Does Public Health Insurance Improve Healthcare Utilization and Health Outcome in China? Semiparametric Estimation of a Panel Data Model with Endogenous Treatment and Nonadditive Individual Heterogeneity

Shuyang Yang *

Department of Economics, Rutgers University

Abstract

This paper assesses the effect of public health insurance on healthcare utilization and health outcome, by evaluating the newest national public medical insurance program in China, Urban Resident Basic Insurance (URBMI). The voluntary enrollment scheme introduces a self-selection bias which may arise from unobserved time-varying factors and time-invariant individual heterogeneity. Moreover, since this is a national program covering a wide range of subpopulations, the effect of enrollment in health insurance may vary largely with individual characteristics. I address both concerns to estimate the causal effect of health insurance. This paper proposes to use a panel data model with endogenous treatment, which incorporates unobserved individual heterogeneity flexibly into the outcome model as an unknown function of observed time-invariant factors. The outcome model of discrete treatment is estimated in the context of a semiparametric setting with triple indexes. I first propose a two-stage semiparametric least square (SLS) method to consistently estimate the model parameters and then conduct a localized 2SLS procedure to recover the quantile treatment effect. Identification, consistency and $\sqrt{N}$-asymptotic normality of estimators for parameters and marginal effects are proved. Estimation result indicates that public health insurance program increases total health care expenditure by 11-16 percents. The marginal treatment effect varies largely across demographic groups. Having health insurance also shows positive impact on health outcomes, especially for children.

Keywords: Public Health Insurance; Program Evaluation; Healthcare Utilization; Semiparametric Estimation; Panel Data;Semiparametric least square; Localized 2SLS

*75 Hamilton Street,Department of Economics, Rutgers University, New Brunswick, New Jersey 08901

1. INTRODUCTION

Healthcare accessibility has been one of the social issues that gained the most public attention in China. By 2006, less than 40% of Chinese citizens were enrolled in medical insurance program of any kind, including private and public, which leaves more than 700 million people uninsured. As a result, unaffordable access to medical services leads to a high risk of medical impoverishment among households. In 2007, the Chinese government launched a new public health insurance program for urban residents in China,Urban Resident Basic Medical Insurance (URBMI), which is designed to cover all disavantaged urban residents who are not eligible for employment-based insurance, including elderly, children, students and unemployed adults.

In this paper, I examine the effect of Chinese Urban Resident Basic Medical Insurance (URBMI) after four years of program implementation, i.e. until 2011, using a panel survey dataset from Chinese Health and Nutrition Survey (CHNS). The evaluation is arranged around two main aspects of potential insurance benefits: healthcare utilization and health condition. By reducing the cost of healthcare, medical insurance is supposed to raise the usage of medical service. Therefore, increasing healthcare utilization is the direct policy goal of a public health insurance program. Ultimately, the increased consumption of healthcare utilization should translate into enrollees' better health status, which is an important measure of the ultimate policy goal of improving social welfare. Hence, I also investigate the impact on health outcomes from this program.

Evaluating the causal effect of enrolling in this public insurance program is not straightforward since the enrollment is voluntary. Endogeneity bias is likely to arise both from time-varying unobserved factors such as changing attitudes towards health and demand shocks on health related consumption, as well as time-invariant individual heterogeneity, such as long run overall health conditions. The argument goes as follows: an individual who values health more is more likely to enroll and at the same time, use health care services more often. On the other hand, an individual with better overall health tends to have lower demand for both health insurance and health care services. Hence, the sources of self-selection into insurance program can be multidimensional and the direction of bias is difficult to predict. Correcting for self-selection bias from multiple sources is the key effort of this paper.

In addition, my preliminary study using linear panel data model with IV and fixed-effect finds that the effect of URBMI varies largely with age, income, education level and geographic region. For example, having URBMI can raise children's healthcare utilization

by as large as 140% and increases the probability of being in good overall health condition by 15-20%, while it shows no significant effect on the elder population on any measures on health care usage and outcome. The heterogeneous marginal treatment effect calls for a more flexible econometric model which can incorporate individual observed and unobserved characteristics into the marginal effect function.

In this paper, I propose to use a general panel data model with endogenous and discrete treatment regressors. The nature of panel data allows one to handle sources of endogeneity which are unobserved but persistent over time ("fixed effect"). Departing from usual additive structure of such unobservables, such as fixed effect specified in linear models, this paper considers a more general class of nonseparable models. The main quantity of interest is the marginal treatment effect, which can vary by individual fixed effects. The model considered in this paper can be written as:

$$Y_{it} = F(X_{it}, a_i, T_{it}) + \epsilon_{it}, t = 1, ..., T \tag{1.1}$$

where $i = 1, 2, ..., N$, Y_{it} is a continuous outcome of individual i at time t, such as health care expenditure, in-patient treatment days, biomakers or subjective health ratings; X_{it} are observed explanatory variables; T_{it} is the discrete treatment variable, which indicates enrollment in URBMI; (a_i, ϵ_{it}) denotes unobservables, which can both be correlated with the enrollment indicator. Specifically, a_i is the time-invariant individual fixed effect, which enters the treatment function nonadditively.

Econometric literatures have extensively explored models with binary treatment and have extended the topic from parametric analysis to nonparametric ones, which allows the incremental effect of the discrete treatment to depend on other exogenous covariates flexibly. The main approach for tackling endogeneity in discrete treatment model is an IV estimator or nonparametric version of 2SLS. Das (2005)[4] developed a two-step estimator which substitutes nonparametric estimation of the instrument in the outcome model to identify the local average treatment effect. In a semiparametric context, Klein et.al(2015)[10] propose an IV estimator which is robust to misspecification of the treatment models, and also develop the distributional results for marginal treatment effect.

This paper builds on the growing literature of discrete treatment model with endogeneity but extends it to a panel data setting. Recent papers have been working on panel data models with fewer parametric restrictions. In most of these studies, individual heterogeneity enter the model in an additive manner. For example, Carroll et.al 2008[5] estimate a nonparametric panel data model with additive fixed-effect using first-differencing. Similarly, Soberon et.al [16] handles a panel data model with discrete treatment and varying

coefficient, using first differencing and local linear regression. Meanwhile, there are other papers study panel data models with nonadditive time-invariant unobservables by imposing various assumptions. Atonji et.al (2005 [1]) impose exchangeability assumptions on the distribution of individual fixed-effects conditioned on values of endogenous variable and use a control approach to estimate the marginal effect. Hoderlein (2012 [6]), which requires conditional independence of time-varying unobservables and endogenous regressors conditioned on individual fixed-effect, to identify local average response on a subpopulation of "stayer", whose explanatory variables stay unchanged over time. However, only continuous endogenous variables are considered in above mentioned models.

This paper also deals with nonseparable model but focus specifically on the nonadditivity of the time-invariant unobservables. It differs from existing literature in several ways. First of all, compared to general treatment effect models, I utilize information from repeated observations within an individual. Hence, marginal treatment effect conditioned on individual's time-invariant characteristics can be estimated, which pins down treatment effect more specifically to an individual. Compared to other panel data models with nonseparable structures, this paper deals with a discrete treatment variable. Many existing estimation methods do not apply, such as control function approach. Lastly, I do not impose restrictions of either exogeneity or stationality on the time-varying unobservables, i.e. ϵ_{it} in Eq (1.1). Hence, the endogeneity in this model can arise from either individual fixed-effect, or other time-varying unobservables, which is more applicable in many empirical analyses.

To identify the model, we adopt a flexible version of the modeling device proposed in Mundlak (1978[14]) and Chamberlain (1984 [3]), which assume that the unobserved individual heterogeneity is related to endogenous regressor only through the time-averages of exogenous variables. Denote $\bar{X}_i \equiv T^{-1}\sum_{t=1}^{T} x_{it}$, the time-average of all explanatory variable X. Assume:

$$a_i = f(\bar{x}_i) + \eta_i \tag{1.2}$$

where $f(.)$ is an unknown function of $\bar{x}_i$, and η_i is an idiosyncratic shock in the fixed-effect, which is independent from the discrete treatment. To be specific, the choice of health insurance purchase can be endogenous because of unobserved underlying health condition of an individual. An individual with a better overall health condition has lower demand for both health insurance and health care service. In the context of our modeling device, unobserved health conditions can be a function of the long run average of observed health measurements, such as blood pressure or physical functionality, or even subjective rating of health, which is represented by $f(\bar{x}_i)$. Meanwhile, there are some unobserved factors

in health conditions, such as genetic determinants. However, since such factors are not shown to the individual, they will not directly play a role in his choice of insurance purchase, i.e. the treatment. Hence, such factors will be modeled as idiosyncratic shock, η_i. Note that this shock is time-invariant as well. A similar approach has been used in various studies with panel data models, such as Semykina and Wooldridge. (2010[15]), and Maurer et.al 2011.[13]

To reduce dimensionality for feasibility of applications with smaller sample size, semiparametric model with triple-indices specification is estimated. We propose a two-step procedure to recover the marginal treatment effect. Due to the discrete nature of treatment variable, the outcome model can be represented in an additive separable potential outcome framework. This allows the implementation of a semiparametric least square (SLS) with "plug-in" propensity score as instrumental variable. Based on the estimated model parameters, a localized 2SLS procedure is conducted to recover the marginal treatment effect function. The rest of this paper will be organized as follow to obtain the result: in Section 2, I will describe the model in semiparametric context, and discuss condition for identification. Section 3 provides estimators for parameters and marginal treatment effect. Large sample properties including consistency and asymptotic distribution are discussed in Section 4. Lastly, a Monte Carlo simulation study of the proposed estimator is shown in Section 5, along with the estimation result on the quantile marginal treatment effect of public health insurance URBMI in China.

2. Econometric Model

Consider the following outcome model of discrete treatment, for individual i at time $t = 1,2,...,T$, denote Y_{it} as the continuous outcome, X_{it} a vector of exogenous variables, and T_{it} as a binary indicator of treatment [1]. We impose a general structure in the model as follows:

$$Y_{it} = F(X_{it}, a_i, T_{it}) + \epsilon_{it} \tag{2.1}$$

where a_i is the time-invariant individual heterogeneity which is not directed observed but can potentially correlate with treatment variable T_{it}. In addition, ϵ_{it} is a time-varying unobserved error term of Y_{it}. Conditional mean independence of ϵ from x is assumed $E(\epsilon|X) = 0$.

[1] assume only one treatment choice here but the model could be easily extended to multiple discrete treatment options

To complete the model specification, we assume a threshold crossing model for the binary treatment. For time period $t = 1, 2, ..., T$, individual chooses treatment at time t if the perceived gross benefit exceed some threshold:

$$T_{it} = 1\{g(Z_{1it}, b_i) > u_{it}\} \tag{2.2}$$

where Z_{it} denotes a vector of exogenous variables. b_i is an unobserved individual fixed effect in the treatment model. There is no additional assumption on the parametric distribution of error term u_{it}. Note that for the model to be identified, Z_{it} contains instrumental variables which are excluded in X_{it}.

Consider the limitation of nonparametric model in applications with smaller sample size, for estimation purposes, semiparametric specification is further imposed on the model. In outcome model, suppose exogenous variables X_{it} enter the model through a linear index $V_{1it} = X_{it}\beta$. Similarly, in the treatment model, Z_{it} take effect through linear index $V_{2it} = Z_{it}\gamma$. Index assumption yields the following model:

$$\begin{aligned} Y_{it} &= F(X_{it}\beta, a_i, T_{it}) + \epsilon_{it} \\ &= F(V_{1it}, a_i, T_{it}) + \epsilon_{it} \end{aligned} \tag{2.3}$$

$$\begin{aligned} T_{it} &= 1\{g(Z_{it}\gamma, b_i) > u_{it}\} \\ &= 1\{g(V_{2it}, b_i) > u_{it}\} \end{aligned} \tag{2.4}$$

2.1. Individual-Specific Heterogeneity

As motivated in Section I, assume that the individual-specific effect, although unobserved, depends on some time-invariant observed factors. Hence, here we model the individual fixed effect as a function of the time average of observed variables. Specifically, the following structure is considered in the outcome model:

$$a_i = f(\bar{X}_i) + \eta_i \tag{2.5}$$

where $\bar{X} = \frac{1}{T}\sum_{t=1}^{T} X_{it}$, a vector of time averaging value of each variable X_{it}. And η_i is a time-invariant shock to the individual fixed effect, which is independent of $\bar{X}_i$.

To construct the individual heterogeneity function, it is neither necessary to include time average of all variables in X_{it}, nor required to include excluded variables from the main model. For example, in previous case of health condition as individual heterogene-

ity, subjective health measures, attitude towards health may serve as better determinant for health condition than education or income. It is also unnecessary to include additional excluded variable. For identification purpose, it is only required for at least 1 variable in X_{it} to be time-varying. For simplification, we assume that $\bar{X}$ and X_{it} contain same set of variable.

The modeling for individual fixed-effect can be easily adopted in a wide range of empirical analysis. For example, in the study of return to college education, decision to attend college is a discrete but endogenous variable which may be correlated with unobserved long run individual ability. Similarly, individual ability can be decomposed into two parts. The first part of ability can be reflected by long run observed cognitive performances, such as average test scores (SAT, GPA etc.), as well as measurable non-cognitive skill sets. However, there can also be a second part of ability which is not reflected by individual's long run characteristics at the moment of college attending decision. For instance, the skill set that predicts academic performance may be different from the skill set that predicts working performance. Nevertheless, by the time of choosing college, abilities for work are not shown to individual and will not directly determine his choice of college attending.

To be consistent with the semiparatric outcome model, linear index assumption is imposed on the fixed-effect structure:

$$a_i = f(\bar{X}_i\alpha) + \eta_i \tag{2.6}$$

The original treatment model can be rewritten as a double-index model:

$$\begin{aligned} Y_{it} &= F(X_{it}\beta, a_i, T_{it}) + \epsilon_{it} \\ &= F(V_{1it}, f(Va_i) + \eta_i, T_{it}) + \epsilon_{it} \end{aligned} \tag{2.7}$$

Similarly, the treatment model follows the same structure for individual heterogeneity:

$$b_i = h(\bar{Z}_i\kappa) + \mu_i \tag{2.8}$$

It can be specified as a double-index model as well:

$$\begin{aligned} T_{it} &= 1\{g(Z_{it}\gamma, b_i) > u_{it}\} \\ &= 1\{g[V_{2it}, h(Vb_i) + \mu_i] > u_{it}\} \end{aligned} \tag{2.9}$$

Such specification has an advantage over traditional approach in which the individual-specific effect enters outcome model additively. It allows the unobserved individual het-

erogeneity to directly enter the treatment function, which enables the marginal effect to be individual specific. Although certain structure is imposed on a_i and b_i, the semiparametric approach still leave room for flexibility in the functional form of fixed effect. Lastly,in contrast to traditional differencing approach for additive specification, this model will preserve the individual fixed-effect during estimation, which will provide additional information specifically on the effect of individual heterogeneity itself.

2.2. IDENTIFICATION

The main identification strategy adopted in this paper is instrumental variable approach. Define the conditional probability function for treatment as:

$$\mathrm{E}(T_{it}|Z_{it},\bar{z_i}) = \mathrm{P}(T_{it}=1|Z_{it},\bar{Z}_i) \equiv p(Z_{it},\bar{Z}_i) \tag{2.10}$$

where $p(Z_{it},\bar{Z}_i)$ is an unknown function of instrumental variables z_{it}. Due to the discrete nature of treatment variable, we can write the outcome model can be represented in a separable manner. Based on the estimation of propensity score, a "plug-in" type of method will be conducted by replacing the endogenous treatment with with instrument to perform semiparametric-least-square. For identification, following assumptions on the instrument are specified:

A. 1 (Exclusion Restriction)**.** *Z is an $h \times 1$ vector of instrumental variable which includes X in the outcome model as subvector, i.e. there exists an additional ex*

A. 2 (Conditional Mean Restriction)**.** $E(\epsilon|z) = 0$.

The conditional mean independence is a weaker assumption than independence of ϵ from z, which is flexible in providing some form of heteroskedasticity from time-varying shocks(Das 2005[4]). Also note that the restriction is conditioned on Z_{it} $\forall\, t = 1,2,...,T$. Under usual index assumption, $p(Z_{it},\bar{Z}_i = E[T|Z_{it},\bar{Z}] = E[T|Z_{it}\gamma,\bar{Z}\kappa] = E[T|V_{2it},V_{bi}] = p(Z_{it},\bar{Z}_i)$, which implies the conditional mean independence of ϵ from the propensity score, i.e. $\mathrm{E}(\epsilon|p) = 0$.

A. 3 (Exogeneity of time-varying error)**.** *$Z \perp \eta$, which implies $p \perp \eta$.*

The outcome model can be rewritten as:

$$\begin{aligned}
Y_{it} &= F(X_{it}\beta, a_i, T_{it}) + \epsilon_{it} \\
&= F(V_{1it}, f(V_{ai}) + \eta_i, T_{it}) + \epsilon_{it} \\
&= \int_\Omega F(V_{1it}, f(V_{ai}) + \eta_i, T_{it}) dF(\eta_i | V_{1it}, V_{ai}, T_{it}) + [F(V_{1it}, f(V_{ai}) + \eta_i, T_{it}) \\
&- \int_\Omega F(V_{1it}, f(V_{ai}) + \eta_i, T_{it}) dF(\eta_i) | V_{1it}, V_{ai}, T_{it}] + \epsilon_{it} \\
&= E(F(V_{1it}, f(Va_i) + \eta_i, T_{it}) | V_{1it}, V_{ai}, T_{it}) + F(V_{1it}, f(V_{ai}) + \eta_i, T_{it}) \\
&- E(F(V_{1it}, f(Va_i) + \eta_i, T_{it}) | V_{1it}, V_{ai}, T_{it}) + \epsilon_{it} \\
&= G(V_{1it}, V_{ai}, T_{it}) + \delta_{it}
\end{aligned} \tag{2.11}$$

where the new error term $\delta_{it} = [F(V_{1it}, f(V_{ai}) + \eta_i, T_{it}) - E(F(V_{1it}, f(Va_i) + \eta_i, T_{it}) | V_{1it}, V_{ai}, T_{it}) + \epsilon_{it}$. Based on **Assumption A.1-A.3**, conditional mean independence applies to the new error term: $E(\delta_{it}|p) = 0$. Substituting (2.10) into the condition yields:

$$E(Y|p) = E[G(T)|p] \tag{2.12}$$

Therefore, in semiparametric setting, the above argument suggests minimizing an objective function, an SLS estimation approach proposed in Ichimura (1993 [7]), exploiting the discrete nature of treatment variable.

$$S(\beta, \alpha) = E\{[(Y - E(Y|V_1, V_a, P_{it})]^2\} = 0 \tag{2.13}$$

This model is hence specified as a triple-index semiparametric model, where (V_{1it}, V_{ai}, P) are the three indexes. An estimated treatment probability $\hat{P}$ will be used to replace P. Since P is an index with stand-along variable, in practice, only two sets of parameters (β, α) will be estimated. Additional index assumption is presented for identification.

A. 4 (Existence of time-varying variable)**.** *$\exists X_{1it}$, such that for some $t = j, k$, $X_{1ij} \neq X_{1ik}$. At least one variable in X is varying across time periods.*

In usual semiparametric model with more than one indices, standard assumption is the existence of one distinct variable in each index. Nevertheless, due to the nature of indices in our model specification, X_{it} in index V_{1it} and $\bar{X}_i$ in index V_a will contain a distinct variable as long as there exist 1 variable X_{1it} in the dataset, which indeed changed through time. With time-varying variable, $X_{1it} \neq \bar{X}_i$, which will serve as the excluded variable for each index. In empirical application, this requires to include at least one variable which is

changing over time, such as income, age, in the individual heterogeneity specification.

3. ESTIMATION

The key parameter in estimation is the marginal treatment effect. To begin with, rewrite the outcome model using a potential outcome framework. Suppose Y_1, Y_0 are the outcomes when treatment T_{it} is 1 or 0 respectively.

$$\begin{aligned} Y_{it} &= Y_{1it}T_{it} + Y_{0it}(1 - T_{it}) \\ &= (Y_{1it} - Y_{0it})T_{it} + Y_{0it} \end{aligned} \tag{3.1}$$

where $Y_{1it} = F(V_{1it}, a_i, T_{it} = 1) + \epsilon_{it}$ and $Y_{0it} = F(V_{1it}, a_i, T_{it} = 0) + \epsilon_{it}$. Hence, the outcome equation reduces down to:

$$\begin{aligned} Y_{it} &= [F(V_{1it}, a_i, T_{it} = 1) - F(V_{1it}, a_i, T_{it} = 0)]T_{it} + F(V_{1it}, a_i, T_{it} = 0) + \epsilon_{it} \\ &= [F(V_{1it}, f(V_{ai}) + \eta_i, T_{it} = 1) - F(V_{1it}, f(V_{ai}) + \eta_i, T_{it} = 0)]T_{it} + F(V_{1it}, f(Va_i) + \eta_i, T_{it} = 0) + \epsilon_{it} \\ &= M(V_{1it}, f(V_{ai}) + \eta_i) * T_{it} + B(V_{1it}, f(V_{ai}) + \eta_i) + \epsilon_{it} \end{aligned} \tag{3.2}$$

The individual marginal effect for individual i at time t is $M(V_{1it}, f(V_{ai}) + \eta_i)$. Since η_i cannot be separately identified from the main model, the object of interest for estimating marginal treatment effect will be:

$$\begin{aligned} MTE &= \int_\Omega M(V_{1it}, V_{ai} + \eta_i) dF(\eta_i) \\ &= \int_\Omega M(V_{1it}, V_{ai} + \eta_i) dF(\eta_i | V_{1it}, V_{ai}) \\ &= E[M(V_{1it}, V_{ai})] \\ &\equiv \bar{M}(V_{1it}, V_{ai}) \end{aligned} \tag{3.3}$$

In this paper, we propose to use a two-step method to estimate the average marginal effect: (1) The index parameters in V_{1it} and V_{ai} will first be estimated consistently from the main outcome model using 2- stage semiparametric least square method (SLS). (2) Based on the estimated indexes, the average marginal effect can be derived using a localized 2SLS approach. Detailed estimation strategy is provided in sessions below.

3.1. ESTIMATING INDEX PARAMTERS

As proposed in Section II, index parameters will be obtain by minimizing an objective function:

$$S(\beta,\alpha) = E\{[(Y - E(Y|V_{1it}, V_{ai}, P_{it})]^2\} = 0 \tag{3.4}$$

In sample analog of $S(\beta,\alpha)$, we minimize:

$$S(\beta,\alpha) = \frac{1}{T\times N}\sum_{i=1}^{N\times T}\{[(Y - \hat{E}(Y|V_{1it}, V_{ai}, P_{it})]^2\} \tag{3.5}$$

where $\hat{E}(Y|V_{1it}, V_{ai}, P_{it})$ is the kernel estimator of true expectation $E(Y|V_{1it}, V_{ai}, P_{it})$, which is given by:

$$\hat{E}(Y|V_{1it}, V_{ai}, P_{it}) = \frac{\sum_{i\neq j}\sum_t y_{it}\left\{\frac{1}{h_1}k\left(\frac{v_{1it}-V_{1jt}}{h_1}\right) * \frac{1}{h_2}k\left(\frac{v_{ai}-V_{aj}}{h_2}\right) * \frac{1}{h_3}k\left(\frac{P_{it}-P_{jt}}{h_3}\right)\right\}}{\sum_{i\neq j}\sum_t \frac{1}{h_1}k\left(\frac{v_{1it}-V_{1jt}}{h_1}\right) * \frac{1}{h_2}k\left(\frac{v_{ai}-V_{aj}}{h_2}\right) * \frac{1}{h_3}k\left(\frac{P_{it}-P_{jt}}{h_3}\right)} \tag{3.6}$$

Note that the propensity score P is unobserved and needs to be estimated.Recall from Section II, $P_{it} \equiv E(T_{it}|Z_{it}, \bar{Z}_i) = E(T_{it}|V_{2it}, V_{bi})$, under index assumption. A semiparametric binary response model with double-index is estimated here using maximum likelihood method to recover the propensity score (Klein 1993,2002 [12],[11]). The estimated propensity $\hat{P_{it}}$ will replace P_{it} in actual estimation process, which constitute a semiparametric version of 2SLS.

The above method is conducted by pooling observations from all time period $t = 1,...,T$. Alternatively, we can obtain the parameter by jointly minimizing the objective function in each period. For a particular period,

$$S_t(\beta,\alpha) = \frac{1}{N}\sum_{i=1}^{N}\{[(Y_{it} - \hat{E}(Y_{it}|V_{1it}, V_{ai}, P_{it})]^2\} \tag{3.7}$$

where $(\hat{E}|V_{1it}, V_{ai}, P_{it})$ is estimated data only in period t, whereas in pooling method all data are used in kernel estimation. Weighting matrix can be imposed on each period's condition, which is beneficial under heteroskedasticity of error term across time periods.

3.2. Estimating Marginal Effect

Replacing the marginal effect function with its mean, the outcome model is equivalent to:

$$\begin{aligned} Y_{it} &= E[M(V_{1it}, V_{ai})] * T_{it} + E[B(V_{1it}, V_{ai})] + [M(V_{1it}, f(Va_i) + \eta_i) - E[M(V_{1it}, V_{ai})] * T_{it} \\ &\quad + [B(V_{1it}, f(Va_i) + \eta_i) - E[B(V_{1it}, V_{ai})]] + \epsilon_{it} \\ &= E[M(V_{1it}, V_{ai})] * T_{it} + E[B(\hat{V}_{1it}, V_{ai})] + \xi_{it} \\ &= \bar{M}(V_{1it}, V_{ai}) * T_{it} + \bar{B}(V_{1it}, V_{ai}) + \xi_{it} \end{aligned} \tag{3.8}$$

where ξ_{it} is the sum of ϵ_{it} and the residuals from demeaning.
The estimation strategy is using localized two-stage-least-square. Using only observations (V_{1it}, V_{ai}) in a neighborhood of (V_{1jt}, V_{aj}), we can develop a local 2SLS estimator for $\bar{M}(V_{1jt}, V_{aj})$ and $\bar{B}(V_{1jt}, V_{aj})$ by replacing regressor T_{it} by its conditional expectation P_{it}, which is estimated from above session. Denote $R_{it} = [P_{it}\ 1]$ and $ME_{jt} = [\bar{M}(V_{1jt}, V_{aj})\ \bar{B}(V_{1it}, V_{ai})]$. Denote $\Delta_{ij} = R_i tME_{it} - R_{it}ME_{jt}$. The localized model can be written as:

$$Y_{jt} = R_{it}ME_{jt} + \Delta_{ij} + \xi_{it} \tag{3.9}$$

The local 2SLS estimator can be calculated as:

$$\hat{ME}_{jt} = [R' D_N(V_{1jt}, V_{aj}) R]^{-1} R' D_N(V_{1jt}, V_{aj}) Y \tag{3.10}$$

where the diagonal matrix $D_N(V_{1jt}, V_{aj})$ represents the weights for localization. To be specific:

$$D_N(V_{1jt}, V_{aj}) = diag\left\{ \frac{1}{h} k\Big(\frac{v_{1jt} - V_{1it}}{h} \Big) * \frac{1}{h} k\Big(\frac{v_{aj} - V_{ai}}{h} \Big) \right\} \tag{3.11}$$

The kernel function imposes heavy weight on observations close to (V_{1jt}, V_{aj}), which serves as the localization device in estimation. Through this process, the estimated marginal effect will be unbiased because the conditional mean of error component is zero.

$$\begin{aligned} E[\Delta_{it} + \xi_{it} | P_{it}] &= E(\Delta_{ij} | P_{it}) + E\Big[\big\{ M(V_{1it}, V_{ai}, \eta_i) - E[M(V_{it}, V_{ai})] \big\} * T_{it} \Big| P_{it} \Big] \\ &\quad + E\Big[B(V_{2it}, V_{ai}, \eta_i) - E[B(V_{1it}, V_{ai})] \Big| P_{it} \Big] + E[\epsilon_{it} | P_{it}] \\ &= 0 \end{aligned} \tag{3.12}$$

The kernel function imposes heavy weight on observations close to (V_{1jt}, V_{aj}), which

serves as the localization device in estimation. Through this process, the estimated marginal effect will consistent because the conditional mean of error component is zero.

Using the estimated index $\hat{V}_{1it}, \hat{V}_{ai}$, and estimated conditional propensity score $\hat{P}$ obtained from first-stage estimation, we can obtain the recover the marginal effect function from localized procedure.

After obtaining the marginal effect at every observation point of $(\hat{V}_{1it}, \hat{V}_{ai})$, follow Klein and Shen [2015 [10]], we summarize this information by quantile marginal effect. Let t_{qj} as an indicator for a particular variable of interest X_{jk} being in quantile q. Define population quantile marginal effect as:

$$\bar{M}_{qj} = \frac{E[t_{qj}\bar{M}(V_{1jt}, V_{aj})]}{E(t_{qj})} \tag{3.13}$$

The sample analog of quantile marginal effect can be presented by:

$$\hat{M}_{qj} = \frac{\sum_{j=1}^{N} \hat{t}_{qj}\hat{M}(\hat{V}_{jt}, \hat{V}_{aj})}{\sum_{j=1}^{N} \hat{t}_{qj}} \tag{3.14}$$

4. Large Sample Theory

4.1. Definitions

Definitions and notations are used in developing asymptotic theories are provided here.

D. 1. Treatment Probability.

Denote $P_{it} \equiv E(T_{it}|Z_{it}, \bar{Z}_i) = E(T_{it}|Z_{it}\gamma, \bar{Z}_i\kappa) = E[T_{it}|V_{2it}(\gamma), V_{bi}(\kappa)]$ under index assumption. The estimated conditional treatment probability is denoted as $\hat{P_{it}} = \hat{P_{it}}((V_{2it}(\gamma), V_{bi}(\kappa))) = \hat{E}[T_{it}|V_{2it}(\gamma), V_{bi}(\kappa)]$, which is given by:

$$\hat{E}(P|V_{2it}, V_{bi}) = \frac{\sum_{i\neq j}\sum_t T_{it}\left\{\frac{1}{h}k\left(\frac{v_{2it}-V_{2jt}}{h}\right) * \frac{1}{h}k\left(\frac{v_{bi}-V_{bj}}{h}\right) * \right\}}{\sum_{i\neq j}\sum_t \left\{\frac{1}{h}k\left(\frac{v_{2it}-V_{2jt}}{h}\right) * \frac{1}{h}k\left(\frac{v_{bi}-V_{bj}}{h}\right) * \right\}} \tag{4.1}$$

D. 2. Kernel.

Assume the total number of indices is m, denoted as $v_1, v_2, ..., v_m$. Define kernel $K \equiv \prod_1^m \frac{1}{h}k\left(\frac{v_{im}-V_{jm}}{h}\right)$, where $k(\cdot)$ is a symmetric density with bounded $\int z^2 k(z)dz$ and $h = O(N^{-r})$.

D. 3. Trimming

Denote λ as quantile fraction and W as a vector of variables and $q(\lambda)$ is the population

quantile vector for λ^{th} quantile. Define sample trimming function as an indicator function representing whether W_i is contained in the specified quantile:

$$\hat{t}_i(\hat{q}) \equiv 1\{\hat{q}(\lambda_1) < W_i < \hat{q}(\lambda_1)\}$$

This trimming function can be used to represent trimming on index as well when W_i denotes estimated indices.

4.2. Large Sample Results

Theorem 1. Consistency.
Denote $\theta = (\beta, \alpha)$, all parameters in outcome model, with $\hat{\theta}$ as the estimator which minimizes (3.5).
Under Assumption (A.1)-(A.5) and definition (D.1)-(D.3):

$$\hat{\theta} \xrightarrow{p} \theta_0$$

Proof. (Here the outline of proof will be provided. For intermediate result and lemma, one can refer to appendix for detail)
The main proof strategy is as follows. Recall from estimation, $\hat{\theta}$ minimizes the objective function $\hat{S}(\theta)$ given as:

$$\hat{\theta} = (\hat{\beta}, \hat{\alpha}) = \underset{\theta}{\operatorname{argmin}}\, \hat{S}(\theta, \hat{P}) = \underset{\theta}{\operatorname{argmin}}\, \frac{1}{T \times N} \sum_{i=1}^{N \times T} \{[Y - \hat{E}(Y|V_{1it}, V_{ai}, \hat{P}_{it})]^2\} \tag{4.2}$$

We can further establish a uniform convergence result:

$$\sup_{\theta} |\hat{S}(\theta, \hat{P}) - E[S(\theta)]| \xrightarrow{p} 0 \tag{4.3}$$

If the uniform limit of the moment condition is uniquely minimized at θ_0, it follows straightly that: $\hat{\theta} \xrightarrow{p} \theta_0$.
For uniform convergence of $\hat{S}(\theta)$, the above upper bound can be written as three pieces: A+B+C

$$\sup_{\theta} |\hat{S}(\theta, \hat{P}) - \hat{S}(\theta, P)| + \sup_{\theta} |\hat{S}(\theta, P) - S(\theta, P)| + \sup_{\theta} |S(\theta) - E[S(\theta, P)]| \tag{4.4}$$

The last piece C goes through straight forwardly because i.i.d. sample mean converges uni-

formly to its expectation.(Amemiya 1984 [2], Klein 1993[12]). Therefore, piece C is $O_p(1)$. For second piece B, there is no generated regressor $\hat{P}$. The difference of $S(\theta,P)$ from $\hat{S}(\theta,P)$ is that it uses the true expectation instead of estimated $\hat{E}$. Note that $|\hat{E}-E| = O(N^{-1/2}h^{-3})$ with three indices. It will converge faster than root-N if $r < 1/6$. With $S(\theta,P)$ a function of E, the uniform convergence result follows directly if $r < 1/6$.
The only piece left unproved is piece A with generated regressor $\hat{P}$. With $\sup_{\gamma,\kappa}|\hat{P}-P| = O_p(N^{-1/2}h^{-2})$ when estimated under specification of two indices. Perform a taylor series expansion on $\hat{S}(\theta,\hat{P})$ on P yields:

$$\hat{S}(\hat{P}) = \hat{S}(P) + \hat{S}'(P^+)(\hat{P}-P) \tag{4.5}$$

Therefore, $\hat{S}(\hat{P}) - \hat{S}(P) = \hat{S}'(P^+)(\hat{P}-P)$. Since $\sup_{\gamma,\kappa}|P^+-P| \xrightarrow{p} 0$, and $\hat{S}'(P) = \hat{G}(P) = \frac{1}{N*T}\sum(Y-\hat{E})\frac{\partial\hat{E}}{\partial P}$, which converges to the true gradient. Following argument similar to Klein (2010)[9], we can show that $\sup_\theta|\hat{S}(\theta,\hat{P}) - \hat{S}(\theta,P)| \xrightarrow{p} P$. Next step is to show that θ_0 is a unique minimizer of $E(S(\theta,P))$. First the expectation can be written as:

$$\begin{aligned}
E[S(\theta)] &= E\{\sum_{i=1}^{NT}[Y_{it} - E_{it}(\theta_0) + E_{it}(\theta_0) - E_{it}(\theta]^2\} \\
&= E\{\sum_{i=1}^{NT}[[Y_{it} - E_{it}(\theta_0)]^2 + [E_{it}(\theta_0) - E_{it}(\theta)]^2] + 2[Y_{it} - E_{it}(\theta_0)] * [E_{it}(\theta_0) - E_{it}(\theta)]\}] \\
&= E\big\{E\{\sum_{i=1}^{NT}[[Y_{it} - E_{it}(\theta_0)]^2 + [E_{it}(\theta_0) - E_{it}(\theta)]^2 + 2[Y_{it} - E_{it}(\theta_0)] * [E_{it}(\theta_0) - E_{it}(\theta)]]\}|X_{it}\beta, \bar{X}\gamma, P\big\} \\
&= E\big\{E\{\sum_{i=1}^{NT}[[\delta_{it}]^2 + [E_{it}(\theta_0) - E_{it}(\theta)]^2 + 2[\delta_{it}] * [E_{it}(\theta_0) - E_{it}(\theta)]]\}|X_{it}\beta, \bar{X}\gamma, P\big\} \\
&= E\big\{E\{\sum_{i=1}^{NT}[E_{it}(\theta_0) - E_{it}(\theta)]^2\}|X_{it}\beta, \bar{X}\gamma, P\big\}
\end{aligned} \tag{4.6}$$

The last step goes through due to conditional mean independence assumption of the instrumental variable P. Hence, for each observation it, θ_0 makes $E_{it}(\theta_0) - E_{it}(\theta) = 0$, which shows that θ_0 is an minimizer of $E[S(\theta)]$. Uniqueness of solution relies on index assumption can shown by similar argument from Ichimura et.al (1991[8]). □

Theorem 2. Normality Under Assumption (A.1)-(A.5) and definition (D.1)-(D.3):

$$\sqrt{N}(\hat{\theta}-\theta_0) \xrightarrow{d} W \sim N(0,\Sigma)$$

where

$$\Sigma \equiv H_0^{-1} E[\sqrt{N} G_0 G_0' \sqrt{N}] H_0^{-1}$$

$$\text{Gradient } G_0 \equiv E[\nabla_\theta S(\theta_{0_0})]$$

$$\text{Hessian } H_0 \equiv E[\nabla_\theta G(\theta_{0_0})]$$

Proof. Starting with the first-order condition (F.O.C) for objective function $\hat{S}(\theta, P)$,denote:

$$\hat{G}(\hat{\theta}) = \frac{1}{NT} \sum_{i=1}^{NT} (Y_{it} - \hat{E}_{it}) \frac{\partial \hat{E_{it}}}{\partial \theta}. \tag{4.7}$$

Using Taylor Expansion on true parameter θ_0, and $\theta^+ \in (\hat{\theta}, \theta_0)$,as an intermediate point, the above gradient can be written as:

$$\hat{G}(\hat{\theta}) = \hat{G}(\theta_0) + \hat{H}(\theta^+)(\hat{\theta} - \theta_0) \tag{4.8}$$

where $\hat{H(\theta)} \equiv \nabla_t heta \hat{G}(\theta)$, and it follows that:

$$\sqrt{N}(\hat{\theta} - \theta_0) = -H^{-1}(\theta^+) \sqrt{N} \hat{G}(\theta_0) \tag{4.9}$$

As outline of a proof strategy, it will be conducted in the following step: (Work in Progress)

- $\sup_\theta |\hat{H}(\theta) - E(H(\theta))| \xrightarrow{p} 0$
- $\theta^+ \xrightarrow{p} \theta_0$, which follows that $H^{-1}(\theta^+) \xrightarrow{p} E(H(\theta_0)) \equiv H_0$
- For $\sqrt{N}\hat{G}(\theta_0) = \sqrt{N} \frac{1}{NT} \sum_{i=1}^{NT} (Y_{it} - \hat{E}_{it}) \frac{\partial \hat{E_{it}}}{\partial \theta}$, the proof strategy is by showing the convergence of the gradient function to the true function without estimated components. With the establishment of such argument and bias reduction mechanism from Klein and Shen[2015[10]], asymptotic normality will follow from standard central limit theorem.

□

Theorem 3. Properties of Quantile Marginal Treatment Effect Under Assumption (A.1)-(A.5) and definition (D.1)-(D.3): quantile marginal effect is consistent and asymptotically normal.

$$\hat{M}_q \xrightarrow{p} M_q$$

$$\sqrt{N}(\hat{M}_q - M_q) \xrightarrow{d} W \sim N(0, \Omega)$$

5. Monte Carlo Result

5.1. Explore forms of F-function

In the first Monte Carlo study, I use different specification of outcome function to explore the estimation method for indexes. The general structure of model is as follow:

$$\begin{aligned} T &= 1\{g(V_{2it}, b_i) > u_{it}\} \\ &= 1\{V_{2it} + b_i > u_{it}\} \\ &= 1\{Z_{it}\gamma + \bar{Z}_i\theta + \mu_i > u_{it}\} \end{aligned} \tag{5.1}$$

$$\begin{aligned} Y_{it} &= F(V_{1it}, a_i, T_{it}) + \epsilon_{it} \\ &= F(X_{it}\beta, f(\bar{X}_i\alpha) + \eta_i, T_{it}) + \epsilon_{it} \end{aligned} \tag{5.2}$$

In this subsection, data is generated to satisfy the following:

1. Error terms in both treatment and outcome models are homoskedastic and standard normally distributed. u_{it} and ϵ_{it} are correlated.
2. X_{1i},X_{2i} and X_{3i} are all normally distributed with expectation 0 and standard deviation 1. X_{1i} is correlated with X_{2i}. X_{3i} is generated to be independent of X_{1i} and X_{2i}.
3. Serial correlation: X_{1it} is correlated with X_{1is}. Same for X_2 and X_3.
4. $Z_{it} = [X_{1it}X_{2it}X_{3it}]$. $X_{it} = [X_{1it}X_{2it}]$

Four variation in specification of outcome model can be explored:

1. Linear Marginal Effect in V_{1it} and a_i:
   $F(V_{1it}, a_i, T_{it}) = 2 * V_{1it} * T_{it} + b_i * T_{it}$
2. Non-linear Marginal Effect in V_{2it} and V_b:
   $F(V_{1it}, a_i, T_{it}) = exp\{2V_{1it} + a_i\} * T_{it} + (V_{1it} + 0.26 * a_i)$
3. Quadratic in index V_{1it} V_{ai}:
   $F(V_{1it}, V_{ai}, T_{it}) = V_{1it}^2 * T_{it} + a_i^2 * T_{it}$

Using bias correction, Sample size $N = 2000$ and repetition of $i = 100$. Window size $r = 1/11$. For each functional specification, I reported the estimates after bias correction.

Specification	Parameter	True Value	Mean	Standard Deviation	Median
(1.1)	β	2	2.26	0.12	2.25
(1.1)	α	1	0.95	0.114	0.95
(1.2)	β	2	2.16	0.15	2.16
(1.2)	α	1	0.91	0.37	0.87
(1.3)	β	2	2.12	0.13	2.12
(1.3)	α	1	0.96	0.27	0.96

Note: Here I set individual fixed effect $b_i = V_{bi}^2 + \eta_i$; Result is similar if $b_i = V_{bi} + \eta_i$.

5.2. SEPERATELY ESTIMATE EACH TIME PERIOD

Due to the fact that X_{it} and $\bar{X}_i$ can be highly correlated in practice if the variation in time dimension is limited, here we use a Monte-carlo study to compare estimation method of pooling data with seperating estimation, which is given by (3.7). The following study uses design (1) above with bias correction. The result confirms the linear issue in index variable specification. As the number of period increases, the estimation gets closer to the true value with lower standard devidation.

Period	Parameter	True Value	Mean	Standard Deviation	Median
2	β	2	2.22	0.12	2.25
2	α	2	1.20	0.24	1.15
3	β	1	2.17	0.14	2.17
3	α	2	1.15	0.13	1.15
4	β	2	2.06	0.07	2.06
4	α	1	1.22	0.10	1.22
5	β	2	2.02	0.09	2.02
5	α	2	1.14	0.16	1.34
6	β	2	1.97	0.09	1.97
6	α	2	1.41	0.18	1.41

6. Empirical Analysis for Chinese Public Health Insurance

6.1. Program Background

The URBMI is a central government directed and local government administered public insurance program, which is similar to the institutional setting of Medicaid program in the US to some extent. The role of Chinese central government in URBMI program includes: initiating the program by offering fiscal support, mandating the implementation by passing a series of legislation, and supervising the details by providing policy guidelines. The detailed design and implementation of URBMI policy, however, are up to the city governments. Therefore, URBMI policy varies considerably across different regions.

As for eligibility and enrollment, URBMI is a voluntary-based program, targeting elderly, children, adults with disability and other non-working urban residents. In most of the cities, enrollment is individual-based, with several exception of household-based requirement to avoid adverse selection. For program financing, URBMI is jointly financed by individual and governments. The average annual cost of insurance is around 250 RMB ($40 USD) for adults, and 120 RMB ($19 USD) for children. Individual contributes less than 50% of the insurance premium while central government and local government share the rest of funding. To ensure program implementation, central government subsidizes at least 80 RMB ($12 USD) per enrollee annually, with extra funding for enrollees with disability or under poverty line. Local level subsidy varies across region but a minimum contribution is required by central government.

Benefit package varies largely across cities in terms of reimbursement rate, payment deductible and reimbursement cap. At the beginning stage, URBMI was only designed to pay for inpatient treatment and outpatient treatment of severe illness such as chronic disease. As the program developed, a much broader range of medical services are included in the insurance coverage, including preventative care. On average, inpatient reimbursement rate is at least 45%, which differs across different levels of medical facilities. The reimbursement is less generous for treatment received in higher-level medical provider. For instance, if a patient is treated in community clinic, the reimbursement rate can be as high as 90%, whereas the cost can only be covered by 45% in city-level big hospital. Moreover, the reimbursement ceiling is set to be 4 6 times the average annual income of urban workers, which is about 25,000-150,000 RMB ($4,100 – $24,600 USD).

6.2. DATA

6.2.1. INDIVIDUAL LEVEL DATA

The dataset used in this paper is the data from Chinese Health and Nutrition Survey (CHNS), which is an ongoing, open cohort survey project conducted by Carolina Population Center at the University of North Carolina at Chapel Hill and the National Institute of Nutrition and Food Safety at the Chinese Center for Disease Control and Prevention. The survey collects rich information on individual and household' s demographic and socioeconomic characteristics, as well as health and nutrition status of both urban and rural population. It also carried out community survey to provide information on community facilities, healthcare provision and public insurance enrollment etc.

For the purpose of my research, I select the data from CHNS dataset in the following manner. First of all, only the last three waves of data, i.e. wave 2006, 2009 and 2011, are selected for analysis. Year 2006 is the year right before URBMI was introduced. Including this wave provides a baseline condition before treatment. Wave 2009 and 2011 data provides the variation after treatment.

The eligible individuals in the survey sample are selected to be policy target population (children under 18, elders above 55 or 60 depending on gender, college student and unemployed adult) who are registered in urban status in Chinese Household Registration System ("Hukou"). The eligible sample contians 4,534 observations in 2006, 4,623 observations in 2009 and 7,279 observations in 2011. Noticing that the sample expanded in 2011 due to an expansion in survey size. In addition to eligible sample, various subsamples are also defined as in Table2 including policy-target sample, elderly sample and children sample. In the empirical analysis, I will mainly use the eligible sample as my study sample, which will be referred to as full sample below. Detail description of different samples used in this paper are provided in Figure A.1.

The key variable of interest is enrollment status into URBMI. By default, in wave2006, no observations were enrolled in URBMI. The trend of enrollment rate in various samples is presented in Figure A.2. After it is first launched in 2007, the enrollment rate among different subsamples exhibits a rapidly increasing trend. The average enrollment rate reached over 30% for all sample, among which sample of children under 18 has the highest enrollment rate of 54% in 2011. The changing enrollment status provides a decent size of variation to utilize in identification.

Other explanatory variables include observable individual, household and commu-

nity characteristics that may potentially affect the outcome variables. There are potentially two sets of controlled variables. For the individual and household level controls, I include basic demographic information including gender, marital status, education level, individual income, as well as household characteristics including household size, per capita household income. I also controls for individual's insurance status on other competing public or private insurance program, which includes commercial medical insurance, UEBMI program, rural cooperative public insurance and other medical subsidy by the government. The second set of controls is community characteristics.

Summary statistics of selected demographic characteristics and insurance status are presented in Table A.1. The first column summarizes the mean and standard deviation of various explanatory variables of the full sample. Column (2) –(5) presents the summary statistics by waves. For wave 2009 and 2011, the comparisons between URBMI enrollee and non-enrollee are also presented, along with test result for difference in sample mean. As indicated by statistics, the characteristics of enrollees significantly differ from those who are not enrolled. Enrollees are more likely to be female, married and less educated. Moreover, URBMI enrollees have significantly lower household incomes. As for the medical insurance status, enrollees in URBMI are less likely to purchase commercial insurance and other public medical insurance. The differential characteristics for insured and uninsured group indicate the existence of self-selection into the URBMI, which may cause biased estimation.

6.2.2. City-level Data: URBMI Insurance Policy

In addition to CHNS individual level dataset, city-level insurance policy variables of URBMI are also used in empirical analysis as instrumental variables. Each city determines its URBMI package for its residents. The policy variables include insurance premium, reimbursement rate for different level of medical providers and treatment type, insurance deductible for different treatment as well as lump-sum reimbursement cap. Within a city, the insurance policy varies in four categories of subpopulation, i.e. elderly, children, college students and adults. In addition, the policy variables are changing from year to year. These policy variables can only be manually collected from local government's legal documents.

Table A.2 reports the summary statistics of URBMI insurance premium and reimbursement cap. In particular the mean of premium and cap is calculated for four groups of urban population, including elderly, children, students and adults. In general, URBMI policy is the most generous to children and college students, with lower enrollment cost

and higher reimbursement cap. Moreover, from 2009 to 2011, the average reimbursement cap largely increased for all subpopulation. However, the regional pattern is harder to generalize. Comparing higher and lower income cities within a province, higher income cities tend to charge higher premium for enrollees but on the other hand, set a more generous reimbursement policy. As for geographical region, cities in Eastern China shows a much higher generosity in reimbursement cap, while middle and western cities have relatively similar polity despite the fact that mid-China is more economically advanced than western part. Hence, the policy summary informally indicates that URBMI policy is somewhat exogenous to a city's characteristics.

6.3. Parametric Results

In this section, the estimation result for the effect of URBMI on healthcare utilization and health outcome variables will be presented. Full sample, i.e. the URBMI eligible sample, is used in estimation throughout this section. For each dependent variable, four model specifications will be analyzed: 1) pooled OLS model for comparison purpose; 2) individual fixed-effect model as the baseline model; 3) fixed-effect 2SLS model with insurance premium as IV; 4)Fixed-effect 3-stage model with insurance premium IV. Throughout this section, only the marginal effect of primary interest are reported and discussed.

Table A.3 presents the estimated coefficient/marginal effect on variable of interest, URBMI enrollment status. I further categorize healthcare utilization into four categories: general healthcare utilization; intensity of treatment; financial burden and treatment for chronic disease.

For Category A: general healthcare choice, it mainly measures the probability of using medical service and general quality of service selected. In particular, it measures the probability of visiting formal medical provider, probability of seeking preventative care, and the choice of medical provider's type. Moreover, these variables are decision variables mainly based on individual's own choice, with less intervention from medical provider. As result, URBMI may have more direct impact on these variables. The regression results from Table6 show that under pooled OLS model, URBMI enrollment shows significant effect on all of the four variables. URBMI enrollment increases the probability of seeking formal healthcare and the probability of seeking preventative care. It also induces enrollees to choose higher quality medical provider, by increasing the probability of visiting city-level hospital and decreasing that of community-level hospital. However, after adding individual fixed effect, URBMI only significantly influence the choice of preventative care. The coefficients

of URBMI for the remaining three variables decrease in magnitude and lose significance. The parity of estimation result between pooled OLS and fixed effect model further confirms the existence of unobserved time invariant factors in error term that are correlated with variable URBMI. Under instrumental variable approach with individual fixed-effect, the impact of URBMI is significant for both the choice of seeking formal healthcare and the choice of preventative care. The magnitude of coefficients also largely increases. Depending on the specific IV model, an enrollment into URBMI is estimated to increase the probability of visiting formal healthcare by 17-21 percentage points, which is higher than the findings in Liu et al. (2012). Meanwhile, enrolling into URBMI will raise the chance of seeking preventative care by 3.8-4.2 percentage points.

Comparing results from different model specification, one thing worth discussing is the sign of endogeneity bias caused by different source of unobserved factors. Noticing that by adding individual fixed-effect into the model, it decreases the size of URBMI effect, which means that time invariant factors in the model tend to bias up the coefficient. However, by using instruments that further adjust for bias from other factors, the magnitude of effect increases largely, meaning that other unobserved factors are likely to have bias down the coefficient. The channel through which these different factors affect the result should be examined with care. One possible explanation is provided as follow: The source of time invariant individual factors are likely to be values towards health, which is arguably persistent across time, while the time varying factors are more likely to capture individual's behavioral characteristics, which are considered more easily changing, especially in a fast developing environment as China. For time invariant factors, a person who values health more is more likely to purchase URBMI, while at the same time more likely to utilize healthcare service. Therefore, this positive chain of correlation is considered to bias up the coefficient.. As for time varying behavioral factors, it is suspected that an individual who enrolled in URBMI is more possible to engage in healthier behavior and therefore has less demand for healthcare utilization due to better health. This chain of connection tends to bias down the effect of URBMI. Comparing the magnitude of bias, time varying factors seems to play a more significant role in the model. However, this is only a possible intuition for explaining the comparative result. To further confirm the underlying mechanism, a more rigorous identification strategy or even a behavioral model should be used in the context.

For category B, it measures the intensity of treatment received. In other words, it provides information on not only whether the individual use healthcare, but also how much quantity or quality of medical service is used. Referring to Table6, after adjusting for endo-

geneity bias, URBMI enrollment only shows statistically significant effect on inpatient days, which is a small decrease for about 0.3 day. The sign of estimated coefficient is counter intuitive in some sense. However, the estimation on inpatient treatment can be imprecise due to the limited observation number in the sample. Only about 10% of individuals who seek healthcare service will received inpatient treatment, which limits the variation in observation. The effect of URBMI on total expenditure is estimated to be a 3 5% positive increase, but with a too large standard deviation to be considered statistically significant. In general, URBMI shows no noticeable impact on this set of total utilization variables. Nevertheless, it should be taken into account that the choice for treatment is a joint decision between both the patient and the medical provider, which is beyond the control of an individual. As a national insurance program, the size of URBMI is very likely to induce supply side change from hospitals and doctors in terms of how they treat patients enrolled in URBMI. Without further exploration of supply-side response, it is difficult to draw a definite conclusion on the effect of URBMI on total utilization of healthcare. In addition to the general measurement of utilization, I also examine treatment usage for individuals with chronic diseases, as shown in Category D. In pooled OLS model, URBMI enrollment indicates a significantly higher adoption rate of treatment for both high blood pressure and diabetes. However, after adjusting for endogeneity bias, the effect disappears and even become negative in some case.

Table A.4 presents the regression results on selected physical examination outcomes. URBMI enrollment significantly decreases the probability of getting high blood pressure test result. In particular, after correcting for bias, enrolling in URBMI is associated with about 16-percentage-points less chance of showing high blood pressure symptoms. The coefficient is considerably large in magnitude. Except for high blood pressure, URBMI enrollment does not impose sizable effect on other examination outcome variables. However, the result is consistent with expectation due to the following two reasons. Firstly, physical examination result is only available for 2009, which is only 2 years after the launce of URBMI at maximum . The effective period of URBMI is still too short for significant impact to take place. Secondly, as discussed in Section 4.1.3, the physical examination conducted in the survey is in the most basic version. The symptoms checked in the examination may not be prevailing in urban area, such as goiter. Hence, there are vary limited variation in the sample to provide a significant coefficient. However, the signs of coefficient for all these variables are negative, which gives some information about the direction of URBMI effect.

The second set of health outcome variables are self-reported, which are presented below in Table A.5. Overall wellbeing is a general self-rating for one's condition, whose pos-

sible answers are multi-categorical but grouped into two categories, positive or negative, for simplification. Under the instrumental variable model, URBMI enrollment raised the probability of a positive wellbeing rating by about 13 percentage points. The effect on psychological wellbeing is even more significant. URBMI increases an individual's chance of feeling happy during the past year by more than 20% under IV model. This finding is consistent with other literature on health insurance. The short-term effect of enrolling into medical insurance includes increasing happiness (Finkelstein et al. 2012), although the underlying mechanism of this effect has not been studied yet.

7. Semiparametric Estimation Result (Work in Progress)

A. Appendix:Tables and Figures

A.1. Summary Statistics

Figure A.1: Sample Definition and Sample Size

Sample	2006	2009	2011	Total
Eligible Sample				
	All survey respondent who registered as "Urban" status (Urban "*Hukou*"), and NOT enrolled in government employee health insurance			
	4,534	4,623	7,279	16,436
	27.59%	28.13%	44.29%	100%
Target Sample				
	Subsample of eligible sample, excluding adults who are employed and aged between 18-60 for male or 18-55 for female			
	2,783	2,869	4,335	9,987
	27.87%	28.73%	43.41%	100%
Elderly Sample				
	Subsample of eligible sample: male respondent age >=60 and female age respondent age >=55 in full sample			
	1,268	1,478	2,354	5,100
	24.68%	28.98%	46.16%	100%
Children Sample				
	Subsample of eligible sample: age <= 18			
	509	456	777	1,742
	29.22%	26.18%	44.60%	100%

Figure A.2: Enrollment Rate in URBMI from 2006 to 2011

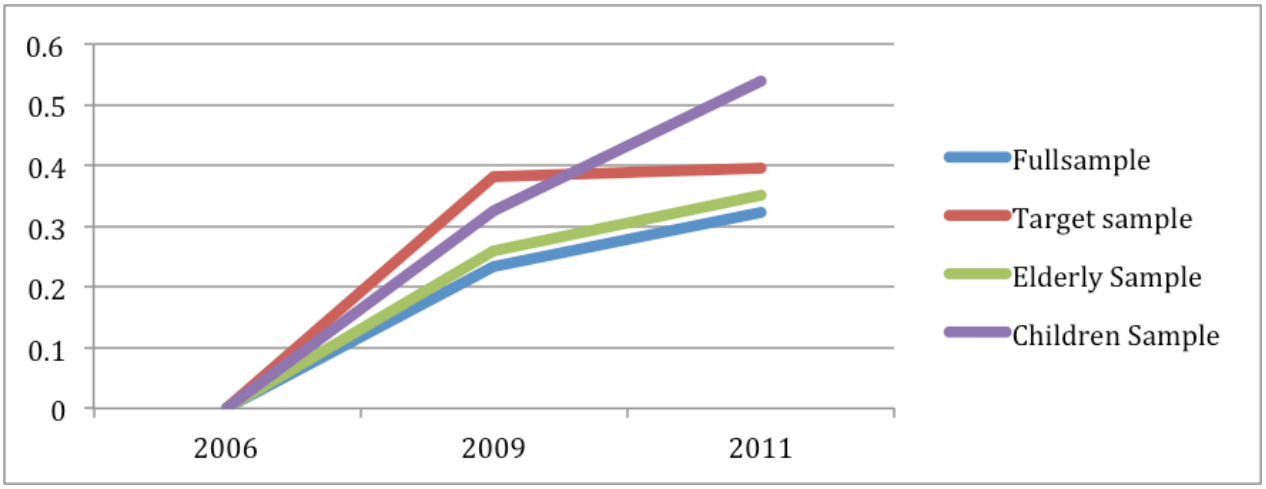

Table A.1: Summary Statistics of Explanatory Variables

Full Sample		Wave2006	Wave 2009			Wave2011		
			Untreated	Treated		Untreated	Treated	
Observation	16,436	4,534	3,535	1,088		5,026	2,253	
Demographic Characteristics								
Female	0.52	0.52	0.5	0.56	***	0.51	0.57	***
Age	46.63	45.59	47.07	47.92		47.01	46.58	
Married	0.73	0.72	0.75	0.69	***	0.77	0.66	***
Education:								
Less than high school	0.59	0.61	0.6	0.71	***	0.49	0.68	***
High school	0.28	0.3	0.3	0.25	***	0.28	0.24	***
College or above	0.13	0.09	0.11	0.04	***	0.23	0.08	***
Household size	3.4	3.43	3.42	3.44		3.35	3.38	
Household income per capita	14917	8917	13965	11591	***	20389	17635	***
Health Insurance Status								
Commercial Insurance	0.06	0.05	0.07	0.04	***	0.07	0.06	**
Employee Insurance	0.38	0.31	0.49	0	***	0.63	0.01	***
Other Medical Subsidy	0.02	0.02	0.02	0.01	***	0.03	0.02	*

Notes: columns with asterisk reports the significance level of sample mean test result, *p<0.10, **p<0.05, ***p<0.01

Table A.2: Summary Statistics of URBMI Policy

	Elderly	Children	Student	Adult
Premium (in 1 RMB)				
All Cities	140.45	38.25	47.58	218.49
All Cities 2009	132.09	32.5	52.87	210.82
All Cities 2011	148.58	39.29	39.6	229.75
High Income Cities	161.89	44.55	60.91	285.23
Low Income Cities	113.02	30.88	31.67	134.69
Eastern China	206.42	49.29	61.79	318.57
Mid-China	107.69	25.31	37.71	167.17
Western China	100.5	43.5	43.5	160.5
Reimbursement Cap (in 1,000 RMB)				
All Cities	66.2	70.75	73.2	66.44
All Cities 2009	46.17	50.61	52.28	46.17
All Cities 2011	84.23	88.88	92.03	84.68
Provincial Capitals	86.9	91.3	95.95	87.35
Lower Income Cities	42.22	46.63	46.63	42.22
Eastern China	89.49	92.72	97.49	89.95
Mid-China	49.21	57.33	59.21	49.21
Western China	62.78	62.89	63	63.11

Table A.3: Effect of URBMI on Healthcare Utilization

Model	Pooled OLS		Fixed Effect		FE 2SLS		FE 3-Stage IV	
Category A: General Healthcare Choice								
Seek formal medical service in past 4 weeks								
	0.0211	***	0.0064		0.1766		0.2176	*
Obs=16,125	-0.008		-0.012		-0.082		-0.118	
Visit city-level hospital in past 4 weeks								
	0.0633	**	-0.038		0.2911		0.0677	
Obs=2,188	-0.03		-0.088		-0.28		-0.205	
Visit community-level hospital in past 4 weeks								
	-0.11	***	-0.014		-0.122		0.0394	
Obs=2,188	-0.032		-0.106		-0.338		-0.244	
Seek preventative care in past 4 weeks								
	0.0135	**	0.0257	***	0.03868	**	0.04205	*
Obs=16,125	-0.006		-0.01		-0.01716		-0.02418	
Category B: Intensity of Utilization								
Inpatient treatment in past 4 week?								
	0.0016		-0.00449		-0.15842		-0.19425	
Obs=16,125	-0.003		-0.00514		-0.15598		-0.20237	
Inpatient days in past 4 week								
	0.0255		-0.03568		-0.24989	**	-0.32601	*
Obs=16,125	-0.037		-0.06923		-0.12392		-0.17314	
Ln (total medical expenditure+1)								
	0.1103	**	0.01184		0.0329		0.05727	
Obs=16,125	-0.046		-0.07631		-0.1366		-0.19057	
Category C: Financial Burden from Healthcare Service								
Ln (out-of-pocket expense+1)								
	-0.09675	**	-0.00638		-0.02281		-0.03324	
Obs=16,125	-0.04342		-0.07285		-0.1304		-0.18193	
Category D: Treatment for Chronic Disease								
Receive high blood pressure treatment								
	0.0993	***	0.029		0.0619		-0.032	
Obs=1,221	-0.037		-0.095		-0.183		-0.165	
Receive diabetes treatment								
	0.1153	*	0.1356		-0.294		-0.267	
Obs=302	-0.066		-0.448		-0.496		-0.86	

Table A.4: Effect of URBMI on Physical Examination Outcomes

Model	Pooled OLS		Fixed Effect		FE 2SLS		FE 3-Stage IV	
High Blood Pressure (wave=2006,2009)								
	-0.0302	*	-0.0919	***	-0.1638	***	-0.15492	***
Obs=8935	-0.0172		-0.031		-0.0619		-0.001	
Obesity (wave=2006,2009)								
	-0.0236		-0.0422		-0.0368		-0.04568	
Obs=8935	-0.0175		-0.0278		-0.0555		-0.06211	
Goiter (wave=2006,2009)								
	6E-05		-0.0004		-0.0048		-0.00552	
Obs=8344	-0.0013		-0.0021		-0.0042		-0.00463	
Angular Stomatitis Symptom (wave=2006,2009)								
	0.0003		-0.0031	*	-0.0011		-0.00274	
Obs=8344	-0.0009		-0.0018		-0.0036		-0.004	
Any symptoms (wave=2006,2009)								
	-0.0055	**	0.0032		-0.007		-0.00894	
Obs=8935	-0.0026		-0.0044		-0.0092		-0.01029	

Table A.5: Effect of URBMI on Self-reported Health Outcomes

Model	Pooled OLS		Fixed Effect		FE 2SLS		FE 3-Stage IV	
Overall wellbeing for the past year (wave=2006,2009)								
	0.0046		-0.0235		0.1355	**	0.1233	*
Obs=8935	-0.0158		-0.02736		-0.057		-0.0639	
Happiness for the past year (wave=2006,2009)								
	0.05	***	0.0547	**	0.2088	***	0.2853	***
Obs=8935	-0.0154		-0.0273		-0.0568		-0.0643	
Feeling sick during past 4 week (wave=2006,2009,2011)								
	-0.0171	***	-0.00771		0.0334		0.0345	
Obs=16089	-0.0086		-0.0128		-0.0353		-0.0349	
Have obvious symptoms during past 4 weeks (wave=2006,2009,2011)								
	-0.0372	***	0.00054		0.0159		0.0064	
Obs=16125	-0.0096		-0.0145		-0.0403		-0.0394	
Suffering from chronic disease (wave=2006,2009)								
	0.0159		-0.0232		-0.0429		-0.0532	
Obs=8935	-0.0132		-0.02031		-0.0424		-0.0473	

B. Appendix: Index Assumption

To obtain the treatment effect, it would be useful to estimate the conditional mean of outcome Y_{it}: (Note: Not sure whether I should condition on the individual heterogeneity as well??)

$$E[Y|X_{it}, T_{it}, b_i] \tag{B.1}$$

To estimate the conditional mean, it can be proceeded to estimate the outcome and treatment model simultaneously by conditioning on all the indexes in the two models and recover the index parameters, i.e. to estimate

$$E[Y|X_{it}, T_{it}, b_i] = E[Y|V_{1it}, V_{2it}, V_{ai}, V_{bi}] \tag{B.2}$$

To further simplify, it would be desirable to conduct a two-stage procedure to estimate the treatment model first and replace true treatment variable by the treatment probability (propensity score). If the following condition holds, the outcome estimation could be

represented by a triple-index model instead of four indexes.

$$E[Y|V_{1it}, V_{2it}, V_{ai}, V_{bi}] = E[Y|V_{2it}, V_{bi}, T_{it}] = E[Y|V_{2it}, V_{bi}, P(T|V_{1it}, V_{ai})] \tag{B.3}$$

where $P(T|V_{1it}, V_{ai}$ is the conditional probability of treatment conditioned on the two indexes assumed in the treatment model.
To verify the validity of the above conditions, it could be proceeded in several steps:

1. Treatment model:
   a) Check $E[T|Z_{it}, Z_{is}] = E[T|Z_{it}, \bar{Z}]$
   b) Check $E[T|Z_{it}, \bar{Z}] = E[T|V_{1it}, V_{ai}] = P[T = 1|V_{1it}, V_{ai}]$
2. Outcome model:
   a) Check $E[Y|X_{it}, T_{it}, b_i] = E[Y|X_{it}, X_{is}, T_{it}, T_{is}]$
   b) Check $E[Y|X_{it}, X_{is}, T_{it}] = E[Y|X_{it}, \bar{X}, T_{it}]$
   c) Check $E[Y|X_{it}, \bar{X}, T_{it}] = E[Y|V_{1it}, V_{2it}, V_{ai}, V_{bi}]$
   d) Check $E[Y|V_{1it}, V_{2it}, V_{ai}, V_{bi}] = E[Y|V_{2it}, V_{bi}, P(T|V_{1it}, V_{ai})]$

B.1. Treatment model index assumption

For treatment model, the conditional mean can be rewritten as follow:

$$\begin{aligned} E[T|Z_{it}, Z_{is}] &= Pr[T = 1|Z_{it}, Z_{is}] \\ &= Pr[u_{it} < g(Z_{it}\beta, a_i)|Z_{it}, Z_{is}] \\ &= Pr[u_{it} < g(Z_{it}\beta, f(\bar{Z}\alpha) + \epsilon_i)|Z_{it}, Z_{is}] \\ &= f_{u_{it}|Z_{it}, Z_{is}}(g(Z_{it}\beta, f(\bar{Z}\alpha))) \end{aligned} \tag{B.4}$$

Where $f_{u_{it}|Z_{it}, Z_{is}}$ is the conditional probability distribution function for error term u_{it}. We can make additional assumptions on the distribution of error term to make the index assumption go through. Assume that all variables in Z_{it} are exogenous to error u_{it}, which is not an unreasonable assumptions to make because endogeneity in the treatment model is not the focus of this paper. Under this assumption, the distribution of u_{it} is independent

of Z_{it}, Z_{is} and $\bar{Z}$ as well.

$$\begin{aligned} E[T|Z_{it}, Z_{is}] &= Pr[T = 1|Z_{it}, Z_{is}] \\ &= f_{u_{it}|Z_{it},Z_{is}}(g(Z_{it}\beta, f(\bar{Z}\alpha))) \\ &= f_{u_{it}}(g(Z_{it}\beta, f(\bar{Z}\alpha))) \\ &= f_{u_{it}|Z_{it},\bar{Z}}(g(Z_{it}\beta, f(\bar{Z}\alpha))) \\ &= E[T|Z_{it}, \bar{Z}] \end{aligned} \tag{B.5}$$

Next step is to verify the index assumption is valid here to summarize the information given by Z_{it} and $\bar{Z}$, i.e. to check $E[T|Z_{it}, \bar{Z}] = E[T|Z_{it}\beta, \bar{Z}\alpha] = E[T|V_{1it}, V_{ai}]$, which follows by regular semiparametric index assumption. As result, the conditional probability of treatment can be represented by $E[T|V_{1it}, V_{ai}] = P[T = 1|V_{1it}, V_{ai}]$

B.2. Outcome model index assumption

For the outcome model, the assumptions to validate steps (a)-(c) listed above will be similar to those given in the treatment model. Here the key step is to check that the two-stage method by plugging in the conditional probability of treatment is valid. The conditional expectation of Y_{it} on all variable Z_{it} will be:

$$\begin{aligned} E(Y_{it}|Z_{it}, Z_{is}) &= E(Y_{it}|Z_{it},, Z_{is}, T_{it} = 1)Pr(T = 1|Z_{it}, Z_{is}) + E(Y_{it}|Z_{it}, Z_{is}, T_{it} = 0)Pr(T = 0|Z_{it}, Z_{is}) \\ &= E(Y_{it}|Z_{it}, Z_{is}, T_{it} = 1)Pr(T = 1|Z_{it}, Z_{is}) + E(Y_{it}|Z_{it}, Z_{is}, T_{it} = 0)[1 - Pr(T = 1|Z_{it}, Z_{is})] \\ &= Pr(T = 1|Z_{it}, Z_{is})[E(Y_{it}|Z_{it}, Z_{is}, T_{it} = 1) - E(Y_{it}|Z_{it}, Z_{is}, T_{it} = 0)] + E(Y_{it}|Z_{it}, Z_{is}, T_{it} = 0) \end{aligned} \tag{B.6}$$

where T_{it} follows the same model above. The key is whether the conditional means $E(Y_{it}|Z_{it}, Z_{is}, T_{it} = 1)$ and $E(Y_{it}|Z_{it}, Z_{is}, T_{it} = 0)$ are functions of $Pr(T|Z_{it}, Z_{is})$. Take $E(Y_{it}|Z_{it}, Z_{is}, T_{it} = 1)$ as

example,

$$
\begin{aligned}
&E(Y_{it}|Z_{it},Z_{is},T_{it}=1) \\
&\qquad = \int_{\eta} F(X_{it}, f(\bar{X})+\eta_i, T_{it}=1)dF(\eta|Z_{it},Z_{is},T=1)+ \\
&\qquad \int_{\epsilon} \epsilon_{it} f(\epsilon|Z_{it},Z_{is},T_{it}=1)d\epsilon \\
&\qquad = \int_{\eta} F(X_{it}, f(\bar{X})+\eta_i, T_{it}=1)dF(\eta)+ \\
&\qquad \int_{\epsilon} \epsilon_{it} f(\epsilon|Z_{it},Z_{is},T_{it}=1)d\epsilon
\end{aligned}
\tag{B.7}
$$

The first term is a function of $Pr(T|Z_{it},Z_{is})$ since T_{it} can be presented as P_{it}+residual. As for, $f(\epsilon|Z_{it},Z_{is},T_{it}=1)$, the conditional pdf for error term ϵ.

$$
\begin{aligned}
f(\epsilon|Z_{it},Z_{is},T_{it}=1) &= f\epsilon|Z_{it},\bar{Z},T_{it}=1 \text{[By imposing assumption2 on error } \epsilon] \\
&= f\epsilon|Z_{it},\bar{Z},g(Z_{1it}\beta,a_i)>u_{it} \\
&= \frac{\int_{-\infty}^{g(Z_{it}\beta,a_i)} f(\epsilon,u|Z_{it},\bar{Z})du}{\iint_{-\infty}^{g(Z_{it}\beta,a_i)} f(\epsilon,u|Z_{it},\bar{Z})d\epsilon du} \\
&= \frac{\int_{-\infty}^{g(Z_{it}\beta,a_i)} f(\epsilon,u|Z_{it},\bar{Z})du}{Pr[u_{it}<g(Z_{it}\beta,a_i)|Z_{it},\bar{Z}]} \\
&= \frac{\int_{-\infty}^{g(Z_{it}\beta,a_i)} f(\epsilon,u|Z_{it},\bar{Z})du}{Pr[T=1|Z_{it},\bar{Z}]}
\end{aligned}
\tag{B.8}
$$

Where $f(\epsilon,u|Z_{it})$ is the joint distribution of ϵ and u. Therefore, $E(Y_{it}|Z_{it},Z_{is},T_{it}=1)$ is also a function of $Pr[T=1|Z_{it},\bar{Z}]$. Similar argument holds for $E(Y_{it}|Z_{it},Z_{is},T_{it}=0)$. As result, the conditional mean of Y can be written as a function of $Pr[T=1|Z_{it},\bar{Z}]$.

$$
E(Y_{it}|Z_{it},Z_{is}) = H(P_{it}) = E(Y_{it}|P_{it}) \tag{B.9}
$$

Therefore, the two-stage method for index estimation should go through.

REFERENCES

[1] Joseph G Altonji and Rosa L Matzkin. Cross section and panel data estimators for nonseparable models with endogenous regressors. *Econometrica,* 73(4):1053–1102, 2005.

[2] Takeshi Amemiya. *Advanced econometrics*. Harvard university press, 1985.

[3] Gary Chamberlain. Panel data. *Handbook of Econometrics*, II:809–837, 1984.

[4] Mitali Das. Instrumental variables estimators of nonparametric models with discrete endogenous regressors. *Journal of Econometrics*, 124(2):335–361, 2005.

[5] Daniel J Henderson, Raymond J Carroll, and Qi Li. Nonparametric estimation and testing of fixed effects panel data models. *Journal of Econometrics*, 144(1):257–275, 2008.

[6] Stefan Hoderlein and Halbert White. Nonparametric identification in nonseparable panel data models with generalized fixed effects. *Journal of Econometrics*, 168(2):300–314, 2012.

[7] Hidehiko Ichimura. Semiparametric least squares (sls) and weighted sls estimation of single-index models. *Journal of Econometrics*, 58(1):71–120, 1993.

[8] Hidehiko Ichimura and Lung-Fei Lee. Semiparametric least squares estimation of multiple index models: single equation estimation. In *Nonparametric and semiparametric methods in econometrics and statistics: Proceedings of the Fifth International Symposium in Economic Theory and Econometrics. Cambridge*, pages 3–49, 1991.

[9] Roger Klein and Chan Shen. Bias corrections in testing and estimating semiparametric, single index models. *Econometric Theory*, 26(06):1683–1718, 2010.

[10] Roger Klein, Chan Shen, et al. Semiparametric instrumental variable estimation in an endogenous treatment model. 2015.

[11] Roger Klein and Francis Vella. Endogenous treatment as a double index semiparametric binary response model. Technical report, working paper, European University Institute, 2003.

[12] Roger W Klein and Richard H Spady. An efficient semiparametric estimator for binary response models. *Econometrica: Journal of the Econometric Society*, pages 387–421, 1993.

[13] Jürgen Maurer, Roger Klein, and Francis Vella. Subjective health assessments and active labor market participation of older men: evidence from a semiparametric binary choice model with nonadditive correlated individual-specific effects. *Review of Economics and Statistics*, 93(3):764–774, 2011.

[14] Yair Mundlak. On the pooling of time series and cross section data. *Econometrica: journal of the Econometric Society*, pages 69–85, 1978.

[15] Anastasia Semykina and Jeffrey M Wooldridge. Estimating panel data models in the presence of endogeneity and selection. *Journal of Econometrics*, 157(2):375–380, 2010.

[16] Alexandra Soberón and Juan Manuel Rodriguez-Poo. Direct semiparametric estimation of fixed effects panel data varying coefficient models. *Available at SSRN 2113037*, 2013.