On Estimating Discrete Choice Models of Product Differentiation with Habits and Market Level Data

Lapo Filistrucchi
EUI
September 2004

Preliminary and Incomplete!

In many markets consumers behaviour is characterized by habits. Whereas recent empirical studies using consumer level data are able to identify the effect of habits show the bias that ignoring it can cause in the estimates of the elasticity of demand with respect to prices, and therefore on the identification of collusive behaviour, discrete choice models of product differentiation which use aggregate level data usually treat observations relating to the same market at different points in time as observations of different markets. Although this might not be problematic in the case of durable goods when consumers in the same market at two points in time are probably different, it is clearly not realistic in the non durable goods case. The econometric issue is relevant for economic policy too as these models are being used more and more often for market analysis in antitrust cases and usually with market level data, which are easier to recover than consumer level ones. This paper investigates the possibility to introduce a habit component in the structural model of utility at the basis of the discrete choice model of product differentiation.

Keywords: discrete choice models, product differentiation, habits, market level data

JEL Classification: C23, L40, C25

1-Introduction

Assessing market power in a market with differentiated goods requires, among other things, the measurement of the residual price elasticity of demand. In order to estimate the latter, discrete choice models of product differentiation are used more and more often by antitrust authorities around the world. As consumer level data are not always available, these models are often estimated on aggregate level data, in such a way as explained in Berry(1994) and Nevo (2000).

Given that the usual available micro panel derives from the observation of sales of many products in many markets but at few points in time, so far the literature has been treating observations relating to the same market at different points in time as observations of different markets. Although this might not appear problematic in the case of durable goods, such as cars[1], when consumers in the same market at two points in time are probably different, it is clearly not realistic in the case of non durable goods, such as cereals[2] or daily newspapers[3].

The issue appears relevant because recent empirical studies in marketing using consumer level data have been able to identify the effect of habits and show the bias that ignoring it can cause in the estimates of the elasticity of demand with respect to prices, although results on the direction of the bias are mixed and depend on the model used[4]. Yet there is no a priori reason to believe that if a bias exists in models which use consumer level data it also exists in models which use market level ones nor that, if it does exist, it necessary has the same direction in the two cases.

But of course whether there is or not a bias and, in case there is, what is its direction and magnitude is crucial to the assessment of market power. At the extreme, if when habits are relevant and are not taken into account the estimated price elasticity is higher then the true one, the observation of a high price may lead to the conclusion that there is collusion in the market; if instead the estimated price elasticity is lower than the true one, observing low prices may mislead into judging that there is competition.

In this paper I therefore investigate the possibility to introduce a habit component in the structural model of utility at the basis of the discrete choice model of product differentiation and to estimate it.

---

[1] As in Brenkers & Verboven (2002).

[2] As in Nevo (2001).

[3] As in Filistrucchi(2004), who recognise the issue, but propose a solution which is unsatisfactory as it is not consistent with the random utility model, as discussed below.

[4] See Erdem (1996), Roy et al (1996), Seetharaman(1999) and Che et al.(2004).

After a brief description in the next section of the simplest model, namely the logit model, in section 3, I discuss the issue of habits and the likely bias deriving from their omission. In section 4 I discuss known ways to deal with the issue of dynamics, and therefore habits, in the estimation of demand, which are however unsatisfactory in a structural approach, while in section 5 I investigate new ways to solve the problem. Section 6 concludes.

2- The model

In order to exemplify the issue at stake, I introduce the simplest discrete choice model of product differentiation, namely the logit model. The same problem arises however also with the nested logit and the mixed logit.

The starting assumption, which is common to fixed coefficients models of product differentiation in general, is the following functional form of consumer i indirect utility from consuming good j at time t I market m:

$$u_{ijtm} = \alpha(y_{itm} - p_{jtm}) + \bar{x}_{jtm}\bar{\beta} + \xi_{jtm} + \varepsilon_{ijtm}$$

where y_{itm} is the income of consumer i at time t, p_{jtm} is the price of good j at time t, $\bar{x}_{jtm}$ is a vector of observed characteristics, ξ_{jtm} is an unobserved (by the econometrician) characteristic, ε_{ijtm} is a mean-zero stochastic term, α is consumers marginal utility from income (and marginal disutility from price) and $\bar{\beta}$ is a vector of taste coefficients.

Such an indirect utility specification assumes a quasi-linear utility function, free therefore of wealth effects, which sounds plausible for many goods, although not for all. It also assumes that both observed and unobserved product characteristics are the same across all individuals in the market, and thus rules out for instance different prices for different consumers in the same market. Finally, the marginal utility from income and the taste parameters are assumed fixed across consumers and, as a result, consumers' heterogeneity enters only through the separable additive random shock ε_{ijtm}.

As consumers may decide not to consume any of the goods considered, an outside good is introduced[5], consuming which yields to consumer i at time t the indirect utility:

$$u_{iotm} = \alpha y_{itm} + \xi_{otm} + \varepsilon_{iotm} \ .$$

Since the outside good is a composite one, its price and its characteristics are not defined. The price of the outside good is then assumed to be equal to zero[6] and all the characteristics are assumed to be unobservable[7] . But as ξ_{ot} is not identified, the standard practice is to set it equal to 0, which, as the term αy_{it} eventually vanishes because it is common to all products, is equivalent to normalizing the mean utility from the outside good to zero[8].

Consumers mean utility, δ_{jtm} from consuming good j at time t in market m is instead given by:

$$\delta_{jtm} = E_i[u_{ijtm}] = \alpha(y_{itm} - p_{jtm}) + \bar{x}_{jtm}\bar{\beta} + \xi_{jtm}$$

Consumers are then assumed to purchase the good which gives them the highest utility and to be never indifferent between buying one or another good.[9]

For convenience, it also assumed that consumers do not choose more than one good, although this behaviour can sometimes be observed. This assumption is common to most empirical studies on differentiated products markets, the usual justification being that assuming otherwise is econometrically very cumbersome and the assumption is instead, at worst, a reasonable approximation. That's because multiple purchases, though by no means uncommon, are not a rule and in any case even if two products are bought together they are then often consumed at different times, so that

[5] In the absence of an outside good, the model would assume consumers to be forced to choose one of the inside goods. and demand would depend only on differences in prices. Therefore, we would assume that a general increase in prices would not decrease aggregate newspapers sales, which would be unfortunate when estimating the elasticity of demand with respect to prices, since in general, when price increases, some consumers buy something else but some other do not buy anything.

[6] In other words, the consumer is assumed to be choosing between buying one of the above newspapers or not buying it, not between buying one of the newspapers above or buying something else. The decision of whether to buy something is not simultaneous.

[7] Or equivalently, both the price and the characteristics are assumed to be unobservable and therefore included in ξ_{otm}.

[8] So that neither the market shares of the outside good nor those of the inside goods respond to changes in the price of the outside good, unless time fixed effects are used. Also, the price of the outside good does not respond to the prices of the inside goods. Nor the prices of the inside goods respond to the price of the outside good.

[9] That is it is assumed there are no ties.

the multiple purchase is just an organisational device. Furthermore, if the potential market size is defined large enough, we might also claim that observed multiple purchases by the same individual are the result not only of his choice but also of somebody else's decision. In particular, if potential market size is defined as total population instead of number of households, the observation that an individual buys two goods might, at least to a certain extent, reflect the fact that he is buying one good for himself and another for another member of his households who asked him to.

Decomposing $\xi_{jtm} = \xi_j + \xi_t + \xi_m + \eta_{jtm}$, with η_{jtm} a random shock independent of ε_{ijtm}, allows me to model product and time specific unobserved characteristics. If ξ_j were considered as an unknown parameter specific to each product j, this would lead to a fixed effects model. If instead ξ_j were assumed to be a random variable with mean μ_ξ and variance σ_ξ, then we would have a random effects model. Whenever the assumption of no correlation between the observed product characteristics and the unobserved product characteristics, which lie at the basis of the random effect estimation, do not appear plausible, it is better to use a fixed effect specification. This choice also allows to better estimate product differentiation, as in this type of models the product fixed effects are usually believed to capture also the vertical component[10]. The inclusion of time fixed effects is instead justified by the necessity to control for the change through time in the utility of the outside good. Given that it is by definition a composite good, there are many reasons why the latter may change in time. It might, for instance, be due to the appearance of new goods or to the change in the characteristics of the goods which are included in the composite outside good. But also changes in the average consumer taste may change the relative utility of the choice to buy the inside goods with respect to the outside one. As its utility is by construction normalised to zero, the absence of time fixed effects or some equivalent control might often raise questions of identification for the estimated coefficients.

We then assume that ε_{ijtm} is i.i.d. across consumers and products and that it is distributed according to a type I extreme value distribution. Assuming ε_{ijtm} to be i.i.d. across consumers rules out, in particular, the possibility that individual specific

[10] See Nevo(2001).

random shocks are correlated across products or equivalently only allows shocks to demand to be correlated across products if they are not individual specific.

All the assumptions above lead to an aggregate logit model[11]. In fact in this model, at a given point in time t, each individual i is defined by a vector of random shocks $\bar{\varepsilon}_{\text{itm}}$ = (ε_{i0tm} ε_{i1tm} ... ε_{iJtm}). As a result, the set of individuals who choose product j at time t is implicitly defined as

$$\text{B}_{\text{jtm}}\,(\text{X}_{\text{tm}}, \bar{p}_{tm,}\bar{\xi}_{tm}, \alpha, \bar{\beta}) = \{\bar{\varepsilon}_{\text{itm}} \mid \text{u}_{\text{ijtm}} \geq \text{u}_{\text{iktm}} \vee_{\text{k}\neq\text{j}}\}$$

where

$$\text{X}_{\text{tm}} = (\bar{\text{x}}_{1tm}, ..., \bar{x}_{Jtm}),\;\; \bar{p}_{tm} = (p_{1tm}, ..., p_{Jtm}) \;\text{ and }\; \bar{\xi}_{tm} = (\xi_{1\text{tm},...,}\xi_{Jtm})$$

The market share of product j at time t in weekday d is therefore given by:

$$\text{S}_{\text{jtm}}\,(\text{X}_{\text{tm}}, \bar{p}_{tm,}\bar{\xi}_{tm}, \alpha, \bar{\beta}) = \text{Prob}\{\text{u}_{\text{ijtm}} \geq \text{u}_{\text{iktm}} \vee_{\text{k}\neq\text{j}}\} = \int_{\text{B}_{\text{jtm}}} dP_{\varepsilon}$$

which leads to

$$s_{jtm} = \frac{\exp(\delta_{jtm})}{1 + \sum_{k\neq 0} \exp(\delta_{ktm})} \;\text{ for any good j}$$

and

$$s_{0tm} = \frac{1}{1 + \sum_{k\neq 0} \exp(\delta_{ktm})} \;\text{ for the outside option}^{12}.$$

It should be noted that the presence of an outside good with market share s_{0tm} means that observations of goods sales are not sufficient to calculate market shares. As a result it is necessary to introduce the concept of potential market size as distinct from the observed market size which would simply be the sum of the observed goods sales. Thus the definitions of market size and market shares are different from the ones commonly used. Potential market size can either be assumed or estimated by parameterising it as depending on some market level data (such as population) which

[11] As it is well-known to the empirical industrial organization literature, the use of an aggregate logit model to estimate demand places restrictive assumptions on own and cross price elasticities or equivalently on own and cross marginal effects of price. A discussion of these restrictive assumption, which can be overcome by using a nested or, even better, a mixed logit model, is outside the scope of this paper. For such a discussion see for instance Berry(1994), Berry et al. (1995) and Nevo(2000).

[12] Note that the term αy_{it} drops out as it is common to all options.

vary across time. In both cases it is useful to test and discuss the robustness of the analysis with respect to its definition.

The own and cross marginal effects of price on market shares are:

$$\frac{\partial s_{jtm}}{\partial p_{jtm}} = -\alpha(1 - s_{jtm}) s_{jtm}$$

and

$$\frac{\partial s_{jtm}}{\partial p_{ktm}} = \alpha s_{ktm} s_{jtm} \text{ with } k \neq j.$$

So that the own and cross price elasticities of the market shares above are respectively:

$$\eta_{jjtm} = \frac{\partial s_{jtm}}{\partial p_{jtm}} \frac{p_{jtm}}{s_{jtm}} = -\alpha p_{jtm}(1 - s_{jtm})$$

and

$$\eta_{jktm} = \frac{\partial s_{jtm}}{\partial p_{ktm}} \frac{p_{ktm}}{s_{jtm}} = \alpha p_{ktm} s_{ktm} \text{ with } k \neq j.$$

The model thus predicts a different demand and therefore different marginal effects and elasticities for each product j, at each time t and in each market m.

Dividing each newspaper market share by the outside good market share, simplifying and taking natural logarithms leads to the following market shares estimation equation:

$$\delta_{jtm} \equiv \ln(s_{jtm}) - \ln(s_{otm}) = \bar{x}_{jtm}\bar{\beta} + \alpha p_{jtm} + \xi_j + \xi_m + \xi_t + \eta_{jtm}$$

Once potential total market size was defined, market shares s_{jtm} can be calculated dividing by it the number of sales of good j at time t in market m. The outside good market share can then be calculated as $s_{otd} = 1 - \sum_j s_{jtd}$.

Once the relevant parameter α has been estimated, the marginal effects and/or the elasticities above can be calculated. From these it is then possible to recover estimates of the mark-ups, for each product j at each time t and in each market m, under different hypothesis on the level and form of competition in these markets. For instance, if the choice variable is p, the benchmark case of Bertrand-Nash competition would imply a mark-up of

$$p_{jtm} - c_{jtm} = -\frac{s_{jtm}}{\frac{\partial s_{jtm}}{\partial p_{jtm}}} = \frac{1}{\alpha\left(1 - s_{jtm}\right)}$$

While the opposite benchmark of perfect collusion would imply

$$p_{jtm} - c_{jtm} = -\frac{s_{jtm}}{\frac{\partial s_{jtm}}{\partial p_{jtm}}} - \sum_{k \neq j} (p_{ktm} - c_{ktm}) \frac{\partial s_{ktm}}{\partial p_{jtm}} \Big/ \frac{\partial s_{jtm}}{\partial p_{jtm}} = \frac{1}{\alpha\left(1 - s_{jtm}\right)} + \sum_{k \neq j} (p_{ktm} - c_{ktm}) s_{ktm} \Big/ (1 - s_{jtm})$$

Substituting the estimated parameter α in the expressions above yields estimates of the mark-ups under Bertrand Nash competition and under perfect collusion. If a rough measure of the real mark-ups, such as the average mark-up in the industry in a given period, is available[13], it can be used to choose the most likely supply model, as discussed in Nevo (2001). Alternatively the demand equation above can be estimated simultaneously first with a pricing equation under Bertrand Nash competition, then with a pricing equation under perfect collusion. It is then possible to test statistically which of the two models of competition better fits the data, as in Ivaldi &Verboven(2003).

3- The bias

Unsurprisingly, estimating the equation above as it is might lead to substantial autocorrelation in the residuals, even if shocks to demand are not correlated through time. My claim is that assessment of market power on the basis of the model above might be strongly biased in markets where consumer habits are strong enough.

If habits are relevant, then there might be more than simple autocorrelation in the random shocks, namely an omitted variable problem. Assuming habits to be myopic, market shares today would not only be a function of the goods price, characteristics and random shocks today, but also of market shares yesterday. Similarly for the ratio of any inside good market share with respect to that of the outside good, which is the dependent variable in the above equation.

---

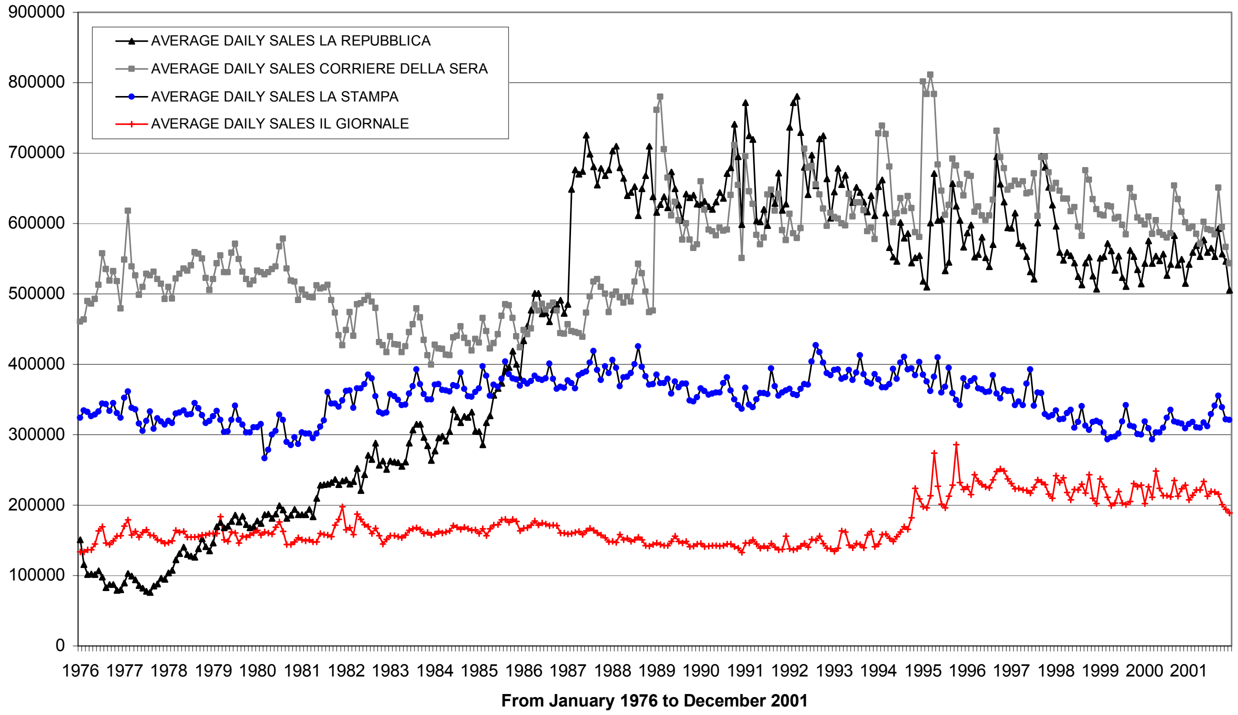
[13] Of course, the more disaggregated is the rough measure, the better it is, as a mark-up is calculated for each product j at each point in time t and in each market m.

Let us consider, for instance, the case of daily newspapers. It is well known in the business that readers tend to exhibit strong habits when it comes to choosing a daily newspaper.

Figure 1 below shows average daily sales in each month from 1976 to 2001 of the four main national dailies in Italy, namely Corriere della Sera, La Repubblica, La Stampa and Il Giornale. There is clearly a strong persistence in the data, although one important feature of daily newspapers characteristics, such as news themselves, vary though time and across products. Such a persistence might therefore be due to switching costs due to habits[14].

As shown in Figure 2, the persistence is clearly still there as we transform the data in such a way as to obtain the dependent variable in the estimation equation above.

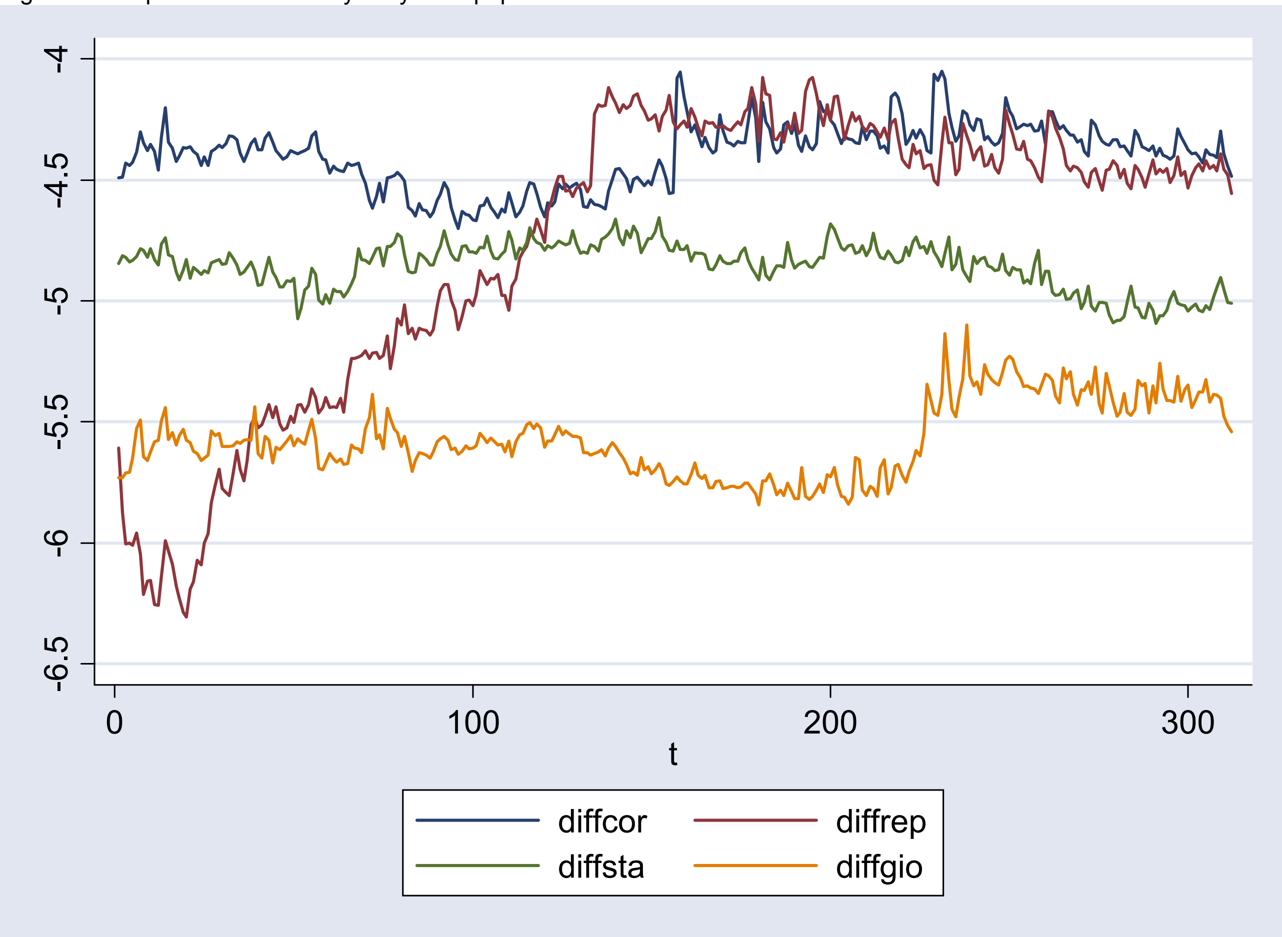
Figure 1 - Corriere della Sera, La Repubblica, La Stampa & Il Giornale on paper



---

[14] It should be noted that the data are sales at the newsstand not subscriptions, so that they are the result of a choice readers make everyday.

Figure 2 – Dependent variable by daily newspaper



4- One common solution

From a purely econometric point of view, all this would naturally lead to estimating the equation with the inclusion of one or more lags of the dependent variable, instead of or in addition to an autocorrelated random shock, which is the usual way to take dynamics, and therefore also habits, into account when estimating a structural model of demand which starts from the specification of an aggregate demand equation.

When a discrete choice model such the one described above is estimated on these data, as in Filistrucchi (2004), they persistence coefficient as measured by a lag of the dependent variable is a highly significant 0.882485 (with a standard error of 0.005072). The marginal disutility from price $-\alpha$ in the model with one lag is estimated to be as opposed -0.0000333 (with a standard error of 0.029) as opposed to -0.0000924 (with a standard error of 0.000037), so that the short-run elasticity is lower in the dynamic model than in the static one, whereas the long-run elasticity is higher.

Yet this specification raises a critical theoretical issue. That's because the inclusion of, for instance, one lag of the dependent variable is a necessary condition of the following specification of indirect consumers utility:

$u_{ijtm} = \alpha(y_{it} - p_{jtm}) + x_{jtm}\overline{\beta} + \xi_{jtm} + \ln(s_{jt-1m}) + \varepsilon_{ijtm}$ for each newspaper j

and

$u_{iotm} = \alpha y_{it} + \xi_{otm} + ln(s_{0t-1m}) + \varepsilon_{iotm}$ for the outside good, a specification which resembles more a way to model a (lagged) network effect than habits and the supposed habit component is not consumer specific.

Another possibility is the inclusion of one or more lags of the explanatory variables. If for instance one lag were introduced, the estimated equation would then appear to derive from the following specification of indirect consumers utility:

$u_{ijtm} = \alpha_0(y_{it} - p_{jtm}) + x_{jtm}\overline{\beta}_0 + \alpha_1(y_{it-1} - p_{jt-1m}) + x_{jt-1m}\overline{\beta}_1 + \xi_{jtm} + \varepsilon_{ijtm}$ for each good j

and

$u_{iot} = \alpha_0 y_{it} + \alpha_1 y_{it} + \xi_{otm} + \varepsilon_{iotm}$ for the outside good, a specification which is again unsatisfactory as the new term is not consumer specific.

5 - Other solutions

Yet to be fully developed.

6 - Conclusion

To be drawn.

References

Anderson Simon P., de Palma André & Thisse Jacques-François, Discrete Choice Theory of Product Differentiation, Cambridge (Massachusetts)-London (UK): The MIT Press, 1992.
Baltagi B. H., Econometric Analysis of Panel Data, New York: Wiley, 2001
Berry S., Estimating Discrete-choice Models of Product Differentiation, The *RAND Journal of Economics*, 1994, 25(2), 242-262.
Berry S., Levinsohn J. & Pakes A., Automobile Prices in Market Equilibrium, *Econometrica*, 1995, 63, 841-890
Brenkers Randy & Verboven Frank, Liberalizing a Distribution System: the European Car Market , mimeo, Catholic University of Leuven, September 2002

Bresnahan Timothy P., Empirical Studies of Industries with Market Power in Schmalensee R. & Willig R.D. Handbook of Industrial Organisation, vol.II, Elsevier, 1989.
Che Hai, Seetharaman P.B. & Sudhir K., Pricing Behaviour In Markets with State Dependence in Demand, mimeo, 2004.
Erdem Tülin, A Dynamic Analysis of Market Structure Based on Panel Data, *Marketing Science*, 1996, 15(4), 359-378.
Filistrucchi Lapo, The Impact of Internet on the Market for Daily Newspapers in Italy, mimeo, EUI, 2004.
Ivaldi Marc & Verboven Frank, Quantifying the Effects from Horizontal Mergers in European Competition Policy, mimeo, 2003.
McFadden Daniel & Train Kenneth, Mixed MNL Models for Discrete Response, mimeo, University of California at Berkeley, May 2000.
Nevo Aviv, A Practitioner's Guide to the Estimation of Random-Coefficients Logit Models of Demand, *Journal of Economics & Management Strategy*, 2000,9(4), 513-548.
Nevo Aviv, Measuring Market Power in the Ready-to-eat Cereal Industry, *Econometrica*, 69(2), March 2001, 307-342.
Nickell S., Biases in Dynamic Models with Fixed Effects, *Econometrica*, 49, 1981, 1417-1426.
Petrin Amil, Quantifying the Benefits of New Products: the Case of the Minivan, Journal of Political Economy,2002,110(4),705-729.
Petrin Amil & Train Kenneth, Omitted Product Attributes in Discrete Choice Models, mimeo, University of Chicago & University of California at Berkeley, December 2002.
Roy Rishin, Chintagunta Pradeep K. and Haldar Sudeep, *Marketing Science*, 1996, 15(3), 280-299.
Train Kenneth, Qualitative Choice Analysis, Cambridge (Massachusetts)-London(UK): The MIT Press,1993
Train Kenneth, Discrete choice methods with simulation **,** Cambridge(Massachussets); New York(NY) : Cambridge University Press, 2003