PRODUCT EXPENDITURE PATTERNS IN THE ECPF SURVEY: AN ANALYSIS USING MULTIPLE GROUP LATENT-VARIABLES MODELS⋆

EVA VENTURA

ALBERT SATORRA

Universitat Pompeu Fabra*

Using data from the Spanish household budget survey, we investigate some aspects of household heterogeneity on several product expenditures. We adopt a latent-variable model approach to evaluate the impact of income on expenditures, controlling for the number of members in the family. Two latent factors underlying repeated measures of monetary and non-monetary income are used as explanatory variables in the expenditure regression equations, thus avoiding possible bias associated to the measurement error in income. The proposed methodology also takes care of the case in which product expenditures exhibit a pattern of infrequent purchases. Multiple-group analysis is used to assess the variation of key parameters of the model across various household typologies. The analysis discloses significant variations across groups on the mean levels of expenditures and on the way income and family size affect expenditures. Asymptotic robust methods are used to account for possible non-normality of the data.

---

1. INTRODUCTION

Our main aim in this study is to develop a general framework to analyze the effects of household heterogeneity on product expenditures. Households differ in income, wealth, age, occupation and education of the head and his/her spouse, size, and many other characteristics. Most of these characteristics vary from the time the household is constituted to the time of its dissolution and consequently the volume and composition of its expenditures also change. A typical household starts with a single individual or with a young couple. After some time children arrive, grow, go to school, and eventually leave the household to live on their own. The members of the original household grow old, retire, and some of them might become members of their offspring's household. Of course there are many other types of household, and each type of them exhibit different patterns of expenditure through their life, according to their specific needs and characteristics. Family role transitions from one stage of life to the other are mainly responsible for those differences.

In the consumer behavior literature, it is assumed that a series of status-changing events produce a series of predictable stages or categories that are associated with systematic patterns of expenditures by consumers. Families are classified into several categories related to particular stages of their development, the stage of life of a household being determined by the age of the head (sometimes the age of the spouse is considered instead), the marital status and the number and age of the children. Of course, the types of families to consider and the stages of life they go trough have changed with time to account for recent cultural and institutional developments. Several authors, Schaninger and Danko (1993) among them, compared a number of alternative family cycle models, with families ranging from «traditional» to «modernized».

Once a complete family typology has been determined, the next step is to write a expenditure system of equations, one linear equation for each product. The different household categories are then introduced in the equations by means of dummy variables in a linear regression, thus allowing testing for household heterogeneity effects on the mean level of expenditures. For example, this is the method followed by Wilkes (1995), who used cross-section data on family budgets and provided empirical verification of changes in household spending across a wide variety of products as households pass from one stage of life to another.

In our study, we want to extend the capabilities of this kind of model. First, we believe that the changes of status of the families affect not only the mean level of expenditures, but also the covariance structure of all the observable variables, and in particular the way that income and family size determine expenditures. For that reason we adopt a multiple group estimation strategy, conducting a separate analysis for each household type and testing afterwards for equalities among particular sets of key parameters.

Second, there is a measurement error problem on the explanatory variables when assessing the effect of income on expenditures. While this type of problem has been treated by several authors (see Summers (1959), Liviatan (1961), Biørn (1992) or Aasness, Biørn and Skjerpen (1993) and (1995)) in various different ways, we choose to circumvent the issue of measurement error in income by adopting a latent-variable model approach. Therefore, two of the explanatory variables in our model will be unobservable factors underlying the various measures of income.

Third, the analysis of products that exhibit a pattern of infrequent purchases requires a specific treatment. In this paper infrequent purchases are treated as censored variables. In a first stage of the analysis we estimate the covariance between the underlying uncensored variable and the rest of the variables of the model. The estimated covariance is integrated then into the standard analysis. This allows us to work with any type of expenditure while keeping the same model framework.

Finally, the nature of the data suggests that its distribution may be non-normal. Our estimation procedure uses asymptotic robust methods and we apply the latest statistical developments related to the multiple group analysis and testing for non-normally distributed data.

Our methodology incorporates these various features into a unified model framework. The paper is organized as follows. Section 2 describes the data, model and the statistical analysis. The results are presented in section 3. Section 4 concludes. The details of the statistical model and asymptotic theory used in the paper are gathered in Appendix 1.

2. METHOD

2.1. Data and household typologies

The data set is taken from the Spanish Continuous Survey of Family Budgets (ECPF, 1996). The sample consists of about 3,200 households per quarter and is rotated in a 12% every quarter.

The survey asks the families to keep a detailed record of all kind of expenditures for a period of one week[1]. For some of the more infrequent purchases the survey ask the families to write down the expenditures realized during the last three months. There are two hundred and fifty eight categories of expenditure. We aggregate some of these categories and build the four types of expenditures that we use in the present analysis:

---

[1]The weeks are chosen randomly over the quarter.

transportation, food, durable and medical expenditures[2]. We select those families that remain in the survey for the last two quarters of data consecutively. A few (less than a 3% of the data) outlier observations have been dropped from the data set using the multiple-outliers detection method of Hadi (1992) implemented in the program Stata (1997). The resulting sample size is around 2,600 households.

The survey also collects information on income perceived during the last tree months by every member of the household. This income is both monetary and non-monetary (mainly due to imputations of home-owned rent, which is also considered as part of consumption expenditures). Note that the various measures of income can only be regarded as a «proxy» of the «true» value of income.

We can identify two main sources of inaccuracy of the reported income. The first one is based in the systematic bias of income and it is known as underreporting. In fact, in our survey families consistently seem to underreport income. The second one has to do with the *reliability* of reported income; i.e. the fraction of variance of the observed income attributable to a random component of measurement error. In the literature of measurement error this second issue is assessed by the so called reliability coefficient, which is defined as the ratio between the variance of the «true» (unobservable) income and the variance of the observable income. It is this second source of error, i.e. a reliability coefficient different than one, the one that can seriously bias OLS estimates of parameters such as the effect of income on expenditures. The latent-variable model approach used in this paper prevents this type of bias.

With regard to the household typologies, we consider the following groups:

1. YOUNG: Young singles or young couples without children. Those are families of one or two (married) members in which the head of the family is less than 65 years old.

2. CHILDREN: Families with young children (at least one child is less than 15 year old). These are families in which the presence of a child is the only common characteristic. Families in which the head of the family is the grandfather are mixed with families constituted by just one couple and some children, or families of single or divorced parents.

3. TEENS: Families with older children (the youngest child is more than 14 years old and less than 25). Again families are mixed, as in the preceding group.

---

[2]We have chosen four diverse categories of spending to illustrate the performance of the model. Many other products could be examined instead.

4. ADULTS: Families constituted exclusively by adults, other than couples or singles. This group includes young couples living with their parents, old couples living with non- emancipated siblings, or just non-related people living together.

5. OLD: Old singles or couples living by themselves. Those are families of one or two (married) members in which the head of the family is more than 64 years old.

Other typologies could be considered and their consumption behavior studied. However, we believe that the presence of children of different ages is the fact that more strongly determines the consumption needs of a household. Furthermore, retirement also influences the patterns of consumption in a powerful way.

2.2. Model

The latent-variable model approach has been used successfully in several areas of empirical investigation. One of the oldest models of this type is the Factor Analysis model, which postulates that the covariance among a set of observable variables is produced by the variation of underlying latent variables (factors). Nowadays, a very popular latent-variable model is LISREL (Jöreskog and Sörbom, 1994). To give a few economic related examples of latent-variable models, we can cite the work of Punj and Staelin (1983) in consumer behavior, the work of Anderson (1985) and Bagozzi (1980) in marketing, Fritz (1986) in management science, or McFatter (1987) in discrimination in salaries. For an introduction to structural equations with latent-variable models see Bollen (1989).

We specify a multiple group latent-variable model that can explain the behavior of most products' expenditures and that takes account of heterogeneity of family behavior on product expenditures. Each type of expenditure is assumed to depend on two factors (latent-variables) which are linearly related to measures of monetary and non-monetary income of the households in different periods of time. The number of members of the household is used as a covariate of the model. In our multiple group set up, the means of expenditures and the income regression coefficients are allowed to vary across household typologies.

In our analysis the spending behavior of families varies not only due to changes in income, but also depending on the stage of life the family is going through at the moment. Such stages are reflected by the a priori defined family typologies. That is, a young single household is thought to show very different consumption patterns from a household with young children, or an old age couple household. It is not just a matter of income, but a matter of preferences, taste, family composition, family needs, and so on. A common model is analyzed for different groups of households, the groups corresponding to different stages of the life of the family, and for different types of product expenditures. The analysis assesses the variation of the parameters of the model across groups, not

only of the intercept parameters but also of the regression coefficients. The intercept parameters determine the mean levels of the variables while the regression coefficients affect the relationships between expenditures, income and number of members of the family.

The specific model to be analyzed is the following:

(1) $\qquad$ $\text{PRODUCT} = \alpha_0 + \beta_1 \text{MEMBER} + \beta_2 F1 + \beta_3 F2 + \varepsilon_0$

(2) $\qquad$ $\text{INCOME1} = \alpha_1 + F1 + \varepsilon_1$

(3) $\qquad$ $\text{INCOME2} = \alpha_2 + F2 + \varepsilon_2$

(4) $\qquad$ $\text{INCOME1}_{-1} = \alpha_3 + \lambda_1 F1 + \varepsilon_3$

(5) $\qquad$ $\text{INCOME2}_{-1} = \alpha_4 + \lambda_2 F2 + \varepsilon_4$

Where: PRODUCT is the product expenditure we want to consider; MEMBER is the number of members in the household; and F1 and F2 are latent-variables underlying two indicators (current and one quarter behind) of reported monetary and non-monetary income respectively. The variables INCOME1 and INCOME1_{-1} refer to monetary income in the current and last periods respectively. Similarly, INCOME2 and INCOME2_{-1} represent non monetary income at the current and one quarter behind periods. The α parameters are the intercepts of the regression equations; the β's are the regression coefficients measuring the effects of two sources of income on expenditure; finally, the λ's are the loading parameters of the observed variables on the different factors. The ε's correspond to the disturbance terms of the regression equations and the factor model equation. Figure 1 shows a path-diagram representation of the model. In the figure the observed variables are enclosed in rectangles and factors are inside a circle. The diamond represents the constant term. Solid arrows represent regression and loading coefficients, while discontinuous ones represent the intercept parameters. Double arrows represent covariances among independent variables.

The above model is a specific case of the Bentler-Week's model implemented in the EQS package (Bentler, 1995). We use the multiple group approach with various levels of constraints across groups that correspond to substantive hypothesis on household heterogeneity effects. The model is estimated by Generalized Least Squares with an optimal weight matrix under normality. We use asymptotic robust standard errors and test statistics to take care for possible non-normality of the data. (See, for example, Satorra (1993), Satorra and Bentler (1994) and Satorra and Bentler (1999) for the theory of asymptotic robustness of LISREL type models). In this paper we have used the statistical package LISREL (Jöreskog and Sörbom, 1994), which in its latest version also provides robust standard errors and t-statistics. To deal with censored and ordinal dependent variables we used the statistical software PRELIS (Jöreskog and
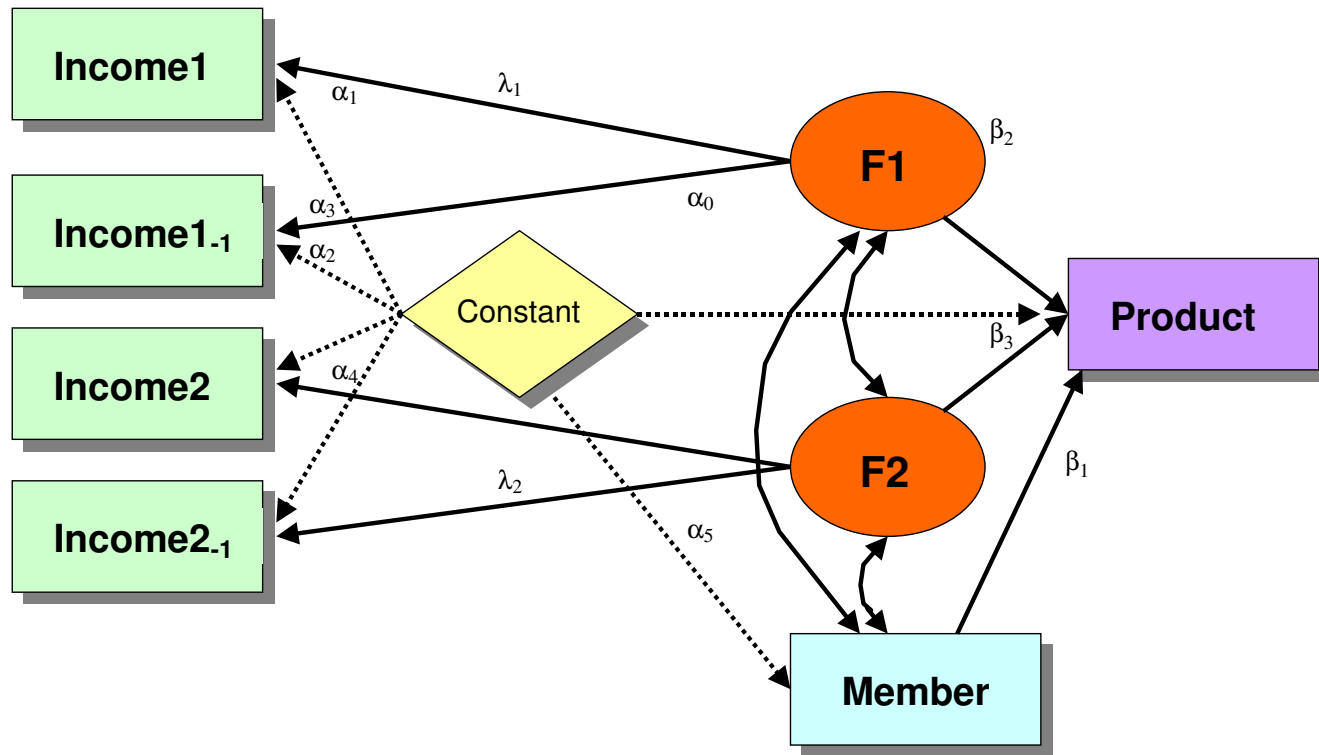
Figure 1. Path Diagram of the Product Expenditures Model.

Sörbom, 1994)[3]. The program code for the PRELIS and LISREL runs are available from the authors upon request.

3. RESULTS

The following subsections describe the statistical results of the analysis for each type of expenditure. An economic interpretation is offered at the end of the section.

3.1. Transportation and Communications Expenditures

Tables 1 to 4 report the parameter estimates and the test statistics of the model presented in section 2, for Transportation, Food, Durable and Medical expenditures respectively. (We do not show standard errors and t-values for those parameters which are known to be significant from a priori grounds, such as the λ's and the α's.). Asymptotic robust t-values are shown within brackets below the parameter estimates. The right columns of the table show the test statistics and the restricted (across-groups) parameter estimates associated to different null hypothesis concerning heterogeneity in family behavior. In all the tables, when a parameter estimate is significantly different than zero or a test statistic rejects the null hypothesis, the corresponding value is emphasized in bold.

Table 1 reports an acceptable fit of the unrestricted model: the chi-square goodness-of-fit of the unrestricted model is 23.87 with 30 degrees of freedom, which corresponds to a P-value of 0.78. We also observe that the intercept of the Transportation equation is basically the same for the first four groups, however it drops dramatically with the last group (old singles living alone or old couples).

Table 1 shows also that the number of members in the household does not affect significantly the transportation and communications expenditures. The coefficient of the first latent-variable −whose indicators are monetary income in the current and previous periods− is highly significant. In contrast, the coefficient of the second latent-variable (indicated by the non-monetary income variables) is significant only for the third (families in which the youngest member is a teenager) and last group (old singles living alone or old couples).

The last two columns serve to analyze household heterogeneity behavior. In these columns we show statistics associated to multiple group analysis for testing various equalities of parameters across groups. We report the value of the difference chi-square test

---

[3]EQS is one alternative commercial software to carry out this type of analysis. See Appendix 1 for more details on the statistical analysis used in this paper.

Table 1. Parameter estimates and test statistics for Transportation expenditures

	Groups					Testing for Household Heterogeneity effects on a:			
						single parameter		set of parameters	
	1	2	3	4	5	Difference test	Restricted parameters	Differences test	Restricted parameters
N	278	896	586	380	426				
β_1 MEMBER	-0.01 (−0.61)	-0.01 (−0.98)	0.02 (0.95)	0.00 (−0.02)	0.04 (1.46)	1.80 (0.77)	0.00 (0.20)	−	−
β_2 F1	**0.11** (5.43)	**0.07** (7.30)	**0.07** (6.27)	**0.09** (3.75)	**0.08** (4.97)	3.87 (0.42)	**0.08** (12.54)	**10.68** (0.03)	**0.08** (12.38)
β_3 F2	−0.02 (−0.63)	0.05 (1.82)	**0.11** (3.16)	−0.01 (−0.18)	**0.09** (2.54)	4.91 (0.30)	**0.06** (3.32)		**0.06** (3.22)
λ_1 F1	**1.15**	**1.03**	**1.23**	**1.53**	**1.22**	4.88 (0.30)	**1.15**	9.71 (0.05)	**1.15**
λ_2 F2	**1.03**	**1.01**	**1.08**	**0.89**	**1.00**	2.46 (0.65)	**1.01**		**1.01**
α_0	**0.49**	**0.68**	**0.65**	**0.51**	**0.10**	**27.98**	**0.37**	**675.17** (0.00)	**0.58**
α_1	**3.86**	**4.93**	**5.37**	**5.02**	**2.31**	**110.41**	**4.37**		**4.86**
α_2	**4.33**	**5.37**	**6.03**	**6.05**	**2.77**	**114.58**	**4.65**		**5.40**
α_3	**0.98**	**1.05**	**1.14**	**1.06**	**0.77**	**36.48**	**1.00**		**1.08**
α_4	**0.97**	**1.06**	**1.15**	**1.04**	**0.78**	**36.29**	**1.00**		**1.08**
α_5	**1.70**	**4.32**	**3.94**	**2.89**	**1.57**	**188.60**	**3.41**		**3.40**
χ^2	23.87 ($P = 0.78$, d.f. $= 30$)								

Numbers in brackets below parameter estimates are asymptotic robust t-values. Numbers in brackets below test statistics are p-values. Bold indicates significant at the 5% level.

Table 2. Parameter estimates and test statistics for Food expenditures

	Groups					Testing for Household Heterogeneity effects on a:			
						single parameter		set of parameters	
	1	2	3	4	5	Difference test	Restricted parameters	Differences test	Restricted parameters
N	284	885	588	377	426				
β_1 MEMBER	**0.45** *(6.52)*	**0.28** *(10.63)*	**0.23** *(6.18)*	**0.29** *(5.48)*	**0.46** *(8.91)*	**13.27** ***(0.01)***	**0.29** *(15.83)*	–	–
β_2 F1	0.03 *(1.48)*	**0.05** *(4.53)*	**0.09** *(4.98)*	**0.05** *(2.39)*	**0.08** *(2.36)*	6.21 *(0.18)*	**0.06** *(7.06)*	**7.01** ***(0.14)***	**0.06** *(7.18)*
β_3 F2	0.05 *(1.18)*	0.05 *(1.38)*	0.05 *(1.16)*	0.04 *(0.80)*	0.01 *(0.24)*	0.48 *(0.98)*	**0.04** *(2.11)*		**0.04** *(2.00)*
λ_1 F1	**1.20**	**1.01**	**1.21**	**1.46**	**1.30**	**11.12** ***(0.03)***	**1.16**	**13.68** ***(0.01)***	**1.16**
λ_2 F2	**1.03**	**1.01**	**0.99**	**0.90**	**0.94**	3.08 *(0.54)*	**0.99**		**0.98**
α_0	0.21	**0.46**	**0.84**	**0.59**	**0.18**	**16.6**	**0.40**	**714.97** ***(0.00)***	**0.53**
α_1	**3.92**	**4.90**	**5.36**	**4.99**	**2.31**	**382.6**	**4.38**		**4.85**
α_2	**4.43**	**5.34**	**6.02**	**6.03**	**2.77**	**138.98**	**4.66**		**5.39**
α_3	**1.00**	**1.04**	**1.15**	**1.06**	**0.77**	**44.98**	**1.00**		**1.08**
α_4	**0.99**	**1.05**	**1.16**	**1.04**	**0.78**	**70.11**	**1.00**		**1.08**
α_5	**1.70**	**4.31**	**3.95**	**2.89**	**1.57**	**226.7**	**3.50**		**3.40**
χ^2	20.37 ($P = 0.91$, d.f. $= 30$)								

Numbers in brackets below parameter estimates are asymptotic robust t-values. Numbers in brackets below test statistics are p-values. Bold indicates significant at the 5% level.

Table 3. Parameter estimates and test statistics for Durable Expenditures

| | | | | | | Testing for Household Heterogeneity effects on a: | | | |
| | Groups | | | | | single parameter | | set of parameters | |
	1	2	3	4	5	Difference test	Restricted parameters	Differences test	Restricted parameters
N	284	904	602	384	426				
β_1 MEMBER	**0.39** *(3.07)*	−0.03 *(−0.56)*	**−0.19** *(−2.65)*	−0.05 *(−0.21)*	−0.02 *(−0.31)*	9.43 *(0.05)*	−0.04 *(−1.17)*	−	−
β_2 F1	**0.04** *(2.06)*	**0.14** *(4.61)*	**0.16** *(4.12)*	**0.20** *(2.21)*	**0.18** *(3.08)*	6.76 *(0.15)*	**0.14** *(7.03)*	**11.91** *(0.02)*	**0.14** *(7.32)*
β_3 F2	**0.17** *(2.34)*	**0.17** *(2.03)*	0.08 *(0.91)*	−0.13 *(−0.58)*	−0.01 *(−0.26)*	5.61 *(0.23)*	0.04 *(1.10)*		0.04 *(1.16)*
λ_1 F1	**1.18**	**1.01**	**1.14**	**1.34**	**1.32**	**14.33** *(0.01)*	1.15	**16.72** *(0.00)*	**1.14**
λ_2 F2	**1.05**	**1.00**	**1.04**	**0.90**	**0.93**	4.27 *(0.37)*	**0.99**		**0.99**
α_0	**−0.37**	**0.72**	**1.36**	0.55	0.22	**17.03**	0.33		0.45
α_1	**3.90**	**4.95**	**5.42**	**5.06**	**2.31**	**155.65**	4.41		4.95
α_2	**4.41**	**5.39**	**6.06**	**6.08**	**2.77**	**159.46**	4.72	**755.87** *(0.00)*	5.50
α_3	**0.99**	**1.06**	**1.15**	**1.06**	**0.77**	**73.84**	1.00		1.09
α_4	**0.98**	**1.07**	**1.16**	**1.04**	**0.78**	**75.76**	1.01		1.09
α_5	**1.70**	**4.31**	**3.94**	**2.90**	**1.57**	**244.42**	3.50		3.48
χ^2	18.77 (P = 0.94, d.f. = 30)								

Numbers in brackets below parameter estimates are Normal theory and asymptotic robust t-values. Numbers in brackets below test statistics are p-values. Bold indicates significant at the 5% level.

Table 4. Parameter estimates and test statistics for Medical Expenditures

	Groups					Testing for Household Heterogeneity effects on a:			
						single parameter		set of parameters	
	1	2	3	4	5	Difference test	Restricted parameters	Differences test	Restricted parameters
N	284	904	602	384	426				
β_1 MEMBER	0.09 (1.05)	−0.02 (−2.00)	0.01 (0.26)	0.00 (−0.11)	−0.01 (−0.36)	2.29 (0.68)	−0.01 (−1.62)	−	−
β_2 F1	**0.03** (2.00)	**0.03** (3.04)	0.03 (0.82)	**0.04** (3.43)	**0.07** (3.33)	2.69 (0.61)	**0.03** (4.39)	5.24 (0.26)	**0.03** (4.33)
β_3 F2	0.02 (0.27)	0.04 (1.24)	0.02 (0.27)	0.07 (1.03)	**0.08** (2.07)	1.48 (0.83)	**0.05** (2.45)		**0.05** (2.45)
λ_1 F1	**1.20**	**0.94**	**1.11**	**1.36**	**1.30**	**11.59** (0.02)	1.10	**16.46** (0.00)	**1.10**
α_2 F2	**1.04**	**1.00**	**1.03**	**0.92**	**0.94**	3.61 (0.46)	**0.99**		**0.98**
α_0	−0.23	**0.19**	−0.03	0.00	−0.08	**19.19**	0.05	**725.24** (0.00)	−0.01
α_1	**3.90**	**4.95**	**5.42**	**5.06**	**2.31**	**132.0**	**4.41**		**4.95**
α_2	**4.41**	**5.39**	**6.06**	**6.08**	**2.77**	**146.92**	**4.72**		**5.50**
α_3	**0.99**	**1.06**	**1.15**	**1.06**	**0.77**	**75.04**	**1.01**		**1.09**
α_4	**0.98**	**1.07**	**1.16**	**1.04**	**0.78**	**77.56**	**1.01**		**1.08**
α_5	**1.70**	**4.31**	**3.94**	**2.90**	**1.57**	**216.74**	**3.39**		**3.48**
χ^2	30.22 ($P = 0.45$, d.f. = 30)								

Numbers in brackets below parameter estimates are asymptotic robust t-values. Numbers in brackets below test statistics are p-values. Bold indicates significant at the 5% level.
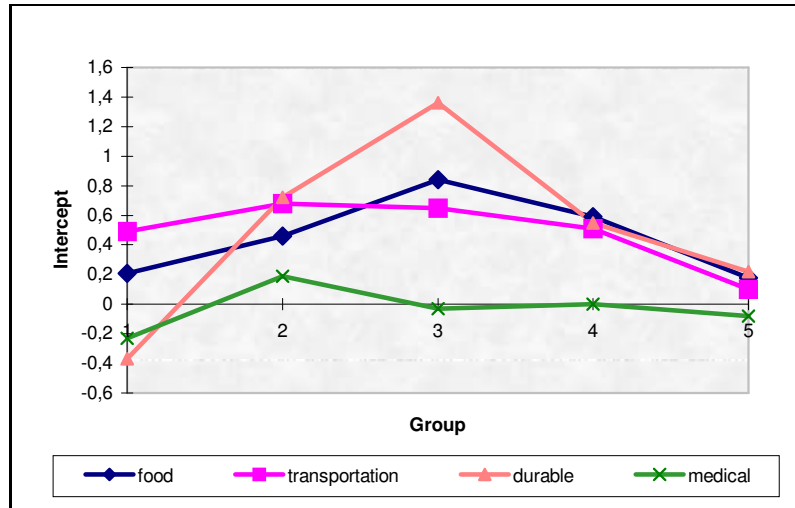
Figure 2. Intercepts of the expenditure equations

statistics and the estimated restricted parameters. The number in brackets below the test statistics is the corresponding P-value. To protect against deviations from normality, these difference test statistics have been computed using the scaled version of the difference chi-square goodness of fit as proposed in Satorra and Bentler (1999). First of all we note that the number of members of the household (MEMBER) does not affect transportation expenditures, therefore it makes no sense to evaluate a household heterogeneity effect on β_1. The same non-significant effect results are observed for the β_3 coefficient, i.e. the impact of non-monetary income on expenditures, with the exception of groups 3 (TEENS) and 5 (OLD) in which we appreciate a significant value. In contrast, we realize that the β_2 coefficients, i.e. the effect of monetary income on transportation, are highly significant for each group. On the other hand, we can not reject the hypothesis of equality of the β_2 parameters across groups (i.e., we do not observe a household heterogeneity effect on the impact of monetary income on Transportation). Household heterogeneity effects are further investigated through the statistics of the last columns of the table.

The variation across groups of the intercept of the regression equation for expenditures (the α_0's parameters) is described in Figure 2. Note the highly significant household heterogeneity effects reflected by the variation of these parameters, which correspond to the variation of expenditures after controlling for family size and unobserved income. In contrast with previous analysis, our model allows for an effect of income and family size that varies across the family groups. Figure 3 is a graphic representation of the variation across household typologies of the intercepts of the measurement equations

105

(parameters α_1 to α_4), i.e. the means of the different income measures. Differences in income related to the family type are clearly appreciated.

3.2. Food expenditures

Table 2 also shows an excellent fit of the model when the product expenditure analyzed is food. The chi-square goodness-of-fit of the unrestricted model is 20.37 (30 degrees of freedom), which corresponds to a P-value of 0.91. In contrast with the transportation expenditures case, now the β_1's (the regression coefficients for MEMBER) are highly significant in each group. We also note a highly significant household heterogeneity effect on β_1, since the hypothesis of equality across groups (a chi-square value of 13.27 for 4 degrees of freedom, P-value of 0.01) is rejected. We also observe significant values for the regression coefficients of the factor associated to monetary income in groups 2 (with children) and 3 (with teenagers), and to less extend in groups 4 (adults) and 5 (old singles and old couples). The values are not significant for the first group (young singles or young couples). The regression coefficient of the factor associated to non-monetary income is clearly non-significant. In conclusion, food expenditures are basically explained by family composition and exhibit a clear household heterogeneity effect through the intercepts and the β_1 coefficient.
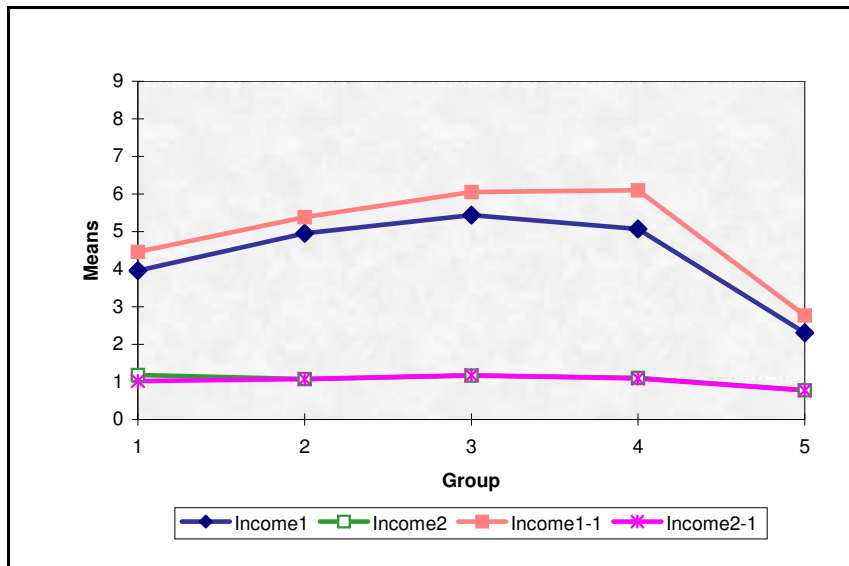

Figure 3. Means of the income measures

106

Table 5. Overview of household heterogeneity effects on parameters

	Transportation		Food		Durables		Medical	
	Single parameters	Set of parameters	Single parameters	Set of parameters	Single parameters	Set of parameters	Single parameters	Set of parameters
β_1 MEMBER	NO	–	YES	–	NO	–	NO	–
β_2 F1	NO	YES	NO	YES	NO	YES	NO	NO
β_3 F2	NO		NO		NO		NO	
λ_1 F1	NO	NO	YES	YES	NO	YES	YES	YES
λ_2 F2	NO		NO		YES		NO	

«YES» indicates that the corresponding test of parameter equality is statistically significant (5% level), and «NO» indicates lack of statistical significance.

In all the equations, the intercept parameters differ significantly across household typologies and therefore the table does not show information on them.

3.3. Durable expenditures

The results for the durable expenditures are shown in Table 3. The chi-square goodness-of-fit of the unrestricted model is 18.77 (30 degrees of freedom), which corresponds to a P-value of 0.94. In this case monetary income is the variable that influences spending in all the cases considered. The value of the β_1 coefficient show that the number of members in the family has a positive effect in the first group, indicating that young couples spend more in durable goods than young singles. Non-monetary income has a weak effect in groups one (YOUNG) and two (CHILDREN). Figure 2 shows the pattern of the intercept of the durable expenditures equation across the different family stages. The household heterogeneity effects are quite evident when we look at the picture. Expenditures rise sharply from group 1 (YOUNG) to group 3 (TEENS) as families are constituted and children are born and grow, and then decrease also quite sharply thereafter. The last columns of the table confirm that the strong household heterogeneity effects are reflected on the intercepts of the equations for which we reject the hypothesis of equality across groups. We can not reject the same hypothesis for the coefficients of the income and number of members variables.

3.4. Medical expenditures

The chi-square goodness-of-fit of the unrestricted model is 30.22 (30 degrees of freedom), which corresponds to a P-value of 0.45. We observe that the t-values associated

with the effect of income on this type of expenditure are much lower than in the previous cases. The β_2 coefficient is not significantly different from zero for the group 3 (TEENS). There is a slight significance of the coefficient of the non-monetary income in group 5 (OLD). Controlling for the number of members of the family becomes unnecessary, since its coefficient is never significantly different from zero in any of the groups. As in the preceding case in which we analyzed durable expenditures, household heterogeneity effects are present through the intercepts of the different equations.

Table 5 gives an overview of the variation across household typologies of the parameters of the model, for the various product expenditures considered.

These results show evidence in favor of the main hypothesis of our work: transitions from one stage of life to another do have an effect on consumer spending. Average expenditures exhibit an inverted U shape, growing steadily from the young stages of life, reaching a maximum for the households with teenager members in them, and declining as the members of the family get old and their offspring leave the household.

Specific types of products follow different patterns of evolution. We observe that the average expenditure in the Transportation and Communication category is about the same for the first four groups of households but it decreases considerably in the old age group. This finding probably reflects the physical difficulties of the oldest members, which forces them to restrict their mobility. It also reflects a particular characteristic of the public transportation system. That is, aged people with the lowest retirement pensions are entitled to lower price transportation fares. Therefore the average amount spent by this group in this category will be lower even if the mobility of the individuals is not diminished. Furthermore, the youngest and the oldest people show a higher propensity to spend in Transportation and Communication out of their income.

The behavior of the expenditures in the Food category is completely different. The average spending grows through the first stages of the life cycle of the families, reaching a maximum when the young members are in the most demanding phase of their physical growth. Thereafter the families diminish their average food spending. As one could expect the number of members in the family strongly determines the total expenditure, but this effect is more important when we compare families with one and two members. Apparently, the addition of one more member to the household causes more consumption variation in these two groups than in the rest. Higher income will also induce higher expenditures in the food category for all the groups, but the effect is rather small.

The most striking differences in spending behavior appear when we examine the Durable goods category. The inverted U shape of the average expenditure through the stages of life is most marked, as can be appreciated in Figure 2. This type of expenditure is strongly influenced by the monetary income, in all the groups. The number of members in the family also affects positively the Durable goods spending of group 1, since new

households are established in the earlier stages of the life cycle and houses have to be furnished. The number of members of the family does not affect spending in Durable goods of the other groups, except however for group 3 which shows a negative and significant regression effect[4].

Finally, the average expenditure in Medical goods and services show a peak for group two: families with small children. One would expect that group five (old age people) should show larger average spending in this category, but again one should remember that most of the medical expenditures have been subsidized until recently by the Spanish government. Old people might be less prone to spend money aside from the Social Security system than young people. This hypothesis would be reinforced by the observation that monetary income also affects medical spending, although not very strongly. In any case, the effect is more apparent in the old age group.

4. CONCLUSION

In the context of Spanish household consumption data, we have analyzed the relationship between product expenditures and income, controlling for family size. A latent-variable model approach was used to assess the impact of income on expenditures, allowing us to circumvent the problem of measurement error present in the income variables. We have also allowed for the case in which expenditures exhibit a pattern of infrequent purchases. The explanatory variables in the regression equations were the number of members in the household and two factors underlying repeated measures of monetary and non-monetary income.

We have found that multiple group analysis is an useful framework through which to specify and test several household heterogeneity hypothesis using classical chi-square tests. Household heterogeneity effects in spending behavior were reflected on the variation of intercept and regression parameters across different family typologies.

We conclude that there are household heterogeneity effects on expenditures, and that these effects vary with the type of expenditure considered. An important finding of our paper is that these household heterogeneity effects have been detected not only on the mean level of consumption but also on the coefficients that assess the impact of income and family size on expenditures.

---

[4]This singular effect deserves further investigation.

5. REFERENCES

Aasness, J., Biørn, E. & Skjerpen, T. (1995). «Distribution of Preferences and Measurement Errors in a Desegregated Expenditure System», *Discussion Papers*, 149. Statistics Norway.

— (1993). «Engel Functions, Panel Data, and Latent-variables», *Econometrica*, 61(6), 1395-1422.

Anderson, J. C. (1985). «A measurement model to assess measure-specific factors in multiple informant research», *Journal of Marketing Research*, 22, 86-92.

Bagozzi, R. P. (1980). *Causal models in marketing*. Wiley, New York.

Biørn, E. (1992). «The Bias of Some Estimators for Panel Data Models with Measurement Errors», *Empirical Economics*, 17, 51-66.

Bollen, K. A. (1989). *Structural Equations with Latent-variables*. John Wiley & sons.

Bentler, P. (1995). *EQS. Structural Equations Program Manual*. University of California, Los Angeles. Encino, CA: Multivariate Software, Inc.

ECPF (1996). *Encuesta Continua de Presupuestos Familiares*, Instituto Nacional de Estadística. Spain

Fritz, W. (1986). «The Lisrel-approach of causal analysis as an instrument of critical theory comparison within management science», in W. Gaul & M. Schader (Eds.), *Classification as a tool of research*, (pp. 142-152). Amsterdam: Elsevier Science.

Hadi, A. S. (1992). «Identifying multiple outliers in multivariate data», *Journal of the Royal Statistical Society*, Series B, 54, 761-771.

Jöreskog, K. G. & Sörbom, D. (1994). *LISREL 7 and PRELIS, A guide to the program and applications*. Chicago: SPSS

Liviatan, N. (1961). «Errors in Variables and Engel Curve Analysis», *Econometrica*, 29, 336-362.

Magnus, J. R. & Neudecker, H. (1991). *Matrix Differential Calculus*, 2nd de. Chichester: Wiley.

McFatter, R. F. (1987). «Use of latent-variable models for detecting discrimination in salaries», *Psychological Bulletin*, 101, 120-125.

Pung, G. N. & Staelin, R. (1983). «A model of consumer information search behavior for new automobiles», *Journal of Consumer Research*, 9, 366-380.

Satorra, A. (1992). «Asymptotic Robust Inferences in the Analysis of Mean and covariance Structures». In: P. V. Marsden (ed.). *e Sociological Methodology 1992* (pp. 249-278). Cambridge, MA: Basic Blackwell.

— (1993). «Asymptotic Robust Inferences in multi-sample analysis of augmented-moment structures». In: C. M. Cuadras and C. R. Rao (Eds.). *Multivariate analysis: Future directions 2* (pp. 211-229). North Holland: Elsevier Science.

Satorra, A. & Bentler, P. M. (1994). «Corrections to Test Statistics and standard errors in covariance structure analysis». In: Alexander von Eye and C. C. Clogg (Eds.). *Latent-variables analysis: applications to developmental research* (pp. 399-419). Thousand Oaks, CA: SAGE .

—— (1999). «A Scaled Difference Chi-Square Test Statistic for Moment Structure Analysis», *UCLA Statistical Series WP*, 120.

Satorra, A. & Neudecker, H. (1994). «On the Asymptotic Optimality of Alternative Minimum-Distance Estimators in Linear Latent-Variable Models», *Econometric Theory*, 10, 867-883.

Schaninger, C. M. & Danko, W. D. (1993). «A Conceptual and Empirical Comparison of Alternative Household Life-cycle Models», *Journal of Consumer Research*, 19, 580-594.

STATACORP. (1997). *Stata Statistical Software: Release 5.0 College Station*, TX: Stata Corporation.

Summers, R. (1959). «A Note on Least Squares Bias in Household Expenditure Analysis», *Econometrica*, 27, 121-126.

Wilkes, R. E. (1995). «Household Life-Cycle Stages, Transitions, and Product Expenditures», *Journal of Consumer Research*, 22, 27-42.

6. APPENDIX 1: ESTIMATION METHOD

The model considered in the paper is a specific case of the following general linear latent-variable model

$$(1) \qquad \eta_i^{(g)} = B^{(g)}\eta_i^{(g)} + \Gamma^{(g)}\xi_i^{(g)}$$

$$(2) \qquad z_i^{(g)} = G^{(g)}v_i^{(g)}, \qquad g = 1,\ldots,G; \quad i = 1,\ldots,n^{(g)}$$

where for each group g, $z_i^{(g)}$ and $n^{(g)}$ are respectively the vector of observable variables and sample size in the g th sample, $v_i^{(g)} = (\eta_i^{(g)\prime}, \xi_i^{(g)\prime})\prime$ is a vector of observable and latent variables, $G^{(g)}$ is a fully specified selection matrix, $B^{(g)}$, $\Gamma^{(g)}$ and the moment matrix $\Phi^{(g)} = E(\xi_i^{(g)}\xi_i^{(g)\prime})$ are parameter matrices of the model. This is the Bentler-Weeks's (e.g., Bentler, 1985) specification of a linear latent-variable model, which is equivalent to the specification in LISREL (Jöreskog and Sörbom, 1995).

A specific model expresses the matrices $B^{(g)}$, $\Gamma^{(g)}$ and $\Phi^{(g)}$, $g = 1,\ldots,G$, as matrix-valued functions of a common vector of parameters θ.

Note that equations (1) and (2) imply

$$(3) \qquad z_i^{(g)} = G^{(g)} \begin{pmatrix} (I - B^{(g)-1})\Gamma^{(g)} \\ I \end{pmatrix} \xi_i^{(g)} = \Lambda^{(g)} \xi_i^{(g)}$$

say, where

$$\Lambda^{(g)} = G^{(g)} \begin{pmatrix} (I - B^{(g)-1})\Gamma^{(g)} \\ I \end{pmatrix};$$

hence, the moment matrices $\Sigma^{(g)} = E z_i^{(g)} z_i^{(g)\prime}$, $g = 1, \ldots, G$, can be expressed as

$$\Sigma^{(g)} = \Lambda^{(g)} \Phi^{(g)} \Lambda^{(g)\prime} = \Sigma^{(g)}(\theta).$$

The analysis proceeds by fitting the matrix-valued functions $\Sigma^{(g)}(\theta)$'s to the sample moment matrices

$$S^{(g)} = \frac{1}{n} \sum_{i=1}^{n_g} z_i^{(g)} z_i^{(g)\prime}, \quad g = 1, \ldots, G.$$

We use the following GLS fitting function:

$$F_{GLS}(\theta) = \frac{1}{2} \sum \frac{n_g}{n} \mathrm{tr} \left\{ \left(S^{(g)} - \Sigma^{(g)} \right) S^{(g)-1} \right\}^2,$$

where $\Sigma^{(g)} = \Sigma^{(g)}(\theta)$ and $n = n_1 + \cdots + n_G$. The minimizer $\hat{\theta}$ of $F_{GLS}(\theta)$ is a minimum-distance estimator that is asymptotically optimal when the $z_i^{(g)}$'s are iid normally distributed (see, e.g., Satorra, 1989).

For general type of distributions, asymptotic robust standard errors and test statistics can be developed. Define:

$$\sigma = \left(\sigma^{(1)}, \ldots, \sigma^{(G)} \right)',$$

with $\sigma^{(g)} = \mathrm{vech}\Sigma^{(g)}$;

$$s = \left(s^{(1)}, \ldots, s^{(G)} \right)',$$

and $s^{(g)} = \mathrm{vech}S^{(g)}$; the Jacobian matrix $R = \left. \dfrac{\partial \sigma}{\partial \theta'} \right|_{\theta = \theta}$; and, finally,

$$V = \mathrm{diag} \left\{ \frac{n_1}{n} V^{(1)}, \ldots, \frac{n^{(G)}}{n} V^{(G)} \right\}$$

with

$$(4) \qquad V^{(g)} = \frac{1}{2} D' \left(\Sigma^{(g)-1} \otimes \Sigma^{(g)-1} \right) D.$$

112

where D is the so called «duplication» matrix of Magnus and Neudecker (1991) and «vech» is the vectorization operator that suppresses the redundant elements due to the symmetry[5]. Under this set-up, the general expression for the variance matrix of estimates is

$$(5) \qquad \operatorname{avar}(\hat{\vartheta}) = \frac{1}{n} J^{-1} R' V \Gamma V R J^{-1},$$

where $J = R'VR$ and Γ is the asymptotic variance matrix of s. The above variance matrix can be estimated substituting V, R and Γ for corresponding consistent estimates. A consistent estimate $\hat{V}$ of V is obtained by substituting in (4) $S^{(g)}$ for $\Sigma^{(g)}$; a consistent estimate $\hat{R}$ of R is obtained by evaluating R at the estimated value $\hat{\theta}$. Finally, an estimate of Γ that is consistent and unbiased under general distribution conditions is

$$\hat{\Gamma} = \operatorname{diag}\left\{ \frac{n}{n^{(1)}} \hat{\Gamma}^{(1)}, \dots, \frac{n}{n^{(G)}} \hat{\Gamma}^{(G)} \right\}$$

where

$$\hat{\Gamma}^{(g)} = \frac{1}{n^{(g)} - 1} \sum_{i=1}^{n^{(g)}} h_i^{(g)} h_i^{(g)}$$

with $h_i^{(g)} = \operatorname{vech}(z_i^{(g)} - s^{(g)})(z_i^{(g)} - s^{(g)})'$ and $s^{(g)} = \operatorname{vech} S^{(g)}$. Under normality of the $z_i^{(g)}$'s, the expression of Γ is such that the estimates' asymptotic variance matrix simplifies to

$$(6) \qquad \operatorname{avar}(\hat{\vartheta}) = \frac{1}{n} J^{-1},$$

an expression which we call the normal theory (NT) form of the variance matrix of estimates. See Satorra (1993) for full details on the derivations of the above results.

The test statistic for the goodness-of-fit of the model is obtained as n times the minimum of the fitting function, i.e. $T = n F_{GLS}(s, \hat{\sigma})$. When the model is true and the distribution assumptions are met, then it can be shown that T is a chi-square statistic of degrees of freedom, where r is the number of independent restrictions implied by the model on the moment matrices. Under general distributional assumptions, a scaled version of this statistic that is approximately chi-square distributed despite non-normality has been developed (Satorra and Bentler, 1994). The scaled statistic is defined as $\overline{T} = c^{-1} T$ where

$$c = r^{-1} \operatorname{tr}\left\{ \left(\hat{V} - \hat{V}\hat{V}\hat{J}^{-1}\hat{R}'\hat{V} \right) \hat{\Gamma} \right\}$$

---

[5]For a symmetric matrix A, $\operatorname{vec} A = D \operatorname{vech} A$ where D is the so-called duplication matrix and «vec» denotes vectorization of a matrix (see Magnus and Neudecker, 1991).

where r is the degrees of freedom of the goodness-of-fit test. The test of specific set of restrictions is carried out using the difference of chi-square goodness-of-fit test. The corresponding version of the Satorra-Bentler scaled chi-square tests applied also to the difference test statistic (see Satorra and Bentler, 1999), and this scaled statistic is the one reported in tables 1 to 4 of the paper.

To cope with variables that show an infrequent purchase pattern (in our paper, durable and medical expenditures), we introduced modifications on the sample matrices to be analyzed. When this occurs, we assume that the observed values of the variable are the result of censoring an underlying normal variable. In this case we modify the matrices $S^{(g)}$ used accordingly. In a first stage of the analysis, the matrices $S^{(g)}$ are computed as consistent estimates of the moment matrix involving the underlying uncensored variables. The PRELIS computer software of (Jöreskog and Sörbom, 1997) produces the modified matrices $S^{(g)}$, with the corresponding modification of the estimate $\hat{\Gamma}$ of Γ. Once we have the new matrices $S^{(g)}$'s and the new estimate $\hat{\Gamma}$, the analysis proceeds using the minimum-distance approach described above.

7. APPENDIX 2: PROGRAM CODE

In this appendix we reproduce PRELIS and LISREL code used in this paper.

PRELIS Code:

```
DA NI=7 NO=2600 MI= -999999 TR=LI
LA
GROUP NMEMB DURABLE TMY_1 TNMY_1 TMY_2 TNMY_2
RA=C:\WINDOWS\TEMP~LS3963.TMP FO; (8F15.6)
OR GROUP
CO NMEMB
CB DURABLE
CO TMY_1
CO TNMY_1
CO TMY_2
CO TNMY_2
CL GROUP 1 = YOUN 2 = CHIL 3 = TEEN 4 = ADUL 5 = OLD
SC 1 =1
OU MA=AM
```

LISREL Code:

MULTIGROUP ANALYSIS. EQUALITY CONSTRAINTS ON THE REGRES-
SION
COEFICIENTS OF
THE TWO FACTORS UNDERLIYING INCOME IN THE PRODUCT
EQUATION.

```
TI HETEROGENEITY EFFECTS ON TRANSPORTATION-GROUP 1
DA NI=7 NO=278 NG=5 MA=CM
LA
NMEMB TRANSPOR TMY_1 TNMY_1 TMY_2 TNMY_2 CONSTANT
CM FI=C:\DATA\LISREL~1\COVARI~1\GROUP1.AM SY
AC FI=C:\DATA\LISREL~1\COVARI~1\GROUP1.ACM
SE
1 2 3 4 5 6 7/
MO NY=7 NE=5 BE=FU,FI PS=SY,FR TE=DI
LE
NMEMB TRANSPOR F1 F2 CONSTANT
FI TE(1,1), TE(2,2),TE(7,7)
FI PS(2,1), PS(3,2), PS(4,2), PS(5,1), PS(5,2), PS(5,3),PS(5,4), PS(5,5)
FR BE(2,1), BE(1,5), BE(2,3), BE(2,4), BE(2,5)
FR LY(3,5), LY(4,5), LY(5,3), LY(5,5), LY(6,4), LY(6,5)
VA 1 LY(1,1) LY(2,2) LY(3,3) LY(4,4) LY(7,5) PS(5,5)
ST 1.0 ALL
PD
OU ME=ML IT=550 AD=OFF SE TV

TI HETEROGENEITY EFFECTS ON TRANSPORTATION-GROUP 2
DA NI=7 NO=896 MA=CM
LA
NMEMB TRANSPOR TMY_1 TNMY_1 TMY_2 TNMY_2 CONSTANT
CM FI=C:\DATA\LISREL~1\COVARI~1\GROUP2.AM SY
AC FI=C:\DATA\LISREL~1\COVARI~1 \GROUP2.ACM
SE
1 2 3 4 5 6 7/
MO
FI TE(1,1), TE(2,2), TE(7,7)
FI PS(2,1), PS(3,2), PS(4,2), PS(5,1), PS(5,2), PS(5,3),PS(5,4), PS(5,5)
FR BE(2,1), BE(1,5), BE(2,3), BE(2,4), BE(2,5)
FR LY(3,5), LY(4,5), LY(5,3), LY(5,5), LY(6,4), LY(6,5)
VA 1 LY(1,1) LY(2,2) LY(3,3) LY(4,4) LY(7,5) PS(5,5)
ST 2.0 ALL
```

```
EQ BE 1 2 3 BE 2 3
EQ BE 1 2 4 BE 2 4
PD
OU IT=550 SE TV AD=OFF

TI HETEROGENEITY EFFECTS ON TRANSPORTATION-GROUP 3
DA NI=7 NO=586 MA=CM
LA
NMEMB TRANSPOR TMY_1 TNMY_1 TMY_2 TNMY_2 CONSTANT
CM FI=C:\DATA\LISREL~1\COVARI~1\GROUP3.AM SY
AC FI=C:\DATA\LISREL~1\COVARI~1\GROUP3.ACM
SE
1 2 3 4 5 6 7/
MO
FI TE(1,1), TE(2,2),TE(7,7)
FI PS(2,1), PS(3,2), PS(4,2), PS(5,1), PS(5,2), PS(5,3),PS(5,4), PS(5,5)
FR BE(2,1), BE(1,5), BE(2,3), BE(2,4), BE(2,5)
FR LY(3,5), LY(4,5), LY(5,3), LY(5,5), LY(6,4), LY(6,5)
VA 1 LY(1,1) LY(2,2) LY(3,3) LY(4,4) LY(7,5) PS(5,5)
ST 3.0 ALL
EQ BE 1 2 3 BE 2 3
EQ BE 1 2 4 BE 2 4
PD
OU IT=550 SE TV AD=OFF

TI HETEROGENEITY EFFECTS ON TRANSPORTATION-GROUP 4
DA NI=7 NO=380 MA=CM
LA
NMEMB TRANSPOR TMY_1 TNMY_1 TMY_2 TNMY_2 CONSTANT
CM FI=C:\DATA\LISREL~1\COVARI~1\GROUP4.AM SY
AC FI=C:\DATA\LISREL~1\COVARI~1\GROUP4.ACM
SE
1 2 3 4 5 6 7/
MO
FI TE(1,1), TE(2,2),TE(7,7)
FI PS(2,1), PS(3,2), PS(4,2), PS(5,1), PS(5,2), PS(5,3),PS(5,4), PS(5,5)
FR BE(2,1), BE(1,5), BE(2,3), BE(2,4), BE(2,5)
FR LY(3,5), LY(4,5), LY(5,3), LY(5,5), LY(6,4), LY(6,5)
VA 1 LY(1,1) LY(2,2) LY(3,3) LY(4,4) LY(7,5) PS(5,5)
ST 1.0 ALL
EQ BE 1 2 3 BE 2 3
EQ BE 1 2 4 BE 2 4
PD
```

OU IT=550 SE TV AD=OFF

TI HETEROGENEITY EFFECTS ON TRANSPORTATION-GROUP 5
DA NI=7 NO=426 MA=CM
LA
NMEMB TRANSPOR TMY_1 TNMY_1 TMY_2 TNMY_2 CONSTANT
CM FI=C:\DATA\LISREL~1\COVARI~1\GROUP5.AM SY
AC FI=C:\DATA\LISREL~1\COVARI~1\GROUP5.ACM
SE
1 2 3 4 5 6 7/
MO
FI TE(1,1), TE(2,2),TE(7,7)
FI PS(2,1), PS(3,2), PS(4,2), PS(5,1), PS(5,2), PS(5,3),PS(5,4), PS(5,5)
FR BE(2,1), BE(1,5), BE(2,3), BE(2,4), BE(2,5)
FR LY(3,5), LY(4,5), LY(5,3), LY(5,5), LY(6,4), LY(6,5)
VA 1 LY(1,1) LY(2,2) LY(3,3) LY(4,4) LY(7,5) PS(5,5)
ST 1.5 ALL
EQ BE 1 2 3 BE 2 3
EQ BE 1 2 4 BE 2 4
PD
OU IT=650 SE TV AD=OFF