



HAL
open science

Dealing with Demand Heterogeneity on Health Care Provider Choice –The Case of Rural China

Martine Audibert, Yong He, Jacky Mathonnat

► **To cite this version:**

Martine Audibert, Yong He, Jacky Mathonnat. Dealing with Demand Heterogeneity on Health Care Provider Choice –The Case of Rural China. *Asian Journal of Economics and Empirical Research*, 2020, 7, pp.258 - 267. 10.20448/journal.501.2020.72.258.267 . halshs-02963761

HAL Id: halshs-02963761

<https://shs.hal.science/halshs-02963761>

Submitted on 11 Oct 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Dealing with Demand Heterogeneity on Health Care Provider Choice – The Case of Rural China

Martine Audibert¹

Yong He²

Jacky Mathonnat³

¹Université Clermont Auvergne, CNRS, CERDI, France.

²Email: yong.he@uca.fr

³Professor Emeritus in Economics, Clermont Auvergne University, Senior Fellow, FERDI, France.

Corresponding Author

Abstract

We built a database of two samples of patients surveyed within the same regions in rural China over a time interval of 18 years, and presumed varying demand heterogeneity due to income increase and people aging. We find that while the mean price and distance negative effects on patients choice were present in both time periods, their differences in heterogeneity, which were confirmed with the mixed multinomial logit (MMNL), could have crucial importance in avoiding erroneous policy making based merely on mean price and distance effects. We also find that while both the multinomial logit (MNL) and the MMNL are able to predict price and distance effects with low heterogeneity, only the MMNL appears able to detect the price effect when heterogeneity is high. These findings suggest using caution when interpreting estimation results with the MNL in cases of high heterogeneity.

Keywords: Price effect, Distance effect, Healthcare choice, Preference heterogeneity, Multinomial logit, Mixed multinomial logit models, Chinese rural households.

JEL Classification: D1; C5; I1.

Citation | Martine Audibert; Yong He; Jacky Mathonnat (2020). Dealing with Demand Heterogeneity on Health Care Provider Choice – The Case of Rural China. Asian Journal of Economics and Empirical Research, 7(2): 258-267.

History:

Received: 20 July 2020

Revised: 24 August 2020

Accepted: 28 September 2020

Published: 7 October 2020

Licensed: This work is licensed under a Creative Commons

Attribution 3.0 License

Publisher: Asian Online Journal Publishing Group

Acknowledgement: All authors contributed to the conception and design of the study.

Funding: This study received no specific financial support.

Competing Interests: The authors declare that they have no conflict of interests.

Transparency: The authors confirm that the manuscript is an honest, accurate, and transparent account of the study was reported; that no vital features of the study have been omitted; and that any discrepancies from the study as planned have been explained.

Ethical: This study follows all ethical practices during writing.

Contents

1. Introduction	259
2. Issues on Heterogeneity in Healthcare Demand	259
3. Estimation Method and Data	261
4. Estimation Results	264
5. Conclusions	267
References	267

Contribution of this paper to the literature

This study contributes to existing literature by providing evidence that the most popular multinomial logit model works only in the case of the choice with homogenous demand, and to avoid estimating bias, the mixed multinomial logit is indispensable when demand becomes heterogeneity.

1. Introduction

The demand for health care and the effects of price and distance on patients' provider choices have been subject to extensive studies that carry important policy implications. For instance, a weak price effect implies a high demand for more expensive and high quality healthcare services, and thus suggests the need for resource allocation towards large and well-equipped hospitals. Studies on distance effect can help to optimize geographical allocation of medical resources. A strong distance effect suggests the need for a more decentralized system with small and nearby healthcare providers.

There is, however, a serious gap on how these effects manifest and are interpreted in the presence of demand heterogeneity among patients. By demand (or choice) heterogeneity, we mean the extent of the difference in demand among the patients in function of the healthcare prices and of the distances of the healthcare providers. In other words, they are the variances of the price and distance effects across the patients. This issue is crucial for health care because some significant changes in heterogeneity could change the sense or the extent of the mean price and (or) distance effects. Difference in demand heterogeneity gives rise to different policy implications.

The first objective of this study, therefore, is to fill the gap in the healthcare literature with the estimation of the effects of price and distance on healthcare provider choice in the presence of various types of demand heterogeneity and illustrate the importance of taking heterogeneity into account in policy making.

We construct two samples of patients from the same villages of nine Chinese provinces over two periods: 1989-1993 and 2004-2006. We focus on the two most important factors that could lead patients' provider choices to become more (or less) heterogeneous: During two periods, the average real income per capita were more than doubled, along with a sharp increase in income inequality, and the average age of patients significantly increased. It is reasonable to assume that an increase in patient income, while causing a weaker price effect, will lead provider choices based on healthcare price to be more heterogeneous, because these choices are henceforth made based on their difference in preferences on quality, and other observed and unobserved aspects of health care. We can also expect that more elderly patients would be less likely to choose a provider based on healthcare price and more likely to choose providers based on proximity. Therefore, choice heterogeneity based on price and distance could be expected to be lower with population aging.

Our second objective is methodological. There have been a number of studies with divergent conclusions on the estimation robustness of two econometric models. The conditional multinomial logit model (MNL) is mostly used for estimating healthcare demand, but is considered by some people as not suitable in cases of high choice heterogeneity. The mixed multinomial logit model (MMNL), which has emerged more recently, offers the possibility of decomposing individual preferences. Comparing the performance between the two models in the presence of preference heterogeneity will allow us to contribute to the existing literature on the relative robustness of these two models.

The first finding of our study is that while the mean price and distance effects on provider choices by patients were present in both time periods, their presumed differences in heterogeneity are confirmed with MMNL testing. More precisely, during the second period, while the heterogeneity of choices based on price increased, mainly due to aging of patients, the heterogeneity of choices based on distance decreased.

The second finding is that, while both the MNL and the MMNL are able to predict price and distance effects with low heterogeneity, only the MMNL detected the price effect when heterogeneity was high. This finding suggests using caution when interpreting estimation results with the MNL in cases of high heterogeneity.

The remainder of the paper is organized as follows. Section 2 introduces the literature on the estimations of price and distance effects and explains why this study focusing on demand heterogeneity in rural China fills a gap in the literature. It also presents the issue on the choice between the MNL and MMNL models and the contributions with this study. Section 3 sets up the econometric model and describes the two-period samples. Section 4 analyzes the results and Section 5 concludes.

2. Issues on Heterogeneity in Healthcare Demand

2.1. Previous Work on Price and Distance Effects

There is abundant literature on healthcare provider choice in developing countries. As explanatory factors, income and price have always received special attention. Results obtained have led to contradictory conclusions. Several studies have found that income and price had significant elasticity on provider choice (Sahn & Stifel, 2000), while others have found that these factors were not important determinants whereas the perceived quality of health care had a greater effect. Several authors have undertaken systematic literature reviews on access to care.

One of the missing points in previous works is the absence of theoretical and empirical considerations on demand heterogeneity. The study on heterogeneity in relation to mean effects could significantly contribute to our understanding of healthcare demand.

This analysis on income and price effects also applies to studies on the distance effect over provider choice. This effect is important in developing countries where access to transport is more costly than in developed countries. The distance effect, especially its heterogeneity, has rarely been examined except by Borah (2006) in the case of India.

2.2. Why Rural China Provides a Good Case for Demand Heterogeneity

Between late 1980 and mid-2000, China achieved some deep social economical changes. The first was general income growth. The average GDP growth of China was 9% between 1989 and 2004, and GDP in 2004 was 8.07 times that of 1989 at current prices and 3.8 times at constant prices. Incomes of rural and urban households in 2004 were 4.88 and 6.86 times those of 1989, respectively, at current prices. Converted to incomes at constant prices, the

income of rural households in 2004 was 2.3 times that of 1989. In our sample, the average per capita income at constant prices and household assets in 2004-2006 are 2.4 and 3.1 times those in 1989-1993, respectively.

An obvious effect of general income growth would be that healthcare choices became more heterogeneous among patients because of the decrease of budget constraints. General income growth could also reduce distance effect and make the impact of distance on preference more heterogeneous because the choices become less constraint by the transports costs.

Another factor that affects choice heterogeneity is population aging. In 1990, 5.57% of the total population were over 65 years old. By 2005, this percentage had nearly doubled to 9.07%, with 9.48% and 8.12% for rural and urban populations, respectively. In our samples, with average age increased from 44 to 56 and the percentage of people over 65 doubled and their health naturally decreased, patients' provider choices would be, in general, less sensitive to price and more affected by some other factors. As elderly people tend not to like to travel, there is a stronger distance effect and reduced heterogeneity of distance impacts.

Along with income growth and aging, several sources of unobservable heterogeneity can potentially evolve and affect changes in demand heterogeneity: on the side of unobserved healthcare provider attributes, there are at least two: 1) factors in non-price competition and 2) transport accessibility. In particular, the perception of non-price competition by the local population is an unobserved variable for research. For transport accessibility, given that distance is an observed variable, the accessibility that varies with specific transportation conditions across the same type of healthcare provider is an unobserved factor.

On the side of personal preference variations of rural patients, there can be several sources. First, there are differences in judgments on the efficiency of Chinese medicine across patients. Second, there are differences in patient perceptions about the effectiveness of the same type of health providers due to their experiences in health care. Third, there are differences in the connections with personal relationship networks. Fourth, there is subjectivity in self-assessment of health. The social, cultural and psychological factors that shape self-assessment clearly vary from patient to patient.

To summarize, with the increases in income and age of our patient samples and enlargement of the choice set, all above-mentioned sources of heterogeneity are subject to significant change, and thus enlarge the extent of heterogeneity in patients' healthcare preferences. The impact of price effect can be expected to be more heterogeneous in the second period, but distance effect is expected to be uncertain, depending on which influence is more important: the income growth that reduces the distance effect but increases its heterogeneity, or the aging of the population that increases the distance effect and reduces its heterogeneity.

2.3. MNL and MMNL in the Presence of Heterogeneity

McFadden's choice model (McFadden, 1974) which relies on the conditional multinomial logit (MNL) model, has long been the leading tool for empirical studies. As the MNL model is based on the Independent and Identically Distributed (IID) assumption, and hence on the Independence of Irrelevant Alternatives (IIA) assumption, its failure to deal with heterogeneity is deemed capable of resulting in inferior model specification, spurious test results and invalid conclusions (Louviere, Hensher, & Swait, 2000). The mixed multinomial logit (MMNL) is similar to the MNL except that it allows parameter estimates to vary across individuals. According to the authors of the MMNL, in the presence of large-scale heterogeneity, the MMNL that relaxes IID will lead to marked improvement in estimations relating to the MNL. The MMNL leads to gaining generality, but the estimation simplicity that characterizes the MNL is lost. Thus, when the MNL is not biased, it is preferred.

To illustrate the difference between the MNL and the MMNL, consider the following utility function:

$$U_{ij} = \alpha_i Z_{ij} + \varepsilon_{ij} = (\alpha + \xi_i) Z_{ij} + \varepsilon_{ij} \quad (1)$$

where U_{ij} is the utility of individual i choosing state j , Z_{ij} and α_i represent all the observed factors and their parameters obtained from the model.

In the first equality of the equation, the coefficient α_i differs across individuals. Like the MNL, the MMNL assumes that the error terms, ε_{ij} , are IID. However, it relaxes the restriction that α is the same for each individual, allowing it to be stochastic instead.

The second equality in Equation 1 expresses another way to look at the MMNL. α_i is perceived as its mean, α , and a deviation around the mean, ξ_i , which differs across individuals. With non-zero error components, $\xi_i Z_{ij}$, utility becomes correlated across alternatives, which relaxes the IIA assumption. Thus, the MMNL incorporates taste variations across individuals. Through attributing each respondent to a random term, taste variations, unobserved heterogeneity in alternatives and unobserved heterogeneous choice sets are allowed (Bhat, 2000B).

The empirical issue is to compare their estimation performance. Most comparisons are on the willingness-to-pay (WTP) for various attributes of the alternatives in which mean coefficients are transformed in terms of WTP. Some authors, such as Horowitz (1980) and Van Den, Kroes, and Verhoef (2009) argued that random unobserved heterogeneity in the marginal utilities does not bias MNL estimates. Carlsson (2003) and Dahlberg and Matias (2003) reached the same conclusion and indicated that there are no conflicting signs with the two models and that the magnitude of the coefficients are very close, with just a few exceptions.

By contrast, Bhat (1998) found that WTP for all attributes are higher with the MMNL than with the MNL, indicating that the MNL underestimates WTP. Revelt and Train (1998) showed significant differences in WTP for some attributes while for others it showed none. Other researchers provided evidence that WTP is higher for some attributes but lower for others with the MMNL than with the MNL. Van Den et al. (2009) found that the MNL underestimates the WTP for travel time compared with the MMNL, but overestimates the WTP for other attributes. Train (1998) showed that the WTP is larger with the MMNL than with the MNL. He concluded that there is probably no general answer to whether or not the MNL gives correct estimates when heterogeneity is present.

Only a few works have reached the conclusion that in the presence of heterogeneity, MNL models lead to estimating failures. Persson (2002) suggested that model choice indeed has implications for the results since the welfare estimates from the two models differ quite remarkably. There are conflicting signs between the MNL and the MMNL.

To conclude, the debate around the potential bias using the MNL in the case of preference heterogeneity is not yet closed. While few works (Borah, 2006; Canaviri, 2007; Harris & Keane, 1998; Hole, 2008; Qian, Pong, Yin, Nagarajan, & Meng, 2009) have used the MMNL in healthcare demand studies, this study could offer a promising case and will provide new evidence on the relative performance between the two methods.

3. Estimation Method and Data

3.1. Model Specification

Let the utility of a patient $i \in [1, I]$ be a function of health status, h , and non-health consumption, x .

$$U = U(h_i, x_i) \tag{2}$$

Health status, h , is determined by the quantity and quality of health care (C), other health inputs (e.g., sanitation) and food consumption (F); and individual attributes such as age, gender, education, state of insurance, and asset (R).

$$h_i = h(C_i, F_i, R_i) \tag{3}$$

Healthcare demand is a function of the price of health care (p) and the distance to the healthcare provider (D). The importance of D is that distance not only implies cost of access, but also reflects to some extent the reputation and quality of providers.

$$C_i = C(p, D) \tag{4}$$

Finally, the other health input, F , is a function of expenditures on these inputs (E_i).

$$F_i = F(E_i) \tag{5}$$

With Equation 2 to 5, we get the indirect utility function expressed in Equation 6 in the case where individual i chooses healthcare provider j , and in which $y_i - p_{ij} - E_i$ is the budget for non-health consumption (y is income).

$$V_{ij}^* = U(h(C_{ij}(p_{ij}, D_{ij}), F(E_i), R_i), y_i - p_{ij} - E_i) \tag{6}$$

Among the healthcare provider alternatives, the patient will choose the one that maximizes his/her indirect utility function. The choice rule is expressed by Equation 7.

$$\begin{aligned} V_{ij} &= 1, \text{ if } V_{ij}^* = \text{Max}(V_{i1}^*, V_{i2}^*, \dots, V_{ij}^*) \\ V_{ij} &= 0, \text{ otherwise} \end{aligned} \tag{7}$$

To make the model amenable to econometric estimation, we must define a functional form of the above indirect utility function. This is expressed by Equation 8 in which the first term on the right is the deterministic component of utility in the function of the above-defined four types of attributes, and the second term is a disturbance term. The term E_i appeared in Equations 5 and 6 is now unobserved and is treated as one part of the error term.

$$V_{ij} = V_{ij}^*(p_{ij}, D_{ij}, y_i, R_i) + \varepsilon_{ij} \tag{8}$$

Equation 8 must be parameterized to allow estimations. The first term can be rewritten as:

$$V_{ij}^*(\cdot) = Z_{ij}\beta_z + X_i\beta_x \tag{9}$$

The X variables are patient-specific characteristics such as age, marital status, insurance status and income. The Z variables are alternative healthcare provider-specific characteristics such as distance, price, healthcare quality and so on. With these defined variables, we get

$$V_{ij} = \alpha_j + \beta_1 p_{ij} + \beta_2 D_{ij} + \beta_3 y_i + \beta_4 R_i + \varepsilon_{ij} \tag{10}$$

The variable p , the healthcare price, and D , the distance to healthcare provider are two provider-specific variables. The y , income and R , individual attributes other than income, are patient-specific variables. Thus, in our econometric estimations, the coefficients of p and D are kept constant across options while those of y and all components of R vary across options.

If Equation 10 is estimated with the MNL, the basic form of the MMNL, and with alternative specific constants α_j and attributes x_{ij} (here, x represents both z and x variables in the Equation 9 the result will be:

$$\text{Prob}(j) = \frac{\exp(\alpha_j + \beta_j' x_{ij})}{\sum_{q=1}^J \exp(\alpha_q + \beta_q' x_{iq})} \tag{11}$$

The difference between the MMNL and the MNL is that in the former, one part of the coefficients is random; in the latter, all coefficients are non-random. In Equation 11, β_j' is composed of β_{ji} with

$$\beta_{ji} = \begin{cases} \beta_j + \sigma_j \eta_{ji} & \text{if random} \\ \beta_j' & \text{if non-random} \end{cases} \tag{12}$$

where β_j is the population mean, η_{ji} is the individual specific heterogeneity, with mean 0 and standard deviation 1, and σ_j is the standard deviation of the distribution of β_{ji} around β_j . The elements of β_{ji} are distributed randomly across individuals with fixed means.

We set both price and distance to healthcare providers as random variables. It would be interesting to estimate the heterogeneity in the preferences for both price and distance.

If the random terms are normally distributed,

$$\beta_{ki} \sim \text{Normal}[\beta_k + \delta_k' w_i, \sigma_k^2] \tag{13}$$

Equation 13 has useful empirical implications and we will return to them in discussing their application. As the usual choice, we will use the normal distribution. Finally, to make our model more realistic, we will allow the two random parameters to be correlated.

3.2. Data, Variables and Characteristics of the Samples

Data are from the CHNS database edited by the Carolina Population Center (CPC, University of North Carolina). The survey covers about 16,000 individuals from more than 3,000 households (about two-thirds from rural and one-third from urban populations) in nine representative provinces. It is a longitudinal survey with seven waves (1989, 1991, 1993, 1997, 2000, 2004, 2006, 2009, and 2011).

The reasons we do not use the data after 2006 are two. First, rural exodus has been accelerating since 2009. According to 2009 survey, in rural area, around 40% of household members left home and worked in cities.

Extending to 2009 may exacerbate the bias in comparisons. Second, healthcare reform could also affect demand heterogeneity. As health insurance reforms began in 2003 but took several years to implement and only reached a real impact after 2006, the obtained results could be interpreted as being impacted by insurance reforms.

We build two samples. Within each sample, income, healthcare prices and supply conditions were not meaningfully evolved, but between them these factors were substantially changed. The number of patients interviewed who were ill was smaller in the first waves than in the last (population aging appears to be the main cause). Thus, to keep some equilibrium between the two samples, we merged three time periods of two-year intervals (1989, 1991 and 1993) for the first sample and two time periods of two-year intervals (2004 and 2006) for the second sample, for a total of 2,117 and 2,594 observations, respectively. The first sample included individuals under 18 years old as the following waves did not. We conducted a logistic regression analog of the Chow test to check whether the healthcare demand of the under-18 differed from that of the over-18 (see (Demaris, 2004)). Results showed that the two models indeed differed. Consequently, observations of individuals under 18 were removed. Finally, our samples included 1,457 rural individuals who reported having been ill in 1989, 1991 or 1993, and 2,594 individuals who reported being ill in 2004 or 2006.

As our data panel included attrition and replacement, we checked the frequency of the patients and whether attrition was non-random. In the 1989-1993 sample, only 11.6% and 0.06% patients were surveyed two and three times; in the 2004-2006' sample, 16.3% patients were surveyed twice. CHNS data collectors have not given more details on attrition. Nevertheless, as Deaton (1997, 19-20) stated, the rate of refusal of participation is lower in developing countries. It must be still lower in rural China since political institutions exert strong control. Thus, we attribute lack of participation on the part of villagers to their physical absence, their moves or their deaths. Therefore, attrition can be regarded as random.

Table 1 presents all variables used and their definitions.

Table-1. Definition of the variables.

Village-C (V)	=1 if the choice of treatment is village clinic; =0 otherwise.
Town-C (T)	=1 if the choice of treatment is township health center; =0 otherwise.
County-H (C)	=1 if the choice of treatment is county or higher level city hospital; =0 otherwise.
Other-type (O)	=1 if the source of treatment is pharmacy, private clinic and other clinic; =0 otherwise.
Self-treatment (S)	=1 if treatment by self is chosen; =0 otherwise.
P_j	Medical expense at constant prices of alternative j after eventual reimbursement by insurance multiplied by 10^{-3} ; j=V, T, C, O, S. The expense of self-care is assumed =0.
Dist0 _j	=1 if distance <0.5 km; =0 otherwise; j=V, T, C, O.
Dist1 _j	=1 if distance >=0.5 km & <3; =0 otherwise; j=V, T, C, O.
Dist2 _j	=1 if distance >=3 km & <10km; =0 otherwise; j=V, T, C, O.
Dist3 _j	=1 if distance >=10 km; =0 otherwise; j=V, T, C, O.
Age	Age of the patient in the wave.
Female	=1 if the patient is female; =0 if male.
Marital	=1 if the patient is married; =0 otherwise.
Edu_level	=1 graduated from primary school; =2 lower middle school degree; =3 upper middle school degree; =4 technical or vocational degree; =5 university or college degree; =6 master's degree or higher.
Nonfarm_job	=1 if the patient's job is not farmer; =0 otherwise.
Farmer	=1 if the patient's job is farmer; =0 otherwise.
No_job	=1 if the patient has not job; =0 otherwise.
No_insured	=1 if the patient is not insured; =0 otherwise.
Urban_insurance	=1 if for family members, the patient's insurance is one of the following types: commercial, free medical, workers compensation, and for the members that are urban employee, pass-way model, block model, catastrophic disease; =0 otherwise.
Cooperative_insurance	=1 if the patient's insurance type is rural cooperative; =0 otherwise.
Other_insurance	=1 if the patient's insurance is other than Urban_insurance and Cooperative_insurance (they include among others Health insurance for women and children, EPI (expanded program of immunization) and insurance for children); =0 otherwise.
Severity	=1 if the illness or injury was answered by the patient as not severe; =2 somewhat severe; =3 quite severe.
Fever	=1 if individual suffered from fever; =0 otherwise.
Chronic	=1 if individual suffered from chronic diseases; =0 otherwise.
Other_diseases	=1 if individual suffered from diseases other than fever and chronic diseases; =0 otherwise.
Hhsize	The number of the household members.
Income	The annual per capita income at constant prices of the household multiplied by 10^{-3} .
Asset	The annual household value of the asset index.
Rural_popu_rate	The share of the rural employees in total labor of the village.
Village_size	The household number of the village multiplied by 10^{-3} .
Suburb	=1 if the village is near a city; =0 otherwise.

Notes: 1) data come from the CHNS database; 2) the first five items (V, T, C, O, S) concern the dependent variable spread in a selected set of healthcare providers; 3) all of the following variables concern the independent variables; 4) with the exception of the first five and the last three, all of the remaining variables are individual-specific attributes; 5) the last three variables were used to take into account environmental features; Rural_popu_rate is a proxy of the development level of the village, Village_size is a proxy of the village clinic's size, and Suburb reflects the proximity of the village to the urban medical infrastructure; 6) Asset and P_j are built with the method described in section 3.2.

The CHNS database provides household per capita annual income at a constant price. As the impact of a household's income and assets on their healthcare provider choice can be quite different, we built an asset index and simultaneously used income and asset to measure income and wealth effects. Following several authors (Filmer &

Kinnon, 2008; Sahn & Stifel, 2000) we used the 9 items, with 4 to 8 modalities for each, and then employed principal components analysis to derive weights (Filmer & Pritchett, 2001) for the asset index.¹

We also wrestled with how to compensate for missing data on healthcare prices. The MMNL requires the prices of all alternative providers, while only the prices of the providers that the patients effectively visited were recorded in the survey. Thus, the prices of alternative providers that patients did not visit needed to be imputed. Following Gertler, Locay, and Sanderson (1987); Gertler and Van Der (1990) and Borah (2006) we used the Stata ICE program created by Royston (2004) to impute the lacking price data. All reported prices were converted at constant prices using the weights given by the CHNS data provider. The chosen predictors of prices included 16 variables: *Age*, *Female*, *Marital*, *Edu_level*, *Nonfarm_job*, *Farmer*, *Income*, *Severity*, *Year*, *Province*, *Urban_insurance*, *Cooperative_insurance*, *Other_insurance*, *Fever*, *Chronic*, and *Hospitalized* (=1 if hospitalized; =0 otherwise). The descriptive statistics of actual plus imputed prices by type of provider are presented in Table 2.

Table 2 calls for some brief comments. First, comparing the two samples, income, asset, education level, healthcare price, the share of patients with insurance, and village size were meaningfully increased over time. Two other increases, linked with population aging, were the *No_job* (composed notably by the retired), and *Chronic*. Second, in general, the share of the big and middle hospitals (township health centers and county hospitals) in chosen healthcare providers increased from 30% to 32% in favor of county hospitals (from 9% to 18%) and to the detriment of township centers (from 21% to 14%). The share of small clinics (village clinics in 1989-1993 and village clinics plus *Other_type* in 2004-2006) decreased from 48% to 33%. Their 15% reduction appears to have benefited self-treatment, which grew 14%.²

Table-2. Descriptive statistics.

Sample distribution by provider choice	1989-1993 (n=1457)				2004-2006 (n=2594)			
	Mean	SD	Min	Max	Mean	SD	Min	Max
Village-C (V)	0.48	0.50	0	1	0.22	0.41	0	1
Town-C (T)	0.21	0.41	0	1	0.14	0.35	0	1
County-H (C)	0.09	0.29	0	1	0.18	0.39	0	1
Other-type (O)					0.11	0.32	0	1
Self-treatment (S)	0.21	0.41	0	1	0.35	0.48	0	1
P_V	0.074	0.078	0	0.477	0.096	0.079	0	0.598
P_T	0.159	0.162	0	0.859	0.207	0.169	0	1.166
P_C	0.466	0.617	0	3.506	0.651	0.597	0	3.808
P_O					0.204	0.318	0	3.972
Dist0_V	1	0	1	1	1	0	1	1
Dist0_T	0.40	0.49	0	1	0.48	0.50	0	1
Dist1_T	0.39	0.49	0	1	0.36	0.48	0	1
Dist2_T	0.21	0.40	0	1	0.15	0.36	0	1
Dist3_T	0	0	0	0	0	0	0	0
Dist0_C	0.13	0.34	0	1	0.23	0.41	0	1
Dist1_C	0.16	0.37	0	1	0.22	0.42	0	1
Dist2_C	0.22	0.41	0	1	0.25	0.43	0	1
Dist3_C	0.49	0.50	0	1	0.30	0.46	0	1
Dist0_O					0.63	0.48	0	1
Dist1_O					0.26	0.44	0	1
Dist2_O					0.09	0.29	0	1
Dist3_O					0.02	0.15	0	1
Age	44.47	15.41	18	92	55.88	15.12	18	97
Female	0.53	0.50	0	1	0.57	0.49	0	1
Marital	0.84	0.37	0	1	0.80	0.40	0	1
Edu_level	0.98	1.06	0	5	1.17	1.21	0	6
Nonfarm_job	0.26	0.44	0	1	0.13	0.35	0	1
Farmer	0.60	0.49	0	1	0.35	0.48	0	1
No_job	0.14	0.35	0	1	0.51	0.50	0	1
No_insured	0.80	0.40	0	1	0.64	0.48	0	1
Urban_insurance	0.15	0.36	0	1	0.10	0.30	0	1
Cooperative_insurance	0.03	0.17	0	1	0.25	0.43	0	1
Other_insurance	0.02	0.13	0	1	0.01	0.10	0	1
Severity	1.71	0.70	1	3	1.70	0.67	1	3
Fever	0.35	0.48	0	1	0.26	0.44	0	1
Chronic	0.13	0.33	0	1	0.34	0.47	0	1
Other_diseases	0.52	0.50	0	1	0.40	0.49	0	1
Hhsize	4.40	1.50	1	13	3.66	1.69	0	13
Income	2.91	2.26	0.45	22.20	7.03	8.03	0.18	210.95
Asset	0.39	0.77	-1.05	3.08	1.20	0.96	-0.62	3.87
Rural_popu_rate	0.52	0.34	0	1	0.41	0.30	0	1
Village_size	0.66	0.74	0.03	6.00	1.01	1.19	0.04	8.00
Suburb	0.28	0.45	0	1	0.24	0.43	0	1

¹ The items are 1) Drinking water (4 choices); 2) Toilet facilities (8 choices); 3) Kind of lighting (5 choices); 4) Kind of fuel for cooking (8 choices); 5) Type of ownership of house (6 choices); 6) Ownership of electrical appliances and other goods (the number of appliances varied between 15 to 18 according to the periods of survey, and this information was absent only in 1989); 7) Means of transportation (5 types); 8) Type of farm machinery (5 types); and 9) Household commercial equipment (6 types). The coefficients of correlation between the obtained *Asset* and *Income* were 0.29 for both periods (1989-1993 and 2004-2006) and were significant at 1%.

² *Other_type* generally includes very small healthcare providers that practice Chinese medicine near a pharmacy, or the retired doctors that open a clinic with elementary equipment. They are far from being a growing alternative force to the three principal healthcare providers.

Notes: 1) data come from the CHNS database; 2) the first five items (*V, T, C, O, S*) concern the dependent variable spread in a selected set of healthcare providers; 3) all of the following variables concern the independent variables; 4) with the exception of the first five and the last three, all of the remaining variables are individual-specific attributes; 5) the last three variables were used to take into account environmental features; *Rural_popu_rate* is a proxy of the development level of the village, *Village_size* is a proxy of the village clinic's size, and *Suburb* reflects the proximity of the village to the urban medical infrastructure; 6) *Asset* and *P_j* are built with the method described in section 3.2.

4. Estimation Results

Tables 3 and 4 contain the regression results of the 1989-1993 and 2004-2006 data samples, respectively. The software used for both models is Nlogit. MMNL estimates are obtained with 100 Halton draws.

4.1. MNL versus MMNL

For both periods, the MMNL yields higher likelihood values and provides improved fits over the MNL (likelihood ratio test is significant at less than 0.01), indicating that the explanatory power of the mixed logit is greater than with the standard logit. Two other measures commonly used to compare competing regression models are the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). These measures account for both the goodness of fit of the model and its parsimony. Each measure penalizes a larger model for using additional degrees of freedom while rewarding improvements in goodness of fit. The BIC places a higher penalty on using degrees of freedom than the AIC. According to the AIC, the MMNL is better while according to the BIC, the MNL is preferred. Thus, the results are not conclusive.

Assuming individual rationality, a negative price effect is expected. Table 3 shows that according to both models in 1989-1993, there were clear price effects. The estimated means are -0.377 (significant at 5%) and -1.606 (significant at 1%) for the MNL and the MMNL, respectively. The coefficients of random variables in the MMNL are consistently of greater magnitude (in absolute terms) than those from the MNL. Revelt and Train (1998) obtained similar results. According to the authors, this is not surprising since a random parameter model decomposes the unobserved portion of utility and normalizes parameters on the basis of part of the unobserved portions.

Table-3. Regression results on 1989-1993 sample.

	MNL			MMNL		
	Village-C	Town-C	County-H	Village-C	Town-C	County-H
PRICE	-0.377 (0.169)**			-1.606 (0.414)***		
Distance1	0.049 (0.155)			-0.135 (0.311)		
Distance2	-0.346 (0.197)*			-0.494 (0.287)*		
Distance3	-0.407 (0.312)			-4.323 (1.749)**		
intercept	0.102 (0.596)	-1.467 (0.727)**	-2.86 (0.943)***	0.091 (0.605)	-1.520 (0.780)*	-3.349 (1.257)***
Age	-0.018 (0.006)***	-0.016 (0.007)**	-0.04 (0.009)	-0.018 (0.006)***	-0.016 (0.007)**	-0.005 (0.011)
Edu_level	0.109 (0.087)	0.070 (0.103)	0.027 (0.130)	0.114 (0.088)	0.072 (0.110)	-0.033 (0.172)
Women	0.432 (0.152)***	0.300 (0.181)*	0.268 (0.235)	0.440 (0.153)***	0.270 (0.193)	0.401 (0.323)
Hhsize	-0.018 (0.048)	0.051 (0.058)	0.049 (0.074)	-0.018 (0.049)	0.047 (0.062)	0.111 (0.095)
Asset	0.382 (0.149)**	0.424 (0.173)**	0.110 (0.220)	0.440 (0.153)***	0.424 (0.184)**	-0.195 (0.301)
Income	-0.027 (0.037)	-0.023 (0.042)	0.059 (0.049)	-0.026 (0.037)	-0.023 (0.045)	0.130 (0.072)*
Severity	0.443 (0.111)***	0.786 (0.129)***	0.938 (0.167)***	0.484 (0.113)***	0.910 (0.148)***	0.977 (0.220)***
Marital	0.356 (0.189)*	0.367 (0.231)	0.576 (0.310)*	0.355 (0.190)*	0.383 (0.248)	0.636 (0.399)
Urban_insurance	-0.082 (0.288)	0.222 (0.342)	0.107 (0.394)	-0.067 (0.293)	0.276 (0.364)	0.221 (0.489)
Cooperative_insurance	0.393 (0.532)	0.294 (0.590)	0.947 (0.702)	0.428 (0.537)	0.307 (0.629)	1.623 (0.926)*
Nonfarm_job	0.459* (0.276)	0.243 (0.331)	0.240 (0.383)	0.462 (0.280)*	0.196 (0.354)	0.244 (0.479)
Farmer	0.129 (0.244)	-0.035 (0.286)	-0.149 (0.362)	0.120 (0.248)	-0.064 (0.310)	-0.273 (0.510)
Fever	-0.184 (0.153)	-0.518 (0.186)***	-0.624 (0.250)**	-0.218 (0.155)	-0.654 (0.204)***	-0.721 (0.336)**
Chronic	0.043 (0.231)	-0.229 (0.276)	-0.103 (0.337)	0.071 (0.233)	-0.308 (0.296)	0.280 (0.423)
Rural_popu_rate	0.403 (0.341)	0.382 (0.401)	-0.068 (0.533)	0.398 (0.344)	0.315 (0.429)	0.088 (0.749)
Village_size	0.227* (0.134)*	-0.107 (0.193)	0.390 (0.158)**	0.237 (0.138)*	-0.123 (0.206)	0.520 (0.181)***
Suburb	-0.088 (0.272)	-0.710 (0.337)**	-0.240 (0.423)	-0.070 (0.276)	-0.718 (0.369)*	-0.091 (0.571)
Province dummies	Yes (omitted)	Yes (omitted)	Yes (omitted)	Yes (omitted)	Yes (omitted)	Yes (omitted)
Wave dummies	Yes (omitted)	Yes (omitted)	Yes (omitted)	Yes (omitted)	Yes (omitted)	Yes (omitted)

SD of parameter distributions		
PRICE		2.100 (0.433)***
Distance1		1.083 (0.700)
Distance2		0.795 (0.545)
Distance3		3.935 (1.070)***
N	1457	1457
Log-likelihood	-1663.084	-1648.234
McFadden Pseudo R2		0.184
Chi Squared	248.499	743.195
Significance level	0.00000	0.00000
AIC	3496.169	3486.467
BIC	3945.320	3988.460

Notes: 1) data come from the CHNS database; 2) *Village-C*, *Town-C*, and *County-H* concern the dependent variable spread in a selected set of healthcare providers; 3) with the exception of the last three, all of the remaining explanatory variables are individual-specific attributes, and their definitions are made in Table 1; 4) the last three variables were used to take into account environmental features. *Rural_popu_rate* is a proxy of the development level of the village, *Village_size* is a proxy of the village clinic's size, and *Suburb* reflects the proximity of the village to the urban medical infrastructure; 5) *Asset* and *Price* are built with the method described in section 3.2; 6) Standard error is in parentheses. *** indicates significance at 1%; ** indicates significance at 5%; and * indicates significance at 10%.

Table-4. Regression results on 2004–2006 sample.

	MNL				MMNL			
	Village-C	Town-C	County-H	Other-Type	Village-C	Town-C	County-H	Other-Type
PRICE	0.010 (0.088)				-0.409 (0.191)**			
Distance1	-0.486 (0.095)***				-0.831 (0.205)***			
Distance2	-0.508 (0.124)***				-0.698 (0.193)***			
Distance3	-0.872 (0.170)***				-1.148 (0.235)***			
Intercept	-1.070 (0.510)**	-1.351 (0.576)**	-1.239 (0.551)**	-1.114 (0.639)*	-1.073 (0.514)**	-1.449 (0.619)**	-1.283 (0.626)**	-1.260 (0.680)*
Age	-0.0005 (0.005)	-0.006 (0.006)	-0.015 (0.005)***	-0.010 (0.006)*	-0.001 (0.005)	-0.006 (0.006)	-0.016 (0.006)***	-0.010 (0.006)*
Edu_level	-0.057 (0.066)	-0.116 (0.073)	-0.101 (0.064)	-0.075 (0.075)	-0.057 (0.067)	-0.119 (0.078)	-0.106 (0.072)	-0.066 (0.078)
Women	0.083 (0.127)	-0.051 (0.142)	-0.038 (0.133)	-0.098 (0.150)	0.083 (0.128)	-0.059 (0.151)	-0.050 (0.150)	-0.091 (0.157)
Hhsize	-0.007 (0.039)	0.009 (0.044)	-0.015 (0.041)	-0.113 (0.049)**	-0.007 (0.039)	0.010 (0.046)	-0.014 (0.046)	-0.125 (0.051)**
Asset	-0.143 (0.092)	0.089 (0.103)	0.202 (0.092)**	0.059 (0.110)	-0.142 (0.092)	0.085 (0.110)	0.241 (0.104)**	0.086 (0.116)
Income	-0.004 (0.009)	-0.005 (0.011)	-0.003 (0.009)	-0.011 (0.012)	-0.002 (0.008)	-0.008 (0.012)	-0.002 (0.010)	-0.012 (0.013)
Severity	0.388 (0.095)***	0.930 (0.103)***	1.124 (0.098)***	0.621 (0.112)***	0.406 (0.097)***	1.042 (0.115)***	1.252 (0.118)***	0.678 (0.119)***
Marital	0.068 (0.151)	0.135 (0.174)	0.281 (0.164)*	0.154 (0.182)	0.076 (0.153)	0.148 (0.186)	0.376 (0.184)**	0.177 (0.190)
Urban_Insurance	-0.354 (0.294)	0.158 (0.281)	0.456 (0.211)**	0.136 (0.278)	-0.347 (0.296)	0.146 (0.302)	0.487 (0.240)**	0.179 (0.291)
Cooperative_insurance	0.257 (0.163)	-0.046 (0.190)	-0.199 (0.191)	0.131 (0.211)	0.247 (0.165)	-0.049 (0.203)	-0.236 (0.215)	0.118 (0.222)
Nonfarm_job	-0.039 (0.211)	0.045 (0.226)	-0.647 (0.211)***	-0.288 (0.241)	-0.055 (0.213)	0.081 (0.242)	-0.722 (0.237)***	-0.343 (0.253)
Farmer	0.082 (0.145)	-0.073 (0.163)	-0.387 (0.171)**	-0.070 (0.182)	0.083 (0.146)	-0.051 (0.175)	-0.452 (0.191)**	-0.067 (0.191)
Fever	0.914 (0.144)***	0.383 (0.168)**	-0.638 (0.183)***	0.664 (0.170)***	0.906 (0.146)***	0.384 (0.180)**	-0.746 (0.203)***	0.702 (0.179)***
Chronic	-0.083 (0.143)	-0.138 (0.154)	-0.182 (0.137)	-0.414 (0.176)**	-0.075 (0.144)	-0.133 (0.165)	-0.189 (0.154)	-0.424 (-0.184)**
Rural_popu_rate	0.612 (0.288)**	-0.165 (0.329)	-0.036 (0.346)	-0.239 (0.376)	0.614 (0.291)**	-0.244 (0.352)	0.009 (0.386)	-0.325 (0.396)
Village_size	-0.104 (0.084)	0.026 (0.081)	0.083 (0.055)	-0.087 (0.077)	-0.103 (0.085)	0.031 (0.085)	0.096 (0.062)	-0.092 (0.081)
Suburb	-0.420 (0.224)*	-1.621 (0.263)***	-0.133 (0.223)	-0.116 (0.248)	-0.399 (0.227)*	-1.653 (0.277)***	-0.191 (0.255)	-0.141 (0.262)
Province dummies	Yes (omitted)	Yes (omitted)	Yes (omitted)	Yes (omitted)	Yes (omitted)	Yes (omitted)	Yes (omitted)	Yes (omitted)
Wave dummies	Yes (omitted)	Yes (omitted)	Yes (omitted)	Yes (omitted)	Yes (omitted)	Yes (omitted)	Yes (omitted)	Yes (omitted)
SD of parameter distributions								
PRICE					1.375 (0.261)***			
Distance1					1.292 (0.338)***			
Distance2					1.131 (0.373)***			

Distance3		0.853 (0.357)**
N	2594	2594
Log-likelihood	-3493.772	-3480.662
McFadden Pseudo R ²		0.166
Chi Squared	954.192	1388.440
Significance level	0.00000	0.00000
AIC	7211.545	7205.324
BIC	3945.320	3988.460

Notes: 1) data come from the CHNS database; 2) *Village-C*, *Town-C*, *County-H*, and *Other-Type* concern the dependent variable spread in a selected set of healthcare providers; 3) with the exception of the last three, all of the remaining explanatory variables are individual-specific attributes, and their definitions are made in Table 1; 4) the last three variables were used to take into account environmental features. *Rural_popu_rate* is a proxy of the development level of the village, *Village_size* is a proxy of the village clinic's size, and *Suburb* reflects the proximity of the village to the urban medical infrastructure; 5) *Asset* and *Price* are built with the method described in section 3.2; 6) Standard error is in parentheses. *** indicates significance at 1%; ** indicates significance at 5%; and * indicates significance at 10%.

The most striking result is that in the 2004–2006 period, unlike with the MMNL, price effect disappeared with the MNL. As shown in Table 4, the coefficient of *Price* is 0.01 and is no longer significant. Nevertheless, in the MMNL model, price effect was present with a coefficient of -0.409 (significant at 5%). Principally on the basis of this difference, we judge that the MNL analysis does not produce logical or consistent signs for price estimates.

The second random variable is distance to healthcare provider. In accordance with the analysis on price effect, the coefficients of *Distances* with the MMNL are higher than with the MNL. However, unlike the price effect in sign and significance, there is not meaningful divergence of distance effects between the two models (except *Distance3* in 1989–1993). In 1989–1993, *Distance1* is insignificant in both the MMNL and the MNL. In 2004–2006, nevertheless, all Distance dummies are significantly negative, indicating that distant healthcare providers are less likely to be chosen.

Given that the variables other than price and distance are patient-specific and do not vary by healthcare provider, they cannot be assigned with a random term; their coefficients with the MMNL and the MNL are similar in sign and extent. As their interpretations are out of the scope of either study on heterogeneity or comparison between the MMNL and the MNL, we choose to show them without comments.

4.2. The Importance of Heterogeneity Analysis in Interpreting Price and Distance Effects

The MMNL provides information on the heterogeneity of provider choice in price and distance. In Tables 3 and 4, the standard deviations of price parameters are 2.1 and 1.375 for the two samples, respectively and both are significant at 1%, indicating that parameters indeed vary in the population. Following Equation 13, we can easily calculate the level of this heterogeneity with the criterion of the percentage of patients for which the coefficients of *Price* are above zero; the result is presented in Table 5. In 1989–1993, while about 80 % of patients followed the rule that when the price rises, the demand falls, it is not observed for 20% of patients. That percentage rose to 38.30% in 2004–2006, indicating that heterogeneity in price preferences meaningfully increased in the second period. Meanwhile, whereas heterogeneity in *Distance2* was unchanged, both *Distance1* and *Distance3* have decreased (from 45.03% to 26%, and 13.59% to 8.92%, respectively).

Table-5. Heterogeneity measured by percentage of patients of which the coefficients of *Price* or *Distances* >0.

Period	1989–1993	2004–2006
Price	22.22%	38.30%
Distance1	45.03%	26.00%
Distance2	26.73%	26.86%
Distance3	13.59%	8.92%

Notes: 1. data come from the CHNS database; 2) *Price* and *Distances* are built with the method described in section 3.2, and their coefficients are found in Tables 3 and 4; 3) Calculated with Equation 13 and using the mean coefficients and SD of parameter distributions from Table 3 and 4.

It is time to explain why the information on choice heterogeneity has crucial importance for our study on price and distance effects, and to what extent the provision of this information allows us to avoid biased interpretations of these effects. Comparing the absolute levels of the coefficients of *Price* between two periods in Tables 3 and 4, we observe a significant decrease of the mean price effect (from -1.606 to -0.409). With this decrease, the appealing policy would be to allocate more resources to high quality services.

However, without checking the heterogeneity level, this policy implication could be problematic. The coefficients of variation (SD/mean), indeed, rise from 1.3 to 3.2, indicating a rise of the heterogeneity level. But, as indicated in Table 5, the percentage of patients who followed the rule that demand falls as price rises only decreased from 77.8% to 61.7%. These indicators on heterogeneity suggest the existence of a majority of patients with modest incomes and imply a persistent need to lower the price of services. In this case, the policy aiming to favor high- and high quality services would be erroneous.

Assume now an evaluation of distance effect for the purpose of a geographical allocation of healthcare resources. In general, a weaker mean distance effect suggests a more geographically concentrated distribution (focusing on several modern and large hospitals to serve the distant population). By contrast, the high mean distance effect indicates a demand for proximity and hence a geographically more decentralized structure. Income growth and aging oppositely affect the mean level of the distance effect. Hence, the distance effect depends on which force: income growth or aging, is more prevalent. Merely focusing on mean effect would likely lead to a biased policy.

If the influence of income growth is higher than aging, the mean distance effect becomes weaker and the concentration solution with more large hospitals could be a relevant answer, because traveling for a longer distance

matters less. But it could be wrong if income inequality becomes high, leading to a larger demand for nearer and cheaper healthcare services.

In the case where the influences of income growth and aging are equal so that the mean level of distance effects is the same, the mean level would be unable to give any insight in the necessary direction of policy change. In such a case, knowing the heterogeneity level and being able to quantitatively identify the sources of the heterogeneity make identifying appropriate solutions possible. For instance, if low heterogeneity of the distance effect is caused by either aging or income inequality (leading to a state of more “poor”), the efficient solution would be to fund more local small healthcare structures. On the contrary, if high heterogeneity is caused by increasing preferences in quality, concentrating on a solution that provides large and modern health care structures would be efficient.

5. Conclusions

We found that over the two periods, mainly due to income increase and aging, the price effect became weaker and more heterogeneous; the distance effect became stronger and less heterogeneous. From a methodological perspective, in both periods, the MMNL provides improved fits over the MNL. This result suggests that, in the presence of important individual preference heterogeneity, researchers should be cautious when interpreting the estimation results produced with the MNL model and that the MMNL model is the most suitable to study healthcare demand.

References

- Bhat, C. R. (2000B). Flexible model structures for discrete choice analysis”, in Handbook of transport modelling, D. Hensher, K.J. Button (eds) (pp. 71-90). Pergamon: Oxford.
- Bhat, C. R. (1998). Accommodating variations in responsiveness to level-of-service measures in travel mode choice modeling. *Transportation Research Part A: Policy and Practice*, 32(7), 495-507. Available at: [https://doi.org/10.1016/s0965-8564\(98\)00011-1](https://doi.org/10.1016/s0965-8564(98)00011-1).
- Borah, B. J. (2006). A mixed logit model of health care provider choice: Analysis of NSS data for rural India. *Health Economics*, 15(9), 915-932. Available at: <https://doi.org/10.1002/hec.1166>.
- Canaviri, J. (2007). A random parameter logit model for modeling healthcare provider choice in Bolivia. Oklahoma State University, MPRA Paper, No. 3263.
- Carlsson, F. (2003). The demand for intercity public transport: the case of business passengers. *Applied Economics*, 35(1), 41-50. Available at: <https://doi.org/10.1080/00036840210158921>.
- Dahlberg, M., & Matias, E. (2003). Relaxing the IIA assumption in locational choice models: A comparison between conditional logit model, mixed logit, and multinomial probit models. Working Paper 2003, 9, Department of Economics Uppsala University.
- Demaris, A. (2004). *Regression with social data modeling continuous and limited response variables*. Hoboken, New Jersey: John Wiley & Sons, Inc.
- Filmer, D., & Pritchett, H. (2001). *Assessing asset indices*. Policy Research Working Paper, n° 4605, The World Bank.
- Gertler, P., Locay, L., & Sanderson, W. (1987). Are user fees regressive?: The welfare implications of health care financing proposals in Peru. *Journal of econometrics*, 36(1-2), 67-88.
- Gertler, P., & Van Der, G. J. (1990). *The willingness to pay for medical care: evidence from two developing countries*. Baltimore, MD: Johns Hopkins University Press for the World Bank.
- Harris, K. M., & Keane, M. P. (1998). A model of health plan choice: Inferring preferences and perceptions from a combination of revealed preference and attitudinal data. *Journal of Econometrics*, 89(1-2), 131-157. Available at: <https://doi.org/10.4159/9780674726307-005>.
- Hole, A. R. (2008). Modelling heterogeneity in patients' preferences for the attributes of a general practitioner appointment. *Journal of Health Economics*, 27(4), 1078-1094. Available at: <https://doi.org/10.1016/j.jhealeco.2007.11.006>.
- Horowitz, J. (1980). The accuracy of the multinomial logit model as an approximation to the multinomial probit model of travel demand. *Transportation Research Part B: Methodological*, 14(4), 331-341. Available at: [https://doi.org/10.1016/0191-2615\(80\)90013-2](https://doi.org/10.1016/0191-2615(80)90013-2).
- Louviere, J., Hensher, D., & Swait, J. (2000). *Stated choice methods and analysis*. Cambridge: Cambridge University Press.
- McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior”, In *Frontiers in Econometrics*” Ed. “ P. Zarembka (pp. 105-142). New York: Academic Press.
- Persson, T. H. (2002). Welfare calculations in models of the demand for sanitation. *Applied Economics*, 34(12), 1509-1518. Available at: <https://doi.org/10.1080/00036840110102752>.
- Qian, D., Pong, R. W., Yin, A., Nagarajan, K., & Meng, Q. (2009). Determinants of health care demand in poor, rural China: the case of Gansu Province. *Health Policy and Planning*, 24(5), 324-334. Available at: <https://doi.org/10.1093/heapol/czp016>.
- Revelt, D., & Train, K. (1998). Mixed logit with repeated choices: households' choices of appliance efficiency level. *Review of Economics and Statistics*, 80(4), 647-657.
- Royston, P. (2004). Multiple imputation of missing values. *The Stata Journal*, 4(3), 227-241.
- Sahn, D. E., & Stifel, D. C. (2000). Poverty comparisons over time and across countries in Africa. *World Development*, 28(12), 2123-2155.
- Train, K. (1998). Recreation demand models with taste differences over people. *Land Economics*, 74(2), 230-239.
- Van Den, B., V., Kroes, E., & Verhoef, E. T. (2009). Choice of train ticket: A study on Dutch travelers. VU University Working Paper.