Empirical Project Report:
Demand Estimation for Antibiotics in France

Mai-Anh Dang, Hoang Tran, Xiahua Wang, Pu Zhao *

Project supervised by: Prof. Pierre Dubois

March 18, 2018

Abstract

In this empirical project, our purpose is to estimate the demand for antibiotics in France. France is one of the countries with highest rate of reimbursement for medicine. Therefore, we might expect the low elasticities of price. Meanwhile, there is negative externalities in consumption due to the antibiotic resistance. We use logit, nested logit and BLP for the estimation. The results are improved and more acceptable than the previous year, when we have standardized the units and augment the existing data set with publicly available data. In addition, we found an interesting fact that the coefficients of lagged animal consumption is always negative but not significant. Further improvement in data is necessary for more persuasive evidences.

1 Introduction

Antibiotics, which is a type of antimicrobial drugs used in treatment and prevention of bacterial infection, have huge contribution to public health. One crucial characteristic of antibiotics is the emergence of resistant pathogens. The overuse of antibiotics leads to the possibility of reduced efficacy. The antibiotics market in France is characterized by the heavy subsidies of health insurance and social security. The high-percentage reimbursement for antibiotics from health insurance and social security would exacerbate the problem of over-consumption, as consumers do not directly bear the cost. It might cause the serious problem of antibiotic resistance that the efficiency of infectious treatments would be reduced. Meanwhile, the restriction in antibiotic consumption would affect the incentives of producers to invest and innovate. Considering the social values created by antibiotics, both over-consumption and under-investment cause severe consequence for the society.

*M2 Econometrics & Empirical Economics, Toulouse School of Economics (2017-2018)

Under the high-percentage reimbursement, we would expect that the demand of antibiotics in France is inelastic to the price. This empirical project aims to examine this assumption by estimating the demand for antibiotics in France. We inherits the results and data from the previous year project [Ussinova et al., 2017], with the data set from the social security with quantity and reimbursed amount for the general regime from 2001-2013.

In this project, we update the data with observations from 2013-2015, and augmented with different publicly available data sources. To improve the quality of estimation, we standardize the quantity in boxes by standard units (dosage). To explore the presence of antibiotic resistance, we include the lagged consumed quantities in animal post-2011 in the model. The antibiotic resistance rates for each antibiotics are also involved to describe the individual characteristics of products. The popular demand estimation approach using the utility function of consumers rather than modeling product space is adopted. The logit, nested logit, and BLP model would be used.

In the next section, we will introduce the literature review and context of French antibiotics market. demand estimations models In **Section 3**, the demand estimation models would be discussed. Data and Results would be presented in **Section 4** and **Section 5**. Finally, the conclusions and suggestions for next year project would be provided in **Section 6**.

2 Context and Literature Review

2.1 Economics of Antibiotics

It is argued that there are at least two potential market failures in the antibiotics markets. First, the negative externalities of antibiotic resistance fails to be internalized by producers and consumers, which leads to the over-consumption than the optimal level. Second, the measures to restrict antibiotic consumption might reduce the incentives of innovations, which leads to the under-investment. Also, the measures to promote the innovations in this market, particularly through the patent system, may motivate producers to maximize the supply and marketing to boost the consumption during the period of patent protection [DiMasi et al., 2003], [DiMasi et al., 2010]. That would exacerbate the issue of antibiotic resistance.

The issue of antimicrobial resistance is considered as a "major European and global challenge", causally related to the death of 25,000 patients per year in the European Union and caused 1.5 billion euros of extra health care cost and productivity loss [1]. The increasing multi-drug resistant tuberculosis and the MRSA (Methicillin-resistant Staphylococcus Aureus) "superbug" is a serious concern for both developed and developing countries in recent years [2].

On the other hand, in recent year, only few new antibiotics are introduced into the market,

[1] According to EU Directorate-General Health and Food Safety: https://ec.europa.eu/health/amr/sites/amr/files/amr_factsheet_en.pdf

[2] In "Project Introduction: Economics of Antibiotics", from Prof. Pierre Dubois, October 2017.

signaling the status of under-investment. In a well-functioning market, price should signal social value of treatments and drugs, to motivate the efficient investment in research and development. However, despite the huge contribution to public health, the prices of antibiotics are relatively low. For illustration, the cost for non-hospitalized patient for community-acquired pneumonia with antibiotics is commonly around EUR 20-200 [Woodhead et al., 2011]. One should take into account the fact that these treatments would prolong life by years, comparing to anticancer drugs (typically cost EUR 20,000 or more for few months more of life. Meanwhile, due to the expected high demand for a large number of patients, health care system would prefer to limit the total expense by controlling the low price of antibiotics. These facts make the antibiotic market not attractive for producers to invest and innovate.

2.2 Medicine Market and Reimbursement in France

French health care system and drugs market is well documented in working papers and reports from the authorities. The price of drugs are fixed by the social security, taking into account features like: innovation, efficiency, fixed and marginal costs of drugs. It is found that the pharmaceutical market in France is insensitive to price. The government spends a significant amount to cover drug expenditure, for 2.11% of GDP [Grandfils et al., 2008]. The demand in this market has these important features:

- Low elasticity of demand to price, due to the high rate of reimbursement
- Preference for new and expensive drugs
- High rate of prescription and consumption

For drug pricing (*see:* [Garau and Mestre-Ferrandiz, 2006]), there are three main schemes.

1. Prices are determined by supply and demand, with completely no price regulations
2. Prices of reimbursable drugs are semi-controlled, determined by factors of: benefit, substitution, realized/predicted demand.
3. Hospital drugs with prices were not regulated before 2003.

The prices of new innovative medicines are proposed by company and approved by the Sickness Insurance. The prices of low-innovative medicines are determined by the comparison of substitutions by Improvement in the Rendered Medical Service (ASMR) criteria. Non-innovative medicine without any significant improvements in medical value may not be reimbursed. The price of generic drugs must be priced at least 40% lower than brand-name drugs.

The health care system in France is covered the medical expenses for all legal residents. One should notice that a certain percentage is reimbursed based on the price set by the system, not the

actual price the patient pays. Besides, the private insurance in France offer the plans to reimburse the remaining amount. There are three regimes of reimbursement: (1) The general one covers the majority of residents in France; (2) The special regime for residents in Alsace Moselle region as they pay more taxes and have higher rate of reimbursement; (3) The regime for elder people who receive the aid from the government.

3 Demand Estimation Models

One straight-forward approach is modeling the demand directly by the product space, $q = D(p, r, \epsilon)$, where q, p, ϵ is $J \times 1$ vectors of quantities, prices, and random shocks for J products. r is vector of other exogenous variables. If $D(p, r, \epsilon) = Ap + \epsilon$, where A is $J \times J$ matrix of parameters to capture the interactions of J products to its own and other (J^2 parameters to estimated). In addition, this aggregate demand approach does not allow us to estimate the distribution of heterogeneity.

Rather than, we would adopt the approach to model the demand on characteristics space, through the individual utility function as a function of attributes of alternatives. It allows explicitly modeling consumer preference, and it is more flexible to model the demand in different market condition (merger or introduction of new products). The indirect utility of the consumer i on products (antibiotics/molecules) j on the market t ($ATC3$ group $\times Year$) is given:

$$U(x_{jt}, \xi_{jt}, I_i - p_{jt}, \tau_i; \theta)$$

where: p_{jt}: price of product j in market t; x_{jt}: vector of observed product characteristics; ξ_{jt}: vector of unobserved product characteristics; τ_i: individual characteristics; I_i: individual income.

Consumer i will choose product j if (for all $k \neq j$):

$$U(x_{jt}, \xi_{jt}, I_i - p_{jt}, \tau_i; \theta) \geq U(x_{kt}, \xi_{kt}, I_i - p_{kt}, \tau_i; \theta)$$

3.1 Logit Model

The baseline model in this project is the Logit model. In which, individual random utility for the $J + 1$ goods ($j = 0, \cdots, J$) is given as below:

$$u_{ij} = x_j\beta - \alpha p_j + \xi_j + \epsilon_{ij}$$

ϵ_{ij} is the stochastic term. It represents that the utility is deterministic, but the choice process is probabilistic. Also, it captures the inability to model individual behavior perfectly. ϵ_{ij} is modeled as an i.i.d random variable under the extreme value distribution:

$$F(\epsilon_{ij}) = \exp(-\exp(-\epsilon_{ij}))$$

We can also write:

$$u_{ij} = \delta_j + \epsilon_{ij}$$

where $\delta_j = x_j\beta - \alpha p_j + \xi_j$ is the common part (mean) of consumer. ξ_j is the unobserved factors in products, which changes the mean by product j. The consumers are only different in ϵ_{ij}.

The probability that consumer i choose product j is:

$$\begin{aligned} P_{ij} &= P(u_{ij} \geq u_{ik}, \forall k \neq j) \\ &= P(\epsilon_{ik} - \epsilon_{ij} \leq \delta_j - \delta_k, \forall k \neq j) = \int P(\epsilon_{ik} \leq \epsilon_{ij} + \delta_j - \delta_k, \forall k \neq j | \epsilon_{ij}) dF(\epsilon_{ij}) \\ &= \frac{\exp \delta_j}{1 + \sum_{k=1}^{J} \exp \delta_k} \in (0,1) \end{aligned}$$

For the outside good: $P_{i0} = \frac{1}{1+\sum_{k=1}^{J} \exp \delta_k}$.

The aggregate market share equals to the individual choice probability. We obtain the standard linear regression model as:

$$\ln\left(\frac{s_j}{s_0}\right) = \delta_j = x_j\beta - \alpha p_j + \xi_j$$

Estimation with different market t:

$$\ln\left(\frac{s_{jt}}{s_{0t}}\right) = \delta_{jt} = x_{jt}\beta - \alpha p_{jt} + \xi_{jt} \tag{1}$$

There is a threat of endogeneity that the error term ξ_{jt} correlated with price, hence OLS estimates of α would be biased towards zero. We would correct the price endogeneity by IVs through 2SLS.

The elasticities is expressed as:

$$\eta_{jk} = \begin{cases} -\alpha p_j (1 - s_j), & \text{if j = k} \\ \alpha p_k s_k, & \text{otherwise} \end{cases} \tag{2}$$

This model has problems that it is lack of heterogeneity, hence the own price elasticities largely depends on price (s_j is small) while cross-price elasticity depend on shares and prices, but not product characteristics.

3.2 Nested Logit Model

To relax the assumptions in Logit model and consider the heterogeneous interactions among products, we consider the Nested Logit Model. We estimate the logit model within each group of products (nests), $g = 0, ..., G$. For products in same nest, the ratio of probabilities is independent of other alternatives (IIA within nest). Yet, outside the nests, the ratio of probabilities are determined by the characteristics of other products in two nests.

The individual random utility for products, where $\zeta_{igt}(\sigma)$ is common for all products in group g. We have: $\sigma_g = 1 - \lambda_g$, where $\lambda_g \in [0,1]$ indicates the degree of independence in unobserved utility among the alternatives in group p.

$$u_{ijt} = x_{ij}\beta - \alpha p_{ij} + \xi_j + \zeta_{igt}(\sigma) + (1 - \sigma)\epsilon_{ijt}$$

With the same process as in logit model, we want to derive the linear regression model. [Berry, 1994] showed that the analytical solution is:

$$\ln\left(\frac{s_{jt}}{s_{0t}}\right) = x_{jt}\beta - \alpha p_{jt} + \sigma_g \ln s_{jt|gt} + \xi_{jt} \tag{3}$$

The own elasticities:

$$\eta_j = \frac{\partial s_j}{\partial p_j} = -\frac{\alpha}{1-\sigma_g} s_j[1 - \sigma_g s_{j|g} + (1-\sigma_g)s_j] \tag{4}$$

The elasticities is:

$$\eta_{jk} = \frac{\partial s_k}{\partial p_j} = \begin{cases} \frac{\alpha}{1-\sigma_g} s_j(\sigma_g s_{j|g} + (1-\sigma_g)s_k) & \text{if } k,\ j \text{ in same group} \\ \alpha s_j s_k & \text{if } k,\ j \text{ in different groups} \end{cases} \tag{5}$$

3.3 Random Coefficients Logit model (BLP)

To allow further flexibility in substitution patterns and relax more assumptions of logit and nested logit model, the utility function is:

$$u_{ijt} = x_{jt}\beta_i + \alpha_i(I_i - p_{jt}) + \xi_{jt} + \epsilon_{ijt}$$

In this equation, the coefficients α_i and β_i is allowed to be varied by i. These consumer-level taste parameters are modeled as:

$$\begin{aligned} \alpha_i &= \alpha + \sum_{r=1}^{d} \pi_{1r} D_{ir} + \sigma_1 v_{i1} \\ \beta_{ik} &= \beta_k + \sum_{r=1}^{d} \pi_{(k+1)r} D_{ir} + \sigma_{k+1} v_{i(k+1),\ \text{for } k=1,\ldots,K} \end{aligned}$$

We have $D_i = (D_{i1}, ..., D_{id})'$ as a $d \times 1$ vector of observed demographic variables. $v_i = (v_{i1}, ..., v_{i(K+1)})'$ as a vector of $K+1$ unobserved consumer characteristics. Denote Π is a $(K+1)\times d$ matrix of parameters and $\sigma = (\sigma_1, ..., \sigma_{K+1})$ is a vector of parameters.

Let $\theta = (\alpha, \beta, \Pi, \sigma)$ as parameters of the model. $\theta_1 = (\alpha, \beta)$ are "linear" parameters, and $\theta_2 = (\Pi, \sigma)$ are "non-linear" parameters. We can decompose utility into "linear" and "non-linear" parts:

$$\begin{aligned} u_{ijt} &= \delta_{jt}(x_t, p_t, \xi_t; \alpha, \beta) + \mu_{ijt}(x_t, p_t, D_i; \Pi, \sigma) + \epsilon_{ijt} \\ &= \delta(x_{jt}, p_{jt}, \xi_{jt}; \theta_1) + \mu(x_{jt}, p_{jt}, D_i, \nu_i; \theta_2) + \epsilon_{ijt} \end{aligned}$$

A consumer with type (D_i, v_i) would choose the product giving the highest utility. The probability consumer i choose product j:

$$\begin{aligned} s_{ijt} &= P[u_{ijt} \geq u_{ikt} \forall k | x_t, \delta_t, p_t, D_i, v_i; \theta] \\ &= \int \mathbf{1}[u_{ijt} \geq u_{ikt} \forall k | x_t, \delta_t, p_t, D_i, v_i; \theta] dF_\epsilon(\epsilon) \end{aligned}$$

We can get aggregate market shares by integrating this probability over consumer attributes:

$$s_{jt}(x_t, \delta_t, p_t; \theta) = \int s_{ijt}(x_t, \delta_t, p_t, D_i, v_i; \theta) dF_D(D) dF_v(v)$$

The price elasticities is:

$$\eta_{jkt} = \frac{\partial s_{jt}}{\partial p_{kt}} \frac{p_{kt}}{s_{jt}} = \begin{cases} -\frac{p_{jt}}{s_{jt}} \int \alpha_t s_{ijt}(1 - s_{ijt}) dP_D(D) dP_v(v), & \text{if } j = k \\ \frac{p_{kt}}{s_{jt}} \int \alpha_i s_{ijt} s_{ikt} dP_D(D) dP_v(v), & \text{otherwise} \end{cases}$$

4 Data

In this paper, we utilize the same data-set in (Ussinova, Taleb el Houda and Fargal, 2017), with some additional information from other sources. The main data-set is the Medicam, which is extracted from Amelie health insurance company. The Medicam data set contains information from general security system regime, which are observed annually from 2001 to 2013. Based on this paper, we extend the time-line to 2015 by merging with new data. The Medicam data set contains:

- Different level of ATC classification system
- CIP13 code for drugs
- Some characteristics of patients such as: age, sex, region, etc.
- Total quantity of drugs (unit: box) sold
- Number of patient
- Potential reimbursed amount from social security system
- Actual reimbursed amount from social security system

Here based on this dataset, we make use the following key variables to define quantities, products and markets:

- **ATC5**: We use ATC5 to identify different molecules;
- **ATC3** $\times$ **YEAR**: We define different market levels by Year and ATC3, namely the therapeutic/pharmacological subgroup of molecules;
- **Standard units**: We use the standard units to define quantities Measurement of the usage of each antibiotics. In particular, the quantities in Boxes in Medicam is converted to standard units. The detailed definition is in Section 4.3.

Lagged Antibiotic Consumption in Animal: For the extension of last year paper, we utilize some additional information from other sources to improve the estimation results. First and foremost, we add the quantity of antibiotics used on animal, which is available from European Surveillance of Veterinary Antimicrobial Consumption (ESVAC), to the Medicam data-set. The key idea is to investigate the presence of antibiotic resistance by study the impact of lagged quantity usage to present quantity usage. The lagged quantity usage on animal will play a role as explanatory variable instead of lagged quantity usage on human, which is likely correlated with price and market share. The intuition behind it is that the more consumption in the past will reduce the consumption/demand today due to the antibiotic resistance, so we expect that the lag animal consumption might have negative coefficients on the demand.

Antibiotic Resistance Rate: Secondly, we try to study whether the antibiotic resistance rate has any significant effect on the demand of antibiotic by adding resistance rate to the data-set. The data is provided by Professor Dubois, contains of 8 different resistance rates for 8 different bacterias, and is merged to the data-set with primary key ATC4.

International Classification of Diseases (ICD): Finally, we try to provide a new classification system parallel with the ATC level, namely International Classification of Diseases (ICD). The main difference between ATC and ICD is that ATC classifies drug by its therapeutic and chemical characteristics, while ICD system categorizes medicine according to the diseases, disorders and injuries. The new system ICD will not only help improving the variation within the data-set but also provide additional solutions for market definition and IV candidates. However, this project will not focus on this information.

4.1 ATC Classification

WHO first published **ATC (Anatomical Therapeutic Chemical) Drug Codes Classification System** in 1976, with a motivation to create a synchronous approach to classify drugs by their anatomical, therapeutic, and chemical characteristics. The classification system divides drugs into different subgroups with similar features as follow:

- **ATC1:** 14 anatomical main groups ("J": anti-infective)
- **ATC2:** therapeutic subgroup (next two digits, J**01**)
- **ATC3:** pharmacological subgroup (next one letter, J01**A**)
- **ATC4:** chemical subgroup (next one letter, J01A**A**)
- **ATC5:** chemical substances/molecules (next two digits, J01AA**07**)

4.2 Data Processing Procedure

In Figure 1, we show the big picture of our data processing procedure. The basic idea is that we append the most recently data to expand our dataset, then we merge other source of data into the original one to include more variables. The original dataset we based is the Medicam (2001 - 2012). This is the raw dataset which contains the information of ATC categorization of each antibiotics, volume of the drug, number of drugs sold, base refund and reimbursed amount by the social security. After we update this data set from the previous year's group, we obtain Medicam (2001 - 2015).

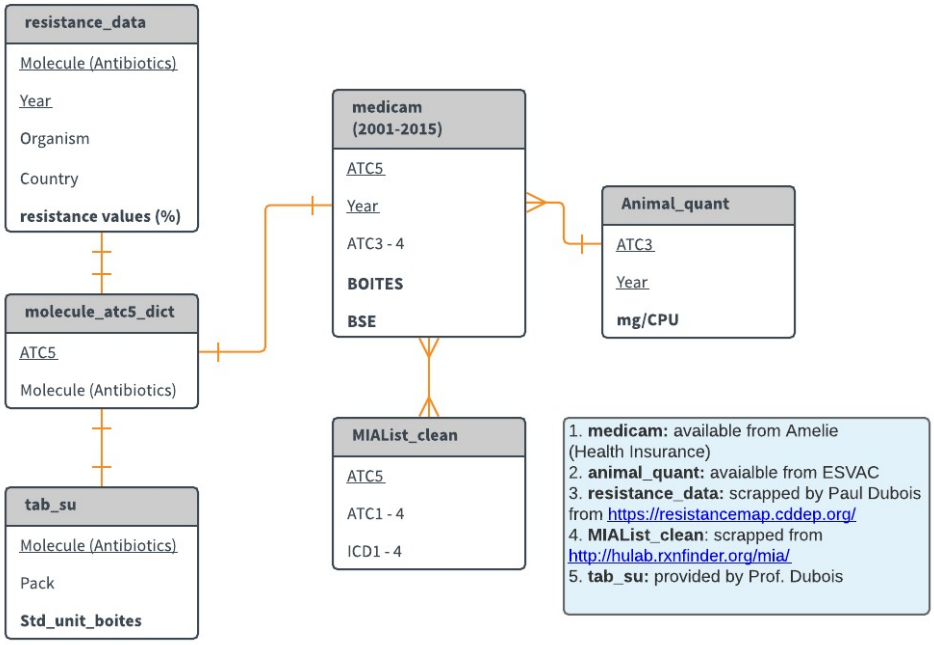

Figure 1: The procedure of the data processing

Then we do some basic calculation to generate the variables for regression. In this step, we generate the standard unit for the usage of each antibiotics. In addition, we define the market share and group all the observation in the same ATC3 category, or we call same market. Further, we compute quantities of outside goods. See Section 4.3 for detail.

In the next step, we try to merge our dataset with other datasets to add more information. To start with, we merge with the usage of antibiotics in the animal market. In this data set, it contains ATC category and quantity of animal usage for the same antibiotics. Therefore the primary key for merging is ATC3. Note that the main problem of merging this dataset is that the records of it start from 2011, which is not consistent with our original dataset. Hence we generate an interaction term $\mathbb{1}\{YEAR \geq 2011\} \times q_{t-1}^{animal}$ in order to expand the data into the range of 2001 – 2015. After filling the missing values, we are able to merge the two datasets.

Then, we merge with the ICD (International Classification of Diseases) Disease Codes Category

dataset. In this dataset, it contains the mapping between ATC and ICD classifications. We can treat ICD categorization as a parallel way of classify drugs as ATC does but based on different classification standard. The problem of merging here is that the mapping is not 1 to 1, one ATC may map to one or more ICDs, and vice versa. Thus we propose a solution of using dummy variable where we generate m new dummy variables where it takes value 1 if one particular ATC-categorized product belongs to ICD level categories and 0 otherwise (m is the number of the different ICD level categories). With this mapping, we can merge with the ATC-setting dataset.

Notice that the level of ATC categories are as ATC3 > ATC4 > ATC5. We use ATC3 to define the market, ATC5 to identify different antibiotic product. However, ICD is a categorization in between ATC3 and ATC5 so that we use this variable for different purposes, such as alternative definition of markets, nests of products, or to construct IVs.[3]

4.3 Key Variables

Standardized Unit of Consumed Quantity: In the original Medicam data-set, the information about antibiotics quantity usage represented by BOITES, which are the amount of box of drugs sold for each specific product. However, due to the fact that boxes contains different amount of usages even for the same medicines, utilizing the number of box as measurement for quantity usage will likely lead to serious measurement errors. For instance, for doliprane, we have both data for the 5g dosage box and 10g dosage box. It is obvious that performing any comparison in terms of boxes will be biased as boxes do not represent the amount of dosage sold. A solution for this problem is to derive standard unit. Standard unit is defined as smallest unit within a box (usually dosage), for which will then be generalized for the same molecule (ATC5 level). Later on in the paper, all the estimations are performed using only standard units, includes price of standard units, market shares of standard units.

For each molecule (defined by ATC5 level), we have the information on standard units obtained from data-set provided by Prof. Dubois. After aggregating the initial Medicam data by ATC5 level, we match it with the standard units data-set. The formula for computing average standard unit per ATC5 follows strictly the last year paper:

$$\frac{\sum_{i \in ATC5_j}(s_j q_i)}{\sum_{i \in ATC5_j}(q_i)} = \frac{(\text{Total Nnumber of Standard Units})_{ATC5_j}}{\sum_{i \in ATC5_j}(q_i)} = \overline{su}_{ATC5_j},$$

where $\bar{su}_{ATC5_j}$ is the standard unit for ATC5 level j; where $j = \{1, ..., J\}$ is the level of ATC5.

Price from Reimbursement: Given the quantity sold, we deduct the Price for each product. As we cannot obtain any official price data, we have to create its proxy by using total potential reimbursement divided by the quantity sold (in standard units). One problem is that given a price

[3] In fact, within the scope of this project, we have not used this information in the results. But, its availability in the data might be useful for the next year group.

of a drug box, the percentage of reimbursement by social security for that box will not be always 100 percent, i.e. actual reimbursement might different from attached box price. Since this is the best proxy we can obtain for the price, we will have to construct it with the assumption that all box of drug will be reimbursed fully by the social security. The formula to calculate product price is presented below:

$$P_{jt} = \frac{BSE_{jt}}{q_{jt}},$$

where BSE_{jt} is the potential reimbursement for product j at time t

Market Share: market in this paper is defined by ATC3 level, while the molecule (alternative) is defined by ATC5. ATC3 refers to the therapeutic and pharmacological characteristics of drugs, and within each market ATC3, a product will be identified by its ATC5 code. For each ATC3 market, a market share for an ATC5 alternatives will be calculated by:

$$s_{jt} = \frac{q_{jt}}{M_{mt}},$$

where M_{mt} is the total quantity of product within a market m at time t, included both the medicine from Medicam data set and outside good. Outside goods are assumed to take account of 20% of total quantity sold: $q_{0mt} = (\sum_{j=1}^{J} q_{jmt}) \times 0.2$

As mentioned in previous part, we use the quantity of antibiotic usage on animal to investigate the existence of antibiotic resistance in France. Antibiotics usage on animal have measurement unit mg/PCU and only available for ATC3 level, instead of ATC5 as we desire. By merging the animal quantity to our data-set through primary key ATC3, we at the same time assume that every product in market $ATC3_j$ will have the same animal quantity usage. It is obviously incorrect since not all medicines that is used for human is used on animal. Nevertheless, as we cannot find any additional information, we have to accept this assumption.

Similar to the animal quantity, our antibiotics resistance rate only available for ATC3 and ATC4. For one specific ATC3/ATC4 level in a given year, we have the resistance rate for at most 8 types of bacterial, measured as percentage. The data available for period from 2002 to 2015 and will be merged to our master data-set through primary key ATC3/ATC4.

4.4 Descriptives

Below **Figure 2** represent the histogram of market shares for all the products available in our data-set. As can be seen, a majority of products account for significantly small market shares, roughly below 3%; while some products will dominate the whole market by taking over 80% of the total market. This wide variance can be explained by examining the attributes of **Figure 4**. **Figure 4** represents the total number of products (total number of ATC5) within each market ATC3. For some markets, for example J01B and J01R, the number of alternatives within each market is relatively low (approximately under 5) might lead to higher percentage of market share

for each alternative; while market like J05A which contains over 40 molecules, thus each molecule will only be able to take smaller piece of pie.

Take a look at **Figure 5**, we can easily observe similar pattern with **Figure 2**. For both market shares and price, since we notice that the variances between Min and Max values are very significant and the extreme values might effects the performance of our estimation, thus we think that it is necessary to control for these 2 factors as below:

- **Control for low market share** (product with market share below 0.1 percent will not be included in our model)
- **Control for extreme unit price** (product with unit price above 100 will not be included in our model)

Table 1: Descriptive Statistics

	Obs	Mean	Std. Dev.	Min	Max
MARKETSHARE_SU	1,145	0.195	0.231	0.010	0.833
PRIXMOY_SU	1,145	7.350	17.373	0.026	135.591
SD_QUAN	1,145	2.10e+07	4.81e+07	6,142.282	4.67e+08
LAG_MG	1,145	1.566	6.581	0.000	64.278
BOITES	1,145	1,622,412	3,760,459	4,202.483	4.13e+07
BSE	1,145	2.20e+07	2.89e+07	17669.18	1.60e+08
REM	1,145	1.66e+07	2.18e+07	9,671.959	1.23e+08

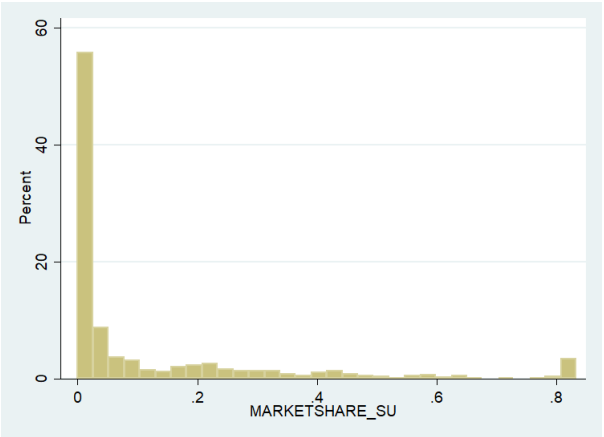

Figure 2: Histogram for market share

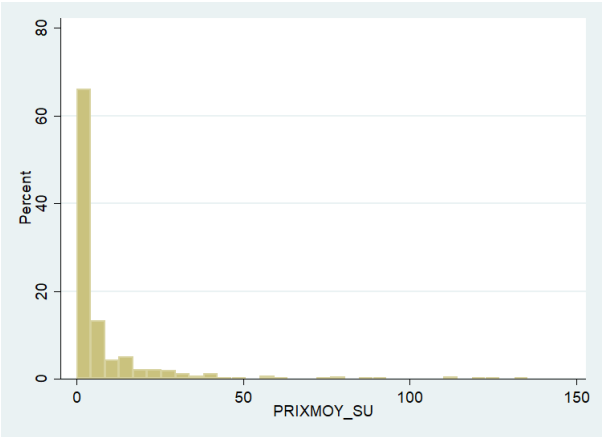

Figure 3: Histogram for price per standard unit

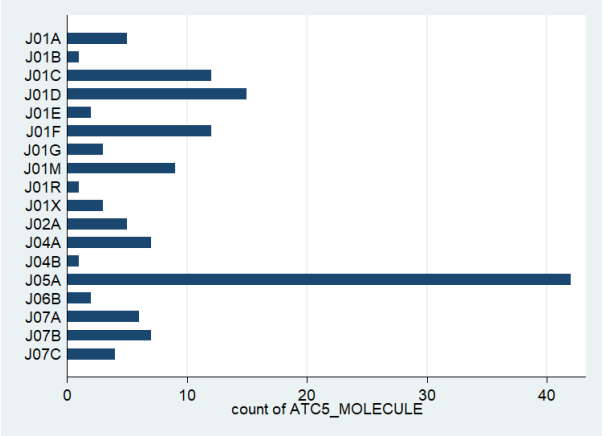

Figure 4: Total number of product by ATC3

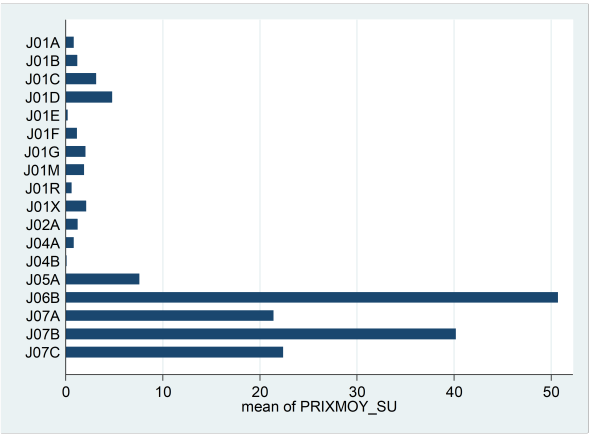

Figure 5: Average product price by ATC3

5 Results

5.1 Instrumental Variables

In section 5, we will observe and discuss the results from Logit Model, Nested Logit Model, and Random-coefficient Logit Model (BLP). In the estimation of demand function, we usually face the endogeneity problem due to simultaneous determination of price and quantities at the equilibrium. We therefore consider to use instrumental variables and 2SLS. The principle of choosing instrumental variables is such that the instrumental variables should be correlated with prices but uncorrelated with the error term, i.e. they need to affect market shares (quantities at the equilibrium) through the prices, but not directly the quantities. In total, we construct 5 different instrumental variables. Their detailed description are summarized in **Table 2**.

Table 2: Different IV Specifications

Regression	Variable(s) Used as Instrument(s)	Description
IV1	INTER_ATC3_ATC5	Interaction term between ATC3 dummies and the number of molecules (ATC5) in it
IV2	COUNTOTHERATC5	Total number of ATC5 except those from one specific category of ATC3 at each time
IV3	PRIXMOY_otherATC3	ATC3-YEAR level average price among other ATC3
IV4	PRIXMOY_otherATC3; COUNTOTHERATC5	Two instruments, as explained above
IV5	INTER_ATC3_ATC5; PRIXMOY_otherATC3	Two instruments, as explained above

IV1 and IV2 are proxies for competition intensity. They affect prices and quantities indirectly through prices. IV3 is the average price of other ATC3 (consider prices in other markets related with but not as same as ATC5 level). IV4 and IV5 combine the above two points together. It is then desired to define clearly the instrumental variables that are used in all these models before proceeding to the results.

Table 3: First Stage Estimates of IV Regression (Logit Model)

Logit (IV1)	coefficient	s.e.	t value	p value
INTER_*	-	-	-	-
LAG_MG	-0.00514	0.033425	-0.15	0.878
F(14,1443) = 2.08, Prob > F = 0.0104				
Logit (IV2)	coefficient	s.e.	t value	p value
COUNTOTHERATC5	-0.15827	0.050851	-3.11	0.002
LAG_MG	-0.01054	0.032892	-0.32	0.749
F(1, 1456) = 9.69, Prob > F = 0.0019				
Logit (IV3)	coefficient	s.e.	t value	p value
PRIXMOY_otherATC3	-3.7525	0.912078	-4.11	0
LAG_MG	-0.01613	0.032459	-0.5	0.619
F(1, 1456) = 24.52, Prob > F = 0.0000				
Logit (IV4)	coefficient	s.e.	t value	p value
PRIXMOY_otherATC3	-3.42102	0.919344	-3.72	0
COUNTOTHERATC5	-0.13162	0.051128	-2.57	0.01
LAG_MG	-0.00269	0.032813	-0.08	0.935
F(2, 1455) = 11.81, Prob > F = 0.0000				
Logit (IV5)	coefficient	s.e.	t value	p value
INTER_*	-	-	-	-
PRIXMOY_otherATC3	-8.58805	1.682326	-5.1	0
LAG_MG	-0.00368	0.033133	-0.11	0.912
F(15, 1442) = 3.72, Prob > F = 0.0000				

Before running the regression using the IVs listed above, we have to check the relevance condition between endogenous variable and all the exogenous variables including IVs. We expect that our IVs have strong correlation with the endogenous variable. Here we report the result of the first stage regression in Logit model as well as Nested Logit model in **Table 3** and **4** respectively. The endogenous variable is the standardized price `PRIXMOY_SU`. In the tables, it lists the coefficients of the IVs in the first stage regression.[4] We can easily find that they are significant, meaning that

[4]For the IV1 and IV5, since the IV `INTER_*` consists of a series of variables, we do not list the result of estimated coefficients in the tables. Instead, we can make inference based on the F exclusion test.

the relationship between endogenous variables and IVs are sufficiently strong. Furthermore, we perform the F test of excluded instruments of each first stage regression. The fact that all the exogenous variables are jointly different from 0 suggests that IVs are relevant.

Table 4: First Stage Estimates of IV Regression (Nested Logit Model)

Nested Logit (IV1)	coefficient	s.e.	t value	p value
INTER_*	-	-	-	-
LAG_MG	-0.0116848	0.0388412	-0.3	0.764
lsjg	-0.5006886	0.1452129	-3.45	0.001
F(14, 1443) = 2.17, Prob > F = 0.0072				
Nested Logit (IV2)	coefficient	s.e.	t value	p value
COUNTOTHERATC5	-0.1596049	0.0582266	-2.74	0.006
LAG_MG	-0.0145376	0.0381955	-0.38	0.704
lsjg	-0.5235715	0.1442595	-3.63	0
F(1, 1456) = 7.51, Prob > F = 0.0062				
Nested Logit (IV3)	coefficient	s.e.	t value	p value
PRIXMOY_otherATC3	-5.19718	1.049591	-4.95	0
LAG_MG	-0.0164432	0.0375361	-0.44	0.661
lsjg	-0.5173166	0.1432738	-3.61	0
F(1, 1456) = 24.52, Prob > F = 0.0000				
Nested Logit (IV4)	coefficient	s.e.	t value	p value
PRIXMOY_otherATC3	-4.891047	1.058492	-4.62	0
COUNTOTHERATC5	-0.1223856	0.0583822	-2.1	0.036
LAG_MG	-0.0033588	0.0380084	-0.09	0.93
lsjg	-0.4996389	0.1433554	-3.49	0.001
F(2, 1455) = 14.49, Prob > F = 0.0000				
Nested Logit (IV5)	coefficient	s.e.	t value	p value
INTER_*	-	-	-	-
PRIXMOY_otherATC3	-10.17179	1.945167	-5.23	0
LAG_MG	-0.0101126	0.0384925	-0.26	0.793
lsjg	-0.5196704	0.143951	-3.61	0
F(15, 1442) = 3.89, Prob > F = 0.0000				

5.2 Logit Results

In **Table 5**, we represent the estimated results by running standard Logit model, as well as a comparison between our model (New) and the model from last year paper (Old). Our model includes the log of market shares (for standard units) as dependent variable; price for standard units, which is endogenous and instrumented by several IVs, and lagged quantity of antibiotics

usage on animals are considered as key explanatory variables. We also impose restrictions such that the models only consider including observations that have market share greater than 0.01 percent and price for standard units below 100.

As can be seen from the table, most of the coefficients are statistically significant at 5%, and all of them are negative. The results mean that an increase in price will lead to a significant drop in demanded quantities for antibiotics, which are reasonable and in line with classical economics theory. It is also noticeable that compare to the Old results, our model with standardized units, sets of IVs, and the control over the sample yields superior performance in terms of magnitudes of coefficients, while still being statistically significant.

While including lagged animal quantities might help the demand to response more intensively to price, the lagged animal quantities itself do not impose notable impacts on demanded quantities. All estimated coefficients, even though consistently negative, are not statistically significant at 10% level. Yet, the negative signs of coefficients aligns with our prediction. Choosing a stronger set of instrumental variables might be a solution to improve the estimation performance. However, it is still an interesting findings, which indicates that there is foundation for our expectation that the past consumed quantities of antibiotics affect the future consumer quantities.

Table 5: Estimates: Logit Models

	OLS		IV1		IV2		IV3		IV4		IV5	
	Old	New	Old	New	Old	New	Old	New	Old	New	Old	New
PRIXMOY_SU	-0.001***	-0.047***	-0.004***	-0.066*	-0.005***	-0.174**	-0.003	-0.108**	-0.005***	-0.131***	-0.003***	-0.023
	(0.00)	(0.01)	(0.00)	(0.03)	(0.00)	(0.08)	(0.00)	(0.05)	(0.00)	(0.04)	(0.00)	(0.02)
LAG_MG		-0.015		-0.005		-0.008		-0.006		-0.007		-0.004
		(0.01)		(0.01)		(0.01)		(0.01)		(0.01)		(0.01)
Constant	-15.271***	0.295	-4.239***	0.034	-4.055***	2.270	-4.351***	0.909	-4.062***	1.377	-4.365***	-0.853
	(2.73)	(0.45)	(0.10)	(0.75)	(0.10)	(1.58)	(0.29)	(1.04)	(0.10)	(0.93)	(0.10)	(0.55)
Observations	1691	1831	1691	1491	1691	1491	1691	1491	1691	1491	1655	1491
Adjusted R^2	0.369	0.426										
R^2	0.392	0.447										
F	16.94	21.28	55.62	30.58	101.9	19.74	1.135	26.82	101.1	24.43	46.64	31.97
df_m	62	67	1	65	1	65	1	65	1	65	1	65
df_r	1628	1763										

Note: standard error in the parentheses; p<0.10 *; p<0.05 **; p<0.01 ***

5.3 Nested Logit Results

We also conduct the demand estimation by nested logit model. The results are reported in **Table 6**. Once again, we obtain a relatively better results comparing the last year project. In fact, all price coefficients are larger in magnitude. Most of them are significant at 10% level. We have IV4, which performs quite well with the price coefficient significant negative at 5%.

For the lagged animal quantities, all coefficients are still not significant at any level. Some models report the positive sign. However, in the model with good IVs that the price coefficients are significant, the lagged animal quantities will have negative sign. Comparing to the results of

logit models, the nested logit presents the results not much different in size and sign.

Table 6: Estimates: Nested Logit Models

	Nested		Nested (IV1)		Nested (IV2)		Nested (IV3)		Nested (IV4)		Nested (IV5)	
	Old	New	Old	New	Old	New	Old	New	Old	New	Old	New
PRIXMOY_SU	0.000	-0.018***	-0.002***	-0.037	-0.003***	-0.142*	-0.002	-0.064*	-0.003***	-0.080**	-0.002***	-0.019
	(0.00)	(0.00)	(0.00)	(0.03)	(0.00)	(0.08)	(0.00)	(0.04)	(0.00)	(0.03)	(0.00)	(0.02)
LAG_MG		0.004		0.000		-0.003		-0.001		-0.001		0.001
		(0.00)		(0.01)		(0.01)		(0.01)		(0.01)		(0.01)
lsjg	0.999***	0.994***	0.846***	0.580***	0.838***	0.522***	0.850***	0.565***	0.839***	0.556***	0.836***	0.590***
		(0.01)		(0.03)		(0.05)		(0.03)		(0.03)		(0.03)
Constant	-0.183	0.506***	-2.246***	-1.019***	-2.179***	-1.229***	-2.274***	-1.073***	-2.182***	-1.104***	-2.334***	-0.982***
	(0.17)	(0.17)	(0.07)	(0.24)	(0.07)	(0.33)	(0.15)	(0.25)	(0.07)	(0.26)	(0.07)	(0.23)
Observations	1691	1831	1691	1491	1691	1491	1691	1491	1691	1491	1655	1491
F	849.3	306.7	1398.3	47.33	1359.5	30.20	1361.7	44.01	1342.6	41.50	10.82	48.53
df_m	63	68	2	34	2	34	2	34	2	34	60	34
df_r												

Note: standard error in the parentheses; p<0.10 *; p<0.05 **; p<0.01 ***

5.4 BLP Results

In this project, we use Stata consistently and apply the command "blp" by which the heterogeneity parameters are estimated using the modified Newton-Raphson method with analytic gradient. It is worth noticing that this Newton-Raphson method leads to the smallest objective function value for both the automobile data used by Berry et al. (1995) and the cereal data used by Nevo (2000a, b).

We then run BLP on the most represented ATC3 (more than 100 observations) including J01C, J01D, J01F, J01M, J05A, and J07B. If including J05A in the BLP, there is an error message: Hessian is not positive semidefinite. This corresponds to the fact that there are much more counts of ATC5 within ATC3 group J05A. The endogenous variable is the price for antibiotics in standard unit, and the instruments are the five we specify above. However, we should notice that not all of these instrumental variables can be used. For example, when we apply IV2 (i.e. COUNT_otherATC5) and IV3 (i.e. PRIXMOY_otherATC3), the error message shows insufficient instruments. The other three sets of instrumental variables correspond to our definitions above.

As we have done for both Logit Model and Nested Logit Model in the above two subsections, we compare the results obtained from our model (New) with those from last year paper (Old). Our IV5 includes Interaction terms and PRIXMOY_otherATC3 (i.e. ATC3-YEAR level average price among other ATC3), while in last year the IV5 includes Interaction terms and PRIXMOY_otherATC5 (i.e. ATC5 level average price among other ATC5, within market (ATC3-YEAR)). This explains why the comparison is not conducted for IV5 in **Table 9** below.

All BLP results are reported below in **Table 7**, **Table 8**, and **Table 9**. In the following paragraphs, we firstly talk about the significance of estimates, then we observe some patterns that how the signs of both price and lagged animal quantities of usage can change or remain the same in terms of certain ATC3 subgroup, and at last we compare the results of price coefficients from

our New models and those from Old models last year.

Taking the first glance of these three tables, we immediately observe that for our New models, although in general the price coefficients tend to be greater in magnitude compared with the estimates obtained from Logit Model and Nested Logit Model, we do not have the statistical significance for all of them. In addition, all estimates for the lagged animal quantities are not significant at any levels as well. The most possible reason for the insignificance is that our instrumental variables are not strong enough.

If we take a closer look at the price coefficients with respect to each ATC3 subgroup across all three BLP models, we can notify the difference regarding whether their signs change across Models when we apply different IVs. For example, the price coefficients for J01D across three models are all positive. The price coefficient for J01M is negative when IV1 is applied but it turns to be positive when IV4 and IV5 are applied. The price coefficients for J01C, J01F, and J07B are all negative across three models.

However, the signs of lagged animal quantities of usage remain consistent for all models with respect to certain ATC3 subgroup. For example, the estimates for lagged animal quantities are all negative for J01C, J01D, J01M, and J07B across three BLP models. The estimates for lagged animal quantities are all positive for J01F across three BLP models.

Regarding the comparison between the Old results and New results for BLP 1 and BLP 2, even though the price was not converted into standard unit last year, it is still interesting to notice that as shown in **Table 7**, the price coefficients for each ATC3 subgroup keep the same sign if we apply IV1, but as reported in **Table 8**, the signs for price coefficients of J01D, J01F, and J01M change when we apply IV4, which has been shown to be the most effective one in both Logit Model and Nested Logit Model.

Table 7: BLP 1 - Instruments: Interaction terms – corresponding to IV1

	J01C		J01D		J01F		J01M		J07B	
	Old	New	Old	New	Old	New	Old	New	Old	New
Mean utility										
cons	30.629	6.776	-140.645	-2.263	-4.242	-0.032	62.985	-0.245	158.584	6.258
	(.)	(91.88)	(.)	(1.68)	(.)	(5.13)	(.)	(239.36)	(.)	(15.05)
LAG_MG		-0.085		-1.733		0.032		-0.914		-3.580*
		(1.74)		(2.30)		(0.06)		(120.03)		(2.17)
PRIXMOY_SU	-6.095**	-20.296	5.729	0.258	-0.584	-1.244	-5.065	-0.556	-7.579	-0.316
	(2.32)	(282.27)	(31.67)	(0.81)	(3.71)	(4.21)	(11.23)	(126.68)	(15.83)	(0.79)
PRIXMOY_SU										
SD	0.500	1.871	0.500	0.000	0.500	0.000	0.500	0.000	0.500	0.060
	(1.12)	(374.51)	(13.15)	(.)	(2.41)	(.)	(6.80)	(.)	(7.85)	(0.41)
Observations	153	93	174	158	143	152	134	131	113	49

Note: Standard error in the parentheses: p<0.10 *; p<0.05 **; p<0.01 ***

Table 8: BLP 2 - Instruments: PRIXMOYATC3 COUNTOTHERATC5 – corresponding to IV4

	J01C		J01D		J01F		J01M		J07B	
	Old	New	Old	New	Old	New	Old	New	Old	New
Mean utility										
cons	-2.140	3.655	-3.542	-4.955	-8.553*	-1.129	10.309	-1.547	-3.791	3.445
	(5.19)	(102.76)	(6.04)	(6.84)	(4.21)	(2.07)	(48.49)	(2.98)	(10.37)	(36.94)
LAG_MG		-0.035		-5.183		0.041		-0.261		-3.694
		(1.79)		(9.13)		(0.04)		(1.68)		(3.61)
PRIXMOY_SU	-0.777	-13.026	-0.069	1.565	0.329	-0.343	-1.699	0.133	-0.183	-0.178
	(1.20)	(294.57)	(0.65)	(3.31)	(0.29)	(1.70)	(6.90)	(1.57)	(1.87)	(1.98)
PRIXMOY_SU										
SD	0.187	0.659	0.000	0.000	0.002	0.000	0.554	0.000	0.107	0.011
	(0.33)	(691.66)	(4.29)	(.)	(3.01)	(.)	(2.26)	(.)	(0.82)	(4.27)
Observations	153	93	174	158	143	152	134	131	113	49

Note: Standard error in the parentheses: p<0.10 *; p<0.05 **; p<0.01 ***

Table 9: BLP 3 - Instruments: Interaction terms and PRIXMOY_otherATC3 – corresponding to IV5

	J01C	J01D	J01F	J01M	J07B
	b/se	b/se	b/se	b/se	b/se
Mean utility					
cons	4.831	-3.741	-1.515	-2.002	4.698
	(6.14)	(2.87)	(2.47)	(2.61)	(16.53)
LAG_MG	-0.054	-3.627	0.044	-0.033	-3.664*
	(0.11)	(4.03)	(0.04)	(1.55)	(2.15)
PRIXMOY_SU	-15.849	0.976	-0.026	0.374	-0.236
	(14.59)	(1.38)	(2.03)	(1.38)	(0.87)
PRIXMOY_SU					
SD	0.000	0.000	0.000	0.000	0.013
	(.)	(.)	(.)	(7.65)	(1.47)
Observations	93	158	152	131	49

Note: Standard error in the parentheses: p<0.10 *; p<0.05 **; p<0.01 ***

5.5 Elasticities

We compute the elasticity from Logit and Nested Logit model as below figure, using our "best" instrumental variable IV4 due to its superior significant impact on demanded quantities. As can be seen, the majority of computed elasticities are extremely close to zero. The reason is that even though the estimated coefficients are significantly improved compare to last year paper, their values are still considerably low. The elasticities results clearly indicate that the French demand for antibiotics is very inelastic. It is reasonable since consumers of antibiotics are sponsored partially or sometimes fully by the government, they tend to not response on price adjustments.

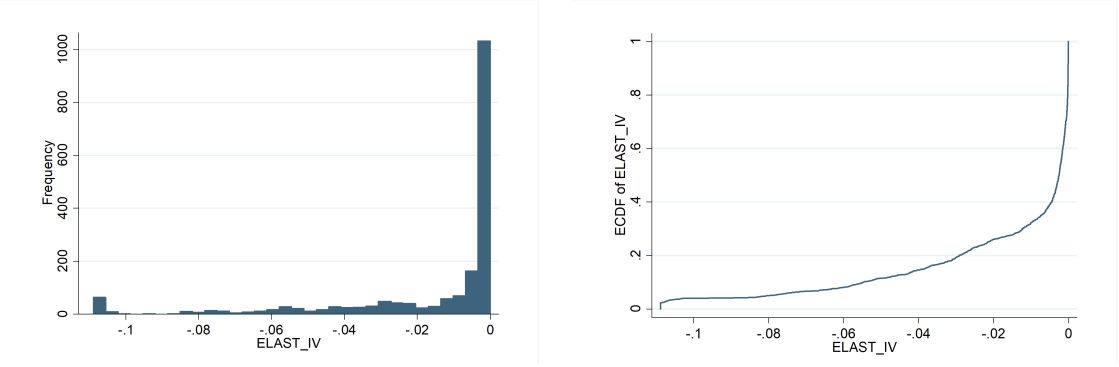

Figure 6: Histogram of elasticities and Cumulative elasticities using IV4

We do not report the elasticities from BLP, due to the inconclusive results. Further attempts are therefore suggested to show the elasticities in our well-defined markets, with the variables of each row and column explicitly shown for analysis.

6 Conclusions and Extensions

The key problem of the last year work is the limited data, thus the estimated price coefficients are low and not robust. In this year, we put efforts to enhance the information by extending the time horizon of the data of reimbursement from Amelie, a French insurance company, and augmenting it with more publicly available data (from ESVAC, WHO, etc.). Other improvements are implemented, such as standardizing the units of quantities, control the spread of price and market shares. The results present the price coefficients higher in magnitude, while still significant. The elasticities are improved comparing to the last year report, but still quite low.

Another interesting findings is that the lagged animal antibiotic usage always have the negative coefficients on human demand, even though not significant. It might be because of the limited in observations, as we only have the data of animal consumptions after 2012. We suggest any further effort to extend this information for the promising evidence of significant negative effects.

During the data processing, we face several problems of mapping data as there are many non-corresponding observations. Also, the cross-mapping issues exists, for example: each ATC5 could be mapped to several ICD. Due to these problems, many collected information has not been used in this project. The information for characteristics of antibiotics are limited.

For the next steps, we propose to augment more characteristic information. One could use the CIP13, specifies box of drug to merge to generic companies, or improve the compatibility of antibiotic resistance rate and ICD code. For the better IVs, we suggest to find the price data in other markets (such as Denmark and Philippines have very good public data) [5].

[5] In the term of this project, we have tried to contact Danish Medicine Agency for price information. Unfortunately, due to the time constraints, we have not been able to access it.

References

[Berry, 1994] Berry, S. T. (1994). Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics*, pages 242–262.

[DiMasi et al., 2010] DiMasi, J. A., Feldman, L., Seckler, A., and Wilson, A. (2010). Trends in risks associated with new drug development: success rates for investigational drugs. *Clinical Pharmacology & Therapeutics*, 87(3):272–277.

[DiMasi et al., 2003] DiMasi, J. A., Hansen, R. W., and Grabowski, H. G. (2003). The price of innovation: new estimates of drug development costs. *Journal of health economics*, 22(2):151–185.

[Garau and Mestre-Ferrandiz, 2006] Garau, M. and Mestre-Ferrandiz, J. (2006). *European Medicines Pricing and Reimbursement: Now and the Future*. Radcliffe Publishing.

[Grandfils et al., 2008] Grandfils, N. et al. (2008). Drug price setting and regulation in france. *Institute for Research*.

[Ussinova et al., 2017] Ussinova, A., Houda, I. T. e., and Fargal, V. (2017). Empirical project on antibiotics in france. *M2 EEE - TSE Project*.

[Woodhead et al., 2011] Woodhead, M., Blasi, F., Ewig, S., Garau, J., Huchon, G., Ieven, M., Ortqvist, A., Schaberg, T., Torres, A., Van Der Heijden, G., et al. (2011). Guidelines for the management of adult lower respiratory tract infections-full version. *Clinical microbiology and infection*, 17:E1–E59.