CrossMark

RESEARCH ARTICLE

Medical expenditure in urban China: a quantile regression analysis

Jianmei Zhao[1] · Hai Zhong[1]

Abstract Many countries have been trying to expand their public health insurance coverage in recent years. To achieve two fundamental policy goals—equity in health care utilization and control of health care costs—policymakers need a better understanding of the underlying determinants of individual health care expenditure beyond the results of mean regressions. In this paper, we apply a quantile regression method to investigate the heterogeneous effects of various determinants of medical expenditure in China. Comparing with the average effects, we find that health care expenditures at the upper end of the distribution are under stronger influences of need factors such as poor health status, and weaker influences of socioeconomic factors and insurance status. On the other hand, health care expenditures at the lower end of the distribution are under stronger influences of socioeconomic factors and insurance status, and weaker influences of need factors. Our study may provide useful information to policymakers for the optimal design of their health care systems, and it may be of particular interests to the health policymakers in China, where is currently still in a period of reshaping its health-care system.

Keywords Medical expenditure · Two-part model · Quantile regression · China

JEL Classification I11 · I18 · H31

Introduction

In recent years, many countries around the world (for example, China, the United States, Kenya, Ghana, Colombia, Mexico, Thailand, Vietnam, various former members of the Soviet

✉ Jianmei Zhao
   jianmei@cufe.edu.cn

   Hai Zhong
   haizhong@cufe.edu.cn

[1] China Academy of Public Finance & Public Policy, Central University of Finance & Economics,
   39 South College Road, Haidian District, Beijing 100081, People's Republic of China

Union, and India) have been trying to expand their public health insurance coverage (Wagstaff et al. 2009; Yip and Hsiao 2009; Kowalski 2010; Duarte 2012). Two fundamental concerns for policymakers on the shaping of their future health care systems are equity in health care utilization and control of health care costs. To achieve either goal, policymakers need a better understanding of the underlying determinants of individual health care expenditure.

In the past several decades, there has been a growing empirical literature on the factors that determine health care expenditures, which include prices of health care and individual characteristics such as age, sex, education, health status, income, and socioeconomic status (e.g., Musgrove 1983; Manning et al. 1987; Wedig 1988; Kenkel 1990; Hitiris and Posnett 1992; Chiappori et al. 1998; DiMatteo and DiMatteo 1998; Eichner 1998; DiMatteo 2003; Vera-Hernandez 2003; Mocan et al. 2004; Kaestner and Dave 2006; Kowalski 2010; Duarte 2012). One limitation of this literature is that most of the evidence collected is based on a mean regression approach.[1] The observed average effect may hide many complex behaviors. In fact, the role of determinant factors at different points of the medical expenditure distribution may have important implications to equity in health care utilization and cost control.

In this paper, we apply a quantile regression method to quantify the heterogeneous impacts of various factors at different segments of the medical expenditure distribution in China. Comparing with the average effects, we find that health care expenditures at the upper end of the distribution are under stronger influences of need factors such as health status, and weaker influences of socioeconomic factors and insurance status. On the other hand, health care expenditures at the lower end of the distribution are under stronger influences of socioeconomic factors and insurance status, and weaker influences of need factors.

Our results may provide useful information to policymakers for the optimal design of their health care systems. For example, as emphasized in Cutler and Zeckhauser (2000), an important difference between real world and optimal insurance policies is that the former almost invariably have a constant coinsurance rate, whereas the latter do not. Several studies show that from the perspective of the tradeoff between risk-reduction and moral hazard, non-linear copayment schedule can be substantially superior (e.g., Blomqvist 1997; Ellis and Manning 2007; Felder 2008). Our results suggest that a regressive co-payment schedule might be preferred due to another concern—equity in health care utilization. Furthermore, this study may have particular interests to the health policymakers in China, where is still in a period of reshaping its health-care system.

The paper proceeds as follows. "Background" section discusses the background of our study. "Data and model" section illustrates the model and data. Endogeneity problems may pose significant challenges to our conclusions, therefore, we give a detailed discussion on the potential selection issues in this section. In "Model" section, we report the econometric results. Finally, "Empirical results" section draws the conclusions.

Background

China is currently still in a period of reshaping its health care system. Prior to the reform, health care in China had been financed primarily through three major public programs: the Cooperative Medical System (CMS), the Government Insurance Scheme (GIS), and the Labor Insurance Scheme (LIS). The CMS was established in rural areas, and was generally funded by contributions from participants but was heavily subsidized by the collective welfare funds. The GIS was primarily for government employees, veterans, educators, and college

---

[1] One exception is Kowalski (2010), in which the author adopts a quantile regression approach.

students, whereas the LIS was for workers of all state-owned, and some non-state owned, enterprises, and was financially supported by the welfare funds of enterprises. These three types of insurance plan together provided near-universal coverage to the Chinese population from the 1950s to the end of 1970s (Zhong 2011).

China has been undergoing its economic reform since the end of 1970s. The communes disappeared after the introduction of Household Responsibility System in the rural economy. Consequently, the CMS collapsed and its coverage fell to less than 5 % of the population by the early 1990s (Zhao 2006). Urban reform was initiated in the mid-1980s, and since then state-owned enterprises (SOEs) have been given substantial financial autonomy. The 1990s was a decade marked by profound shifts in industrial and enterprise structures in urban China. Since the early 1990s, non-state enterprises (including private, foreign, joint-venture, and mixed ownership) have emerged as important players in the Chinese economy. The proportion of the labor force employed in the state sector has fallen continuously. Meanwhile, the SOEs have performed poorly due to soft budget constraints and many other institutional problems. Radical reforms for the SOEs were introduced in the middle and late 1990s. To vitalize the Chinese economy, a wave of ownership restructuring for the SOEs occurred in which many loss making enterprises were bankrupted, merged with other enterprises, transformed into joint-stock companies, or privatized. As a result, large numbers of workers in the SOEs were laid off. Because LIS schemes pool risks at firm rather than national level, many workers in the SOEs who kept their jobs found their employers unable to maintain the plans due to the firm's poor economic performance. As a result, health insurance coverage declined considerably in urban areas. By 1998, nearly 50 % of the urban population lacked insurance coverage (Wagstaff and Lindelow 2008). Meanwhile, both the absolute and relative costs of health care have risen sharply due to providers' profit-seeking behavior and the resultant waste and inefficiency (Yip and Hsiao 2008).

In the response to the public discontent with unaffordable health care services, the Chinese government has initiated several public health insurance programs since 1998, which try to cover different segments of the population. The New Cooperative Medical Scheme (NCMS) was introduced in the rural areas. The new health insurance scheme that tries to cover all formal-sector workers in urban areas was known as Urban Employee Basic Medical Insurance (UEBMI). Compared with the old urban health insurance systems (LIS and GIS), the UEBMI expanded coverage to private sector employees and pools its risk at the municipal level, which provides more stable financing. The final goal of UEBMI is to replace LIS and GIS in the cities. The fundamental rules of UEBMI is designed at provincial level, and managed by each city's Social Insurance Bureau (SIB). Local SIBs have a certain degree of discretion regarding the choices of deductibles, services covered, reimbursement methods, and methods of payment to providers. UEBMI coverage has expanded continuously since 1998. It was extended to over 180 million urban enrollees by 2007 (Lin et al. 2009). The GIS is the best public health insurance available in China in terms of generosity. Although the final goal of the health-care reforms is to gradually subsume LIS and GIS into UEBMI, the real power holders have incentives to protect their interests collectively. Consequently, the process of reform to GIS has been slow, it was still in existence in most urban areas until very recently. The percentage of the population covered by the private and 'other' insurance plans in urban areas has increased significantly during this period (Wagstaff and Lindelow 2008). Despite the existence of these insurance schemes, a significant proportion of urban residents without formal employment remained outside of the health-care insurance system in 2007 (Lin et al. 2009).

Data and model

Data

The data used in this study come from the Rural–Urban Migration in China and Indonesia (RUMiCI) project, which include longitudinal surveys in both countries. The RUMiCI China survey, also called Chinese Household Income Project (CHIP)[2] comprises three randomly selected samples[3]: the rural household sample surveys households living in rural areas with rural Hukou (rural sample), the urban household sample covers households living in urban areas with Urban Hukou (urban sample), and the migrant sample comprises households working in urban areas with rural Hukou (migrant sample). In this study we employ the urban sample of the most recent 2009 RUMiCI survey data, in which we exclude a small number of observations with multiple insurance plans.

The 2009 RUMiCI urban survey was conducted from nine provinces: Shanghai, Jiangsu, Zhejiang, Anhui, Henan, Hubei, Guangdong, Chongqing, and Sichuan, which includes 5000 households and 14,859 individuals. The 2009 RUMiCI urban sample collects detailed information regarding demographic characteristics of members of the household, their income, earnings, job situation, as well as medical expenditures and types of health insurance information. Among valid records, 23.82 % individuals are with zero-medical consumption. Definition and summary statistics of selected variables are presented in Table 1.

Model

Estimating the effects of determinants of medical expenditure requires the treatment of large fraction of zero values. Researchers have dealt with this issue by estimating with either selection model or two-part models. The adoption of the selection model is criticized by Dow and Norton (2003), who argued that no such selection problem exists when the observed zero expenditure is the true consumption, as opposed to latent potential outcomes. In addition, the objection to the selection model also centers around the fact that they assume a bivariate normal distribution between the error terms, which are known to be sensitive to departures from normality (Duan et al. 1983; Goldberger 1983). Along with the selection model, the use of the two-part model assumes the independent process of decision to medical spending and the decision on the level of expenditure. Although it is not satisfactory on the grounds of this potentially restrictive assumption (Hay and Olsen 1984; Maddala 1985), it is demonstrated to be a robust estimator even if the true model is a selection type (Duan et al. 1983). Additionally, the two-part model allows an investigation as to whether variables of interest have larger impacts on the participation or consumption decisions (Manning et al. 1995). Therefore, two-part models are frequently employed benchmarks in health economics research that involves observations with a cluster of zero (Doorslaer and Wagstaff 2000).

The first stage in the two-part model is to run a logit regression to check factors affecting the likelihood of urban residents' care-seeking behavior. The participation equation is:

$$\Pr ob(D_i = 1) = \exp\{Z_i \alpha\}/[1 + \exp\{Z_i \alpha\}] \tag{1}$$

Table 1 Definition and summary statistics of selected variables

Variable	Definition	Mean or %
MedExp	Natural logged total medical expenditure before reimbursements	4.60
Income	Natural logged monetary value of income	7.58
Age	Age of individual	42.09
Gender	Gender=0 for male	50.05%
	Gender=1 for female	49.95%
ID	Current work identity, ID=1for working or employed person	49.62%
	ID=2 for retired or people unable to work	27.62%
	ID=3 for unemployed	7.23%
	ID=4 for full-time homemaker, students, or persons who are waiting for job assignment or entering a higher school	15.52%
Minority	Nationality, minority=1 for Han majority	98.58%
	Minority=2 for minorities	1.42%
MariStatus	Marital Status, MariStatus=1 for the single	21.94%
	MariStatus=2 for married	73.31%
	MariStatus=3 for divorced	1.85%
	MariStatus=4 for widowed	2.61%
	MariStatus=5 for others	0.28%
Education	Schooling years	10.79
Health	Self-reported health status, health=1, very good	11.55%
	Health=2, good	51.18%
	Health=3 general	32.43%
	Health=4 poor	4.08%
	Health=5 very poor	0.76%
Disable	Disable=1 disability affects work and living	1.18%
	Disable=2 without disability	98.82%
InsType	Types of insurance plans, InsType=1: GIS	10.70%
	InsType=2: UEBMI	69.01%
	InsType=3: self-purchased commercial health insurance	3.25%
	InsType=4: medical expenditure afford with one's pocket money	17.05%

where D_i is a dichotomous variable defined as 1 if individuals has positive health care spending, and 0 otherwise. Z_i are elements that influence the decision of medical consumption for urban residents. In the second stage, given the positive medical consumption, a log-linear model is created to analyze the effects of various factors on the conditional mean of medical expenditure.

$$Ln(Y_i|Y_i > 0, X_i) = X_i\beta + \varepsilon_i \ \ \varepsilon_i \sim [0, \sigma_i^2] \tag{2}$$

similarly, X_i represents decisive factors affecting the level of medical consumption. ε_i is the error term.

Departure from the existing literatures, our paper argues that valuable information could be concealed by merely considering the effects of various determinants on the mean value of medical consumption. Consequently, we employ a quantile regression to further inspect heterogeneity of impact factors, and to mine the disparities of effects of the same factor on different medical consumption groups. The quantile regression is outlined as:

$$Q_\tau(Y_i \,|Y_i > 0, X_i) = X_i' \, \beta_\tau \quad \tau \in (0, 1) \tag{3}$$

by solving the below minimization problem,

$$\beta_\tau \in \operatorname*{arg\,min}_{\beta_\tau \in R^k} \left(\sum_{(i:Y \geq X'\beta_\tau)} \tau \left|Y_i - X_i'\beta_\tau\right| + \sum_{(i:Y < X'\beta_\tau)} (1 - \tau) \left|Y_i - X_i'\beta_\tau\right| \right) \quad \tau \in (0, 1)$$

we get coefficient vector β_τ at different quantiles on the medical expenditure distribution.

In each of the regression, we include the following set of independent variables: income, age, education, gender, working status, ethnic minority, marital status, self-reported health status, whether suffer a physical disability, type of health insurance plan. Income may be positively related to medical expenditures. The variable of income is defined as the natural logged monetary value of income. Age is defined by years. It is associated with chronic health decline and need for health care services. Males and females may have different attitudes and needs for health care services. Gender is indicated by a dummy variable in which female is 1 and male is 0. Working status is defined by a series of dummy variables to indicate whether the respondent has held a full-time job in the past year, retired, or being unemployed. Labor force participation implies a higher time cost for seeking health care, which may have a negative influence on medical expenditure. Ethnic minority is a dummy variable to indicate whether the respondent is a member of ethnic minority, who may have different culture or attitudes for seeking health care. Marital status is defined by a series of dummy variables to indicate whether the respondent is single, married, divorced, or widowed. Levels of education are associated with attitudes to health care and good health behaviors, which may have certain influence on health care utilization. We therefore include schooling years into the regressions. Health status should have strong influence on seeking health care. Respondent's self-reported health status is defined by a series of five categorical variables: very good, good, general, poor, and very poor. We also include a dummy indicator for those who suffer a physical disability that affects their normal work or living. Finally, types of health insurance plan expressed as a series of dummy variables are used to indicate whether the respondent has GIS, UEBMI, self-purchased commercial insurance and no insurance at all.

Potential endogeneity and selection issues

Endogeneity problems may pose significant challenges to our analysis. One concern at here might be that individuals may self-select into different insurance status, i.e., individuals who expect higher medical expenditure may choose more generous insurance plan. As discussed in "Background" section, the two major public health insurance plans (GIS and UEBMI) are all employer-sponsored and managed by local governments, and thus self-selection into these two types of health insurance might not be a significant issue. As one way to deal with the potential selection issue, in addition to the full sample regressions, we will also exclude those observations that have private health insurance plans from the quantile regression as a sensitivity analysis. One may still argue that certain individuals could purposefully select into particular types of job, and then these individuals will have different types of insurance, for

Table 2 GIS and UEBMI holders

Variable	Definitions of variable	GIS	UEBMI
Income	Natural logged total household income	7.80	7.59
Age	Age of individual	50.13	45.34
Education	Schooling years	11.61	10.64
Gender	Gender = 0 for male;	54.96 %	47.58 %
	Gender = 1 for female	45.04 %	52.42 %
Disable	=1 disability affects work and living	1.10 %	1.47 %
	=2 without disability	98.90 %	98.53 %
Health	Health = 1: very good	11.94 %	8.97 %
	Health = 2 good	49.85 %	46.22 %
	Health = 3 general	30.39 %	38.77 %
	Health = 4 poor	6.82 %	5.17 %
	Health = 5 very poor	1.00 %	0.87 %
MariStatus	MariStatus = 1 for the single	6.93 %	13.54 %
	MariStatus = 2 for married	87.85 %	81.03 %
	MariStatus = 3 for divorced	1.81 %	1.90 %
	MariStatus = 4 for widowed	3.21 %	3.23 %
	MariStatus = 5 for others	0.20 %	0.30
Minority	1 for majority	98.89 %	98.97 %
	2 for minority	1.11 %	1.03 %
ID	ID = 1 for working or employed person	52.46 %	51.56 %
	ID = 2 for retired or people unable to work	42.43 %	33.67 %
	ID = 3 for laid-off workers, or unemployed	1.40 %	5.59 %
	ID = 4 for full-time homemaker or students	3.71 %	9.18 %
n	Number of observations	1900	7233

Source calculated from the 2009 RUMiCI data

example, GIS. This clearly can bias the results. In Table 2, we compare various demographic, socioeconomic, and health factors between the two groups of individuals. Generally speaking, we cannot observe systematic differences in those characteristics between the GIS holders and UEBMI holders. As another way to address this concern, we will also run the regressions separately for the GIS holders and UEBMI holders to see whether we can reach the same conclusions.

Another concern might be related to the omitted variable problem. Residents in different regions may have different income levels and face different supply-side factors of health care services, which could have significant impacts on medical expenditures. Moreover, as we discussed in the second section, various health insurance plans are mainly managed at province level, and thus people in different localities may face different insurance policies. To address this concern, we add a group of provincial dummies into the quantile regression.

Against the conclusions in Duan et al. (1983, 1984, 1985), Leung and Yu (1996) argue that if the true model is a selection type, the two-part model might be problematic. Therefore, we will also run a Heckman selection model to check whether results are markedly different between the two modeling choices.

Empirical results

Two-part model

Table 3 presents the two-part model estimates.

Income and education have been suggested in previous studies as important predictors of one's socioeconomic status. In this study, although income does not significantly affect people's likelihood of using health care service, once people with higher income get sick, they expend more on their health care. Better educated persons are more likely to visit the hospitals, but their medical expenditures are not significantly different from others. We also find that both age and gender significantly affect the likelihood of individual's visiting hospitals and the costs on medical expenditure. Older people are more likely to visit the hospital and

Table 3　The two-part model

	Logit model		Log-linear model	
	Coef.	SD	Coef.	SD
Income	0.020	(0.045)	0.277***	(0.028)
Age	0.014***	(0.004)	0.024***	(0.002)
Education (years of education)	0.024**	(0.006)	0.005	(0.007)
Gender (female vs male)	0.432***	(0.063)	0.277***	(0.038)
Current work identity: the benchmark is ID = 1,working or employed individuals				
ID = 2: retired or inability to work group	0.168	(0.178)	0.198**	(0.089)
ID = 3: unemployed group	0.414	(0.269)	0.236*	(0.141)
ID = 4: not ready for job group	0.272	(0.314)	0.583***	(0.204)
Self-reported health status: benchmark is health = 1, very good				
Health = 2, good	0.053	(0.087)	0.166***	(0.057)
Health = 3, general	0.451***	(0.100)	0.562***	(0.062)
Health = 4, poor	1.730***	(0.436)	1.760***	(0.140)
Health = 5, very poor	0.355	(0.546)	2.250***	(0.321)
Disability (normal vs disable)	−0.631	(0.654)	−0.349***	(0.266)
Types of health insurance: benchmark is InsType = 1, GSI				
InsType = 2, UEBMI	−0.027	(0.094)	−0.054	(0.056)
InsType = 3, self-purchased commercial insurance	−0.033	(0.245)	−0.127	(0.148)
InsType = 4, no insurance	−0.392***	(0.127)	−0.499***	(0.082)
Marital status: benchmark is MariStatus = 1, single				
MariStatus = 2, married	0.648***	(0.099)	0.036	(0.071)
MariStatus = 3, divorced	0.403*	(0.221)	−0.030	(0.141)
MariStatus = 4, widowed	0.983***	(0.432)	0.125	(0.204)
MariStatus = 5, others	0.578	(0.455)	−0.197	(0.278)
Minority, minority vs Han	−0.182	(0.261)	0.347**	(0.170)
Constant	−0.272	(0.758)	2.720***	(0.359)
Sample size	n = 12,905		n = 9831	

Significant degree at *0.1, **0.05 and ***0.01, respectively. Standard deviations are in the parenthesis

their medical expenditure is higher than that of those younger. Meanwhile, female are more inclined to access health care service, and their medical expenditure is larger than that of male on average.

We break urban residents' work identity (ID) into four categories and consider the working or employed individuals as the benchmark, other people are divided into groups of inability to work, unemployed group, and not ready for job group.[4] Relative to the working employees, individuals in other work identity groups show no significant differences in their frequency to access the medical service. Nonetheless, they uniformly have higher medical expenditure than the benchmark.

Health status (Health) is one of the most important factors that determine people's propensity to use health care. With a deteriorating trend of self-reported health status, people's medical expenditure increases progressively. People who report their health to be general and poor tend to have a higher frequency of hospital visit. We also find that people with a physical disability neither visit hospital more nor expend higher on their medical expenditure.

For the health insurance plans (InsType), individuals with their medical and health services publicly funded by the state or work unit are regarded as the benchmark. People with UEBMI and commercial health insurance plans are not remarkably different from the benchmark in either their probability of health care utilization or medical expenditure. However, individuals who have to pay their health cares by their own have a lower frequency of hospital visiting and spend much less on their medical cares relative to the benchmark.

Marriage status (MariStatus) also influences the likelihood of hospital visiting. Relative to the single person, married, divorced, and widowed people access the health care service more often. Nonetheless, their medical expenditure does not significantly depend on their marriage status.

Quantile regression

Table 4 lists the heterogeneous quantile coefficients of various determinants of the medical expenditure, and offers supplemental information in understanding how those determinants influence people's medical consumption. The Wald test reported in the notes below Table 4 indicates the presence of quantile effects.

It can be seen from Table 4 that the marginal effects for work identities (ID), marital status (MariStatus), and minority status do not have a clear pattern across the medical expenditure distribution, and their mean estimates reflect their impacts on medical expenditure approximately as their coefficients spread randomly around the mean coefficient. However, when the marginal effects from certain determinants display a regular pattern, policy design simply based on the mean coefficients can be incomplete. For example, the coefficient regarding the gender indicates an obvious decreasing pattern along the medical expenditure distribution, reflecting that its mean coefficient significantly underestimates female's health care expenditure at the lower end of health care consumption, and overestimates the difference of medical payments between male and female at higher quantiles along the expenditure distribution. Meanwhile, age positively contributes to the medical expenditure with decreasing marginal effects. Therefore, the mean estimate underestimates the age effect for subgroups with lower medical expenditure, and overestimate age effect for individuals with higher health care spending.

---

[4] Among them, inability to work group includes formally retired people or individuals who are unable to work. Unemployed group is composed of lay-off workers and purely unemployed individuals. Finally, the group named not ready for job contains full-time homemakers, students, and other idle persons.

Table 4 The quantile regression

	5th	10th	25th	50th	75th	90th	95th
Income	0.291***	0.274***	0.317***	0.298***	0.262***	0.217***	0.208***
	(0.051)	(0.039)	(0.030)	(0.030)	(0.034)	(0.036)	(0.069)
Age	0.021***	0.027***	0.026***	0.027***	0.020***	0.013***	0.021***
	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	(0.006)
Education	−0.007	0.009	0.006	0.015**	0.007	−0.017**	0.003
	(0.013)	(0.010)	(0.008)	(0.007)	(0.009)	(0.007)	(0.016)
Gender	0.312***	0.313***	0.321***	0.286***	0.228***	0.180***	0.281***
	(0.065)	(0.054)	(0.041)	(0.042)	(0.051)	(0.045)	(0.105)
ID = 2	0.228*	0.083	0.271***	0.227***	0.302***	0.150	−0.086
	(0.136)	(0.109)	(0.096)	(0.087)	(0.098)	(0.133)	(0.300)
ID = 3	0.229	0.128	−0.011	0.280*	0.222	0.440***	0.532
	(0.168)	(0.189)	(0.177)	(0.162)	(0.315)	(0.088)	(1.330)
ID = 4	0.415***	0.331	0.473	0.573***	0.941***	0.367***	0.495**
	(0.109)	(0.205)	(0.325)	(0.073)	(0.107)	(0.077)	(0.228)
Health = 2	−0.068	−0.043	0.127*	0.232***	0.233***	0.177***	0.267*
	(0.102)	(0.066)	(0.069)	(0.062)	(0.067)	(0.048)	(0.139)
Health = 3	0.180	0.287***	0.548***	0.659***	0.665***	0.614***	0.633***
	(0.125)	(0.079)	(0.074)	(0.066)	(0.073)	(0.054)	(0.154)
Health = 4	0.838**	0.901***	1.350***	2.050***	2.100***	2.410***	2.530***
	(0.332)	(0.213)	(0.312)	(0.144)	(0.500)	(0.384)	(0.206)
Health = 5	1.390***	1.890***	2.040***	2.500***	2.350***	2.770***	2.640***
	(0.291)	(0.247)	(0.509)	(0.114)	(0.877)	(0.164)	(0.298)
Disability = 2	−0.837	−0.570*	−0.303	−0.313***	−0.025	−0.471	−0.121
	(1.940)	(0.310)	(0.641)	(0.090)	(0.729)	(0.319)	(2.120)
InsType = 2	−0.022	0.064	−0.050	−0.145**	−0.080	−0.061	0.145
	(0.135)	(0.066)	(0.070)	(0.061)	(0.079)	(0.061)	(0.127)
InsType = 3	0.204	0.124	−0.225	−0.348	−0.100	−0.168	−0.239
	(0.350)	(0.187)	(0.167)	(0.212)	(0.374)	(0.164)	(0.156)
InsType = 4	−0.503***	−0.473***	−0.418***	−0.579***	−0.443***	−0.460***	−0.385
	(0.152)	(0.110)	(0.101)	(0.082)	(0.096)	(0.051)	(0.243)
MariStatus = 2	−0.083	−0.104	0.002	0.045	0.179**	0.193***	0.204
	(0.101)	(0.103)	(0.067)	(0.076)	(0.089)	(0.067)	(0.193)
MariStatus = 3	−0.112	−0.337***	−0.049	−0.054	0.336	0.092	−0.042
	(0.123)	(0.117)	(0.127)	(0.156)	(0.227)	(0.131)	(0.315)
MariStatus = 4	−0.111	−0.511***	0.254	0.167	0.184	0.410	0.423
	(0.142)	(0.196)	(0.630)	(0.106)	(0.119)	(0.365)	(0.391)
MariStatus = 5	0.248	−0.009	0.285***	−0.207	−0.236	−0.788***	−1.140***
	(0.174)	(0.177)	(0.079)	(0.158)	(0.167)	(0.146)	(0.357)
Minority = 2	−0.141	−0.382**	0.501	0.534***	0.290	0.261	0.564
	(0.164)	(0.161)	(0.724)	(0.111)	(0.342)	(0.842)	(0.523)

Significant degree at *0.1, **0.05 and ***0.01, respectively. The Wald test for testing the quantile effects shows that F value = 4.35 and p value < 2e−16

The estimated income elasticity (Income) in the quantile regression indicates a decreasing pattern, as reflected in its coefficient of 0.317 at the first quantile versus that of 0.262 at the third quantile. The mean estimate underestimates the income elasticity at lower levels of health care expenditure, and overvalues income elasticity for those with higher medical payments. The disparity of income elasticity on the medical expenditure distribution implies that, income imposes greater effects on health care consumption when the medical expenditure is at small scale, and this effect is reduced as the medical expenditure becomes larger.

Referring to the health status, the quantile regression reveals that, with the deterioration of individual's health condition, his/her medical expenditure increases accordingly. For those who report their health condition to be general, poor, or very poor, the marginal effects of health status along the health expenditure distribution indicates an increasing pattern, implying that the mean coefficient seriously underestimates the health care costs for those with high medical expenditure and overestimates the health care costs for those with low medical expenditure. Therefore, the mean estimates are insufficient to reflect marginal effect of health status on the medical expenditure. In addition, different from the indistinctive effects of disability on medical expenditure, the quantile regression reports that, at the 10th and 50th quantile, people who suffer certain kind of disability spend more on their health care than a normal person.

Previous studies suggest that in China, publicly funded health insurance program motivates inefficient use of health care resources (Gao et al. 2001; Liu et al. 2002). Results in our study indicate that health care expenditures are closely related to insurance status. Comparing with the people with health care insurance, individuals without such benefits consume much less medical resources, especially for those with lower medical expenditures.

One important conclusion that can be drawn from results of the quantile regression is that medical expenditures at the upper end of the distribution are mainly driven by need factors such as poor health status; one the other hand, medical expenditures at the lower end of the distribution are under stronger influences of non-need factors such as income, gender, and insurance status.

Potential endogeneity problems may have significant influences on this conclusion. First, residents in different regions may have different income levels and face different supply-side factors of health care services, which could have significant impacts on medical expenditures. Moreover, as we discussed in the second section, fundamental rules on various health insurance plans are mainly designed at provincial level, and thus people in different provinces may face different insurance policies. To address this concern, we add a group of provincial dummies into the quantile regression, and the results are reported in Table 5.[5]

Another concern might be that individuals may self-select into different insurance status. To address this concern, as discussed in "Data and model" section, we first exclude those observations that have private health insurance plans from the quantile regression. The results are reported in Table 6. We also run the regressions separately for the GIS holders and UEBMI holders, and report the results in Tables 7 and 8.

Results in Tables 4 and 5 are different. It is clear that adding province dummies changes the coefficients. Similarly, when we run the regression using different subsamples, we reach different sets of coefficients in Tables 6, 7, and 8.[6] However, generally speaking, we still can observe the same pattern as in Table 4 in all the cases. That is, medical expenditures at the

---

[5] To control the potential city-specific effect and heteroscedasticity problem (errors clustered at city level), we have tried to add a group of city dummies instead of the provincial dummies, and reached similar results. To save more degrees of freedom, we run the regressions with provincial dummies.

[6] From Tables 7 and 8, an interesting finding is that the marginal effects of age for GIS holders are relatively stable across health care expenditure distribution, and much stronger than that for UEBMI holders. Rising

Table 5 Quantile regression with province dummies

	5th	10th	25th	50th	75th	90th	95th
Income	0.191***	0.190***	0.180***	0.152***	0.184***	0.189***	0.080
	(0.042)	(0.020)	(0.028)	(0.033)	(0.034)	(0.049)	(0.059)
Age	0.024***	0.024***	0.027***	0.028***	0.023***	0.014***	0.019***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.005)
Education	0.006	0.014	0.018***	0.029***	0.008	−0.006	0.013
	(0.010)	(0.009)	(0.007)	(0.008)	(0.009)	(0.013)	(0.015)
Gender	0.365***	0.269***	0.314***	0.267***	0.216***	0.201***	0.206**
	(0.054)	(0.048)	(0.039)	(0.043)	(0.047)	(0.066)	(0.088)
ID = 2	0.300***	0.305*	0.178	0.184*	0.151	0.179	0.149
	(0.081)	(0.161)	(0.134)	(0.100)	(0.095)	(0.142)	(0.251)
ID = 3	0.095	0.045	0.099	0.199	0.277	0.598***	0.653
	(0.070)	(0.096)	(0.166)	(0.278)	(0.393)	(0.148)	(0.711)
ID = 4	0.102	0.137	0.516	0.673***	0.612***	0.390	0.148
	(0.153)	(0.306)	(0.630)	(0.201)	(0.068)	(0.404)	(0.447)
Health = 2	0.002	0.025	0.048	0.155***	0.229***	0.293***	0.364**
	(0.078)	(0.079)	(0.048)	(0.056)	(0.066)	(0.103)	(0.154)
Health = 3	0.303***	0.399***	0.397***	0.524***	0.630***	0.719***	0.778***
	(0.089)	(0.089)	(0.053)	(0.063)	(0.079)	(0.110)	(0.167)
Health = 4	0.760***	0.860***	1.220***	1.870***	1.990***	2.560***	2.480***
	(0.120)	(0.315)	(0.290)	(0.218)	(0.262)	(0.222)	(0.174)
Health = 5	1.030*	1.680***	2.040***	2.330***	2.210**	2.610***	2.570*
	(0.597)	(0.441)	(0.755)	(0.093)	(0.887)	(0.238)	(1.420)
Disability = 2	−0.620	−0.388*	−0.275	−0.411***	−0.186	−0.342	−0.066
	(1.780)	(0.207)	(0.624)	(0.136)	(0.675)	(0.405)	(1.500)
InsType = 2	−0.127*	−0.088	−0.108*	−0.166***	−0.125*	−0.109	−0.062
	(0.077)	(0.070)	(0.060)	(0.059)	(0.067)	(0.102)	(0.106)
InsType = 3	−0.058	−0.088	−0.068	−0.261	0.011	−0.121	−0.351
	(0.107)	(0.147)	(0.143)	(0.185)	(0.216)	(0.267)	(0.253)
InsType = 4	−0.455***	−0.524***	−0.472***	−0.527***	−0.412***	−0.416***	−0.465***
	(0.097)	(0.098)	(0.081)	(0.101)	(0.090)	(0.128)	(0.125)
MariStatus = 2	−0.104	−0.107	0.045	0.110	0.092	0.193*	0.173
	(0.095)	(0.082)	(0.057)	(0.079)	(0.079)	(0.099)	(0.210)
MariStatus = 3	−0.246	−0.156	−0.015	0.136	0.070	0.100	0.003
	(0.230)	(0.214)	(0.139)	(0.158)	(0.174)	(0.173)	(0.329)
MariStatus = 4	−0.626***	−0.450	0.122	0.333***	0.306	0.272*	−0.012
	(0.209)	(0.480)	(0.357)	(0.128)	(0.279)	(0.154)	(0.264)
MariStatus = 5	−0.278	−0.361***	0.091	−0.356	−0.657	−0.998***	−1.540***
	(0.482)	(0.093)	(0.096)	(0.252)	(0.513)	(0.334)	(0.329)
Minority = 2	0.001	−0.303	0.340	0.611***	0.379**	0.328	0.728***
	(0.125)	(0.191)	(0.452)	(0.114)	(0.170)	(0.777)	(0.241)

Significant degree at *0.1, **0.05 and ***0.01, respectively; coefficients of province dummies are not reported; the Wald test for testing the quantile effects shows that F value = 4.28 and p value < 2e−16

Table 6 Quantile regression without self-purchased insurance(with province dummies)

	5th	10th	25th	50th	75th	90th	95th
Income	0.183***	0.215***	0.163***	0.159***	0.206***	0.196***	0.090
	(0.027)	(0.032)	(0.034)	(0.033)	(0.030)	(0.043)	(0.063)
Age	0.025***	0.024***	0.025***	0.028***	0.024***	0.015***	0.020***
	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.005)
Education	0.012	0.013*	0.018**	0.031***	0.010	−0.007	0.009
	(0.010)	(0.008)	(0.007)	(0.008)	(0.009)	(0.012)	(0.013)
Gender	0.372***	0.277***	0.299***	0.267***	0.227***	0.204***	0.226**
	(0.053)	(0.043)	(0.041)	(0.042)	(0.045)	(0.064)	(0.089)
ID = 2	0.343***	0.334***	0.211*	0.197**	0.156*	0.164*	0.121
	(0.069)	(0.121)	(0.125)	(0.084)	(0.090)	(0.089)	(0.224)
ID = 3	0.105	0.063	0.148	0.213	0.423	0.589***	0.629
	(0.081)	(0.090)	(0.108)	(0.273)	(0.277)	(0.169)	(0.740)
ID = 4	0.073	0.161	0.547	0.670***	0.613***	0.389	0.152
	(0.120)	(0.198)	(0.572)	(0.200)	(0.074)	(0.499)	(0.225)
Health = 2	0.019	0.030	0.073	0.152***	0.223***	0.301***	0.362***
	(0.063)	(0.072)	(0.051)	(0.057)	(0.060)	(0.112)	(0.136)
Health = 3	0.321***	0.410***	0.442***	0.527***	0.631***	0.735***	0.754***
	(0.071)	(0.080)	(0.056)	(0.063)	(0.073)	(0.117)	(0.150)
Ealth = 4	0.764***	0.920***	1.230***	1.900***	2.060***	2.560***	2.490***
	(0.117)	(0.331)	(0.258)	(0.164)	(0.264)	(0.215)	(0.165)
Health = 5	1.040***	1.630	1.660*	2.020***	2.240***	2.610***	2.580***
	(0.087)	(1.710)	(0.989)	(0.283)	(0.800)	(0.967)	(0.233)
Disability = 2	−0.569	−0.441**	−0.320	−0.440***	−0.068	−0.334	−0.186
	(1.780)	(0.182)	(0.582)	(0.075)	(0.653)	(0.524)	(1.270)
InsType = 2	−0.122*	−0.075	−0.121*	−0.168***	−0.111*	−0.103	−0.049
	(0.074)	(0.066)	(0.072)	(0.061)	(0.066)	(0.085)	(0.094)
InsType = 4	−0.442***	−0.505***	−0.490***	−0.526***	−0.387***	−0.407***	−0.434***
	(0.100)	(0.072)	(0.091)	(0.102)	(0.095)	(0.115)	(0.106)
MariStatus = 2	−0.104	−0.133*	0.054	0.120	0.076	0.169*	0.183
	(0.098)	(0.080)	(0.059)	(0.076)	(0.074)	(0.100)	(0.203)
MariStatus = 3	−0.239	−0.119	−0.070	0.131	0.136	0.082	−0.006
	(0.262)	(0.256)	(0.142)	(0.155)	(0.173)	(0.156)	(0.365)
MariStatus = 4	−0.647***	−0.418***	0.066	0.329***	0.310	0.254	−0.010
	(0.129)	(0.130)	(0.385)	(0.113)	(0.255)	(0.160)	(0.236)
MariStatus = 5	−0.333	−0.378**	0.136	−0.362	−0.674	−1.020***	−1.460***
	(0.599)	(0.155)	(0.235)	(0.270)	(0.504)	(0.310)	(0.238)
Minority = 2	0.023	−0.301***	0.037	0.625***	0.373**	0.342	0.736***
	(0.084)	(0.053)	(0.442)	(0.239)	(0.160)	(0.728)	(0.119)

Significant degree at *0.1, **0.05 and ***0.01, respectively; coefficients of province dummies are not reported; the Wald test for testing the quantile effects shows that F value = 6.86 and p value < 2e−16

Table 7 GIS subgroup quantile regression with province dummies

	5th	10th	25th	50th	75th	90th	95th
Income	0.480***	0.274***	0.091	0.210***	0.300***	0.409***	0.342**
	(0.143)	(0.059)	(0.069)	(0.061)	(0.106)	(0.111)	(0.149)
Age	0.038***	0.021***	0.027***	0.041***	0.035***	0.036***	0.027***
	(0.009)	(0.006)	(0.005)	(0.006)	(0.007)	(0.009)	(0.008)
Education	−0.069*	0.002	0.014	0.011	−0.037	−0.023	−0.012
	(0.036)	(0.016)	(0.012)	(0.019)	(0.024)	(0.026)	(0.037)
Gender	0.442***	0.269***	0.210**	0.312***	0.254**	0.428***	0.468***
	(0.146)	(0.091)	(0.096)	(0.107)	(0.102)	(0.135)	(0.156)
ID = 2	0.610	0.346	0.634***	0.243	0.373	0.673**	1.680
	(0.389)	(0.626)	(0.125)	(0.266)	(0.274)	(0.306)	(1.710)
ID = 3	3.020***	2.290***	1.580***	0.905***	0.347**	−0.322*	−1.060***
	(0.734)	(0.132)	(0.223)	(0.159)	(0.165)	(0.186)	(0.356)
ID = 4	2.770***	2.080	0.224	1.530***	1.340	0.948	0.484
	(0.927)	(2.390)	(2.730)	(0.207)	(1.300)	(0.610)	(1.470)
Health = 2	−0.745**	−0.357***	0.009	0.092	0.176	0.202	1.000
	(0.353)	(0.104)	(0.087)	(0.114)	(0.171)	(0.126)	(0.831)
Health = 3	−0.514**	−0.140	0.367***	0.465***	0.578***	0.723***	1.360
	(0.253)	(0.153)	(0.134)	(0.166)	(0.170)	(0.184)	(0.828)
Health = 4	0.673	0.742	2.390***	2.180***	2.710***	3.120**	3.210***
	(0.518)	(2.360)	(0.139)	(0.660)	(0.455)	(1.270)	(0.841)
Health = 5	−1.070	−0.701	0.519	1.560***	1.910***	2.750***	4.050**
	(0.698)	(0.720)	(0.457)	(0.388)	(0.180)	(0.530)	(1.750)
Disability = 2	0.166	0.465**	0.407***	0.151	0.066	−0.018	−0.425
	(0.498)	(0.207)	(0.147)	(0.264)	(0.143)	(0.495)	(1.060)
MariStatus = 2	0.048	0.181	0.414	−0.065	0.219	−0.560	−1.420
	(1.280)	(0.587)	(0.327)	(0.427)	(0.303)	(0.513)	(1.090)
MariStatus = 3	4.200***	3.630***	2.940***	2.060***	1.180***	−0.069	−1.360
	(0.478)	(0.262)	(0.213)	(0.298)	(0.249)	(0.557)	(1.080)
MariStatus = 4	0.360	−0.149	−0.717***	−1.290***	−1.900***	−2.770***	-4.860*
	(0.610)	(0.205)	(0.181)	(0.335)	(0.192)	(0.562)	(2.690)
MariStatus = 5	0.051	−0.381	1.260	0.910***	0.384	0.285	−0.066
	(0.155)	(0.294)	(2.600)	(0.304)	(0.579)	(0.233)	(0.855)
Minority = 2	−2.160***	−1.210***	−0.712***	−0.491**	0.136	0.782***	2.420
	(0.387)	(0.126)	(0.128)	(0.201)	(0.196)	(0.246)	(2.450)

Significant degree at *0.1, **0.05 and ***0.01, respectively; coefficients of province dummies are not reported; the Wald test for testing the quantile effects shows that F value = 10.30 and p value < 2e−16

upper end of the distribution are mainly driven by some need factors and medical expenditures at the lower end of the distribution are under stronger influences of some non-need factors.

Footnote 6 Continued

health expenditures on the elderly is a major concern for policymakers in the developed world. Our results indicate that it might be closely related to features of health insurance plans.

Table 8 UEBMI subgroup quantile regression with province dummies

	5th	10th	25th	50th	75th	90th	95th
Income	0.188***	0.124***	0.151***	0.135***	0.112***	0.057	0.053
	(0.047)	(0.041)	(0.036)	(0.035)	(0.029)	(0.062)	(0.067)
Age	0.024***	0.025***	0.028***	0.028***	0.022***	0.007	0.008
	(0.004)	(0.003)	(0.003)	(0.003)	(0.004)	(0.005)	(0.007)
Education	−0.008	0.014	0.012	0.033***	0.026**	0.015	0.007
	(0.012)	(0.009)	(0.009)	(0.009)	(0.010)	(0.016)	(0.015)
Gender	0.377***	0.240***	0.293***	0.234***	0.179***	0.083	0.068
	(0.068)	(0.049)	(0.046)	(0.045)	(0.053)	(0.078)	(0.099)
ID=2	0.258**	0.299**	0.183*	0.109	0.111	0.164	0.031
	(0.130)	(0.129)	(0.110)	(0.101)	(0.087)	(0.147)	(0.129)
ID=3	0.179**	0.324	0.107	0.138	0.689	1.080*	1.370
	(0.088)	(0.444)	(0.146)	(0.313)	(0.690)	(0.625)	(1.040)
ID=4	−0.454	−0.248	−0.100	0.742***	0.545	−0.215	0.313
	(0.948)	(0.474)	(0.967)	(0.104)	(0.473)	(0.153)	(1.840)
Health=2	0.073	0.130***	0.098*	0.167**	0.230***	0.294**	0.510***
	(0.115)	(0.048)	(0.057)	(0.067)	(0.079)	(0.130)	(0.163)
Health=3	0.364***	0.537***	0.443***	0.527***	0.609***	0.757***	0.912***
	(0.122)	(0.077)	(0.063)	(0.072)	(0.090)	(0.137)	(0.160)
Health=4	0.804***	0.978***	1.210***	1.850***	2.000***	2.620***	2.690***
	(0.217)	(0.255)	(0.175)	(0.251)	(0.350)	(0.387)	(1.030)
Health=5	1.070***	2.020	1.640*	2.120***	2.110***	2.430***	2.840***
	(0.308)	(1.890)	(0.906)	(0.192)	(0.173)	(0.476)	(0.262)
Disability=2	−0.938***	−0.600**	−0.500***	−0.394	−1.070	−0.835***	−1.010
	(0.276)	(0.249)	(0.165)	(0.312)	(1.160)	(0.282)	(1.060)
MariStatus=2	−0.150	−0.182	−0.007	0.148	0.102	0.322**	0.360
	(0.110)	(0.115)	(0.064)	(0.091)	(0.086)	(0.164)	(0.365)
MariStatus=3	−0.313*	−0.481**	−0.182	0.154	0.034	0.174	0.467
	(0.182)	(0.199)	(0.262)	(0.118)	(0.241)	(0.298)	(0.611)
MariStatus=4	0.029	−0.233	0.087	0.269**	0.159	0.206	0.538
	(0.242)	(0.186)	(0.477)	(0.126)	(0.457)	(0.247)	(0.448)
MariStatus=5	−0.159	−0.197	0.107	−0.307	−0.546**	−1.060***	−1.450***
	(0.434)	(0.138)	(0.099)	(0.306)	(0.214)	(0.199)	(0.496)
Minority=2	−0.369	−0.371***	−0.667	0.053	0.186*	1.210*	0.755**
	(0.296)	(0.070)	(0.506)	(0.531)	(0.106)	(0.683)	(0.356)

Significant degree at *0.1, **0.05 and ***0.01, respectively; Coefficients of province dummies are not reported; the Wald test for testing the quantile effects shows that F value=7.94 and p value < 2e−16

To highlight the pattern in Tables 4, 5 and 6, we summarize the results for several important variables in Table 9. In the view of space consideration, we only report the results for the most important need factors (general, poor, and very poor health) and two non-need factors (income and gender) at 5th, 25th, 75th, and 95th quantile. From Table 9, we can clearly observe that the influences of income and gender on medical expenditures become weaker,

Table 9 A summary of results

	Non-need factors						Need factors								
	Full sample (without provincial dummies)		Full sample (with provincial dummies)		GIS and UEBMI holders (with provincial dummies)		Full sample (without provincial dummies)			Full sample (with provincial dummies)			GIS and UEBMI holders (with provincial dummies)		
	Income	Gender	Income	Gender	Income	Gender	Health=3	Health=4	Health=5	Health=3	Health=4	Health=5	Health=3	Health=4	Health=5
5th	0.291***	0.312***	0.191***	0.365***	0.183***	0.372***	0.180	0.838***	1.390***	0.303***	0.760***	1.030*	0.321***	0.764***	1.040***
25th	0.317***	0.321***	0.180***	0.314***	0.163***	0.299***	0.548***	1.350***	2.040***	0.397***	1.220***	2.040***	0.442***	1.230***	1.660***
75th	0.262***	0.228***	0.184***	0.216***	0.206***	0.227***	0.665***	2.100***	2.350***	0.630***	1.990***	2.210**	0.631***	2.060***	2.240***
95th	0.208***	0.281***	0.080	0.206**	0.090	0.226***	0.633***	2.530***	2.640***	0.778***	2.480***	2.570*	0.754***	2.490***	2.580***

Significant degree at * 0.1, ** 0.05 and *** 0.01, respectively

Table 10 Results from Heckman model and two-part model		Heckman model	Two-part model
	Income	0.247***	0.277***
		(0.030)	(0.028)
	Age	0.029***	0.024***
		(0.003)	(0.002)
	Education	0.015*	0.005
		(0.008)	(0.007)
	Gender	0.407***	0.277***
		(0.043)	(0.038)
	Identy = 2	0.216**	0.198**
		(0.102)	(0.089)
	Identy = 3	0.342**	0.236*
		(0.160)	(0.141)
	Identy = 4	0.725***	0.583***
		(0.224)	(0.204)
	Health = 2	0.171***	0.166***
		(0.063)	(0.057)
	Health = 3	0.681***	0.562***
		(0.069)	(0.062)
	Health = 4	2.090***	1.760***
		(0.163)	(0.140)
	Health = 5	2.340***	2.250***
		(0.361)	(0.321)
	Disability	−0.480	−0.349
		(0.307)	(0.266)
	InsType = 2	−0.060	−0.054
		(0.062)	(0.056)
	InsType = 3	−0.124	−0.127
		(0.166)	(0.148)
	InsType = 4	−0.627***	−0.499***
		(0.091)	(0.082)
	MariStatus = 2	0.289***	0.036
		(0.078)	(0.071)
	MariStatus = 3	0.134	−0.030
		(0.157)	(0.141)
	MariStatus = 4	0.446*	0.125
		(0.234)	(0.204)
	MariStatus = 5	0.078	−0.197
		(0.311)	(0.278)
	Minority = 2	0.253	0.347**
		(0.189)	(0.170)
	Constant	1.930***	2.720***
		(0.407)	(0.359)
Significant degree at * 0.1, ** 0.05 and *** 0.01, respectively	Observations	6112	4398

and the influences of general, poor and very poor health on medical expenditures become stronger at the upper end of the distribution.

Finally, we compare results of the two-part model to results obtained from a Heckman selection model in Table 10, and we find no systematic differences.

Conclusion

In this paper, we apply a quantile regression method to investigate the heterogeneous effects of various determinants of medical expenditure in China. Comparing with the mean effects, we find that health care expenditures at the upper end of the distribution are under stronger influences of need factors such as poor health status, and weaker influences of socioeconomic factors and insurance status. On the other hand, health care expenditures at the lower end of the distribution are under stronger influences of socioeconomic factors and insurance status, and weaker influences of need factors.

Many countries have been trying to expand their public health insurance coverage in recent years. To achieve two fundamental policy goals—equity in health care utilization and control of health care costs—policymakers need a better understanding of the underlying determinants of individual health care expenditure beyond the results of mean regressions. This study may provide useful information to policymakers for the optimal design of their health care systems. Health care consumers in many countries face a linear copayment rate after they reach a given deductible. Our results suggest that a regressive co-payment schedule may promote equity in health care utilization and reduce overall health care costs to the society. A higher copayment rate at the lower end of the medical expenditure distribution may reduce health care demands that are not driven by need factors. On the other hand, medical expenditures at the upper end of the distribution are under weaker influences of health insurance compared to the need factors. A lower co-payment rate may not cause large increases of expenditures but improve equity. Although it has different cost implications to the public insurer and individuals, it may reduce overall health care costs from the perspective of the society. This result may be of particular interests to the health policymakers in China, where is currently still in a period of reshaping its health-care system.

Our study has two limitations. First, due to data limitation, we are unable to examine the effects of more detailed features of health insurance plans (such as deductibles, copayment rate, etc.), which may have important implications to the design of a health insurance plan and deserve some further studies. Secondly, although "Model" section shows that the potential endogeneity problems may not significantly change our fundamental conclusions; as a major challenge to most empirical studies, the endogeneity issue deserves some further exploration when it is possible.

References

Blomqvist, A. (1997). Optimal non-linear health insurance. *Journal of Health Economics*, *16*(3), 303–321.
Chiappori, P. A., Durand, F., & Geoffard, P. Y. (1998). Moral hazard and the demand for physician services: First lessons from a French natural experiment. *European Economic Review*, *42*, 499–511.
Cutler, D., & Zeckhauser, R. (2000). The anatomy of health insurance. In A. Culyer & J. Newhouse (Eds.), *Handbook of health economics*. Amsterdam: North Holland.

DiMatteo, L. (2003). The income elasticity of health care spending: A comparison of parametric and nonparametric approaches. *The European Journal of Health Economics*, *4*(1), 1–20.

DiMatteo, L., & DiMatteo, R. (1998). Evidence on the determinants of Canadian provincial government health expenditures: 1965–1991. *Journal of Health Economics*, *17*(2), 211–228.

Dow, W. H., & Norton, E. C. (2003). Choosing between and interpreting the heckit and two-part models for corner solutions. *Health Services & Outcomes Research Methodology*, *4*, 5–18.

Duarte, F. (2012). Price elasticity of expenditure across health care services. *Journal of health economics*, *31*, 824–841.

Duan, N., Manning, W. G., Morris, C. N., & Newhouse, J. P. (1983). A comparison of alternative models for the demand for medical care. *Journal of Business and Economic Statistics*, *1*, 115–126.

Duan, N., Manning, W., Morris, C., & Newhouse, J. (1984). Choosing between the sample-selection model and the multi-part model. *Journal of Business and Economic Statistics*, *2*, 283–289.

Duan, N., Manning, W., Morris, C., & Newhouse, J. (1985). Comments on selectivity bias. *Advances in Health Economics and Health Services Research*, *6*, 19–24.

Eichner, M. (1998). The demand for medical care: What people pay does matter. *The American Economic Review*, *88*, 117–121.

Ellis, R., & Manning, W. (2007). Optimal health insurance for prevention and treatment. *Journal of Health Economics*, *26*(6), 1128–1150.

Felder, S. (2008). To wait or to pay for medical treatment? Restraining ex-post moral hazard in health insurance. *Journal of Health Economics*, *27*(6), 1418–1422.

Gao, J., Tang, S., Tolhurst, R., & Rao, K. (2001). Changing access to health services in urban China: Implications for equity. *Health Policy and Planning*, *16*, 302–312.

Goldberger, A. S. (1983). Abormal selection bias. In S. Karlin & T. Amemiya (Eds.), *Studies in econometrics? Time series and multivariate statistics*. New York: Academic Press.

Hay, J. W., & Olsen, R. J. (1984). Let them eat cake: A note on comparing alternative models of the demand for medical care. *Journal of Business and Economic Statistics*, *2*, 279–282.

Hitiris, T., & Posnett, J. (1992). The determinants and effects of health expenditure in developed countries. *Journal of Health Economics*, *11*, 173–187.

Kaestner, R., & Dave, D. (2006). *Health insurance and ex ante moral hazard: Evidence from medicare*. NBER working paper no. 12764.

Kenkel, D. (1990). Consumer health information and the demand for medical care. *Review of Economics and Statistics*, *72*, 587–595.

Kowalski, A. (2010). *Censored quantile instrumental variables estimates of the price elasticity of expenditure on medical care*. NBER working paper no. 15085.

Leung, S. F., & Yu, S. (1996). On the choice between sample selection and two-part models. *Journal of Econometrics*, *72*, 197–229.

Lin, W., Liu, G. G., & Chen, G. (2009). The urban resident basic medical insurance: A landmark reform towards universal coverage in China. *Health Economic*, *18*(S2), S83–S96.

Liu, G. G., Zhao, Z., Cai, R., Yamada, T., & Yamada, T. (2002). Equity in health care access to: Assessing the urban health. *Social Science & Medicine*, *55*, 1779–1794.

Maddala, G. S. (1985). A survey of the literature on selectivity bias as it pertains to health care market. In R. Scheffler & L. Rossiter (Eds.), *Advances in health economics and health services research*. Greenwich, CT: JAI Press.

Manning, W. G., Blumberg, L., & Moulton, L. H. (1995). The demand for alcohol: The differential response to price. *Journal of Health Economics*, *14*, 123–148.

Manning, W. G., Newhouse, J., Duan, N., Keeler, E., & Leibowitz, A. (1987). Health insurance and the demand for medical care: Evidence from a randomized experiment. *The American Economic Review*, *77*(3), 251–277.

Mocan, N. H., Tekin, E., & Zax, J. S. (2004). The demand for medical care in urban China. *World Development*, *32*, 289–304.

Musgrove, P. (1983). Family health care spending in Latin America. *Journal of Health Economics*, *2*, 245–257.

Van Doorslaer, E., Wagstaff, A., et al. (2000). Equity in the delivery of health care in Europe and the U.S. *Journal of Health Economics*, *19*, 553–583.

Vera-Hernandez, M. (2003). Structural estimation of a principal-agent model: Moral hazard in medical insurance. *The RAND Journal of Economics*, *34*(4), 670–693.

Wagstaff, A., & Lindelow, M. (2008). Can insurance increase financial risk? The curious case of health insurance in China. *Journal of Health Economics*, *27*, 990–1005.

Wagstaff, A., Lindelow, M., Gao, J. X., & Qian, J. (2009). Extending health insurance to the rural population: An impact evaluation of China's new cooperative medical scheme. *Journal of Health Economics*, *28*, 1–19.

Wedig, G. J. (1988). Health status and the demand for health: Results on price elasticities. *Journal of Health Economics*, *7*, 151–163.

Yip, W. S., & Hsiao, W. C. (2008). The Chinese health system at a crossroads. *Health Affairs*, *27*, 460–468.

Yip, W. S., & Hsiao, W. C. (2009). Non-evidence-based policy: How effective is China's new cooperative medical scheme in reducing medical impoverishment? *Social Science & Medicine*, *68*, 201–209.

Zhao, Z. (2006). Income inequality, unequal health care access, and mortality in China. *Population and Development Review*, *32*, 461–483.

Zhong, H. (2011). Effect of patient reimbursement methods on health-care utilization: Evidence from China. *Health Economics*, *20*, 1312–1329.