The University of Manchester

MANCHESTER 1824

Economics Discussion Paper Series

Integrability of Demand Accounting for Unobservable Heterogeneity: A Test on Panel Data

Mette Christensen

EDP-0713

Economics
School of Social Sciences
The University of Manchester
Manchester M13 9PL

August 2007

Integrability of Demand Accounting for Unobservable Heterogeneity: A Test on Panel Data*

Mette Christensen †

Abstract

In recent years it has become apparent that we must take unobservable heterogeneity into account when conducting empirical consumer demand analysis. This paper is concerned with integrability (that is, whether demand is consistent with utility maximization) of the conditional mean demand (that is, the estimated demand) when allowing for unobservable heterogeneity. Integrability is important because it is necessary in order for the estimated price and income effects to be used for welfare analysis. Conditions for conditional mean demand to be integrable in the presence of unobservable heterogeneity are developed in the literature. These integrability conditions have testable implications for panel data. In this paper we exploit a unique long panel data set to test these conditions. We make use of the panel structure to identify a very flexible specification of unobservable heterogeneity: We model individual demands as an Almost Ideal Demand system and allow for unobservable heterogeneity by allowing all the parameters of the demand system to be individual-specific. We test the conditions for integrability of the conditional mean demand of this demand system. We find that that they can not be rejected. This means that the conditional mean demand generated by a population of consumers with different preferences described by different Almost Ideal Demand systems is consistent with utility maximization.

*I am grateful to Walter Beckert, Richard Blundell, Martin Browning, Mette Ejrnaes, Stefan Hoderlein, Arthur Lewbel, Birgitte Sloth, Frank Windmeijer, Allan Würtz and seminar participants at University of Copenhagen, Mannheim University, The 11th International Conference on Panel Data 2004, The Econometric Society European Meeting 2004, Essex University, IFS, Cemmap, Queen Mary College, Birkbeck College, Aarhus University and University of Manchester for many helpful suggestions and discussions. Financial support from the European Community's Human Potential Programme under contract HPRN-CT-2002-00235 [AGE] is gratefully acknowledged. All errors are mine.

†Department of Economics, University of Manchester, and The Institute for Fiscal Studies, London. E-mail: mette.christensen@manchester.ac.uk

JEL: C33, D12, D60
Keywords: Integrability, demand, unobservable heterogeneity, panel data

1 Introduction

Demand system estimation provides estimates of price and income elasticities, as well as estimates of the effects of demographic variables on demands. These elasticities and effects are important inputs into many policy analyses; for example the analysis of the effects of income or commodity taxes on market demands and the implications these effects have for welfare. Indeed, one of the main motivations for estimating demand systems is to facilitate welfare analysis. For this it is necessary that demands are consistent with consumer theory; that is, it is necessary that demands are integrable[1].

In this paper we exploit a unique data set to test conditions that are necessary for the conditional mean demand function across consumers to be integrable, assuming that individual consumers all separately maximise utility. The conditional mean demand is of great importance in empirical demand analysis because this is what is estimated: when we estimate a demand system, what we estimate is always the average demand, conditional on observables. To see this, consider the usual way of modelling and estimating demand systems. The usual way of modelling demand systems is by an additive model: The demand of each individual is modelled as the sum of a systematic component, i.e. some function, and an additive error term. The systematic component is functionally dependent on observables, like prices, incomes and observable demographics, and is common for all individuals: Different individuals have different values of observables, but the systematic component is the same function for all individuals. The error term is functionally independent of observables and is specific to each individual. The systematic component is then estimated from data, typically employing the assumption that the additive error term has conditional mean zero. This, together with the conditional mean zero condition on error terms, implicitly defines the systematic component to be the average budget share function. This means that what is in fact being estimated is the conditional mean budget share function. In order to use the estimated coefficients from a demand system for welfare analysis we thus need the conditional mean demand to be integrable.

There is a vast literature on demand system estimation and this literature provides an abundance of different ways of specifying the conditional mean demand to be estimated, both parametrically and nonparametrically. Widely used examples of parametric specifications are the Almost Ideal Demand system (introduced by Deaton and Muellbauer (1980)) and the Quadratic Almost Ideal Demand system (introduced by Banks, Blundell and Lewbel (1997)). Nonparametric specifications include Härdle and Jerison (1988), Lewbel (1991), Blundell, Duncan and Pendakur (1998) and Blundell, Chen and

[1] By the term "integrability" we will understand that demand is generated from maximising a utility function subject to a linear budget constraint.

Kristensen (2007). The data used for demand system estimation is nowadays typically cross sectional household expenditure survey data, that is, data at the household level, where each household is observed only once. A couple of decades ago demand systems were estimated using aggregate data; thus Deaton and Muellbauer (1980) used aggregate data. Regardless of whether data is aggregate level data or household level data, and regardless of whether the demand system is parametrically or nonparametrically estimated, what is always estimated is the conditional mean demand.

When modelling demand systems in the usual way, unobservable individual-level heterogeneity is assumed to be captured in the additive, individual-specific error term, which is not taken into account when the integrability properties of the demand system are analysed. In recent years, however, it has become widely acknowledged that we must take unobservable heterogeneity into account when conducting empirical consumer demand analysis. In an early paper, Brown and Walker (1989) consider the additive demand model and show that in order for integrability to hold at both the individual and the average level, the individual error terms must be functionally dependent on prices and/or incomes. In other words: If we want integrability to hold both at the individual level and at the average level, we can not employ the usual modelling of demand systems with additive independent error terms. In contrast to the usual way of modelling demand systems, the random utility approach takes unobservable heterogeneity specifically into account by letting a random component either enter the individual utility maximization problem directly, or letting it enter the demand system. The idea is that each consumer has his or her own value of the random component (coefficient) and hence the distribution of the random coefficients represent the distribuition of preference heterogeneity. Each consumer knows his or her own value, but this is unknown to the researcher. In a recent paper, Lewbel (2001) adopts the random utility approach to the usual additive model and derives conditions under which the conditional mean demand is integrable, assuming that individual demands are integrable. The conditions imply the result of Brown and Walker (1989). These integrability conditions are conditions on a matrix whose elements are covariances across individuals (households) between individual income responses and individual demands and hence they have testable implications for panel data: The conditions can be tested as properties of an estimate of the matrix of covariances. Because each element of the matrix is a covariance across individuals (households) between individual income responses and individual demands, this matrix can be estimated from panel data by a simple two step procedure. In a first step, the time series variation of the data can be used to estimate the individual income responses and the individual demands for all individuals. In a second step, the cross sectional variation in the data can be used to estimate the elements of the matrix of covariances as the sample covariances across households of the estimated individual effects.

In this paper, we exploit the unique long time series dimension of a Spanish panel data set on households expenditures to empirically test the integrability conditions developed in Lewbel (2001), following the two step procedure outlined above. The data is the Encuesta Permanente de Consumo (the ECP), which is a 6 year long data set with quarterly information on household expenditures, on prices and on demographics, collected by the Spanish National Bureau of Statistics in the period 1978-83 [2]. The ECP is to the best of our knowledge the longest real panel on households consumption covering a wide range of commodity groups we have available. The exceptional long time-series dimension allows us to estimate the individual income responses and individual demands that are needed for each household in order to construct an estimate of the matrix of covariances. It also allows us to be far more flexible in our specification of preference heterogenity than other studies estimating demand systems with unobservable heterogeneity.

We take a semiparametric approach and model the demand of each household by an Almost Ideal Demand system. We introduce preference heterogeneity by allowing all parameters to be household-specific, but we impose no parametric restrictions on the distribution of preference parameters. Identification of such a flexible heterogeneity scheme is possible only because we have this long panel data. The idea is to view the time series dimension for each household as a repetetion of the same thing: Demand systems are models of consumer behavior that describe how consumers allocate total expenditures to consumption goods *within* the period, given that consumers have already allocated a given amount of total expenditure to each period by solving an intra-temporal optimastion problem[3]. In other words, demand systems are static models and thus we have no dynamics in our model. We test for integrability and find that we can not reject the integrability conditions. This finding implies that a set of completely heterogeneous Almost Ideal Demand systems generate an integrable conditional mean demand.

Among our other findings are however some strong rejections of homogeneous consumer behavior: We strongly reject that different consumers have identical income effects. That is, we find strong evidence of preference heterogeneity in marginal propensities to spend. We also strongly reject that the coefficients on seasonal dummies are identical across consumers. This means that different households adjust their budget shares differently over the year; one interpretation could be that some households simply like winter holidays whereas others prefer summer holidays. Both these rejections of homogeneous consumer behavior is in stark contrast to what is usually assumed in demand system estimation: Usually, it is assumed that these behavioral coefficients are identical across

[2] The data was kindly provided by Lola Collado. Many thanks to her for answering numerous queries about the data. Also thanks to José M. Labeaga for help with this data.

[3] Demand systems are thus the second stage of a two stage budgeting process, see Deaton and Muellbauer (1980).

consumers - at most it is assumed that income responses can vary with observable demographics (the classic example is to allow income repsonses to be different according to number of children in the household). But our finding suggests that this is not enough. Our findings thus also adds to the growing body of research which shows the importance of taking unobservable heterogeneity into account.

Few other papers have looked at integrability of demand systems when accounting for unobservable heterogeneity. Hoderlein (2004) derives nonparametric tests of negativity and symmetry in a random utility setting which is more general than that of Lewbel (2001). Amongst other features, Hoderlein (2004) considers a model that is more general than the additive model. The framework of Hoderlein (2004) nests the framework of Lewbel (2001) and thus also nests our model. But none of these papers contain any empirical applications. Brown and Matzkin (1998) construct a random utility model by letting a random component enter the direct utility function and then derive the demand equations from the utility maximisation problem. Their paper contains no empirical application. Beckert (2005) estimates the demand for internet services, allowing for preference heterogeneity in a Cobb-Douglas utility framework. This model automatically generates a conditional mean demand which is integrable because individual preferences are Cobb-Douglas preferences. Calvet and Comon (2003) estimate an Almost Ideal Demand system with unobserved heterogeneity, but since they have cross sectional data, they are forced to being very restrictive in their heterogeneity specification; e.g. they can only identify a linear scheme with one heterogeneity parameter per individual and one per good, whereas panel data allows at least one heterogeneity parameter per individual per good. In our model, we furthermore allow heterogeneous price and income responses.

The integrability conditions in Lewbel (2001) bear a strong resemblance to the conditions for integrability of the unconditional average demandin Muellbauer (1975) and Mas-Colell (1985). The matrix of covariances that appear in Lewbel (2001) is roughly speaking a conditional version of the matric of covariances that appear in Mas-Colell's work on aggregation. It is somewhat surprising that conditioning on observables like income does not provide more structure.

It is worth pointing out that while this paper deals with the question of obtaining integrability at the average level, assuming integrability at the individual level prevails, a different type of question one could ask is "Can we obtain integrability - or just the Weak Axiom of Revealed Preference - at the average level without assuming integrability at the individual level, but instead by assumptions on the *distributions* of individual behavior?". This is the question asked in a strand of the theoretical demand aggregation literature by amongst others Hildenbrand (1994) and Grandmont (1992). The results of Hildenbrand and Grandmont have an important feature in common with the conditions in Brown and Walker (1989) and Lewbel (2001). Namely that what is needed is what

Hildenbrand and Grandmont denote *behavioral heterogeneity* in preferences, which are distributional assumptions on consumers' *behavior,* i.e. distributional assumptions on how consumers respond to changes in for example income. In other words, they also require that unobservable heterogeneity is functionally dependent on prices and/or incomes.

The rest of this paper is organised as follows. In Section 2 we present the integrability conditions and discuss what they imply for the ways in which we can introduce preference heterogeneity into demand systems in a way that is consistent with consumer theory. In Section 3 we formulate our theoretical model and provide a theoretical discussion of the integrability conditions in the context of our model. Section 4 presents the econometric model to and section 5 contains the empirical analysis and the results. Section 6 discusses and concludes.

2 The Integrability Conditions Accounting for Unobservable Heterogeneity

In this section we present and discuss the theoretical conditions for integrability of a demand system with unobservable heterogeneity developed in Lewbel (2001).

2.1 Demand Systems and Integrability: The Usual Way

In order to put the integrability conditions for a demand system with unobservable heterogeneity into the right perspective, we first consider the usual way of modelling demands in empirical demand analysis. Usually, we specify an additive model: The demands are modelled as the sum of a function of observables (like prices and income) and an error term. The function of observables is common for all households, whereas the error term is household-specific and does not depend on observables. The error term then capture, among other things, unobservable heterogeneity and is typically assumed to have conditional mean zero. Let N denote the number of goods, let w_h denote the vector of budget shares for household h, let $p = (p_1, ..., p_N)'$ denote the vector of prices for the N goods, let $\ln p$ denote the vector of log prices, let x_h denote total expenditure for household h, let z_h denote a K-dimensional vector of observable characteristics of household h (e.g. demographic characteristics) and let ε_h denote the error term specific to household h. Then the usual additive model can be written

$$w_h = G(\ln p, \ln x_h, z_h) + \varepsilon_h,$$

where

$$E\left[\varepsilon \left| \ln p, \ln x, z\right.\right] = 0.$$

The function $G(\cdot)$ is then estimated from data. There is a vast literature on the estimation of demand systems, with both parametric and nonparametric specifications of $G(\cdot)$. One of the most well-known examples of a parametric form of $G(\cdot)$ is the Almost Ideal Demand system (hereafter denoted the AID system), introduced by Deaton and Muellbauer (1980b) and since used in numerous applications. The data used for demand system estimation is nowadays typically cross sectional household expenditure survey data, but until a couple of decades ago, household-level data were not that common and demand systems would be estimated from aggregate level data[4]. Regardless of whether data is aggregate level data or individual-level cross sectional data, and regardless of whether $G(\cdot)$ is specified parametrically or nonparametrically, the additive structure of the individual budget share function together with the zero mean condition implicitly defines $G(\cdot)$ as the *conditional mean* budget share function. This means that what is in fact being estimated is always an *average* (namely the conditional mean) budget share function.

Now, in order to be able to use the estimated price - and income elasticities from this demand system for welfare analysis, the estimated demand must be consistent with consumer theory, i.e. the conditional mean demand must be integrable. In order to perform welfare analysis at the individual level, we will also need that the behavior of each individual is consistent with consumer theory, i.e. that the individual demands $G+\varepsilon$ are integrable. Traditionally, empirical demand analysis conducts integrability analysis on G without taking the unobservable heterogeneity ε into account[5]. However, if one wants to explicitly interpret unobservable heterogeneity as containing preference heterogeneity, it seems natural to also require that individual demands are integrable. The first question that comes to mind is then whether it is possible that individual demands, $G+\varepsilon$, as well as the estimated demand, namely the conditional mean demand G, are integrable in the usual additive model? The answer to this question is no (Brown and Walker (1989)): If G as well as $G+\varepsilon$ are integrable, ε must be functionally dependent on prices and/or incomes (i.e. the additive error terms must be heteroskedastic)[6]. As a consequence of this result the literature on demand systems therefore turned to formulating demand systems that allow for unobservable heterogeneity to be functionally dependent on prices and incomes.

[4]The AID system was, when it was first introduced in Deaton and Muellbauer (1980(b)), estimated on aggregate level data.

[5]Quoting Brown and Walker (1989): As Barten (1977) remarks, "disturbances are usually tacked on to demand equations as a kind of afterthought".

[6]With the exception of homothetic preferences. If $w_{ih} = \alpha_{ih} + \varepsilon_{ih}$, where $E[\varepsilon_i] = \alpha_{ih} - \bar{\alpha}_i = 0$, where $\bar{\alpha}_i$ is the mean of the α_{ih}'s, then both the individual and the conditional mean demand is integrable (because they are both Cobb-Douglas). However, there seems to be an overall consensus in the literature on demand systems that homothetic preferences are too restrictive to realistically describe consumer behavior.

The natural way to do this seems to be to adopt the random utility hypothesis as an approach for randomisation (e.g. Brown and Matzkin (1998), Beckert (2002)).

2.2 Demand Systems and Integrability: Accounting for unobservable Heterogeneity

Lewbel (2001) adopts the random utility approach to the additive demand system model allowing for general heteroskedasticity of the error term and derive conditions under which the conditional mean demand is integrable, given that individual demands are integrable. These conditions have testable implications for panel data which is what we utilize in this paper. In order to formulate Lewbel's conditions, consider a sample of H households, $h = 1, ..., H$. Throughout, we assume independence across households. Let N, w, $\ln p$, $\ln x$ and z be as before. Let η denote an L-dimensional vector of unobservable characteristics with $L \geq N$[7]. Let g denote the individual budget share function of household h, and let $F(\eta \mid \ln x, \ln p,\ z)$ denote the conditional distribution of the unobserved characteristics in the population, conditional on observable characteristics. We can then write individual budget shares as

$$\begin{aligned} w &= g(\ln p, \ln x, z, \eta) \\ &= G(\ln p, \ln x, z) + \nu(\ln p, \ln x, z, \eta), \end{aligned} \tag{1}$$

where G is defined as the conditional mean:

$$\begin{aligned} G(\ln p, \ln x, z) &= E\left[\, w \mid \ln p, \ln x, z\,\right] \\ &= \int g(\ln p, \ln x, z, \eta) dF(\eta \mid \ln p, \ln x, z\,). \end{aligned} \tag{2}$$

The definition of G then implies that

$$E\left[\nu \mid \ln p, \ln x, z\right] = 0.$$

Notice that this formulation of g in itself imposes no restrictions on individual budget shares: We can choose g to be any budget share function, calculate G from (2) and then construct $\nu(\cdot)$ as the residual $g - G$. Obviously, this formulation nests the usual model.

[7]In order to ensure that the model produces a non-degenerate distribution of budget shares, it is necessary that there are at least as many unobservables per individual as there are goods (Beckert (2006)).

Before turning to the integrability conditions, let us comment on how unobservable heterogeneity enters in this framework as compared to how it enters in the usual model. In the usual model, preference heterogeneity is implicitly assumed to be captured in the ε's. Since ε is functionally independent of observables, price - and income effects are restricted to have the same functional form for all households: When differentiating the budget share function with respect to prices or income, there is no contribution from ε. This means that preference heterogeneity can only enter as level effects in the usual model. In other words, the usual model does not allow for unobservable heterogeneity in the marginal effects. The formulation in (1) has the household-specific error term as a function of both unobservables *and* observables (as was shown is necessary for integrability by Brown and Walker (1989)). This means that preference heterogeneity enters not just as level effects as in the usual model, but also as slope effects. For example, two households with identical income levels and identical observable characteristics can have different responses to a change in income. In other words, the result of Brown and Walker (1989) means that in order to ensure integrability both at the individual level of the conditional mean in models that allow for preference heterogeneity, it is necessary that preference heterogeneity enters not just as level effects (i.e. that some households persistently have a high budget share for some good and others a low budget share independently of prices and income levels), but also in the marginal effects (i.e. that different households respond differently to changes in prices or in their incomes, all other things being equal).

The error term being functionally independent on prices and/or incomes, however, only provides necessary conditions for integrability of the conditional mean demand. Sufficient conditions are provided in Lewbel (2001). Define the N by N matrices s and $\tilde{s}$ by

$$\begin{aligned} s(\ln p, \ln x, z) &= \frac{\partial g(\ln p, \ln x, z)}{\partial(\ln p)'} + \frac{\partial g(\ln p, \ln x, z)}{\partial(\ln x)} g(\ln p, \ln x, z)' \\ \tilde{s}(\ln p, \ln x, z) &= s(\ln p, \ln x, z) + g(\ln p, \ln x, z) g(\ln p, \ln x, z)' - \operatorname{diag} g(\ln p, \ln x, z). \end{aligned}$$

These are the budget share analogs to the Slutsky matrix. From classical demand theory we know that the four conditions that ensure integrability of a continuously differentiable budget share function are: Adding up (budget shares add up to one), homogeneity (the budget share function is homogeneus of degree zero in prices and income), symmetry (that s is symmetric[8]) and negativity (that $\tilde{s}$ is negative semidefinite). The corresponding budget share analogs to the Slutsky matrix for the conditional mean demand G are defined

[8] Note that $\tilde{s}$ is symmetric if and only if s is symmetric.

similarly

$$
\begin{aligned}
S(\ln p, \ln x, z) &= \frac{\partial G(\ln p, \ln x, z)}{\partial(\ln p)'} + \frac{\partial G(\ln p, \ln x, z)}{\partial(\ln x)} G(\ln p, \ln x, z)' \\
\tilde{S}(\ln p, \ln x, z) &= S(\ln p, \ln x, z) + G(\ln p, \ln x, z) G(\ln p, \ln x, z)' - \operatorname{diag} G(\ln p, \ln x, z).
\end{aligned}
$$

We assume that individual demands are integrable. In addition, the following independence assumption is invoked:

$$
F_\eta \equiv F(\eta \,|\ln x, \ln p, z\,) = F(\eta \,|z\,), \tag{3}
$$

which, roughly speaking, states that preferences are stochastically independent of prices and income. Under this independence assumption, $\tilde{S}$ can be written

$$
\tilde{S} = E\left[\tilde{s}\right] - M - \operatorname{Var}\left[g\right], \tag{4}
$$

where

$$
M = \{M_{ij}\}_{i,j=1,\dots,N} = \left\{ \operatorname{Cov}\left[\frac{\partial g_i}{\partial(\ln x)}, g_j \right] \right\}_{i,j=1,\dots,N}
$$

and where Var$[g]$ is the variance-covriance matrix of g. For G to be integrable, G must satisfy adding up, homogeneity, symmetry and negativity. Adding up follows directly because adding up is satisfied at the individual level, and homogeneity follows from homogeneity at the individual level in conjunction with the independence assumption. Since a variance-covariance matrix is always symmetric and positive definite, the negative of the variance matrix $-$Var$[g]$ is also symmetric and negative definite. $E\left[\tilde{s}\right]$ is symmetric and negative semidefinite because $\tilde{s}$ is symmetric and negative semidefinite, which follows from integrability at the individual level. Therefore $\tilde{S}$ is symmetric if and only M is symmetric, and $\tilde{S}$ is negative semidefinite if M is positive semidefinite. Note that the integrability conditions are sufficient conditions: Symmetry of M is necessary and sufficient, whereas the positive semidefiniteness of M is only sufficient.

2.3 The Matrix of Covariances

M is a matrix of covariances (not a variance-covariance matrix!). It expresses, roughly speaking, how income effects (i.e. marginal propensities to consume) vary with budget shares in response to changes in unobservables. Mas-Colell, Whinston and Green (1995) interpret the positive semidefiniteness of M as consumers with higher than average consumption of one commodity also tend to spend a higher than average fraction of their last unit of income on that commodity. The expression for the Slutsky matrix of the conditional mean demand, $\tilde{S}$, in (4) displays clearly how M comes about: The Slutsky matrix for the conditional mean demand is not equal to the conditional average of the Slutsky matrices for the individual demands. There is "something" left over, and this "something" is precisely the matrix of covariances M. Integrability at the individual level ensures that $E[\tilde{s}]-\mathrm{Var}[g]$ is well-behaved (negative semidefinite), but imposes no structure whatsoever on M. This is very similar to the results found in the studies of demand aggregation. Here the averageing of demands is unconditional and thus aggregation happens over consumers with different incomes. But the same matrix of covariances (only now the covariances are unconditional) occurs in the special case where the income distribution is fixed (see the aggregation chapter in Mas-Colell, Whinston and Green (1995)) and also in the case of more general income distributions (see Mas-Colell (1985)). As Mas-Colell, Whinston and Green (1995) remark "the source of the aggregation problem rests squarely with the wealth effects on the consumption side". The result of Lewbel (2001) shows us that conditioning on income does not aid in getting rid of this problem caused by the income effects.

2.4 The Distribution of Unobservable Heterogeneity

There are two obvious cases in which the integrability conditions are met. One is the case where M is the zero matrix. This case is for example implied by Gorman aggregation: If all consumers have identical income effects for each commodity, then each element of M is the covariance between a constant and a random variable, which is always zero. Another case is where M is so close to being the zero matrix that the negative semidefiniteness of $E[\tilde{s}]-\mathrm{Var}[g]$ is large enough to make $\tilde{S}$ negative semidefinite. This case can be interpreted as budget shares and income effects varying very little with unobservables, i.e. that there is only little dispersion in preferences across consumers.

The integrability result presupposes that the distribution of unobservables conditional on observable demographics are independent of prices and incomes ((3)). This assumption has recently been examined empirically: Calvet and Comon (200?), Labeaga and Puig (2003), Browning and Collado (2006) and Christensen (200?) all contain empirical evidence suggesting that this assumption may not hold in the data for all commodities. Calvet and Comon (200?) use the FES. The FES is a cross sectional data set and hence the authors are forced to rely on a restictive identificaiton scheme (they assume that preference heterogeneity can be described by one random parameter per good and one random parameter per individual) which may account for their strong finding; they find that the majority of the observed variation in budget shares is due to preference heterogeneity and hence that income effects can only explain very little of the observed differences in budget shares. Labeaga and Puig (2003), Browning and Collado (2006) and Christensen (200?) all use panel data (the same panel data set, namely the ECPF) and hence they can allow for more flexible specifications of unobservable heterogeneity than Calvet and Comon. Differing models and tests are employed in the three papers, but the findings are in concordance with each other, namely that for some, but not all, goods, there is evidence of correlated heterogeneity. When employing a similar test for the data set used in this paper we find no evidence against the independence assumption [(see Section ?)].

3 How to Formulate Preference Heterogeneity and the Research Question

To estimate the matrix of covariances M, we must first specify the individual budget share function g and the distribution of preference heterogeneity F. g can be specified either parametrically or nonparametrically. We choose a parametric specification, because this allows us to model the unobservable household-specific characteristics η by the unknown parameters of the individual budget share functions, which fascilitates identification. More precisely, we model the preference heterogeneity by taking g to be a parametric demand system and allowing some or all the parameters to be different across households. This amounts to specifying g as a variable-coefficient model and accordingly view the coefficients as random variables with a conditional distribution, conditional on observables [9]. This conditional distribution of the coefficients is then the distribution of preference heterogeneity. For example, if we took g to have the most simple Working-Leser form, that is, $g(\ln x) = \alpha + \beta \ln x$, we would view α and β as random variables with a conditional distribution across households such that each household h had its own intercept α_h and its own slope β_h. The values α_h and β_h would be known to household h, but unknown to the econometrician. In estimation, we can choose at one extreme to specify the distribution of coefficients completely nonparametrically, placing no restrictions on its form. In this case, we would estimate the distribution of coefficients as the empirical distribution of the estimated realizations of the random variable underlying the distribution of coefficients. Or, in other words, we would estimate each coefficient for each household. This approach is very general. The cost of this generality is that it involves estimating a large number of parameters; in the example above with g having the Working-Leser form, we would have to estimate one α and one β for each household. At the other extreme, we can choose to model the distribution of coefficients completely parametrically, for example by a normal distribution; in the example with g having the Working-Leser form, this would involve

[9] We will refer to the parameters both as "coefficients" and as "parameters".

estimating only five parameters: The mean and variance of α, the mean and variance of β and the covariance of α with β. In between the two extremes lies a whole range of (semi-parametric) possibilities, like for example assuming a mixture distribution for the coefficients a la Heckman and Singer (1984).

We choose the nonparametric extreme and thus place no restrictions on the distribution of coefficients. The reason for choosing the fully nonparametric approach to the modelling of the distribution of preference heterogeneity is that any restriction on this distribution would be completely ad hoc; this type of model has never been estimated on demand panel data for a complete set of goods before, and so there are no suggestions in the literature about what is reasonable to assume about the distribution of preference heterogeneity. Moreover, as we will show later in an example, distributional assumptions can actually imply that the matrix of covariances is symmetric and positive semidefinite. In other words, by imposing distributional assumptions, we risk imposing integrability of the conditional mean demand by assumption. Obviously, this would be highly undesirable. As mentioned earlier, the drawback of the fully nonparametric approach is that it involves a large number of parameters to be estimated, and therefore we will expect less precise estimates from this approach than from a parametric or a semiparametric approach.

Because we assume independence across households, the model in which all parameters are household-specific can be estimated by estimating a demand system separately for each household. Obviously, this is only feasible because our panel data set has large T. Our approach is similar to the idea underlying the mean-group estimator in Pesaran and Smith (1995): Like them, we also estimate a model for each household (each group), but where Pesaran and Smith (1995) are interested in the average regression coefficient, we are interested in a different function of the estimated coefficients, namely in the matrix of covariances of the income effects with the budget shares, M.

In Section 3.1 we specify g and F, in Section 3.2 we calculate the conditional mean demand for our model, and in Section 3.3 we calculate the object of interest for the integrability test, the matrix of covariances M, for our model. Then we are finally able to state the research question in precise terms. In Section 3.4 we go beyond the model and give some examples of model specifications that in themselves lead to M being symmetric and positive semidefinite. Note that all that follows depend on the choice of g and F, since the matrix of covariances M is specific to these choices.

3.1 Individual Demands: An Almost Ideal Demand System with Household-Specific Parameters and a Nonparametric Distribution of Parameters

We choose g to be an AID system and introduce randomness in demands by making the parameters household-specific. Then the budget share equations for the N goods for household h are in their most general form with all parameters being household-specific given by

$$w_{ih} = \alpha_{ih} + \textstyle\sum_j \gamma_{ijh} \ln p_j + \beta_{ih} \left[\ln x - \ln P(p)\right], \qquad i = 1, ..., N,$$

where $P(p)$ is the price index given by[10]

$$\ln P(p) = \textstyle\sum_k \alpha_{kh} \ln p_k + \dfrac{1}{2} \textstyle\sum_k \sum_l \gamma_{klh} \ln p_k \ln p_l.$$

One can choose a greater or lesser degree of heterogeneity by restricting various parameters to be identical across households. We will consider various specifications in estimation and we return to a discussion of this point in Section 3.4. Since we assume that the demand of each individual is generated from utility maximisation, we have the usual restrictions on the parameters of an AID system, but now for each h [11]:

$$\begin{aligned} \textstyle\sum_{i=1}^n \alpha_{ih} &= 1, \quad \textstyle\sum_{i=1}^n \beta_{ih} = 0, \quad \textstyle\sum_{i=1}^n \gamma_{ijh} = 0 \text{ for all } j & (5) \\ \textstyle\sum_{j=1}^n \gamma_{ijh} &= 0 \text{ for all } i & (6) \\ \gamma_{ijh} &= \gamma_{jih} \text{ for all } i, j. & (7) \end{aligned}$$

Let θ denote the vector of the parameters in the demand system, that is, $\theta = ((\alpha_1, ..., \alpha_N), (\beta_1, ..., \beta_N), (\gamma_{11}, ..., \gamma_{NN}))$. The conditional distribution of preference heterogeneity is the conditional distribution of the unknown parameters θ, conditional on log total expenditure, log prices and demographics. Denote it by $F_\theta \equiv F(\alpha, \beta, \gamma \,|\, \ln x, \ln p, z)$. Thus, the independence assumption (3) translates in our model into that $\theta = ((\alpha, \beta, \gamma))_{i,j=1}^N$ is conditionally independent of $\ln x$ and of $\ln p$ for all i, j, conditional on demographics z. Denote the conditional means in the marginal distributions of the different parameters by

$$E_\theta[\beta_i \,|\, z] = \mu_{\beta_i}, \quad E_\theta[\alpha_i \,|\, z] = \mu_{\alpha_i}, \quad E_\theta[\gamma_{ij} \,|z] = \mu_{\gamma_{ij}},$$

$i, j = 1, ..., N$, where E_θ denotes integration with respect to F_θ.

[10] Since the intercept parameter, usually denoted α_0, in the price index is not identified we omit it wthout any loss of generality.

[11] See Deaton and Muellbauer (1980a) section 3.4 or Christensen (2005) Chapter 4 Appendix A.

3.2 The Conditional Mean Demand

The conditional mean demand for commodity i is calculated from (2) as

$$\int g_i(\ln x, \ln p; \theta) dF_\theta = G_i(\ln x, \ln p).$$

The conditional mean budget share function G_i for commodity i is thus in its most general form given by, where we denote the price index by $P(p;\theta)$ in order to emphasize that the price index is a function of the parameters and thus household-specific,

$$\begin{aligned} G_i(\ln x, \ln p) &= E_\theta[g_i(\ln p, \ln x, z; \theta) \,|z] \\ &= \mu_{\alpha_i} + \mu_{\beta_i} \ln x - E_\theta \left[(\beta_i \ln P(p;\theta)) \,|z\right] \\ &\quad + \textstyle\sum_j \mu_{\gamma ij} \ln p_j, \end{aligned}$$

where we used the independence assumption in the last equality[12]. If we restrict the price effects ($\gamma_{ij}, i = 1, .., N, j = 1, ..., N$) to be identical across households, the conditional mean demand is given by

$$\begin{aligned} G_i(\ln x, \ln p) &= \mu_{\alpha_i} + \mu_{\beta_i} \ln x - E_\theta \left[\beta_i \textstyle\sum_{k=1}^N \alpha_k \ln p_k \,|z\right] \\ &\quad - \mu_{\beta_i} \frac{1}{2} \textstyle\sum_k \sum_l \gamma_{kl} \ln p_k \ln p_l + \sum_j \gamma_{ij} \ln p_j. \end{aligned}$$

Note that in neither specification is G in itself an AID system because of the term $E_\theta \left[(\beta \ln P(p;\theta)) \,|z\right]$. If it was not for this term, G would be an AID system with parameters given by the conditional mean of the corresponding parameters of the individual AID systems. But because β is not necessarily independent of $\ln P(p;\theta)$, the mean of the product of β with $\ln P(p;\theta)$ is not necessarily equal to the product of the mean of β and the mean of $\ln P(p;\theta)$, and so G is not an AID system. From this observation it is clear that this must be a general point: When introducing unobservable heterogeneity in preferences by letting the parameters of a parametric demand system vary across individuals, the conditional mean demand will not have the generic form of the individual

[12] The additive error term for commodity i for household h, $\nu_{ih}(\ln x, \ln p)$, is given by $\nu_{ih}(\ln x, \ln p) = \left(\alpha_{ih} - \mu_{\alpha_i}\right) + \left(\beta_{ih} - \mu_{\beta_i}\right) \ln x - \left(\beta_{ih} \ln P(p;\theta) - E_\theta \left[\beta_i \ln P(p;\theta) \,|z\right]\right) + \left(\sum_j \left(\gamma_{ijh} - \mu_{\gamma_{ij}}\right) \ln p_j\right)$, from which it is clear that this model has error terms that are functionally dependent on both prices and total expenditure.

demands if the individual demands are nonlinear in the parameters (i.e. are nonlinear in the unobservable preference heterogeneity). From this statement it is now clear that a sufficient condition for the conditional mean demand to be an AID system is that the β_i's are conditionally independent of the parameters in the price index, α_i and γ_{ij} for all $i, j = 1, ..., N$:

Example 1 ***Independent income effects***

Assume that the income effect for commodity i, β_i, is independent of all the other parameters in the price index, α_j and γ_{ij}, $j = 1, ..., N$, for all $i = 1, ..., N$. Then the conditional mean demand G is itself an AID system:

$$G_i(\ln x, \ln p) = \mu_{\alpha_i} + \mu_{\beta_i} \ln x - E_\theta\left[\beta_i \ln P(p;\theta)\right] + \textstyle\sum_j \mu_{\gamma_{ij}} \ln p_j,$$

and since β_i is independent of α and γ,

$$G_i(\ln x, \ln p) = \mu_{\alpha_i} + \mu_{\beta_i} \ln x - E_\theta\left[\beta_i\right] E_\theta[\ln P(p;\theta)] + \textstyle\sum_j \mu_{\gamma_{ij}} \ln p_j,$$

and since

$$\begin{aligned} E_\theta[\ln P(p;\theta)] &= E_\theta[\textstyle\sum_k \alpha_{ik} \ln p_k + \frac{1}{2} \sum_k \sum_l \gamma_{kl} \ln p_k \ln p_l] \\ &= \textstyle\sum_k \mu_{\alpha_{ik}} \ln p_k + \frac{1}{2} \sum_k \sum_l \mu_{\gamma_{kl}} \ln p_k \ln p_l, \end{aligned}$$

then defining

$$\ln P(p;\mu_\theta) \equiv \textstyle\sum_k \mu_{\alpha_k} \ln p_k + \frac{1}{2} \sum_k \sum_l \mu_{\gamma_{kl}} \ln p_k \ln p_l,$$

we can write G_i on the following form for all $i = 1, ..., N$

$$G_i(\ln x, \ln p) = \mu_{\alpha_i} + \mu_{\beta_i} \ln x - \mu_{\beta_i} \ln P_{\mu_\theta} + \textstyle\sum_j \mu_{\gamma_{ij}} \ln p_j.$$

Thus, G is an AID system with parameters given as the conditional means of the corresponding individual parameters.

3.3 The Matrix of Covariances and the Research Question

The (i, j)'th entry in the matrix of covariances, M, is the covariance between the partial derivative of the budget share function for commodity i with respect to $\ln x$ and the budget share for commodity j across households. Since in our model the partial derivative of the budget share function for household h for commodity i with respect to log total expenditure is β_{ih}, $h = 1, ..., H$, the (i, j)'th entry of M for our model is given by, using the independence assumption of parameters being conditionally independent of prices and total expenditures (3)

$$
\begin{aligned}
M_{ij} &= \operatorname{Cov}\left[\beta_i, g_i(\ln p, \ln x, z; \theta) \,|z\right] \\
&= \operatorname{Cov}\left[\beta_i, \alpha_j \,|z\right] + \operatorname{Cov}\left[\beta_i, \beta_j \,|z\right] \ln x \\
&\quad -\operatorname{Cov}\left[\beta_i, \beta_j \ln P_\theta(p) \,|z\right] + \operatorname{Cov}\left[\beta_i, \textstyle\sum_k \gamma_{jk} \ln p_k \,|z\right]. \qquad (8)
\end{aligned}
$$

Now we are finally able to state the research question in precise terms. We have formulated a demand system with unobservable heterogeneity which is nested within the framework of Lewbel (2001). The integrability conditions in Lewbel (2001) then yield that the conditional mean demand for this demand system is integrable if the matrix of covariances M is symmetric and positive semidefinite. Our model is a demand system in which each household has its own AID system. The research question is thus whether a set of AID systems with all parameters being household-specific generates an integrable conditional mean demand. Taking a closer look at M, we see that without further assumptions, M is not necessarily symmetric. A sufficient condition for symmetry of M is that each of the four terms in (8) is symmetric. But only the second term, $\operatorname{Cov}\left[\beta_i, \beta_j \,|z\right]$, is symmetric without further assumptions. Thus, G is not necessarily integrable.

3.4 Alternative assumptions on the distribution of preference heterogeneity

As mentioned earlier, making assumptions on the distribution of preference heterogenity can have the unfortunate consequence that it makes M symmetric. In this section we give two examples of such assumptions:

Example 2 ***Identical income effect coefficients in the AID system***
Let the budget share equations for household h be given by

$$
g_i(\ln p, \ln x; \theta_h) = \alpha_{ih} + \beta_i \ln x - \beta_i \ln P(p; \theta) + \textstyle\sum_j \gamma_{ijh} \ln p_j, \qquad i = 1, ..., N,
$$

for each $h = 1, ..., H$. Then the partial derivative of the budget share function for commodity i with respect to log total expenditure is β_i for all households. Since M_{ij} is the covariance of the income effect for commodity i with the budget share of commodity j across households, and since the income effect is the same for all households, M_{ij} is the covariance between a constant and a random variable, so M_{ij} is zero for all i, j. This trivially implies that M is both symmetric and positive semidefinite.

This example shows that identical income effects for all households for all commodities in the AID system implies that the matrix of covariances is symmetric and positive semidefinite, i.e. identical income effects imply that G is integrable.

Example 3 ***Income effects independent of the other parameters of the demand system***

Suppose that β_i *is conditionally independent of* (α_j, γ_{ij}), $j = 1, ..., N$, *conditional on* $\ln x$, $\ln p$ *and* z *for all* i *and consider again* M_{ij}. *This independence assumption, together with all parameters being independent of* $\ln x$ *and* $\ln p$, *implies that the first and last terms in* M_{ij} *are zero:*

$$C \quad ov[\beta_i, \alpha_j \,|z] = 0$$

$$C \quad ov\left[\beta_i, \textstyle\sum_k \gamma_{jk} \ln p_k \,|z\right] = 0.$$

Furthermore, the third term becomes

$$\begin{aligned}
Cov\left[\beta_i, \beta_j \ln P(p;\theta) \mid z\right] &= E_\theta\left[\beta_i\beta_j \ln P(p;\theta) \mid z\right] - E_\theta\left[\beta_i \mid z\right] E\left[\beta_j \ln P(p;\theta) \mid z\right] \\
&= E_\theta\left[\beta_i\beta_j \mid z\right] E_\theta\left[\ln P(p;\theta) \mid z\right] - E_\theta\left[\beta_i \mid z\right] E_\theta\left[\beta_j \mid z\right] E_\theta\left[\ln P(p;\theta) \mid z\right] \\
&= \left(E_\theta\left[\beta_i\beta_j \mid z\right] - E_\theta\left[\beta_i \mid z\right] E_\theta\left[\beta_j \mid z\right]\right) E_\theta\left[\ln P(p;\theta) \mid z\right] \\
&= Cov\left[\beta_i, \beta_j \mid z\right] E_\theta\left[\ln P(p;\theta) \mid z\right],
\end{aligned}$$

i.e. M_{ij} *reduces to*

$$M_{ij} = Cov\left[\beta_i, \beta_j \mid z\right] \left(\ln x - E_\theta\left[\ln P(p;\theta) \mid z\right]\right),$$

i.e. M *is symmetric. Furthermore,* M *is the product of the conditional variance-covariance matrix of* $\beta = (\beta_1, \beta_2, \beta_3)$ *conditional on* z *and the number* $(\ln x - E_\theta\left[\ln P(p;\theta) \mid z\right])$. *The conditional variance-covariance matrix is positive definite which implies that* M *is positive semidefinite if* $\ln x - E_\theta\left[\ln P(p;\theta) \mid z\right]$ *is greater than or equal to zero.* $P(p;\theta)$ *is a price index and thus between* 0 *and* 1, *hence* $\ln P(p;\theta)$ *is less than or equal to zero, hence the mean of* $\ln P(p;\theta)$ *is also less than or equal to zero, which implies that* $\ln x - E_\theta\left[\ln P(p;\theta) \mid z\right]$ *is greater than zero for all values of total expenditure greater than 1. Since total expenditure is always much larger than 1,* M *is also positive semidefinite.*

Note that these examples are specific to this particular model where g is an AID system. If chosing a different form of g, one would have to re-examine which additional assumptions on the distribution of preference heterogeneity have unfortunate consequences for *that* choice of g[13]. We chose the AID system as our basic functional form, because the AID system is one of the most used parametric demand systems in the literature and because the income effects in that model are simply the β's which simplifies all our estimations and interpretations.

4 The Econometric Model

We model individual demand as an AID system with all parameters being household-specific and allow for heteroskedasticity across households (i.e. that each household has its own covariance matrix). Since we have assumed independence across households and since we have made all parameters of the demand equations household-specific, there are no cross-equation restrictions across households and so estimation of the demand system amounts to estimating an AID system separately for each household. The AID system for a given household h is given by the budget share equations

$$w_{iht} = \alpha_{ih} + \beta_{ih} \left[\ln x_{ht} - \ln P_{ht}(p) \right] + \sum_{j=1}^{N} \gamma_{ijh} \ln p_{jt} + \sum_{k=1}^{K} \delta_{ikh} z_{kht} + \varepsilon_{iht}, \tag{9}$$

$i = 1, ..., N,\ \ t = 1, ..., T$, with the price index given by

$$\ln P_{ht} = \alpha_{0h} + \sum_{k=1}^{N} \alpha_{hk} \ln p_{kt} + \frac{1}{2} \sum_{j=1}^{N} \sum_{k=1}^{N} \gamma_{kjh} \ln p_{kt} \ln p_{jt},$$

for $t = 1, ..., T$, and where z_{ht} is a K-dimensional vector of demographics for household h at time t and ε_{iht} is the idiosyncratic error term for household h, commodity i at time t. Integrability at the individual level is imposed by imposing the restrictions (5), (6) and (7) on the parameters for each household: Adding up, (5), is satisfied by leaving out one commodity. We leave out the N'th commodity for each household and thus end up with a system of $N-1$ equations. Given that adding up is satisfied, (6) implies that homogeneity is satisfied. (6) is imposed by replacing the N log prices with the $N-1$ homogeneous (or

[13] If for example chosing g to be the QUAID system, identical income effects do not imply symmetry of M, see Lewbel(2001).

relative) log prices given by[14]

$$\ln \widetilde{p}_{jt} = \ln p_{jt} - \ln p_{Nt}, \;\; j = 1, ..., N-1. \tag{10}$$

The error term structure of the demand system for household h is given by

$$\begin{aligned} E\left[\varepsilon_{iht}\right] &= 0, && \text{all } i, t \\ E\left[\varepsilon_{iht}\varepsilon_{jhs}\right] &= \sigma^2_{ijh}, && \text{all } i, j, \; t = s \\ E\left[\varepsilon_{iht}\varepsilon_{jhs}\right] &= 0, && \text{all } i, j, \; t \neq s. \end{aligned}$$

5 Empirical Analysis

In this section we estimate the matrix of covariances M and test whether M is symmetric. The estimation of M is carried out in two steps. Recall that entry (i, j) in M is the covariance of the budget share for commodity i with the income effect for commodity j. In the first step, we estimate budget shares and income effects for each of the commodities for *each* household. In the second step we calculate the covariance of the budget share for commodity i with the income effect for commodity j as the sample covariance across households of the estimated budget share for commodity i with the estimated income effect for commodity j.

5.1 Data

The data we use is a unique Spanish panel data sets on household expenditures, the Spanish Permanent Survey of Consumption (Encuesta Permanente de Consumo, hereafter the ECP), collected by the Spanish National Bureau of Statistics (Instituto Nacional de Estadistica). It is a real panel of sizeable longitudinal length: the ECP covers the years 1978-83 with quarterly information on all recorded variables with households staying in the survey between 6 and 24 quarters. All households in the data set are headed by a married couple and may contain children or other adults cohabiting in the household. The data set contains information on consumption expenditures for a wide range of commodity groups, price indices for these commodities as well as a variety of demographic variables such as labour market status, occupation, education level of the husband, the ages of the different household members and housing tenure. The version of the ECP we have

[14] $\sum_{j=1}^{N} \gamma_{ij} = 0$ implies that $\gamma_{iN} = -(\gamma_{i1} + ... + \gamma_{iN-1})$, which in turn implies that $\sum_{j=1}^{N} \gamma_{ij} \ln p_j = \gamma_{i1} \ln p_1 + ... + \gamma_{iN-1} \ln p_{N-1} - (\gamma_{i1} + ... + \gamma_{iN-1}) \ln p_N = \gamma_{i1}(\ln p_{i1} - \ln p_{iN}) + ... \gamma_{iN-1}(\ln p_{iN-1} - \ln p_{iN}) = \sum_{j=1}^{N-1} \gamma_{ij} \ln \widetilde{p}_j$.

available consists of 1641 households with more than 70 percent staying in the sample for at least 9 consecutive quarters. The total number of observations is 21.668. For complete lists of the variables recorded in the ECP, see Table A.1 and Table A.2 in Appendix A.

This Spanish household expenditure data set is exceptional in that it is a real panel with detailed information on a full range of commodities and prices, as well as on demographics. To the best of my knowledge, this is the longest real panels on household consumption expenditures covering a wide range of commodity groups, we have available. For comparison, the British Family Expenditure Survey (the FES) is a cross sectional data set and so does not provide the same possibility of taking account of unobservable heterogeneity. The American Consumer Expenditure Survey (the CEX) contains a rotating panel with information on several commodity groups, but it only has information on 4 quarters per household. Moreover, the consumption information in the panel part of the CEX consists of recall data only, and rescent research shows that the best method for collecting accurate consumption expenditure data is a combination of diary and recall information, see Battistin (2003). In contrast, the Spanish data consists of a combination of diary and recall information with a grouping of commodities into which are recorded as diary information and which are recorded as recall information very close to the one recommended in Battistin (2003). The Panel Study of Income Dynamics (the PSID) is also a real panel and is of considerable length, but it only contains consumption expenditure information on food (food in and food out) and is therefore not well suited for analyzing demand choices. The British Household Panel Survey (the BHPS) is another real panel, but also this panel has insufficient information on consumption expenditures: The BHPS only records expenditures on household appliances and electronics and has one food question which asks how much the household approximately spends on food and groceries each week [15]. The European Community Household Panel Survey (the ECHP) has no records on consumption expenditures [16]. Finally, there exists a panel survey for Japan (The Japanese Panel Survey on Consumers, the JPSC) which contains consumption information. However, the JPSC asks all households about expenses only for the calendar month of September [17], which makes it impossible to control for seasonal variation in demands.

In this paper, we use a subsample of the ECP. We select a sample consisting of the households that are observed for all 24 quarters and where the husband is all the time employed as a wage earner with a permanent job. This results in quite a small sample of the full data set, but there are two good reasons for this selection: Firstly, we select

[15] Browning, Crossley and Weber (2003) p. F563.

[16] The ECHP asks the household questions like "can you afford" various expenses, see Browning, Crossley and Weber (2003) p. F563, but this type of information is not useful for estimating demand systems.

[17] Browning, Crossley and Weber (2003) p. F563.

the households observed for all 24 quarters in order to have as many observations as possible per household, since we need to estimate a demand system separately for each household [18]. Secondly, we select the full-time employed wage earners with permanent jobs because by only modelling the demand (and not modelling the labor supply) we have implicitly assumed separability between the consumption of goods and the labor supply, and there seems to empirical evidence against this separability assumption: The empirical findings of Browning & Meghir (1991) show rejections in the FES. By selecting out the unemployed, the part-time employed and the employed with temporary jobs, we increase the probability that none of the households in our sample are making labor market descisions during the sample period which makes the assumption of separability between consumption and labor supply more plausible. This sample selection leaves us with a balanced panel of 87 households, each observed for 24 consecutive quarters yielding a total of 2088 observations. The table below illustrates the sample selection:

	Number of households	Number of observations
Total sample	1641	21668
In the survey for 24 quarters	249	5976
In the survey for 24 quarters and husband full-time employed as wage earner with permanent job	87	2088

Even though this is a somewhat small sample, it is worth emphasizing at this point that this is still the longest time-series available on expenditures on a full range of goods following the *same* individuals over time. It would not be possible to estimate all the parameters of an AID system separately for each household with any other existing data set than this [19].

The sample we have selected is clearly more homogeneous than the full data set and so we would expect less unobserved heterogeneity in this sample than in the full data set. To get an idea of whether there are selection problems, we compare the selected sample to the full data set. Summary statistics are presented in Appendix A. We see from Table A.3 and Table A.4 that there are fewer households where the husband has very low education level

[18] The original sample of the ECP covers three more quarters than our sample. The total number of households (including both employed and unemployed) for 27 quarters is 413 (Labeaga (1993) p. 107). There is thus potential scope for increasing each individual time series with 3 observations.

[19] As mentioned previously, the ECP is to the best of our knowledge the longest real panel data set on household expenditures which also includes reliable demographics.

(illiterate or less than 5 years of school), more with primary school, the same percentage has secondary school and more have a university degree. Thus, our sample has a higher education level than is the case in the full data set. The division of owners and renters of hte main house is exactly the same as in the full data set, but when it comes to the second house there is a difference: More households in our sample have a second house than in the full data set (the fraction owning a second house is 10 percentage points higher in our sample). This could suggest that the households in our sample is on average slightly more wealthy. This is somewhat confirmed when looking at total expenditures (Table A.5 and Table A.6): The average is 182.365 pesetas (standard deviation of 134.273) in the full data set, but 218.735 in our sample (standard deviation of 152.268). With regard to the expenditures on the individual goods, then the mean expenditure for Food at home is lower whereas the mean expenditure for Foodout is higher in our sample than in the full data set. The average ages of the husband, respectively the wife, is lower in our sample, but the average number of children and cohabiting adults is the same. In summary, our sample has slightly younger, more educated and slightly more wealthy households that eat more out than is the case in the full data set. Table A.7 provides a more detailed impression of the types of households in our sample. Almost all households (80 households) have either children and/or other cohabiting members. Most of the households either have children throughout or get them at some point during the survery period (73 households) and many households have cohabiting adults (60). Only 7 households (less than 10 percent of our sample) have no cohabitation at all.

When estimating a demand system separately for each household, we can only identify effects from those variables that change over time within the household. The level of the husband's education is time invariant. The dummy for housing tenure is time invariant for 83 of the 87 households. We therefore leave these two variables out of the demand system estimations. As for the variables describing the size and composition of the household, we control only for the level effect of the family size in the demand system estimation. 36 households experience changes to their family size during the survey period, so we include family size as an explanatory variable in the demand system estimations for those 36 households only.

Next we turn to the commodity groups. To keep things simple, we construct 4 aggregate commodity groups from the 12 of the 14 commodities, leaving out Rent and Durables. The definition of the 4 commodity groups, together with their mean and standard deviations, are displayed in the table below.

Commodity group	Definition	Mean	Standard deviation
Food, alcohol & tobacco	Food at home + Foodout + Alcohol & tobacco	.54	.14
Clothing	Clothing	.10	.07
Utilities	Energy + Services + Medication + Education + Transportation + Other	.29	.13
Leisure activities	Leisure + Holidays	.07	.06

The commodity group Food, Alcohol and Tobacco constitues a larger percentage of the total budget than is usually the case for food. It is abnormally large because the budget share for Food at home is abnormally large; an average budget share for Food of around 30% is what is usually observed in other expenditure surveys. The reason for this is that Spain went through a recession during the years where the ECP was collected [20].

We construct the log prices for each of the composite commodity groups as the weighted average of the log prices for the goods in that particular commodity group, with the weights being the average budget share for the goods. To take an example, the log price for the commodity group Food, Alcohol & Tobacco is constructed as

$$\begin{aligned}\ln\left(p_t^{\text{Food, Alcohol\&Tobacco}}\right) &= 2\cdot\left(\frac{1}{H}\sum_{h=1}^{H} w_{ht}^{\text{Food at home}}\right)\cdot\ln\left(p_t^{\text{Food at home}}\right)\\ &\quad+\left(\frac{1}{H}\sum_{h=1}^{H} w_{ht}^{\text{Alcohol\&Tobacco}}\right)\cdot\ln\left(p_t^{\text{Alcohol\&Tobacco}}\right)\end{aligned}$$

When estimating the demand system, the left out good will be Utilities, and therefore the relative prices are the prices of Food, Alcohol&Tobacco, Clothing and Leisure&Holidays

[20] M. D. Collado (1998).

relative to the price of Utilities:

$$
\begin{aligned}
\ln \widetilde{p}_{1t} &= \ln\left(p_t^{\text{Food, Alcohol\&Tobacco}}\right) - \left(p_t^{\text{Utilities}}\right) \\
\ln \widetilde{p}_{2t} &= \ln\left(p_t^{\text{Clothing}}\right) - \left(p_t^{\text{Utilities}}\right) \\
\ln \widetilde{p}_{3t} &= \ln\left(p_t^{\text{Leisure\&Holidays}}\right) - \left(p_t^{\text{Utilities}}\right)
\end{aligned}
$$

The variation over the sample period in relative prices is shown in Figure 1 in Appendix A. As the graph shows there is a lot of independent variation between the three relative prices which means that all price parameters (the price effects) should be well identified.

In the initial step in the estimation of the AID systems we deflate total expenditures with a Stone price index. We calculate household-specific Stone price indeces, i.e. the Stone price index for household at time t is given by

$$
\begin{aligned}
\ln P^*_{ht} = {} & w_{ht}^{\text{Food, Alcohol\&Tobacco}} \cdot \ln p_t^{\text{Food, Alcohol\&Tobacco}} + w_{ht}^{\text{Clothing}} \cdot \ln p_t^{\text{Clothing}} \\
& + w_{ht}^{\text{Leisure\&Holidays}} \cdot \ln p_t^{\text{Leisure\&Holidays}} + w_{ht}^{\text{Utilities}} \cdot \ln p_t^{\text{Utilities}}.
\end{aligned}
$$

One of the most (if not the most) important parameter for our test of integrability is the income effect for the different goods. The individual income effects are identified off variation in total expenditure *within* the household. And the larger the variation in total expenditure is within a household, the more precise will the estimate of the income effect for that household be. We therefore examine whether there is sufficient variation in total expenditure within the household. Firstly, a simple linear random effects panel data estimation shows the fraction of variance in the error term which is due to the individual-specific part, which is a measure of the within-variation in the data. As can be seen form the table below, this fraction is between 20 and 50 percent:

Commodity	Food, Alcohol and Tobacco	Clothing	Leisure and Holidays	Utilities
Fraction of variance due to individual-specific part	.4308	.2077	.1779	.4923

Secondly, the estimate of the matrix of covariances M from the individual-specific estimates will also depend on the between-variation in the data. We regress each of the budget shares on the set of household dummies; the R^2s from these regressions then measures how much of the variation in budget shares is explained by the between variation:

Commodity	Food, Alcohol and Tobacco	Clothing	Leisure and Holidays	Utilities
R^2	.4794	.2301	.2189	.4843

Finally, scatterplots of the variance against the mean as well as the interquartile range against the median are shown in Figure 2 and 3 in Appendix A. The two scatterplots look similar, suggesting there are no problems with outliers. And since the mean and the variance (and the median and the interquartile range) are completely uncorrelated this suggests that there is just as much variation in total expenditures within the richer households as there is within the poorer households.

5.2 Estimation of the Matrix of Covariances

The first step consists in estimating an AID system separately for each household. The second step consists in estimating the matrix of covariances of the income effects with the levels of the budget shares as the sample covariance across households of the estimated income effects with the estimated levels of the budget shares.

5.2.1 Step 1: Estimation of the Demand System

Estimating an AID system separately for each household amounts to estimating (9) subject (**??**), (6) and (7) for each household. The AID system is nonlinear in parameters, because the price index contains parameters. If there were no parameters to be estimated in the price index, the system would be linear in parameters. The usual way of estimating the AID is an iterative procedure which exploits this: In an initial step, the parametric price index is replaced by the Stone price index and the resulting linear model is estimated. The parameter estimates of this initial step are then used to calculate the parametric price index, which is then inserted in place of the Stone price index. A new set of parameters is estimated and the parametric price index is recalculated. This iterative procedure is continued until the parameter estimates converge (converge in the sense that they do not change from iteration to iteration) [21].

Adding up is satisfied by leaving out one good. The left out good is Utilities, so the three commodity groups are "Food, Alcohol & Tobacco", "Clothing" and "Leisure & Holidays". The included explanatory variables are log total expenditure, relative log prices, quarterly dummies and a constant term. For the households that experience changes to their family size during the survey period we include also family size among the explanatory variables (we can only identify both the constant intercept and the coefficient on

[21] We set $\alpha_{0h} = 0$ for all h, since α_0 in the AID system is not identified.

family size for a given household if there are changes to family size within that household).

Homogeneity (10) and symmetry (**??**) are imposed in each iteration in the following way. Suppressing the household index, let $w_i = (w_{i1}, ..., w_{iT})'$ denote the vector of budget shares for commodity i, let $\ln x = (\ln x_1, ..., \ln x_T)'$ denote the vector of deflated log total expenditure [22], let $\ln \widetilde{p}_i = (\ln \widetilde{p}_{i1}, ..., \ln \widetilde{p}_{iT})'$ denote the vector of the relative log prices for good i, let $Dq1$, $Dq2$ and $Dq3$ denote the quarterly dummies for the quarters 1, 2 and 3 respectively, let $z = (z_1, ..., z_T)'$ denote the vector of family size, let $\varepsilon_i = (\varepsilon_{i1}, ..., \varepsilon_{iT})'$ denote the vector of error terms for commodity i, let $\mathbf{1}$ denote the T-vector of ones and let θ denote the vector of parameters. Then the AID system for a given household with adding up, homogeneity and symmetry imposed can be written as

$$\begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} = \begin{bmatrix} X & P & Dq & \mathbf{1} & Z \end{bmatrix} \theta + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{bmatrix}$$

where the data is organised as follows

$$X = \begin{bmatrix} \ln x & 0 & 0 \\ 0 & \ln x & 0 \\ 0 & 0 & \ln x \end{bmatrix}, \quad P = \begin{bmatrix} \ln \widetilde{p}_1 & \ln \widetilde{p}_2 & \ln \widetilde{p}_3 & 0 & 0 & 0 \\ 0 & \ln \widetilde{p}_1 & 0 & \ln \widetilde{p}_2 & \ln \widetilde{p}_3 & 0 \\ 0 & 0 & \ln \widetilde{p}_1 & 0 & \ln \widetilde{p}_2 & \ln \widetilde{p}_3 \end{bmatrix}$$

and

$$Dq = \begin{bmatrix} Dq1 & Dq2 & Dq3 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & Dq1 & Dq2 & Dq3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & Dq1 & Dq2 & Dq3 \end{bmatrix}$$

and

$$\mathbf{1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad Z = \begin{bmatrix} z & 0 & 0 \\ 0 & z & 0 \\ 0 & 0 & z \end{bmatrix}$$

and where accordingly the parameters are organised as $\theta = (\beta_1, \beta_2, \beta_3, \gamma_{11}, \gamma_{12}, \gamma_{13}, \gamma_{22}, \gamma_{23}, \gamma_{33}, \delta_1^{dq1}, \delta_2^{dq1}, \delta_3^{dq1}, \delta_1^{dq2}, \delta_2^{dq2}, \delta_3^{dq2}, \delta_1^{dq3}, \delta_2^{dq3}, \delta_3^{dq3}, \alpha_1, \alpha_2, \alpha_3, \delta_1^z, \delta_2^z, \delta_3^z)'$ and where the error term structure is given by

$$E\left[\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{bmatrix} \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{bmatrix}'\right] = \begin{bmatrix} \sigma_{11}^2 & \sigma_{12}^2 & \sigma_{13}^2 \\ \sigma_{12}^2 & \sigma_{22}^2 & \sigma_{23}^2 \\ \sigma_{13}^2 & \sigma_{23}^2 & \sigma_{33}^2 \end{bmatrix} \otimes I_{24}$$

[22] In the first iteration, log total expenditure is deflated with the Stone price index. In the following iterations, log total expenditure is deflated with the parametric price index calculated with the parameter estimates of the previous iteration. The point is that in each iteration, the price index is fixed which implies that the model is linear in parameters in each iteration.

We have 24 quarters for each household, so the vector of all three budget shares $(w_1, w_2, w_3)'$ is 72×1, the matrix of regressors $\begin{bmatrix} X & P & Dq & \mathbf{1} & Z \end{bmatrix}$ is 72×21 if family size is not included and 72×24 if family size is included, and accordingly the vector of parameters θ is 21×1 if family size is not included and 24×1 if family size is included and the vector of error terms $(\varepsilon_1, \varepsilon_2, \varepsilon_3)'$ is 72×1.

The AID system is estimated for each household by the iterative procedure described above. In each iteration we estimated the model by OLS (which is consistent, but not efficient due to the cross-equation retrictions imposed by symmetry). The outcome of this first step estimation is 87 sets of parameter estimates of $\theta : \widehat{\theta}_h,\ h = 1, ..., 87$.

For the whole demand system for all households, the total number of observations is $72 \cdot 87 = 6264$, and since family size is included for 36 households, the total number of parameters to be estimated is $36 \cdot 24 + 51 \cdot 21 = 1935$.

5.2.2 Step 2: Calculation of the matrix of covariances

The figures needed for estimating the matrix of covariances are the estimated income effects and the estimated budget shares for each of the three commodities for each of the 87 households. The estimated income effects are simply the estimates of the coefficients on log total expenditure, $\left(\widehat{\beta}_{ih}\right)_{i=1 h=1}^{3 \quad 87}$. We estimate the budget share for good j for household h by the mean over time of the predicted budget shares for good j for household h, i.e.

$$\hat{w}_{ih} = \frac{1}{T} \sum_{t=1}^{T} \hat{w}_{iht}, \qquad i = 1, ..., 3,\ \ h = 1, ..., H.$$

The covariances of income effects with budget shares can now be calculated as the sample covariances across households of the estimated income effects with the estimated budget shares. Entry (i, j) of M is thus calculated as

$$M_{ij} = \frac{1}{H} \sum_{h=1}^{H} \hat{\beta}_{ih} \hat{w}_{jh} - \left(\frac{1}{H} \sum_{h=1}^{H} \hat{\beta}_{ih} \right) \left(\frac{1}{H} \sum_{h=1}^{H} \hat{w}_{jh} \right) \qquad i, j = 1, ..., 3.$$

5.3 Empirical Results

In this section we present and discuss the estimation results of the first and second stage estimations and present the test results from the tests of symmetry and positive semidefiniteness of the matrix of covariances M.

5.3.1 Estimation Results from Step 1

The output from the first stage estimations is 87 sets of estimates of the parameters of the AID system, one set for each household. Firstly, we compare our estimation results with the usual estimates from an AID system (i.e. the estimates from a pooled AID model) to get an idea of whether our individual-specific estimates are reasonable. To this end, note that the demand for good i in the pooled AID system evaluated at unit prices (i.e. in the base year) is given by

$$w_i = \widetilde{\alpha}_i + \widetilde{\beta}_i \ln x, \quad i = 1, 2, 3,$$

where $\widetilde{\alpha}_i$ and $\widetilde{\beta}_i$ are the parameters of the pooled system. And that the aggregate demand for our system of household-specific AID systems, evaluated at unit prices, is given by

$$G(\ln x) = \mu_\alpha + \mu_\beta \ln x,$$

where μ_α and μ_β are the means of the corresponding household-specific parameters α_{ih} and β_{ih}, $h = 1, ..., H$. This means that the Engel curve for good i in the base year resulting from the estimation of the pooled AID model should be similar to the "Engel curve" for good i in the base year with intercept and slope coefficients given by the mean of the estimated household-specific intercepts and slopes. The estimates of the intercept and slope for the pooled model as well as the means of the household-specific estimates of intercepts and slopes are given in the table below [23]:

	Pooled AID	Mean of household-specific AID systems
Intercept coefficient for good 1 (α_1)	1.57835	1.39704
Intercept coefficient for good 2 (α_2)	.32207	.00392
Intercept coefficient for good 3 (α_3)	-.22152	-.08746
Slope coefficient for good 1 (β_1)	-.08883	-.07080
Slope coefficient for good 2 (β_2)	-.01795	.00980
Slope coefficient for good 3 (β_3)	.02426	.01251

The two sets of Engel curves for the three goods are shown in Appendix B. As can be seen, the "Engel curve" with the mean estimates is similar to the Engel curve for the pooled model for all three goods. We therefore conclude that our household-specific estimates are reasonable and now turn to examining them in more detail.

[23] The corresponding table containing all the parameter estimates is shown in Table C.1 in Appendix C.

Histograms of the parameter estimates are depicted in Appendix B. It appears that there are quite large differences in the parameter estimates across households, suggesting that there is considerable unobservable heterogeneity in our sample. We test whether this is indeed the case for a selected set of parameters. The most important parameters for our integrability test are the income effects β_1, β_2 and β_3, because, as shown previously in Example 2 in Section 3, identical income effects across households for each good implies that the matrix of covariances M is symmetric (exact linear aggregation or Gorman aggregation). We test for identical income effects across households by an asymptotic F-test. The null hypothesis is

$$\mathrm{H}_0 : \quad \text{For each } i : \beta_{ih} = \beta_i \text{ for all } h,$$

which imposes $3 \cdot 86 = 258$ restrictions. Let SSR_{ur} denote the sum of squared residuals in the unrestricted model and let SSR_r denote the sum of squared residuals in the restricted model. The F test statistic is then given by

$$F = \frac{6274}{1935} \cdot \frac{SSR_r - SSR_{ur}}{SSR_{ur}}$$

which is asymptotically chi-square with 258 degrees of freedom. We get a test statistic of $F = 807$ which results in a p-value of nearly zero $(2.36 \cdot 10^{-67})$, so we strongly reject income effects being identical across households. This means that if the matrix of covariances M turns out to be symmetric, it is *not* because we have Gorman aggregation.

We also test whether the effects of the quarterly dummies are identical across households. There are two reasons for being interested in testing this. Firstly, since the quarterly dummies represent macro shocks one could think their effects to were similar across households. Secondly, if the seasonal pattern effects are in fact identical across households it would reduce the total number of parameters to be estimated by 774 to 1161, which would increase precision of the parameter estimates considerably. We test for identical seasonal patterns across households by an asymptotic F-test. The null hypothesis of identical seasonal pattern effects is

$$\mathrm{H}_0 : \quad \text{For each good } i \text{ and quarter } j,\ i, j = 1, 2, 3 : \delta_{ih}^{dqj} = \delta_i^{dqj} \text{ for all } h,$$

which imposes $9 \cdot 86 = 774$ restrictions, so the F test statistic is asymptotically chi-square with 774 degrees of freedom. We get a test statistic of $F = 1363$, which results in a p-value of nearly zero $(4.275 \cdot 10^{-35})$, so we strongly reject that the seasonal patterns in demands are identical across households. This means that different households adjust their budget shares differently over the year. One example of this could be the budget share for Leisure&Holiday: Some households may prefer to spend more on winter holidays than on summer holidays, and vice versa. Another example could be the budget share for Food, Alcohol & Tobacco: Some households may spend relatively more on celebrating christmas, relative to their consumption of Food, Alcohol & Tobacco during the rest of the year, than other households.

Recall that one of our arguments for modelling the distribution of preference heterogeneity nonparametrically was that any parametric restrictions would be totally unfounded. *If* we were to model the distribution of preference heterogeneity parametrically, an obvious choice of a parametric form would be the normal distribution. The histograms of the parameter estimates are the naive non-parametric density estimates of the marginal distributions of the parameters. As can be seen from the histograms of the parameter estimates in Appendix B, not all the marginal distributions are close to fit the normal, and indeed, skewness and kurtosis tests for normality strongly rejects normality for the parameters β_2, β_3, α_2, α_3, γ_{22}, γ_{23} and γ_{33} [24]:

Parameter	β_1	β_2	β_3	α_1	α_2	α_3	γ_{11}	γ_{12}	γ_{13}	γ_{22}	γ_{23}	γ_{33}
p-value	.49	0	.05	.36	0	0	.10	.54	.21	0	0	0

This suggests that modelling the distribution of preference heterogeneity by a normal distribution would be a misspecification. However, eyeballing the histograms, it seems that it would be fair to assume that all the marginal distributions are unimodal. So, if we were to model the distribution of preference heterogeneity parametrically, we should choose a distribution with unimodal marginal distributions and that allows for skewness in the marginal distributions.

Next, we examine whether β_i is correlated with the set of all α's and γ's for any i. The reason we are interested in this specific correlation is that, as shown in Example 3 in Section 3, if the income effect for commodity i is conditionally independent of all the α's and the γ's for all i, then the matrix of covariances M is symmetric. We examine this by regressing β_i on the α's and γ's, controlling for mean family size, and then testing if any of the α's or γ's are significant. Results are reported in the table below, where * indicates that the variable is significant:

[24] The null hypothesis is that the variable is normal. The test statistic is based on tests for whether skewness and kurtosis are significantly different from those of the normal. The test statistic is $\chi(2)$.

	α_1	α_2	α_3	γ_{11}	γ_{12}	γ_{13}	γ_{22}	γ_{23}	γ_{33}	$\{\alpha_1, \alpha_1, \alpha_1\}$	$\{\gamma_{11}, \gamma_{12}, \gamma_{13}$ $\gamma_{22}, \gamma_{23}, \gamma_{33}\}$
β_1	*					*				*	
β_2		*								*	
β_3			*			*		*	*	*	*

The income effect for good i is significantly correlated with the intercept for good i for all goods. Furthermore, all three price effects are significantly correlated with the income effect for Leisure&Holidays. This we take as evidence of income effects not being uncorrelated with the α's and γ's. This implies that income effects are not independent of the α's and γ's. This in turn means that if the matrix of covariances, M, turns out to be symmetric, it is *not* because the income effects are independent of all the other parameters in the demand system.

An underlying assumption of the integrability result in Lewbel (2001) about the conditional distribution of preference heterogeneity is that it is conditionally independent of prices and total expenditure, conditional on demographics. In our model this translates into the conditional distribution of (α, β, Γ) being independent of prices and total expenditures, conditional on demographics. We have not imposed this assumption in estimation because of our completely non-parametric approach, but we check it by regressing each of the parameter estimates on the means of deflated log total expenditure, controlling for mean family size. As can be seen from the t statistics from these regressions, the mean total expenditure is not significant for any of the parameters:

Parameter	β_1	β_2	β_3	α_1	α_2	α_3	γ_{11}	γ_{12}	γ_{13}	γ_{22}	γ_{23}	γ_{33}
t[25]	.04	.12	-1.46	-.08	-.41	1.57	-.86	.12	-1.54	.61	1.09	.65

This we take as evidence in favor of the independence assumption [26].

In summary, we have found firstly that when comparing our household-specific estimates with the pooled estimates it seems that our estimates are reasonable. Secondly, we have found that there is a significant large amount of unobserved heterogeneity in preferences: Tests of identical income effects and identical seasonal effects across consumers were strongly rejected. We also found that if we were to model the distribution of preference heterogeneity parametrically by some assumed disitribution, our non-parametric

[25] t-statistic on mean total expenditure.

[26] Also, Hausman tests for random effects versus fixed effects do not reject the random effects specification for any of the commodities. Although the models are different, this also points to that the unobservable heterogeneity is uncorrelated with prices and total expenditure.

approach suggests that this distribution should have unimodal marginal distributions, allow for skewness in the marginal distributions and allow for correlations between the coefficients (that is, allow for that for example β_1 is correlated with γ_{13}).

5.3.2 Estimation Results from Step 2

The estimate of the matrix of covariances M is (all elements are multiplied by 1000)[27]:

$$\widehat{M} = \begin{bmatrix} 2.8394 & .4043 & -.7723 \\ -.4596 & -.1278 & .4398 \\ .8685 & -.2064 & .1450 \end{bmatrix}$$

When eyeballing these numbers, the first thing one notices is that the off-diagonal elements seem similar in absolute magnitude, but have opposite signs. We have not been able to find any interpretation of the *magnitudes* of the elements of M.

5.3.3 Tests of Symmetry and Positive Semidefiniteness

In this section we test whether the matrix of covariances is symmetric and positive semidefinite. In order to make inference we need the variance-covariance matrix of $\widehat{M}$. We estimate the variance-covariance matrix of $\widehat{M}$ by bootstrap.

The bootstrap is a method for estimating the distribution of an estimator or a test statistic or features of that distribution by resampling from the data. The bootstrap treats the data as if they were the population and requires sampling with replacement from the data. Bootstrap sampling can be done in different ways, either by resampling directly from the data (non-parametric bootstrap) or by resampling from a model estimated from the data (parametric bootstrap).

We do a non-parametric bootstrap and sample from the data. We sample in clusters, that is, we sample households and for each household we sample, we sample all three budget shares. By sampling households we maintain the (true) covariance between time periods within a household. By sampling all three budget shares we ensure that the adding up restriction on budget shares holds in each bootstrap sample. We draw $B =$ 10.000 bootstrap samples and calculate the variance-covariance matrix of $\widehat{M}$ as the sample

[27] Row 1 is Food, Alcohol & Tobacco; row 2 is Clothing and row 3 is Leisure & Holidays.

covariance of the boostrap samples. Let M also denote the vectorized matrix of M :

$$M = \begin{bmatrix} M_{11} \\ M_{21} \\ M_{31} \\ M_{12} \\ M_{22} \\ M_{32} \\ M_{13} \\ M_{23} \\ M_{33} \end{bmatrix}$$

Let $\mathrm{V}(M)$ denote the variance-covariance matrix of M. To take an example, the $(2,3)$'th element of $\widehat{\mathrm{V}}(\widehat{M})$ is then calculated as

$$\widehat{\mathrm{Cov}}(\widehat{M}_{21}, \widehat{M}_{31}) = \frac{1}{B-1}\sum_{b=1}^{B}\left(\widehat{M}_{21}^{b}\widehat{M}_{31}^{b}\right) - \left(\frac{1}{B-1}\sum_{b=1}^{B}\widehat{M}_{21}^{b}\right)\left(\frac{1}{B-1}\sum_{b=1}^{B}\widehat{M}_{31}^{b}\right).$$

Our bootstrap estimate of the variance-covariance matrix of M is (all elements are multiplied by 10^6 so that they correspond in magnitude to $1000 \cdot M$) :

$$\hat{\mathrm{V}}(M) = \begin{bmatrix} 2.7167 & -.5694 & -.0744 & -.1544 & .0707 & .0587 & -.3588 & .1252 & -.0061 \\ & .8748 & -.1760 & .0757 & -.1267 & .0375 & .1369 & -.1049 & .04132 \\ & & .8127 & .0231 & .0436 & -.1056 & .0035 & .0305 & -.0915 \\ & & & .3131 & -.0761 & -.0433 & .0202 & -.0060 & .0036 \\ & & & & .1244 & -.0023 & -.0071 & -.0023 & -.0113 \\ & & & & & .1125 & -.0010 & -.0088 & -.0116 \\ & & & & & & .1863 & -.0386 & -.0244 \\ & & & & & & & .0586 & .0004 \\ & & & & & & & & .0589 \end{bmatrix}$$

The estimate of M is repeated below, now with the bootstrapped standard errors in parentheses:

$$\widehat{M} = \begin{bmatrix} \underset{(1.6482)}{2.8394} & \underset{(.5596)}{.4043} & \underset{(.4316)}{-.7723} \\ \underset{(.9353)}{-.4596} & \underset{(.3527)}{-.1278} & \underset{(.2421)}{.4398} \\ \underset{(.9015)}{.8685} & \underset{(.3354)}{-.2064} & \underset{(.2426)}{.1450} \end{bmatrix}$$

As can be seen immediately, none of the elements of $\widehat{M}$ are significantly different from zero on a 5 percent significance level. This means that on a 5 percent significance level we can not reject that M is a matrix of zeroes in which case M is trivially symmetric and positive semidefinite. Below we report the χ^2-statistics of the nine individual tests:

$$H_0 : M_{ij} = 0, \quad i, j = 1, .., 3,$$

where * indicates that the element is significant on a 10 percent significance level[28]:

Element in $\widehat{M}$	$\widehat{M}_{11}$	$\widehat{M}_{21}$	$\widehat{M}_{31}$	$\widehat{M}_{12}$	$\widehat{M}_{22}$	$\widehat{M}_{32}$	$\widehat{M}_{13}$	$\widehat{M}_{23}$	$\widehat{M}_{33}$
χ^2-statistic	2.965*	.241	.927	.523	.132	.380	3.204*	3.302*	.360

Recall that we are interested in whether M is symmetric. To this end, the diagonal elements of M are thus irrelevant. Two of the off-diagonal elements of $\widehat{M}$, however, are different from zero on a 10 percent significance level, namely $\widehat{M}_{13}$ and $\widehat{M}_{23}$. We therefore do the separate t-tests for whether the off-diagonal elements are different. The null hypothesis is:

$$H_0 : M_{ij} - M_{ji} = 0, \quad i \neq j.$$

A rejection of H_0 for just one set of (i, j) is enough to reject symmetry of M. The t-statistics for these three tests are, again with * indicating significance at the 10 percent significance level:

Element	$M_{12} - M_{21}$	$M_{13} - M_{31}$	$M_{23} - M_{32}$
χ^2-statistic	.721	2.716*	2.214

For completeness, we report a joint test of whether M is the zero matrix. The null hypothesis for this test can be written

$$H_0 : RM = 0,$$

where $R = I_9$ and M denotes the vectorized matrix of M. The Wald test statistic for H_0 is then given by

$$W_0 = \left[R\widehat{M}\right]' \left[R\,\mathrm{V}(\widehat{M})\,R'\right]^{-1} \left[R\widehat{M}\right],$$

which is asymptotically chi sqaure with 9 degrees of freedom[29]. We get $\widehat{W}_0 = 8.388$, i.e. we can not reject that M is the zero matrix, even on a 10 percent significance level.

This means that we can not reject that M is symmetric, because we can not reject that M is the zero matrix. This in turn trivially implies that M is positive semidefinite.

[28] The critical value on 10% significance level in the $\chi^2(1)$ distribution is 2.706.

[29] The critical value in the $\chi^2(9)$ distribution at significance level 5% is 16.919 and on significance level 10% it is 14.684.

6 Conclusion and Further Work

In this paper we have presented a panel data test of integrability of the conditional mean demand in a random utility model. A uniquely long panel data set on household expenditures allowed us to model individual demands as a set of completely heterogeneous Almost Ideal Demand systems by allowing all the parameters of the demand system to be household-specific. We do not reject symmetry and negativity of the conditional mean demand, our test result in a p-value of 0.76. Hence we conclude that a set of completely heterogeneous Almost Ideal Demand systems generates a conditional mean demand that is integrable.

The value of having integrable conditional demands is that it fascilitates welfare analysis. Suppose for example that the government lowers the VAT on healthy foods; from the demand system estimates alone we can assess quantitatively how demands will change, but in order to assess how the welfare of the consumers change, we need a utility framework. Given that integrability is not rejected, we thus know that the conditional mean demand is generated by utility maximisation of some utility function. This paper, however, does not provide this utility function. We saw in Section 3 that the conditional mean demand generated by a set of completely heterogeneous Almost Ideal Demand systems is not in itself an Almost Ideal Demand system, hence the utility function underlying the conditional mean demand is not that of an Almost Ideal Demand system. This leaves us with the question: Given that integrability is not rejected, what can we use this conditional mean demand for? We derived an expression for the conditional mean demand in Section 3.2. Using this expression we can calculate price and income elasticities for the conditional mean demand and so we can do positive analysis. We can not perform welfare analysis directly using the conditonal mean demand without knowing the form of the utility function generating it, but we *can* perform welfare analysis at the individual level, since we have estimated an Almost Ideal Demand system for each household. Hence, one could compare results of policy analyses based on the usual pooled model (i.e. not taking unobservable heterogeneity into account) to results based on the separate estimates for each household; for example, calculations of the compensating variations following a price change. A future version of the paper will contain such calculations. These will be illustrative rather than deep policy analysis, given the selection criteria of the sample, but they illustrate the difference between taking unobservable heterogeneity into account and not taking it into account.

Finally, a discussion of the power of our test of integrability is in place. Looking at the standard errors of $\widehat{M}$, it is clear that symmetry and positive definiteness of $\widehat{M}$ are not rejected because the standard errors are large enough to not reject that M is the zero matrix. Though some people may argue that with a sample size of 87 households, a significance level of 10% is appropriate, in which case our test would reject integrability.

A first inclination would be to abandon the non-parametric modelling approach for the distribution of unobservable heterogeneity and instead impose a distributional assumption, e.g. joint normality, on the distribution of unobservable heterogeneity and estimate the model by maximum likelihood instead. This approach, however, entails several problems. Firstly, our theoretical examples in Section 3 showed how assumptions on the distribution of preference heterogeneity could lead to integrability of the conditional mean demand by assumption. This means that if we were to chose a distributional form of the preference parameters, we need to allow for a general covariance-structure in that distribution (i.e. that we need to allow at least for the income effects to correlate with other parameters). And our tests of normality of the marginal distributions of parameters indicate that we may need a joint distribution that allows for more skewness than the normal does. All estimates of the marginal distributions, however, clearly feature a unimodal distribution. Finding such a distribution would not be a big problem. A more complicated issue would be to find a distribution from which the coefficients could be randomly drawn and yet always yield integrability at the individual level. It is not at all clear which distribution would ensure this. This highlights the motivation for the chosen modelling strategy of estimating every household separately instead of imposing a random utility model across households.

Another way to potentially obtain more efficient estimates would be to restrict some of the parameters to being identical across households. For example, one could restrict the price effects, the seasonal effects and the coefficients on household size to being identical across households. This would reduce the number of parameters to be estimated from 1935 to 540. Alternatively, one could also consider heterogeneity in fewer goods; for example, one could reduce the number of goods considered to two (one inside good and one outside good), which would reduce the number of parameters to be estimated from 1935 to 179.

Finally, the large standard errors could quite simply be a small sample problem. This suggestion can be examined in two ways. Either via simulation: One could simulate a large data set on which the test rejects and then select a smaller sample (of our sample size) and check if the test then fails to reject on the smaller sample. Or one could increase the sample size by including the remaining households that are observed for all 24 quarters. This would increase the number of households from 87 to 249. The implicit assumption being made by this is that labour market status is exogenous in the allocation of total expenditures to the different goods. Further work will look into some or all of these different ways of obtaining more effecient estimates.

References

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

[2] Beckert, W. (2005): "Estimation of Heterogeneous Preferences, with an Application to Demand for Internet Services", *The Review of Economics and Statistics*, 87(3), 495-502.

[3] Blundell, R., Chen, X. and Kristensen, D. (2007): "Nonparametric IV estimation of shape-invariant Engel curves", *Econometrica* forthcoming, CeMMAP Working Paper WP15/03.

[4] Blundell, R., Duncan, A. and Pendakur, K. (1998): "Semiparametric Estimation and Consumer Demand", *Journal of Applied Econometrics*, Vol. 13, No. 5, 435-462.

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

[6] Browning, M., Crossley, T. and Weber, G. (2003): "Asking Consumption Questions in General Purpose Surveys", *The Economic Journal*, 113, Nov, F540-F567.

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

[8] Calvet, L. and Comon, E. (2003): "Behavioral Heterogeneity and the Income Effects", *The Review of Economics and Statistics,* Vol. 85(3), 653-669.

[9] Christensen, M. (2005): Essays in Empirical Demand Analysis: Evidence from Panel Data, PhD Thesis, University of Copenhagen, Red Series No. 111, ISBN 87-91342-28-7.

[10] Collado, M. D., (1998): Separability and Aggregate Shocks in the Life-Cycle Model of Consumption: Evidence from Spain, *Oxford Bulletin of Economics and Statistics*, Vol. 60, 227-247.

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

[12] Deaton, A., and J. Muellbauer (1980b): "An Almost Ideal Demand System", *American Economic Review*, Vol. 70, No.3, 312-326.

[13] Grandmont, J.-M., (1992): "Transformations of the Commodity Space, Behavioral Heterogeneity, and the Aggregation Problem", *Journal of Economic Theory*, Vol. 57, No. 1, 1-35.

[14] Hardle, W. and Jerison, M. (1991): "Cross Section Engel Curves over Time", *Recherches Economiques de Louvain*, 57, 391-431.

[15] Heckman J., and B. Singer (1984): "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data", *Econometrica*, Vol. 52, No. 2. pp. 271-320.

[16] Hildenbrand, W., (1994): "Market Demand: Theory and Empirical Evidence", Princeton University Press.

[17] Hoderlein, S., (2004): "Nonparametric Demand Systems, Instrumental Variables and a Heterogeneous Population", mimeo, Mannheim University.

[18] Horowitz, J., (2000): The Bootstrap, Handbook of Econometrics, Vol. 5, J.J. Heckman and E.E. Leamer, eds., Elsevier Science B.V., 2001, Ch. 52, pp. 3159-3228.

[19] Labeaga, J. M., and Puig (2003): "Demand system estimation with individual heterogeneity: an analysis using panel data on households", mimeo.

[20] Lewbel, A. (1991): "The Rank of Demand Systems: Theory and Nonparametric Estimation", *Econometrica*, 59, 711-730.

[21] Lewbel, A., (1999): "Consumer Demand Systems and Household Equivalence Scales", Handbook of Applied Econometrics, Blackwell Handbooks in Economics, H. Pesaran and M. Wickens (eds).

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

[23] Mas-Colell, A., (1985): "The Theory of General Economic Equilibrium: A Differentiable Approach", Econometric Society Monographs, Cambridge University Press.

[24] Mas-Colell, A., M. D. Whinston and J. R. Green (1985): "Microeconomic Theory", Oxford University Press.

[25] Muellbauer, J., (1975): Aggregation, Income Distribution and Consumer Demand, *Review of Economic Studies*, Vol. 62, 525-543.

[26] Pesaran, H., and R. Smith (1995): "Estimating long-run relationships from dynamic heterogeneous panels", *Journal of Econometrics*, Vol. 68, 79-113.

A Data

Table A.1. - *Consumption Expenditures in the ECP*

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	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 A.2. - *Demographic Variables in the ECP*

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: Husband's type of 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
heduc: Husband's education	0: illiterate 1: primary school 2: secondary school, first level 3: secondary school, second level 4: secondary school, second level, professional studies 5: university degree (3 years) 6: university degree (5 years) and PhD's 7: less than 5 years of school
tenure: Housing tenure	1: renters 2: home owners 3: free accomodation 4: not documented, presumably missing
tenure2: Housing tenure of second house	1: renters 2: home owners 3: free accomodation 4: not documented, presumable missing 6: does not have a second house

Table A.3 - *Summary statistics in full data set*

Husband's employment status:	**Percent**	**Frequency**
Employed	77,97	16.895
Retired	18,66	4044
Unemployed or out of labor force	3,37	729
Total	100	21.668
Husband's occupational status		
Wage earners with permanent job	63,56	10.738
Wage earners with temporary job	7,82	1321
Self-employed	28,61	4834
Nonpaid work	0,01	2
Total	100	16.895
Husband's education level:		
Illiterate	3,24	701
Less than 5 years of school	22,47	4869
Primary school	60,05	13.011
Secondary school	10,45	2264
University degree	3,79	823
Total	100	21.668
Housing tenure of main house:		
Renters	19,3	4319
Home owners	76,18	16.506
Other (free accomodation or missing)	3,89	843
Total	100	21.668
Housing tenure of second house:		
Does not have a second house	90,18	19.541
Own second house	8,57	1856
Rent or free accomodation or missing	1,25	271
Total	100	21.668

Table A.4 - *Summary statistics in sample*

Husband's education level:	**Percent**	**Frequency**
Illiterate	2.11	44
Less than 5 years of school	12.60	263
Primary school	69.40	1449
Secondary school	9.53	199
University degree	6.37	133
Total	100	2088
Housing tenure of main house:		
Renters	21.50	449
Home owners	73.99	1545
Other (free accomodation or missing)	4.41	94
Total	100	2088
Housing tenure of second house:		
Does not have a second house	80.51	1681
Own second house	17.10	357
Rent or free accomodation or missing	2.25	47
Total	100	2088

Table A.5 - *Sample means of expenditures and household composition in full data set* [30]

	Mean	Std.Dev
Food at home	100.434	56.319
Alcohol & tobacco	8.640	9.713
Clothing	21.336	22.790
Rent	13.298	49.246
Energy	17.331	19.622
Services	4.711	7.447
Medication	4.552	10.666
Transportation	9.198	15.989
Leisure	7.635	12.820
Education	6.431	14.459
Foodout	2.202	13.591
Holidays	6.398	29.807
Other	17.702	29.807
Durables	32.006	91.541
Total Expenditure	182.365	134.273
Husband's age	51,24	11,89
Wife's age	48,61	11,95
Number of children	1,27	1,36
Number of adults	2,55	0,80
Total household size	3,97	1,54

[30] These are pooled means and standard deviations, based on all 21668 observations.

Table A.6 - *Sample means of expenditures in sample*

	Mean	Std.Dev
Food at home	82 493	44 185
Alcohol & tobacco	7315	8650
Clothing	17 454	15 619
Rent	12 000	46 021
Energy	16 786	16 583
Services	3268	2624
Medication	3855	8607
Transportation	7815	11 334
Leisure	7044	9369
Education	6509	11 842
Foodout	16 011	27 199
Holidays	4946	13 819
Other	2125	14 097
Durables	31 114	91 290
Total expenditure	218 735	152 268
Husband's age	46.49	7.21
Wife's age	44.17	7.84
Number of children	1.63	1.35
Number of adults	2.65	2.84
Total household size	4.29	1.49

Table A.7 - *Summary statistics of demographics in sample*

Cohabitation	Number of Households
Children or cohabiting members in the household for some or all of the survey period	80
Household consists of only the husband and wife throughout the survey period	7
Total	87
Children	
Children in the household throughout the survey period	59
Children arriving in the household during the survey period	14
No children	14
Total	87
Cohabiting adults	
Adults cohabiting in the household throughout the survey period	19
Adults arriving in the household during the survey period[31]	41
No adults cohabiting in the household	27
Total	87

[31] An arrival of an adult can also be a child that turns 18 and then counts as an adult instead.

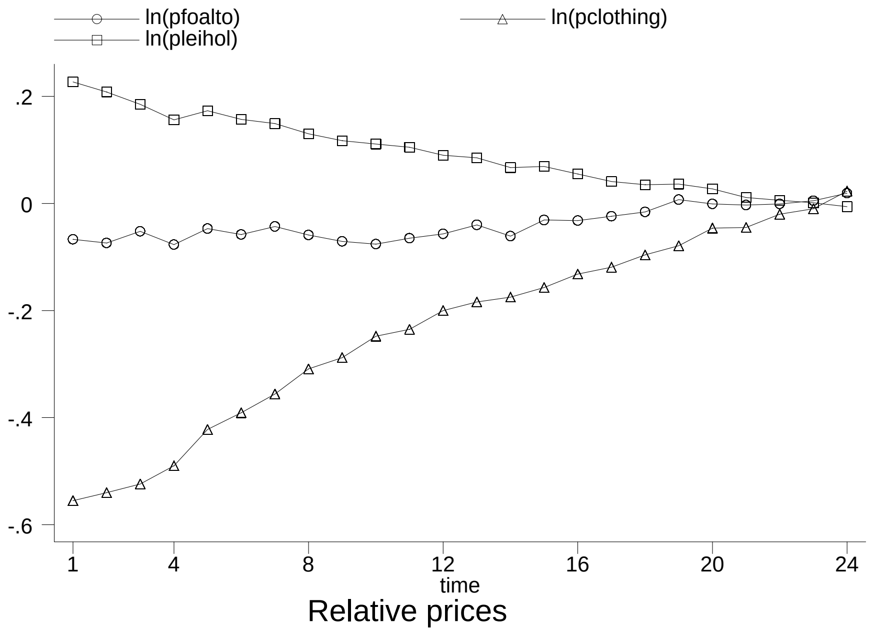


Figure 1: The variation in relative prices during the sample period.

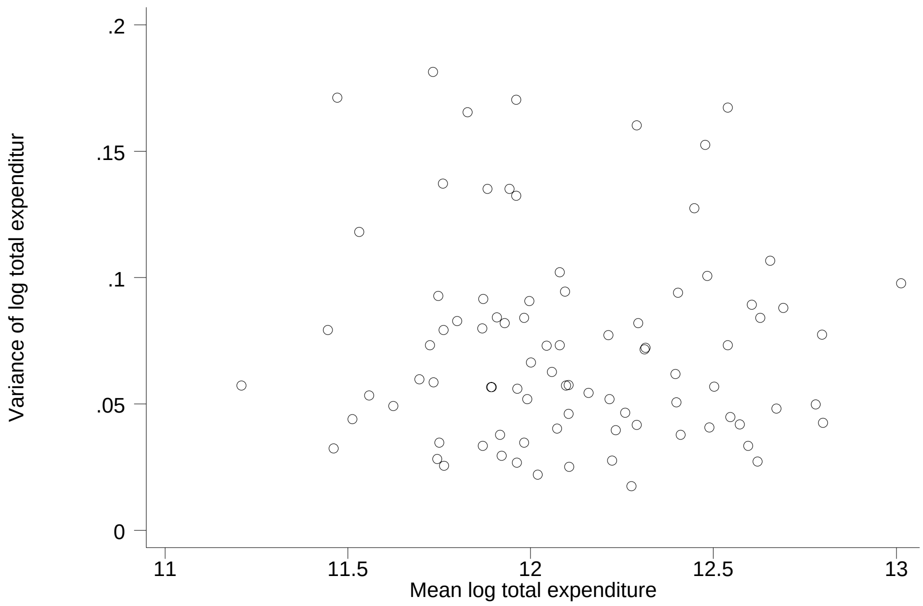


Figure 2: The correlation between the variance of log total expenditure within the household and the mean of log total expenditure within the household (thus based on 87 data points).

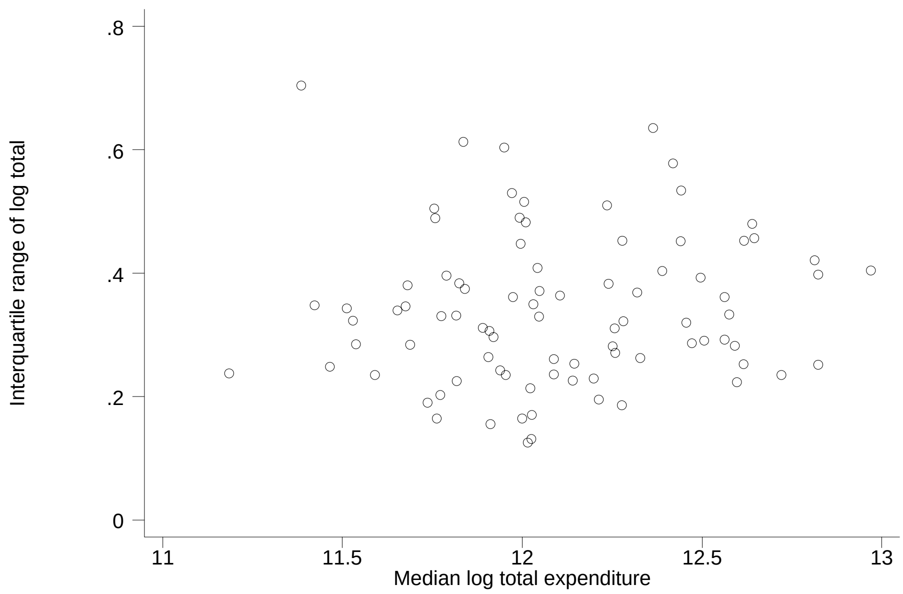


Figure 3: The correlation between the interquartile range of log total expenditure within the household and the median of log total expenditure within the household (thus based on 87 data points).

B Empirical Results

Table B.1 - *Parameter estimates from pooled AID and means of household-specific estimates*

	Pooled AID	Means of household-specific estimates
β_1	-.08883	-.07080
β_2	-.01795	.00980
β_3	.02426	.01251
γ_{11}	-.11937	-.49027
γ_{12}	-.04830	.07321
γ_{13}	-.08557	.01700
γ_{22}	-.03626	-.14315
γ_{23}	-.01537	-.03564
γ_{33}	-.10925	-.22473
δ_1^{dq1}	.00034	.00470
δ_1^{dq2}	.00684	.01175
δ_1^{dq3}	.01137	.01732
δ_2^{dq1}	-.02460	-.02281
δ_2^{dq2}	-.02031	-.02045
δ_2^{dq3}	-.04861	-.05061
δ_3^{dq1}	.00460	.00335
δ_3^{dq2}	.00280	.00213
δ_3^{dq3}	.00961	.00769
δ_1^{fam}	.00780	-.01091
δ_2^{fam}	.00357	.00287
δ_3^{fam}	-.00300	.00262
α_1	1.57835	1.39704
α_2	.32207	.00392
α_3	-.22152	-.08746

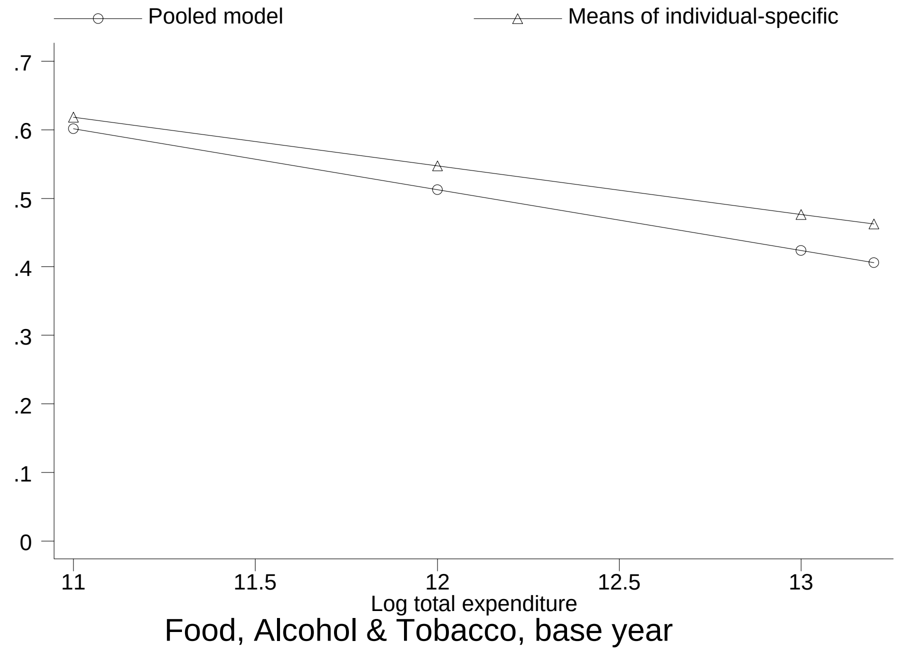


Figure 4: Engel curve for the estimated pooled model and "Engel curve" for the model with parameters given by the means of the individual-specific estimated parameters for the commodity Food, Alcohol & Tobacco.

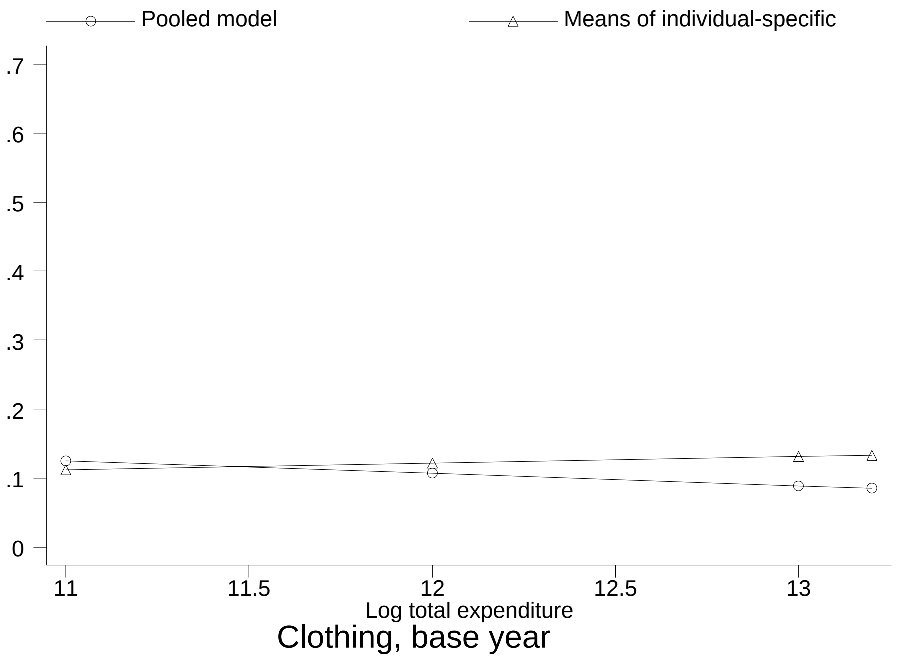


Figure 5: Engel curve for the estimated pooled model and "Engel curve" for the model with parameters given by the means of the individual-specific estimated parameters for the commodity Clothing.

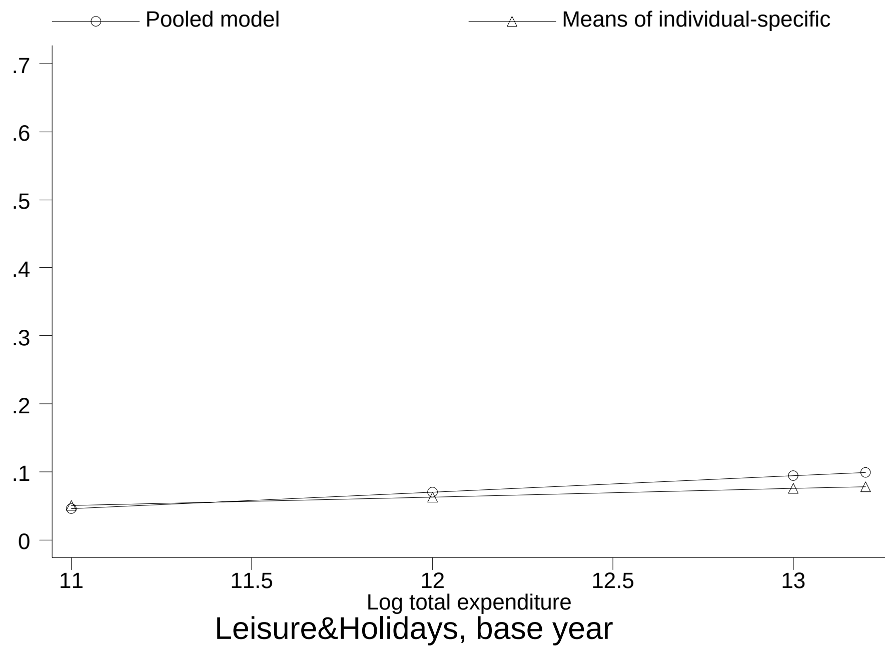


Figure 6: Engel curve for the estimated pooled model and "Engel curve" for the model with parameters given by the means of the individual-specific estimated parameters for the commodity Leisure & Holidays.

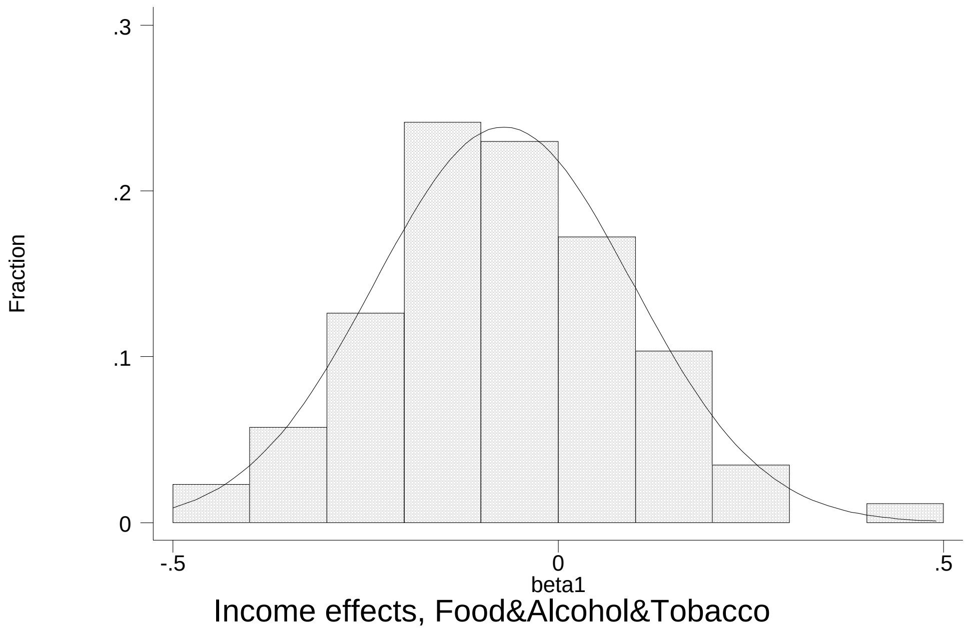


Figure 7: Histogram of the estimated individual-specific income effects $(\widehat{\beta}_{1h})_{h=1,\ldots,87}$ toghether with the normal distribution for the commodity Food, Alcohol&Tobacco.

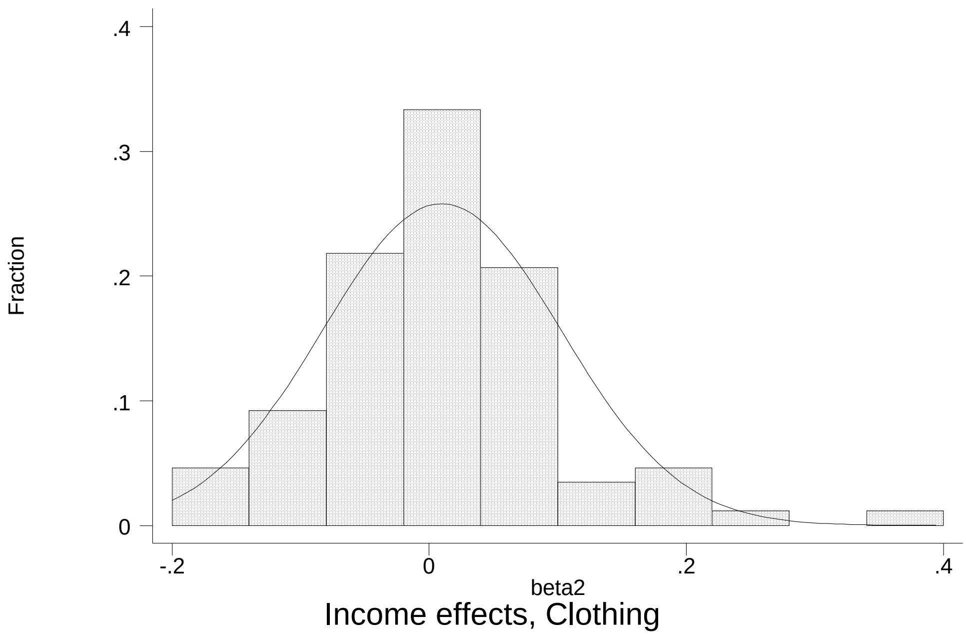


Figure 8: Histogram of the estimated individual-specific income effects $(\widehat{\beta}_{2h})_{h=1,\ldots,87}$ toghether with the normal distribution for the commodity Clothing.

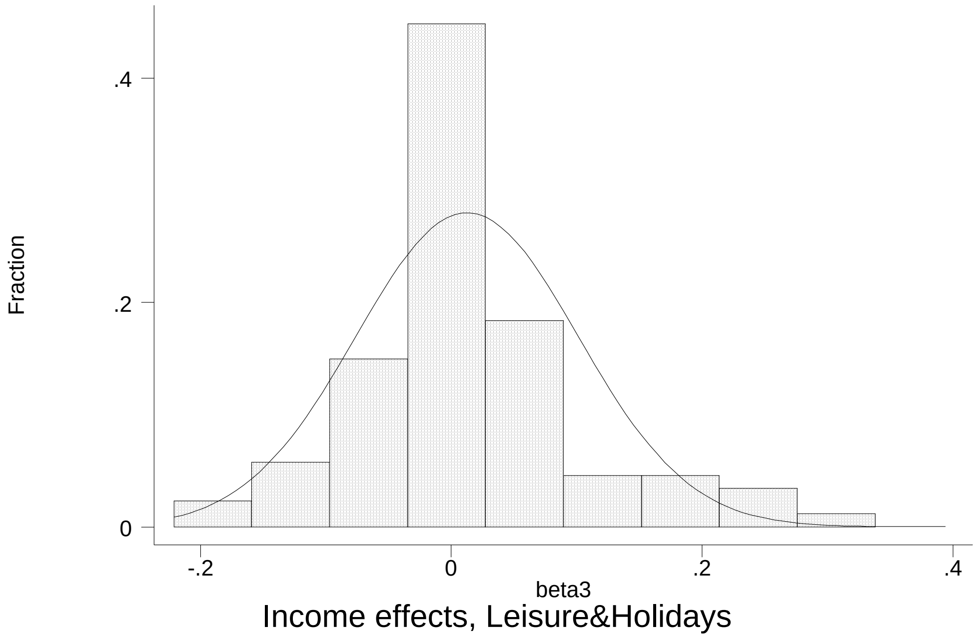


Figure 9: Histogram of the estimated individual-specific income effects $(\widehat{\beta}_{3h})_{h=1,\ldots,87}$ toghether with the normal distribution for the commodity Leisure&Holidays.

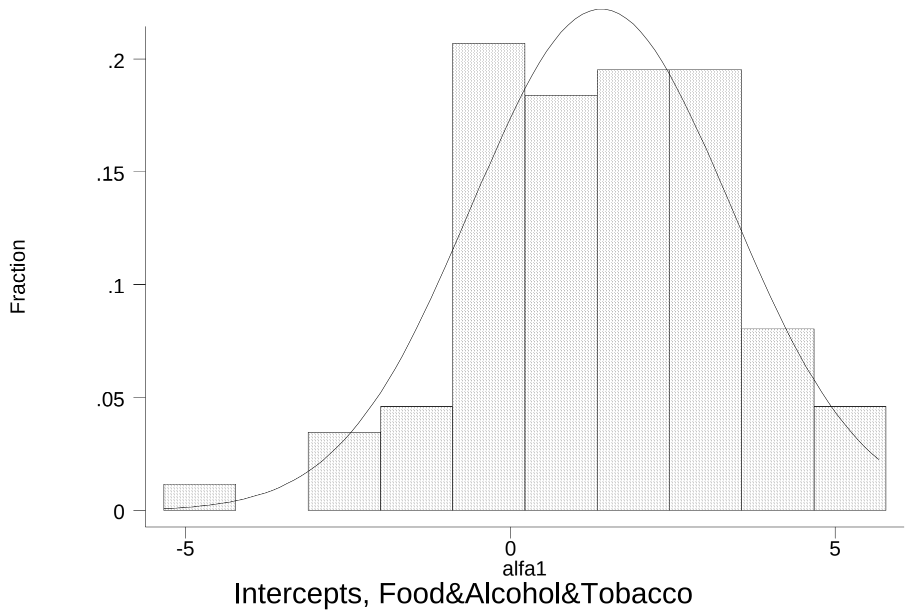


Figure 10: Histogram of the estimated individual-specific constant terms $(\widehat{\alpha}_{1h})_{h=1,\ldots,87}$ toghether with the normal distribution for the commodity Food, Alcohol&Tobacco.

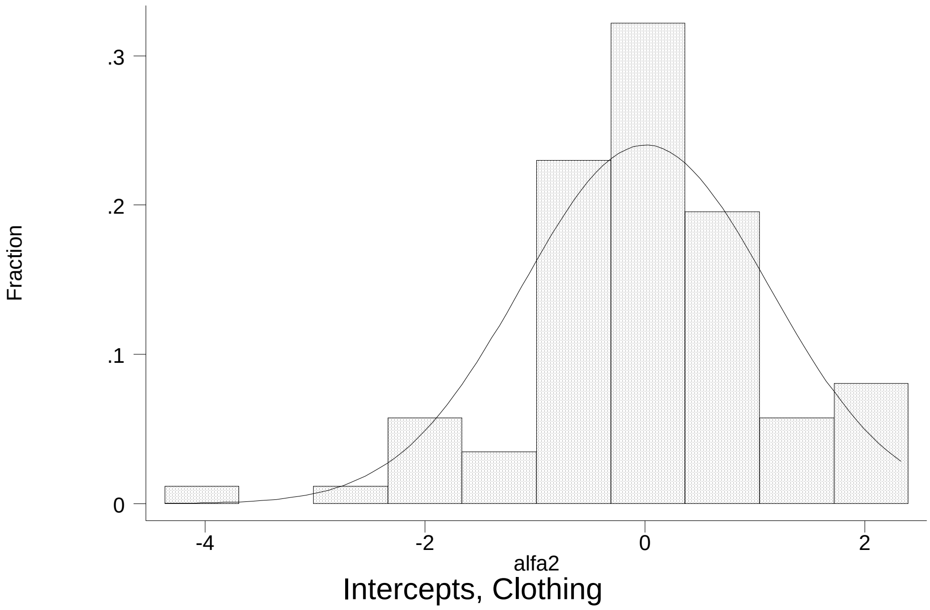


Figure 11: Histogram of the estimated individual-specific constant terms $(\widehat{\alpha}_{2h})_{h=1,\ldots,87}$ toghether with the normal distribution for the commodity Clothing.

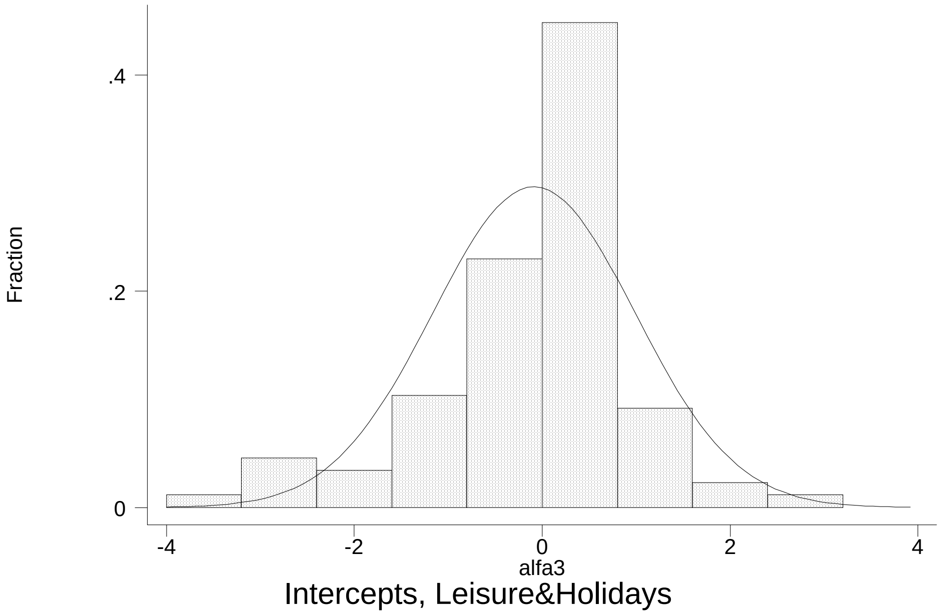


Figure 12: Histogram of the estimated individual-specific constant terms $(\widehat{\alpha}_{3h})_{h=1,\ldots,87}$ toghether with the normal distribution for the commodity Leisure & Holidays.

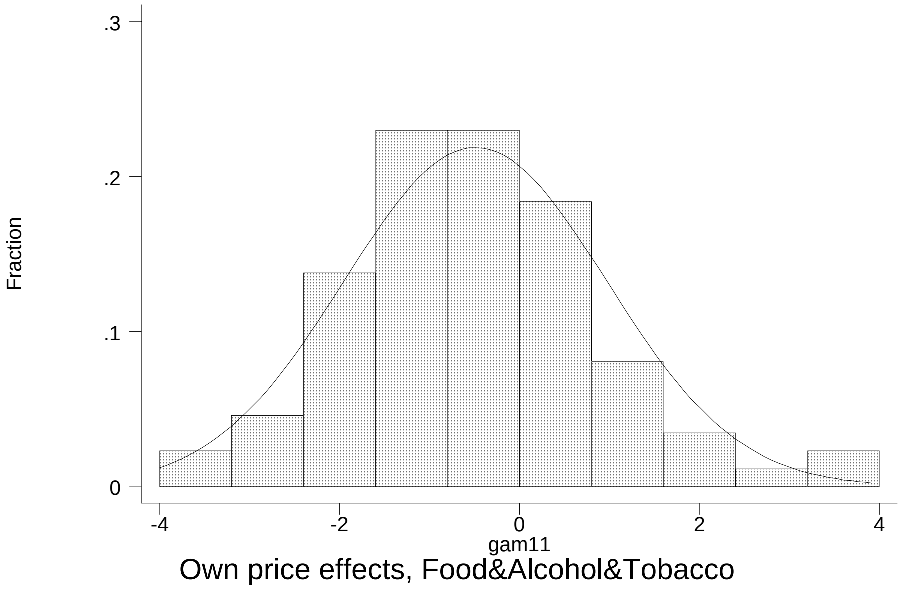
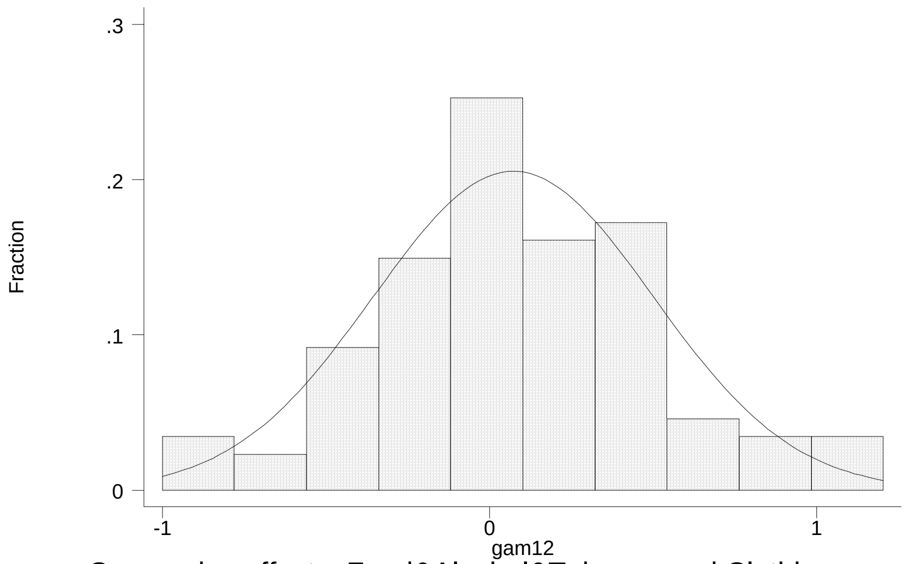


Figure 13: Histogram of the estimated individual-specific own price effects $(\widehat{\gamma}_{11h})_{h=1,\dots,87}$ toghether with the normal distribution for the commodity Food, Alcohol&Tobacco.



Cross price effects, Food&Alcohol&Tobacco and Clothing

Figure 14: Histogram of the estimated individual-specific cross price effects $(\widehat{\gamma}_{12h})_{h=1,\ldots,87}$ toghether with the normal distribution for the commodities Food, Alcohol&Tobacco and Clothing.

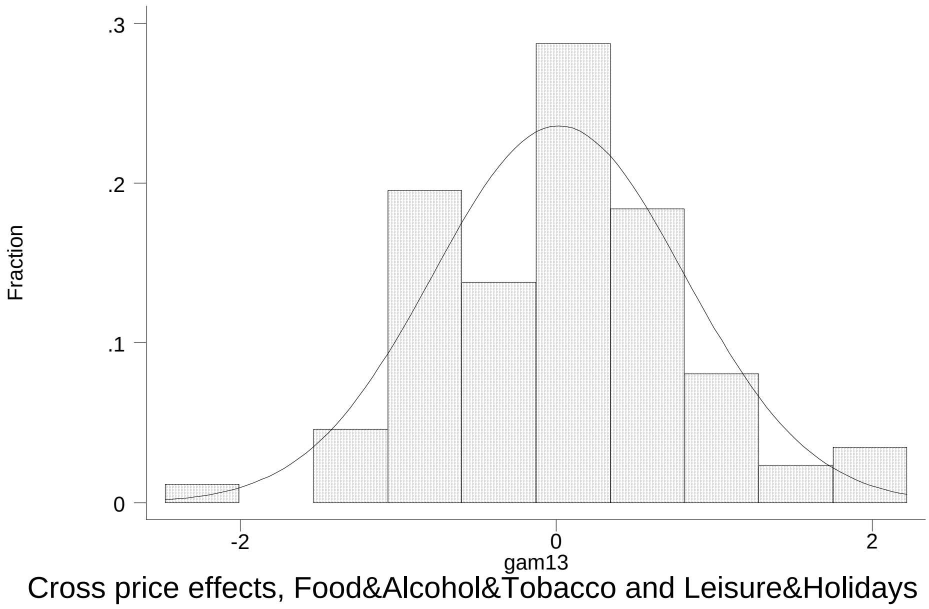


Figure 15: Histogram of the estimated individual-specific cross price effects $(\widehat{\gamma}_{13h})_{h=1,\ldots,87}$ toghether with the normal distribution for the commodities Food, Alcohol&Tobacco and Leisure & Holidays.

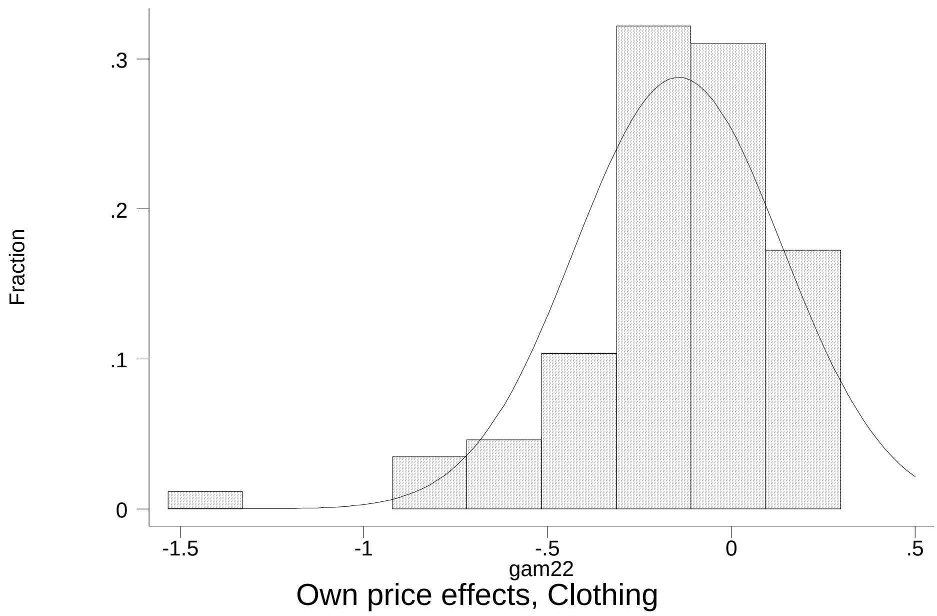


Figure 16: Histogram of the estimated individual-specific own price effects $(\widehat{\gamma}_{22h})_{h=1,\ldots,87}$ toghether with the normal distribution for the commodity Clothing.

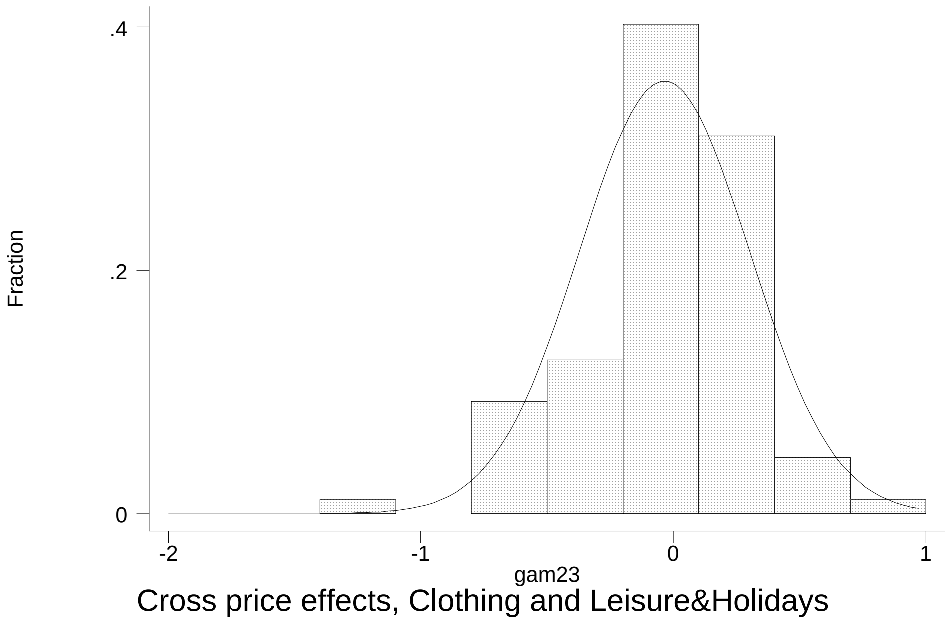


Figure 17: Histogram of the estimated individual-specific cross price effects $(\widehat{\gamma}_{23h})_{h=1,\ldots,87}$ toghether with the normal distribution for the commodities Clothing and Leisure&Holidays.

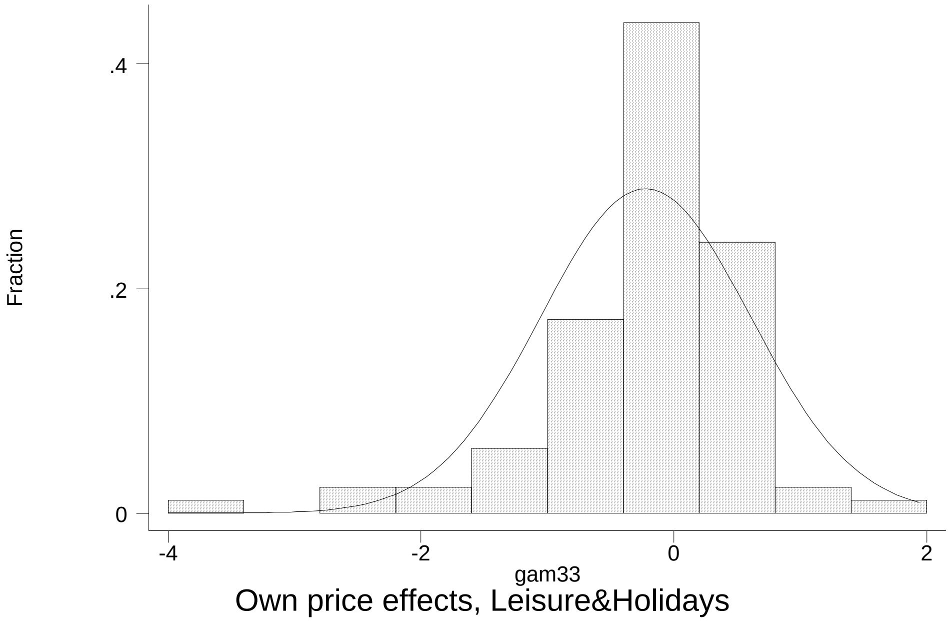


Figure 18: Histogram of the estimated individual-specific own price effects $(\widehat{\gamma}_{33h})_{h=1,\ldots,87}$ toghether with the normal distribution for the commodity Leisure & Holidays.

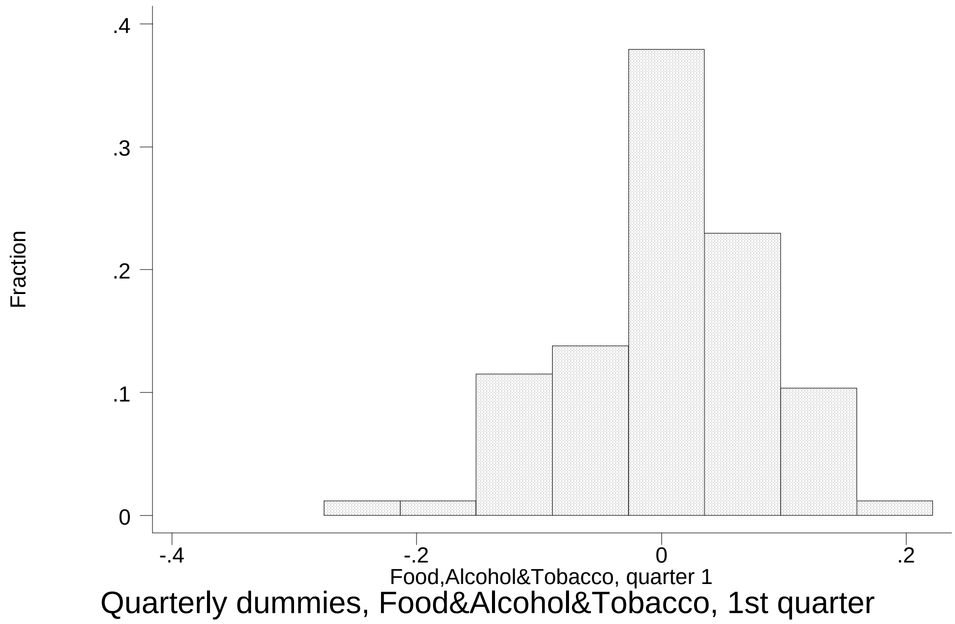


Figure 19: Histogram of the estimated individual-specific coefficient on the quarterly dummy for quarter 1 $(\widehat{\delta}_{11h})_{h=1,\ldots,87}$ for the commodity Food, Alcohol&Tobacco.

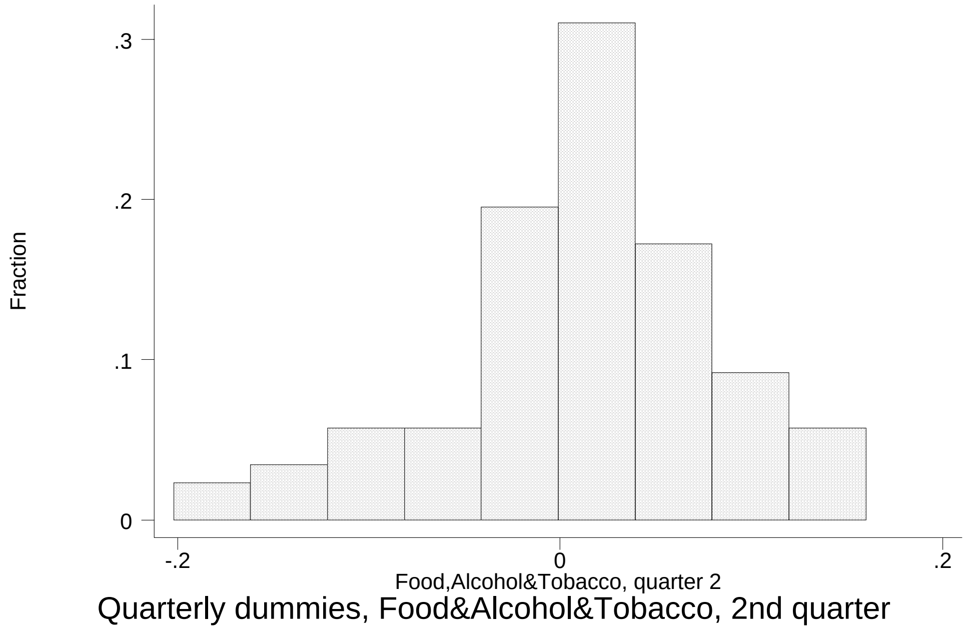


Figure 20: Histogram of the estimated individual-specific coefficient on the quarterly dummy for quarter 2 $(\widehat{\delta}_{11h})_{h=1,\ldots,87}$ for the commodity Food, Alcohol&Tobacco.

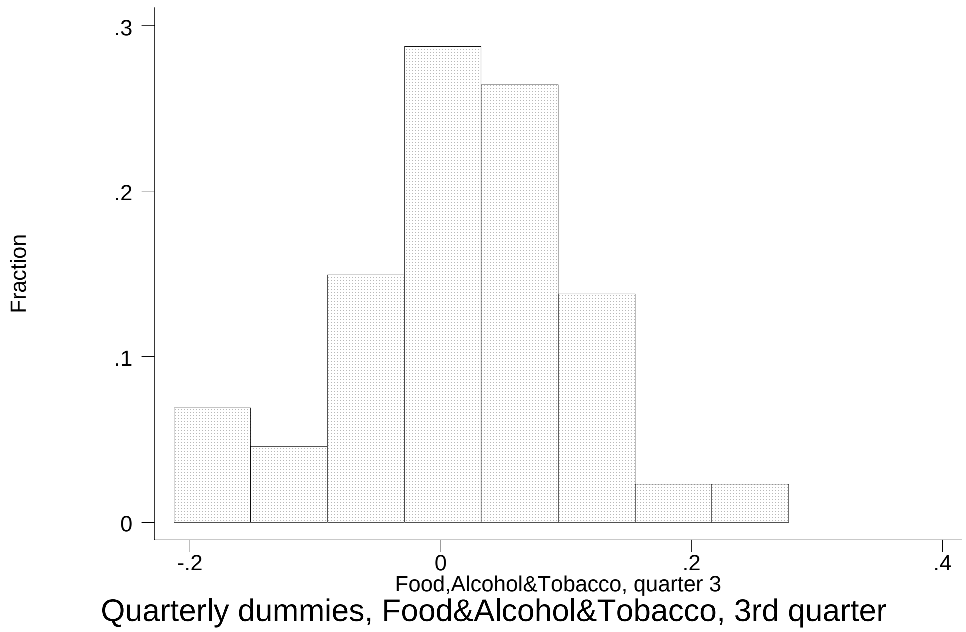
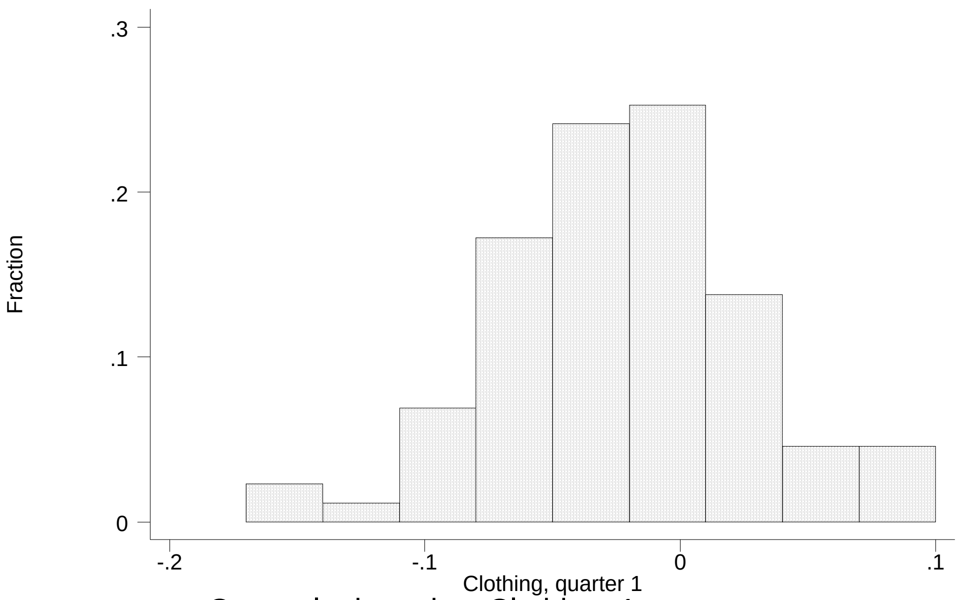


Figure 21: Histogram of the estimated individual-specific coefficient on the quarterly dummy for quarter 3 $(\widehat{\delta}_{11h})_{h=1,\ldots,87}$ for the commodity Food, Alcohol&Tobacco.



Quarterly dummies, Clothing, 1st quarter

Figure 22: Histogram of the estimated individual-specific coefficient on the quarterly dummy for quarter 1 $(\widehat{\delta}_{11h})_{h=1,\ldots,87}$ for the commodity Clothing.

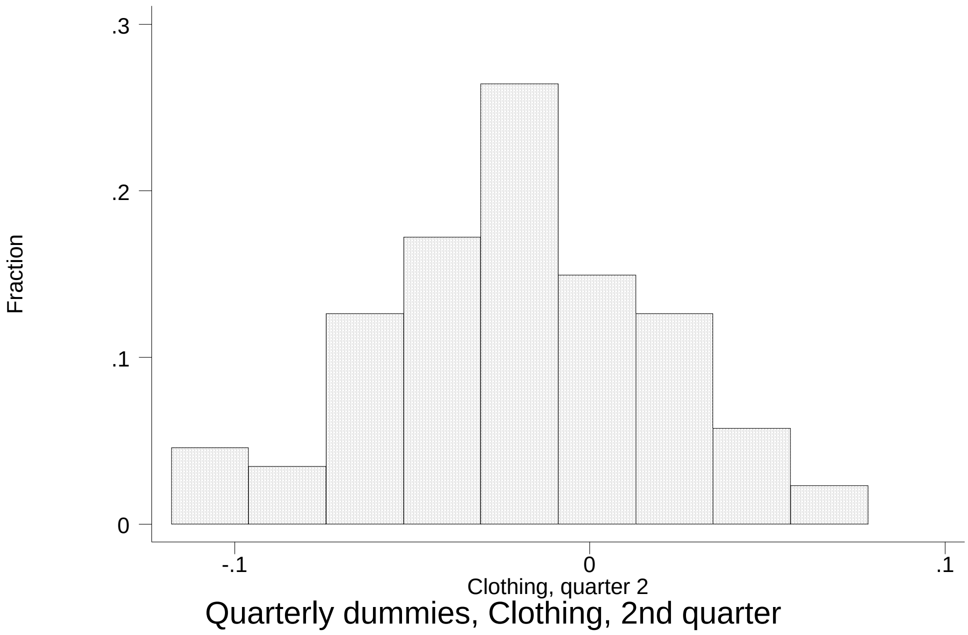


Figure 23: Histogram of the estimated individual-specific coefficient on the quarterly dummy for quarter 2 $(\widehat{\delta}_{11h})_{h=1,\ldots,87}$ for the commodity Clothing.

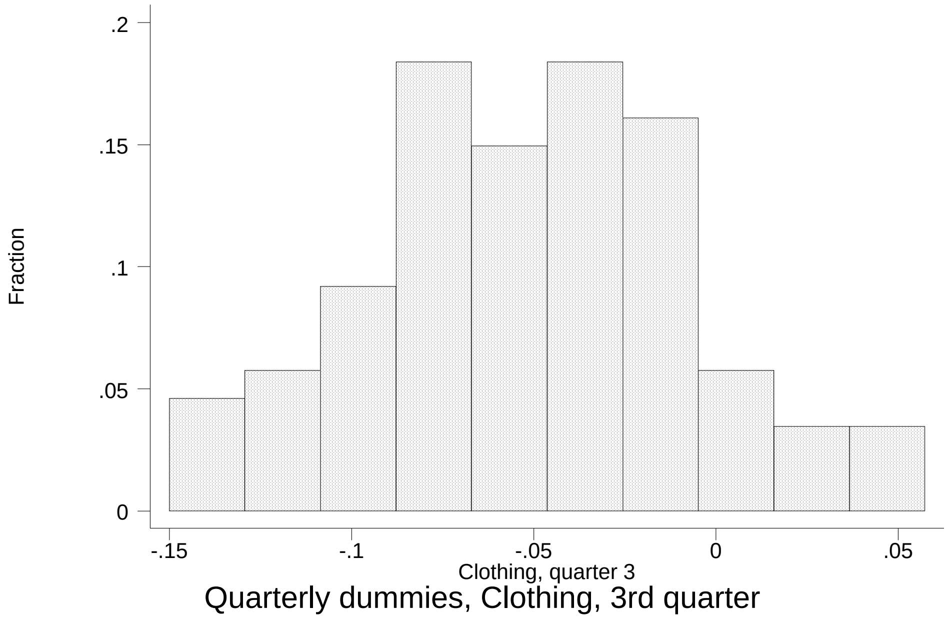


Figure 24: Histogram of the estimated individual-specific coefficient on the quarterly dummy for quarter 3 $(\widehat{\delta}_{11h})_{h=1,\ldots,87}$ for the commodity Clothing.

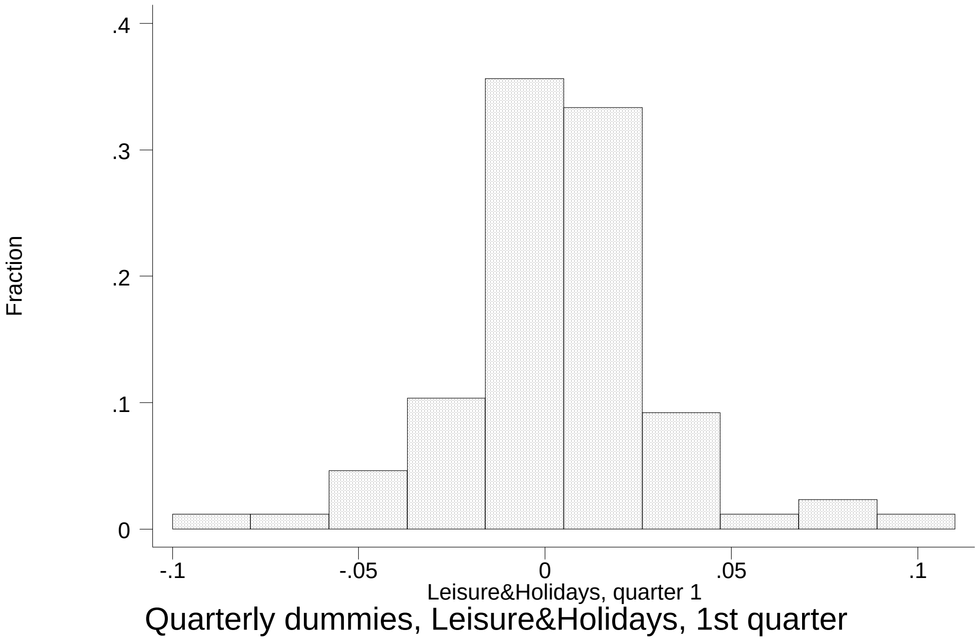


Figure 25: Histogram of the estimated individual-specific coefficient on the quarterly dummy for quarter 1 $(\widehat{\delta}_{11h})_{h=1,\ldots,87}$ for the commodity Leisure&Holidays.

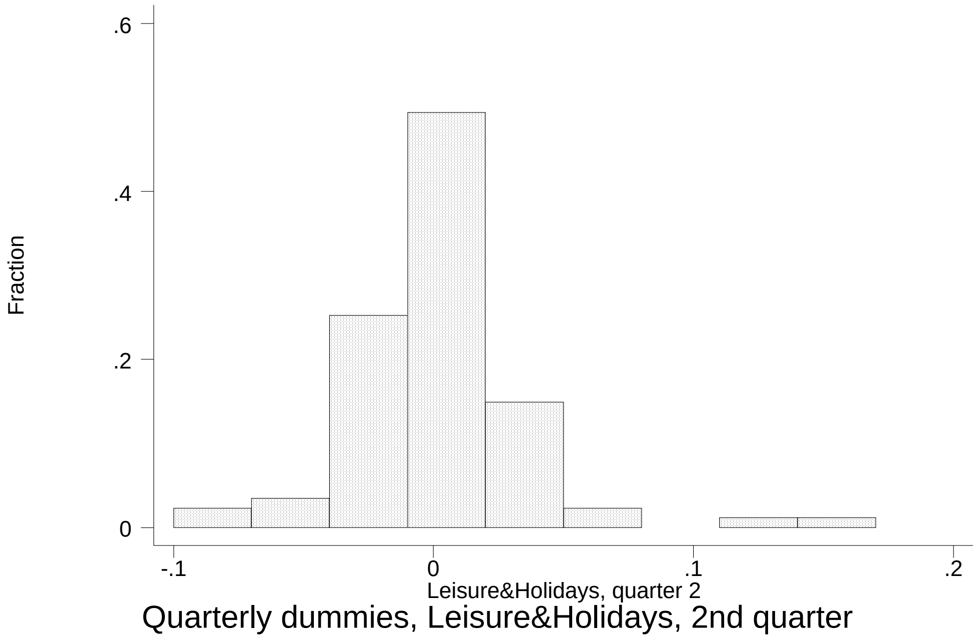


Figure 26: Histogram of the estimated individual-specific coefficient on the quarterly dummy for quarter 2 $(\widehat{\delta}_{11h})_{h=1,\ldots,87}$ for the commodity Leisure&Holidays.

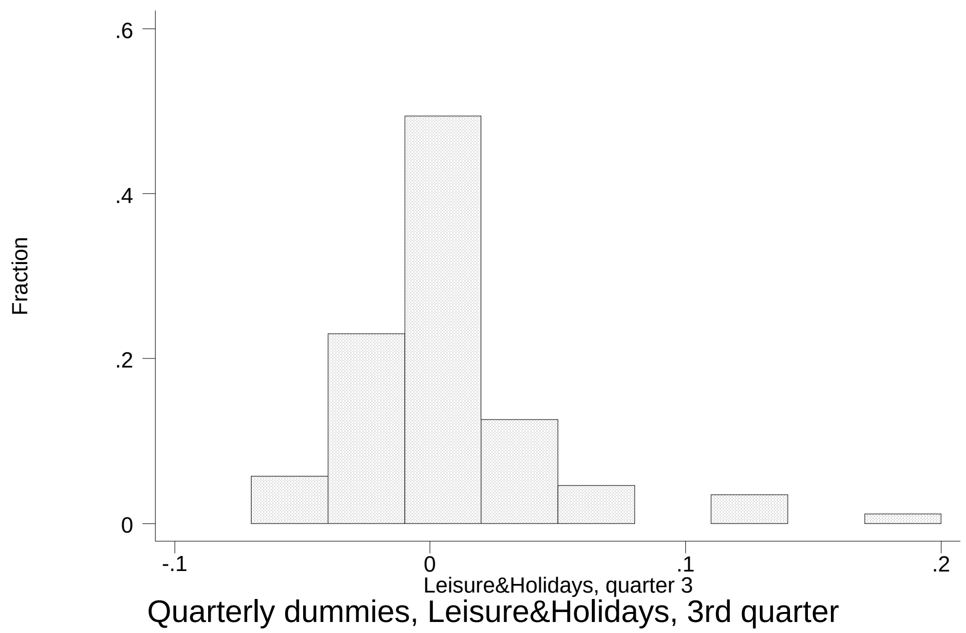


Figure 27: Histogram of the estimated individual-specific coefficient on the quarterly dummy for quarter 3 $(\widehat{\delta}_{11h})_{h=1,\ldots,87}$ for the commodity Leisure&Holidays.