{rj} = \sum_{k \in r} \beta_{lk}, G_r = \sum_{j \in r} \gamma_j, \) and \( \epsilon_r = \sum_{j \in r} \sum_{k}^{M} \beta_{lk} \mu_l \). The homogeneity and symmetry restrictions of the disaggregated AIDS model, (1), are then inherited by the parameters of (3). That is, if \( \sum_{r}^{M} \beta_{mj} = 0 \) for homogeneity in (1), the assumptions ensuring SHA guarantee that \( \sum_{r}^{R} B_{rj} = 0 \). Further, if the symmetry restrictions, \( \beta_{mj} = \beta_{jm}, m \neq j \) hold for (1), then \( B_{rj} = B_{jr}, r \neq j \). This is a special case of the result stated in Lewbel’s (1988) Theorem 3.

To ensure that the foregoing properties of the parameters of (3) hold, the assumption that the \( k_j \) are constant is crucial. If the \( k_j \) are variable across observations, rather than being absorbed in the constants, \( A_r \), terms like \( \sum_{j \in r} \sum_{k}^{M} \beta_{jk} k_j \) will appear as part of the disturbances, \( \epsilon_r \), in (3). It is highly likely that variable \( k_j, (j \in r) \) will be correlated with \( \ln P_r \). Standard estimators of the parameters of (3) will then be inconsistent, and the outcomes of hypothesis tests on these parameters will be misleading.

For example, Nicol (1985) found that rejections of homogeneity in a model of highly aggregated expenditures was probably due to nonconstancy of \( k_j \). These rejections occurred when nondurable goods were constructed by aggregating: (a)
semi-durable goods with nondurables; but not when forming a durable goods category with: (b) durables and semi-durables. (Further details are available from the author on request.)

3. GROUPING EXPENDITURES

Several methods exist by which expenditures can be combined to form broad groups. One of the classic studies in this area is Theil (1954), who also suggested the use of "best linear index numbers" (Theil 1960). Kloek and de Wit (1961) mathematically related these index numbers to the principal components technique, and suggested an extension to reduce the bias in Theil's (1960) indices.

More recently, Fisher (1969) discussed the application of cluster analysis to the aggregation problem in economics. The notion that the separable structure of preferences could be used as a similarity or distance measure was raised there also. Pudney (1981) is an example of the empirical implementation of this approach.

The main aim of this paper is to suggest a suitable way of grouping goods to meet the requirements of SHA. Cluster analysis can provide such a grouping. A common criticism of the use of cluster analysis to group categories of expenditures, however, is that some of the resulting aggregates are not economically meaningful. For example, a SHA procedure might suggest grouping some durable goods with some nondurable goods, some durable goods with some services and some nondurable goods with some services. The interpretation of the derivatives of the SHA demand functions with respect to own-price and cross-price aggregates would often be unclear. This is not a "problem" due to the cluster analysis procedure itself, however, but is usually a difficulty associated with other complications which have suggested the use of SHA via cluster analysis. That is, it might not always be desirable or possible to estimate a disaggregated demand system, owing to considerations such as parsimony, or the inability to empirically identify all the parameters of a disaggregated demand system using highly correlated data. Estimating sub-systems of demand equations comprising the sub-group expenditure categories included in the SHA would not be an appropriate approach either. This would result in specification error, due to the omission of prices of nonseparable goods outside the sub-group demand system being estimated. (The existence of SHA does not imply that such aggregates are even weakly separable from one another.)

The creation of SHA via cluster analysis can thus be viewed as a useful expedient by which some parameters of a demand model can be estimated fairly precisely and, in the context of this study, as a method permitting the testing of economic hypotheses in a way which is not influenced by the aggregation of expenditures. Also, if desired, it would be possible to estimate derivatives of sub-group demands with respect to the SHA price indices. This would provide information on the responsiveness of demands for sub-group goods to changes in their prices. The linear association between the prices of goods in the SHA could then provide additional information to help derive estimates of own-price responses, net of the effects of the response to changes in other goods prices in the SHA. In any event, the gross response estimates would provide a bound on the actual net response.
SHA requires that relative prices should be constant, up to a random element. An appropriate choice of a distance measure for cluster analysis, therefore, would be one which reflects the constancy of relative prices. For prices, \( p_j \) and \( p_k \), \( i \neq k = 1, \ldots, M \), denote the relative price ratio at observation \( t \) as \( \phi_{jk} = \frac{p_{jt}}{p_{kt}} \). If \( \phi_{jk} \) is constant, it can be denoted simply as \( \phi_{jk} \). Constant \( \phi_{jk} \) are unlikely to be observed since these are realisations of a random variable. The variability of \( \phi_{jk} \) about their sample mean will thus depend on how relative prices vary in the population.

One way to test for statistically significant deviations from constancy of \( \phi_{jk} \) is to split a sample in two parts and compute alternative estimates of the sample mean. Denote \( \phi_{jk,l} \) and \( \sigma^2_{jk,l} \) as the population means and variances respectively of the sub-samples, \( l = 1, 2 \). A constant relative price, \( \phi_{jk} \), can then be represented by the null hypothesis, \( H_0: \phi_{jk,1} = \phi_{jk,2} \). When this hypothesis holds (for all \( j \neq k \)), then

\[
\begin{align*}
t_{jk} &= \frac{(\bar{\phi}_{jk,1} - \bar{\phi}_{jk,2})}{(\frac{S_{\phi,1}^2}{n_1} + \frac{S_{\phi,2}^2}{n_2})^{0.5}} \sim t(n_1 + n_2 - 2),
\end{align*}
\]

where \( \bar{\phi}_{jk,l} = \frac{1}{n_l} \sum_{t} \phi_{jk,l} \), \( l = 1, 2 \); \( S_{\phi,1}^2 = \frac{(n_1 - 1) \cdot S_{\phi,1,1}^2 + (n_2 - 1) \cdot S_{\phi,1,2}^2}{n_1 + n_2 - 2} \); \( S_{\phi,2}^2 = \frac{1}{(n_1 - 1)} \sum_{t} (\phi_{jk,l} - \bar{\phi}_{jk,l})^2 \); and \( n_l, l = 1, 2 \) are the two sub-sample sizes. The \( t \) statistic in (4) is the standard test statistic for significant differences in population means (here, the difference is zero, under \( H_0 \)) when the population variances, \( \sigma^2_{jk,l}, l = 1, 2 \), are identical.

For any sample, the \( t_{jk} \) can be computed. Large positive and negative deviations of \( t_{jk} \) are viewed symmetrically as reflecting statistically significant deviations from constancy of \( \phi_{jk} \). A cluster analysis technique which uses the \( t_{jk} \) as distance measures will thus group goods with relatively low \( t_{jk} \).

The interpretation of the \( t_{jk} \) is similar in spirit to the procedure used by Lewbel (1988) to test for statistically significant deviations of \( \ln p_j(j \in r) \) from \( \ln P_r \). In that approach, the sample is split and alternative estimators of the \( k_j \) are computed, using the two sub-samples, so

\[
\bar{k}_{j,l} = \frac{1}{n_l} \sum_{t} (\ln p_{j,t} - \ln P_{r,l}), \quad j \in r \text{ and } l = 1, 2.
\]

Lewbel (1988) then tests for statistically significant deviations in these \( \bar{k}_{j,l} \) across sub-samples using Chow tests. This procedure, however, is not a direct test of the constancy of relative prices, only of statistically significant deviations of intra-group prices from aggregate group prices. Since one of the aims of this paper is to suggest procedures for the construction of appropriate aggregates, focusing directly on within-sample relative price movements requires direct analysis of these relative prices. Consequently, the use of the \( t_{jk} \) statistics as a distance measure in a cluster analysis procedure is proposed. In the next section, the cluster analysis techniques used, and the clusters created will be discussed.

4. CLUSTER ANALYSES OF DISAGGREGATED GOODS

In this paper, goods are grouped by cluster analysis where the “cases” are prices faced by households. The intent is to create groups of goods for which intra-group
prices are similar, given the sample data. “Complete linkage,” hierarchical agglomerative cluster analysis is used, since it is the most appropriate for this paper. That is, grouped goods are no more than a given distance apart, based on the $t_{jk}$ defined previously. Starting from a total of $M$ goods, clustering continues until $R$ groups are created.

In many cluster analysis applications, clustering ceases conditional on an ad hoc “stopping rule.” This rule is to stop grouping when a large change in the distance between cases to be merged would occur. Here, clustering ceases at a stage when a statistically significant change in the $t_{jk}$ would occur. Consequently, clustering should cease immediately before a step where the “fusion coefficient” (the $t_{jk}$ between goods grouped at a certain step) rises by a statistically significant amount. Mojena (1977) proposes the stopping rule that clustering ceases at step $i$ when $T_{i+1} > \bar{t} + \kappa \cdot S_{\bar{t}}$, where $T_{i+1}$ is the $(i + 1)$th fusion coefficient; $\bar{t}$ and $S_{\bar{t}}$ are the mean and standard deviation respectively of the distribution of $T_i$; and $\kappa$ is a constant. In practice, Mojena (1977) states that $\kappa = 2.75$–$3.50$ gives good results (in identifying the true number of clusters) in simulation experiments.

There is a total of $\{M(M - 1)/2\} + 1$, which can be arranged in a symmetric, $M \times M$ matrix, denoted $T_M$. The diagonal elements of $T_M$, denoted $T_{M_i}$, are zero for all $j$, and the off-diagonal elements are $|t_{jk}|$. An attractive feature of $T_M$ is that its off-diagonal elements are distributed as $t$-statistics, each having $n_1 + n_2 - 2$ degrees of freedom. Thus, using the previously defined notation for fusion coefficients, $F_i = t_{jk} \cdot A \cdot N(0, 1)$. Suppose clustering continues until step $i$, and $t_{jk,i+1} \geq 3.29$. Then the probability of observing fusion coefficients greater than 3.29 is approximately 0.001. This means that members of any cluster will be within about three standard deviations of one another when clustering ceases.

The data used for clustering were the prices of thirteen goods and services faced by the four household groups described in Nicol (1985). These households were drawn from the 1978, 1982, 1984 and 1986 Canadian Family Expenditure Surveys. The total numbers of households of types (1) through (4) were 3257, 3226, 4313 and 1166 respectively. The expenditure categories were food (FO), household operation (HO), household goods (HG), vehicle purchase (AU), gasoline (GA), other auto operation (OA), public transit (PT), personal care services (PE), personal care supplies (PU), recreation (RE), tobacco products (TO), alcoholic beverages (AL) and clothing (CL). (More detailed information on these categories is available from the author on request.)

The price data were constructed from indices of inter-city retail price differences (published in Consumer Prices and Price Indexes, Statistics Canada, Ottawa, Canada). Details of the construction method and the actual price indices are available on request.

Two approaches to grouping goods could be taken. Households could be viewed as having identical SHA, and clustering applied to all households simultaneously. Alternatively, each household group could be regarded as possessing distinct SHA, and clustering conducted separately for each group. The latter approach could generate different aggregates across household groups. These differences would be attributable to intertemporal and geographic changes in the distribution of household groups. Such differences could be very important empirically, in terms of
generating deviations from SHA. That is, economic conditions are highly likely to affect the distribution of household groups by time and place. This distribution in turn influences demand and supply conditions for goods and services, hence affecting relative prices. Whether these effects lead to statistically significant deviations from SHA is, however, an empirical question. Consequently, clustering was applied to all household groups simultaneously, and to each household group separately. This allows for the possibility that statistically significant deviations in SHA across household groups can be detected.

Clusters were constructed using the $T_M$ matrices as distance measures. As mentioned earlier, clustering ceases at step $i$ when $t_{jk_{r1}} \geq 3.29$. This results in eight aggregates for household groups (1) and (4), and nine aggregates for household groups (2) and (3). However, in the case of clustering with the sample of all households together, the lowest valued element in $T_M$ was 3.72, and the next lowest, 5.66. This indicates no aggregates should be formed, according to the criterion proposed. Evidently, the variability in relative prices across household groups (1) through (4) combined is too pronounced to permit any SHA. Since the elements of $T_M$ are distributed as $t$ statistics, this shows there are statistically significant deviations in SHA across household groups. This is reinforced by the expenditures included in the SHA formed when each household group is analysed separately, the aggregates for which are presented in Table 1. (The order in which goods are listed in the table indicates the stage at which clustering took place.) That is, there are significant differences across household groups as to where clustering occurs for specific goods, and in the goods belonging to clusters for any given household group.

The result that goods which comprise SHA differ across household groups is very interesting. This shows that appropriate expenditure aggregates should differ across samples. It also suggests that employing aggregates such as durables, nondurables and services probably tends to ignore the correct structure of SHA within samples. Evidence for this assertion is provided by the fact that expenditures included in the clusters in Table 1 tend to be a mixture of what would usually be considered to be durables, nondurables and services.

Another purpose of this paper is to determine whether standard aggregates or cluster analysis gives aggregates most closely conforming to SHA. Since the $T_M$
matrices focus directly on relative price deviations within samples, one would expect these to yield superior results. This hypothesis can be investigated by two methods. First, the Chow tests suggested by Lewbel (1988) can be conducted for the different aggregates. Alternatively, a demand model can be estimated and hypotheses tested using the two types of aggregate. The latter topic is the subject of Section 5. The former is considered here.

The Chow tests described in Lewbel (1988) were conducted based on the alternative estimators of $k_{j,l}$, $l = 1, 2$, defined in (5). Under the null hypothesis that $k_j$ is constant, $j \in r$, an $F$-statistic can be computed. Two sets of $F$-statistics of the foregoing type were computed. These related to “standard” aggregates (durables, nondurables and services, (a) and (b), referred to in Section 2); and $T_M$ matrix-based clusters for each household group. Rejections of the hypothesis of constant $k_j$ were recorded for: 99 out of 104 tests; and 8 out of 29 tests respectively. In all cases of rejection, the upper-tail probability values were approximately zero. These results indicate that the $T_M$ matrices for separate household groups perform best in finding aggregates with relatively stable $k_j$.

5. Testing Demand Systems with Aggregated and Disaggregated Data

The AIDS specifications in (1) and (3) was estimated (as appropriate) conditional on three classifications of goods and services. The classifications were: standard aggregates, (a) and (b), described in Section 2; aggregates generated by cluster analysis using $T_M$ matrices; and the thirteen disaggregated expenditure categories themselves. The objective was to determine how hypothesis tests such as homogeneity and symmetry are affected in the light of different aggregation regimes. The results of Section 4 suggest that fewer rejections would occur when using aggregates based on cluster analysis.

The disaggregated expenditure data were extracted from the 1978, 1982, 1984 and 1986 Canadian Family Expenditure Surveys, for household groups (1) through (4). Budget-shares ($w_m$ and $w_r$) and total expenditures ($y$) were then constructed for each of the expenditure classifications. Dummy variables were also constructed for households, indicating regions of residence. (The regions were: Atlantic Provinces; Québec; Ontario, Prairie Provinces; and British Columbia.) The omission of regional variables would be highly likely to induce specification error. Since regional effects are related to geographic price indices, such a specification error could influence the type of effects this study is trying to isolate.

Four models of demand were estimated, for each type of aggregate and household group. These were: an unconstrained AIDS model, including intercept-shift dummy variables; a constrained AIDS model, excluding regional effects; a homogeneity-constrained AIDS model, including regional effects; and a homogeneity plus symmetry constrained AIDS model, including regional effects. Each of the constrained models was tested against the unconstrained AIDS model, including regional effects. The nature of the homogeneity and symmetry restrictions were outlined in Section 2.

The first systems estimated were those associated with standard aggregates, (a) and (b). In a singular system with $R = 3$ equations, the unrestricted model (including four intercept-shift dummy variables per estimable equation) has 18 parameters. There
are eight regional dummy variable restrictions; two homogeneity restrictions; and three homogeneity plus symmetry restrictions. Assuming the system disturbances, \( \varepsilon = (\varepsilon_1, \ldots, \varepsilon_R)^T \sim N(0, \Sigma) \), where \( \text{rank}(\Sigma) = 2 \), \( 2[\ln L(\hat{\theta}) - \ln L(\overline{\theta})] \overset{\Delta}{=} \chi^2(q) \) under the null hypothesis. Here, \( \ln L(\hat{\theta}) \) and \( \ln L(\overline{\theta}) \) are maximised values of the unrestricted and restricted log-likelihood function of the system respectively, and \( q \) is the number of restrictions. The estimation method (for these aggregates as well as the results reported later) was by seemingly unrelated regression, subject to restrictions where appropriate, using SHAZAM, version 6.04 (White 1978).

The results of hypothesis tests for the standard aggregates are presented in Table 2. It was found that the regional dummy variables were statistically significant explanatory variables, whether using standard aggregates (a) or (b). As in Nicol (1985), however, it was found that homogeneity and symmetry were rejected when employing (a) (even with regional dummy variables included) but not always when using (b). This is what the results in Section 4 predicted: standard aggregates (for these data) do not typically exhibit SHA.

The models were then estimated using aggregates generated via cluster analysis with \( T_M \) matrices as distance measures. Test results are presented in Table 3. Regional effects were again found to be statistically significant, but homogeneity and symmetry were not rejected. This indicates that clustering with \( T_M \) matrices appears to be capable of identifying aggregates which do not exhibit sufficient deviations from SHA to affect hypothesis tests such as homogeneity and symmetry.

The last set of hypothesis tests was constructed on systems estimated with disaggregated data. These results are presented in Table 4. In this case, upper-tail probability values for the regional dummy variable hypothesis tests were much larger than previously. This indicates that regional dummy variables could perhaps be excluded for such disaggregated goods. Tests of homogeneity and symmetry conditional on the inclusion of regional effects indicated these restrictions were not rejected by the data, in general.

To summarise, it appears that even the specification error of omitting regional effects has less impact, the less the data are aggregated. However, the greater the degree of aggregation, and the more this aggregation tends to ignore the structure of SHA, the greater the impact on hypothesis tests. Similar results were obtained when using cluster analysis with correlation coefficients between the price variables as a distance measure. The only notable difference was that slightly different groupings were generated, but the resulting SHA worked equally well, in terms of having little impact on the outcome of hypothesis tests. (Further details of these last results are available from the author on request.)

6. SUMMARY AND CONCLUSIONS

In this paper, the possibility that aggregation of goods which fail to exhibit SHA could affect the outcome of hypothesis was explored. Cluster analysis was used to determine the goods most likely to belong to groups of SHA. Standard aggregates such as durables, nondurables and services were also constructed. The cluster analysis made use of statistically-based stopping rules, to determine the appropriate number of SHA, or clusters, which is not a common approach.
<table>
<thead>
<tr>
<th>Model</th>
<th>No. of Restrictions</th>
<th>Log. of Likelihood</th>
<th>Test Statistic</th>
<th>Log. of Likelihood</th>
<th>Test Statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unrestricted</td>
<td>—</td>
<td>4816.443</td>
<td></td>
<td>4747.735</td>
<td></td>
</tr>
<tr>
<td>Ex. Regional Effects</td>
<td>8</td>
<td>4793.950</td>
<td>44.986</td>
<td>4732.057</td>
<td>31.356</td>
</tr>
<tr>
<td>HG-Constrained</td>
<td>2</td>
<td>4808.155</td>
<td>16.576</td>
<td>4747.380</td>
<td>0.710</td>
</tr>
<tr>
<td>HG &amp; SY-Constrained</td>
<td>3</td>
<td>4804.612</td>
<td>23.662</td>
<td>4747.364</td>
<td>0.742</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model</th>
<th>No. of Restrictions</th>
<th>Log. of Likelihood</th>
<th>Test Statistic</th>
<th>Log. of Likelihood</th>
<th>Test Statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unrestricted</td>
<td>—</td>
<td>6237.950</td>
<td></td>
<td>6569.507</td>
<td></td>
</tr>
<tr>
<td>Ex. Regional Effects</td>
<td>8</td>
<td>6221.267</td>
<td>33.366</td>
<td>6554.507</td>
<td>30.806</td>
</tr>
<tr>
<td>HG-Constrained</td>
<td>2</td>
<td>6216.148</td>
<td>43.604</td>
<td>6565.654</td>
<td>7.706</td>
</tr>
<tr>
<td>HG &amp; SY-Constrained</td>
<td>3</td>
<td>6210.812</td>
<td>54.276</td>
<td>6565.304</td>
<td>8.406</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model</th>
<th>No. of Restrictions</th>
<th>Log. of Likelihood</th>
<th>Test Statistic</th>
<th>Log. of Likelihood</th>
<th>Test Statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unrestricted</td>
<td>—</td>
<td>8767.474</td>
<td></td>
<td>9396.961</td>
<td></td>
</tr>
<tr>
<td>Ex. Regional Effects</td>
<td>8</td>
<td>8739.425</td>
<td>56.098</td>
<td>9359.528</td>
<td>74.866</td>
</tr>
<tr>
<td>HG-Constrained</td>
<td>2</td>
<td>8689.901</td>
<td>155.146</td>
<td>9385.755</td>
<td>22.412</td>
</tr>
<tr>
<td>HG &amp; SY-Constrained</td>
<td>3</td>
<td>8680.541</td>
<td>173.866</td>
<td>9384.527</td>
<td>24.868</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model</th>
<th>No. of Restrictions</th>
<th>Log. of Likelihood</th>
<th>Test Statistic</th>
<th>Log. of Likelihood</th>
<th>Test Statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unrestricted</td>
<td>—</td>
<td>2429.961</td>
<td></td>
<td>2551.202</td>
<td></td>
</tr>
<tr>
<td>Ex. Regional Effects</td>
<td>8</td>
<td>2421.186</td>
<td>17.550</td>
<td>2539.113</td>
<td>24.178</td>
</tr>
<tr>
<td>HG-Constrained</td>
<td>2</td>
<td>2398.894</td>
<td>62.134</td>
<td>2549.220</td>
<td>3.964</td>
</tr>
<tr>
<td>HG &amp; SY-Constrained</td>
<td>3</td>
<td>2380.296</td>
<td>99.330</td>
<td>2547.042</td>
<td>8.320</td>
</tr>
</tbody>
</table>

Note: The test statistics are distributed as $\chi^2$, their degrees of freedom being the number of restrictions in each case. The critical values of these statistics, at a significance level of 0.001, are 26.1, 13.8 and 16.3 for 8, 2 and 3 degrees of freedom respectively.

It was found that aggregates more closely exhibiting SHA properties generated fewer rejections when conducting standard hypotheses tests on demand systems. Also, there appeared to be statistically significant evidence that the goods compris-
TABLE 3
TESTS BASED ON AGGREGATES CREATED USING $T_M$ MATRICES

<table>
<thead>
<tr>
<th>Model</th>
<th>No. of Restrictions</th>
<th>Household Group 1</th>
<th>Household Group 2</th>
<th>Household Group 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Log. of Likelihood</td>
<td>Test Statistic</td>
<td>Log. of Likelihood</td>
</tr>
<tr>
<td>Unrestricted</td>
<td>—</td>
<td>27061.73</td>
<td>—</td>
<td>42517.37</td>
</tr>
<tr>
<td>Ex. Regional Effects</td>
<td>28</td>
<td>27029.70</td>
<td>64.06</td>
<td>42482.25</td>
</tr>
<tr>
<td>HG-Constrained</td>
<td>7</td>
<td>27059.97</td>
<td>3.52</td>
<td>42513.01</td>
</tr>
<tr>
<td>HG &amp; SY-</td>
<td>28</td>
<td>27045.05</td>
<td>33.36</td>
<td>42495.96</td>
</tr>
<tr>
<td>Constrained</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model</th>
<th>No. of Restrictions</th>
<th>Household Group 4</th>
<th>Household Group 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Log. of Likelihood</td>
<td>Test Statistic</td>
</tr>
<tr>
<td>Unrestricted</td>
<td>—</td>
<td>13751.19</td>
<td>—</td>
</tr>
<tr>
<td>Ex. Regional Effects</td>
<td></td>
<td>13733.06</td>
<td>36.26</td>
</tr>
<tr>
<td>HG-Constrained</td>
<td>7</td>
<td>13745.33</td>
<td>11.72</td>
</tr>
<tr>
<td>HG &amp; SY-</td>
<td>28</td>
<td>13736.85</td>
<td>28.68</td>
</tr>
<tr>
<td>Constrained</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: The test statistics are distributed as $\chi^2$, their degrees of freedom being the number of restrictions in each case. The critical values of these statistics, at a significance level of 0.001, are 56.9, 24.3 and 62.5, 26.1 and 69.0 for 28, 7, 32, 8 and 36 degrees of freedom respectively.

The results therefore show that it is both possible and desirable to analyse the SHA properties of data, prior to estimation of demand systems. The techniques proposed in Sections 3 and 4 provide a means by which this can be achieved. It is also apparent that cluster analysis performs rather well in determining SHA which appear to have a minimal impact on the outcome of important hypothesis tests. It would be of interest to conduct a simulation study, to ascertain if these results are generally valid. Also, further demand analysis of effects of inappropriate aggregation on elasticities and concavity would be of interest.

The foregoing results are also suggestive of some unrelated areas of fruitful research possibilities. In particular, many studies of aggregate consumption analysis, wage determination, labour supply and numerous macroeconomic models make use of aggregate price indices. These studies typically employ a single index to capture price effects. Given the results observed here, it is of interest to re-evaluate studies which employ aggregated goods and price indices. The implicit assumption being made there is that relative prices across goods are stable over time and place. This is definitely not the case for observed data, as seen in this
Table 4
TESTS BASED ON DISAGGREGATED EXPENDITURE CATEGORIES

<table>
<thead>
<tr>
<th>Model</th>
<th>Restrictions</th>
<th>Household Group 1</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Log. of Likelihood</td>
<td>Test Statistic</td>
<td></td>
</tr>
<tr>
<td>Unrestricted</td>
<td>—</td>
<td>58057.26</td>
<td>—</td>
<td>67644.49</td>
</tr>
<tr>
<td>Ex. Regional Effects</td>
<td>48</td>
<td>58020.42</td>
<td>73.68</td>
<td>67610.71</td>
</tr>
<tr>
<td>HG-Constrained</td>
<td>12</td>
<td>58051.05</td>
<td>12.42</td>
<td>67640.01</td>
</tr>
<tr>
<td>HG &amp; SY- Constrained</td>
<td>78</td>
<td>58011.63</td>
<td>91.26</td>
<td>67599.47</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model</th>
<th>Restrictions</th>
<th>Household Group 2</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Log. of Likelihood</td>
<td>Test Statistic</td>
<td></td>
</tr>
<tr>
<td>Unrestricted</td>
<td>—</td>
<td>67644.49</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>Ex. Regional Effects</td>
<td>48</td>
<td>67610.71</td>
<td>67.56</td>
<td></td>
</tr>
<tr>
<td>HG-Constrained</td>
<td>12</td>
<td>67640.01</td>
<td>8.96</td>
<td></td>
</tr>
<tr>
<td>HG &amp; SY- Constrained</td>
<td>78</td>
<td>67599.47</td>
<td>90.04</td>
<td></td>
</tr>
</tbody>
</table>

Note: The test statistics are distributed as $\chi^2$, their degrees of freedom being the number of restrictions in each case. The critical values of these statistics, at a significance level of 0.001, are 84.0, 32.9 and 122.4 for 48, 12, and 78 degrees of freedom respectively.

study. The effect of relaxing the fixed relative price assumption on the results of estimating the models mentioned above would thus seem to be of significant value.

University of Regina, Canada

REFERENCES

CONSUMER PRICES AND PRICE INDEXES, Catalogue No. 62-010 (Ottawa: Statistics Canada).
FISHER, W. D., Clustering and Aggregation in Economics (Baltimore: Johns Hopkins Press, 1969).
You have printed the following article:

The Effect of Expenditure Aggregation on Hypothesis Tests in Consumer Demand Systems
Christopher J. Nicol
Stable URL:
http://links.jstor.org/sici? sici=0020-6598%28199105%2932%3A2%3C405%3ATEAO%3E2.0.CO%3B2-6

This article references the following linked citations. If you are trying to access articles from an off-campus location, you may be required to first logon via your library web site to access JSTOR. Please visit your library's website or contact a librarian to learn about options for remote access to JSTOR.

**References**

**An Almost Ideal Demand System**
Angus Deaton; John Muellbauer
Stable URL:
http://links.jstor.org/sici? sici=0002-8282%28198006%2970%3A3%3C312%3AAAIDS%3E2.0.CO%3B2-Q

**An Empirical Method of Approximating the Separable Structure of Consumer Preferences**
S. E. Pudney
Stable URL:
http://links.jstor.org/sici? sici=0034-6527%28198110%2948%3A4%3C561%3AAEMOAT%3E2.0.CO%3B2-H

**A General Computer for Econometric Models--Shazam**
Kenneth J. White
Stable URL:
http://links.jstor.org/sici? sici=0012-9682%28197801%2946%3A1%3C239%3AAGCFEM%3E2.0.CO%3B2-0