INSTITUT NATIONAL DE LA STATISTIQUE ET DES ETUDES ECONOMIQUES
Série des Documents de Travail du CREST
(Centre de Recherche en Economie et Statistique)

n° 2002-02

Panel and Pseudo-Panel Estimation of Cross-Sectional and Time Series Elasticities of Food Consumption :

The Case of American and Polish Data

G. J. DUNCAN[1]
F. GARDES[2]
P. GAUBERT[3]
C. STRARZEC[4]

[1] Northwestern University, USA.
[2] CREST-LSM, Université de Paris 1 Panthéon-Sorbonne.
[3] Université du Littoral, Matisse (Université Paris 1)
[4] CNRS-TEAM, INSEE.

Panel and Pseudo-Panel Estimation of Cross-Sectional and Time Series Elasticities of Food Consumption: The Case of American and Polish Data[1]

François Gardes, CREST-LSM, Université de Paris I Panthéon-Sorbonne

Greg J. Duncan, Northwestern University

Patrice Gaubert, Université du Littoral, Matisse (Un.Paris I)

Christophe Starzec, CNRS - TEAM, INSEE

Keywords. Panel, pseudo panel and cross section estimations.

Classification JEL: D12, C33, C31

Abstract.

The problem addressed in this article is the bias to income and expenditure elasticities estimated on pseudo-panel data caused by measurement error and unobserved heterogeneity. We gauge empirically these biases by comparing cross-sectional, pseudo-panel and true panel data from both Polish and American expenditure surveys. Our results suggest that unobserved heterogeneity imparts a downward bias to cross-section estimates of income elasticities of at-home food expenditures and an upward bias to estimates of income elasticities of away-from-home food expenditures. Within and first-difference estimators suffer less bias, but only if the effects of measurement error are accounted for with instrumental variables. Pseudo-panel data provide cross-sectional estimates with less likely bias than estimates based on individual cross-sectional data. The magnitude of the differences in elasticity estimates across methods of estimation is roughly similar in U.S. and Polish-based expenditure data. Contrary to typical cross-sectional estimates, income elasticities for food at home and food away in United States are very similar in magnitude.

Résumé

L'article analyse les biais sur les élasticités de la consommation alimentaire par rapport à la dépense totale et au revenu telles qu'elles sont estimées sur des données de pseudo-panel. L'origine de ces biais est l'erreur de mesure propre aux pseudo-panels et l'hétérogénéité inobservée. Ces biais sont évalués en utilisant trois types de données: un vrai panel, un pseudo-panel et les coupes transversales provenant des enquêtes polonaises de budget de famille et d'un panel américain (PSID). Nos résultats suggèrent que l'hétérogénéité inobservée implique un biais allant dans le sens d'une sous-estimation des élasticités-revenu de l'alimentation à domicile obtenues à partir de la coupe transversale et une sur-estimation de l'élasticité-revenu de l'alimentation à l'extérieur. Les estimateurs "within" et de différences premières subissent un biais plus faible mais uniquement dans le cas où l'effet des erreurs de mesure est neutralisé par l'instrumentation. Les données de pseudo-panel donnent les estimateurs "cross-section" avec moins de biais que les estimateurs obtenus à partir des données individuelles transversales. Les écarts entre les élasticités obtenues selon les méthodes d'estimation utilisées sont proches pour les données américaines et polonaises. Contrairement à ce que l'on trouve habituellement à partir des estimations sur les données transversales, les élasticités-revenu pour l'alimentation à domicile et à l'extérieur aux Etats Unis ont des niveaux très proches.

[1] We gratefully acknowledge helpful suggestions from participants at seminars at the University of Michigan, Northwestern University, Université de Paris I, Université de Genève, Erudite, CREST, Journées de Microéconomie Appliquée (1998), Congress of the European Economic Association, International Conference on Panel Data, as well as research support from Inra, Inrets and Credoc. The Polish data were available thanks to Professor Górecki, University of Warsaw, Department of Economics.

I. Introduction

A multitude of data types and econometric models can be used to estimate demand systems. Data types include aggregate time series, within-group time series, cross-sections, pseudo-panels using aggregated data, and cross sections and panels using individual data.

Static and dynamic models have been developed for these data, with each adopting a different approach to problems caused by unobserved heterogeneity across consumption units or time period of measurement as well as the cross-equation restrictions imposed by consumption theory. The use of different types of data helps reveal the nature of the biases they impart to estimates of income and expenditure elasticities.

This article addresses the issue of bias to income and expenditure elasticities caused by errors of specification, measurement and omitted variables in grouped and individual-based models. We gauge these biases by estimating static expenditure models using cross-sectional, pseudo-panel and true panel data from both Polish and U.S. expenditure surveys. It is, to our knowledge, the first comparison between cross-sectional, pseudo-panel and panel estimations based on the same data set. The use of one of our two data sets (The Panel Study of Income Dynamics - PSID) is motivated by the numerous expenditure studies based on it (Altug and Miller, 1990; Altonji, Siow, 1987; Hall, Mishkin,1982; Naik, Moore, 1996; Zeldes, 1989). Our second data set is from Poland in the late 1980s, which enables us to capitalize on large income and price variations during the transition period in Poland.

Section 2 presents a background discussion. The econometric problems and methods used are presented in Section 3. The data are described in the fourth section, with results presented in the fifth section and discussed in the sixth section.

II. Background

No matter how complete, survey data on household expenditures and demographic characteristics lack explicit measures of all of the possible factors that might bias the estimates of income and price elasticities. For example, the value of time differs across households and is positively related to a household’s observed income. Since consumption activities (e.g., eating meals) often involve inputs of both goods (e.g., groceries) and time (e.g., spent cooking and eating), households will face different (full) prices of consumption even if the prices of the goods-based inputs are identical. If, as is likely in the case of meals prepared at home, these prices are positively associated with income and themselves have a negative effect on consumption, the omission of explicit measurement of full prices will impart a negative bias to the estimated income elasticities. The same argument can be applied to the case of virtual prices arising, say, from liquidity constraints that are most likely in low-income households (Cardoso and Gardes, 1997).

Panel data on households provide opportunities to reduce these biases, since they contain information on changes in expenditures and income for the same households. Differencing successive panel waves nets out the biasing effects of unmeasured persistent characteristics. But while reducing bias due to omitted variables, differencing income data is likely to magnify another source of bias: measurement error. Altonji and Siow (1987) demonstrate the likely importance of measurement error in the context of first-difference consumption models by showing that estimates of income elasticities are several times higher when income change is instrumented than when it is not.

Deaton (1986) presents the case for using “pseudo-panel” data to estimate demand systems. He assumes that the researcher has independent cross sections with the required expenditure and demographic information and shows how cross sections in successive years can be grouped into comparable demographic categories and then differenced to produce many of the advantages gained from differencing individual panel data.

We evaluate implications of alternative approaches to estimating demand systems using two sets of household panel expenditure data. The two panels provide us with data needed to estimate static expenditure models in first difference and “within” form. However, these data can also be treated as though they came from independent cross sections and from grouped rather than individual-household-level observations. Thus we are able to compare estimates from a wide variety of data types. Habit persistence and other dynamic factors give rise to dynamic models (e.g., Naik and Moore, 1996). We estimated dynamic versions of the static models using usual instrumentation methods (Arrelano, 1989) and found that elasticity estimates were quite similar to those estimated for the static models that are presented in this paper. However, since these dynamic versions may suffer from specification and econometric problem, we do not consider them in this paper.

True panel and pseudo-panel methods each offer advantages and disadvantages for handling the estimation problems inherent in expenditure models. A first set of concerns center on measurement error. Survey reports of household income are measured with error; differencing reports of household income across waves undoubtedly increases the extent of error. Instrumental variables can be used to address the biases caused by measurement error (Altonji and Siow, 1987). Like instrumentation, aggregation in pseudo-panel data helps to reduce the biasing effects of measurement error, so we expect that the income elasticity parameters estimated with pseudo-panel data to be similar to those estimated on instrumented income using true panel data. Since measurement error is not likely to be serious in the case of demographic variables like location, age, social category, and family composition, we confine our instrumental variables adjustments to our income and total expenditure predictors. Measurement errors in our dependent expenditure variable are included in model residuals and, unless correlated with the levels of our key independent variables, should not bias the coefficient estimates.

Second, the aggregation inherent in pseudo-panel data produces a systematic heteroskedasticity. This can be corrected by decomposing the data into between and within dimensions and computing the exact heteroskedasticity on both dimensions. But since the heteroskedastic factor depends on time, correcting it by GLS makes individual specific effects vary with time, thus canceling the spectral decomposition in between and within dimensions. This can result in serious estimation errors (Gurgand, Gardes and Bolduc, 1997). The approximate correction of heteroskedasticity that we use consists in weighting each observation by a heteroskedasticity factor that is a function of cell size. Thus the LS coefficients computed on the grouped data may differ slightly on those estimated on individual data. This approximate and easily implemented correction consists of using GLS on the within and between dimensions with a common variance-covariance matrix computed as the between transformation of the heteroskedastic structure due to aggregation.

Third, unmeasured heterogeneity is likely to be present in both panel and pseudo-panel data. In the case of panel data the individual-specific effect for household h is $\alpha(h)$, which is assumed to be constant through time. In the case of pseudo-panel data, the individual-specific effects for a household (h) belonging to the cell (H) at period *t*, can be written as the sum of two orthogonal effects: $\alpha(h,t)=\mu(H) + \upsilon(h,t)$. Note that the second component depends on time since the individuals composing the cell H change through time.

The specific effect μ corresponding to the cell H ($\mu(H)$) represents the influence of unknown explanatory variables W(H), constant through time, for the reference group H, which is defined here by the cell selection criteria. $\upsilon(h,t)$ are individual specific effects containing effects of unknown explanatory variables Z(h,t). In the pseudo-panel data the aggregated specific effect $\zeta(H)$ for the cell H is defined as the aggregation of individual specific effects:

$$\zeta(H,t)=\sum \gamma(h,t)*\alpha(h,t) = \mu(H) + \sum\gamma(h,t) * \upsilon(h,t)$$

where t indicates the observation period and γ is the weight for the aggregation of h within cells. Note that the aggregate but not individual specific effects depend on time.

The within and first difference operators estimated with panel data cancel the individual specific effects $\alpha(h)$. The component $\mu(H)$ is also canceled on pseudo-panel data by the same operators, while the individual effect $\upsilon(h,t)$ may be largely eliminated by the aggregation. Thus it can be supposed that the endogeneity of the specific effect is greater on individual than on aggregated data, as aggregation cancels a part of this effect.

Therefore, with panel data the within and the first-differences operators suppress all the endogeneity biases. With pseudo-panel data the same operator suppresses the endogeneity due to μ, but not that due to $\sum\gamma(h,t) * \upsilon(h,t)$. For each individual this part of the residual may be smaller relatively to μ, as cell homogeneity is increased. Conversely, the aggregation into cells is likely to cancel this same component υ across individuals, so that it is not easy to predict the effect of the aggregation on the endogeneity bias.

Our search for robust results is facilitated by the fact that the two panel data sets we use cover extremely different societies and historical periods. One is from the United States for 1984-1987, a period of steady and substantial macroeconomic growth. The second source is from Poland for 1987-1990, a turbulent period that spans the beginning of Poland’s transition from a command to free-market economy.

III. Specification and econometrics of the consumption model

Data constraints force us to estimate a demand system on only two commodity groups over a period of four years: food consumed at home and food consumed away from home. Away-from-home food expenditures are rare in the Polish data and produced unreliable estimates, but we keep them in order to compare them with PSID estimates. We use the Almost Ideal Demand system developed by Deaton and Muellbauer (1980), with a quadratic form for the natural logarithm of total income or expenditures in order to take into account non-linearities. Note that the true quadratic system proposed by Banks et al. (1997) implies much more sophisticated econometrics if the non-linear effect of prices is taken into account. It may be difficult to estimate precisely the price effect because of the short estimation period. Thus we estimate the linearized version of QAIDS with the Stone index. The additivity constraint is automatically imposed by OLS.

The possible correlation between the residuals of food at home and food away suggests the use of Seemingly Unrelated Regression. We tested this possible correlation on Polish data and we found no significant difference between OLS and SUR estimations. For the U.S. in the mid 1980s, relative prices vary much less within than between annual survey waves. We account for price effects and other macro-economic shocks with survey year dummies for the US data. For the Polish data each individual was given a price index differentiated by the social category and the quarter of the year in which he was surveyed.

Our model takes the following form:

$$w_{ht}^{i} = a^{i} + b^{i}\ln\left(Y_{ht} / p_{t}\right) + c^{i} / e(p)\left[\ln\left(Y_{ht} / p_{t}\right)\right]^{2} + Z_{ht}d^{i} + u_{ht}^{i} \qquad (1)$$

with w^{i}_{ht} the expenditure budget share on good i by household h at time t, Y_{ht} its income (in the case of U.S. data and logarithmic total expenditure in the case of the Polish expenditure panel), p_t the Stone price index, Z_{ht} a matrix of socio-economic characteristics and survey year or quarter dummies and e(p) is a factor ensuring the integrability of the demand system, which was estimated by the convergence procedure proposed by Banks et al. (1997). When using total expenditure data from the Polish panel, the allocation of income between consumption and saving can be ignored, and total expenditures can be considered as a proxy for permanent income. Our U.S. data do not provide information on total expenditure, so that income elasticities are computed on the basis of total household disposable income.

Our cross-sectional estimates of equation (1) are based on data on individual households from each available single-year cross-section (1984-1987 in the case of the PSID and 1987-1990 in the case of the Polish expenditure survey).

First differencing and within operators are common procedures employed to eliminate biases caused by persistent omitted variables, and we use our panel data to obtain first-difference and within estimates of our model. Following Altonji and Siow (1987), we estimate our models both with and without instrumenting for change in log income or expenditures. Instrumenting income from the PSID is necessary because of likely measurement errors observed in such income data. We also instrumented the total expenditure from the Polish surveys because measurement errors for both total expenditures and food expenditures are likely correlated.

In the QAIDS specification the classical errors-in-variables cannot hold for the squared log income term if it holds for the log of income. As far as we know this problem has not been solved conveniently so we simply used the square of the instrumented income checking that a separate instrumentation of the squared term does not change significantly the results.

We found two types of correlations for cross-sections and first-differences: between individuals in cross-sections and between periods in first-differences. We handled this problem by estimating separately for each period with a robust OLS method. For the within estimation, all autoregressive processes on the residuals (for instance resulting from partial adjustment in exogenous variables) are taken into account, as suggested by Hsiao (1986, p.95-96), by estimating the system of equations written for the successive periods.

Pseudo-panel estimates. We grouped data for pseudo-panels using six age cohorts and two or three education levels. The grouping of households (h,t) in the cells (H,t) gives rise to the exact aggregated model:

$$\sum_{h\in H}\gamma_{ht}w^i_{ht} = w^i_{Ht} = \left(\sum_h \gamma_{ht}X_{ht}\right)A^i + \alpha^i_H + \sum_h \gamma_{ht}\varepsilon^i_{ht}$$

with $\gamma_{ht} = \frac{Y_{ht}}{\sum_{h\in H} Y_{ht}}$ under the hypothesis $\alpha^i_h = \alpha^i_H$ for $h \in H$ (a natural hypothesis, according to the grouping of households into a same H cell). A heteroskedasticity factor $\delta_{Ht} = \sum_{h\in H}\gamma^2_{ht}$ arises for the residual ε^i, which is due to the change of cells sizes (as $\gamma \cong \frac{1}{|H|}$ if the two grouping criteria homogenize the household's total expenditures).

The PSID data provide an indicator of to which of four random subsamples a given case belongs. We use these indicators to aggregate cases for the four different years. This prevents the same household from being included in the same cell in more than one period (in which case the aggregation would just correspond to grouped panel data).

For Polish data, all households (after filtering for some outliers defined on cross-section estimations) in the cross-sectional component of each survey are used for the pseudo-panelization; panel households belonging to the surveys are excluded. Sample sizes for each year are around 27,000 households, which is much larger than for the PSID data.

The PSID cells sizes vary from 9 to 183 households with a mean of 65.5 and from 8 to 60, with a mean of 25.1 for the Polish data. Fourteen of the 72 cells constituting the pseudo-panel in the PSID contain less than 30 households, and represent only 4 % of the entire sample. As the correction for the heteroskedasticity on the pseudo-panel data consists in weighting each cell by weights close to its size, the estimation without these small cells gives the same results as those for the whole. For each cell the size variation through time is much less important, so that the heteroskedasticity factor due to the grouping is quite invariant through time. Thus, we correct for it by simple generalized least squares based on the average heteroskedasticity factor of each cell:

$$\sum_{t=1}^{T}\sum_{h\in H}\frac{1}{T}\gamma^2_{ht} = \delta_H$$

It is clear that the residuals for two adjacent equations estimated in first differences, ($u^i_{h,t}-u^i_{h,t-1}$) and ($u^i_{h,t-1}-u^i_{h,t-2}$), are systematically correlated. Since all specifications are estimated by Zellner's seemingly unrelated regressions, our procedures take into account the correlation between the residuals of the two food components.

Price effects are taken into account by period dummies for the PSID and by price elasticities for Poland. The age of the household's head, and family size and structure are also taken into account in the estimations. Adding other control variables such as head's sex, education level, wealth, and employment status in the PSID had very little effect on the estimates. We selected only age and family structure variables for the PSID to make the estimations comparable to the results based on the Polish data.

Correction for grouped heteroskedasticity may still leave some heteroskedasticity for the estimations at the individual level. We test for this by regressing the squared residuals on a quadratic form of explanatory variables, thus correcting it when necessary by weighting all observations by the inverse absolute residual. The coefficient on the squared income is generally significant, but QAIDS estimates are very close to AIDS.

IV. Data

The Panel Study of Income Dynamics. Since 1968, the PSID has followed and interviewed annually a national sample that began with about 5,000 U.S. families (Hill, 1992). The original sample consisted of two sub-samples: i) an equal-probability sample of about 3,000 households drawn from the Survey Research Center's dwelling-based sampling frame; and ii) a sample of low-income families that had been interviewed in 1966 as part of the U.S. Census Bureau's Survey of Economic Opportunity and who consented to participate in the PSID.

When weighted, the combined sample is designed to be continuously representative of the nonimmigrant population as a whole. To avoid problems that might be associated with the low-income sub-sample, our estimations based on individual-household data are limited to the (unweighted) equal-probability portion of the PSID sample. To maximize within-cell sample sizes, our pseudo-panel estimates are based on the combined, total weighted PSID sample. We note instances when pseudo-panel estimates differed from those based on the equal-probability portion of the PSID sample.

Since income instrumentation requires lagged measures from two previous years, our 1982-87 subset of PSID data provides us with data spanning five cross sections (1983-1987). We use only four years 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 "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 PSID provides information on two categories of expenditure: food consumed at home and food consumed away from home and has been used in many expenditure studies (e.g., Hall and Mishkin, 1982; Altonji and Siow, 1987; Zeldes, 1989; Altug and Miller, 1990; Naik and Moore, 1996). These expenditures are reported by the households as an estimation of their yearly consumption so reporting a zero consumption can be considered as a true case of zero consumption, eliminating the need for selection bias adjustment.

To adjust expenditures and income for family size we use the Oxford equivalence scale: 1.0 for the first adult, 0.8 for the others adults, 0.5 for the children over 5 years old and 0.4 for those under 6 years old. Our expenditure equations also include a number of household structure variables to provide additional adjustments for possible expenditure differences across different family types.

Disposable income is computed as total annual household cash income plus food stamps minus household payments of alimony and child support to dependents living outside the household and minus income taxes paid. The household's expenditure on food purchased with food stamps is included in our measure of at-home food expenditure. As instruments for levels of disposable income we follow Altonji and Siow (1987) in including three lags of quits, layoffs, promotions and wage-rate changes for the household head, we construct our wage rate measure from a question sequence about rate of hourly pay or salary that is independent of the question sequence that provides the data on disposable household income) as well as changes in family composition other than the head, marriage and divorce/widowhood for the head, city size and region dummies. For first-difference models, the change in disposable income is instrumented using the first-difference of instrumented income in level.

Means and standard deviations of the PSID variables are presented in Appendix Table 1; coefficients and standard errors from the first stage of the instrumental variables procedure are presented in Appendix Table 2.

<u>The Polish expenditure panel</u>. Household budget surveys have been conducted in Poland for many years. In the analyzed period (1987-1990) the annual total sample size was about 30 thousand households; this is approximately 0.3% of all the households in Poland. The data were collected by a rotation method on a quarterly basis. The master sample consists of households and persons living in randomly selected dwellings. To generate it, a two stage, and in the second stage, two phase sampling procedure was used. The full description of the master sample generating procedure is given by Lednicki (1982).

Master samples for each year contains data from four different sub-samples. Two sub-samples began their interviews in 1986 and ended the four-year survey period in 1989. They were replaced by new sub-samples in 1990. Another two sub-samples of the same size were started in 1987 and followed through 1990.

Over this four-year period in every annual sample it is possible to identify households participating in the surveys during all four years who collectively form a four-year panel. There is no formal identification possibility (by number) of this repetitive participation, but tests allows to specify with a very high probability the four year participants. The checked and tested number of households is 3,707 (3,630 after some filtering). The available information includes socio-demographic characteristics of households and individuals, as well as details on incomes and expenditures. Food expenditures are reported for three consecutive months each year, so we again considered zero expenditure as a true no-consumption case.

Comparisons between reported household income and record-based information showed a number of large discrepancies. For employees of state-owned and cooperative enterprises (who constituted more than 90% of wage-earners until 1991), wage and salary incomes were checked at the source (employers). Kordos and Kubiczek (1991) estimate that employees' income declarations for 1991 were 21% lower, on average, than employers' declaration.

Generally, the proportion of unreported income is decreasing with the level of education and increasing with age. In cases where declared income was lower than that reported by enterprises, household's income was increased to the level of the reported one. Since income measures are used only to form instrumental variables in our expenditure equations, the remaining measurement error is likely to cause only minor problems.

Appendix Table 3 presents descriptive information on the Polish data, while Appendix Table 4 presents coefficients from the instrumental-variables equation. The period 1987-1990 covered by the Polish data is unusual even in Polish economic history. It represents the shift from the centrally planned, rationed economy (1987) to a relatively unconstrained fully liberal market economy (1990). GDP grew by 4.1% between 1987 and 1988, but fell by .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.

This means that the consumers' market reactions could have been highly influenced by these unusual situations. This is most likely the case of the year 1989 when uncertainty, inflation, market disequilibrium and political instability reached their highest level. Moreover, in 1989 and 1990 individuals were facing a huge real income fluctuations as well as a very strong change in relative prices. This particularly unstable situation had as a result the atypical consumption behaviors of households facing a subsistence constraint. This may be the case of very low income households, over 30% of whom faced a dramatic decrease in purchasing power.

V. Results

Estimates from our various models are presented in Tables 1 (PSID) and 2 (Polish surveys). Respective columns show income (for PSID; total expenditure for Polish data) elasticity estimates for between, cross section, within and first-difference models. Results are also presented separately for models in which income (total expenditure) is and is not instrumented using the models detailed in Appendix Tables 2 and 4. We expect the between estimates to be similar to the average of cross-section estimates. Compared to the within, the first-difference estimates may be biased by greater measurement error, but the specific effects may be better taken into account whenever they change within the period.

Looking first at the PSID results for at-home food expenditures, it is quite apparent that elasticity estimates are very sensitive to adjustments for measurement error and unmeasured heterogeneity. Cross-sectional estimates of at-home income elasticities are low (between .15 and .30) but statistically significant with or without instrumentation (when performing robust estimations). The between estimates effectively average the cross sections and also produce low estimates of elasticities. Pseudo-panel data produces similar elasticities for between and cross-sections estimates. Despite some variations between the different estimations, the relative income elasticity of food at home is around .20 based on this collection of methods.

Within and first difference estimates of PSID-based income elasticities are close to zero without instrumentation and around .40 with instrumentation. Pseudo-panel within and first-differences estimates are somewhat smaller (around .3). A Hausman test strongly rejects (p-value<.01) the equality of within and between estimates (Table 3). Since the within and first difference models adjust for persistent heterogeneity and the instrumentation adjusts for measurement error, .40 to .50 is our preferred range of estimates for the income elasticity of at-home food expenditures in the United States. This range spans values that are half again as large as the corresponding between and cross-sectional parameter estimates, which suggests that failure to adjust for heterogeneity imparts a considerably downward bias to cross-sectional estimates. PSID-based pseudo-panel data also produce a significantly (according to a Hausman test) higher elasticity estimates for first-differences as compared with between and cross-sections models although there is little consistency across the full set of pseudo-panel estimates. Pseudo-panel estimates fall somewhat below our preferred .40 to .50 range.

Expenditure elasticities for at-home food estimated with the Polish data are much higher in value than the income elasticity estimates based on PSID data.[2] Higher elasticities are to be expected for a country in which food constitutes a share of total expenditures that is three times higher than in the U.S. (Appendix Tables 1 and 3). Their consistency probably stems from the smaller degree of measurement error in the Polish expenditure as opposed to the PSID income data. Time series Polish pseudo-panel estimates are somewhat smaller than estimates based on the microdata. On the whole the estimations on Polish data also produce higher within and first-differences elasticities than between and cross-sections.

PSID-based income elasticities for away-from-home food expenditures are quite different from and even more sensitive to specification than the at-home elasticity estimates. Between and cross-sectional estimates are around 1.0 in both individual and pseudo-panel data. In contrast to the case of at-home expenditures, adjustments for heterogeneity through use of within and first difference estimates produce much lower estimates. We speculate on why this might be the case in our discussion section.

Polish food-away expenditures are relatively rarely reported in the survey and they are very low even when compared with other countries with comparable income levels. Moreover, about 70% of this expenditure in the observed periods are spent on highly subsidized business canteens and cafeterias. So the estimations should be compared with caution with the food away expenditure estimation in other countries.

Price data in Poland enabled us to compute price elasticities. Quarterly price indices for four social categories were computed from monthly GUS (Polish Main Statistical Office) publication Biuletyn Statystyczny and imputed at the individual level to the data set. The variability of the prices both over time and over social categories provides good estimates of the direct compensated elasticity for food at home (around minus one). Contrary to the income elasticities, the cross-section estimates of direct price elasticities are close to those obtained from time-series. As prices change between quarters and households of different types the only explanation of an endogeneity bias would be a correlation between household types and level

[2] Note that the Polish elasticities are computed on total expenditures so that they must be multiplied by the income elasticity of total expenditures, which is around .7, to be compared to the PSID income elasticities. Before grouping into cells ten households have been suppressed. The criterion was the prediction of food consumption in cross-section estimations. On the panel data, robust estimations produced by suppressing some outliers gave similar results as those obtained for the whole panel.

of prices. Such a correlation is less probable than the one between the relative income and the specific component of food consumption (which produces the endogeneity bias on income elasticities). However, if systematically high (or low) price level is correlated with household's type (which can be the case for instance on segmented markets) and if these types of households are characterized by a systematically positive or negative specific consumption, the endogeneity bias can appear. This is not the case of Poland in 1987-1990.

VI. Discussion

We have attempted to assess the bias to income and expenditure elasticity estimates caused by inattention to measurement error and unobserved heterogeneity. In the case of the U.S. at-home food expenditure elasticity, our preferred estimate is around .40. Failure to adjust for unmeasured heterogeneity and, in some cases, measurement error appears to impart a substantial downward bias to this estimate. These adjustments operate in the same direction for estimates of the at-home expenditure elasticity found in the Polish data.

In the case of U.S. away-from-home food expenditures, our preferred elasticity estimate is less certain but similar in magnitude to the .40 elasticity for at-home expenditures. Surprising here is the magnitude and sign (upward) of the apparent bias inherent in both individual and pseudo-panel estimates that do not adjust for unobserved heterogeneity.

Why should unmeasured heterogeneity induce an upward bias in away-from-home expenditures in the U.S. and Polish data? Earlier, we speculated that a likely downward bias for the at-home food elasticity estimates may be caused by failure to account for the fact that the value of time differs across households and is positively related to the household's observed income. The time input to producing at-home meals lead households to face different (full) prices of consumption even if the prices of the goods-based inputs are identical. If these prices are positively associated with income and themselves have a negative effect on consumption, their omission will impart a negative bias to the estimated income elasticities. Note that the bias seems somewhat less pronounced on pseudo-panel data. The correlation between income and the specific effect may be decreased by aggregation, as suggested in section 2.

In the case of expenditures for meals consumed in restaurants, there are large variations in the mixture of food and service components. Time spent consuming full-service restaurant meals is typically longer than time spent consuming fast-food restaurant meals, but in this case higher-income households may well attach a more positive value to such time than low income households. Failing to control for this source of heterogeneity will probably impart a positive bias to estimates of income elasticities.

Elasticity estimates from the Polish data also proved somewhat sensitive to adjustments for heterogeneity. This is not surprising given the very different price regimes in the Polish economy during this period, with one, presumably low set of prices faced by the many farm families who have the option of growing their own food; official, subsidized prices set by the Polish government; and higher black market prices for the same or similar products. In fact, until 1988 the official prices of stables such as bread were set so low that very few farm families grew food for their own consumption, when queuing was not a problem.

Our estimates are based on a static consumption models and risk of bias due to the omission of dynamic factors such as habit persistence. We investigated this by estimating a dynamic version of our model that included lagged consumption. To obtain cross-sectional estimates of our dynamic model, we treated our data as though they came from three independent two-year panels (1984-85 through 1986-87 in the case of the PSID and 1987-88 through 1989-90 in the case of the Polish expenditure survey). We estimate this dynamic model as a system using SURE and both with and without instrumentation for log total income or expenditure. We found that the coefficient on the lagged dependent variables were significant in these dynamic models, but their inclusion changed the values of the short-run income elasticity very little.

VII. Conclusion

An important result of this work is that pseudo panel estimates are often close to estimates based on genuine panel data. Large and similar apparent endogeneity biases were found in both countries. Cross-section estimations produce elasticities that are systematically higher for food at home and lower for food away. Moreover it seems that the aggregation lowers the endogeneity bias for food consumption. In further research these results should be verified on a complete set of expenditure data. Once corrected for the biases, food at home and food away income elasticities become very close to each other in the U.S. data, a result that seems reasonable to us and highlights possible errors that can arise from estimations using cross-sectional data.

REFERENCES

Altug, S. and Miller, R., "Household choices in equilibrium", *Econometrica* 58 (1990), 543-570.

Altonji, J. and Siow, A., "Testing the response of consumption to income changes with (noisy) panel data", *Quarterly Journal of Economics* (1987), 293-328.

Arellano, M., "A note on the Anderson-Hsiao estimator for panel data", *Economic Letters* 31 (1989), 337-341.

Baltagi, H. B., *Econometric analysis of panel data* (New York: Wiley, 1995).

Banks, J., Blundell, R. and Lewbel, A., "Quadratic Engel curves and consumer demand", *The Review of Economics and Statistics*, 79 (1997), 527-539.

Blanciforti, L. and Green, R., "An almost ideal demand system incorporating habit effect", *Review of Economics and Statistics* 65 (1983), 511-515 .

Cardoso, N. and Gardes, F., "Estimation de lois de consommation sur un pseudo-panel d'enquêtes de l'INSEE (1979, 1984, 1989)", *Economie et Prévision* 5 (1996), 111-125.

Gardes, F., Langlois, S. and Richaudeau, D. "Cross-section versus time-series income elasticities", *Economics Letters* 51 (1996), 169-175.

Gorecki, B. and Peczkowski, M., "Polish Household Panel", Dept. of Economics, Warsaw University (1992).

Gurgand, M., Gardes, F. and Bolduc, D., "Heteroskedasticity in pseudo-panel", working paper, Lamia, Université de Paris I, (1997).

Hall, R. and Mishkin, F., "The sensitivity of consumption to transitory income: estimates from panel data on households", *Econometrica* 50 (1982), 261-281.

Hill, M.S., *The Panel Study of Income Dynamics, a User's Guide*, (Newburry Park, SAGE Publications, 1992).

Hsiao, C., *Analysis of panel data*, (Cambridge: Cambridge University Press, 1986).

Kordos, J., "Towards an Integrated System of Household Surveys in Poland", *Bulletin of International Statistical Institute*, 51 (1985).

Kordos, J. and Kubiczek, A., "Methodological Problems in the Household Budget Surveys in Poland", GUS, Warsaw,1991.

Lednicki B., "Dobor proby i metoda estimacji w rotacyjnym badaniu gospodarstw domowych (Sample Design and Method of Estimation in Rotation Survey of Households)", GUS, Warsaw, 1982.

Naik, N. Y. and Moore, M. J., "Habit Formation and Intertemporal Substitution in Individual Food Consumption ", *The Review of Economics and Statistics*, 78 (1996), 321-328.

White, H., "A Heteroskedasticity consistent covariance matrix estimator and a direct test for heteroskedasticity", *Econometrica*, 48-4 (1980), p. 817-838.

Zeldes, S.P. , "Consumption and liquidity constraints : an empirical investigation", *Journal of Political Economy*, 97 (1989), 305-345.

Table 1: Income elasticities for food at home and away from home: Panel Study of Income Dynamics (1984-87)

	Between	Cross-Sections[1]	Within	First-differences
Panel				
Food at home				
Not instrumented	0.167 (.014)	0.140 (.031)	-0.019 (.031)	-0.029 (.058)
Instrumented				
- whole pop.	0.010 (.029)	-0.030 (.066)	0.363 (.104)	0.473 (.165)
- robust est.[2]	0.187 (.023)	0.134 (.056)	0.383 (.104)	0.447 (.157)
Food away				
Not instrumented	0.963 (.018)	0.872 (.041)	-0.025 (.080)	0.062 (.091)
Instrumented				
- whole pop.	1.046 (.042)	0.966 (.101)	0.375 (.175)	0.431 (.282)
- robust est.	1.050 (.042)	0.960 (.101)	0.387 (.189)	0.447 (.282)
Surveys	1984-85-86-87b			
n[3]	9720			
Control variables	Age, equivalence scale and its square[4]			
Pseudo-panel				
Food at home				
Not instrumented	0.311 (.045)	0.265 (.056)[5]	0.240 (.095)	0.382 (.114)
Food away				
Not instrumented	1.387 (.068)	1.265 (.083)[5]	0.800 (.150)	0.847 (.172)
Surveys	1984-85-86-87			
N	90			
Control variables	Log of Age and its square, equivalence scale and its square			

Note: All standard errors have been adjusted for heteroskedasticity by White's (1980) method and for the instrumentation of income.

1: cross-section elasticities computed as the means of cross-sectional estimates obtained on each cross-sectional survey

2 : the robust estimates are obtained by deleting all households for which the residuals of the food at home consumption function are greater than 2.6 times the root mean standard error. This eliminates about 0.5% of the sample.

3 : 12 cases with missing data were eliminated when instrumenting income.

4 : Adding control variables such as wealth and household members' employment status does not affect the estimates substantially.

5. Average of estimates for the four surveys

Table 2: Total expenditure elasticities for food at home and away from home: Polish surveys (1987-90).

	Between	Cross-Sections°°	Within	First-differences
Panel				
Food at home				
Not instrumented	0.579 (.004)	0.536 (.005)	0.466 (.006)	0.451 (.007)
Instrumented	0.494 (.012)	0.567 (.010)	0.755 (.012)	0.788 (.016)
Food away				
Not instrumented	1.119 (.067)°°°	1.239 (.091)	2.618 (.518)°°°	1.460 (.181)
Instrumented	1.216 (.119)°°°	1.326 (.148)	4.195 (.993)°°°	1.315 (.198)
N	14520	14520	14520	10890
Control variables	Log of Age, proportion of children, Education level, Location, Log of relative price for all commodities, quarterly dummies			
Pseudo-panel				
Food at home				
Not instrumented	0.452 (.022)	0.549 (.023)	0.542 (.023)	0.865 (.036)
Food away				
Not Instrumented	1.001 (.002)	0.636 (.246)	0.120 (.347)°	0.360 (.370)
N	224			
Control variables	Log of Age, proportion of children, Location, Log of relative prices for food, quarterly and year dummies			

Note: All standard errors have been adjusted for heteroskedasticity by White's (1980) method and for the instrumentation of Total Expenditures by the usual method. AIDS estimates. The estimation of a Quadratic AI demand system by iteration on the integrability parameter (see Banks et al., 1999) gives very similar results, except for case °°°. Filtering data for outliers (like for PSID) did not change significantly the results.

° Estimation without 1969.

°° Average of estimates for the four surveys

°°° QAIDS estimates: for not instrumented income: 1.128 (.064) for Between, 1.457 (.139) for

Within; for instrumented income: 1.252 (.118) for Between, 1.645 (.200) for Within.

Table 3: Hausman test for income parameters (Food at home)

	Panel		Pseudo-panel
Instrument	without IV	with IV	Without IV
PSID	77.8	4.0	0.5°
Poland	254.9	245.7	8.0
Poland°°	250.2	280.0	.-

The test is computed by the usual quadratic form, distributed as a $\chi 2$: $(\beta_b - \beta_w)'$ $P(V^{-1})$ $(\beta_b - \beta_w)$ where $\beta = (\beta_{ly})$ or $\beta = (\beta_{ly}, \beta_{ly^2})$ for the quadratic estimation on the Polish panel, $V=V_b+V_w$ corresponds to all the explanatory variables and P is the projection on log income or expenditure, and its square. Note that a test with V as a matrix 2x2 computed only for the two income variables would be biased. χ^2 bounds for 1 degree of freedom at 1 % : 6.63; 5%: 3.84; 10%: 2.71; for 2 degrees 9.21, 5.99, 4.61.

° 3.36 for food at home and away together.

°° QAIDS estimation

Appendix Table 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)

Appendix Table 2: Regression Coefficient and Standard Errors for Instrumental Variables Equation for Income Level for the PSID (Dependent Variable: Natural Logarithm of Disposable Family Income in 1987, 1986, 1985)

Independent Variable	Coefficient (Standard Error)	Independent Variable	Coefficient (Standard Error)
Quit $_t$	-.049 (.014)	Birth	.001 (.015)
Quit $_{t-1}$	-.036 (.014)	Age Head	.073 (.004)
Quit $_{t-2}$	-.019 (.014)	Age Head squared	-.0007 (.00004)
Lay off $_t$	-.066 (.021)	Wage growth*Quit $_t$	-.063 (.033)
Lay off $_{t-1}$	-.103 (.022)	Wage growth*Quit $_{t-1}$	-.003 (.035)
Lay off $_{t-2}$	-.053 (.020)	Wage growth*Quit $_{t-2}$	-.005 (.032)
Promoted $_t$	.025 (.022)	Wage growth*Lay off $_t$	-.097 (.047)
Promoted $_{t-1}$	.047 (.021)	Wage growth*Lay off $_{t-1}$	-.093 (.051)
Promoted $_{t-2}$	.017 (.021)	Wage growth*Lay off $_{t-2}$	.005 (.042)
Unemp hrs $_t$	-.271 (.044)	Wage growth*Promoted $_t$	.043 (.074)
Unemp hrs $_{t-1}$	.043 (.043)	Wage growth*Promoted $_{t-1}$	-.149 (.075)
Unemp hrs $_{t-2}$	.139 (.040)	Wage growth*Promoted $_{t-2}$	-.035 (.069)
Hrs lost ill $_t$	.266 (.044)	Region 1-2	.037 (.021)
Hrs lost ill $_{t-1}$	.387 (.047)	Region >3	.034 (.028)
Hrs lost ill $_{t-2}$	.446 (.049)	City Size 1-2	-.016 (.026)
Wage growth $_t$	.050 (.024)	City Size >3	-.054 (.026)
Wage growth $_{t-1}$	-.005 (.022)	Education Head	.138 (.006)
Wage growth $_{t-2}$	-.005 (.016)	Wage	.012 (.003)
Divorce	.037 (.029)	Wage $_{t-1}$	.022 (.003)
Marriage	.202 (.030)		

Notes: The instrumentation of log income was made initially by pooling the five surveys with 39 instrumental variables (IV) giving the multiple correlation coefficient R^2 of 50%. Income was proved to be endogenous by Wu-Hausman test, but the set of the IV was not asymptotically independent of the consumption function residuals estimated with the instrumented income. Thus instruments are not valid.

By eliminating 12 IV we obtain significant positive result for both Wu-Hausman and IV validity tests. The R^2 coefficient decreases to .33.

Appendix Table 3: Means and standard deviations of variable used in the Polish panel analyses

	1987		1988		1989		1990	
	Level	Dif.	Level	Dif.	Level	Dif.	Level	Dif.
Budget share for food at home	0.508 (.14)	-	0.484 (.15)	-0.024 (.14)	0.486 (.18)	0.003 (.17)	0.554 (.15)	0.068 (.17)
% with at-home share > 0	100	-	100	100	100	100	100	100
Budget share for food away from home	0.006 (.02)	-	0.006 (.03)	-0.001 (.02)	0.005 (.02)	-0.001 (.02)	0.005 (.03)	.0002 (.03)
% with away-from-home share = 0	28.4	-	29.7	-	26.9	-	20.5	-
ln household expenditure	10.65 (.45)	-	11.17 (.49)	0.50 (.38)	12.25 (.79)	-0.18 (.62)	14.14 (.50)	-0.03 (.58)
ln head's age	3.789 (.33)	-	3.809 (.32)	0.020 (.16)	3.824 (.32)	0.014 (.15)	3.842 (.32)	0.019 (.15)
ln family size	1.140 (.59)	-	1.121 (.60)	-0.019 (.24)	1.095 (.61)	-0.026 (.21)	1.081 (.61)	-0.014 (.22)

Appendix Table 4: Regression Coefficient and Standard Errors for Instrumental Variables Equation for Total Expenditure Level and Change for the Polish Expenditure Panel

Independent Variable	Coefficient (Standard Error)
log income	.374 (.071)
children % in family	-.270 (.024)
log age	1.704 (.311)
log age squared	-.0299 (.042)
location (ref: countryside)	- - - - - - - - -
large city	.041 (.014)
average city	.028 (.014)
small city	.017 (.019)
social category (ref: wage earners)	- - - - - - - - - --
wage earners-farmers	-.096 (.015)
pensioneers	-.047 (.017)
farmers	-.196 (.018)
education	-.028 (.003)
log income squared	.004 (.003)

Notes: The hypothesis of endogeneity of total expenditure cannot be accepted by Wu-Hausman test. However, strong theoretical arguments (total expenditure contains measurement errors that give rise to a correlation between the residuals and the total expenditure used as an explanatory variable) leads us to instrument the income.