A semi-parametric test of relative income effects

François Gardes

CREST-LSM, Ensai, Campus de Ker Lann, 35170, Bruz, France (e-mail :gardes@univ-paris1.fr). This version, March 2002. Special thanks are due to Greg Duncan, Patrice Gaubert, B. Górescki and Christophe Starzec for making available the data for the U.S. and Poland. I aknowledge comments and suggestions from Franck Arnaud, Andrew Clark, Marc Diaye, David Margolis, Philippe Merrigan, Claude Montmarquette, Christophe Starzec and participants to seminars in Cirano (Montréal), Crest-Ensai, Universities of Caen and Lausanne. The paper was written while the author was invited at the University of Montréal.

JEL: C14, D12, D31. Key words: relative income, household, panel

A semi-parametric test of relative income effects[1]

Abstract

This test of the Duesenberry hypothesis on social interactions is based on the aggregation of the endogeneity bias of cross-section estimates. It involves no restriction on the specification of the relative income effect. An application on US and Polish panels confirms the existence of Duesenberry's demonstration effect.

1. Introduction

The relative income hypothesis proposed by Duesenberry (1948) has been relatively neglected in the empirical literature, perhaps because it was primarily used for the macroeconomic consumption analysis, and finally replaced by the Permanent Income Hypothesis. Some recent articles take up this hypothesis explaining its

[1] Special thanks are due to Greg Duncan, Patrice Gaubert, B. Górescki and Christophe Starzec for making available the data for the U.S. and Poland. I aknowledge comments and suggestions from Franck Arnaud, Andrew Clark, Marc Diaye, David Margolis, Philippe Merrigan, Claude Montmarquette, Christophe Starzec and participants to seminars in Cirano (Montréal), Crest-Ensai, Universities of Caen and Lausanne. The paper was written while the author was invited at the University of Montréal.

JEL: C14, D12, D31. Key words: relative income, household, panel

theoretical foundations (Bagwell-Bernheim, 1996; Dybvig, 1995). In recent research (Clark-Oswald, 1996; Kapteyn et al., 1997), this type of interdependence of preferences is generally proved to be consistent with the empirical evidence (though serious interpretation problems arise, as shown by Manski in his discussion of the reflection problem, Manski, 1995). It deserves more thorough appreciation from the micro point of view, for it builds a bridge, as a Veblen effect, between sociological and economic explanations of households behavior.

Parametric tests of the relative income model can be criticized because of the hypotheses necessary to define relative income and specify consumption functions. For instance, Duesenberry, in his macro applications, defines relative income as the largest past income observed by the households, while it is measured, in his micro analysis, as the relative position of households within a given population. No doubt that the first definition is quite remote from the original micro concept, while the second, sociological, definition requires indicating the reference population by objective characteristics (if no subjective information from the household is available), and specifying the relative income effect on consumption.

Section 2 discusses the testing of relative income effects, section 3 describes the test and section 4 applies it on the U.S. Panel Study of Income Dynamics and on a Polish panel.

2. Framework:

Suppose the relative income position of the household within its reference population can be measured by (y_h-my_h), with y_h household h's income (per consumption unit) and my_h the average income in its reference population. A relative income effect can be revealed by relating this relative income position to the similar residual of the household consumption over the average consumption of the reference population: c_h-mc_h.

Table 1. Correlation between food expenditures and relative income

Survey	1987	1988	1989	1990	average
my	0.496	0.495	0.444	0.668	0.526
	(0.036)	(0.039)	(0.034)	(0.039)	(0.019)
ys	0.533	0.499	0.443	0.551	0.506
	(0.009)	(0.009)	(0.009)	(0.009)	(0.005)

Data-set: Polish Panel; 52 reference populations have been defined according to five cohorts, three education levels of the head and four quarters (some populations have been grouped to obtain a sufficient size).
logarithmic specification with average total expenditure for the reference population: my and the residual total expenditure for household h: $ys_h=y_h-my$. *Control variables*: proportion of children, log of head's age, location, education level, relative prices, 16 quarter dummies as proxies of macroeconomic shocks (similar results for linear AI Demand System specification).

Such an estimation is presented in Table 1, for the Polish surveys. The elasticity between residual income and consumption is not systematically smaller than the elasticity as concerns average income, which contradicts the observation that (relative) poor households may spend more money for food that the (relative) rich. By the way, the income elasticities as concerns the average income for the reference population and the residual income, are not significantly different.

Moreover, this correlation does not truly indicate a relative income effect because the residual consumption may be influenced by all the determinants of the individual consumption which are not used as criteria to define the reference populations or as explanatory variables, because they are absent in the data-set. These latent variables may be also correlated to the endogenous household income, thus creating an indirect relation between consumption and income. Thus, it seems necessary to take into account, by panel estimation, the specific component of consumption which cannot be explained by the explanatory variables present in the data-set, but which persists over time and characterize the household consumption.

Indeed, when estimating a consumption function on panel data, the specific consumption of some household h can be identified as the permanent component of the residual. The correlation between the permanent part of the individual error and the relative income position of the household within its reference population, indicates Duesenberry effects. However, it is well known that the computation of such individual effect is difficult. Thus, we propose to estimate directly the endogeneity bias on panel data due to the relative income effect. On grouped data, the remaining correlation between the specific effect and income variables can no longer represent Duesenberry effects since all relative positions within reference populations have been cancelled by aggregation. Thus, the difference between the estimation on individual data and the estimation on grouped data is the part of the correlation which disappears when the data is grouped into cells corresponding to the reference populations. This estimation procedure clearly shows the role of aggregation in reducing this endogeneity bias and what statistical hypotheses are necessary to estimate the relative income effect. Moreover, it gives an indicator of the importance of the Duesenberry effect (for instance the ratio between the bias obtained on individual or grouped data) which may be used to compare different groupings and to define the reference populations (the vignettes to which the individuals compare themselves) and the relative impact of each consumption on this definition.

3. Theory

Consider a model of consumption for individual h at time t:

$$z_{ht} = X_{ht}\,\beta + u_{ht} \text{ with } u_{ht} = \alpha_h + \varepsilon_{ht}. \qquad (1)$$

When estimating with panel data, Mundlak (1978) shows that between estimates are biased if the specific effects α_h are correlated with the explanatory variables: $E(\alpha_h|X_{ht}) \neq 0$. The specific effect α_h may be related to the between form of the explanatory variables, such that:

$$\alpha_h = BX_{ht}\pi + \zeta_h \text{ with } \zeta_h \sim N(0,\sigma_\zeta^2), \text{ i.i.d., } E(\zeta_h, BX_{ht}) = 0. \qquad (2)$$

Thus, the model can be written in between form (i.e. for the average over periods):

$$Bz_{ht} = BX_{ht}\,\beta + BX_{ht}\,\pi + \zeta_h + B\varepsilon_{ht} \qquad (3)$$

which implies for the between estimates of Bz_{ht} over BX_{ht}:

$$E(\beta_b|X) = E[(X_{ht}'BX_{ht})^{-1}\; X_{ht}'Bz_{ht}] = \beta + \pi$$

while the within estimate (computed on the difference between equation (1) and equation (3)) is unbiased: $E(\beta_w|X) = \beta$ as the within operator suppresses all specific effects, and therefore the endogeneity biases caused by the correlation between these effects and the explanatory variables. Such endogeneity biases are shown to exist in the estimation of consumption functions for at least half of the commodities (Gardes et al., 1996, 2002).

Aggregating data can be operated to correct for measurement errors, or to build pseudo-panel data in order to estimate dynamic models when only separate surveys are available. Such grouping may change the endogeneity biases, but in a way which is difficult to predict by considering directly the aggregation of (1), (2), (3). In order to analyze how aggregation affects the endogeneity bias, we propose to split the individual effect α_h between the collective specific effect (common to all individuals in the sub-population) and the residual effect specific to the individual.

Suppose that the estimation is performed on a population H of individuals h = 1 to N surveyed within the whole population $\mathscr{H}$ ($H \subset \mathscr{H}$). Sub-populations are defined by crossing characteristics k_j, j=1 to J such that: $\mathscr{H}_i = \{h \in \mathscr{H} / k_j(h)=c_j(i) \text{ for all } j\}$ with $c_j(i)$ taking all possible items or values for characteristics k_j. H_i is thus defined as $\mathscr{H}_i \cap H$[2].

Suppose that the first explanatory variable is the logarithmic individual income y_h. We make the usual hypotheses on the distributions of income and specific effects for individuals: (H1) $h \in H_i \Rightarrow y_h \sim N(y_i,\sigma^2_{yi})$ and $\alpha_h \sim N(\mu_i,\sigma^2_{\alpha i})$, i.i.d., with $y_i = E(y_h|h \in H_i)$, $\mu_i = E(\alpha_h|h \in H_i) < \infty$.

[2] We suppose this grouping according to a-priori exogenous criteria (age and education) is exogenous to the household consumption.

The average y_i in H_i is computed by regressing y_h on the vector of characteristics K: $y_{Hi} = Ka_i + \xi_i$ so that $y_{Hi} = 1/n_i (\Sigma_{h\in Hi}\ y_h)$ with n_i the number of individuals in H_i. So the distribution of the mean is: $y_{Hi} \sim N(y_i,\sigma^2_{yi}/n_i)$. We define the *specific income* (which may be considered as the relative income[3] of individual h in its reference population $\mathcal{H}_i$) as $ys_h = y_h - y_{Hi}$ so that $ys_h \sim N(0, \sigma^2_{yi} - \sigma^2_{yi}/n_i)$.

By the same reasoning, $\mu_{Hi} = 1/n_i (\Sigma_{h\in Hi}\ \alpha_h)$ and $\mu_{Hi} \sim N(\mu_i,\sigma^2_{\alpha i}/n_i)$.

Consider now the decomposition of the specific effect into the specific effect of the reference population H_i and an individual effect: $\alpha_h = \mu_{Hi}+\nu_h$. We obtain the distribution for ν as:

$$\nu_h \sim N(0,\sigma^2_{\mu i}(1-1/n_i)).$$

The covariance on individual data between α_h and some explanatory variable y (here log-income or total expenditures) can be decomposed into the reference population components and the true individual components:

$$A=E\{(y_h-Ey).(\alpha_h-E\alpha)\} = E\{ [(y_{Hi}-y_i) + (y_i-y) + ys_h].[(\mu_{Hi}-\mu_i) + \mu_i + \nu_h]\}$$

This expression is shown in Appendix I to reduce asymptotically to the sum of two of the nine terms of its decomposition, so that

[3] Note that ys_h corresponds to the log-ratio of household's income and average income y_{Hi} if incomes are defined in logarithm.

$$\frac{A}{V(y)} = \pi = (\beta_b - \beta_w)_{panel} = p\ (\beta_b - \beta_w)_{\text{grouped data}} + (1-p)\ \gamma$$

where $p = \frac{V(y_{Hi})}{V(y)}$ and γ is the coefficient resulting from the correlation between the specific effect ν_h of individual h and her specific (relative) income[4]. Thus, this coefficient γ and its standard error can be computed in terms of the difference between the estimates of β on individual and grouped data in the between and within dimensions:

$$\gamma\ (\nu_h/ys_h) = \frac{1}{1-p}\ \{(\beta_b - \beta_w)_{panel} - p.(\beta_b - \beta_w)_{\text{grouped data}}\}$$

4. Empirical application

Since 1968, the Panel Study of Income Dynamics has followed and interviewed annually a national sample that began with about 5,000 U.S. families. We use only four years (1984-1987) in the estimation of the consumption equation to be comparable with the Polish data. In all cases the data are restricted to households in which the head did not change over the six-year period and to households with major imputations on neither food expenditure nor income variables (in terms of the PSID's

[4] For p=1, each cell contains only one household, so that the panel and the pseudo-panel coincide. For p=0, all the population is grouped into one cell, and ys is the difference between y and its average on the whole population, so that γ just indicates the endogeneity bias estimated on the panel.

"Accuracy" imputation flags, we excluded cases with codes of 2 for income measures and 1 or 2 for food at home and food away from home measures).

In order to construct cohorts for the pseudo-panels, we defined a series of variables based on the age and education levels of the household head. Specifically, we define : i) 6 cohorts of age of household head: under 30 years old, 30-39, 40-49, 50-59, 60-69, and over 69 years old; and ii) three levels of education of household head: did not complete high school (12 grades), completed high school but no additional academic training, and completed at least some university-level schooling. The population is randomly divided into four sub-samples, each of which is used to aggregate data for the different years. This prevents the same household from being included in the same cell in more than one period. The PSID cells sizes vary from 9 to 183 households with a mean of 65.5 (see Gardes et al., 2002, for details).

The annual Polish expenditure surveys contain about 30 thousand households, which represent approximately 0.3% of all households in Poland. On every annual sub-sample between 1987 and 1990, it is possible to identify 3707 households participating in the surveys during all four years and interviewed in the same quarter for each year for their expenditures, income and various socio-economic variables. The period covered by the Polish panel is unusual even in Polish economic history. It represents the shift from a centrally planned, rationed economy (1987) to a relatively unconstrained fully liberal market economy (1990). Real GDP grew by 4.1% between 1987 and 1988, but fell by 0.2% between 1988 and 1989 and by 11.6% between 1989 and 1990. Price increases across these pairs of years were 60.2%, 251.1% and 585.7%, respectively. Thus, the transitory years 1988 and 1989 produced a period of a very high inflation and a mixture of free-market, shadow and administrated economy (see Gardes et al., 2002, for a presentation of the data-set and the estimation procedures, and Lednicki, 1982, Górecki, 1992, for a full description of the master sample generating procedure). The pseudo-panel is built on the whole surveys from 1987 to 1990. It contains 224 cells with 107 households per cell in average.

Identification for the estimation on Between transformed data needs more cells than the number of regressors: 18 cells are used for the PSID and 224 for the Polish panel (for four years), compared to 9 and 35 explanatory variables. The precision of the estimators depend on the number of cells, but the errors of measurement (due to the fact that the cells contain different households for two periods) decrease with the cells size, so that a trade-off exists between numerous cells in the pseudo-panel, but with possible errors of measurement, and a small number of great cells without much error of measurement[5]. A priori, the estimators may be more efficient in the Polish case, with a great number of numerous cells. The specification uses the linear Almost Ideal Demand System for the PSID. For the Polish panel, we dispose of prices for four professional groups and each quarter and year, so that it is possible to estimate the price elasticities and the integrability coefficient of the quadratic system (see Banks et al., 1995, for the estimation method by convergence on the integrability coefficient).

[5] Also the use of limits in probability in the proof (Appendix I) requires numerous cells.

Table 2. Income elasticities and relative income effects

		Food at Home		Food Away
		Polish Panel	PSID	PSID
Individual data[a]	Between	0.494	0.186	1.050
		(0.012)	(0.027)	(0.142)
	Within	0.755	0.507	0.443
		(0.012)	(0.112)	(0.205)
Pseudo-Panel data[b]	Between	0.452	0.311	1.387
		(0.022)	(0.045)	(0.068)
	Within	0.542	0.240	0.800
		(0.023)	(0.095)	(0.150)
$(\beta_B - \beta_w)$:	Ind. Data	-0.261	-0.321	0.607
		(288.10^{-6})	(0.0132)	(0.0438)
	Pseudo-Panel	-0.090	0.071	0.587
		(0.00101)	(0.0111)	(0.0271)
$p = \frac{V(y_{Hi})}{V(y)}$		0.3294	0.2731	0.2731
Budget Share		0.508	0.138	0.0328
γ		-0.1753	-0.0646	0.0202
		(0.0151)	(0.0097)	(0.0137)
Student for γ		11.60	2.09	1.36
Relative income Elasticity		0.655	0.532	1.616
		(0.030)	(0.070)	(0.418)

[a] panel data (3630 households for the Polish panel, 2430 households for the PSID).

[b] *Polish Pseudo-panel*: 224 cells according to six cohorts, three education levels of the head, location and four quarters (some cells are grouped to obtain a sufficient cell size).

PSID Pseudo-panel: 18 cells according to six cohorts and three education levels of the head.

Specifications: PSID: Almost Ideal Demand System specification with instrumented income and 8 control variables: equivalence scale and its square, log of head's age and its square, 4 survey dummies as proxies of relative prices and macroeconomic shocks.

Polish surveys: QAIDS specification with instrumented total expenditure and its square and 33 control variables: proportion of children, log of head's age, location, education level, relative prices, 16 quarter dummies as proxies of macroeconomic shocks; estimation by convergence on the integrability parameter.

Results:

(i) $\gamma_{v/ys}$ are negative and significant for food at home, both in US and Poland. It indicates a *negative Duesenberry effect on food consumption*: relative poor in their reference population (i.e. having a negative specific income ys_h) have a greater food budget share than relative rich-belonging to another reference population-which have the same total income and similar control variables. Note that the relative income elasticity for food at home is similar in both countries, contrary to the income elasticities which are much greater, as expected, for Polish consumers.

(ii) For food away, $\gamma_{v/ys}$ is significantly positive, which indicates that relative rich households have a greater budget share of food away from home than the relative poor.

(iii)

The relative income elasticities for food at home are greater than the between, which indicates that these two types of elasticity do not measure exactly the same effects: the cross-section effects of incomes differences contain indeed relative income effects, but also the influence of long term changes in the average income of the reference populations which may be recovered by comparing relative income coefficients and the total cross-section coefficients.

(iv) The estimation of the relative income effect is much better for the Polish panel, as was expected for this much larger data-set.

Conclusion

The estimations confirm the existence of a demonstration effect for food in both countries and for the reference populations defined by age cohorts, education and location. The effect is negative for food at home and positive for food away from home, which corresponds to the predictions. A more important effect could be obtained by defining more precisely the reference populations. Moreover, another application of these results could be to better distinguish homogenous groups for clustering. The magnitude of the correlation coefficients $\gamma_{v/ys}$ between the residual of the explained and exogenous variables, may serve as indicators for the homogeneity of the reference populations, thus giving a criteria to define them.[6].

The test is said to be semi-parametric because the specific household consumption is computed as a residual by panel analysis, thus taking indirectly into account the influence of all latent variables which are constant through time. On the other hand, relative income is predicted by the criteria used to define the reference

[6] For instance, households being grouped into sub-populations according to some exogenous criteria, their specific expenditures for a set of commodities and the correlation coefficients $\gamma_{v/ys}$ can be recovered by panel analysis. These specific residuals being in a second step used to optimally group the households (so that the specific effect is homogenous within each cell and can be removed efficiently by a pseudo-panel estimation), the correlation coefficients $\gamma_{v/ys}$ estimated in the second step are used to compare the homogeneity of this second grouping to the first.

populations, so the method is parametric as concerns this definition[7]. On the whole, the test is much more general than usual parametric tests and based on independent definitions of relative income and consumption.

References:

Bagwell, L.S., Bernheim, B.D. "Veblen Effects in a Theory of Conspicuous Consumption", *American Economic Review*, June 1996, 86 (3), pp. 349-373.

Banks, J., Blundell, R. and Lewbel, A., "Quadratic Engel Curves and Consumer Demand", *The Review of Economics and Statistics*, November 1997, 79(4), pp. 527-539.

Clark, A.E., Oswald, A.J., "Satisfaction and Comparison Income", *Journal of Public Economics*, September 1996, 61(3), pp. 359-381.

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

Dybvig, P.H., "Duesenberry's Ratcheting of Consumption: Optimal Dynamic Consumption and Investment given Intolerance for any Decline in Standard of Living", *Review of Economic Studies*, April 1995, vol. 62(2), pp. 287-313.

Gardes, F., Langlois, S., Richaudeau, D., "Cross-Section and Time-Series Income Elasticities", *Economics Letters*, May 1996, 51(2), pp. 169-175.

Gardes, F., Duncan, G.J., Gaubert, P., Starzec, C., "A Comparison of Consumption Models Estimated on American and Polish Panel and Pseudo-Panel data", *Cahiers de Recherche du CREST*, 2002.

[7] Income would be also predicted by panel analysis, using life cycle models, but this would require

Górecki, B., Peczkowski, M., ”Polish Household Panel”, Department of Economics, Warsaw University, 1992.

Kapteyn, A., Van de Geer, S., Van de Stadt, H., Wansbeek, T., “Interdependent Preferences: an Econometric Analysis”, *Journal of Applied Econometrics*, December 1997, 12(6), pp. 665-686.

Lednicki, B., “Dobor Proby i Metoda Estimacji w Rotacyjnym Badaniu Gospodarstw Domowych (Sample Design and Method Estimation in Rotation Survey of Households)”, GUS, Warsaw, 1982.

Manski, C.F., *Identification Problems in the Social Sciences*, Harvard University Press, Cambridge, MA, 1995.

Mundlak, Y., “On the Pooling of Time Series and Cross Section data, *Econometrica*, January 1978, 46(1), pp.69-85.

---

much longer series than the two panels.

Appendix I : Decomposition of the endogeneity bias

We examine the limits in probability, when $H_i \uparrow \mathscr{H}$, of the nine cross-products of the endogeneity bias for individual data:

$$A=E\{(y_h-Ey).(\alpha_h-E\alpha)\}=E\{((y_{Hi}-Ey_i)+(Ey_i-Ey)+ys_h).((\mu_{Hi}-\mu_i)+\mu_i+\nu_h))\}.$$

Supposing as usual that all means and variances are bounded, the limits in probability are equivalent to the limits in mean square, so the proof is made for limits in probability.

Note that plim $y_{Hi}=Ey_i= y_i$,plim $\mu_{Hi}=\mu_i$, plim $(Ey_i-Ey)=Ey_i-Ey,=y_i-y$,

$H_i \uparrow \mathscr{H}$

plim $p_{Hi}=p_i$=proportion of the sub-population $\mathscr{H}$ in the whole population $\mathscr{H}$ and plim$(p_h/p_{Hi})=1/N_i$ with N_i=Card($\mathscr{H}$).

(i) plim $(\Sigma_{h\in H}\, p_h.1_{Hi}.(y_{Hi}-Ey_i).(\mu_{Hi}-\mu_i))$=plim $(\Sigma_i\Sigma_{h\in Hi}\, p_h.(y_{Hi}-Ey_i).(\mu_{Hi}-\mu_i))$

$=\Sigma_i$plim p_{Hi}.plim $(y_{Hi}-Ey_i)$.plim $(\mu_{Hi}-\mu_i)$ =0.

(ii) plim $(\Sigma_{h\in H}\, p_h.(y_{Hi}-Ey_i).\mu_i)=\Sigma_i p_i$.plim $((y_{Hi}-Ey_i).\mu_i))\leq(\text{Sup}_i\mu_i)\Sigma_i p_i$.plim $(y_{Hi}-Ey_i)$=0.

(iii) plim $(\Sigma_{h\in H}\, p_h.(y_{Hi}-Ey_i).\nu_h)=\Sigma_i\, p_i$ plim $(y_{Hi}-Ey_i).\Sigma_{h\in Hi}$ plim $(p_h/p_{Hi})\nu_h$

$=\Sigma_i p_i.\text{plim}\ (y_{Hi}\text{-E}\ y_i).(\ 1/N_i)\Sigma_{h\in Hi}\ \nu_h=0$ as $p_h/p_{Hi}=1/N_i$ and $\Sigma_{h\in Hi}\nu_h=0$.

(iv) $\text{plim}\ (\Sigma_{h\in H}\ p_h.(Ey_i\text{-}Ey).(\mu_{Hi}\ \text{-}\ \mu_i))=\Sigma_i\ p_{Hi}.\text{plim}\ (y_i\text{-}Ey).\text{plim}\ (\mu_{Hi}\text{-}\mu_i)=0$.

(v) $\text{plim}\ \Sigma_{h\in H}\ p_h.(Ey_i\text{-}Ey)\mu_i=\Sigma_i\ p_{Hi}.\text{plim}\ (Ey_i\text{-}Ey).\text{plim}\ \mu_i=\Sigma_i p_{Hi}.(Ey_i\text{-}Ey).\mu_i$

$=\Sigma_i p_{Hi}.(Ey_i\text{-}Ey).(\mu_i\text{-}\mu)$ = covariance on grouped data as $\Sigma_i p_{Hi}.Eyi \rightarrow Ey$.

$H_i \uparrow \mathscr{H}$

(vi) $\text{plim}\ \Sigma_{h\in H}\ p_h.(Ey_i\text{-}Ey).\nu_h=\Sigma_i p_{Hi}.\text{plim}\ (Ey_i\text{-}Ey).\Sigma_{h\in Hi}\ \text{plim}\ (p_h/p_{Hi}).\nu_h$

$=(\Sigma_i p_{Hi}.(Ey_i\text{-}Ey)).(1/N_i).\Sigma_{h\in Hi}\ \nu_h=0$ as $\Sigma_{h\in Hi}\nu_h=0$.

(vii) $\text{plim}\ \Sigma_{h\in H}\ p_h.ys_h.(\mu_{Hi}\text{-}\mu_i)=\Sigma_i\ \text{plim}\ (\mu_{Hi}\text{-}\mu_i).\Sigma_{h\in Hi}\ \text{plim}\ ((p_h/p_{Hi}).ys_h)$

$=\Sigma_i\ \text{plim}\ (\mu_{Hi}\text{-}\mu_i).(1/N_i).\Sigma_{h\in Hi}\ ys_h=0$.

(viii) $\text{plim}\ \Sigma_{h\in H}\ p_h.ys_h.\mu_i=\Sigma_i(p_i.\mu_i.\Sigma_{h\in Hi}\ \text{plim}\ ((p_h/p_{Hi}).ys_h))=\Sigma_i(p_i.\mu_i.\Sigma_{h\in Hi}\ (1/N_i).ys_h)$

$=\Sigma_i(p_i.\mu_i.(1/N_i).\Sigma_{h\in Hi}\ ys_h)=0$.

(ix) $\text{plim}\ \Sigma_{h\in H}\ p_h.ys_h.\nu_h$=covariance due to relative specific effects on the individual data.

Thus we obtain: A = (v)+(ix) and

$$\frac{A}{\text{V(y)}}=\frac{(v)}{V(y_{Hi})}\cdot\frac{\text{V}(\text{y}_{\text{Hi}})}{\text{V(y)}}+\frac{\text{V(y) - V}(\text{y}_{\text{Hi}})}{\text{V(y)}}\cdot\frac{(ix)}{V(ys)}$$

so that, according to (2):

$$\frac{A}{V(y)} = \pi = (\beta_b - \beta_w)_{panel} = p\,(\beta_b - \beta_w)_{grouped\ data} + (1-p)\,\gamma$$

with $p = V(y_{Hi}) / V(y)$ and $\gamma = (ix) / V(ys)$ = correlation coefficient of the specific effect ν_h in equation (1) over the specific (relative) income.

Appendix II : Description of the data sets.

Table 2.1: Means and standard deviations of variable used in the PSID analyses

	1983	1984		1985		1986		1987	
	Level	Level	Dif.	Level	Dif.	Level	Dif.	Level	Dif.
Budget share for food at home	.147 (.103)	.144 (.098)	-.003 (.084)	.129 (.095)	-.015 (.086)	.137 (.100)	.008 (.082)	.134 (.096)	-.003 (.081)
% with at-home share = 0	0.0	0.0	53.2	0.0	74.0	0.0	41.5	0.0	51.3
Budget share for food away from home	.033 (.040)	.034 (.038)	.001 (.034)	.031 (.038)	-.003 (.033)	.033 (.041)	.002 (.032)	.033 (.034)	.001 (.033)
% with away-from-home share =0	9.5	8.9	5.7	9.6	5.5	10.3	5.5	8.9	5.7
ln household income	9.9254 (.648)	9.9985 (.657)	.0731 (.280)	10.1714 (.716)	.1729 (.320)	10.1238 (.686)	-.0475 (.308)	10.1671 (.694)	.0432 (.299)
ln age Head	3.7044 (.377)	3.7306 (.368)	.0262 (.013)	3.7573 (.359)	.0267 (.013)	3.7801 (.351)	.0228 (.012)	3.8044 (.343)	.0242 (.012)
ln family size (Oxford scale)	.6741 (.404)	.6837 (.401)	.0096 (.162)	.6896 (.405)	.0060 (.168)	.6894 (.409)	-.0002 (.159)	.6912 (.410)	.0018 (.171)

<u>Table 2.2</u>: Mean and standard deviation in the Polish panel

	1987	1988	1989	1990
Budget share for	0.508	0.484	0.486	0.554
food at home	(.14)	(.15)	(.18)	(.15)
Budget share for	0.0006	0.0006	0.0005	0.0005
food away from home	(.02)	(.03)	(.02)	(.03)
% with away-from	28.4	29.7	26.7	20.5
home share > 0				
Ln household	10.65	11.17	12.25	14.14
Expenditure	(.45)	(.49)	(.79)	(.50)
Ln head's age	3.789	3.809	3.824	3.842
	(.33)	(.32)	(.32)	(.32)
Ln family size	1.140	1.121	1.095	1.081
	(.59)	(.60)	(.61)	(.61)