DOCUMENT DE TREBALL

XREAP2007-04

CROSS-SECTION DATA, DISEQUILIBRIUM SITUATIONS
AND ESTIMATED COEFFICIENTS: EVIDENCE
FROM CAR OWNERSHIP DEMAND

Anna Matas; Josep-Lluis Raymond

XARXA
DE REFERÈNCIA
EN ECONOMIA APLICADA

Cross-section data, disequilibrium situations and estimated coefficients: evidence from car ownership demand

Anna Matas*

Departament d'Economia Aplicada, Universitat Autònoma de Barcelona,

08193 Bellaterra, Barcelona, Spain

anna.matas@uab.es

Josep-LLuis Raymond

Departament d'Economia i Història Econòmica, Universitat Autònoma de Barcelona,

08193 Bellaterra, Barcelona, Spain

josep.raymond@uab.es

*Corresponding author. Tel.: +34.93.581.1578; Fax: +34.93.581.2292

Abstract

The objective of this paper is to analyse to what extent the use of cross-section data will distort the estimated elasticities for car ownership demand when the observed variables do not correspond to a state equilibrium for some individuals in the sample. Our proposal consists of approximating the equilibrium values of the observed variables by constructing a pseudo-panel data set which entails averaging individuals observed at different points of time into cohorts. The results show that individual and aggregate data lead to almost the same value for income elasticity, whereas with respect to working adult elasticity the similarity is less pronounced.

1. Introduction

The use of disaggregate data to estimate transport demand models has been common practice for the last three decades. The specification of demand models at the decision-making unit has proven to be the appropriate level on which to base transport demand theory.

However, the use of cross-section data has been questioned on the grounds that individuals do not instantaneously adjust their behaviour to changes in the explanatory variables and, in consequence, the observed situation at any point in time will not correspond to a state equilibrium for some individuals in the sample[1]. When the unobserved disequilibrium factors are correlated with the explanatory variables in the equation the estimated coefficients will be inconsistent and, hence, the cross-section relationship will not be appropriate to approximate the behavioural response to changes in the explanatory variables.

According to the available literature (Goodwin, 1990), the lack of instantaneous adjustment may be due to several factors, such as incomplete information, searching costs, and various constraints that prevent immediate response. The existence of such factors may result in habit effects and lead to asymmetry in response or hysteresis.

Car ownership is one of the markets in which evidence has been provided both about lags in adjustment to changes in the contributing factors and asymmetrical response patterns[2]. The purpose of this paper is to analyse to what extent the use of cross-section data will distort the estimated elasticities in this specific context.

As is well known, disequilibrium situations can be modelled by estimating a dynamic model using panel data at individual level. Nevertheless, even ignoring the

---

[1] See, for instance, Goodwin et al (1990), Kitamura (1990), Kitamura and Bunch (1990) and Kitamura (2000).
[2] See, for instance, Dargay (2001), Dargay and Hanly (2004) and Goodwin (1997).

econometric problems related to the estimation of dynamic, disaggregate models[3], cross-section data is very frequently the only available information.

Our proposal consists of approximating the equilibrium values of the observed variables by constructing a pseudo-panel data set which entails averaging individuals observed at different points of time into cohorts according, among other variables, to the year of birth of the individual. The underlying assumption is that by averaging across individuals the disequilibrium situations will tend to be cancelled out and the estimated coefficients will represent long term behaviour better than cross-section individual data. In our view, this assumption can be sustained for car ownership demand. The main contributing factors to car ownership are: household income, the size and structure of the household, the working positions of its members, the cost of car ownership, the residential location and the quality of public transport. In the real world, these variables change continuously and given that there is a lag in response, the observed number of cars per household may differ from that desired by some families. However, given the characteristics of the contributing factors it may well be that changes are not in the same direction for all households. If this is so, averaging individual observations will tend to reduce the disequilibrium values in the observed variables.

The results of the pseudo-panel model will be compared with those of the cross-section analysis in terms of elasticities in order to assess the accuracy of cross-section data for policy analysis.

Averaging observations, however, poses its own problem. Specifically, the variability in regressors is reduced, and in turn, in the limit, may lead to severe multicollinearity. In order to solve this problem, our approach consisted of grouping individuals into cohorts defined on the basis of common shared characteristics. Under

---

[3] Among them, the limited variability in regressors for the sampled individuals when the temporal dimension is small.

the assumption that individuals in the same cohort have common characteristics, it is possible to treat the average for the cohorts as individual observations. Our proposal relies on the availability of repeated cross-section data at different points of time, which is far more frequent than a true panel data set[4].

2. The data

The study relies on data from the Spanish Household Surveys (EPF) for 1980, 1990 and 2000, with sample sizes of 23,696, 20,927 and 28,963 observations respectively. The dependent variable is the number of cars per household which has been specified according to four alternatives: zero, one, two, and three or more cars.

The explanatory variables include socio-economic, demographic and residential location variables. With respect to the former the variables included in the equation are: household income (proxied by annual household expenditure), number of working adults and the sex of the head of the family. Several studies have proven that car ownership is also influenced by a generation effect[5]. The estimation of a car ownership equation for three different years makes it possible to test the existence of a generation effect, over and above growth in income and changes in socio-economic variables, by grouping observations in accordance with the date of birth of the head of household. Initially, we formed 8 cohorts by grouping individuals born in the same decade. Nevertheless, the results showed that the estimated coefficients were not statistically different after the generation born in the forties. In the final specification we differentiated between three groups: those born before 1930, those born in the thirties, and the rest of the sample.

---

[4] The use of pseudo panels has been common in car ownership demand. For instance, Madré (1990), Dargay and Vythoulkas (1999) and Dargay (2002).
[5] For instance, Madré (1990) and Madré and Pirotte (1997).

The effect of residential location was captured through two variables: municipality size and region of residence. We divided municipalities into four categories: very large (those with population over one million), large (those with populations between one million and half a million), medium (those with between 10,000 and 500,000 inhabitants), and small (those with less than 10,000 inhabitants). The size of the municipality can be seen as a proxy for a range of variables affecting car ownership. For instance, different access to public transport or spatial distribution of activities. Secondly, the data showed that there was an additional effect depending on the region of residence.

The estimation of a car ownership equation with a cross-section sample when all households are faced by the same prices makes it impossible to include a price variable. However, given that we have a sample at three different points in time the effect of price is captured by the constant term in the equations and its variation over time[6].

The pseudo panel was formed by grouping households into cohorts according to three variables: year of birth of the head of the household, municipality size and region of residence. In order to guarantee that the number of observations in each cohort was high enough, households were grouped by periods of ten years. The same cohorts were defined for each year in the sample.

3. Estimated models

The selected specification to estimate the car ownership demand equation with individual data was the ordered probit model, whereas the pseudo panel equation was estimated according to a tobit model to avoid predictions lower than zero. The same

---

[6] For a more detailed description of the data and variables of the study see Matas and Raymond (2005).

explanatory variables were included in both models. Given that the estimated coefficients obtained in an ordered probit and tobit model are not directly comparable, the comparison was made in terms of elasticities for both formulations.

Table 1 shows the results for the ordered probit equation estimated with a sample of 73,586 observations. The dependent variable takes four values: zero, one, two or three or more cars. For each of the dummy variables one category has to be excluded from the equation in order to avoid perfect multicollinearity. The excluded categories, and hence the reference categories, are: small municipalities, year 1980, the cohort corresponding to those born before 1930 and the region of Andalusia. As can be observed in Table 1, all of the variables take the expected sign and are, in general, highly significant.

With respect to the pseudo panel, the number of observations after grouping households was 851. The dependent variable is now a continuous variable, which always takes positive values. In the tobit model the starting point is:

$$Y_i = X_i' \cdot \beta + \varepsilon_i \quad if : X_i' \cdot \beta + \varepsilon_i \geq 0$$
$$Y_i = 0 \qquad\qquad if : X_i' \cdot \beta + \varepsilon_i < 0$$

After averaging observations:

$$\overline{Y}_j = \overline{X}_j' \cdot \beta + \overline{\varepsilon}_j \quad if : \overline{X}_j' \cdot \beta + \overline{\varepsilon}_j \geq 0$$
$$\overline{Y}_j = 0 \qquad\qquad if : \overline{X}_j' \cdot \beta + \overline{\varepsilon}_j < 0$$

In order to ensure homoskedasticity in the random disturbances, all variables are weighted by the square root of the number of observations in the respective cohort.

Again, as shown in Table 2, the coefficients take the expected sign and are significantly different from zero.

4. Elasticities

In order to compare the results of the two formulations used we computed the elasticity of the expected number of cars with respect to two variables: total expenditure, as a proxy for permanent income, and the number of working adults in the household. An interesting point in selecting these two variables is that whereas observed expenditure can be considered a good proxy for household permanent income, the number of working adults in the household can reflect situations of transitory disequilibrium depending on the business cycle. For instance, in periods of recession households can face a situation of transitory unemployment.

Elasticity values correspond to an average of individual elasticities for the whole sample and are computed for a unit percent increase in the explanatory variables. As shown in Table 3, income-elasticities are almost identical for both types of data, whereas elasticity with respect to working adults is slightly higher in the pseudo panel formulation. It should also be noted that the results agree with those found in the literature.

In order to provide a more detailed comparison, we proceed by constructing a density function for each elasticity value. Let the starting point be the asymptotical distribution of the coefficient estimated in the corresponding model:

$$\hat{\beta} \rightarrow N(\beta, \sigma_{\beta}^2)$$

Given $\hat{\beta}$, it is possible to approximate a univocal function between the estimated elasticity and the estimated *beta* coefficient of the form:

$$Estimated\ elasticity = \delta + \gamma \cdot \hat{\beta}$$

Hence, the distribution of the elasticity will be:

Estimated elasticity $\rightarrow N(\delta + \gamma \cdot \hat{\beta}, \gamma^2 \sigma_{\hat{\beta}}^2)$

After computing the mean and standard deviation for each individual elasticity, it is possible to construct the density function by simulation. Figures 1 and 2 present the results for income and working-adult elasticities, respectively. When comparing the distributions we should note that averaging across individuals leads to a loss of efficiency given the reduction in the number of observations. On the contrary, as long as averaged observations are a better approximation for equilibrium values, a source of inconsistency in the estimation is reduced.

5. Conclusions

The empirical results show that in our study individual and aggregate data lead to almost the same value for income elasticity. With respect to working-adult elasticity, the similarity is less pronounced. This outcome can be at least partially explained by the fact that in a sample of cross-section individuals disequilibrium errors are probably higher for the number of working adults in a family than for total expenditure, as stated above. Effectively, if the problem of disequilibrium values is more severe for observed working adults than for total expenditure, the potential inconsistency problem will probably be higher for working adult elasticity than for income elasticity. Nevertheless, this explanation should only be considered a suggestion given that if regressors are not orthogonal, the inconsistency in one of the estimated coefficients will partially affect the consistency of the remaining coefficients.

If averaging individual observations into cohorts with similar characteristics does effectively reduce the effect of disequilibrium in individual values, the results

show that in our sample these potential disequilibria do not substantially affect the estimated elasticities. This evidence also agrees with Dargay (2002) in the context of car ownership demand. A possible explanation might be that in our case the disequilibria that affect cross-section data are of little importance in comparison to the variability observed in the explanatory variables across individuals. However, this does not generally need to be the case. Therefore, in order to reduce the effects of disequilibrium values in individual observations, our proposal can lead to a reasonable approximation to the relationship between variables when panel data are not available. As long as the equation estimated using grouped data provides a reliable approximation to the relationship between equilibrium values, our approach can be considered a way to approximate the long term relationship between variables.

6. References

Dargay, J.M. (2001): "The effect of income on car ownership: evidence of asymmetry", *Transportation Research Part A*, 35, 807-821.

Dargay, J.M. and P.C. Vythoulkas (1999): "Estimation of a dynamic car ownership model", *Journal of Transport Economics and Policy*, 33, 287-302.

Dargay, J.M. (2002): "Determinants of car ownership in rural an urban areas: a pseudo-panel analysis", *Transportation Research Part E*, 38, 351-366.

Dargay, J. M., and M. Hanly (2004): "Volatility of car ownership, commuting and mode and time in the UK", *Proceedings of the World Conference on Transport Research*, Istanbul.

Goodwin, P.B., R. Kitamura and H. Meurs (1990): "Some principles of dynamic analysis of travel demand", in Jones, P. (ed.) *Developments in dynamic and activity-based approaches to travel analysis*, Oxford Studies in Transport, Gower Publishing Company, Aldershot.

Goodwin, P.B. (1997): "Have panel surveys told us anything new?", in Golob, T.F., R. Kitamura and L. Long (eds.) *Panels for transportation planning*, Kluwer Academic Publishers, Boston.

Kitamura, R. (1990): "Panel analysis in transportation planning: an overview", *Transportation Research Part A*, 24, 401-415.

Kitamura, R. and D.S. Bunch (1990): "Heterogeneity and state dependence in household car ownership: a panel analysis using ordered-response models with error components", *Transportation and Traffic Theory*, in Koshi, M. (ed.) Elsevier.

Kitamura, R. (2000): "Longitudinal methods", in Hensher, D.A. and K.J. Button (eds.) *Handbook of Transport Modelling*, Elsevier.

Madré, J.L. (1990): "Long-term forecasting of car ownership and car use", *Developments in dynamic and activity-based approaches to travel analysis* in Jones, P. (ed.), Oxford Studies in Transport, Aldershot, Avebury.

Madré, J.L and A. Pirotte (1997): "Regionalisation of car-fleet and traffic forecast", *Understanding travel behaviour in an era of change*, in Stopher, P. and M. Lee-Gosselin (eds.), Pergamon, Oxford.

Matas, A. and J.L. Raymond (2005): "Economic development and changes in car ownership patterns", *European Transport Conference* 2005, Strasbourg.

Table 1. Estimation results of the ordered probit model

	Coefficient	Std. Error	z-Statistic
Ln(total expenditure)	1.091296	0.00989	110.338
Working adults	0.30087	0.005876	51.20384
Sex (men=1)	0.573185	0.014272	40.16229
Very large municipalities	-0.415007	0.027663	-15.00207
Large municipalities	-0.250281	0.02308	-10.84413
Medium municipalities	-0.124665	0.011428	-10.90828
Dummy 1990	0.283598	0.012503	22.68166
Dummy 2000	0.491549	0.012091	40.65266
Cohort 1930-1939	0.357539	0.01373	26.04027
Cohort 1940-1980	0.516534	0.012368	41.76399
Aragon	0.172403	0.024391	7.068282
Asturias	0.042384	0.02823	1.501378
Baleares	0.455118	0.029964	15.18891
Canarias	0.16504	0.025896	6.373054
Cantabria	-0.007398	0.033834	-0.218642
Castilla y León	0.11079	0.018703	5.923602
Castilla la Mancha	0.085612	0.022343	3.831709
Cataluña	0.216913	0.020054	10.81637
Valencia	0.364779	0.020029	18.21293
Extremadura	0.110479	0.027847	3.967338
Galicia	0.105488	0.02103	5.016163
Madrid	0.029486	0.028378	1.039074
Murcia	0.165161	0.030288	5.453012
Navarra	0.114233	0.034574	3.303974
País Vasco	-0.100706	0.023074	-4.364495
La Rioja	0.078129	0.035032	2.230206
Limit 1	11.7703	0.095301	123.5072
Limit 2	14.03372	0.099319	141.2999
Limit 3	15.40391	0.101799	151.3163
Observations	73586		
Log likelihood	-51641.23		
Schwarz criterion	1.407978		
Pseudo-R2	0.291184		

Table 2. Estimation results of the tobit model

	Coefficient	Std. Error	z-Statistic
Constant term	-4.263321	0.284123	-15.00519
Ln (total expenditure)	0.439019	0.031142	14.09715
Working adults	0.146005	0.018207	8.019133
Sex (men=1)	0.252662	0.059162	4.270692
Very large municipalities	-0.179583	0.026901	-6.675773
Large municipalities	-0.099911	0.023965	-4.169097
Medium municipalities	-0.051223	0.012126	-4.224348
Dummy 1990	0.115968	0.010031	11.56111
Dummy 2000	0.219038	0.011331	19.33067
Cohort 1930-1939	0.098162	0.013232	7.418295
Cohort 1940-1980	0.179234	0.012227	14.65942
Aragon	0.071212	0.021246	3.351769
Asturias	0.019883	0.01634	1.216829
Baleares	0.19448	0.025394	7.658483
Canarias	0.058611	0.017423	3.363983
Cantabria	-0.000216	0.0257	-0.0084
Castilla y Leon	0.045011	0.017069	2.637008
Castilla la Mancha	0.038871	0.017409	2.232878
Cataluña	0.096953	0.017983	5.391237
Valencia	0.153445	0.017368	8.834931
Extremadura	0.04616	0.020428	2.259679
Galicia	0.032266	0.017033	1.894395
Madrid	0.018044	0.02166	0.833075
Murcia	0.060626	0.024322	2.492621
Navarra	0.059524	0.030079	1.978927
País Vasco	-0.033833	0.016687	-2.027545
La Rioja	0.039223	0.024972	1.570648
Scale parameter	0.901933	0.028078	32.12207
Observations	851		
Left censored at zero	58		
Uncensored	793		
Log likelihood	-1091.116		
Schwarz criterion	2.786289		

Table 3. Estimated car ownership elasticities

	Ordered Probit	Tobit
Permanent income	0.5648	0.5632
Working adults	0.1744	0.1988

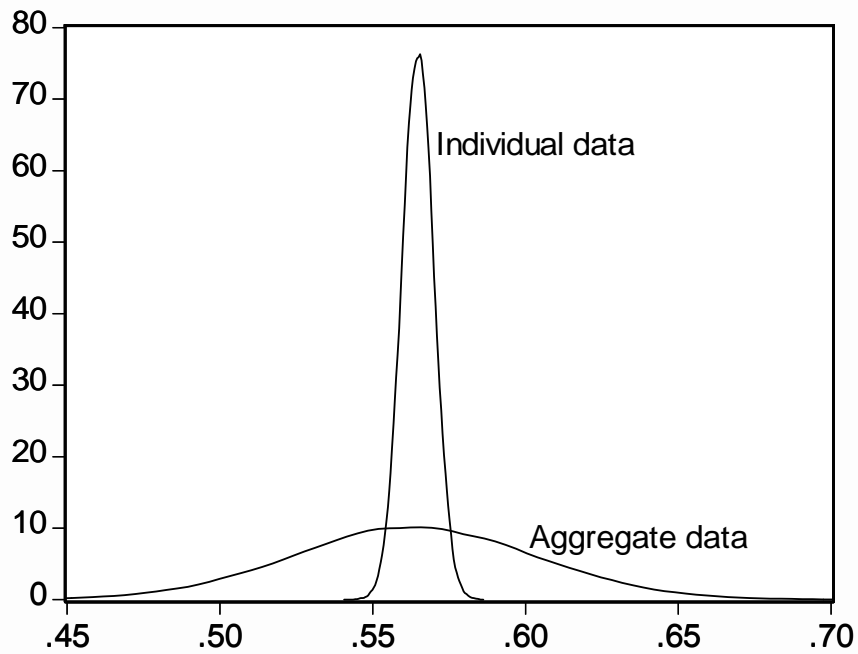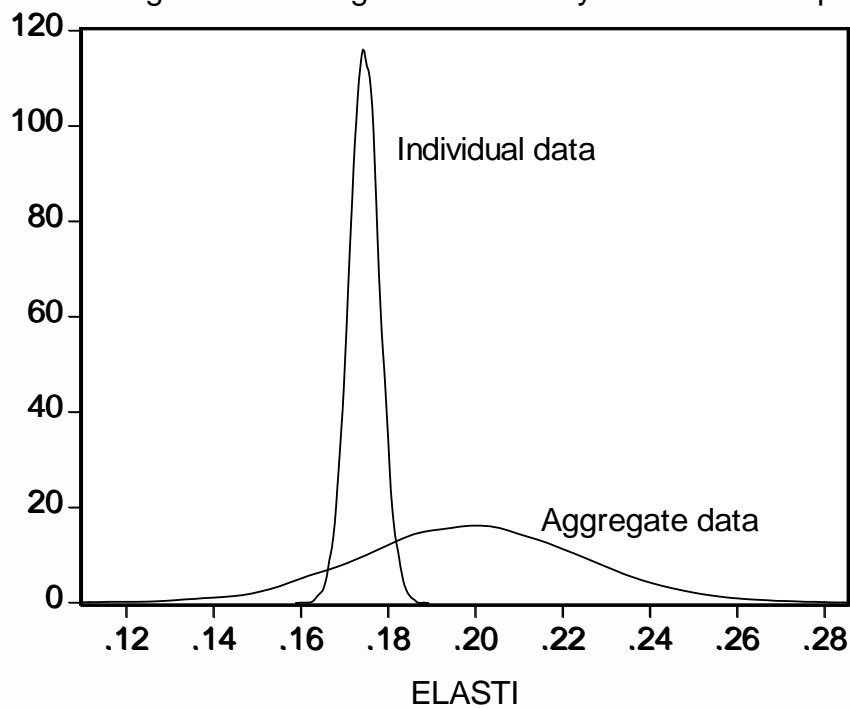## Figure 1. Income elasticity of car ownership

Figure 2. Working-adults elasticity of car ownership

CREAP2006-01
Matas, A. (GEAP)**; Raymond, J.Ll.** (GEAP)
"Economic development and changes in car ownership patterns"
(Juny 2006)

CREAP2006-02
Trillas, F. (IEB)**; Montolio, D.** (IEB)**; Duch, N.** (IEB)
"Productive efficiency and regulatory reform: The case of Vehicle Inspection Services"
(Setembre 2006)

CREAP2006-03
Bel, G. (PPRE-IREA)**; Fageda, X.** (PPRE-IREA)
"Factors explaining local privatization: A meta-regression analysis"
(Octubre 2006)

CREAP2006-04
Fernàndez-Villadangos, L. (PPRE-IREA)
"Are two-part tariffs efficient when consumers plan ahead?: An empirical study"
(Octubre 2006)

CREAP2006-05
Artís, M. (AQR-IREA)**; Ramos, R.** (AQR-IREA)**; Suriñach, J.** (AQR-IREA)
"Job losses, outsourcing and relocation: Empirical evidence using microdata"
(Octubre 2006)

CREAP2006-06
Alcañiz, M. (RISC-IREA)**; Costa, A.; Guillén, M.** (RISC-IREA)**; Luna, C.; Rovira, C.**
"Calculation of the variance in surveys of the economic climate"
(Novembre 2006)

CREAP2006-07
Albalate, D. (PPRE-IREA)
"Lowering blood alcohol content levels to save lives: The European Experience"
(Desembre 2006)

CREAP2006-08
Garrido, A. (IEB)**; Arqué, P.** (IEB)
"The choice of banking firm: Are the interest rate a significant criteria?"
(Desembre 2006)

CREAP2006-09
Segarra, A. (GRIT)**; Teruel-Carrizosa, M.** (GRIT)
"Productivity growth and competition in spanish manufacturing firms:
What has happened in recent years?"
(Desembre 2006)

CREAP2006-10
Andonova, V.; Díaz-Serrano, Luis. (CREB)
"Political institutions and the development of telecommunications"
(Desembre 2006)

CREAP2006-11
Raymond, J.L.(GEAP)**; Roig, J.L..** (GEAP)
"Capital humano: un análisis comparativo Catalunya-España"
(Desembre 2006)

CREAP2006-12
Rodríguez, M.(CREB)**; Stoyanova, A.** (CREB)
"Changes in the demand for private medical insurance following a shift in tax incentives"
(Desembre 2006)

CREAP2006-13
Royuela, V. (AQR-IREA)**; Lambiri, D.**; **Biagi**, **B.**
"Economía urbana y calidad de vida. Una revisión del estado del conocimiento en España"
(Desembre 2006)

CREAP2006-14
Camarero, M.; Carrion-i-Silvestre, J.LL. (AQR-IREA).;**Tamarit**, **C.**
"New evidence of the real interest rate parity for OECD countries using panel unit root tests with breaks"
(Desembre 2006)

CREAP2006-15
Karanassou, M.; Sala, H. (GEAP).;**Snower** , **D. J.**
"The macroeconomics of the labor market: Three fundamental views"
(Desembre 2006)

2007

XREAP2007-01
Castany, L (AQR-IREA)**; López-Bazo, E.** (AQR-IREA)**.**;**Moreno** , **R.** (AQR-IREA)
"Decomposing differences in total factor productivity across firm size"
(Març 2007)

XREAP2007-02
Raymond, J. Ll. (GEAP)**; Roig, J. Ll.** (GEAP)
"Una propuesta de evaluación de las externalidades de capital humano en la empresa"
(Abril 2007)

XREAP2007-03
Durán, J. M. (IEB)**; Esteller, A.** (IEB)
"An empirical analysis of wealth taxation: Equity vs. Tax compliance"
 (Juny 2007)

XREAP2007-04
Matas, A. (GEAP)**; Raymond, J.Ll.** (GEAP)
"Cross-section data, disequilibrium situations and estimated coefficients: evidence from car ownership demand"
 (Juny 2007)