The Rank and Model Specification of Demand Systems: An Empirical Analysis Using United States Microdata

by

Christopher J. Nicol
Department of Economics
University of Regina
Regina, Saskatchewan, S4S 0A2
CANADA

Internet: Chris.Nicol@URegina.CA;
http://www.econ.uregina.ca

December 31, 1999

The support of the Social Sciences and Humanities Research Council of Canada is gratefully acknowledged. I thank James Farley for outstanding research assistance, participants at the presentation of an earlier version of this paper at the Annual Meetings of the Canadian Economics Association, Ottawa, Ontario, June, 1998 for their helpful comments and the referees for their patience in reviewing a much longer earlier version of the paper. Any errors are my responsibility.

Abstract

A rank three demand system incorporating labour force participation, non-separability of demands from excluded goods and non-exact aggregation in income and household characteristics is estimated with United States Consumer Expenditure Survey microdata. A unique price data set is used in conjunction with these microdata, which permits the analysis of systematic errors in price variables, and their effects on hypothesis tests and estimated elasticities.

It is found that errors in price variables bias test results for the rank three hypothesis in particular towards rejection. Other test results are affected, but to a lesser degree. It is also found that estimating smaller systems of demand equations, even when conditioning on excluded goods yields significantly different test results and estimated elasticities.

Another important conclusion is that model specification is statistically significantly different for households of varying family sizes and housing tenure statuses. This indicates that caution should be exercised in interpreting demand system estimation results based on samples with households with heterogeneous characteristics.

Keywords:

errors in variables; labour supply; microdata; rank three demands; separability.

1 Introduction

Recent research in applied demand analysis has focused on a variety of model specification issues. For example: whether demand systems are rank two or three (Fry and Pashardes, 1992 and Banks, Blundell and Lewbel, 1997 and Lyssiotou, Pashardes and Stengos, 1999); the separability of commodity demands from labour supply (Browning and Meghir, 1991 and Kaiser, 1993); and the role of household characteristics variables in demand models (Blundell, Pashardes and Weber, 1993, Dickens, Fry and Pashardes, 1993 and Micheline, 1999). These studies have indicated that all such features are important determinants of demand. In these applications, however, not all of these modelling considerations are controlled for simultaneously. That is, one or more of these features are *not* explicitly modelled in all of the applications mentioned. Since the various modelling aspects have been found to be important determinants of demand, applications which focus on one or more but ignore others gives cause for concern. This was highlighted in Dagenais (1994), where the effects of multiple model specification errors were investigated. It was found that, when more than one model specification error was present, correcting for only one of them induced *larger* biases and inconsistencies than if one did not control for any of them.

It is also well-known that errors in variables can give rise to serious distortions in estimation results when these errors are ignored (see Cragg, 1994, for a recent survey article on this issue). When demand systems are estimated using microdata, it can be said almost unequivocally that such research routinely abstracts from the presence of gross errors in the price variables used. For example, applied demand research using United Kingdom (UK) data identifies price parameters solely through the variation in prices over time, though it is known that regional variation in prices exists in the UK. Unfortunately, however, there do not exist reliable, official data for the UK reflecting these regional price differences. On the other hand, research employing Canadian microdata has used regional and temporal price variation to identify price parameters, yet there still exist errors in these price variables associated with using regional as opposed to city-level price indices. Again, however, accounting for city-level price differences is not possible on the basis of available officially published data. The approaches of empirical researchers who have used the UK and Canadian data in the above ways are thus "best practice" given the shortcomings of available data.

It is therefore of some interest to conduct a detailed study which focuses on all of the determinants of demand which have been found to be important (as described above), and on the indicated errors in variables problem, in a comprehensive estimation strategy. Such a study could shed light on the possible effects the interactions of these features might have. This requires a very rich data set, covering a geographic region where prices vary considerably. In addition, it would be helpful if variations in prices were well documented.

The United States has a long series of microdata sets which have seen limited use in demand analysis, partly owing to the problem that inter-regional price comparisons are not published by government agencies. There are consumer price data for a wide range of cities, states and regions, but these compare prices across time for specific cities or regions; not *across* cities at a *point* in time. In this paper, price data are constructed based on the US consumer price index (CPI) and American Chamber of Commerce Research Association (ACCRA) data. These latter price data comprise indices for a range of goods and services across US cities at points in time, and are comparable across those cities at those times, although not through time.

US Consumer Expenditure Survey (CEX) microdata are used to analyse the model specification issues discussed earlier, and also the possible influence on estimation results of errors in price variables. Given the unique nature of the price data used, it is possible to analyse the effects on hypothesis tests and estimated elasticities of induced errors in price variables, where the nature of the errors are similar to the kinds encountered by researchers who have used the UK and Canadian microdata. The importance of labour force participation effects, the rank of the demand system and the effects of separability of the goods in the demand model from other goods can also be analysed.

It is found that test results and estimated elasticities are significantly different when estimated via a separable demand sub-system, as opposed to estimating a larger system where all goods are included directly. This is the case even when the separable sub-system controls for or conditions on the omission of goods not included directly. It is also found that a rank two demand specification is adequate for several of the household groups considered. This result is not inconsistent, however, with earlier research. In those cases, the apparent rank of the estimates demand systems were shown to depend on the inclusion or exclusion of households in the tails of the income distribution (for example, Lewbel, 1991), and the endogeneity of explanatory variables (Lyssiotou, Pashardes and Stengos, 1999). The finding of rank three demands for other household types is also supported by our results, although rank tests are significantly influenced toward rejection of the rank two hypothesis when errors in price variables are present.

With respect to estimated elasticities, these are affected significantly both by errors in price variables and the size of the system estimated. In terms of other model specification matters, the results in this paper bear out earlier findings regarding the importance of labour force participation effects as determinants of demand, the effects of age of head of household and several other variables typically found to affect demand patterns.

The remainder of the paper is structured as follows. In Section 2, the model specification is discussed. Some relevant empirical literature is also discussed, which gives direction to the initial model specification. The data used are discussed in Section 3. Section 4 gives details of the estimated models, hypothesis tests conducted, and provides comparisons of the estimated elasticities under the various estimation conditions. Section 5 summarises and concludes.

2 Model Specification

Suppose "goods" over which consumers make spending decisions can be partitioned into three types. Goods of direct interest, denoted q (with their prices, p); labour force variables, ℓ (with their prices, r); and other goods of indirect interest, c (with their prices, p_c). For convenience in what follows, let $P = [p, r, p_c]^T$ and $Q = [q, \ell, c]^T$. Also, let z denote a vector of demographic or household characteristics variables.

If preferences can be represented by a utility function, $U_h[q, \ell, c]$, where h indexes households, when a group of households share the same preferences, given the same set of household characteristics, z, the utility function can be re-written $U^*[q, \ell, c; z]$. Under ideal circumstances, maximisation of $U^*[q, \ell, c; z]$ subject to an appropriate budget constraint yields the *unconditional* demand functions for the elements of q, ℓ and c. These demand functions depend on z, so are specific to households of a particular type. Or, in other words, they are *conditional* demands, relative to the vector of household characteristics, z.

Alternatively, consider the households' cost function using the same notation and definitions for Q, P and z

$$\ln C^*[p, r, p_c, u; z] = a(p, r, p_c; z) + \frac{b(p, r, p_c; z)}{[f(u) - g(p, r, p_c; z)]} \tag{1}$$

which can be re-arranged to yield the indirect utility function

$$\ln V^*[p, r, p_c, y; z] = f^{-1}\{\frac{b(p, r, p_c; z)}{[\ln y - a(p, r, p_c; z)]} + g(p, r, p_c; z)\} \tag{2}$$

This model specification is based on that in Fry and Pashardes (1992) (which is equivalent to expressions in Lewbel, 1991, and implicit in Gorman, 1981), with the inclusion here of the distinction between q, ℓ and c, and conditioning on z. Fry and Pashardes (1992) show that this indirect utility function yields the budget-share system

$$w_i = a_i'(p, r, p_c; z) + \frac{b_i'(p, r, p_c; z)}{b(p, r, p_c; z)}\{\ln(y) - a[p, r, p_c; z]\} + \frac{g_i'(p, r, p_c; z)}{g(p, r, p_c; z)}\{\ln(y) - a[p, r, p_c; z]\}^2 \quad (3)$$

where $a_i'(p, r, p_c; z), b_i'(p, r, p_c; z)$ and $g_i'(p, r, p_c; z)$ are derivatives of $a(p, r, p_c; z)$, $b(p, r, p_c; z)$ and $g(p, r, p_c; z)$ with respect to $\ln p_i$.

The model in (3) above is a quadratic generalisation of the Almost Ideal (AI) demand system of Deaton and Muellbauer (1980). The quadratic AI (QAI) system nests the AI model, and is less restrictive than the quadratic expenditure system (QES) proposed by Howe, Pollak and Wales (1978). The most general form of the QES was presented in van Daal and Mierkes (1989). Lewbel (1991) developed a general representation for the QAI system, and this has been used in empirical applications by Banks, Blundell and Lewbel (1992 and 1997) and Fry and Pashardes (1992). One of the attractions of the QAI model is that it typically outperforms the popular AI system. Furthermore, it is a rank three system, which can be seen from Lewbel's (1991) definition of the rank of a demand system as the minimum number of functions, $g_m(\ln P, \ln Y; z)$ for all $\ln P$ and $\ln Y$, and for which

$$w_i = d_i^*[\ln P, \ln Y; z] = \sum_{m=1}^{M} a_{mi}^*(\ln P; z) g_m^*(\ln P, \ln Y; z), \quad (4)$$

where $\ln P$ is the vector of the logs of prices, including the prices of the elements of ℓ and c, and Y also includes expenditures on these "goods". Note that $a_{mi}^*[\ln P; z]$ and $g_m^*[\ln P, \ln Y; z]$ in the notation of (4) are chosen to coincide with the notation for these objects in Lewbel (1991), and are *not* to be confused with $a[q, r, p_c; z]$ and $g[p, r, p_c; z]$ of (1) and (2) above.

A system such as (4) has a rank of three when $M = 3$. Rank three systems are attractive since, not only have they been found to be supported empirically (relative to rank two systems such as the AI model) as was already mentioned, but also because a rank of three is the maximum number of linearly independent columns in a matrix formed with functions such as $a_{mi}^*[\ln P; z]$, and for which the associated demand system will be exactly aggregable (Gorman, 1981, and Lewbel, 1991). Exactly aggregable demand systems are often of interest since parametric restrictions which hold for such systems at the level of households' demands imply that aggregate household data could be used to identify empirically all model parameters.

The rationale for partitioning Q into q, ℓ and c in this application is that certain empirical counterparts of the consumers' complete demand system arising from an indirect utility function such as (2) are not estimable for several reasons. In particular, elements of c could be durable goods. Data reflecting the service flows from these goods would be necessary for estimation and empirical identification of the parameters of the demand model. In general these flows differ from expenditures made on the same goods per unit time, however, and it is expenditures, not flows, which household expenditure surveys usually measure. In addition, with respect to labour force participation, household expenditure surveys have limited information on labour force participation variables, the elements of ℓ. This suggests that it would be fruitful to focus on the estimation of demand systems conditional not only on z, but also on certain other variables. These variables' effects could then be taken into account in a general way, without specifying an explicit functional relationship for this behaviour. The demand relationship for goods of direct interest, q, can be as complex as desired.

A conditional cost function, $C[p,u;\ell,c,z] = \min_q[p \cdot q | U(q;\ell,c,z) = u]$, can be defined relative to ℓ, c and z as conditioning variables. The properties of such functions are discussed in Pollak (1969) and Browning (1983). The conditional, compensated demand functions, q_i are the derivatives of $C[p,u;\ell,c,z]$ with respect to p_i and can be denoted $q_i = f_i[p,u;\ell,c,z]$, $i = 1,\ldots,n$. Also, let y denote total expenditure on the n goods, $q = [q_1,\ldots,q_n]^T$. Note that n does *not* include goods which comprise expenditures on goods in the vector, c, and nor does y. Given this modified conditional cost function, a variation of the QAI model can be specified as follows

$$\ln C[p,u;\ell,c,z] = a(p;\ell,c,z) + \frac{b(p;\ell,c,z)}{[f(u) - g(p;\ell,c,z)]} \tag{5}$$

with associated indirect utility function

$$\ln V[p,y;\ell,c,z] = f^{-1}\{\frac{b(p;\ell,c,z)}{[\ln(y) - a(p;\ell,c,z)]} + g(p;\ell,c,z)\} \tag{6}$$

yielding the budget-share system

$$w_i = a_i'(p;\ell,c,z) + \frac{b_i'(p;\ell,c,z)}{b(p;\ell,c,z)}\{\ln(y) - a[p;\ell,c,z]\} + \frac{g_i'(p;\ell,c,z)}{g(p;\ell,c,z)}\{\ln(y) - a[p;\ell,c,z]\}^2 \tag{7}$$

where $a_i'(p;\ell,c,z), b_i'(p;\ell,c,z)$ and $g_i'(p;\ell,c,z)$ are the derivatives of $a(p;\ell,c,z), b(p;\ell,c,z)$ and $g(p;\ell,c,z)$ with respect to $\ln p_i$. A rank three model in this modified conditional demand environment would then take the general form

$$\begin{aligned} w_i &= d_i(\ln p, \ln y;\ell,c,z) \\ &= \sum_{m=1}^{M} a_{mi}(\ln p;\ell,c,z) g_m(\ln p, \ln y;\ell,c,z) \end{aligned} \tag{8}$$

where $\ln p$ is the vector of logs of the elements of p. If this conditional demand system has rank $M = 3$, then it can be written

$$\begin{aligned} w_i &= a_{1i}(\ln p; \ell, c, z) g_1(\ln p, \ln y; \ell, c, z) + a_{2i}(\ln p; \ell, c, z) g_2(\ln p, \ln y; \ell, c, z) + \\ & \quad a_{3i}(\ln p; \ell, c, z) g_3(\ln p, \ln y; \ell, c, z) \end{aligned} \tag{9}$$

To operationalise empirical implementation of (7) and thus (9), functional forms must be specified for $a(p; \ell, c, z)$, $b(p; \ell, c, z)$ and $g(p; \ell, c, z)$. Following Fry and Pashardes (1992), and allowing for conditioning on the variables, ℓ, c and z, these functions are parameterised as follows

$$a(p; \ell, c, z) = \alpha_0 + \sum_{i=1}^{n} \alpha_i(\ell, c, z) \ln p_i + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \gamma_{ij} \ln p_i \ln p_j \tag{10}$$

which is the familiar functional form for the counterpart function $a[p]$ from the AI system, but here conditioned to depend on ℓ, c and z,

$$b(p; \ell, c, z) = \beta_0 \prod_{i=1}^{n} p_i^{\beta_i(\ell, c, z)} \tag{11}$$

again taking the form of $b[p]$ from the AI system, but incorporating the effects of ℓ, c and z, and

$$g(p; \ell, c, z) = b(p; \ell, c, z) \cdot \lambda(p; \ell, c, z) \tag{12}$$

where

$$\lambda(p; \ell, c, z) = \lambda_0 + \sum_{i} \lambda_i(\ell, c, z) \ln p_i \tag{13}$$

The specifications of (12) and (13) are convenient, in that these yield budget-shares without across-equation restrictions on quadratic expenditure effects, and the inclusion of λ_0 removes effects of price normalisation on the other parameters of the model (Fry and Pashardes, 1992, p. 5). The following linearisations of the functions $\alpha_i(\ell, c, z), \beta_i(\ell, c, z)$ and $\lambda_i(\ell, c, z)$ complete the specification

$$\alpha_i(\ell, c, z) = \sum_{j=1}^{n} \sum_{k=1}^{K} [\alpha_{jk} v_k] \tag{14}$$

$$\beta_i(\ell, c, z) = \sum_{j=1}^{n} \sum_{k=1}^{K} [\beta_{j0} + \beta_{jk} v_k] \tag{15}$$

$$\lambda_i(\ell, g, z) = \lambda_i \tag{16}$$

where the vector $\nu = [\nu_1, \ldots, \nu_K]^T$ is used to represent ℓ, c and z, for notational convenience. The influences of ℓ, c and z are therefore confined to $\alpha_i(\ell, c, z)$ and $\beta_i(\ell, c, z)$. Given the above parameterisations for (10)–(16), the following budget-share system can be obtained:

$$\begin{aligned} w_i &= \sum_{k=1}^{K} \alpha_{ik}\nu_k + \sum_{j=}^{n} \gamma_{ij} \ln p_j + [\beta_{i0} + \sum_{k=1}^{K} \beta_{ik}\nu_k][\ln(y) - a(p,\nu)] + \\ & \{\lambda_i + [\beta_{i0} + \sum_{k=1}^{K} \beta_{ik}\nu_k][\lambda_0 + \sum_{j=1}^{n} \lambda_j \ln p_j]\}[\ln(y) - a(p,\nu)]^2 + \mu_i \end{aligned} \qquad (17)$$

The random term, μ_i, denotes a stochastic disturbance such that $[\mu_1, \ldots, \mu_n]^T \sim \mathrm{N}(0, \Omega)$. The covariance matrix of μ is singular, so only $n-1$ equations of the system need to be estimated, the parameters of the n'th being recoverable by the adding-up and integrability conditions indicated below. Empirical considerations relating to this stochastic specification will be discussed in Section 4.

Integrability conditions which the parameters of (17) must satisfy in order to be consistent with the indirect utility function, (6), are: $\sum_i^n \alpha_{i1} = 1, \sum_i^n \alpha_{ik} = 0, k = 2, \ldots K, \sum_i^n \gamma_{ij} = 0, \sum_i^n \beta_{ik} = 0$ all k, and $\sum_i^n \lambda_i = 0$, to satisfy adding up; $\sum_j^n \gamma_{ij} = 0$ to satisfy homogeneity; and $\gamma_{ij} = \gamma_{ji}, \forall i \neq j$, to satisfy symmetry of substitution effects. A negativity condition is also required of the Slutsky matrix of the model, but this cannot be satisfied globally for this class of system. In the empirical application which follows, the integrability conditions are imposed (apart from negativity), since earlier work indicated that homogeneity and symmetry restrictions could not be rejected for these data (Nicol, 1995). Additional discussion of negativity conditions for this application will also be provided in Section 4.

The model, (17), encompasses a variety of effects, all of which have been found to be important determinants of demand. Whether these effects should enter the model simultaneously, or whether some are capturing more than one influence can therefore be explored, given the above specification. In particular, it is important to note that rank three effects *and* the significance or otherwise of conditioning goods, c, are separately identifiable effects, given (17). This is an important point since conditional demands can give the appearance (empirically) of being rank two models if the specification is not sufficiently general. In other words, if $a_{3i}[\ln p; \ell, c, z]$ and $g_3[\ln p, \ln y; \ell, c, z]$ in (9) are not specified appropriately, then the effects of the inclusion of c can mask rank three effects. This is not possible in the present specification, (17), as can be seen by abstracting from the effects of ℓ and z, and writing (17) as depending on $\ln p$, $\ln y$ and c,

$$\begin{aligned} w_i &= \tilde{\alpha}_i(\ln p) + \{\beta_{i0} + \beta_{i1} \cdot (p_c \cdot c)\}\{\ln(y) - a[p]\} + \\ &\quad \{\lambda_i + [\beta_{i0} + \beta_{i1} \cdot (p_c \cdot c)]\lambda(\ln p)\}\{\ln(y) - a[p]\}^2 + \mu_i \end{aligned} \tag{18}$$

where $p_c \cdot c$ are expenditures on the goods included in the vector c. Collecting terms in $p_c \cdot c$ yields

$$\begin{aligned} w_i &= \tilde{\alpha}_i(\ln p) + \beta_{i0}\{\ln(y) - a[p]\} + \{\lambda_i + \beta_{i0}\lambda(p)\}\{\ln(y) - a[p]\}^2 + \\ &\quad \beta_{i1}\{(\ln(y) - a[p]) + \lambda(\ln p)(\ln(y) - a[p])^2\}\{p_c \cdot c\} + \mu_i \end{aligned} \tag{19}$$

which, by appropriate re-definition of the four right hand side terms yields

$$\begin{aligned} w_i &= a_{1i}(\ln p)g_1(\ln p, \ln y) + a_{2i}(\ln p)g_2(\ln p, \ln y) + a_{3i}(\ln p)g_3(\ln p, \ln y) + \\ &\quad \beta_{i1}\{g_2(\ln p, \ln y) + \lambda(\ln p)g_3(\ln p, \ln y)\}\{p_c \cdot q_c\} + \mu_i \end{aligned} \tag{20}$$

Equation (20) is thus a conditional, rank three demand system. It contains the standard terms for a rank three demand system (on the first line of equation (20) above), plus the additional term involving $p_c \cdot c$. One can contrast the first line of equation (20) above with Equation (10) of Banks, Blundell and Lewbel (1997), a recently estimated rank three specification. In the notation above, $a_{1i}[\cdot]g_1[\cdot]$ corresponds to the "function" $\alpha_i + \sum_{j=1}^{n} \gamma_{ij} \ln p_j$ in Banks, Blundell and Lewbel (1997), which is independent of income or total expenditures. If only this function were included, the demands would be rank one, or homothetic, with budget-shares independent of income. The rank two term in Banks, Blundell and Lewbel (1997) involves $\{\ln(y) - a[p]\}$, or $g_2[\cdot]$ in the notation above, whereas the rank three term involves $\{\ln(y) - a[p]\}^2$, or $g_3[\cdot]$ in the notation above.

Many hypotheses can be tested using (17), via zero restrictions on the parameters, α_{ik}, β_{ik} and λ_i. Restricting the model to be a rank two system requires that $\lambda_i = 0$, for all i (λ_0 can be non-zero and the model would continue to be rank two, as long as all $\lambda_i = 0$). Separability of commodity demands from labour supply requires that the α_{ik} and β_{ik} parameters associated with labour force participation variables, ℓ, be zero, and so on. Also, the interaction of household characteristics variables (in the ν vector) with $[\ln(y) - a(p, \nu)]$ yields a non-exactly aggregable demand system. To control for the most important household characteristics effects (as evidenced in the work of Barnes and Gillingham, 1984, and Nicol, 1989), six different household group sub-samples, based on family size and housing tenure status, were employed. Other variables, such as labour force participation effects, were included directly via the vector, ν. Details of the overall approach will be explained further in Section 4.

Elasticity estimates were also calculated, given the following parameterisations for income elasticities (η_i), compensated price (ϵ^*_{ij}) and uncompensated price (ϵ_{ij}) elasticities respectively (following Fry and Pashardes, 1992)

$$\eta_i = \{\beta_0 + \sum_k^K \beta_{ik}\nu_k + 2[\lambda_i + (\beta_0 + \sum_k^K \beta_{ik}\nu_k) \cdot \lambda_0](\ln(y) - \alpha_0)\}/w_i + 1 \quad (21)$$

$$\epsilon^*_{ij} = [1/w_i][\gamma_{ij} + (\beta_0 + \sum_k^K \beta_{ik}\nu_k)\lambda_j(\ln(y) - \alpha_0)^2 + w_i(\eta_i - 1)(w_j - \sum_k^K \alpha_{jk}\nu_k)] + w_j - \delta_{ij} \quad (22)$$

$$\epsilon_{ij} = \epsilon^*_{ij} - w_j\eta_i \quad (23)$$

where $\delta_{ij} = 1$ if $i = j$ and 0 otherwise. It should be noted that these elasticity formulae take into account evaluation at normalised prices, $p = [1, \ldots, 1]^T$ and $y = 1$, which are zero when logs of p_i and y are taken, and that evaluation further requires a value for the vector ν. The vector of sample means was used for this purpose, although the means of elasticities for specific households were also computed, averaged, and compared with the former elasticities.

3 Data

The expenditure data for this study are drawn from the 1980–81, 1982–83, and the annual, 1984–1992 *Interview Survey Public-Use Tapes* of the CEX for the United States. The procedure for collecting data from households in these CEX samples was as follows. Each sample was split into three monthly rotation groups. Households in a rotation group were then interviewed in the "same" month of each quarter for five consecutive quarters, reporting their spending patterns for the preceding quarter. This yields one rotation group reporting their quarterly expenditures every month. Households can therefore be matched to monthly price data, since it is known when a household reports, and for which quarter. In contrast, in the Canadian case, households report expenditures for a whole year, so this limits the price information which can be used when working with the FAMEX.

At any time, there are approximately 5000 households in the CEX Interview Surveys. Pooling data from the eleven indicated surveys permits construction of large samples of different household groups. The characteristics of households within any one group can be chosen to be as similar as desired. Key household characteristics variables are also used to split samples into sub-groups having relatively similar characteristics within the group, but having a range of characteristics across groups. The choice of such "stratifying variables" is based on experience in applied demand research where such variables have been associated with statistically significant differences in model parameters across groups.

Family size and housing tenure have been found to be of such importance in other demand studies (for example, Barnes and Gillingham, 1984, and Nicol, 1989), as to merit stratifying households into groups based on these variables. Household types were therefore classified according to three different family sizes: married couples without children; married couples with one child; and married couples with two children. Also, two types of housing tenure status were used to further classify these households: renter households; and home-owners with mortgages. While it is possible to construct samples with larger family sizes and samples with home-owner families without mortgages, such samples are significantly smaller, and are not the focus of this paper. For all six household types, only those with age of head 18–65 and with no self-employed members were included in the sub-samples. Each of the respective household sizes and housing tenure statuses are distinguished in what follows by the identifiers: MOR0–MOR2 for home-owners; and REN0–REN2 for renters, the ending digit indicating the number of children in the households.

The next step in creating the data sets to be used is the selection of expenditure categories of interest. This choice is governed by a number of considerations. The CEX surveys define disaggregated expenditures into various categories, so these categories are the minimum level of disaggregation one can work with. However, the kinds of categories in the surveys are similar to those used in surveys in other countries (such as Canada and the UK). Consequently, the expenditure categories used in this study can be chosen to be as close as possible to those in other studies.

The choice of expenditure categories is also affected by the availability of price data with which to match expenditures. In addition, the more categories included in a demand system, the greater the number of parameters which have to be estimated. Such estimation is difficult in a nonlinear setting. Large systems can be made smaller by aggregating goods. However, inappropriate aggregation of expenditures can lead to misleading inferences (Nicol, 1991, provides evidence on this in a homogeneity and symmetry testing context). Consequently, a small, disaggregated demand system is preferable from this perspective. There is then the danger, however, of excluding non-separable goods from such a system. Even including the effects of "other goods" (in a vector such as c) in the form of a conditional demand system could lead to misleading inferences, for a variety of reasons. For example, the way in which these other goods influence demand behaviour could be intimately related to their distribution by expenditure type within the other goods category. This distributional information is completely obscured when employing an aggregate other good as an explanatory variable. There has been no research on the kinds of effects which might arise, so there is no way of knowing *a priori* whether one need be concerned over the nature and magnitude of these possible effects.

Given the above considerations, a maximum of nine expenditure categories were chosen for inclusion in the largest possible demand system to be estimated. These were: food (F), alcoholic beverages (A), clothing (C), gasoline and fuel (G), other automobile expenditures (O), public transportation (P), household operation (H), personal care spending (E) and health care spending (S). All other expenditures were dealt with as an aggregated other goods category, or conditioning good (CG). Section 2 above denoted the goods in such a category by the vector, c. The expenditure categories included in CG were: shelter; household furnishings; auto purchase; tobacco, entertainment; education and other expenditures. Thus, CG $= p_c \cdot c$, the inner-product of c with its price vector, p_c. Complete details of these expenditure categories are contained in the *Interview Survey Public-Use Tape Documentation.*

A nine equation demand system with the complexity of the QAI model presented above contains many parameters. For the specification, (17), there are 158 parameters in the nine-equation system. The variability in prices of the data used in this study permits the empirical identification of all these parameters for the six household sub-groups, MOR0–MOR2 and REN0–REN2. Such identification is not routinely possible when using microdata for other countries with more limited price variability, where only smaller systems can be estimated. It is, however, of interest to analyse the impact of estimating such smaller systems, even when a variable such as CG is included, for the reasons alluded to earlier. To this end, not only was the nine equation system (FACGOPHES) estimated, but so also were four other systems. These were three, three equation systems, which included: F, A and C (the FAC system); G, O and P (the GOP system); and H, E and S (the HES system). In addition, a six-equation system including F, A, C, G, O and P (the FACGOP system) was estimated. For these reduced systems, CG not only included those expenditures indicated above for the FACGOPHES system, but also expenditures for the nine equation system not included as equations for the reduced systems. For example, for the FAC system, CG would include expenditures on G, O, P, H, E and S, as well as those expenditures included in CG for the

FACGOPHES system.

Estimating the range of five systems indicated above means that test results and estimated elasticities can be compared for individual systems or goods across estimation situations. In particular, test results for rank three demands can be compared for the FAC versus FACGOPHES system and so on, and estimated elasticities based on these two systems can be compared for F, A and C. Although all such results were computed, in this paper comparisons are confined to those between the FACGOPHES and FAC systems. Interested readers can review other comparisons by referring to an earlier version of this paper (Nicol, 1998), available at

http://www.econ.uregina.ca/nicolc/papers/RankUS

As the model specification discussion in Section 2 indicated, labour force participation effects were to be included in the demand equations. The CEX data contain information on the labour force participation status of adult household members. Consequently, these effects were introduced as labour force participation dummy variables. One variable was included for each of the adult male and female household members. In addition, these dummy variables were interacted with other variables on the right hand side of the estimating equations, as indicated in equation (17).

Other household characteristics effects included in the vector ν were age of the head of the household, a tobacco consumption dummy variable and a vehicle ownership dummy variable. These variables have been found to be important determinants of demand in other studies. (See, for example, Banks, Blundell and Lewbel, 1997).

Given the six samples of households from the CEX (MOR0–MOR2 and REN0–REN2), these households have next to be matched to the price vectors which they faced for the goods directly included in the demand system. There are several other variables in the CEX which influence how this matching can be done. These are the variables giving household location information. The relevant variables are: region of residence (REG); city population size in region of residence (CITY); and state of residence (STATE). There are four REG locations defined in the CEX: Northeast (NE); Midwest (MW); South (SO); and West (WE). Also, five CITY sizes are defined. These change slightly over the eleven CEX data sets used. Also, from 1980–85, the following states were *not* covered by the CEX: Delaware, Idaho, Nevada, New Hampshire, North Dakota, Oklahoma, South Dakota, Vermont and Wyoming. From 1986–92, the following states were *not* covered: Montana, Nevada, North Dakota, Rhode Island, South Dakota, Vermont and Wyoming.

For certain households, some or all of REG, CITY and STATE were suppressed by the Bureau of Labor Statistics, in the interests of confidentiality of the survey respondents. However, using these three variables when not suppressed, it was possible to identify the city in which a household lived for many observations. In surveys prior to 1985, one could determine the households' cities of residence for twenty-seven US cities by cross-tabulating on the REG, CITY and STATE variables. These cities are listed in Table A1 of the Appendix, along with a series of price indices, which will be discussed below. From the 1986 survey onwards, however, the CITY variable was suppressed for all households in WE states. These states were: Alaska, Arizona, California, Colorado, Hawaii, Oregon, Utah and Washington. In those cases, it was still possible to identify city of residence of households living in the twenty cities in the NE, MW and SO regions listed in Table A1. Also, state of residence was provided for households in WE states, and their city of residence could be inferred as one of a small sub-set of cities within each of these states.

The change in reporting of household location variables from 1986 and thereafter had implications for the way that price data could be constructed. For households in cities in the NE, MW and SO, city of residence could be determined for the whole period, 1980–1992. These households could therefore be assigned city prices. On the other hand, for households in WE from 1986–92, only state of residence could be determined exactly, so these households had to be matched to state level prices. This resulted in the introduction of errors in price variables for households in the WE region for such observations. These types of errors are commonly seen in other data sets (for example, Canadian FAMEX-based data sets, and United Kingdom FES-based data sets), as was discussed earlier. Consequently, the effects of these errors in price variables can be assessed by comparing results based on different sub-samples for the non-WE regions after 1986, and for all regions for data from 1980–85. In addition, therefore, to distinguishing amongst the six household types, MOR0–MOR2 and REN0–REN2, estimation was also carried out: using four-region data from 1980–1992; three-region data from 1980–1992 (excluding households from WE); and four-region data for households from all twenty-seven cities from 1980–1985.

The CEX also indicate the month in which a household is interviewed. In principle, a household could thus be assigned a price vector reflecting when they made their expenditures. With twelve monthly price observations per year in thirteen years for up to twenty-seven cities, price data would a lot of information, and are highly variable within a given sample. Price effects could then be determined precisely. However, in the case of the CEX, although detailed city-level CPI data are available and these can be matched to households at the city level, CPI data do not reflect how prices change over time for the specific cities. That is, city-level CPI's do not indicate whether prices in New York are higher than in Los Angeles for the same good, such as F. Fortunately, an extensive database of price indices for six categories each year, and for many cities, is available. These indices are published by the *American Chamber of Commerce Researchers Association* (ACCRA). The expenditure categories covered are: grocery items, housing, utilities, transportation, health care and miscellaneous goods and services. This is a smaller number of categories than are included in the FACGOPHES system. A representative sample of the ACCRA data (which also lists the twenty-seven cities covered in this study) is presented in Table A1 of the Appendix. These data, and others like them, were used in conjunction with city level CPI data to construct inter-city price indices for the cities in Table A1. The ACCRA categories were assigned to the FACGOPHES categories as closely as possible, given the expenditures included in the respective categories. These assignments are given in the Appendix Notes to Table A1.

To provide an impression of the impact of using ACCRA-adjusted CPI prices, rather than the CPI prices themselves, Figures 1–3 below show the behaviour of these series from 1980–1996. The comparisons are for F, A and C prices in New York city relative to Los Angeles, based on CPI only versus ACCRA-adjusted CPI prices.

From these figures, it can be seen that the series for the different goods shift by different amounts, following ACCRA-based adjustments to city prices, depending on which good is considered. In addition, the relationship for C prices between the two cities is actually *reversed*, relative to the corresponding comparisons for F and A. These graphs therefore serve to demonstrate that one can expect to observe markedly different results, depending on the price series used (CPI only, versus ACCRA-adjusted CPI prices).

Once the city and state level ACCRA-adjusted CPI price data were constructed as described above, they were then matched to appropriate households in the various samples. Further details of the price construction procedure, matching to households, and extraction of actual households from the complete CEX are available on request.

4 Estimation and Results

4.1 Exogeneity of Explanatory Variables

It is becoming increasingly common in applied demand studies using microdata for instrumental variables estimation procedures to be used. This is because of concern over purchase infrequency in some bodies of data, which calls into question the exogeneity of total expenditures (y). The inclusion of labour force participation effects also lends strong support to this approach, owing to the possible endogeneity of such variables. Also, in the current study, several other included right-hand side variables could, by their nature, be non-exogenous. In particular, the conditioning of demands on expenditures outside the demand system requires inclusion of the $p_c \cdot c$ variable, and allowing for the effects of tobacco consumption and vehicle ownership requires inclusion of dummy variables to model these effects.

In all systems estimated, therefore, total expenditures, labour force participation of adult male and female household members, other expenditures $p_c \cdot c$, vehicle ownership and tobacco consumption were included as right-hand side variables, but were accounted for through the use of instrumental variables. Other right-hand side variables which were included but did not require instruments were age of the head of the household and prices.

In these circumstances, the additional instrumental variables required to implement estimation were: age squared of adult male and female household members; four regional dummy variables (for NE, NC, SO and WE); linearly independent squares and cross-products of price variables; a time trend; income after taxes and its square; income after taxes interacted with all adult age variables; personal taxes; government transfer payments; seven occupation dummy variables each for male and female adult members; and seven education dummy variables each for male and female adult members. The occupational dummy variables were categorised as working in the following areas: managerial; technical; service; farming, forestry, fishing; production, craftsmen, repair; operators, fabricators, labourers; armed forces; and not working. Educational dummy variables were defined as having schooling to the levels: elementary; some high school; high school graduate; some college; college graduate; more than four years of college; and no schooling.

Given the above set of instrumental variables, estimation of the model (17) was carried out for all of the household groups and expenditure category demand systems indicated earlier. The GMM procedures in TSP, Version 4.3 were employed, allowing for heteroskedasticity of unknown form in the computation of the estimated variance-covariance matrices of these systems, where possible.

Attention is confined in the discussion which follows to the FAC and FACGOPHES systems, as mentioned earlier.

For the FAC system, there are thirty-five parameters to be estimated for the complete system. Depending on the sample being employed, there can be up to sixty-five over-identifying restrictions (when all instruments are linearly independent). The total number of restrictions ranged from fifty-seven to sixty-five, depending on the data set used. That is, for some household groups, there were instances of no household members in certain occupation types or schooling types. Sargan (1983) or J tests of over-identifying restrictions indicated non-rejection of the over-identifying restrictions, with upper-tail probability values for these test statistics ranging from 0.08 to 0.94. The outcomes of these tests were unaffected whether the the data covered households in: cities alone from all four regions (data from 1980–85); cities alone from the three regions, NE, NC and SO (data from 1980–92); or data covering households in cities in NE, NC and SO, and households with State identifiers for WE (data from 1980–92). These tests of over-identifying restrictions thus do not appear to be sensitive to errors in price variables.

With respect to the FACGOPHES system, there are 158 independent parameters in the complete, eight equation (singular) system. Given the instrument set discussed earlier, and the included exogenous variables, there was a total of up to 386 linearly independent over-identifying restrictions, and a range from 354–386. Two out of six test statistics indicated rejection of the over-identifying restrictions in the results where errors in price variables were suspected to be a potential for problems (with four-region data from 1980–92). However, there were no rejections of the over-identifying restriction tests for other results, with upper-tail probability values here ranging from 0.01 to 0.41. This result is in contrast to that for the FAC system, mentioned above. Recall, however, that the FAC system includes non-FAC expenditures as a composite, other goods expenditure category, $p_c \cdot c$. As was mentioned above, it is possible that this means of handling non-separability of excluded goods ($p_c \cdot c$) from included goods could have unexpected effects on the outcome of hypothesis tests. The difference between over-identifying restriction test outcomes across the FAC and FACGOPHES systems could therefore be an illustration of such effects.

To conclude, however, with respect to tests of over-identifying restrictions, and the exogeneity of the selected instruments, these results would appear to indicate that the choice of instruments is appropriate. A question arises, however, with respect to possible effects on these (and other) test results of treating a large block of expenditures as an aggregate "conditioning good". Further discussion on this will be taken up in the succeeding sub-sections.

4.2 Model Specification Tests

One of the purposes of this paper is to determine the importance of a series of factors which have previously been found to influence consumer behaviour. In addition, by comparing test outcomes across demand systems including different categories of expenditures, the impact of this factor can also be highlighted in terms of its effect on test outcomes. This will provide evidence for applied researchers on the need to model certain effects, while possibly being able to abstract from others, through the use of a relatively more parsimonious model specification.

Another avenue for exploration is the analysis of the effects of errors in price variables on test outcomes, amongst other things (to be discussed later in this section). As discussed above, the data being used in this study presents an unusual opportunity to obtain insights on this issue. In particular, the ability to identify city of residence for households living in three of the regions of the United States from 1980–1992 allows for more precise estimation of responses than has previously been possible in the applied demand literature.

The model specification tests to be conducted involve various zero restrictions. Attention in this paper is focused on test results for the FACGOPHES and FAC systems. Additional test results for other systems estimated are presented in Nicol (1998), available at the URL (uniform resource locator) indicated above.

The restriction that $\lambda = [\lambda_0, \lambda_1, \ldots, \lambda_n]^T = 0$, reduces the system to one of rank two demands. In Tables 1–5 which follow, this is referred to as the hypothesis "Excluding Q". The importance of age of head ("Excluding A"), labour force participation ("Excluding L"), conditioning goods ("Excluding C"), vehicle ownership ("Excluding V") and tobacco consumption ('Excluding T") as determinants of demand can also be analysed by tests of exclusion restrictions for these respective variables.

Test results in Tables 1 and 2 are for the FACGOPHES system, for households MOR0–MOR2 and REN0–REN2 respectively. For the six hypothesis tests, three sets of results are presented for each household type. The first block of results in each of these tables are based on data where errors in price variables are present, and all four regions of the United States from 1980–92 are covered. These results indicate rejection of the null hypotheses: Excluding Q, A, L, C, and V for all six household groups. Also, rejection of the Excluding T hypothesis in four cases is observed. In other words, on the basis of these tests, one would be compelled to conclude that demand is rank three for all households types. Furthermore, age of head, labour force participation, conditioning goods and vehicle ownership are unambiguously important determinants of demand. It is particularly interesting to note in the context of these results that not only is the rank three hypothesis supported, but so also is the hypothesis that the conditioning good ($p_c \cdot c$) should be included. This confirms the earlier analytical proposition that these two effects are separately identifiable, and that indeed the effects are also empirically identified, based on these data.

As was discussed in Section 3 above, the price data from 1980–1992 for the four regions of the United States, NE, MW, SO and WE, contain errors in the price variables for households in WE. In particular, it is only possible to identify state of residence for these households. Hence, the prices which households in WE are faced with in the data from 1986–1992 are state level prices. However, from 1980–1985, it is possible to identify city of residence for all households. In addition, it is possible to identify city of residence for all households in NE, MW and SO from 1980–1992.

The second block of results in Tables 1 and 2 contain test results based on estimation of the FACGOPHES system, but for households resident in the four regions from 1980–1985 only. Thus, the data in these samples for MOR0–MOR2 and REN0–REN2 comprise households in the twenty-seven cities identified in Table A1 of the Appendix. In contrast with the results with errors in prices, it is found here that there are *no* rejections of the Excluding Q (rank two) hypothesis. For the Excluding A, L, C and V hypothesis, there are no more than three rejections each, and these tend to be confined to the MOR0–MOR2 household groups. On the basis of these results, it thus appears that the earlier results seem to be biased towards rejection of all hypotheses, presumably due to the presence of the errors in prices. It is also of interest to note that, for REN0–REN2 households, these results indicate that the rank two hypothesis *and* exclusion of $p_c \cdot c$ are not rejected.

The above results provide an interesting contrast with the earlier results, in the same tables. One might be tempted to conclude that the errors in variables described above are the factor driving these differences. This is probably true to some extent, but it must also be borne in mind that there are certain confounding influences to be considered. The test results in the mid-sections of Tables 1 and 2 are based on city-level data, but the data sets themselves are subsets of the data which are used to obtain the earlier test results in the top sections of the same tables. Thus, the second set of test results are based on less information, even though *some* of the information in the larger data sets is error prone. This alone could be the reason for the results, so further analysis is warranted.

In the bottom sections of Tables 1 and 2, test results are provided which are based on estimation with data from 1980–1992, with households resident in the three regions, NE, MW and SO. That is, households resident in the cities in these three regions, and identified in Table A1 of the Appendix. These samples are larger than those based on the 1980–1985 data alone, even though the latter are for four regions. The pattern of rejections here is closer to that for the 1980–1992 four-region data results. It is still the case, however, that there are less rejections for the test results based on three-region data. Variables which are the most additional important determinants of demand according to this last series of results are age of head, labour force participation and vehicle ownership. Rank three effects and $p_c \cdot c$ appear to be less important, although these effects cannot be discounted for all household types. These results, in contrast to those in the top sections of Tables 1 and 2, thus seem to indicate that the errors in price variables tend to bias all test results towards rejection, for the various hypotheses under consideration.

The comparison of test results in the top and bottom sections of Tables 1 and 2 are compelling in terms of the apparent effects of the errors in variables. However, it should be borne in mind that the first set of results are based on data which includes households from the WE region from 1980–1992, resident in cities. There is therefore a confounding influence possible between these two sets of results, as the data are not identical, even though the more expansive data set includes observations on prices which are error prone.

To remove the above confounding influence, one can construct price data from 1980–1992 for the three-region households, but average these prices across regions. Thus, each household could be faced with a price vector representing the regional price for the goods in FACGOPHES, for the period when they were making their expenditures. These data can then be used to re-estimate the model for the various household groups, and tests results such as those in Tables 1 and 2 computed. Table 3 contains test results computed on this basis. That is, any differences between these results and those in the last sections of Tables 1 and 2 are purely associated with the induced errors in price variables generated through the creation of the alternative price series described above.

Table 3 indicates similar results for the finer, three-region based test results in Tables 1 and 2 with respect to the Excluding A, L, V and T hypotheses. However, the tests for rank two demands and exclusion of $p_c \cdot c$ indicate a greater number of rejections for these hypotheses than was the case with the same data without the errors in price variables. The results in Table 3 thus confirm that errors in price variables have a significant impact on the outcome of tests for exclusion of the additional types of determinants of demand which are the focus of this paper. In particular, these errors in variables bias test results towards rejection.

A final sequence of results is the analogous series of tests which appeared in Tables 1 and 2, but for the FAC system. These results are contained in Tables 4 and 5, for MOR0–MOR2 and REN0–REN2 respectively. These results indicate fewer rejections of the various hypotheses in general. The majority of the rejections are concentrated in the top block of the two tables, where the data used for estimation included the errors in price variables already discussed. In this case there are ten out of a possible eighteen rejections. The mid-sections of the tables (1980–1985, four-region data) indicate only three rejections, while the bottom sections (1980–1992, three-region data) indicate five. It is therefore apparent that the categories included in the demand system to be estimated has implications for the outcome of hypothesis tests. This is true even here where the goods excluded from the system directly are treated as a composite, conditioning good, $p_c \cdot c$.

The broad conclusions which emerge from this discussion are therefore as follows. Age of head, labour force participation, vehicle ownership and tobacco consumption are important additional determinants of demand which ought to be modelled. These determinants are over and above household size and housing tenure considerations. For home-owners households, the rank of demands appears to be three, and goods excluded from the system directly can also be important for this group. However, these considerations seem to be less important for renter households. This is consistent with earlier research, for example, Lewbel (1991), where Engel functions were found to be rank two when households from the extremes of the income distribution were excluded. All of these test results are influenced significantly by the presence of errors in price variables. The largest effects appear to be with respect to the rank hypothesis, and the exclusion of conditioning goods, $p_c \cdot c$. Finally, although estimating a system with a large number of equations is computationally demanding, this appears to dominate estimation of a much smaller system, even when one controls for excluded goods. That is, the test results are qualitatively different when based on the relatively smaller system.

4.3 Estimated Elasticities

Uncompensated price and income elasticities were computed based on the parameter estimates of each of the five systems, six household types and three regional groupings, given the formulae in equations (21)–(23) of Section 2. The test results in the previous section indicated that the FACGOPHES system was least likely to be influenced by specification error. Thus, attention is concentrated in this section on elasticity estimates obtained from FACGOPHES-based results. Readers interested in elasticity estimates based on the other systems are referred to Nicol (1998), at the URL indicated above.

Tables 6 and 7 contain own-price and income elasticity estimates respectively, for households MOR0–MOR2, REN0–REN2, for the three different types of temporal and regional data discussed above. The first three columns of each table give own-price and income elasticity estimates for 1980–1992 four-region data. Five of these own-price elasticity estimates are non-negative, but their standard errors (reported below the elasticity estimates) are large relative to the elasticities themselves. Variations in values across family sizes, as well as housing tenure, are apparent in these estimates. This confirms that classifying households by these variables is necessary if one is to obtain a useful picture of consumption responses to exogenous changes. With respect to income elasticities, F and G are uniformly income inelastic whereas C is uniformly income elastic with these data.

The last six columns of Tables 6 and 7 give the opportunity to examine the effects of errors in price variables on estimated elasticities. Columns four to six give results for 1980–1985, four-region data, and columns seven to nine, results for 1980–1992, three-region data. Since, in general, income elasticities are more precisely estimated than own-price elasticities, comparisons between income elasticities with and without errors in variables are of more interest. Again, one can see that F and G are income inelastic, while C is income elastic. The estimates are influenced by errors in price variables, but not to the extent that this makes a qualitative difference to them. That is, income elasticity standard errors are not small enough that differences could be viewed as statistically significantly different.

To summarise, the extent to which elasticities take economically meaningful values, particularly own-price elasticities, is encouraging, given the size of the system being estimated. This is possible because of the significant amount of price variability in these data. In addition, the model is extremely general in terms of the influences which are modelled. Thus, the possibility of inaccurate elasticity estimates due to model specification error has been minimised.

The elasticities presented in Tables 6 and 7 are evaluated at sample mean household characteristics. An alternative approach often used in this area is to evaluate elasticities for each household, and then average them. In addition, using this second approach, since compensated elasticities are available, one can assess the number of data points at which the curvature conditions associated with the Slutsky matrix are satisfied. In the context of this (and many other) demand model(s), however, complete Slutsky matrices can only be calculated for a limited number of households, since w_i is a divisor in the expression for compensated elasticities, (22). An alternative, therefore, is to compute compensated own-price elasticities for each household for which this is possible, and determine the proportion which are negative for each good. These estimates are given in Table 8, for households MOR0–MOR2 and REN0–REN2, using the 1980–1992, three-region data. As is to be expected, these estimates are more variable than those for (uncompensated) own-price elasticities in Tables 6 and 7. However, there are only five out of fifty-four cases for which less than half of the households do not have negative compensated own-price elasticities. Since this kind of functional form cannot satisfy curvature conditions at all possible data points, these results are encouraging, as they show that Slutsky elasticities are well-behaved in a significant range of points for which data are observed.

In conclusion, the evidence provided in this paper suggests that a demand system cannot be specified so that all types of households and considerations can be handled simultaneously. That is, there are important differences in demand patterns and the determinants of these patterns across households which cannot be assumed away. Findings that a particular model works well in the context of a data set comprising homogeneous households does not mean that the same model will perform equally well for other household groups. In particular, while demands might appear to be rank three for a given household group, as has been seen, this cannot be assumed to be generally true. In addition, while labour force participation variables are important determinants of demand for the MOR0–MOR2 households identified in this paper, this does not appear to be true for REN0–REN2 households. The only evidence in favour of this hypothesis was obtained when significant errors in the variables being used were present. Furthermore, these errors in price variables tended to have important effects on the outcomes of several hypothesis tests. The nature of these errors in variables, unfortunately, tend to pervade available data for various countries. Thus, such results must be interpreted with caution.

5 Summary and Conclusions

In this paper, a rank three demand system was estimated. The model also controlled for the non-separability of labour force participation effects, non-separability of other goods not included directly in the demand system, and the influence of a variety of household characteristics effects. Systems with from three equations to nine equations were estimated, for six different household types. Given the nature of the data available, it was also possible to analyse the effects of errors in price variables on the outcome of hypothesis tests and estimated elasticities.

The possible rank three nature of demands appeared to be significantly affected by the presence of errors in price variables. Test results in the absence of these errors indicated some evidence in support of the rank three hypothesis. This evidence was limited to two out of the six household types considered.

Various other influences such as labour force participation were also found to be important determinants of demand. However, this finding was not universal for all households types considered in this study. Again, there was evidence errors in price variables on these test results. The empirical evidence in this paper therefore appears to indicate that there is no completely general model of demand which applies to all households of different types, even where differences in corresponding parameters across these systems are permitted.

Estimated elasticities were also computed for the different models. The values arising from the largest system estimated (nine equations) were extremely encouraging, with elasticities taking economically meaningful values in general. There was evidence, however, that the presence of errors in price variables influenced these elasticities, although not to the extent of changing their qualitative nature.

The data used in this study were selected from a series of US CEX surveys. However, not all available data were used in this particular application. The results obtained here were designed to help give guidance for future research. In particular, for certain household types, some variables included as determinants of demand here were not found to be statistically significant. Also, smaller demand sub-systems did not perform nearly as well as the largest possible system, and errors in price variables were seen to have important effects on hypothesis test outcomes. In future work, these factors can therefore be taken into account, to obtain more precise estimates of parameters of interest, elasticities, equivalence scales, and other demand-related estimates.

References

American Chamber of Commerce Researchers Association, Inter-City Price Indices for the United States. Various years.

Banks, J., R. Blundell and A. Lewbel (1992), "Kernel Regression, Quadratic Logarithmic Engel Curves and Welfare Measurement." Institute for Fiscal Studies Working Paper No. 92/14.

Banks, J., R. Blundell and A. Lewbel (1997), "Quadratic Engel Curves and Consumer Demand." *The Review of Economics and Statistics*, **79**, 527–539.

Barnes, R. and R. Gillingham (1984), "Demographic Effects in Demand Analysis: Estimation of the Quadratic Expenditure System." *Review of Economics and Statistics*, **66**, 591–601.

Blundell, R., P. Pashardes and G. Weber (1993), "What Do We Learn About Consumer Demand Patterns From Micro Data." *American Economic Review*, **83**, 570–597.

Browning, M. (1983), "Necessary and Sufficient Conditions for Conditional Cost Functions." *Econometrica*, **51**, 851–856.

Browning, M. and C. Meghir (1991), "The Effects of Male and Female Labor Supply on Commodity Demands." *Econometrica*, **59**, 925–951.

Bureau of Labor Statistics, Washington DC. Interview Survey Public-Use Tapes, Consumer Expenditure Surveys, 1980–1992.

Cragg, J. G. (1994), "Innis Lecture: Making Good Inferences from Bad Data." *Canadian Journal of Economics*, **27**, 776–800.

Dagenais, M. (1994), "Parameter Estimation in Regression Models With Errors in Variables and Autocorrelated Disturbances." *Journal of Econometrics* **64**, 145–163.

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

Dickens, R., V. Fry and P. Pashardes (1993), "Non-linearities and Equivalence Scales." *Economic Journal*, **103**, 359–368.

Fry, V. and P. Pashardes (1992), "An Almost Ideal Quadratic Logarithmic Demand System for the Analysis of Micro Data." Discussion Paper No. 25, City University of London.

Gorman, W. M. (1981), "Some Engel Curves." *Essays in the Theory and Measurement of Consumer Behaviour in Honour of Richard Stone*, A. Deaton, editor, Cambridge University Press.

Howe, H., R. Pollak and T. Wales (1978), "Theory and Time-Series Estimation of the Quadratic Expenditure System." *Econometrica*, **47**, 1231–1247.

Kaiser, H. (1993), "Testing for Separability Between Commodity Demand and Labour Supply in West Germany." **18**, 21–56.

Lewbel, A. (1990), "Full Rank Demand Systems." *International Economic Review*, **31**, 289–300.

Lewbel, A. (1991), "The Rank of Demand Systems: Theory and Non-parametric Estimation." *Econometrica*, **59**, 711–730.

Michelini, C. (1999), "The Estimation of a Rank 3 Demand System With Demographic Demand Shifters from Quasi-unit Record Data of Household Consumption." *Economics Letters*, **65**, 17–24.

Nicol, C.J. (1989), "Testing a Theory of Exact Aggregation." *Journal of Business and Economic Statistics*, **7**, 259–265.

Nicol, C.J. (1991), "The Effect of Expenditure Aggregation on Hypothesis Tests in Consumer Demand Systems." *International Economic Review*, **32**, 405–416.

Nicol, C.J. (1995) "Model Specification Issues in Consumer Demand Systems Using United States Microdata". Department of Economics Working Paper No. 56, University of Regina.

Lyssiotou, P., P. Pashardes and T. Stengos (1999), "Testing the Rank of Engel Curves with Endogenous Expenditure." *Economics Letters*, **64**m 61–65.

Pollak, R. (1969), "Conditional Demand Functions and the Implications of Separability." *Quarterly Journal of Economics*, **83**, 70–78.

van Daal, J. and A. H. Q. M. Mierkes (1989), "A Note on the Quadratic Expenditure Model." *Econometrica*, **57**, 1439–1443.