2002s-87

How Large is Your Reference Group?

François Gardes, Claude Montmarquette

Série Scientifique
Scientific Series

Montréal
Novembre 2002

CIRANO
Centre interuniversitaire de recherche
en analyse des organisations

CIRANO

ISSN 1198-8177

How Large is Your Reference Group*

François Gardes[†], *Claude Montmarquette*[‡]

Résumé / *Abstract*

L'interdépendance des préférences telle que spécifiée dans les études économétriques de consommation ou de choix individuel conduit à des estimateurs biaisés. Dans cette étude, nous présentons de nouvelles spécifications économétriques qui prennent en contre ce problème et qui permettent un estimé de la taille du groupe de référence. Ce dernier élément est ignoré dans les écrits actuels et s'avère très important pour juger des biais d'estimation. Nous montrons à l'aide de données françaises sur le niveau relatif et subjectif de pauvreté que ce groupe de référence est vraisemblablement de très petite taille.

We discuss how specifications of interdependent preferences found in the literature yield biased estimates of parameters of the underlying consumption or choice models. We present new specifications which alleviate this problem and permit an estimation of the size of the reference group. This last point, a key element affecting the estimation biases, has been overlooked in most studies. Using French individual data on the reported subjective poverty level, we show that the reference group is likely to be very small.

Mots clés: Interdépendance des préférences, estimateurs biaisés, taille du groupe de référence.

Keywords: *Interdependent preference, biased estimates, size of the reference group.*

Codes JEL : C5, D0, I3

---

* Our thanks to Alain Carpentier, Pierre Garrouste and Louis Lévy-Garboua for helpful suggestions and comments. We also thank Marie-Christine Giroux for her editorial assistance.

† CREST-LSM, Université Paris 1 (Panthéon-Sorbonne)

‡ CIRANO and Department of Economics, University of Montreal

1. Introduction

Subsequent to Duesenberry's classic relative income hypothesis (1949, chapter 3), Pollak and Wales (1992), Kapteyn, van de Geer, van de Stadt and Wansbeek (1997) have proposed to account for the interdependency of individual preferences by adding to the individual's consumption function the average demand of a reference group or population. Recently, Aronsson, Blomquist and Sacklén (1999) have studied how individuals' choices of hour of work are influenced by the average hours of work in a social reference group. Gaviria and Raphael (2001) have analyzed school-based peer effect and juveniles to engage in drug use, alcohol drinking, cigarette smoking, church going and dropping out of high school.

Following Manski's seminal contribution (1993) on the problem of identification, the definition of reference groups, and the endogenous interaction and reflexion problem involved in neighborhood and peer-group effects, various strategies have been suggested to cope with these issues. Gaviria and Raphael have instrumented the average behavior of the reference group to solve the endogeneity problem. Aronsson et al. used panel data or data from several points in time to disentangle the effects of interdependent behavior and preference variation across groups.

Our approach relates to these works. We show how estimation error affects the parameters of the underlying model, as well as the associated endogenity problem and the correlated error terms in cases where the interdependence of preferences is not considered. The originality of our paper is to provide, in the context of cross section data, an estimation of the size of the reference group. This question is a key element affecting the estimation biases and has been overlooked in the literature. Using French individual data on the reported subjective poverty level, we show that the reference group is likely to be very small. The study among interacting members of an industrial oligopoly and the growth of fiscal frauds and tax evasions linked to the number of delinquents in a community are two examples that can benefit from this research.

In Section 2, we present our approach. In Section 3, we propose different methods to estimate the interaction effect and the size of the reference group. In Section 4, we discuss the results of these methods applied to the reported subjective level of poverty or the minimum income declared by French households « to be sufficient to make ends meet ».

2. Accounting for the interdependency

Suppose two households h and h' who mutually consider the current consumption (or decision) of the other when choosing their own consumption of some good *x*. For h:

$$x_h = Z_h\beta + \eta x_{h'} + \varepsilon_h, \tag{1}$$

where Z_h is a set of explanatory variables and β the corresponding parameters. ε_h is the error term. With the current consumption of the other household h', $x_{h'}$ obtained from (1) and substituted into equation (1) yields the reduced form:

$$x_h = (1/(1-\eta^2))[(Z_h + \eta Z_{h'})\beta + (\eta\varepsilon_{h'} + \varepsilon_h)]. \tag{2}$$

Equation (2) shows that the assumption of interdependent preferences implies that (i) the residual errors are correlated between households; (ii) the current consumption of each household depends on the determinants of the other household's current consumption; (iii) the parameters of vector β are biased by a factor $1/(1-\eta^2)$. Estimating equation (1) by OLS and ignoring $x_{h'}$ or falsely including $\eta Z_{h'}$ in the constant term underestimates the effect of a given change in one element of Z_h by 1% if $\eta = 0.1$, and 178% if $\eta =$ 0.8. Moreover, estimating equation (1) with $x_{h'}$ implies an endogeneity bias as $x_{h'}$ depends on x_h. Finally, the estimators of β's are inefficient because of the correlated error terms.

More generally, let the consumer h be influenced by the choices of a reference group $\Re_h$ (containing n_h consumers h', including h). Consider the reciprocity assumption:

$(H1): h' \in \Re_h \Rightarrow h \in \Re_{h'}$, a situation that occurs whenever the reference group is defined by socio-economic variables. Therefore, $\Re_{h'} \cap \Re_{h'} \neq \varnothing \Rightarrow \Re_h = \Re_{h'}$.

Equation (1) becomes:

$$x_h = Z_h\beta + X_h\eta + \varepsilon_{ih}, \qquad X_h = \{x_{h'}, h' \in \Re_h, h' \neq h\}, \tag{1'}$$

with η a set of $(n_h - 1)$ parameters corresponding to all households in $\Re_h$ except h, that is $\Re'_h = \Re_h \setminus \{h\}$.

Our second assumption, H2, supposes that all the influences of households h' in $\Re_h$ on household h are taken into account with the average consumption $x(\Re_h)$ of the reference group $\Re_h$:

(H2): $X_h = \{x_{\Re_h}\} = \left\{ \sum_{h' \in \Re'_h} \alpha_{h'} x_{h'} \right\}$, for $h' \in \Re'_h$. $\alpha_{h'}$ is the reference weight measuring the importance the consumer h attaches to consumer h' spending on the good *x*.

Compared to equation (1), $\{x_{\Re_h}\}$ replaces $x_{h'}$ and x_h appears in the right side through the direct and indirect dependencies of $x_{h'}$ on x_h (as $x_{h'}$ depends on all x_k, $k \in \Re_h$, which depend also on all x_k, $k \in \Re_h$). We write the following system of n_h equations:

$$x_h = Z_h \beta + \eta \sum_{h' \in \Re'_h} \alpha_{h'} x_{h'} + \varepsilon_h \, . \qquad (3)$$

With (H3), we assume that $\alpha_{h'}$ is a constant in the reference group $\Re_h$ i.e. all consumers pertaining to $\Re_h$ assign the same weight $\alpha_{h'}$ to the expenditure made by consumer h'. Therefore: $\alpha_{h'} = 1/(n_h - 1)$.

In order to estimate the size of the reference group, we assume (H4), $n_h = n$, that is the reference group is of equal size for each household.

From (H1) the reference groups, in finite number, are disjunct and cover the total household population.

Consider the matrix $M = M_r(n,n)$ composed of one in the diagonal and $-\eta\alpha_{h'}$ elsewhere, for a typical *r* reference group. The system of equations (3) for all households in this reference group is simply:

$$Mx_h = Z_h \beta + \varepsilon_h \, . \qquad (4)$$

We obtain the inverse matrix M^{-1} (see Appendix):

$$M^{-1} = \{I_h - [\eta/(1-\eta)(1-n)]\mathbf{1}\}/[1-\eta/(1-n)]. \tag{5}$$

I_h is the identity matrix of size n and **1** the (n,n) matrix of one. It is possible to write (4) as:

$$x_h = M^{-1}Z_h\beta + M^{-1}\varepsilon_h. \tag{6}$$

We can extent this formulation to all reference groups by concatenation.

Estimating equation (6) on Z_h instead of $M^{-1}Z_h$ gives rise, as seen earlier, to: (i) an autocorrelation of the residuals between equations for different consumers in the same reference group; (ii) a specification bias due to the omission of the explanatory variables for the other consumers in the reference group; (iii) if $1 > \eta > 0$ an over-evaluation of β, by a factor $\theta = \{[1-\eta/(1-\eta)(1-n_h)]/(1-\eta/(1-n_h))\}^{-1} > 1$. It is increasing with the interdependent effect η, and decreasing with the size of the reference group n_h.

Our approach yields a reduced form without referring to strong hypotheses to avoid the endogeneity problem. In contrast, Pollak and Wales (1992), suppose that the consumptions of reference households influence the consumption choices of other households with a delay equal to the period of observation (generally one year). It seems rather that the demonstration effect is more rapid. Indeed all explanatory variables are submitted to different adjustment delays that should be taken into account with a dynamic specification. Kapteyn et al. (1997) also had to make several strong hypotheses to obtain their reduced form which characterizes the distribution of the weights $\alpha_{h'}$. For instance, the covariance between the total weight $\alpha_{h'}$ and the mean of x over the same reference population is supposed to be positive if the individual belongs to a population with an above-average level of x. Most hypotheses are not independently tested but embedded in their final specification.

3. Methods to estimate the interaction effect and the size of the reference group

In Section 2, we have laid out the basic econometrics elements and challenges that arise when dealing with the question of individuals' interdependent preferences. In this section, we consider various methods to estimate the interdependence coefficient η and the size of the reference group *n*. The methods are to some extent complementary and rely on different assumptions, some more restrictive than others.

Method A is a simple OLS on equation (3) directly. We fix the interdependence coefficient η to a starting value (let say $\overline{\eta} = 0.1$) to obtain $\hat{\beta}^{(1)}$ for the vector of coefficients. Next, summing equation (3) over the reference populations $\Re_h$ yields the following aggregate equation, $x_{\Re_h} = \overline{Z}_{\Re_h}.\hat{\beta}^{(1)} + \eta\, x_{\Re_h} + \overline{e}_{\Re_h}$ recognizing that the average of $x_{\Re_h}$ over $\Re_h$ is simply $x_{\Re_h}$. Thus, it is possible to estimate η from $x_{\Re_h} = \frac{1}{1-\eta}\overline{Z}_{\Re_h}\hat{\beta}^{(1)} + \zeta_{\Re_h}$ with $\zeta_{\Re_h} = \frac{1}{1-\eta}\overline{\varepsilon}_{\Re_h} + \xi_{\Re_h}$. This estimate of η is then used to re-estimate β from equation (3) until convergence on η. This is a very simple method to obtain η, which uses the entire sample and avoids specifying the reference group size n.

The reference group size is itself an interesting issue that depends on the type of problem, which is addressed in our interdependency framework. For example, the subjectivist conception of the individual's information, as proposed by Hayek (1948, 1952), implies a rather small size of the reference group. According to Hayek, social influences come from interactions between individuals (no social behavior exists as such), but individual behaviors are so heterogeneous that nobody can take into account all of the interactions that exist between a person and their acquaintances. Hayek's point of view is not psychological and does not concern imitation behavior. It relies only on the cost of acquiring and treating the information. So, one must consider a limited number of people interacting through compatible actions. In fact, these influencing persons represent typical attitudes and can be considered as proxies for the attitudes of all corresponding acquaintances.

Method B allows an evaluation of the size of the reference group. We rewrite equation (3) separating the variables chosen to identify the reference population (for example, age and education), associated with the vector Z^1, from the vector Z^2 of the variables that are different from the grouping criteria (for instance income or the number of workers in the household)

$$x_h = Z_h{}^1 \beta^1 + Z_h{}^2 \beta^2 + \eta\, x_{\Re_h} + \varepsilon_h \,. \tag{7}$$

The usual spectral decomposition of the variance in the panel data analysis into the between and the within dimensions, is applied to the clustering structure of our cross-section data set.
This is naturally implied when studying interdependent preferences or behavior.
In this spatial specification the corresponding between transformation is: $By_h = \frac{1}{n_h} \sum_{h' \in \Re_h} y_{h'}$. The within transformation is: $Wy_h = y_h - By_h$. Note that the variables in Z^1 have 0 within component, while the variance of the variables in Z^2 contains non-null between and within components.

Rewriting equation (7) in the between and within dimensions, we obtain:

$$\begin{cases} Bx_h = \frac{1}{1-\eta} BZ_h{}^1 \beta^1 + \frac{1}{1-\eta} BZ_h{}^2 \beta^2 + \frac{1}{1-\eta} B\varepsilon_h, \\ \\ Wx_h = WZ^2 \beta^2 + W\varepsilon_h, \end{cases}$$

as $WZ^1 = 0$ and by assuming $Wx_{\Re'_h} = 0$ (assumption H).

Assumption (H) cancels the interaction term and permits a direct estimation of the β^2 in the within equation. By simultaneously estimating the between and within equation with the same β^2 for variables in Z^2, and by comparing the between and within estimates of the coefficients of Z^2, we obtain an estimate of the interaction coefficient η.

The general within equation with $Wx_{\Re'_h} = -\frac{1}{n-1} Wx_h$ is[1]:

$$Wx_h = \frac{1}{1+\frac{\eta}{n-1}} WZ_h^2 \beta^2 + \frac{1}{1+\frac{\eta}{n-1}} W\varepsilon_h$$

This general specification of the within equations yields a relation between the interaction coefficient and the size of the reference group obtained by the ratio of the coefficient γ of Z^2 in the between and within dimensions:

$$\gamma = \frac{\hat{\beta}_B}{\hat{\beta}_W} = \frac{1}{1-\eta}\left[1+\frac{\eta}{(n-1)}\right]. \tag{8}$$

Fixing η yields an estimate of the size of the reference group.

Another method (Method C) to estimate the size of the reference group conditional on a given η consists in comparing the coefficients of Z_h and $\bar{Z}_h$ in equation (6)[2] using the definition of M^{-1} in (5).

Taking the ratio π of the coefficients of z^k and $\bar{z}^k$, for each variable z^k in Z, we obtain:

$$\pi = \frac{1-\eta}{\eta} + \frac{1}{n-1} \tag{9}$$

Finally, a last method (Method D) considers equation (6) directly. Equation (6) is first estimated by OLS regressions calibrating n at predetermined values (n = 2, 5, 10). Then the corresponding η's are estimated by GLS.

To summarize. Methods A and D (fixing n) estimate the interdependency parameter η. Methods B (without hypothesis H) and C link the interdependency parameter and the size of the

[1] $Wx_{\Re'_h} = x_{\Re'_h} - Bx_{\Re'_h} = x_{\Re'_h} - Bx_h$. As $x_{\Re'_h} = \frac{n}{n-1} Bx_h - \frac{x_h}{n-1}$, one obtains: $Wx_{\Re'_h} = -\frac{1}{n-1}(x_h - Bx_h)$.

[2] Equation (6) writes $x_h = \frac{1-\lambda}{\mu} Z_h \beta - \frac{\lambda}{\mu}(n-1)\bar{Z}_{\Re'_h}\beta + M^{-1}\varepsilon_h$ where $\lambda = \frac{\eta}{(1-\eta)(1-n)}$ and $\mu = 1 - \frac{\eta}{1-n}$. The ratio π of the coefficient of z_h^k and $\bar{z}_h^k$, for each variable z_h^k in Z_h, is equal to $\frac{1-\eta}{\eta} + \frac{1}{n-1}$. With 2 k estimated coefficients, the 3 parameters λ, μ, n are over-identified.

reference group that permits an estimate of n by fixing η at a predetermined value. Method B (with hypothesis H) gives a direct estimate of η.

4. An empirical application with French data

Consider the following specification for equation (3):

$$\ln(y_{\min h}) = \beta_1 \ln(y_h) + Z_h \beta_2 + \eta \sum_{h' \in \Re'_h} \alpha_{h'} \ln(y_{\min h'}) + \varepsilon_h , \qquad (10)$$

where $y_{\min h}$ is the "minimum income declared necessary for household h to make ends meet". y_h is the income of h. The interdependence effect is captured with the subjective minimum income of individuals in the reference group. Z_h represents other explanatory variables. To define a subjective poverty line, Van Praag et al. (1982) use a lognormal indirect utility of income to obtain a double-log specification similar to equation (10). β_1, the income elasticity of the poverty line, provides an estimate of what minimum income is required to maintain a given level of utility as the income of h increases. Gardes and Loisy (1997) interpret this elasticity as an index of the pressure of needs.

Equation (11) is the corresponding reduced form of equation (6):

$$1n(y_{\min h}) = \theta' \beta_1 1n(y_h) - \lambda.(1-n).\beta_1 1n(y_{\Re'_h}) + \theta' Z_h \beta_2 - \nu.(1-n) Z_{\Re'_h} \beta_2 + \theta' \varepsilon_h - \lambda \varepsilon_{\Re'_h} \qquad (11)$$

$1n(y_{\Re'_h})$ and $Z_{\Re'_h}$ are averages of the variables $1n(y_h)$ and Z_h over $\Re'_h$. $\theta' = \theta^{-1} = [1-(\eta/(1-\eta)(1-n))]/[1-\eta/(1-n)]$ and $\nu = [\eta/(1-\eta)(1-n)]/[1-\eta/(1-n)]$. The residuals heteroskedasticity and spatial autocorrelation depend on $\varepsilon_{\Re'_h}$.

The restrictions embedded in equation (11) on the parameters of the socio-economic characteristics Z for household h and its reference group can be easily tested. Using a simple

OLS on equation (11), the Fisher test for the constraints is 2.20, a value that accepts the restriction at 1%.[3]

The regressions are run with the 1995 French Insee Family Expenditures Survey. The reference populations must be defined by a priori exogenous criteria to *instrument* the consumption of the influencing persons in $\Re_h$ that are not observed directly. They are defined in this data-set by 3 age groups of the household head, 7 level of education and 3 family types.[4]

The interaction coefficient η and the size of the reference group n are first estimated using methods A and B. With method A, η is estimated by iterative convergence using equation (10) on individual data and on the reference populations. In column 1 of Table 1, the results are presented after 10 iterations with starting values of $\overline{\eta} = 0.1$ and 0.25. At convergence, the estimate of the interaction parameter is $\hat{\eta}$ = 0.221 (0.0026). For the income elasticity: $\hat{\beta}$ = 0.661 (0.013). Column 2 of Table 1 refers to method B in the general case described in Section 3 (see equations (7) and (8)). The corresponding ratio of between and within estimates for the constrained coefficients of variables in Z^2 that are excluded for the groupings of the reference populations (specifically the variables log of income, number of workers in the family, number of unemployed) is estimated at $\hat{\beta}_B / \hat{\beta}_W$ = 1.387. With equation (8) and $\hat{\eta} = 0.221$, estimated by method A, we obtain $\hat{n}_h = 3.73\,(0.77)$.[5] Column 3 of Table 1 is the results of method B under the assumption H. We obtain by direct comparison of the within and between estimates of the coefficients for variables in Z_h^2, $\hat{\eta} = 0.279\,(0.34)$. With method C to calibrate η, and by comparing the coefficients of Z_h and $Z_{\Re'_h}$, we derive (see equation (9)): $\hat{\pi} = (1-\eta)/\eta + 1/(n-1) = 2.937$ and therefore $\hat{n} = 3.83\,(0.54)$.[6]

The different estimates of η and n are similar, suggesting that: (i) the interaction coefficient is significantly positive with a value of around 0.25; (ii) the reference groups are quite

---

[3] Those restrictions can be tested for all variables that are not used to define the reference group. F(6,8833) = 2.80.

[4] The same criteria were also used to instrument income in the regressions. Therefore the usual test of the validity of the instrument has permitted to test also the exogeneity of the definition of the reference populations.

[5] $1.387 = \frac{1}{1-\eta}\left[1 + \frac{\eta}{n-1}\right]$.

[6] To use the estimate obtained under hypothesis (H) to fix η in the general case is in itself contradictory.

narrow and contain, for the basic needs indicated by the minimum income question, only two or three persons influencing each individual.

[Insert Table 1, about here]

In Table 2, we present the results of the estimations of equations (10) and (11) under various conditions. The first two columns are coefficients estimates where we ignore the econometric problems associated with the interaction effect. In the 1st column, we simply set $\eta = 0$ and obtain an income elasticity estimate of 0.729 (0.013). In column 2, we directly estimate the interaction coefficient by OLS: the value of 0.312 (0.024) is higher than the estimates discussed earlier. Columns 3 to 5 relate to equation (11). They are GLS regressions calibrating the size of the reference group at 2, 5 and 10. All of the estimation results show that the minimum income is indexed on the actual income of the households, with an income elasticity of the poverty line around 0.6. Thus, any increase in the household's income increases its needs by more than half. Estimates of η, the interdependency effect, are always statistically significant and positive with values between 0.26 and 0.35. These values are consistent with those obtained previously. Accounting for the interdependency significantly changes the income elasticity of the poverty line by about 20%. Part of the influence of income changes on the poverty line acts through the general increase of income for the reference groups: noting $g_{\Re'_h}$ the change in income for the reference population and g_h the change in income for the household h, the income elasticity amounts to $\beta_1 = 0.6$ for $g_h > 0$ and $g_{\Re'_h} = 0$, and $\beta_1/(1-\eta) = 0.85$ for $g_h = g_{\Re'_h} > 0$. Note that the individual elasticity β_1 varies from one estimation to the other, while the total elasticity, encompassing the individual income change and the interdependent effect, is more stable.

[Insert Table 2, about here]

The bias of the estimation problems related to specification (1) with regards to the importance of the interdependency effect can be measured, either by the parameter θ (which reflects the bias due to the omission of variables $Z_{\Re'_h}$) or by comparing the estimation of β in (10) with the parameters of equation (11). The average over-estimation θ of the income elasticity of minimum income (and for the other parameters of the explanatory variables) is 11% for $n = 2$. It is consistent with the bias obtained directly as an estimated parameter of equation (11). The over-estimation diminishes when the reference group contains more households: it is only 1% for $n = 10$. This result may indicate that the distance between households pertaining to the same reference group increases with the size of this group, thus diminishing the interaction effects between them (by a greater factor than the increase of the size, which multiplies the interactions among a greater number of agents). The ratios $\beta(10, \eta = 0) / \beta(11)$ also indicate a significant over-estimation, generally higher than with θ because of the misspecification of equation 10.

5. Conclusion

Interdependent preferences within a reference group are a simple specification of social interactions, which is generally well supported by the data. In this paper, we show, along with many authors, how ignoring this issue could introduce important biases in the coefficients of regressions variables. We present a simple approach to the interdependency question and we propose various methods of estimation. One originality of our research is that it approximates the size of the reference group, a major element in the importance of estimation biases and a question largely ignored in the literature. Applied to a sample of French households, a 10% bias in the regression parameters of a reported subjective poverty model is found when one ignores the interdependency question. We also show that the overall influence of income changes on the poverty line, through the income variations of the individual and of its reference group, is about 0.85 relatively to 0.60 for the income variations of the individual. Finally, the reference group involves few persons: our various estimates suggest a reference group of about 2 to at most 5 persons, a figure consistent with Hayek's point of view on the subject.

Table 1

Parameters estimates of the subjective poverty model

Method of estimation	A	B (general case)	B (assumption H)***
Equation	(10)*	(7) and (8)	(7)
Interaction η	0.221 (0.003)	Fixed at 0.221	0.279 (0.034)
Size n		3.73 (0.77)	3.83 (0.54)
Income elasticity β_1	0.61 (0.013)		
π			2.937 (0.334)
β_B / β_W		1.387 (0.034**)	

Explanatory variables: instrumented income per unit of consumption, log of the number of units of consumption and its square, logarithmic age of the head, number of employed and number of unemployed.

Reference populations defined by 3 age groups, 7 education levels and 3 family types.

Filtering: The sample is screened by deleting all households whose relative position in the income and total expenditure distributions differs by more than 50 centils and for whom the minimum income declared is greater by 30% than its own income. Overall this excluded about 6% of the initial sample.

Degrees of freedom: 8834 for equation (10).

*10 iterations.** The standard error is approximated by the delta method. It must be taken with caution as the ratio of β's follows a Cauchy distribution of a theoretically unknown variance.*** Interaction coefficient η estimated by method C.

Source: 1995 Insee Family Expenditures Survey. See Gardes-Loisy, 1997, for a description of the data.

Table 2
Additional coefficients estimates and comparisons

Equation	$(10, \eta = 0)$	(10)	(11)	(11)	(11)
	OLS	OLS	GLS	GLS	GLS
Size of the Reference Group: n	In the data-set	In the data-set	2	5	10
Individual Log income: β_1	0.729 (.013)	0.633 (.015)	0.573 (.033)	0.636 (.017)	0.645 (.016)
η	0	0.312 (.024)	0.350 (.053)	0.272 (.032)	0.261 (.029)
θ	-	-	1.140 (.049)	1.024 (.006)	1.010 (.003)
$\beta_1 /(1-\eta)$	-	0.922	0.881	0.874	0.873
$\beta(10, \eta = 0) / \beta(11)$	-	-	1.27	1.15	1.13
$\beta(10) / \beta(11)$	-	-	1.11	1.00	0.98
R^2 or pseudo	0.4026	0.4141	0.3291	0.3444	0.3467

References

Aronsson, R, S. Blomquist and H. Sacklén. 1999. "Identifying Interdependent Behaviour in an Empirical Model of Labour Supply", *Journal of Applied Econometrics*, 14, 607-626.

Duesenberry, J.S. 1949. *Income, Saving and the Theory of Consumer Behavior*, Harvard University Press, Cambridge, M.A.

Gardes, F., and C. Loisy. 1997. "La Pauvreté selon les Ménages: une Evaluation Subjective et Indexée sur leur Revenu", *Economie et Statistique*, n° 308-309-310, 95-112.

Gaviria A. and S. Raphael. 2001. "School-Based Peer Effect and Juvenile Behavior", *The Review of Economics and Statistics,* 83(2), 257-268.

Hayek, F. von. 1948. *Individualism and Economic Order*, The University of Chicago Press.

Hayek, F. von. 1952. *Scientism and the Study of Society*, The Free Press, Glencoe, Illinois.

Kapteyn, A., S. Van de Geer, H. Van de Stadt, and T. Wansbeek. 1997. "Interdependent Preferences: an Econometric Analysis", *Journal of Applied Econometrics*, 12, 665-686.

Manski, C.F., 1993, "Identification of endogenous Social Effects: the Reflection Problem", *Review of Economic Studies*, 60, 531-542.

Pollak, R.A., and T.J. Wales. 992. *Demand System Specification and Estimation*, Oxford University Press, Oxford.

Van Praag, B.M.S., A,J.M. Hagenaars and H. Van de Stadt. 1982. "Poverty in Europe", *The Review of Income and Wealth*, 28(3), 345-359.

Appendix

Inversion of matrix M

Consider the (n,n) matrices $M^{-1} = (m_{ij})$, and **1** = (1). From the text $M = I_n + \eta/(1-n)$. (**1** $-I_n$).

Note that the matrix (m_i) with columns formed by the sums $m_{i.} = \sum_j m_{ij}$, writes M^{-1}.**1**. Also $(m_{.j}) =$ **1**.M^{-1} is the matrix with columns of $m_{.j} = \sum_i m_{ij}$, and **1**.M^{-1}.**1** = $\sum_{ij} m_{ij} = n_h^2(m)$ with m the average of m_{ij}. Therefore,

$$I_n = MM^{-1} = (m_{ij}) + [\eta/(1-n)].(m_{.j}) - [\eta/(1-n)]\cdot(m_{ij}). \quad \text{(A1)}$$

By pre-multiplying the first equality by **1**, we obtain:

$$\mathbf{1} = \mathbf{1}.MM^{-1} = \mathbf{1}.M^{-1} + \mathbf{1}.(\mathbf{1}.M^{-1})\eta/(1-n) - \mathbf{1}.M^{-1}\eta/(1-n)$$

$$= (m_{.j}).\left[1 + \frac{\eta}{1-n}(n-1)\right] = (1-\eta).(m_{.j}), \text{ as } \mathbf{1}.\mathbf{1} = n.\mathbf{1}.$$

This implies, $(m_{.j}) = \left(\frac{1}{1-\eta}\right)$.**1**. Post-multiplying $.MM^{-1}$ by **1** gives $(m_{i.}) = \left(\frac{1}{1-\eta}\right)$. **1**. With M symmetric with identical triangular coefficients, M^{-1} is also symmetric. Thus,

$$m_{i.} = m_{.j} = [1/(1-\eta)]. \quad \text{(A2)}$$

Using (A2), equation (A1) writes: $M^{-1} = \{I_n - [\eta/((1-\eta)(1-n))]\cdot\mathbf{1}\}/(1-\eta/(1-n))$.

With η fixed, the variance-covariance matrix is:

$$V(M^{-1}\varepsilon) = E\{(M^{-1}\varepsilon)(\varepsilon' M^{-1})\} = E(\varepsilon\varepsilon')M^{-2}.$$

Therefore, $M^{-2} = \delta^2(I_n - \gamma.\mathbf{1})\ (I_n - \gamma.\mathbf{1}) = \delta^2(I_n + \gamma(n\gamma - 2).\mathbf{1})$ with $\delta = 1/(1-\eta/(1-n))$ and $\gamma = \eta/((1-\eta)(1-n))$, so that the transformed error is heteroskedastic and autocorrelated between consumers.