A Demand System with Social Interactions: Evidence from CEX

Giacomo Pasini*

School for Advanced Studies in Venice
and
University of Venice

This version: May 12, 2006

Abstract

A Quadratic Almost Ideal Demand System that allows for social interactions is described and then estimated on CEX data. Social interactions are introduced as mean budget shares and depend on peer membership and visibility. Peer identification is obtained by means of a similarity index which measures the probability of group membership. Reflection problem is tackled directly and therefore estimation is carried on with a Generalized Spatial 2SLS that deal with two types of endogeneity: the first is due to contemporaneous choices of households, the second is due to contemporaneous choice of goods. The results support the hypothesis that total expenditure allocation to budget shares depends both on social interaction and visibility.

JEL Classification: H31, C31, C49

Keywords: Social Interactions, Demand System, Similarity, Endogeneity, Spatial Econometrics

1 Introduction

People are social animals: they do not live in isolation, almost any economically relevant action and choice is taken in a particular social environment, and behavior of others are likely to influence individual activities. Even if this can be considered a common sense statement, traditional economic models of individual behavior assume that agents choose in perfect isolation and

*Address: Dipartimento di Scienze Economiche, Cannaregio 873 – 30121 Venice, Italy. Telephone: +39 041 2349231; e–mail giacomo.pasini@unive.it; personal homepage: http://venus.unive.it/pasinigi

preferences are not directly influenced by the behavior of others. Nevertheless the idea that peer effects do matter attracted a number of economists in different fields, that tried to include social interactions in models of educational attainment, job search, crime and deviating behavior, early pregnancy and many others[1]. Unfortunately, most of the empirical evidence is drawn from specific datasets or natural experiments, therefore limiting the validity of the results to particular sub–populations.

Interdependent preferences were considered also in consumption literature: if Mr Smith buys a brand new car to keep up with Mr Jones, this means that Mr Smith preferences are influenced by Mr. Jones' one. The question is whether social interactions matter in consumption choices: is it reasonable to think that at least for some goods consumption choices of friends, colleagues or in general peers have a role in individual choices? This paper aims to shed some light on this issue.

This study is mainly empirical: although a complete characterization of preferences is not provided, social interactions will be explicitly allowed for and introduced as a conditioning factor in a demand system. The objective is to assess their relevance using a US–wide survey as the Consumers' Expenditures Survey, CEX. The results suggest that Social Interactions do matter.

The introduction of peer effects in an empirical consumption model rises two econometric issues: the definition of the relevant reference group for each individual, and a particular kind of endogeneity, called reflection problem by Manski [16]. The estimation strategy proposed in this paper tackles both of them directly. The idea is to use a measure of similarity to identify peer membership and on this basis re–define the demand system as a Spatial Autoregressive Model (SAR).

Section 2 describes the Economic Model - the Quadratic Almost Ideal Demand System (QUAIDS) proposed by Banks, Blundell and Lewbell [2] - the separability assumptions needed to restrict the attention to demand systems, the inclusion of conditioning factors and how social interactions are modelled. In section 3 the dataset is described, the following one is devoted to the estimation strategy and results. Section 5 concludes.

2 The Economic Model

The framework on which consumption behavior is modelled is the Life Cycle Hypothesis of Modigliani. The model describes consumers' choices as the maximization of an expected intertemporal utility function under an appropriate budget constraint. The utility function depends on consumption of durables and non–durables in each period and hours of work on each period.

[1] A useful review is Brock, Durlauf [5]. Young and Durlauf [11] tried to put the recent literature within a common framework.

In order to reduce this general problem to a treatable one, an intertemporal separability assumption is needed.

To be specific, it is assumed that the objective function is intertepo-rally additive in consumption of non–durable goods. It is well known that this assumption implies two–stage budgeting: in the first stage households equates the discounted marginal utility of each period and determines total non–durables expenditures, hours of work and durables' consumption of each period. In the second stage consumers allocate total expenditures to each non–durable good conditional on leisure and durables choice of the first stage. This allocation process can be described by means of a demand system.

The second key assumption is that social interaction matters only at the second stage. As to say, saving decisions are not affected by others' behavior, therefore peer group effect on consumption is conditional on total expenditure and enter in the demand system, yet not in the Euler equation describing the first–stage.

While intertemporal separability is a standard assumption even if it's not innocuous (see as an example Browning, Meghir [6] for a discussion on labor supply and non–durables consumption separability), the second one is not and it's crucial in this paper. Binder and Pesaran [3] propose a theoretic life–cycle model where social interactions' impact on optimal consumption depend on intertemporal considerations. However, they do not rule out the possibility that social interactions matter also in total expenditure allocation, and even if they infer that intertemporal considerations should be more relevant then static ones, their paper is purely theoretic, so still there is no empirical evidence on the relative importance of peer effects on savings and consumption allocation. Further on, the second assumption can be substituted by the following: social interactions effects on savings and on consumption are separable. In this way social interactions in first stage are not ruled out. The key point is that whatever the assumption it is meaningful to concentrate the attention on the demand system.

2.1 The Demand System: QUAIDS

The starting point is the Quadratic Almost Ideal Demand System (QUAIDS) of Banks, Blundell and Lewbell [2]. This is a quadratic extension of the well–known Almost Ideal Demand System (AIDS) of Deaton and Muellbauer [10], shares all its features plus it allows for heterogeneous Engel curves. QUAIDS can be seen as a quadratic local approximation of almost any demand system that is exactly aggregable, meaning that it's linear in (functions of) total expenditure. Define

I number of consumption goods;

H number of consumers;

m total expenditure;

w_i expenditure share on good i;

p_i price of good i and $\boldsymbol{p}$ prices' vector;

The budget share for good i by household h is

$$w_i^h = \alpha_i + \sum_{j=1}^{I} \gamma_{ij} \ln p_j + \beta_i \ln \left[\frac{m^h}{a(\boldsymbol{p})}\right] + \frac{\lambda_i}{b(\boldsymbol{p})} \left(\ln \left[\frac{m^h}{a(\boldsymbol{p})}\right]\right)^2 \tag{1}$$

where

$$\begin{aligned} \ln a(\boldsymbol{p}) &= \alpha_0 + \textstyle\sum_{i=1}^{I} \alpha_i \ln p_i + \frac{1}{2} \sum_{i=1}^{I} \sum_{j=1}^{I} \gamma_{ij} \ln p_i \ln p_j \\ b(\boldsymbol{p}) &= \textstyle\prod_{i=1}^{I} p_i^{\beta_i} \end{aligned}$$

$a(\boldsymbol{p})$ and $b(\boldsymbol{p})$ are price aggregators: the former takes a translog form, the latter a Cobb–Douglas. It's relevant for estimation purposes to discuss properties and possible restrictions on these price aggregators: conditional on $a(\boldsymbol{p})$ and $b(\boldsymbol{p})$ demands are linear in prices and quadratic in total expenditure. Restrictions on $b(\boldsymbol{p})$ have to do with the rank of the demand system, which Lewbell [15] defines as the dimension of the space spanned by its Engel curves. Therefore, (1) has a rank lower or equal to 3. Banks, Blundell and Lewbell [2] prove that in any rank 3 exactly aggregable demand system the squared term's coefficient must be price dependent, i.e. $b(\boldsymbol{p})$ cannot be constant. The authors refer to Gorman (1981) where it is proved that the maximum possible rank for any exactly aggregable demand system is 3. Therefore, there's no gain adding cubic and higher terms to the demand equations. They also show that empirical Engel curves estimated on British data indicates that the demand system has rank 3. Note that (1) nests QUAIDS with constant $b(\boldsymbol{p})$, which is simpler to estimate at the price of restricting Engel curves' shape. This latter model itself nests AIDS. Blundell et al. [4] obtain a good fit with a QUAIDS where $b(\boldsymbol{p})$ is set to 1 and therefore rank is 2. In this paper the choice is to write a general rank 3 QUAIDS with social interactions, but then carry out the estimation setting $b(\boldsymbol{p}) = 1$[2].

2.2 Properties of Demand Systems

In order to be a demand system, (1) must respect adding up, zero–homogeneity in $\boldsymbol{p}$ and m simultaneously, symmetry and negative semi–definiteness of the Slutsky matrix of compensated price elasticities. All of them but for Slutsky

[2]Estimation has been carried on also restricting to AIDS. Results (which are not reported) suggest that as long as the interest is in social interactions' effect, conclusions are qualitatively similar

matrix negative semi–definitness (which therefore has to be checked ex–post) can be modelled in terms of linear restrictions on the parameters:

$$\sum_{i=1}^{I} \alpha_i = 1; \quad \sum_{i=1}^{I} \gamma_{ij} = 0; \quad \sum_{i=1}^{I} \beta_i = 0; \quad \sum_{i=1}^{I} \lambda_i = 0 \tag{2}$$

$$\sum_{j=1}^{I} \gamma_{ij} = 0; \tag{3}$$

$$\gamma_{ij} = \gamma_{ji} \ \forall i, j \tag{4}$$

(2) implies adding up; (2) and (3) together imply zero–homogeneity. Conditions (2) and (4) together imply Slutsky symmetry. Among them, if price aggregators were known only (4) would set cross–equations restrictions. This observation will be useful for estimation: conditioning on preliminary estimates of $a(\boldsymbol{p})$ and setting $b(\boldsymbol{p}) = 1$ it's possible to impose adding up and homogeneity (i.e. restriction (2) and (3)) and estimate the system equation by equation.

2.3 Demographics

With household data consumer preferences must be allowed to depend on individual characteristics, i.e. demographics $\boldsymbol{z}$[3] must enter (1) as conditioning factors. Therefore the coefficients α_i, β_i, λ_i can be thought as household–h specific: they are re–written as polynomials in $\boldsymbol{z}$ to make demographics' effect explicit. Note also that $\boldsymbol{z}$ include deterministic time–dependent variables (seasonal/year dummies). Then, $\forall i \neq 0$:

$$\alpha_i^h = \alpha_{i0} + \sum_{k=1}^{K} \alpha_{ik} z_k^h \tag{5}$$

$$\beta_i^h = \beta_{i0} + \sum_{k=1}^{K} \beta_{ik} z_k^h \tag{6}$$

$$\lambda_i^h = \lambda_{i0} + \sum_{k=1}^{K} \lambda_{ik} z_k^h \tag{7}$$

This is the most general formulation including demographics. The three polynomials need not to depend on all the K elements of $\boldsymbol{z}$: it is enough to set a–priori (or test ex–post) the relevant parameters equal to zero. Substituting them in (1):

[3] $\boldsymbol{z}$ is a K dimensional vector, where K is the number of observable individual characteristics

$$
\begin{aligned}
w_i^h = \alpha_{i0} &+ \sum_{k=1}^{K} \alpha_{ik} z_k^h \\
&+ \sum_{j=1}^{I} \gamma_{ij} \ln p_j \\
&+ \beta_{i0} \ln \left[\frac{m^h}{a(\boldsymbol{p},\boldsymbol{z})}\right] + \sum_{k=1}^{K} \beta_{ik} \left(z_k^h \ln \left[\frac{m^h}{a(\boldsymbol{p},\boldsymbol{z})}\right]\right) \\
&+ \frac{\lambda_{i0}}{b(\boldsymbol{p},\boldsymbol{z})} \left(\ln \left[\frac{m^h}{a(\boldsymbol{p},\boldsymbol{z})}\right]\right)^2 + \sum_{k=1}^{K} \frac{\lambda_{ik}}{b(\boldsymbol{p},\boldsymbol{z})} \left(z_k^h \left(\ln \left[\frac{m^h}{a(\boldsymbol{p},\boldsymbol{z})}\right]\right)^2\right)
\end{aligned}
\tag{8}
$$

where also the price aggregators are household–dependent. Restrictions (2) must be rewritten in terms of the new parameters:

$$
\begin{array}{l}
\sum_{i=1}^{I} \alpha_{i0} = 1; \\
\sum_{i=1}^{I} \alpha_{ik} = 0 \quad \forall k = 1, \dots K; \\
\sum_{i=1}^{I} \gamma_{ij} = 0; \\
\sum_{i=1}^{I} \beta_{ik} = 0 \quad \forall k = 0, \dots K; \\
\sum_{i=1}^{I} \lambda_{ik} = 0 \quad \forall k = 0, \dots K
\end{array}
\tag{9}
$$

2.4 Social Interactions

Social Interactions' effect can be defined as follows: "the propensity of an individual to behave in some way varies with the prevalence of that behavior in some reference group containing the individual" (Manski [16]). This definition is as broad as possible and in a demand analysis framework it has been previously called preference interdependence (Alessie, Kapteyn [1]), meaning that consumer's preferences are influenced by the behavior of others.

Manski makes three hypotheses to explain this empirical observation:

1. Endogenous effects: the propensity of an individual to behave in some way is affected by the behavior of the group. That is, demand of good i of consumer h changes with the average demand of good i by other people in his reference group;

2. Contextual effect: the propensity of an individual to behave in some way is affected by the exogenous characteristics of the group. That is, demand for good i by household h depends on the average total expenditure or on the average characteristics in $\boldsymbol{z}$ of individuals in the reference group.

3. Correlated effects: individuals in the same group tend to behave similarly because they have similar (unobserved) individual characteristics.

Endogenous and contextual effect are then "economically meaningful" social interactions' effects, while correlated effect reflects an omitted variable problem, and therefore it is not a social effect of the variety we want to identify.

Manski sets up a general linear–in–means model where the output y depend linearly on the averages on the reference group of the output itself, of the independent variables and of the unobserved attributes. The presence of the average output variable on the right–hand–side of the regression equation rises what the author calls the "reflection problem", which does not allow to separately identify endogenous and contextual effects. Nevertheless, in the reduced form of the model it is possible to identify a composite parameter capturing truly social interactions' effects separately from correlated effects.

The aim of this paper is to detect whether or not there is any significant effect of social interactions on demand. To keep things as easy and tractable as possible, the assumption is that there are no contextual effects. In other words the effect of the peers is fully captured by the average demand in the reference group. This hypothesis is somewhat unavoidable: the demand system is linear–in–means, therefore without assuming out contextual effect it's possible to estimate just the reduced form in which social effects are captured by one social effects' composite parameter.

Now define the "mean budget share" of good i for household h as

$$\tilde{w}_i^h := \sum_{n=1}^{N} \delta_{in}^h w_i^n \tag{10}$$

$\tilde{w}_i^h$ is a weighted average of individual demands for good i, w_i^n. The reference weights δ_{in}^h capture the importance household h attaches to consumption of good i by family n. Assume without loss of generality that $\delta_{ih}^h = 0$.[4]

Alessie and Kapteyn [1] defined (10) as "mean perceived budget share". In their model the reference weights are individual parameters, as to say that heterogeneity in preference interdependence among agents depend on differences in the perception of other households' demand. In this terms, it can be interpreted as a framing problem: unobserved individual characteristics determining reference weights lead households to "measure" differently.

In this paper the assumption is that consumers observe correctly other households' expenditures, and the reference weights are determined by the "similarity" between agents and the "visibility" of good i:

$$\delta_{in}^h = \theta_i \pi_n^h \tag{11}$$

Where θ_i measures "visibility" of commodity i and $\Pi = \left[\pi_n^h\right]$ is the $H \times H$ matrix whose elements represent pair–wise similarities between households.

[4] It's just a rescaling: if $\delta_{ih}^h \neq 0$ the system can be written in terms of $\ddot{w}_i^h = (1 - \delta_{ih}^h) w_i^h$.

In this context similarity has no direction, i.e. $\pi_n^h \equiv \pi_h^n$, therefore Π is symmetric and with zeros on the diagonal.

The motivation behind similarities is peer identification: the behavior of consumer n can have an impact on consumer h's choices only if they belong to the same peer. A microeconomic data–set with both direct information about reference groups and the required detail about expenditure patterns would provide a measure of peer membership, but unfortunately such data are not available. Without direct observation, the best the researcher can do is to infer the probability that two individuals belong to the same reference group from available information as physical residence, family characteristics, race, education and so on. The underling hypothesis is that similarity is a valid measure of reference group membership, and therefore δ_{in}^h will be high if households h and n are likely to be in the same peer, vice versa it will be low. Case [8] sets up a model where mean demand depends on physical proximity: individuals belong to the same peer if they live in the same neighborhood. Conley [9] provides tools to estimate models with generic economic distances, possibly measured with error.

The second factor determining reference weights is visibility: it's reasonable to think that consumers care more about peer members' expenditure in clothing rather than in toothpaste, i.e. social interactions effect matter more for visible goods' demand rather than for non–visible ones. There are two possible motivations: first, individuals may not be able to observe peer members' consumption of non–visible goods as groceries or underwear. Second, visibility may be a valuable characteristic of goods itself. Heffetz [13] characterizes a class of utility functions that depend on conspicuousness of goods: the idea is that consumption has a direct effect on individual utility, but also an indirect *social* effect resulting from peers observing his choice.

Now plugging (11) into (10)

$$\tilde{w}_i^h = \theta_i \bar{w}_i^h \quad \text{where } \bar{w}_i^h = \sum_n^N \pi_n^h w_i^n$$

It is possible to add social interactions in (8) as a conditioning factor defining each α_{i0} as a polynomial in $\tilde{w}_i^h$:

$$\alpha_{i0} = \tilde{\alpha}_{i0} + \sum_{j=1}^{I} (\tilde{\alpha}_{ij}\theta_i)\bar{w}_j^h \tag{12}$$

Note it is implicitly assumed that social interactions change intercepts but not slopes. Restrictions (9) has to be modified as well:

$$
\begin{array}{l}
\sum_{i=1}^{I} \tilde{\alpha}_{i0} = 1; \\
\sum_{i=1}^{I} \tilde{\alpha}_{ij} = 0 \quad \forall j = 1, \dots I; \\
\sum_{i=1}^{I} \alpha_{ik} = 0 \quad \forall k = 1, \dots K; \\
\sum_{i=1}^{I} \gamma_{ij} = 0; \\
\sum_{i=1}^{I} \beta_{ik} = 0 \quad \forall k = 0, \dots K; \\
\sum_{i=1}^{I} \lambda_{ik} = 0 \quad \forall k = 0, \dots K
\end{array} \tag{13}
$$

At this point to obtain the complete demand system unobservables u_i^h are needed. Estimation will be done in a GMM framework, so no particular distributional assumption across goods will be done. Nevertheless unobservable factors may have the same structural dependence as demands (correlated effect), therefore the h dimension of the error term will be modelled as follows:

$$
u_i^h = \rho \sum_{n=1}^{N} \pi_n^h u_i^n + \epsilon^h \tag{14}
$$

All the I equations constituting the demand system to be estimated are then obtained adding (14) and substituting (12) into (10):

$$
\begin{aligned}
w_i^h = {} & \tilde{\alpha}_{i0} + \phi_{i1}\bar{w}_1^h + \cdots + \phi_{iI}\bar{w}_I^h \\
& + \sum_{k=1}^{K} \alpha_{ik} z_k^h \quad + \quad \sum_{j=1}^{I} \gamma_{ij} \ln p_j \\
& + \beta_{i0} \ln \left[\frac{m^h}{a^h(\boldsymbol{p}, \boldsymbol{z}, \bar{\boldsymbol{w}})} \right] + \sum_{k=1}^{K} \beta_{ik} z_k^h \ln \left[\frac{m^h}{a^h(\boldsymbol{p}, \boldsymbol{z}, \bar{\boldsymbol{w}})} \right] \\
& + \frac{\lambda_{i0}}{b^h(\boldsymbol{p}, \boldsymbol{z})} \left(\ln \left[\frac{m^h}{a^h(\boldsymbol{p}, \boldsymbol{z}, \bar{\boldsymbol{w}})} \right] \right)^2 \\
& + \sum_{k=1}^{K} \frac{\lambda_{ik}}{b^h(\boldsymbol{p}, \boldsymbol{z})} z_k^h \left(\ln \left[\frac{m^h}{a^h(\boldsymbol{p}, \boldsymbol{z}, \bar{\boldsymbol{w}})} \right] \right)^2 \\
& + u_i^h
\end{aligned} \tag{15}
$$

where $\phi_{ij} = \tilde{\alpha}_{ij}\theta_i$. θ_i are not separately identifiable from $\tilde{\alpha}_{i1}$ for all i. This lack of identifiability will complicate interpretation: pure social interaction effect, captured by $\tilde{\alpha}_{ij}$ may well have a different sign and different magnitude from visibility effect, θ_i.

The price aggregators depend now on all the conditioning factors:

$$\ln a^h(\boldsymbol{p}, \boldsymbol{z}, \bar{\boldsymbol{w}}) = \alpha_0 + \sum_{i=1}^{I} \ln p_i \left(\tilde{\alpha}_{i0} + \sum_{k}^{K} \alpha_{ik} z_k^h + \sum_{j}^{I} \phi_{ij} \bar{w}_j^h \right) + \sum_{i=1}^{I} \sum_{j=1}^{I} \gamma_{ij} \ln p_i \ln p_j \tag{16}$$

$$b^h(\boldsymbol{p}) = \prod_{i=1}^{I} p_i^{\beta_{i0} + \sum_{k=1}^{K} \beta_{ik} z_k^h} \tag{17}$$

3 The data: Consumer Expenditure Survey (CEX) and Consumer's Price Index (CPI)

CEX is a detailed survey on individual expenditures. There are quarterly data from 1980 until 2002 on approximately 600 consumption categories. This survey is issued by the Bureau of Labor Statistics, that is the Office which publishes the CPI price indexes. The long and detailed repeated cross–sections dataset under analysis is obtained merging together CPI prices and CEX expenditures. CEX provides also a large number of demographic details about individuals, but as pointed out in the previous section there are no direct questions about reference groups. The claim is that the information is adequate to compute similarities among individuals.

In particular, 10 years of data are considered - from 1993 until 2002 - since in this period the state of residence identifier is available. For non–disclosure problems the variable STATE is suppressed for some observations in a subset of states and it is suppressed for all the observations on some other states. All the observations from those states are dropped, so we are left with observations from Arizona, Arkansas, California, Colorado, Connecticut, District of Columbia, Florida, Hawaii, Illinois, Missouri, New Hampshire, New Jersey, Pennsylvania, South Carolina, Utah, Virginia and Washington. The heterogeneous distribution of those states across US still allows to draw population–wide inference (see figure 1).

Data are summed up at yearly level, and only households with four consecutive quarterly observations are considered. At the end the sample consists of 14,272 observations. In the appendix means and standard deviations are reported for a set of relevant demographics on the selected subsample and on the US–wide sample. Differences suggest the sample is representative for the US population.

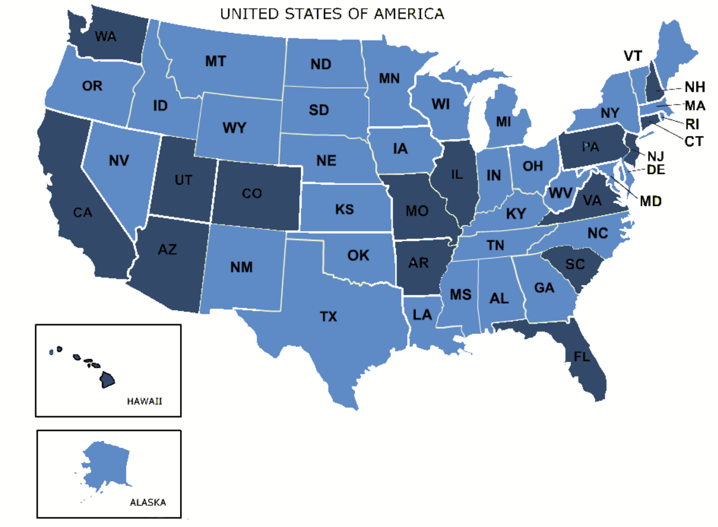

Figure 1: selected States are dark–blue

4 Estimation Strategy

The estimation strategy is based on the one that Banks and Blundell [2] and Blundell et al. [4] used. However, an extension is needed in order to deal with the reflection problem. The estimation is divided into three steps:

1. Π Matrix estimation: similarities are measured on the basis of a set of geographical and demographic individual characteristics.

2. Equation–by–equation estimates: parameters on each equation are estimated after imposing adding–up and homogeneity restrictions (13) and (3). Using the Generalized Spatial 2SLS (GS2SLS) procedure of Kelejian and Prucha [14] the reflection problem is taken into proper account. GS2SLS estimator is a GMM spatial estimator within the class defined by Conley [9]. The author proves that as long as estimates in step–1 are imprecise measurements of true group membership probabilities, but they are not mis–measurement, step–2 estimates are consistent[5].

3. Restricted system estimation: a Minimum Distance estimator is applied to step–2 estimates of parameters to impose cross–equation restrictions (4).

[5] An imprecise measure is a measure that is correct up to a certain level, as home–work place traveling distances up to city detail but not beyond. A mis–measurement is a truly incorrect distance, as a transformation applied to true distances

4.1 Similarity Matrix estimation

The claim is that two individuals are likely to belong to the same peer and therefore possibly influence each others' choices if they live close, they are observed in two not–too–distant points in time and they share some household's characteristics. Further on, a short physical distance is considered a prerequisite for peer membership.

Given these assumptions similarity between agents h and n, d_n^h, follows a lexicographic order and it is computed as follows:

1. Two individuals are assumed not to belong to the same peer if they live in different States, or in the same State but in two cities with different population size, or if one is observed before 1997 and the other after that date. Therefore, pairs of individuals h, n with those characteristics have similarity d_n^h equal to 0.

2. Otherwise, if h and n live in the same State in two cities with the same population–size and they are both observed either before 1997 or after that date, d_n^h is equal to a matching similarity measure constructed as follows:

   - A set of 0/1 dummy variables is created starting from the following variables: Family composition, 5 years–wide age class of household head, race, marital status, origin (ancestry) of household head, highest educational attainment, presence of children younger than 18 in the family, gender.
   - the index is equal to

$$d_n^h = \frac{\sum 1\text{–}1 \text{ matches}}{\# \text{ of } 0/1 \text{ dummies}}$$

Finally this similarity measure has no direction by construction therefore $d_n^h \equiv d_h^n$ and as previously explained it is re–parametrized in order to have $d_n^n = 0$ (zeros on the diagonal).

This procedure provide an estimate of similarities that is by construction imprecise: the physical distance information are quite poor if compared with other datasets used in social–interactions empirical literature (eg Topa [17]). The matching similarity identifies individuals living in two equally big cities (possibly the same city) in the same State. Note also that matching similarities are considered as exogenous and given in the successive steps of the procedure.

In order to check that these similarities didn't simply capture State, population size and year effects, an OLS regression of π_n^h on the full set of year, state and population dummies, plus their interactions is run. Results [6] shows

[6] which are not reported but are available upon request

that interactions' parameters are significantly different from zero, suggesting that similarities are more informative than a simple set of dummies.

4.2 Equation–by–Equation estimation

The demand system is non–linear, but each equation in (15) is linear conditional on $a(\boldsymbol{p},\boldsymbol{z})$ and $b(\boldsymbol{p},\boldsymbol{z})$. The second step uses this conditional linearity to estimate the model without imposing the cross–equation restrictions (4) but allowing for within–equation ones (13) and (3). $a(\boldsymbol{p},\boldsymbol{z})$ is approximated with an household–level Stone price Index. $b(\boldsymbol{p},\boldsymbol{z})$ is set equal to 1. As already explained this choice reduces the rank of the demand system to 2 according to Lewbell's definition.

Two endogeneity issues have to be addressed: first, total expenditure $\ln m^h$ and $(\ln m^h)^2$ are endogenous along the i dimension, i.e. they are endogenous due to the contemporaneous allocation of total expenditure to different goods by each household. Second, in each equation describing the budget share of good i, mean budget share $\bar{w}_i^h$ is endogenous along the h dimension, meaning it's endogenous due to the contemporaneous choice of the H households of each good. These issues can be solved using a proper Instrumental Variables' procedure: endogeneity of total expenditure can be treated with standard 2SLS, the Generalized Spatial 2SLS (GS2SLS) proposed by Kelejian and Prucha [14] is needed to account for endogeneity of mean budget shares. The resulting procedure requires that $\ln m^h$ and $(\ln m^h)^2$ are regressed on the exogenous variables and their predicted values are used as instruments. Then GS2SLS is applied instead of the standard second step to account for endogeneity of $\bar{w}_i^h$. GS2SLS is itself an iterative procedure. To see the basic steps and to underline the fact that endogeneity is along the h dimension, rewrite demand for good i (15) in matrix notation:

$$\begin{aligned} \boldsymbol{w}_i^h &= X^h\boldsymbol{\beta} + \phi_i \Pi \boldsymbol{w}_i^h + \boldsymbol{u}_i^h \\ \boldsymbol{u}_i^h &= \rho \Pi \boldsymbol{u}^h + \boldsymbol{\epsilon}_i^h \end{aligned} \tag{18}$$

This is written as a spatial autoregressive model, where $\boldsymbol{w}^h$ is the $H \times 1$ vector of observation on expenditure share on good i; X^h is the $H \times K^*$ matrix that contains observations on the exogenous variables in Z^h, the predicted values of total expenditure and squared total expenditure, prices, $\bar{w}_j^h$, $\forall j \neq i$[7] and iterations among Z^h and predicted values for $\ln m^h$ and $(\ln m^h)^2$. Π is treated as a $H \times H$ matrix of known constants, ρ and ϕ_i are scalar spatial autoregressive parameters.

Now rewriting model (18) as[8]

[7] All the mean budget shares $\bar{w}_j^h$ $\forall j \neq i$ are considered as exogenous in ith budget share equation. Therefore the set of variables in X^h changes for each equation. The overall set of regressors doesn't change preserving adding up, since in the ith equation $\bar{w}_i^h$ is instrumented.

[8] indexes h are omitted

$$
\begin{aligned}
\boldsymbol{w}_i &= D\boldsymbol{\eta} + \boldsymbol{u}_i \\
\boldsymbol{u}_i &= \rho\Pi\boldsymbol{u}_i + \boldsymbol{\epsilon}_i
\end{aligned}
\tag{19}
$$

where $D = (X, \Pi\boldsymbol{w}_i)$, $\boldsymbol{\eta} = (\boldsymbol{\beta}', \phi_i)'$, $\epsilon \sim IID(0, \sigma^2)$. The model can furthermore be transformed into

$$
\boldsymbol{w}_i^*(\rho) = D^*(\rho)\boldsymbol{\eta} + \boldsymbol{\epsilon}_i \tag{20}
$$

where $\boldsymbol{w}_i^*(\rho) = \boldsymbol{w}_i - \phi_i\Pi\boldsymbol{w}_i$, $D^*(\rho) = D - \rho\Pi D$. The estimation procedure is based on three steps:

- compute a 2SLS estimator for $\boldsymbol{\eta}$ in (19), $\hat{\boldsymbol{\eta}}$, using as instruments for $\Pi\boldsymbol{w}_i$ the matrix $(X, \Pi X)$;
- use $\hat{\boldsymbol{\eta}}$ to estimate $\hat{\rho}$ and $\hat{\sigma}^2$ with GMM[9]
- use $\hat{\rho}$ and $\hat{\sigma}^2$ to compute $\boldsymbol{\eta}_{KP}$, a feasible 2SLS of $\boldsymbol{\eta}$ in (20) and its variance–covariance matrix $\hat{V}(\boldsymbol{\eta}_{KP})$

As already noted Conley [9] proves that if Π is an imprecise but non mis–measured matrix of similarities GS2SLS lead to consistent estimates. Therefore, using it as a second step in the overall procedure both endogeneities are taken into account and $\boldsymbol{\eta}_{KP}$ is consistent.

The system is estimated for 8 consumption categories: Alcohol at home (ALH), Alcohol out (ALO), Food at Home (FDH), Food out (FDO), Clothing excluding underwear (CLO), Underwear (UND), Motor Fuel (GAS), other non durables (OTH). Some of those consumption categories have a relevant presence of zero expenditures among the 14,242 observations:

	zero occurrences	
	freq.	perc.
ALH	6,497	45.52
ALO	6,505	45.58
FDH	6	0.04
FDO	740	5.18
CLO	1,097	7.69
UND	2,798	19.60
GAS	964	6.75
OTH	2	0.01

Given the type of aggregates chosen, these zero occurrences are likely to correspond to purchase infrequency[10]. As pointed out by Blundell et al.

[9] details on moment conditions are in Kelejian and Prucha [14]

[10] There may be undetected data quality problems: the under garments figure seems unreasonable given that data are year–level aggregates.

[4] it means that there is a conceptual difference between consumption and expenditure: the latter is not simply the empirically measured counterpart of the former. This difference affects both the dependent variables in the demand system and total consumption, arising a potential measurement error problem due to omitted variables. Nevertheless the estimation procedure removes the issue: budget shares are all treated as endogenous and therefore total expenditure is instrumented.

But for gasoline and other goods, the other consumption aggregates are chosen to check whether social interactions have different marginal effects on goods with a different visibility. Alcohol demand is maintained despite the particularly high zero occurrences because of its relevance from a tax policy point of view. OTH is omitted from the estimation to satisfy adding–up. Prices are monthly US–wide price indexes series for each category (OTH price is the overall price index) referring to the last month of each yearly observation. Base year is 2000. All indexes are then divided by OTH price to impose homogeneity. Because of two–stage budgeting hypothesis occupation is not instrumented: job–market participation is considered non–separable from overall consumption in the first stage, but when households have to decide about consumption allocation the job–market decision is already taken, and therefore it's predetermined with respect to budget shares' allocation. The same reasoning goes through for durables. The next table reports estimates for own mean budget shares parameters for the first six consumption categories:

Visible goods				Non visible goods		
FDO	ϕ_{FDO}	0.1657		FDH	ϕ_{FDH}	0.0819
FDO	std.err	0.019		FDH	std.err	0.060
FDO	t–stat	8.71		FDH	t–stat	1.37
ALO	ϕ_{ALO}	-0.0244		ALH	ϕ_{ALH}	-0.0050
ALO	std.err	0.008		ALH	std.err	0.008
ALO	t–stat	-2.90		ALH	t–stat	-0.61
CLO	ϕ_{CLO}	-0.1068		UND	ϕ_{UND}	-0.4177
CLO	std.err	0.027		UND	std.err	0.032
CLO	t–stat	-3.98		UND	t–stat	-13.13

Estimated parameters are generally significantly different from 0, and they varies significantly across different types of goods and between visible and non–visible goods of the same type. Parameters are not marginal effects, since also the price aggregator $a(\boldsymbol{p})$ depend on ϕ. Nevertheless, the correction in marginal effects is small:

Marginal effects

FDO	0.16714		FDH	0.08800
ALO	-0.02668		ALH	-0.00570
CLO	-0.10743		UND	-0.42329

The main result of this paper is the significance of 5 out of 6 parameters reported in the previous tables[11]: social interaction and visibility together do matter in consumption choices. Visibility itself seems to be relevant: estimates are different within pairs ALH/ALO, FDH/FDO, CLO/UND[12]. Food Out has a parameter twice the Food at Home one, intuitively a less visible category. In this case common–sense is supported by previous results by Heffetz [13], who ranked the same aggregates in terms of visibility. Alcohol at home is not significant, while alcohol out becomes significant and negative. The sign could depend on a stigma attached to alcohol consumption due, as an example, to bad health effects: in this case, the social interactions effect in this case is negative. Difference in significance is coherent with the stigma interpretation: a person could be convinced to buy less drinks in public while the less visible expenditure for alcohol at home may well not depend by the negative social interaction effect[13].

By visibility considerations common sense suggest suggest a lower social interactions' parameter for underwear than for clothing. Anyway in this case interpretation of the sign is not straightforward: a reasonable prior seems to be that social interactions have positive effects on apparel expenditure.

The magnitude of ρ's estimates, the spatial autoregressive parameters on unobservables, suggests that the spatial correction on $\boldsymbol{u}$ is meaningful as well:

$\hat{\rho}$, spatial autoregressive parameter

CLO	0.01706		UND	0.02954
FDO	0.02373		FDH	0.02617
ALO	0.01800		ALH	0.01848

It's interesting to see that but for the apparel ones there isn't much difference across goods of the same type in ρ's estimates: this result together with the sign on ϕ parameters on apparel suggest that on those consumption categories there may be some non modeled effect. Complete estimation results can be found in the appendix.

4.3 Minimum Distance estimation

The final step consists in applying a minimum distance estimator to the previously obtained $\boldsymbol{\eta}_{KP}$. The cross–equation restrictions (Slutsky matrix symmetry) can be expressed as

[11] 6 out of 7 considering the gasoline equation

[12] Pairs of consumption categories are similar but for visibility, but it cannot be tested whether differences in ϕ are due only to visibility.

[13] Note that Heffetz [13] ranks ALH as more visible than ALO. Lack of a full preferences' characterization leave space to alternative interpretations of the results.

$$\boldsymbol{\eta} - \boldsymbol{S}\boldsymbol{\theta} = \mathbf{0} \tag{21}$$

Where η is an $r \times 1$ dimensional vector while θ is $q \times 1$, with $r > q$. Symmetry restrictions are all linear. As in GMM estimation, to impose those restrictions OMD chooses $\boldsymbol{\theta}_{OMD}$ as to minimize

$$Q(\boldsymbol{\theta}) = [\boldsymbol{\eta}_{KP} - \boldsymbol{S}\boldsymbol{\theta}]'\hat{V}(\boldsymbol{\eta}_{KP})^{-1}[\boldsymbol{\eta}_{KP} - \boldsymbol{S}\boldsymbol{\theta}] \tag{22}$$

The three steps procedure has an implicit assumption on the parameters' space at the equation–by–equation estimation step: parameters on different equations are assumed to be uncorrelated, therefore $V(\boldsymbol{\eta}_{KP})$ is block–diagonal. Cross–equation restrictions refer only to prices' parameters γ_{ij}, this implies that but for $\hat{\gamma}_{ij}$ equation–by–equation estimates and their standard errors are the final estimates. Therefore, considering only the seven consumption categories (remember OTH is omitted for adding–up), $r = 49$ while $q = 28$, the number of unique elements of a 7×7 symmetric matrix. Further on, γ_{ij} do not depend on $\tilde{w}_i^h$, therefore also the marginal effects on mean budget shares are unchanged after OMD estimation.

The minimized value of the objective function, $Q(\boldsymbol{\theta}_{OMD})$ is asymptotically distributed as a central χ^2 with $r - q$ degrees of freedom. This provides a test for Slutsky symmetry[14]. The test rejects Slutsky symmetry ($Q(\boldsymbol{\theta}_{OMD}) = 40143.76$). Given the linearity of (21) the estimate of Covariance matrix of OMD is:

$$\hat{V}(\boldsymbol{\theta}_{OMD}) = H\left(S'\hat{V}(\boldsymbol{\eta}_{KP})^{-1}S\right)^{-1} \tag{23}$$

Where $H = 14272$ is the sample size. As for the unrestricted estimates, most of $\hat{\theta}_{ij}$ are non–significant. Complete restricted estimates of prices' parameters matrix $\Gamma = [\gamma_{ij}]$ are reported in the appendix.

5 Conclusions

The aim of this work was to assess whether consumption choices depend on social interactions. To do so Social Interactions were introduced in a Quadratic Almost Ideal Demand System as a conditioning factor. The novelty of the paper is in the estimation procedure: social interactions are captured with mean budget shares, that depend on probability of peer membership and visibility of each good. Peer membership identification is a major econometric issue once estimation is not performed with natural experiment or ad–hoc data sets. In this paper it is achieved constructing a similarity index, which measures the probability of belonging to the same peer for each

[14]Proof of asymptotic properties of OMD estimators can be found in Cameron, Trivedi [7] and in Ferguson [12]

couple of observations. This formulation allows to re–write each budget share equation as a Spatial Autoregressive model in order to adapt tools taken from the Spatial Econometrics literature: the endogeneity of mean budget shares that arises from the reflection problem is tackled using a Generalized Spatial 2SLS procedure.

Results support the initial hypothesis that social interactions are relevant in consumption allocation. Further on, they suggest a non–trivial role for visibility of different goods.

Future research should address two open issues which limit interpretation of estimation results: first, in this linear–in–means model pure social interaction and visibility are not separately identifiable. Second, in the literature there isn't a model that provides a structural characterization of preference dependence on social interactions and visibility. Another related field is the empirical investigation of an intertemporal consumption model with social interactions.

References

[1] R. Alessie and A. Kapteyn. Habit formation, interdependent preferences and demographic effects in the almost ideal demand system. *The Economic Journal*, 101(406):404–419, May 1991.

[2] J. Banks, R. Blundell, and A. Lewbel. Quadratic engel curves and consumer demand. *Review of Economics and Statistics*, LXXIX(4):527–539, November 1997.

[3] M. Binder and M. Pesaran. Life–cycle consumption under social interactions. *Journal of Economic Dynamics & control*, 25:35–83, 2001.

[4] R. Blundell, P. Pashardes, and G. Weber. What do we learn about consumer demand patterns from micro data? *The American Economic Review*, 83(3):570–597, June 1993.

[5] W. A. Brock and S. N. Durlauf. Interactions–based models. *NBER Technical Working Paper*, August 2000.

[6] M. Browning and C. Meghir. The effects of male and female labor supply on commodity demands. *Econometrica*, 59(4):925–951, July 1991.

[7] A. Cameron and Pravin K. Trivedi. *Microeconometrics - Methods and Applications*. Cambridge University Press, New York, USA, 2005.

[8] A. C. Case. Spatial patterns in household demand. *Econometrica*, 59(4):953–965, July 1991.

[9] T. G. Conley. Gmm estimation with cross sectional dependence. *Journal of Econometrics*, 92(1):1–45, 1999.

[10] A. Deaton and J. Muellbauer. An almost ideal demand system. *The American Economic Review*, 70(3):312–326, June 1980.

[11] S. N. Durlauf and P. H. Young, editors. *Social Dynamics.* MIT Press, London, 2001.

[12] T. S. Ferguson. A method of generating best asymptotically normal estimates with application to the estimation of bacterial densities. *The Annals of Mathematical Statistics*, 29(4):1046–1062, 1958.

[13] O. Heffetz. Conspicous consumption and the visibility of consumer expenditures. Priceton University, November 2004.

[14] H. H. Kelejian and I. R. Prucha. A generalized spatial two–stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *Journal of Real Estate Finance and Economics*, 17(1):99–121, 1998.

[15] A. Lewbell. The rank of demand system theory and nonparametric estimation. *Econometrica*, 59:711–730, 1991.

[16] C. F. Manski. Identification of endogenous social effects: The reflection problem. *Review of Economic Studies*, 60(3):531–542, July 1993.

[17] G. Topa. Social interactions, local spillovers and unemployment. *Review of Economic Studies*, 68(261-295), 2001.

A Codebook and Descriptive Statistics

Var name	Variables description
ALH	alcoholic beverages for home use
ALO	alcoholic beverages at restaurants, bars, cafeterias, cafes, etc
FDO	dining out at restaurants, drive-thrus, etc, excl. alcohol; incl. food at school
FDH	food and nonalcoholic beverages at grocery, specialty and convenience stores
CLO	clothing and shoes, not including underwear, undergarments, and nightwear
UND	underwear, undergarments, nightwear and sleeping garments
GAS	gasoline and diesel fuel for motor vehicles
OTH	Other non durables expenses
CAR	the purchase of new and used motor vehicles such as cars, trucks, and vans
JWL	jewelry and watches
HSE	rent, or mortgage, or purchase, of their housing;
	home furnishings and household items;
	homeowners insurance, fire insurance, and property insurance
TOTEXP	total expenditure
p ALH	Alcoholic beverages at home price index
p ALO	Alcoholic beverages away from home price index
p FDO	Food away from home price index
p FDH	Food at home price index
p CLO	Apparel price index
p UND	Women's apparel (underwear prices are not available 1993-1996) price index

Var name	Variables description
p GAS	Motor fuel price index
p OTH	All items price index
h ALH	log price ALH-log price OTH
h ALO	log price ALO-log price OTH
h FDO	log price FDO-log price OTH
h FDH	log price FDH-log price OTH
h CLO	log price CLO-log price OTH
h UND	log price UND-log price OTH
h GAS	log price GAS-log price OTH
stone	$\sum_{\{X=ALH,ALO,FDO,FDH,CLO,UND,GAS\}} X \ln(X)$
IYEAR 1994	year dummy
IYEAR 1995	year dummy
IYEAR 1996	year dummy
IYEAR 1997	year dummy
IYEAR 1998	year dummy
IYEAR 1999	year dummy
IYEAR 2000	year dummy
IYEAR 2001	year dummy
IYEAR 2002	year dummy
IQTR 2	quarter 2 dummy
IQTR 3	quarter 3 dummy
IQTR 4	quarter 4 dummy
IREGION 2	North Central dummy
IREGION 3	South dummy
IREGION 4	West dummy
IOCCUP1 2	Technical, sales, and administrative support occupations dummy
IOCCUP1 3	Service occupations dummy
IOCCUP1 4	Farming, forestry, and fishing occupations dummy
IOCCUP1 5	Precision production, craft, and repair occupations dummy
IOCCUP1 6	Operators, fabricators, and laborers dummy
IOCCUP1 7	Armed forces dummy
IOCCUP1 8	Self-employed dummy
IOCCUP1 9	Not working dummy
IOCCUP1 10	Retired dummy
SEX REF	Sex of reference person
AGE REF	age of reference person
YR EDREF	year of education reference person
IMARITAL1 2	Widowed dummy
IMARITAL1 3	Divorced dummy
IMARITAL1 4	Separated dummy
IMARITAL1 5	Never married dummy
PERSLT18	"Number of children less than 18 "
PERSOT64	Number of persons over 64 in CU
IREF RACE 2	Black
IREF RACE 3	American Indian, Aleut, Eskimo
IREF RACE 4	Asian or Pacific Islander
m ALH	mean budget share of ALH
m ALO	mean budget share of ALO
m FDO	mean budget share of FDO
m FDH	mean budget share of FDH
m CLO	mean budget share of CLO
m UND	mean budget share of UND
m GAS	mean budget share of GAS
m OTH	mean budget share of OTH
lnx	$\log TOTEXP - stone$

Var name	Variables description
lnx2	$(\log TOTEXP - stone)^2$

	Estimation Subsample				US–wide sample	
	mean	**sd**	**min**	**max**	**mean**	**sd**
ALH	169.4168	323.9644	0	9689	156.0034	305.6665
ALO	148.1916	349.6133	0	8596	137.3304	328.3154
FDO	1496.894	1924.96	0	54991	1410.301	1766.066
FDH	3946.552	2184.401	0	22452	3787.429	2100.249
CLO	810.556	1061.452	0	33948	801.5828	1021.236
UND	138.5562	199.0201	0	2964	137.63	196.3205
GAS	1176.581	933.4128	0	9270	1172.394	925.0178
OTH	2.57E+07	3.96E+07	0	1.06E+09	11044.81	8904.229
CAR	3223.62	7905.023	0	95580	3278.012	8008.563
JWL	168.4439	1900.58	0	210000	148.0257	1271.566
HSE	5398478	1.31E+07	0	5.07E+08	3728.37	4086.647
TOTEXP	28370.56	20634.27	707.9996	743532.3	27190.09	19419.9
p ALH	99.06309	4.604702	90.89744	105.641		
p ALO	98.36219	7.944797	82.3299	110.7195		
p FDO	98.52624	6.234391	86.00479	107.7153		
p FDH	98.59102	5.979608	84.1852	106.0734		
p CLO	102.4084	3.366371	93.61198	107.571		
p UND	105.9989	5.466155	92.67873	118.8631		
p GAS	98.71063	13.41501	74.24512	130.373		
p OTH	99.134	6.049358	85.95972	107.4052		
h ALH	0.0000973	0.0168857	-0.0223212	0.0579662		
h ALO	-0.0092608	0.0216242	-0.0502381	0.0312734		
h FDO	-0.0062909	0.0066385	-0.021008	0.0055633		
h FDH	-0.0054863	0.008038	-0.0208597	0.0149212		
h CLO	0.0338519	0.0876056	-0.1308093	0.2179918		
h UND	0.0675184	0.10364	-0.1408286	0.3083668		
h GAS	-0.0115424	0.1079955	-0.2719941	0.2138472		
stone	2.497275	0.7220481	0.0668289	4.423194		
IYEAR 1994	0.0697169	0.2546783	0	1	0.0757231	0.2645582
IYEAR 1995	0.0647422	0.2460789	0	1	0.0719397	0.2583916
IYEAR 1996	0.032301	0.1768045	0	1	0.033804	0.1807268
IYEAR 1997	0.1103559	0.3133439	0	1	0.1111995	0.3143833
IYEAR 1998	0.109375	0.3121201	0	1	0.1131186	0.316742
IYEAR 1999	0.1144899	0.3184165	0	1	0.117231	0.3216997
IYEAR 2000	0.1625561	0.3689731	0	1	0.1545716	0.3615008
IYEAR 2001	0.1566704	0.3635025	0	1	0.1515284	0.3585681
IYEAR 2002	0.1619254	0.3683953	0	1	0.1525977	0.3596042
IQTR 2	0.2383688	0.4261008	0	1	0.2442221	0.4296309
IQTR 3	0.2378083	0.425756	0	1	0.2391501	0.4265704
IQTR 4	0.2698991	0.4439227	0	1	0.2744071	0.4462212
IREGION 2	0.1617152	0.3682023	0	1	0.2673338	0.4425741
IREGION 3	0.2397001	0.4269154	0	1	0.33878	0.4733014
IREGION 4	0.3462024	0.4757753	0	1	0.1927622	0.3944733
IOCCUP1 2	0.1403447	0.3473565	0	1	0.1390267	0.3459792
IOCCUP1 3	0.1122478	0.3156821	0	1	0.1133105	0.3169763
IOCCUP1 4	0.0073571	0.0854602	0	1	0.00817	0.0900192
IOCCUP1 5	0.0519198	0.221873	0	1	0.0533242	0.2246822
IOCCUP1 6	0.0818386	0.2741282	0	1	0.0947498	0.2928731
IOCCUP1 7	0.0044142	0.0662952	0	1	0.0032625	0.0570259
IOCCUP1 8	0.0349636	0.183694	0	1	0.0395065	0.1947994

	Estimation Subsample				US–wide sample	
	mean	**sd**	**min**	**max**	**mean**	**sd**
IOCCUP1 9	0.0985846	0.298114	0	1	0.1012474	0.3016602
IOCCUP1 10	0.2282791	0.4197382	0	1	0.2136258	0.4098712
SEX REF	1.430143	0.4951133	1	2	1.432433	0.4954205
AGE REF	51.36848	17.06942	17	94	50.8984	16.92091
YR EDREF	13.82112	2.813901	0	18	13.70314	2.809938
IMARITAL1 2	0.1219871	0.3272824	0	1	0.1206032	0.32567
IMARITAL1 3	0.1340387	0.3407058	0	1	0.1321727	0.3386831
IMARITAL1 4	0.0298487	0.1701756	0	1	0.0279644	0.164873
IMARITAL1 5	0.1387332	0.34568	0	1	0.1363674	0.343183
PERSLT18	0.7101317	1.131586	0	10	0.7067032	1.108377
PERSOT64	0.3805353	0.6572266	0	4	0.3587389	0.6448471
IREF RACE 2	0.1053111	0.3069646	0	1	0.115257	0.3193362
IREF RACE 3	0.0058156	0.0760406	0	1	0.007512	0.0863468
IREF RACE 4	0.0557035	0.2293562	0	1	0.0325977	0.1775836
m ALH	0.1526445	0.1480501	0.0002917	0.6559903		
m ALO	0.118248	0.1004168	0.000288	0.4505704		
m FDO	1.222685	1.146591	0.0045779	5.28772		
m FDH	3.911378	3.657658	0.0235374	16.5382		
m CLO	0.6206222	0.5814182	0.0016867	2.735063		
m UND	0.1158841	0.1091201	0.0003363	0.5074397		
m GAS	1.085815	1.054143	0.0055939	4.869476		
m OTH	9.659403	8.539694	0.0612457	39.68477		
lnx	7.558408	0.9215498	3.685857	11.09327		
lnx2	57.97873	14.35368	13.58554	123.0606		

B Equation–by–equation estimation results

	GAS	GAS	GAS	UND	UND	UND	CLO	CLO	CLO
	beta	std.err	t-stat	beta	std.err	t-stat	beta	std.err	t-stat
m ALH	-0.5394	0.0134	-40.14	0.5218	0.0299	17.47	0.2704	0.0265	10.20
m ALO	0.0778	0.0029	26.77	-0.1202	0.0063	-19.04	0.0185	0.0058	3.17
m FDO	0.0070	0.0009	7.44	-0.0073	0.0021	-3.42	-0.0050	0.0018	-2.77
m FDH	0.0994	0.0056	17.77	0.0989	0.0132	7.48	0.3516	0.0452	7.77
m CLO	-0.1220	0.0234	-5.22	0.0424	0.0075	5.64	-0.1068	0.0268	-3.98
m UND	0.0196	0.0033	5.88	-0.4177	0.0318	-13.13	-0.0434	0.0057	-7.61
m GAS	0.2967	0.0138	21.45	-0.6437	0.0647	-9.94	-0.1519	0.0105	-14.43
m OTH	-0.0167	0.0005	-34.55	0.0133	0.0011	12.46	0.0074	0.0010	7.80
CONSTANT	0.3340	0.1766	1.89	-2.8171	0.3459	-8.14	-1.6873	0.4084	-4.13
h ALH	-0.0097	0.0453	-0.21	0.0489	0.0885	0.55	-0.0288	0.1050	-0.27
h ALO	0.0692	0.0561	1.23	-0.0696	0.1097	-0.63	-0.2550	0.1300	-1.96
h FDO	0.0502	0.1046	0.48	-0.2224	0.2045	-1.09	0.3154	0.2425	1.30
h FDH	-0.0261	0.0449	-0.58	0.1444	0.0879	1.64	0.0119	0.1041	0.11
h CLO	0.0351	0.0310	1.13	-0.0196	0.0607	-0.32	-0.0495	0.0719	-0.69
h UND	-0.0075	0.0184	-0.41	-0.0236	0.0359	-0.66	-0.0104	0.0425	-0.25
h GAS	0.0067	0.0050	1.34	0.0011	0.0098	0.11	0.0003	0.0116	0.03
IYEAR 1994	0.0007	0.0016	0.42	-0.0007	0.0032	-0.22	0.0067	0.0038	1.77
IYEAR 1995	0.0017	0.0025	0.68	-0.0028	0.0049	-0.57	0.0030	0.0058	0.52
IYEAR 1996	0.0001	0.0030	0.05	-0.0021	0.0059	-0.35	0.0080	0.0070	1.15
IYEAR 1997	0.0012	0.0031	0.40	-0.0068	0.0061	-1.11	0.0071	0.0072	0.98
IYEAR 1998	-0.0100	0.0034	-2.92	0.0072	0.0067	1.07	0.0143	0.0080	1.80
IYEAR 1999	-0.0110	0.0035	-3.11	0.0074	0.0069	1.08	0.0159	0.0081	1.95
IYEAR 2000	-0.0108	0.0039	-2.79	0.0106	0.0076	1.40	0.0157	0.0090	1.75
IYEAR 2001	-0.0104	0.0046	-2.25	0.0127	0.0090	1.40	0.0148	0.0107	1.39
IYEAR 2002	-0.0104	0.0049	-2.11	0.0111	0.0097	1.15	0.0129	0.0114	1.13
IQTR 2	-0.0008	0.0007	-1.18	0.0008	0.0013	0.59	0.0019	0.0016	1.19
IQTR 3	-0.0005	0.0006	-0.86	0.0018	0.0011	1.62	0.0011	0.0013	0.80

	GAS beta	GAS std.err	GAS t-stat	UND beta	UND std.err	UND t-stat	CLO beta	CLO std.err	CLO t-stat
IQTR 4	-0.0010	0.0006	-1.57	0.0036	0.0013	2.85	0.0013	0.0015	0.88
SEX REF	0.0008	0.0003	2.30	-0.0014	0.0006	-2.12	-0.0082	0.0008	-10.61
IREGION 2	-0.0228	0.0008	-27.15	0.0272	0.0018	15.10	0.0203	0.0017	11.70
IREGION 3	-0.0047	0.0007	-6.71	0.0074	0.0015	5.03	0.0046	0.0015	3.09
IREGION 4	-0.0048	0.0007	-6.53	0.0195	0.0015	12.60	0.0009	0.0015	0.58
IOCCUP1 2	0.0000	0.0005	0.08	0.0011	0.0010	1.12	-0.0048	0.0012	-3.92
IOCCUP1 3	-0.0007	0.0006	-1.19	0.0053	0.0011	4.93	-0.0096	0.0013	-7.48
IOCCUP1 4	0.0008	0.0017	0.47	0.0126	0.0034	3.71	-0.0026	0.0040	-0.65
IOCCUP1 5	-0.0003	0.0007	-0.36	0.0111	0.0014	7.87	-0.0061	0.0017	-3.65
IOCCUP1 6	0.0001	0.0006	0.08	0.0106	0.0012	8.47	-0.0079	0.0015	-5.37
IOCCUP1 7	0.0005	0.0022	0.23	0.0086	0.0043	2.00	0.0026	0.0051	0.52
IOCCUP1 8	-0.0004	0.0008	-0.47	0.0034	0.0016	2.10	-0.0077	0.0019	-4.01
IOCCUP1 9	-0.0001	0.0006	-0.20	-0.0020	0.0011	-1.74	-0.0120	0.0013	-8.86
IOCCUP1 10	-0.0003	0.0007	-0.49	0.0078	0.0013	5.89	-0.0024	0.0016	-1.53
AGE REF	0.0038	0.0016	2.45	-0.0051	0.0031	-1.66	-0.0088	0.0036	-2.42
YR EDREF	-0.0096	0.0062	-1.55	0.0062	0.0121	0.52	0.0168	0.0143	1.17
IMARITAL1 2	0.0107	0.0009	11.85	-0.0013	0.0018	-0.73	-0.0020	0.0021	-0.99
IMARITAL1 3	0.0080	0.0007	11.78	-0.0004	0.0013	-0.26	-0.0013	0.0015	-0.84
IMARITAL1 4	0.0078	0.0010	7.74	-0.0012	0.0020	-0.63	-0.0007	0.0023	-0.29
IMARITAL1 5	0.0092	0.0008	11.13	-0.0020	0.0016	-1.24	0.0037	0.0019	1.98
PERSLT18	0.0199	0.0181	1.10	-0.0826	0.0354	-2.33	-0.0655	0.0419	-1.56
PERSOT64	-0.0840	0.0424	-1.98	0.0007	0.0830	0.01	0.0671	0.0982	0.68
IREF RACE 2	0.0079	0.0006	13.19	-0.0179	0.0012	-14.84	-0.0162	0.0013	-12.15
IREF RACE 3	0.0099	0.0020	5.01	-0.0180	0.0039	-4.65	-0.0256	0.0046	-5.62
IREF RACE 4	-0.0013	0.0008	-1.53	0.0003	0.0017	0.19	0.0043	0.0018	2.33
lnx	-0.0894	0.0464	-1.92	0.7848	0.0909	8.63	0.4747	0.1074	4.42
it lnx AGE	-0.0010	0.0004	-2.45	0.0012	0.0008	1.53	0.0021	0.0009	2.25
it lnx LT18	-0.0043	0.0046	-0.93	0.0200	0.0091	2.20	0.0134	0.0107	1.25
it lnx OT64	0.0214	0.0109	1.96	-0.0031	0.0213	-0.14	-0.0195	0.0253	-0.77
it lnx EDU	0.0025	0.0016	1.53	-0.0020	0.0031	-0.65	-0.0032	0.0037	-0.85
lnx2	0.0057	0.0030	1.89	-0.0524	0.0059	-8.83	-0.0319	0.0070	-4.56
it lnx2 AGE	0.0001	0.0000	2.40	-0.0001	0.0001	-1.40	-0.0001	0.0001	-2.05
it lnx2 LT18	0.0002	0.0003	0.82	-0.0012	0.0006	-2.13	-0.0007	0.0007	-1.01
it lnx2 OT64	-0.0013	0.0007	-1.90	0.0003	0.0014	0.20	0.0013	0.0016	0.80
it lnx2 EDU	-0.0001	0.0001	-1.45	0.0001	0.0002	0.67	0.0001	0.0002	0.60
CAR	-0.0213	0.0124	-1.72	0.1940	0.0243	7.99	0.1371	0.0287	4.78
JWL	0.0196	0.0127	1.55	-0.0349	0.0247	-1.41	0.2166	0.0293	7.38
HSE	0.0095	0.0034	2.79	-0.0752	0.0067	-11.29	-0.0508	0.0079	-6.46
ρ	0.0244			0.02954			0.01706		
σ_ϵ^2	0.0007			0.00004			0.00046		

	FDH beta	FDH std.err	FDH t-stat	FDO beta	FDO std.err	FDO t-stat
m ALH	-0.0420	0.0577	-0.73	-0.3301	0.0184	-17.91
m ALO	-0.0295	0.0124	-2.38	0.0056	0.0013	4.34
m FDO	0.0362	0.0244	1.48	0.1657	0.0190	8.71
m FDH	0.0819	0.0599	1.37	0.1247	0.0076	16.34
m CLO	-0.3162	0.1007	-3.14	-0.0990	0.0320	-3.09
m UND	-0.0073	0.0139	-0.53	0.0253	0.0043	5.83
m GAS	0.0325	0.0041	7.87	0.0152	0.0042	3.60
m OTH	-0.0086	0.0021	-4.14	-0.0120	0.0007	-17.82
CONSTANT	-4.4741	0.7302	-6.13	-0.8036	0.2471	-3.25
h ALH	-0.1158	0.1872	-0.62	-0.0740	0.0634	-1.17
h ALO	0.2204	0.2318	0.95	-0.0354	0.0785	-0.45
h FDO	-0.8672	0.4324	-2.01	0.3020	0.1464	2.06
h FDH	0.1574	0.1857	0.85	-0.0054	0.0629	-0.09

	FDH beta	FDH std.err	FDH t-stat	FDO beta	FDO std.err	FDO t-stat
h CLO	-0.0293	0.1282	-0.23	-0.0246	0.0434	-0.57
h UND	-0.0671	0.0758	-0.89	0.0205	0.0257	0.80
h GAS	-0.0331	0.0206	-1.60	0.0103	0.0070	1.47
IYEAR 1994	-0.0047	0.0067	-0.70	0.0016	0.0023	0.70
IYEAR 1995	-0.0149	0.0104	-1.43	0.0018	0.0035	0.51
IYEAR 1996	-0.0235	0.0124	-1.89	0.0012	0.0042	0.29
IYEAR 1997	-0.0344	0.0128	-2.68	0.0004	0.0044	0.10
IYEAR 1998	-0.0280	0.0142	-1.97	-0.0061	0.0048	-1.27
IYEAR 1999	-0.0220	0.0146	-1.51	-0.0070	0.0049	-1.42
IYEAR 2000	-0.0291	0.0160	-1.82	-0.0078	0.0054	-1.44
IYEAR 2001	-0.0392	0.0191	-2.05	-0.0102	0.0065	-1.58
IYEAR 2002	-0.0409	0.0204	-2.00	-0.0120	0.0069	-1.75
IQTR 2	0.0050	0.0028	1.78	-0.0006	0.0009	-0.69
IQTR 3	0.0026	0.0024	1.10	0.0002	0.0008	0.28
IQTR 4	0.0036	0.0027	1.36	-0.0015	0.0009	-1.68
SEX REF	0.0037	0.0014	2.67	-0.0002	0.0005	-0.44
IREGION 2	0.0044	0.0036	1.24	-0.0130	0.0012	-11.17
IREGION 3	-0.0064	0.0030	-2.17	-0.0038	0.0010	-3.86
IREGION 4	0.0072	0.0031	2.32	-0.0027	0.0010	-2.67
IOCCUP1 2	-0.0035	0.0022	-1.62	-0.0025	0.0007	-3.44
IOCCUP1 3	0.0124	0.0023	5.45	-0.0054	0.0008	-6.93
IOCCUP1 4	0.0348	0.0072	4.86	-0.0024	0.0024	-0.97
IOCCUP1 5	0.0121	0.0030	4.07	-0.0035	0.0010	-3.50
IOCCUP1 6	0.0097	0.0026	3.66	-0.0035	0.0009	-3.91
IOCCUP1 7	0.0053	0.0090	0.59	-0.0006	0.0031	-0.18
IOCCUP1 8	0.0003	0.0034	0.09	-0.0021	0.0012	-1.77
IOCCUP1 9	0.0360	0.0024	14.96	-0.0053	0.0008	-6.54
IOCCUP1 10	0.0146	0.0028	5.22	-0.0028	0.0009	-3.00
AGE REF	-0.0007	0.0065	-0.11	-0.0011	0.0022	-0.48
YR EDREF	0.0495	0.0256	1.94	0.0025	0.0086	0.28
IMARITAL1 2	-0.0055	0.0037	-1.49	0.0097	0.0013	7.75
IMARITAL1 3	-0.0021	0.0028	-0.76	0.0071	0.0009	7.52
IMARITAL1 4	0.0107	0.0042	2.55	0.0089	0.0014	6.30
IMARITAL1 5	0.0090	0.0034	2.62	0.0100	0.0012	8.64
PERSLT18	0.3011	0.0748	4.02	0.0928	0.0253	3.66
PERSOT64	-0.0712	0.1753	-0.41	-0.0321	0.0594	-0.54
IREF RACE 2	0.0083	0.0025	3.33	0.0075	0.0008	9.04
IREF RACE 3	-0.0079	0.0082	-0.97	-0.0013	0.0028	-0.48
IREF RACE 4	0.0140	0.0035	4.05	-0.0003	0.0012	-0.22
lnx	1.3148	0.1920	6.85	0.2245	0.0650	3.45
it lnx AGE	0.0004	0.0017	0.26	0.0002	0.0006	0.37
it lnx LT18	-0.0658	0.0192	-3.43	-0.0210	0.0065	-3.24
it lnx OT64	0.0166	0.0451	0.37	0.0080	0.0153	0.52
it lnx EDU	-0.0165	0.0066	-2.50	-0.0005	0.0022	-0.22
lnx2	-0.0913	0.0125	-7.28	-0.0152	0.0042	-3.57
it lnx2 AGE	0.0000	0.0001	-0.30	0.0000	0.0000	-0.26
it lnx2 LT18	0.0038	0.0012	3.11	0.0012	0.0004	2.94
it lnx2 OT64	-0.0010	0.0029	-0.34	-0.0005	0.0010	-0.52
it lnx2 EDU	0.0012	0.0004	2.87	0.0000	0.0001	0.22
CAR	0.3560	0.0513	6.94	0.0518	0.0174	2.98
JWL	-0.1879	0.0523	-3.59	0.1986	0.0177	11.21
HSE	0.0581	0.0141	4.13	-0.0189	0.0048	-3.96
ρ	0.02617			0.02373		
σ_ϵ^2	0.00491			0.00148		

	FDH	FDH	FDH	FDO	FDO	FDO
	beta	std.err	t-stat	beta	std.err	t-stat

	ALO	ALO	ALO	ALH	ALH	ALH
	beta	std.err	t-stat	beta	std.err	t-stat
m ALH	0.0009	0.0017	0.51	-0.0050	0.0082	-0.61
m ALO	-0.0244	0.0084	-2.90	0.0019	0.0017	1.12
m FDO	0.0011	0.0005	2.04	0.0000	0.0005	-0.09
m FDH	0.0022	0.0030	0.75	0.0025	0.0029	0.86
m CLO	-0.0157	0.0131	-1.20	-0.0295	0.0130	-2.27
m UND	-0.0019	0.0017	-1.09	-0.0025	0.0017	-1.51
m GAS	0.0649	0.0091	7.17	0.0409	0.0097	4.22
m OTH	-0.0006	0.0003	-2.21	-0.0003	0.0003	-1.19
CONSTANT	0.0872	0.1196	0.73	-0.2467	0.1161	-2.12
h ALH	-0.0656	0.0307	-2.14	-0.0275	0.0298	-0.92
h ALO	0.0555	0.0381	1.46	-0.0471	0.0369	-1.28
h FDO	0.0048	0.0710	0.07	0.0330	0.0689	0.48
h FDH	0.0738	0.0305	2.42	0.0014	0.0296	0.05
h CLO	0.0100	0.0210	0.48	-0.0137	0.0204	-0.67
h UND	0.0129	0.0125	1.04	-0.0049	0.0121	-0.41
h GAS	0.0057	0.0034	1.69	-0.0017	0.0033	-0.51
IYEAR 1994	0.0004	0.0011	0.34	0.0004	0.0011	0.39
IYEAR 1995	-0.0008	0.0017	-0.45	-0.0007	0.0017	-0.40
IYEAR 1996	-0.0008	0.0020	-0.39	-0.0005	0.0020	-0.25
IYEAR 1997	-0.0006	0.0021	-0.28	-0.0012	0.0020	-0.57
IYEAR 1998	-0.0010	0.0023	-0.44	-0.0015	0.0023	-0.68
IYEAR 1999	-0.0011	0.0024	-0.46	-0.0018	0.0023	-0.76
IYEAR 2000	-0.0007	0.0026	-0.28	-0.0015	0.0026	-0.59
IYEAR 2001	-0.0004	0.0031	-0.14	-0.0029	0.0030	-0.96
IYEAR 2002	0.0008	0.0033	0.25	-0.0022	0.0032	-0.69
IQTR 2	-0.0010	0.0005	-2.20	0.0006	0.0004	1.38
IQTR 3	0.0003	0.0004	0.67	0.0001	0.0004	0.38
IQTR 4	-0.0015	0.0004	-3.35	0.0004	0.0004	0.96
SEX REF	-0.0026	0.0002	-11.66	-0.0029	0.0002	-13.10
IREGION 2	-0.0010	0.0005	-2.02	-0.0003	0.0005	-0.52
IREGION 3	0.0005	0.0004	1.11	0.0001	0.0004	0.16
IREGION 4	0.0002	0.0005	0.47	0.0009	0.0005	1.87
IOCCUP1 2	-0.0009	0.0004	-2.55	-0.0003	0.0003	-0.76
IOCCUP1 3	-0.0008	0.0004	-2.23	0.0002	0.0004	0.57
IOCCUP1 4	-0.0022	0.0012	-1.84	0.0048	0.0011	4.19
IOCCUP1 5	-0.0007	0.0005	-1.49	0.0009	0.0005	1.90
IOCCUP1 6	-0.0013	0.0004	-3.00	-0.0002	0.0004	-0.50
IOCCUP1 7	0.0006	0.0015	0.39	-0.0004	0.0014	-0.30
IOCCUP1 8	-0.0018	0.0006	-3.11	-0.0013	0.0005	-2.30
IOCCUP1 9	-0.0018	0.0004	-4.63	0.0003	0.0004	0.73
IOCCUP1 10	-0.0004	0.0005	-0.83	0.0012	0.0004	2.77
AGE REF	-0.0023	0.0011	-2.21	-0.0021	0.0010	-2.02
YR EDREF	-0.0008	0.0042	-0.20	0.0067	0.0041	1.64
IMARITAL1 2	0.0015	0.0006	2.42	0.0015	0.0006	2.61
IMARITAL1 3	0.0018	0.0005	3.96	0.0018	0.0004	4.12
IMARITAL1 4	0.0021	0.0007	3.10	0.0021	0.0007	3.24
IMARITAL1 5	0.0036	0.0006	6.48	0.0022	0.0005	4.02
PERSLT18	-0.0431	0.0123	-3.51	-0.0166	0.0119	-1.40
PERSOT64	-0.0143	0.0288	-0.50	-0.0015	0.0279	-0.05
IREF RACE 2	-0.0023	0.0004	-5.99	-0.0011	0.0004	-2.87

	ALO beta	ALO std.err	ALO t-stat	ALH beta	ALH std.err	ALH t-stat
IREF RACE 3	-0.0034	0.0013	-2.57	-0.0016	0.0013	-1.21
IREF RACE 4	-0.0018	0.0005	-3.33	-0.0018	0.0005	-3.41
lnx	-0.0123	0.0314	-0.39	0.0751	0.0305	2.46
it lnx AGE	0.0005	0.0003	1.85	0.0005	0.0003	1.77
it lnx LT18	0.0097	0.0031	3.07	0.0034	0.0031	1.12
it lnx OT64	0.0036	0.0074	0.49	0.0001	0.0072	0.02
it lnx EDU	0.0003	0.0011	0.29	-0.0017	0.0011	-1.62
lnx2	0.0003	0.0021	0.15	-0.0053	0.0020	-2.67
it lnx2 AGE	0.0000	0.0000	-1.59	0.0000	0.0000	-1.58
it lnx2 LT18	-0.0005	0.0002	-2.72	-0.0002	0.0002	-0.90
it lnx2 OT64	-0.0002	0.0005	-0.49	0.0000	0.0005	0.00
it lnx2 EDU	0.0000	0.0001	-0.34	0.0001	0.0001	1.59
CAR	0.0142	0.0084	1.69	0.0279	0.0082	3.42
JWL	0.0258	0.0086	3.01	0.0255	0.0083	3.05
HSE	0.0008	0.0023	0.35	0.0016	0.0022	0.71
ρ	0.01800			0.01848		
σ_ϵ^2	0.00013			0.00012		

C OMD estimates of prices' parameters

	ALH	ALO	FDO	FDH	CLO	UND	GAS
ALH	-0.427						
	(0.91)						
ALO	0.311**	-0.021					
	(0.14)	(0.79)					
FDO	0.256**	0.022	0.113				
	(0.06)	(0.05)	(1.76)				
FDH	-0.019	-0.041	-0.112	0.121			
	(0.33)	(0.25)	(0.53)	(5.53)			
CLO	-0.249	0.035	-0.018	-0.063	-0.039		
	(1.37)	(0.62)	(0.2)	(3.99)	(2.74)		
UND	0.061	-0.132	-0.141	0.033	-0.089	0.040	
	(0.2)	(0.18)	(0.19)	(0.76)	(0.48)	(2.86)	
GAS	0.251	-0.606*	0.356**	0.023	-0.451	-0.006	0.064
	(0.92)	(0.32)	(0.11)	(0.3)	(0.95)	(0.36)	(1.45)

std errors in parenthesis. * denotes 10% significance level, ** 5%
χ^2 specification test: 40143.76 [d.f. = 21]