Heterogeneity in Consumer Demands and the Income Effect: Evidence from Panel Data[*]

Mette Lunde Christensen [†]
preliminary version

February 2002

Abstract

Empirical studies of consumer demand show that budget shares vary considerably with income. This variation can mainly be attributed to two economic effects: An *income effect* (consumers change their budget shares when their income changes) and a *taste effect* (consumers with different incomes have different tastes). In this paper we estimate a system of consumer budget share functions using a unique panel data set on consumer expenditures. The panel structure allows us to disentangle the income effect from unobservable individual heterogeneity in tastes and test for whether there is a taste effect. Our main finding is that the variation in budget shares seems to be caused by a combination of the income effect and the taste effect for most commodities. We discuss the implications of this finding for the consumer demand models which use cross-sectional data.

JEL: C33, D12

Keywords: Engel curves, unobservable heterogeneity, panel data

1 Introduction

Empirical studies of consumer demand on cross-sectional show that budget shares for a given commodity vary considerably with income. This variation in budget shares can, after controlling for demographics, be the result of two economics effects: An income effect, i.e. that consumers change their budget shares when their income changes. Or a taste effect, i.e. that consumers with different incomes have different tastes. At one extreme, if all the variation in budget shares is due solely to the income effect, then all consumers would have identical income-dependent budget share functions. At the other extreme, if all the variation in budget shares is due to the taste effect, then all consumers would have different income-independent budget share functions. Evidently, consumers have different income-independent budget share functions if the variation in budget shares is due to a combination of the income effect and the taste effect, then consumers have different income dependent budget share functions[1].

The appropriate data set on which to analyse the question of which of the two effects causes the variation in budget shares is a panel data set of consumer expenditures, since panel data enables us to control for unobserved individual heterogeneity in taste. But data sets on consumer expenditures are typically cross-sectional data, e.g. the British Family Expenditure Survey (FES) or the American Expenditure Survey (CES). Therefore, empirical specifications of budget share functions typically assume that the variation in budget shares is a pure income effect[2]. In case, though, of unobserved individual taste-heterogeneity, which is correlated with income, the cross-sectional estimates will be biased.

In this paper, we estimate a system of consumer budget shares on the Spanish Permanent Survey of Consumption (Encuesta Permanente de Consumo, hereafter EPC), which is a panel data set on household expenditures for Spanish households in the period 1978-1983. The data set contains quarterly information on consumer expenditures on 13 different commodities[3].

---

[1]When consumers' budget share functions can be generated from utility maximisation, the case of a pure income effect corresponds to consumers having identical non-homothetic preferences, and the case of a pure taste effect corresponds to consumers having different homothetic preferences.

[2]E.g. the Almost Ideal Demand System introduced by Deaton & Muellbauer (1980) and the Quadratic Almost Ideal Demand System introduced by Banks, Blundell & Lewbel (1997).

[3]It is, to the best of my knowledge, one of the longest panel consumer surveys with

We model consumer budget shares by a Seemingly Unrelated Regressions Model (hereafter SUR) where the budget share for each good consists of an explained part plus an unexplained part. The explained part is a function of total expenditure, prices and demographics and has the same functional form for all households. The functional form is taken to be linear in log total expenditure and log total expenditure squared, i.e. a quadratic form. The unexplained part is composed of a household-specific effect, which is constant over time, and the usual error term. The household -and commodity specific effect represents unobserved heterogeneity in tastes across households for the different commodities. Hence, the model captures the income effect through the explained part, it captures the taste effect through the household - and commodity specific effect, and it controls for demographic composition.

First, we perform the Hausman specification test for each of the commodities in order to assess whether the household -and commodity specific effects are correlated with the explanatory variables, among those which is total expenditure. We find that for most commodities this is the case. Further analysis suggests that the household -and commodity specific effects are indeed correlated with whether a household is on average rich or poor, i.e. there is a taste effect. Secondly, we find that there is an income effect for all commodities except for the commodity group alcohol & tobacco. Thus, the overall conclusion from these tests is that the variation in budget shares seems to be due to a combination of the income effect and the taste effect. Comparisons of the income elasticity estimates from the panel data model with the income elasticity estimates obtained from estimating a cross-sectional model on the same data indicates that the presence of the taste effect causes a bias in the cross-sectional estimates.

To the best of my knowledge, there are only few empirical papers in the literature[4] dealing with the correlation between income and preferences .

Aasness, Biørn & Skjerpen (1993) considers consumer demands on panel data. They estimate a fully intagrable demand system on a Norwegian panel which covers 2 years with yearly information and find that preferences are correlated with income[5]. However, they assume that Engel curves are linear in income which might be a problem when testing for the correlation between

---

expenditure information on this many commodities.

[4]There is a larger literature on random utility models, i.e. models in which income and preferences are assumed to be independent, e.g. Brown & Matzkin (1998).

[5]Thus, their panel is much shorter that our panel. The ideal panel data set would have the length of 30-40 years such that it contains all the changes in the households' income.

preferences and income. We return to this issue of choice of functional form in Section 2.

There is one study considering the issue of taste and its correlation with income in a cross-sectional framework, namely Calvet & Comon (2000). That paper uses a functional form which is linear in log total expenditure and specifies taste-heterogeneity in a special way. By doing this, identification of both the income effect and the taste effect is achieved. The empirical finding of the paper is that almost all the variation in budget shares is due to the taste effect. We will estimate a linear model on our data in order to compare our conclusions with those in Calvet & Comon (2000).

The final step that will complete this preliminary version of the paper consists in examining whether the correlation between taste and income can be explained by means of observable variables. The hope is that this may lead to suggestions of ways to correct for the bias caused by the taste effect in cross-sectional models.

The paper is organised as follows. In Section 2 we present the model and the estimation method we use. In Section 3 we describe the data set and in Section 4 we present the empirical results. Section 5 concludes and discusses further work.

2 Model and Estimation Method

In this section we first present the model and the way we estimate it. Then we link our analysis to that of Calvet & Comon (2000) and describe why some of their findings can be tested in our model.

We model the demand system of budget shares for the N commodities as a Seemingly Unrelated Regressions panel data model with household - and commodity-specific effects, which are constant over time:

$$ w_{iht} = \alpha_i + \beta_i \left(\frac{\ln x_{ht}}{\ln P_t^*} \right) + \gamma_i \left(\frac{\ln x_{ht}}{\ln P_t^*} \right)^2 + z'_{iht} \delta_i + \eta_{ih} + \varepsilon_{iht}, \qquad (1) $$

where i denotes commodities ($i = 1, ..., N$), h denotes households ($h = 1, ..., H$) t denotes time ($t = 1, ..., T$), and where

$$
\begin{aligned}
w_{iht} &= \text{budget share for commodity } i \text{ for household } h \text{ at time } t\,[6] \\
x_{ht} &= \text{total expenditure for household } h \text{ at time } t \\
z_{iht} &= K\text{-dimensional observable vector of demographic variables}
\end{aligned}
$$

4

for household h at time t as well as the vector of the prices of the N different commodities

$\alpha_i, \beta_i, \delta_i$ = parameters

η_{ih} = household-specific effect for household h relating to commodity i

ε_{iht} = error term for commodity i for household h at time t

and where P_t^* is a Stone price index given by

$$\ln P_t^* = \sum_{i=1}^{N} \overline{w}_{it} \ln p_{it}, \qquad t = 1, ..., T,$$

where $\overline{w}_{it}$ denotes the average budget share for commodity i at time t, i.e. $\overline{w}_{it} = (1/H) \sum_h w_{iht}$.

Since the budget shares add up to 1 by construction ($\sum_{i=1}^{N} w_{iht} = 1$ for each h, t) also the right hand side must add up to one (for each h, t). This places the following linear restrictions on the model,

$$\sum_{i=1}^{N} \alpha_i = 1, \qquad \sum_{i=1}^{N} \beta_i = 0, \qquad \sum_{i=1}^{N} \gamma_i = 0,$$

$$\sum_{i=1}^{N} \delta_{ik} = 0, \qquad k = 1, ..., K$$

$$\sum_{i=1}^{N} \eta_{ih} = 0, \qquad \sum_{i=1}^{N} \varepsilon_{iht} = 0.$$

To get identification of the intercepts α_i's and the household -and commodity-specific effects, we furthermore need that

$$\sum_{h=1}^{H} \eta_{ih} = 0.$$

The first set of restrictions will be satisfied by leaving out one of the N equations, and the last restriction will be satisfied by construction.

With this model we have a specification of the budget shares which captures the income effect through the term $\alpha_i + \beta_i \left(\frac{\ln x_{ht}}{\ln P_t^*} \right) + \gamma_i \left(\frac{\ln x_{ht}}{\ln P_t^*} \right)^2 + z'_{iht} \delta_i$ and which captures the taste effect through the η_{ih}'s.

5

We will refer to the term capturing the income effect as an QUAID-type specification, since it is the QUAID system [10] with the original price index $\ln P_t$ replaced by the Stone price index $\ln P_t^*$, where

$$\ln P_t = \alpha_0 + \sum_{j=1}^{N} \alpha_j \ln p_{jt} + \frac{1}{2} \sum_{j=1}^{N} \sum_{k=1}^{N} \gamma_{jk} \ln p_{jt} \ln p_{kt}, \qquad t = 1, ..., T$$

By using the Stone price index instead of the original price index we get a model which is linear in its parameters and thus far easier to estimate. In our model commodity i is a luxury at total expenditure level $\ln x$ if and only if $(\beta_i + 2\gamma_i \ln x)/w_i > 0$ [11].

The specification which leaves out the quadratic term, i.e.

$$w_{iht} = \alpha_i + \beta_i \left(\frac{\ln x_{ht}}{\ln P_t^*} \right) + z_{iht}' \delta_i,$$

we will refer to as an AID-type specification, since it is the AID system [12] with the original price index $\ln P_t$ replaced by the Stone price index. In the model with an AID-type specification, commodity i is a luxury if and only if β_i, i.e. the income effect is the same at all levels of total expenditure.

The QUAID-type specification allows different slopes (i.e. different income effects) at different levels of total expenditure, which the AID-type specificationd does not. The reason we choose the QUAID-type specification is that misspecifying the model by using a linear specification might drive the conclusion of the Hausman test to be the rejection of the random effects model. To see this why misspecification can drive the conclusion of the Hausman test, consider the following example. Suppose we have two households, one with high total expenditure and the other with low total expenditure, and suppose that thre is great variation in the total expenditure for the rich households but almost no variation in total expenditure for the poor household. Then suppose the true specification is an QUAID-type, but we misspecify the model by using an AID-type. The linear AID-type fits one straight line through the data and since the estimator which exploits the

---

[10] The QUAID system was introduced in Banks, Blundell & Lewbel (1997).

[11] Letting q_i denote the quantity of commodity i we have that $\frac{\partial w_i}{\partial \ln x} = \frac{\partial (p_i q_i / x)}{(\partial x)/x} = \frac{\partial (p_i q_i / x) x}{(\partial x)} = \frac{(p_i (\partial q_i / \partial x) x - p_i q_i) x}{x^2} = p_i \left(\frac{\partial q_i}{\partial x} \right) - \frac{p_i q_i}{x}$, so $\frac{1}{w_i} \frac{\partial w_i}{\partial \ln x} = \frac{x}{p_i q_i} \left(\frac{\partial q_i}{\partial x} \right) - 1$, i.e. $e_i = \frac{1}{w_i} \frac{\partial w_i}{\partial \ln x} + 1$.

[12] The AID system was introduced in Deaton & Muellbauer (1980b).

panel structure (i.e. the within variation) gives much more weight to the rich hushehold than to the poor household whereas the OLS weighs all data points equally, the distance between these two estimators becomes big because of the linear structure. And since the Hausman test basically compares the fixed effects model (estimated only with the use of within variation) with the random effects model (which is basically OLS) and judge in favor of the fixed effects model if the distance between the two estimators is big, a rejection of the random effects model is due to the misspecification.

A remark on integrability is in place here. The demand system specified in (1) is not integrable as it stands. If the integrability requirement we want to impose is that the individual budget share functions $w_{iht}(x_{ht}, p_t, \eta_h) = \alpha_i + \beta_i \left(\frac{\ln x_{ht}}{\ln P_t^*} \right) + z'_{iht} \delta_i + \eta_{ih}$ as well as the conditional mean value budget share function, $E[w \,|p, x, z]$ both are integrable, then necessary and sufficient conditions are found in Lewbel (2001). A necessary condition for integrability is that the individual-specific terms η_{ih} are functionally dependent on prices or income.

We consider three econometric specifications of the basic SUR model (1). The fixed effects model, which assumes that the household - and commodity specific effects are correlated with the explanatory variables and thus can be considered as fixed parameters; the random effects model which assumes that the household - and commodity specific effects are uncorrelated with the explanatory variables and thus must be considered as random variables; and the pooled regression model which ignores the panel aspect and treats all observations as drawn from different households. The fixed effects model is given by (1) with error term structure

$$
\begin{aligned}
E[\varepsilon_{iht}] &= 0 \\
E[\varepsilon_{iht}\varepsilon_{jks}] &= \begin{cases} \rho_{ij}\sigma_\varepsilon^2, & i \neq j, \ h = k, \ t = s \\ 0, & \text{otherwise} \end{cases}
\end{aligned}
$$

where ρ_{ij} denotes the correlation through the error terms between equation i and equation j. Letting Σ denote the $(N-1) \times (N-1)$ covariance-matrix of the correlations between the equations and letting ε denote the $(N-1)HT$-dimensional vector, we can write the covariance-matrix in matrix form, using the Kronecker product, as

$$
E[\varepsilon\varepsilon'] = \Sigma \otimes \ \sigma_\varepsilon^2 \left(I_H \otimes I_T \right), \tag{2}
$$

where I_H denotes the $H \times H$ identity matrix.

The random effects model is given by (1) with error term structure

$$
\begin{aligned}
E[\eta_{ih}] &= 0 \\
E[\eta_{ih}\eta_{jk}] &= \left\{ \begin{array}{ll} \nu_{ij}\sigma_\varepsilon^2, & i \neq j,\ h = k \\ 0, & \text{otherwise} \end{array} \right. \\
E[\varepsilon_{iht}] &= 0 \\
E[\varepsilon_{iht}\varepsilon_{jks}] &= \left\{ \begin{array}{ll} \rho_{ij}\sigma_\varepsilon^2, & i \neq j,\ h = k,\ t = s \\ 0, & \text{otherwise} \end{array} \right.
\end{aligned}
$$

i.e. in matrix form

$$
\Omega = V \otimes \sigma_\eta^2 \left(I_H \otimes J_T \right) + \Sigma \otimes \sigma_\varepsilon^2 \left(I_H \otimes I_T \right),
$$

where V denotes the correlation between equations through the η-term, and J_T is a $T \times T$-matrix of ones.

Since the model is a SUR-model with identical regressors in the different equations, we can apply Zellner's result to yield that both models are consistently and efficiently estimated by equation-by-equation estimation [13]. The random effects model is estimated by GLS, and the fixed effects model is estimated using the within-estimator, both models equation by equation.

In Calvet & Comon (2000), the budget share equations are given as Grandmont's homothetic α-transformation of the AID system. This yields budget share equations

$$
\begin{aligned}
w_{ih} = {} & \alpha_i + \beta_i \ln x_h + \sum_{j=1}^N \gamma_{ij} \ln p_j \\
& - \beta_i \left\{ \alpha_0 + \sum_{k=1}^N \alpha_k \ln p_k + \frac{1}{2} \sum_{j=1}^N \sum_{k=1}^N \gamma_{kj} \ln p_k \ln p_j \right\} + \beta_i \eta_h,
\end{aligned}
$$

where the term $\beta_i \eta_h$ represents unobserved individual taste-heterogeneity. The heterogeneity scheme specified in Calvet & Comon (2000) can be interpreted in the following way: If $\beta_i > 0$ then commodity i is a luxury, so $\eta_h > 0$ implies that household h likes all luxuries more than the average household

---

[13] See proof in Appendix A.1.

and likes all necessities less than the average household. This heterogeneity scheme thus allows only two different kinds of taste: Either a household likes all luxuries more than the average household ($\beta_i > 0$), or a household likes all luxuries less than the average household ($\beta_i < 0$). Thus it restricts taste to have a common linear structure across commodities where the linear structure is given by the directions of the income effects for the commodities. By estimating (1), leaving out the quadratic income term, we should be able to establish whether our data supports the conclusion in Calvet & Comon (2000) that almost all variation in budget shares is caused by the taste effect [14].

3 Data

The data we use is the Spanish Permanent Survey of Consumption (Encuesta Permanente de Consumo, hereafter EPC) carried out by the Spanish National Bureau of Statistics (Instituto nacional de Estadistica) in the period 1978-1983. The data set is an unbalanced panel with quarterly information and it thus covers 24 periods. Some households drop out of the survey, and a part of the sample is renewed each period. The data set contains information on consumption expenditures and demographic characteristics of each of the households as well as the prices of the commodities in each period. The commodity groups are: food eaten at home, food eaten outside home, alcohol & tobacco, clothing, rent, energy, services, medical expenses, transportation, leisure, education, durables and other, i.e. a total of 13 different commodity groups [15]. All households consist of married couples, most of them with children or an additional number of people in other age groups living in the household. The household identifier is the husband. The household characteristics recorded in the data set are the family size, the number of people in different age groups living in the household, the husband's age and the wife's age, the family's housing tenure of the first and an eventually

---

[14] Also note that the model (1) with the general heterogeneity scheme η_{ih} is *not* identifed on repeated cross-sections, but with $\eta_{ih} = \beta_i \eta_h$ it is: First sum the budget share equations over households, then the heterogeneity term cancels out, exactly because it is decomposed into a commodity-specific component and household-specific component. Therefore, a test of $\eta_{ih} = \beta_i \eta_h$ is also a test of the assumption that identifies the model on repeated cross-sections.

[15] For the commodity group "food eaten outside home" there is no price recorded, so we use the price for "food eaten at home".

second house, the husband's education level, his employment status and his type of employment. For a detailed description of the demographic variables as well as the variables constructed from these, see table A2.1 and A2.2 in Appendix A.2.

The rest of this section contains description and summary statistics of the selected sample. All tables are placed in Appendix A.2.

For our empirical analysis we select a sample consisting of the observations where the husband is employed as a wage earner with a permanent job [16]. We do this to avoid issues concerning the assumption of separability between the consumption of goods and the labor supply, which there seems to be empirical evidence against (see e.g. Browning & Meghir (1991) which presents empirical evidence of rejection of separability). One can easily imagine examples where the separability assumption is violated: if a person goes to work, this person might have higher transportation costs and need more clothes than a person not going to work. On the other hand, a person not working might stay more at home and therefore use more energy for heating. By including only the households where the husband is employed (and employed in a certain type of job) we do not have to worry about modelling the labor supple. When selecting the observations where the husband is a wage earner with a permanent job, we get that 64 households only appear once in the sample. These 64 households will therefore not have any effect on the estimates, hence we trim them. Thus, our final sample is an unbalanced panel of 909 households with total number of periods in the sample varying between 2 and 24 periods, yielding a total of 10674 observations [17]. We expect the sample selection bias arising from this selection to be in the direction of less individual heterogeneity, since we have selected a sample which is more homogenous than the original data set. Table A2.3 shows the distribution of the households in our sample on the total number of periods in which the husband is employed as a wage earner with a permanent job. Almost half of the households in the sample are observed more than 11 times covering a time span of almost three years. Table A2.4 shows that the observations seem to be more or less evenly spread over the time period.

Next, we present some summary statistics for the household characteristics in our sample.

---

[16] Referring to the definition of demographics, we thus select the observations with hempl=1 and hgact=3.

[17] Note, that the observations in our sample for a given household need not to be consecutive.

The household size[18] varies between 2 and 10 people, 2 meaning that the household consists only of the husband and the wife. The average size of a household is 4.3 persons. Table A2.5 shows the distribution of average household size. As can be seen from the table, 75 households consist only of the husband and the wife for the entire survey period, while 163 households are on average of the size 2-3 persons, 301 households are on average of the size 3-4 persons and so on. Households of between 2 and 4 persons on average make up more than 70% of the sample, and households of more than 6 persons on average only account for around 6% of the households. Thus, the very large households are not so numerous in our sample. The average number of children (age 0-17) living in a household is 1.6, and the average number of adults (age 18 or above) is 0.7[19]. Tables A.26 and A2.7 show the distribution of the average number of children and the distribution of the average number of adults. The tables show that almost 80% of the households in the sample have children living in the household at some point in time during the survey period, whereas the corresponding number for adults living in the household is around 50%. Thus, it is more common that households have children living in the household than they have adults living in the households. The mean age of the husband in our sample is 46.3 years, and the mean age of the wife is 43.9 years. With respect to housing tenure, then 558 of the households are home owners in all periods, 122 households are home owners for some, but not all, periods and 229 households are never home owners, but rent or have free accomodation[20]. With respect to education level of the husband, then for 847 of the households (corresponding to 93% of the sample) the husband's education level is low, i.e. for the vast majority of our sample the husband is lowly skilled[21] and thus the sample is very homogenous with respect to this household characteristic.

Finally, we consider the different commodity groups and the total expenditure. We leave out the commodity groups rent, durables and other. Both

---

[18]We define household size as famsize+2.

[19]We define the variable "children" as age02+age36+age713+age1415+age1617 and the variable "adult" as age1824+age2564+age65.

[20]Almost no households have free accomodation, so not being a home owner almost always means being a renter.

[21]There are many obvious errors in the records of the variable heduc in the sense that the education level often changes for the same household. We only use the constructed varible Dheduc and therefore we only correct this variable. For the households recording changes Dheduc, we choose the education level reported most times.

rent and durables are a different kind of commodity than the rest of the commodities in that none of them are flexible on the short sight: if the rent changes one typically adjusts one's expenses for rent not until after at least some periods (has to find a new house and move). Similarly with durables, then a durable good is more considered as an investment. This leaves us with ten commodities. The table below gives an impression of which commodity groups that account for a large share of the total expenditure and which account for only a small share of the total expenditure on average:

Commodity group	Mean budget share
Food at home	0.503
Foodout	0.084
Alcohol & tobacco	0.042
Clothing	0.111
Energy	0.094
Services	0.024
Medication	0.022
Leisure	0.037
Education	0.034
Transportation	0.049
Total	1.000

Food at home is clearly the biggest commodity group, with foodout, clothing and energy coming second in size.

Since a great share of zeroes in the expenditures for a given commodity group will bias the estimates for this commodity group, we find the share of zeroes in consumption expenditures to see if we face this particular risk of bias:

Commodity group	Share of zeroes
Food at home	0.0018
Foodout	0.0947
Alcohol & tobacco	0.0880
Clothing	0.0111
Energy	0.0032
Services	0.0101
Medication	0.1574
Leisure	0.1201
Education	0.3284
Transportation	0.0944

As can be seen from the table, education, medication and leisure are the commodity groups with the largest share of zeroes with education being the largest in that almost one third of the sample does not spend income on education. This indicates that there might be a problem for the three commodity groups mentioned[22].

Estimators in fixed effects models are sensitive to how the explanatory variables vary within the individual. The explanatory variable which we are mainly interested in is total expenditure and we therefore examine how total expenditure varies within a household[23]. The distribution of the coefficient of variation (the standard deviation divided by the mean) in deflated total expenditure is:

	Percentiles
1%	.0612
5%	.0924
10%	.1105
25%	.150
50%	.2079
75%	.2681
90%	.3511
95%	.4204
99%	.6067

and the mean of this distribution is .2246. This means that the percentage variation relative to the mean of total expenditure within the households is on average 22%. This shows that there seems to variation in total expenditure within households, and the average variation seems too high to be only measurement error.

4   Empirical analysis

In this section we present the empirical results. First we perform the Hausman specification test for each commodity in order to assess whether the household -and commodity specific effects are correlated with the explana-

---

[22]A way to solve this would be to switch to a tobit-type model. This is another project.

[23]For a table of the distribution of mean total expenditure across households, see **Table A2.9**.

tory variables[24]. We find that for most commodities the random effects model is rejected. Having established this, the question is if the fixed effects are indeed correlated with total expenditure or only with the other explanatory variables. We examine this by regressing the fixed effects on the set of the means of the explanatory variables, to see if there is an effect from mean total expenditure. We find that mean total expenditure is highly significant. We discuss the implications for the bias of results obtained in cross-sectional models which assume that taste-heterogeneity is random. Secondly, we examine if the income effect is present; we find that it is. Finally we do the Hausman tests and the comparison of the fixed effects model with the cross-sectional model, using the AID-type specification in order to compare our conclusion with the conclusion in Calvet & Comon (2000).

We estimate the model (1) as described in Section 2 for the commodities food at home, foodout, alcohol & tobacco, clothing, energy, services, medication, leisure, education and transportation; as explanatory variables we use deflated log total expenditure (deflated with the Stone price index), the prices of the ten commodities, Famsize, Dheduc, Dtenure and quarterly dummies. This amounts to fifteen explanatory variables[25].

First we carry out the Hausman specification tests. We do this for the entire sample, as well as for subsamples where we trim the upper 1% percentile, respectively the upper 5% percentile, of households that have the highest variation in deflated total expenditure (within a household)[26]. We do this to make sure that our results are not driven by only 1% or 5% of the sample, since our estimation method is very sensitive to variations within households. The results of the Hausman tests for the three samples are reported in Table 4.1[27]. We test if there is an income effect present in the data by testing the joint significance of $(\ln x/P^*)$ and $(\ln x/P^*)^2$. These test results are also reported in Table 4.1. Finalle we test if the quadratic total expenditure term is significant. These test results are reported in Table 4.2.

Table 4.1: Hausman test results, QUAID-type specification

---

[24]The Hausman specification test is a test for the random effects model against the fixed effects model. The test exploits that under the null of random effects, both the fixed effects estimator and the random effects estimator is consistent, but only the random effects estimator is efficient.

[25]Here, one would expect total expenditure to be endogenous. We discuss an instrumental variables approach to this problem in Section 5.

[26]Sample size after trimming 1% (5%) is 899 (863) households.

[27]Table reports $\chi^2_{(15)}$-statistics, and numbers in parentheses are test probabilities.

Commodity group	Test of RE against FE, whole sample	Test of RE against FE, 99% incl	Test of RE against FE, 95% incl	Test of income effect
Food at home	146.5 (0)	143.86 (0)	137.27 (0)	sign
Foodout	61.29 (0)	43.51 (.0002)	85.88 (0)	sign
Alcohol & tobacco	196.05 (0)	76.1 (0)	112.62 (0)	not sign
Clothing	80.13 (0)	39.2 (.0010)	83.38 (0)	sign
Energy	28.59 (.0268)	30.14 (.0173)	57.99 (0)	sign
Services	98.24 (0)	90.05 (0)	73.87 (0)	sign
Medication	13.47 (.6380)	14.33 (.5739)	17.67 (.3435)	sign
Leisure	73.68 (0)	75.57 (0)	118.75 (0)	sign
Education	94.86 (0)	97.29 (0)	90.82 (0)	sign
Transportation	36.3 (.0016)	36.55 (.0024)	43.64 (.0002)	sign

As **Table 4.1** shows, leaving out the households with most variation in deflated total expenditure from the estimations do not change the results much: On a 1% significance level, the fixed effects model is the correct specification for all commodities except medication in all three samples and except for energy in the two larger samples [28]. On a 5% significance level, the fixed effects model is the correct specification for all commodities except medication in all three samples. Since the conclusions of the Hausman tests are so similar for the three samples on a 1% significance level and identical on a 5% significance level, it seems like the estimation results are not driven by a few households with high variation in deflated total expenditure and thus we include all 909 households in the further analysis.

---

[28]The F-test of all individual-specific effects being zero is rejected in all cases and therefore we do not report test sizes for this test.

Table 4.2: Test results of t-tests of significance of the quadratic term:

Commodity	All observations included	99% included	95% included
Food at home	68.86 (0)	86.19 (0)	60.80 (0)
Foodout	68.14 (0)	32.46 (0)	30.67 (0)
Alcohol & tobacco	3.69 (.0547)	3.55 (.0597)	5.68 (.0172)
Clothing	5.79 (.0161)	12.34 (.0004)	7.8 (.0052)
Energy	.75 [29] (.3855)	6.83 [30] (.0090)	2.72 [31] (.0994)
Services	8.68 (.0032)	13.39 (.0003)	8.04 (.0046)
Medication	1.51 [32] (.2197)	6.67 [33] (.0098)	5.69 [34] (.0171)
Transportation	4.29 (.0383)	.97 (.3245)	1.54 (.2151)
Leisure	2.91 (.0879)	0 (.9817)	.46 (.4993)
Education	3.64 (.0564)	.74 (.3901)	2.85 (.0915)

This table shows that the quadratic term is significant for food at home, foodout and services in all three samples. Furthermore, it is significant at a 5% significance level for clothing and transportation when all observations are included. It thus seems justified to include the quadratic term.

From the Hausman tests we conclude that the fixed effects model is the correct specification except for energy and medication (only for medication at a 5% significance level). This means that there is unobservable individual-specific heterogeneity which is correlated with the set of explanatory variables

---

[29] REmodel: $\chi^2(1)$-test: 1.16 (.2811)
[30] REmodel: $\chi^2(1)$-test: 9.02 (.0027)
[31] REmodel: $\chi^2(1)$-test: .66 (.0309)
[32] REmodel: $\chi^2(1)$-test: 1.47 (.2259)
[33] REmodel: $\chi^2(1)$-test: 7.24 (.0071)
[34] REmodel: $\chi^2(1)$-test: 5.14 (.0234)

present in the sample for all commodities except clothing, energy and medication. This strongly suggests that there is a taste effect present in the data set for these seven commodities. However, the Hausman tests only show that the household-specific effects are correlated with all the explanatory variables, among which is deflated total expenditure. We therefore calculate the fixed effects and regress them on mean deflated log total expenditure together with the other observables to see which are significant for the fixed effects. Results are in the table below ("yes" means that the variable is significant on a 1% significance level):

Table 4.3

Fixed effect for commodity group	Mean total expenditure	Mean family size	Dheduc	Dtenure
Food at home	yes	yes	yes	no
Foodout	no	yes	no	no
Alcohol & tobacco	yes	no	no	no
Clothing	yes	yes	yes	no
Energy	yes	yes	no	no
Services	yes	yes	yes	no
Medication	yes	yes	no	no
Leisure	yes	yes	no	no
Education	yes	no	yes	no
Transportation	yes	yes	no	no

This shows that mean deflated log total expenditure is significant for all commodities, except foodout, i.e. whether a household is rich or poor on average seems to have some explanatory power in explaining the taste-heterogeneity. In other words, this indicates that there is a taste effect present in the sample.

So, the overall conclusion of this empirical analysis is that there seems to both a taste effect and an income effect present in the sample. For some commodities, both effects are present, for other commodities only one effect is present. This further indicates reason to be worried about the bias this correlation between taste and income might create in cross-sectional consumer demand models.

For our sample, we try to get an idea how this biases the estimates of the income effects one would get from a cross-sectional data set by estimating the

pooled model (i.e. ignoring the panel aspect). Estimating the pooled model by OLS, correcting for heteroscedasticity, is as far as I can see the closest I can get to what would have been done had the data set been cross-sectional instead of a panel[35]. In the table below is reported the estimates of the income elasticities minus 1 evaluated at the mean observed budget share and the mean total expenditure:

Table 4.4:

Commmodity group	FE-model	Cross-sectional model
Food at home	-.0742	-.1967
Foodout	.5067	.5079
Alcohol & tobacco	.0461	-.1027
Clothing	.0808	-.0272
Energy	-.0954	.0706
Services	-.6151	-.0814
Medication	.1626	.3145
Transportation	.1727	.3667
Leisure	.1288	.4704
Education	-.2682	.4878

Thus, a commodity is a luxury if the reported number is positive and a necessity if the reported number is negative. In the fixed effects model we get that food at home, energy, services and education are necessities, while in the cross-sectional model we get that food at home, alcohol & tobacco, clothing and services are necessities. In other words, taking account of the taste effect reverses four commodities to being a luxury (necessity) if the commodity was a necessity (luxury) in the cross-sectional model. Thus, the cross-sectional estimators are biased.

Therefore, an interesting question is whether it is possible to model some of the correlation between taste and total expenditure by means of observables. We do some informal examinations. From the table with results from the regressions of the fixed effects on all explanatory variables (**Table 4.3**) it seems that the variable family size has the greatest explanatory power, while the other observable variables in our sample do not seem to explain the fixed effects very well. We therefore regress the fixed effects on mean family size,

---

[35]We use the correction of the standard errors to allow for heteroscedasticity, in order to avoid the assumption that all observations are iid, since we know this is not true.

stratifying on different types of families in order to control for different family composition. We construct four groups of family types. First recall the definitions of children as people aged 0-17 years incl, and the definition of adults as people aged 18 and above. Then we define one type by families which only have children living in the household, another type by families only having adults living in the household, a third type as households having both children and adults living in the household and then a residual group of the households which consist of only the man and the wife. The results of the regression are reported in the table below. The numbers reported are R^2 of the regression of the fixed effects on mean family size for each group:

Fixed effect for commodity group	Households only with children[36]	Households only with adults[37]	Households with both children and adults[38]
Food at home	.1399	.0053	.1463
Foodout	.0863	.0003	.0155
Alcohol & tobacco	.0009	.0033	.0019
Clothing	.0015	.0103	.0022
Energy	.0514	0	.0134
Services	.0003	0	.0029
Medication	.0055	.0462	.0281
Leisure	.1030	.0018	.1101
Education	.0004	.0012	.0614
Transportation	.0095	.0413	.0007

This indicates that family size unfortunately only has little explanatory power.

Next we turn to a model with an AID-type specification in order to compare conclusions with Calvet & Comon (2000). The result of the Hausman test for the three samples are reported below:

---

[36] There are 349 households in this group.

[37] There are 99 households in this group.

[38] There are 171 households in this group.

Table 4.5: Hausman test results, AID-type specification

Commodity group	Test of RE against FE, whole sample[39]	Test of RE against FE, 99%	Test of RE against FE, 95%
Food at home	131.4 (0)	139.0 (0)	128.5 (0)
Foodout	37.9 (.0009)	38.0 (0)	43.5 (.0001)
Alcohol & tobacco	23.0 (.0849)	76.1 (0)	38.1 (.0009)
Clothing	35.8 (.0019)	21.8 (.1123)	30.5 (.0102)
Energy	28.4 (.0191)	29.2 (.0154)	23.5 (.0738)
Services	87.1 (0)	81.0 (0)	67.1 (0)
Medication	13.5 (.5655)	13.8 (.5404)	18.8 (.2242)
Leisure	73.3 (0)	76.2 (0)	132.5 (0)
Education	92.6 (0)	93.0 (0)	84.7 (0)
Transportation	36.3 (.0016)	36.0 (.0018)	31.8 (.0069)

As the table shows, leaving out the households with most variation in deflated total expenditure from the estimations also inthis does not change the results much: On a 1% significance level, the fixed effects model is the correct specification for all commodities except clothing, energy and medication in all three samples [40]. Since the conclusions of the Hausman tests are exactly the same for the three samples on a 1% significance level, we need not worry if the estimation results are driven by a few households with high variation in deflated total expenditure. The Hausman test concludes that the fixed effects model is the correct specification except for clothing, energy and

---

[39] RE stands for random effects model and FE stands for fixed effects model.

[40] The F-test of all individual-specific effects being zero is rejected in all cases and therefore we do not report test sizes for this test.

medication. Next we look at the income effects for the AID-type model. The estimates from the fixed effects model as well as the estimates from running OLS with robust standard errors are reported in the table below:

Table 4.6: Estimates, AID-type specification.

Commodity group	$\widehat{\beta}^{\text{fixed effect}}$	t-statistic	$\widehat{\beta}^{\text{pooled}}$	t-statistic
Food at home	-.03508 (.00418)	-8.398 (0)	-.09799 (.00380)	-25.012 (0)
Foodout	.04109 (.00257)	15.982 (0)	.04199 (.00221)	19.0 (0)
Alcohol & tobacco	.00177 (.00117)	1.522 (.128)	-.00446 (.00095)	-4.680 (0)
Clothing	.00857 [41] (.00273)	3.134 (.002)	-.00255 (.00202)	-1.258 (.208)
Energy	-.00905 [42] (.00224)	-4.047 (0)	.00632 (.00174)	3.621 (0)
Services	-.01490 (.00066)	-22.534 (0)	-.00236 (.00066)	-3.581 (0)
Medication	.00343 [43] (.00143)	2.393 (.017)	.00676 (.00106)	6.375 (0)
Leisure	.00494 (.00139)	3.543 (0)	.01757 (.00092)	19.0 (0)
Education	-.00899 (.00144)	-6.232 (0)	.01663 (.00119)	13.92 (0)
Transportation	.00823 (.00196)	4.191 (0)	.01810 (.00146)	12.4 (0)

This table shows that the income effect is significant for all commodities. Also, comparing the estimates for the two models, we find that there are differences. Clothing and alcohol & tobacco are luxuries in the fixed effects model, but neccessities in the pooled model and vice versa for energy and education. Except for foodout, the numerical values of the two estimates seem to differ.

The overall conclusion from estimating the AID-type model is that we do

---

[41] Random effects estimate: 0.00356 (.00220).
[42] Random effects estimate: -0.00500 (.00202).
[43] Random effects estimate: 0.00522 (.00118).

not reach the same conclusions as did Calvet & Comon (2000). They found that almost all variation was due to the taste effect; we found that there is an income effect for all commodities and a taste effect for some commodities.

5   Discussion and Further Remarks

In this paper we estimated a system of consumer budget share equations on a panel data set of consumption expenditures, thus taking account of individual heterogeneity in tastes. We found indications that the variation in budget shares with income is caused by a combination of a the income effect and the taste effect. Preliminary results indicate that the taste effect, i.e. the correlation between income and tastes can not be well explained by the observable variables in our sample, and that the taste effect causes bias in cross-sectional estimates of the income effect. Therefore, it would be interesting to see if it is possible to explain some of the taste effect by observable variables and thereby construct ways to correct for this bias in cross-sectional models. This will probably involve some difficulties with respect to identification, since the observable variables will already be included in the empirical specification.

   As mentioned earlier, the empirical analysis in this preliminary version of the paper does not deal with the issue of possible measurement error in explanatory variables. But especially total expenditure is very likely to be endogenous. Errors of measurement can, when exploiting the within variation in data for estimation, result in even more biased estimators than OLS estimators using cross-sectional data alone. And since the Hausman test basically compares the fixed effects model (estimated only with the use of within variation) with the random effects model (which is basically OLS) and judge in favor of the fixed effects model if the distance between the two estimators is big, then measurement error might drive the conclusion of the Hausman test to be that the appropriate model is a fixed effects model, when in fact it is not. Therefore we want to estimate our model in a way which takes account of measurement error in total expenditure in order to exclude the possibility that the conclusion about whether there is a taste effect or not is not driven by measurement error. In the following we propose how to do instrumental variables (IV) estimation in order to deal with this problem [44].

---

[44]This discussion only concerns iid measurement error.

To illustrate the idea, suppose for simplicity that the equation of interest is

$$y_{it} = X_{it}\beta + u_{it}, \qquad i = 1, ..., n, \ t = 1, ..., T. \qquad (3)$$

We want to test if u is correlated with X and if so, we want to estimate β in a way that takes this correlation into account. Suppose we have an instrument for X, then a way to test if the correlation is present is to do a residual augmented regression: First regress X on its instrument and calculate the residuals from this regression. Second, regress y on X and the residuals from the first stage. The t-test of the coefficient on the residuals being zero is then a test of exogeneity of X : The coefficient on the residuals should be insignificant if X is exogenous, i.e. if u and X are uncorrelated.

Now, split the error term u_{it} up in an individual-specific part η_i and the usual error term ε_{it} :

$$u_{it} = \eta_i + \varepsilon_{it} \qquad i = 1, ..., n, \ t = 1, ..., T. \qquad (4)$$

We want to find out if there is a taste effect, i.e. in terms of (3) and (4) to find out if $E\left[\eta \mid X\right] \neq 0$, and we want to exclude the possibility that the conclusion is driven by measurement error in total expenditure, i.e. we want to take account of the possibility that $E\left[\varepsilon \mid X\right] \neq 0$. Let the matrix A be the first-differencing transformation matrix, i.e. AX is the matrix of the first-differences of X :

$$A = \begin{bmatrix} -1 & 1 & 0 & & ... & 0 \\ 0 & -1 & 1 & 0 & ... & 0 \\ \vdots & & & & & \vdots \\ \vdots & & & & & \vdots \\ 0 & & ... & 0 & -1 & 1 \end{bmatrix}$$

Using the idea described above we propose the four IV estimations in the table below:

Model	Assumptions	Instruments
Model 1	$E\left[\varepsilon \mid X\right] = 0$ and $E\left[\eta \mid X\right] = 0$	X
Model 2	$E\left[\varepsilon \mid X\right] = 0$ and $E\left[\eta \mid X\right] \neq 0$	$A'(AA')^{-1}AX$
Model 3	$E\left[\varepsilon \mid X\right] \neq 0$ and $E\left[\eta \mid X\right] = 0$	X_{-1} [45]
Model 4	$E\left[\varepsilon \mid X\right] \neq 0$ and $E\left[\eta \mid X\right] \neq 0$	ΔX_{-1} [46]

---

[45] X_{-1} is the matrix of lagged levels of X.

[46] $\Delta X_{-1} = [X_{t-1} - X_{t-2}]_{t=1}^{T}$.

Under each of the assumptions listed about the correlation between X and the error terms, each of the listed instrumental variables is a valid instrument for X : $A'(AA')^{-1}AX$ is uncorrelated with η (since it consists of differences across time in X within individuals and thus it can by definition not be correlated with something which is time-invariant) but not uncorrelated with ε; X_{t-1} is uncorrelated with ε_t for each t and not uncorrelated with η; ΔX_{-1} is uncorrelated with both η and ε since it is both lagged and in differences. *Model 1* is the pooled regression model and the IV estimator is the OLS estimator; *Model 1* thus covers the case where there is no correlation between X and η, i.e. there is no taste effect, and no correlation between X and ε, i.e. there is also no measurement error. *Model 2* covers the case where there is a taste effect, but no measurement error. *Model 3* covers the case where there is measurement error, but no taste effect. *Model 4* covers the case where there is both a taste effect and measurement error. Thus, doing residual augmented regression in each of these models enables us to detect if there is measurement error and if so, take it into account when testing for a taste effect.

The reason for choosing the instrument $A'(AA')^{-1}AX$ instead of the more tractable AX in *Model 2* is that *Model 2* is the fixed effects model and with this instrument the IV estimator is the efficient fixed effects estimator. To see this, first note that a consistent and efficient estimator of β in *Model 2* is the within estimator [47]

$$\widehat{\beta}_W = (X'M_D X)^{-1} (X'M_D y).$$

Secondly, note that estimating β in *Model 2* by IV using the instruments $A'(AA')^{-1}AX$ yields the IV estimator

$$
\begin{aligned}
\widehat{\beta}_{IV} &= \left[\left(A'(AA')^{-1}AX\right)' X\right]^{-1} \left[\left(A'(AA')^{-1}AX\right)' y\right] \\
&= \left(X'A'(AA')^{-1}AX\right)^{-1} \left(X'A'(AA')^{-1}Ay\right).
\end{aligned}
$$

Thirdly, note that one can estimate β in *Model 2* efficiently by applying GLS on the first-differenced data (i.e. by premultiplying (3) with A and running GLS on the transformed model using covariance-matrix $V(A\varepsilon) = AA'$) which yields

$$\widehat{\beta}_{fd,GLS} = \left((AX)'(AA')^{-1}(AX)\right)^{-1} \left((AX)'(AA')^{-1}(Ay)\right)$$

---

[47] $M_D = I - D(D'D)^{-1}D$, where D is the matrix of the dummies for the individual-specific effects.

$$= \left(X'A'(AA')^{-1}AX\right)^{-1}\left(X'A'(AA')^{-1}Ay\right),$$

i.e. $\widehat{\beta}_{IV} = \widehat{\beta}_{fd,GLS}$. But since $\widehat{\beta}_{W}$ is algebraically identical to $\widehat{\beta}_{fd,GLS}$ [48], the circle is closed, and thus the estimator obtained by IV estimation with $A'(AA')^{-1}AX$ as instruments for X, is equivalent to the efficient fixed effects estimator [49].

A revised version of the paper will use these proposed IV estimators and thus take account of measurement errors.

References

[1] J. Aasness, E. Biørn & T. Skjerpen (1993): *Engel Functions, Panel Data and Latent Variables*, Econometrica, 61, 6, 1395-1422.

[2] J. Banks, R. Blundell & A. Lewbel (1997): *Quadratic Engel Curves and Consumer Demand*, The Review of Economics and Statistics, nov, 4, 527-539.

[3] D. Brown & R. Matzkin (1998): *Estimation of Nonparametric Functions in Simultaneous Equations Models, with an Application to Consumer Demand*, working paper, Northwestern University.

[4] B. Brown & M. Walker (1989): *The Random Utility Hypothesis and Inference in Demand Systems*, Econometrica, vol. 57, no. 4, 815-829.

[5] M. Browning & C. Meghir (1991): *The Effects of Male and Female Labor Supply on Commodity Demands*, Econometrica, vol. 59, no. 4, 925-951.

[6] L. Calvet & E. Comon (2000): *Behavioral Heterogeneity and the Income Effect*, Harvard Institute of Economic Research Discussion paper no. 1892

[7] A. Deaton & J. Muellbauer (1980a): *Economics and Consumer Behavior*, Cambridge University Press.

[8] A. Deaton & J. Muellbauer (1980b): *An Almost Ideal Demand System*, American Economic Review, vol. 70, no.3, 312-326.

---

[48] For proof, see Verbeek (1995).

[49] Note that the easier constructed $M_D X$ is not a valid instrument for X, since $M_D X$ is not uncorrelated with ε.

[9] J. M. Grandmont (1992): *Transformations of the Commodity Space, Behavioral Heterogeneity and the Aggregation Problem*, Journal of Economic Theory, vol. 57, 1-35.

[10] C. Hsiao (1986): *Analysis of Panel Data*, Econometric Society Monographs, Cambridge University Press.

[11] A. Lewbel (2001): *Demand Systems with and without Errors*, American Economic Review, vol. 91 no. 3, 611-618.

[12] M. Verbeek (1995): *Alternative Transformations to Eliminate Fixed Effects*, Econometric Reviews, 14(2), 205-211.

[13] A. Zellner (1962): *An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias*, Journal of American Statistical Association, vol. 57, 348-368.

A Appendix

A.1 Proof of system estimation being equivalent to equation by equation estimation in a SUR panel data model with identical regressors

Consider the fixed effects model

$$y_{iht} = x'_{iht}\beta_i + \eta_{ih} + \varepsilon_{iht}, \quad i = 1,...,N, h = 1,...,H, t = 1,...,T,$$

where $x_{iht} \in \mathbb{R}^K$. Stacking observations by commodities this model can be written

$$Y_i = X_i\beta_i + \eta_i + \varepsilon_i, \qquad i = 1,....,N, \tag{5}$$

where Y_i is $HT \times 1$-dimensional, X_i is $HT \times K$-dimensional, $\eta_i = (\eta_{i1}\iota_T,...,\eta_{iH}\iota_T)$ where $\iota_T = (1,...,1)'$ is of dimension T and ε_i is $HT \times 1$-dimensional. Stacking all the observations the model can be written

$$Y = X\beta + \eta + \varepsilon$$

with covariance-matrix $E[\varepsilon\varepsilon'] = \Sigma \otimes \sigma_\varepsilon^2 (I_H \otimes I_T)$, where Σ $N \times N$-dimensional is the covariance-matrix for the correlation between commodities.

We transform the model by a transformation that eliminates the fixed effects, i.e. we multiply the observations for each individual by a matrix A satisfying $A\iota_T = 0$. Hence, in (5) we multiply by $(I_H \otimes A)$. Then we have the model

$$\widetilde{Y}_i = \widetilde{X}_i \beta_i + \widetilde{\varepsilon}_i, \qquad i = 1, ..., N,$$

where $\widetilde{Y}_i = (I_H \otimes A) Y_i$, $\widetilde{X}_i = (I_H \otimes A) X_i$ and $\widetilde{\varepsilon}_i = (I_H \otimes A) \varepsilon_i$. The covariance-matrix for the transformed model is given by

$$E[\widetilde{\varepsilon}_i \widetilde{\varepsilon}_i'] = \sigma_\varepsilon^2 I_H \otimes B,$$

where B, $T \times T$-dimensional, is the covariance-matrix for the correlation within $\widetilde{\varepsilon}_{ih}$ arising from the transformation. It is not necessary to know B for the point we want to make here. Stacking all the observations, the transformed model can be written

$$\widetilde{Y} = \widetilde{X}\beta + \widetilde{\varepsilon}, \tag{6}$$

where $\widetilde{Y} = (I_{NH} \otimes A) Y$, $\widetilde{X} = (I_{NH} \otimes A) X$ and $\widetilde{\varepsilon} = (I_{NH} \otimes A) \varepsilon$ and the covariance-matrix is

$$E[\widetilde{\varepsilon}\widetilde{\varepsilon}'] = \sigma_\varepsilon^2 (\Sigma \otimes I_H \otimes B).$$

For notational convinience, we set $\Omega = \Sigma \otimes I_H \otimes B$, i.e. $E[\widetilde{\varepsilon}\widetilde{\varepsilon}'] = \sigma_\varepsilon^2 \Omega$ and we set $C = I_H \otimes B$, i.e. $E[\widetilde{\varepsilon}_i \widetilde{\varepsilon}_i'] = \sigma_\varepsilon^2 C$ and $\Omega = \Sigma \otimes C$. We can then write

$$\begin{aligned} C^{-1} &= I_H \otimes B \\ \Omega^{-1} &= \Sigma^{-1} \otimes C^{-1}, \end{aligned}$$

where

$$\Sigma = \begin{bmatrix} \rho_{11} & \cdots & \rho_{1N} \\ : & & : \\ \rho_{1N} & \cdots & \rho_{NN} \end{bmatrix}, \qquad \Sigma^{-1} = \begin{bmatrix} \rho^{11} & \cdots & \rho^{1N} \\ : & & : \\ \rho^{1N} & \cdots & \rho^{NN} \end{bmatrix}$$

Now note that GLS on a single equation i yields

$$\widehat{\beta}_i^{\text{single-GLS}} = \left[\widetilde{X}_i' C^{-1} \widetilde{X}_i \right]^{-1} \left[\widetilde{X}_i' C^{-1} \widetilde{Y}_i \right], \quad i = 1, ..., N.$$

Note that with identical regressors, i.e. $\widetilde{X}_i = \widetilde{X}_j$ for all i, j, the single-equation-estimator becomes

$$\widehat{\beta}_i^{\text{single-GLS}} = \left[\widetilde{X}' C^{-1} \widetilde{X} \right]^{-1} \left[\widetilde{X}' C^{-1} \widetilde{Y}_i \right], \quad i = 1, ..., N. \tag{7}$$

27

A consistent and efficient estimator of β for the system in (6) is given by

$$
\begin{aligned}
\beta &= \left[\widetilde{X}'\Omega^{-1}\widetilde{X}\right]^{-1}\left[\widetilde{X}'\Omega^{-1}\widetilde{Y}\right] \\
&= \left[\widetilde{X}'\left(\Sigma^{-1}\otimes C^{-1}\right)\widetilde{X}\right]^{-1}\left[\widetilde{X}'\left(\Sigma^{-1}\otimes C^{-1}\right)\widetilde{Y}\right] \\
&\quad \begin{bmatrix} \rho^{11}\widetilde{X}_1'C^{-1}\widetilde{X}_1 & \ldots & \rho^{1N}\widetilde{X}_1'C^{-1}\widetilde{X}_N \\ & \vdots & \\ \rho^{N1}\widetilde{X}_N'C^{-1}\widetilde{X}_1 & \ldots & \rho^{NN}\widetilde{X}_N'C^{-1}\widetilde{X}_N \end{bmatrix}^{-1} \begin{bmatrix} \sum_{j=1}^N \rho^{1j}\widetilde{X}_1'C^{-1}\widetilde{Y}_j \\ \vdots \\ \vdots \\ \sum_{j=1}^N \rho^{Nj}\widetilde{X}_N'C^{-1}\widetilde{Y}_j \end{bmatrix} \\
&= \begin{bmatrix} \rho^{11}\widetilde{X}'C^{-1}\widetilde{X} & \ldots & \rho^{1N}\widetilde{X}'C^{-1}\widetilde{X} \\ & \vdots & \\ \rho^{N1}\widetilde{X}'C^{-1}\widetilde{X} & \ldots & \rho^{NN}\widetilde{X}'C^{-1}\widetilde{X} \end{bmatrix}^{-1} \begin{bmatrix} \sum_{j=1}^N \rho^{1j}\widetilde{X}'C^{-1}\widetilde{Y}_j \\ \vdots \\ \vdots \\ \sum_{j=1}^N \rho^{Nj}\widetilde{X}'C^{-1}\widetilde{Y}_j \end{bmatrix},
\end{aligned}
$$

where the last equality follows from that regressors are identical. Now, from (7) it follows that

$$
\left[\widetilde{X}'C^{-1}\widetilde{Y}_i\right] = \left[\widetilde{X}'C^{-1}\widetilde{X}\right]\widehat{\beta}_i^{\text{single-GLS}},
$$

which inserted into the expression for β yields

$$
\begin{aligned}
\widehat{\beta}^{system} &= \begin{bmatrix} \rho^{11}\widetilde{X}'C^{-1}\widetilde{X} & \ldots & \rho^{1N}\widetilde{X}'C^{-1}\widetilde{X} \\ & \vdots & \\ \rho^{N1}\widetilde{X}'C^{-1}\widetilde{X} & \ldots & \rho^{NN}\widetilde{X}'C^{-1}\widetilde{X} \end{bmatrix}^{-1} \begin{bmatrix} \sum_{j=1}^N \rho^{1j}\left[\widetilde{X}'C^{-1}\widetilde{X}\right]\widehat{\beta}_j^{\text{single-GLS}} \\ \vdots \\ \vdots \\ \sum_{j=1}^N \rho^{Nj}\left[\widetilde{X}'C^{-1}\widetilde{X}\right]\widehat{\beta}_j^{\text{single-GLS}} \end{bmatrix} \\
&= \begin{bmatrix} \rho^{11}\left(\widetilde{X}'C^{-1}\widetilde{X}\right)^{-1} & \ldots & \rho^{1N}\left(\widetilde{X}'C^{-1}\widetilde{X}\right)^{-1} \\ & \vdots & \\ \rho^{N1}\left(\widetilde{X}'C^{-1}\widetilde{X}\right)^{-1} & \ldots & \rho^{NN}\left(\widetilde{X}'C^{-1}\widetilde{X}\right)^{-1} \end{bmatrix} \begin{bmatrix} \left(\widetilde{X}'C^{-1}\widetilde{X}\right)\sum_{j=1}^N \rho^{1j}\widehat{\beta}_j^{\text{single-GL}} \\ \vdots \\ \vdots \\ \left(\widetilde{X}'C^{-1}\widetilde{X}\right)\sum_{j=1}^N \rho^{Nj}\widehat{\beta}_j^{\text{single-GL}} \end{bmatrix}
\end{aligned}
$$

Now Zellner's result deirectly applies to yield

$$
\widehat{\beta}_i^{system} = \beta_i^{\text{single-GLS}}, \qquad i = 1, \ldots, N.
$$

Thus, a system of equations of fixed effects model can be estimated consistently and efficiently by estimating each equation consistently and efficiently separately.

A.2 Summary statistics

Table A2.1.a:

Variable name	Definition
Hempl	Husband's employment status: 1 employed 2 unemployed 3 military service having worked before 4 retired 5 living out of property rents 6 student (none) 7 housewife (none) 8 others
Hgact	Type of husband's employment: 0 missing 1 entrepreneurs or self-employed with employees 2 entrepreneurs or self-employed without employees 3 wage earners with a permanent job 4 wage earners with a temporary job 5 working in family business without salary 6 other
Hage	Age of the husband
Wage	Age of the wife
Famsize	Number of people living in the household in addition to the husband and wife
Age02	Persons aged 0-2 years living in the household
Age36	Persons aged 3-6 years living in the household
Age713	Persons aged 7-13 years living in the household
Age1415	Persons aged 14-15 years living in the household
Age1617	Persons aged 16-17 years living in the household
Age1824	Persons aged 18-24 years living in the household
Age2564	Persons aged 25-64 years living in the household
Age65	Persons aged 65 years and above living in the household

Table A2.1.b:

Variable name	Definition
Dheduc	Indicator variable for husband's education level taking the value zero if husband's education is low (illiterate, less than 5 years at school, primary school or secondary school) and taking the value on if husband's education is high (secondary school: second level & professional studies, university degree (3 or 5 years) and PhD's
Dtenure	Indicator value for housing tenure taking the value one if household is home owner and zero if household is renter or has free accomodation

Table A2.2:

Commodity group	Description
Food at home	Food at home
Alcohol & tobacco	Alcoholic drinks and tobacco
Clothing	Clothing
Rent	Renting (including the money paid to the owner and the water, but not electricity, heating etc.)
Energy	Electricity, heating and petrol
Services (Grh)	Furniture and appliances repairing, products for cleaning, money paid to people for cleaning the house and other household services
Medication	Medical expenses
Transportation	Car repairing, public transportation and communications (phone, mail, etc)
Leisure	Books, cinemas and other entertainments
Education	Education
Foodout	Restaurants and cafeterias
Holiday	Holidays
Other	Hairdressing, non-durables for personal care (soap, cosmetics etc), pocket money given to children, other services
Durables	Durables (cars, furniture, tv, etc)

Table A2.3:

Total number of periods a household is in the sample	Frequency of households	Cumulative frequency of households	Percent
1	64	.	
2	19	19	
3	12	31	
4	10	41	
5	18	59	
6	127	186	
7	162	348	
8	38	386	
9	52	438	
10	46	484	
11	42	526	
12	31	557	
13	48	605	
14	46	651	
15	24	675	
16	22	697	
17	23	720	
18	28	748	
19	21	769	
20	15	784	
21	8	792	
22	12	804	
23	18	822	
24	87	909	
Total	973	909	

Table A2.4:

Period	Frequency of households
7801	367
7802	384
7803	411
7804	492
7901	491
7902	521
7903	455
7904	524
8001	526
8002	527
8003	463
8004	535
8101	530
8102	522
8103	431
8104	467
8201	466
8202	466
8203	405
8204	393
8301	358
8302	341
8303	300
8304	299
Total	10674

Table A2.5:

Average household size	Frequency of households	Percent
2	75	8.25
]2, 3]	163	17.93
]3, 4]	301	33.11
]4, 5]	189	20.79
]5, 6]	123	13.53
]6, 7]	33	3.63
]7, 8]	14	1.54
]8, 9]	9	0.99
]9, 10]	2	0.22
Total	909	100

Table A2.6:

Average number of children in household	Frequency of households	Percent
0	207	22.77
]0, 1]	187	20.57
]1, 2]	297	32.67
]2, 3]	127	13.97
]3, 4]	63	6.93
]4, 5]	17	1.87
]5, 6]	8	0.88
]6, 7]	3	0.33
Total	909	100

Table A2.7:

Average number of children in household	Frequency of households	Percent
0	425	46.75
]0, 1]	272	29.92
]1, 2]	144	15.84
]2, 3]	62	6.82
]3, 4]	5	0.55
]4, 5]	1	0.11
Total	909	100

Table A2.9:

Frequency of households	Interval of mean total expenditure (pesetas)	Size of interval
90	37 775.18 - 126 775.5	89 000
90	126 992.7 - 147 830	21 000
90	148 201.6 - 167 437	20 000
90	167 680.2 - 185 946.1	18 200
90	186 466.7 - 212 982.1	26 500
90	213 181.6 - 232 706.9	19 500
90	133 164.1 - 259 184.1	26 000
90	259 514 - 290 343.5	31 000
90	290 925.6 - 341 412.4	50 000
90	341 732.9 - 479 914.2	138 000
9	480 942.9 - 770 177.5	289 000