ESSAYS ON HEALTH OUTCOMES, ECONOMIC WELLBEING AND MISMEASURED DISCRETE HEALTH VARIABLES

by

NING LI

A dissertation submitted to the

School of Graduate Studies

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Economics

Written under the direction of

Roger Klein

Carolyn Moehling

And approved by

____

____

____

____

New Brunswick, New Jersey

October, 2019

ABSTRACT OF THE DISSERTATION

Essays on Health Outcomes, Economic Wellbeing and Mismeasured Discrete Health Variables

By NING LI

Dissertation Director:

Roger Klein

Carolyn Moehling

This thesis studies the labor supply effects of health shocks for aging Americans. To address the mismeasured binary health variables used in the labor supply equation, this thesis develops a new theoretical approach to the non-classical measurement error. In addition to Chapter 1 which provides an overview of the thesis, there are three primary chapters explaining the theoretical development and the empirical studies.

Chapter 2 theoretically addresses the estimation bias due to the misclassification of a binary regressor in treatment models. Different from the assumption of a valid instrument in the literature, this paper allows the potential instrument to be correlated with the measurement error. In such a general setting, I propose an estimator relying only on those extreme observations that are free of misclassification. This proposed estimator is proven to have large sample properties and much better performance than OLS and traditional IV estimates in finite samples.

Chapter 3 uses the method proposed in Chapter 2 to handle the binary, misclassified health variable in studying the labor supply effects of health shocks. Extracting information on true health from objective health measures, for example functional limitations

and doctors' diagnoses, this new method relies on such information to dynamically select observations that are free of misclassification. Using the 2012 wave of the Health and Retirement Study (HRS), this paper primarily examines the labor supply effects of health shocks for men and women aged 45-61. This study finds that individuals in middle age will greatly reduce their labor supply when experiencing health shocks and that the estimation results are very sensitive to the health measures used.

In Chapter 4, I examine how an individual's labor supply responds in the short- and long-run to a negative shock to her spouse's health. I propose an optimal instrument strategy with fixed effects to study labor supply effects of spousal health shocks. Analysis of the 1996-2012 data from the Health and Retirement Study (HRS) indicates that in the short run, both husbands and wives change their labor supply very little when their spouses become ill. However, in the long run, husbands adjust their labor supply in response to their wives' health problems. As the duration of wives' health problems increases by two years, husbands work 165 fewer hours per year.

Acknowledgements

My sincere gratitude goes to my main advisors, Professors Roger Klein and Carolyn Moehling. I started my research with Professor Moehling when I was a fresh Ph.D. student with an interest in empirical studies. Since the first day of becoming her student, she has instructed me to skillfully investigate facts behind data, improved my scientific writing through her commenting line by line, embraced all my ideas and at the same time paid close attention to how I developed them in a rigorous way. She encouraged and accompanied me through difficult times. She is so understanding and supportive and I am very lucky to have her as my advisor. Professor Klein has also given enormous guidance to me in terms of not only how to do good research but also how to enjoy the process of doing it. Whenever I was lost in the scientific maze, he always pointed out the hidden door. Working with him was very enjoyable, since the conversation with him made me think of research from a new perspective and motivated me to go through so many otherwise hopeless days. I was also deeply impressed by his attitude towards life: live life to the fullest. Observing his practice of this attitude in every single day inspired me to bravely pursue dreams that I would otherwise have given up because of fear of failure. With his help, I want to become an economist with great curiosity to endlessly search for and engage in important research and most importantly, he makes me realize that I will.

I would also like to thank Professors John Landon-Lane and Chan Shen, who gave me many helpful comments during the process I finished this thesis. Their suggestions enlighten me on future research, providing directions to which I plan to extend my current work.

I am also grateful to my friends, Sung Lee, Weiqi Xiong, Yuanting Wu, Yilin Wu, Mark Avery and Junyi Chen. We shared happiness and struggles during the Ph.D. study. The friendship with them is precious. I would like to thank my job market comrades, Yixiao Jiang, Geoff Clarke and Izumi Ryuichiro. We shared information and encouraged each other when we went on the market. It was great relief for me to have

them together with me to face the competitive job market.

Last but not least, I would like to express my deepest gratitude to my husband, Hanxiong Wang. With him, life is beautiful and I am a much better person than I would be otherwise. My final gratitude goes to my parents and brother for their endless support and love.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Building on the seminal work of Grossman (1972), health has been regarded as part of human capital that plays an important role in determining an individual's labor market behavior. As people age, the risk of health problems increases. To assess the economic impact of declining health status with aging, this thesis studies the labor supply effects of health shocks for Americans age 45 years and older. When studying the effect of health shocks on economic outcomes, the measurement error of binary health variables is inherently non-classical, making the traditional IV strategy fail to obtain a consistent estimate. This issue has attracted econometricians' attention for decades and has remained an open question. Therefore, this thesis develops a new theoretical approach to the measurement error of binary health variables in labor supply equations.

In Chapter 2, I address the estimation bias due to the misclassification of a binary regressor in treatment models. As a true "0" can only be misclassified, if at all, to a "1", and vice versa, the true binary variable and the measurement error will be correlated. Accordingly, misclassification of a binary regressor is distinct from the classical measurement error of a continuous independent variable. As a result, conventional OLS and IV methods fail to obtain a consistent treatment effect. Failing to address misclassification leads to biased estimates, and hence to potentially misleading recommendations. An extensive body of literature has studied misclassification by assuming a constant misclassification rate, which is a quite strong assumption, or by at most allowing the probability of misclassification to depend on other covariates. Identification relies on the existence of an instrument that is uncorrelated with the measurement error. However, such a valid instrument will be difficult to find and may not exist, because the measurement error will generally be related to the true binary unobservable. No previous research examines the case where the potential instrument is allowed to be correlated with the measurement error. This paper relaxes the zero correlation between the instruments and the measurement error by assuming that potential instruments indicate information not only on the true binary regressor but also on the misclassification probabilities. For example, objective health measures (functional limitations or doctors' diagnoses) as potential instruments are correlated with the general self-rated health status that is of empirical interest and subject to measurement error. At the same time, it is reasonable to believe that individuals with more functional limitations and doctors' diagnoses are more likely to be sick and less likely to misreport their health status as healthy. The same logic applies to those with fewer functional limitations and doctors' diagnoses. Based on this argument, I assume that misclassification probabilities depend on an index as a linear combination of individuals' characteristics and objective health measures and that there is accurate reporting (i.e. no misclassification) for extreme values of the index. To implement this intuition, I define a data dependent high probability set containing extreme index values. This set is determined so as to optimize a bias/variance tradeoff. Based on some assumptions on tail conditions, I prove identification, consistency and asymptotic normality. I run Monte Carlo simulations to test the validity of the estimator in finite sample studies. The results suggest the proposed estimator significantly improves the estimation compared to OLS and IV methods, in terms of having a much smaller mean square error.

Chapter 3 uses the method proposed in Chapter 2 to handle the binary, misclassified health variable in studying the labor supply effects of health shocks. Extracting information on true health from objective health measures, for example functional limitations and doctors' diagnoses, this new method relies on such information to dynamically select observations that are free of misclassification. Based on these "correct" observations, an IV estimator is consistent. Using the 2012 wave of the Health and Retirement Study (HRS), this paper primarily examines the labor supply effects of health shocks for men and women aged 45-61. When examining the measure of self-reported health status, the results suggest that men (women) will reduce labor supply by 2,299 (1,929) hours per year when they rate their health as "Fair" or "Poor" and that OLS and traditional IV estimators suffer from considerable attenuation biases compared to the proposed estimator. When examining the measure of work-limiting health, the results suggest that the traditional IV strategy and the proposed new technique produce similar estimates, while the OLS estimate biases towards zero. As a comparison, I also examine the sample aged 62-70 when the elderly seriously consider retirement. The estimation results indicate significant reductions in hours of work when in poor health but generally smaller than those for 45-61 year olds.

Besides the reduced work hours due to one's own health shocks, an individual may also respond in labor supply to her spouse's health problems. To study the spouse's labor market behavior in the face of the other's health declines is equally important in assessing the economic circumstances of aging households. In Chapter 4, I examine how an individual's labor supply responds in the short- and long-run to a negative shock to her spouse's health. The predicted effect of health shocks on spouse's labor supply is theoretically ambiguous. On one hand, poor health often reduces productivity in household production and requires time spent in care giving by the spouse. In this case, the spouse remains the primary home producer and caregiver, which pulls her away from the labor market. On the other hand, it has been well established that health declines reduce own labor supply, resulting in an income loss. To compensate for this loss, the spouse may need to earn more money and increase her labor supply. I focus not only on the short-run effect soon after a negative health shock to a spouse but also on the long-run effect of how a spouse adjusts her labor supply over time in response to her partner's health shock. The long-run effect may be larger than the short-run effect, especially for the elderly who are more likely to suffer from chronic conditions and so become more cautious than the young when making decisions about labor supply. In this paper, I propose an optimal instrument strategy with fixed effects to study labor supply effects of spousal health shocks. Analysis of the 1996-2012 data from the Health and Retirement Study (HRS) indicates that in the short run, both husbands and wives change their labor supply very little when their spouses become ill. However, in the long run, husbands adjust their labor supply in response to their wives' health problems. As the duration of wives' health problems increases by two years, husbands work 165 fewer hours per year. This research reveals that the impact on time allocation is greater for wives' poor health than husbands'. A wife's health decline will gradually pull her husband away from the market. Such households, therefore, are at greater risk for financial hardship.

Chapter 2

Identifying and Estimating Regression Models with a Misclassified, Binary Regressor

2.1 Introduction

Many empirical studies examine the causal effect of a binary treatment on economic outcomes; some important examples are the wage effects of college education or union status and the labor supply effects of a job training program or a health problem. The measurement error in such binary treatment variables creates an econometric issue distinct from the classical measurement error. In particular, the measurement error in a binary treatment variable is by its nature negatively correlated with the true part of the mismeasured variable (Aigner, 1973). As a result, it is difficult to find valid instruments to resolve the discrete measurement error. Conventionally, this type of measurement error is referred to as the misclassification.

This paper considers the identification and estimation of regression models where a binary regressor is subject to misclassification. In contrast to other papers on misclassification, this paper uses a more general and applicable assumption about the misclassification process. Most previous studies gained partial or point identification based on the assumption that the probabilities of misclassification were constant across the population or, at most, dependent on the covariates from the outcome equation. As discussed below, their point identification results assumed not only that misclassification probabilities depended on outcome covariates, but also that a valid instrument existed. In view of Aigner's results, such an instrument will be difficult to find and may not exist. The selection of a valid instrument should be very sophisticated such that the instrument is correlated with the true unobservable but uncorrelated with the measurement error, given the inherent correlation of the latter two terms. No previous work considered the case where the probabilities of misclassification are functions of not only the covariates from the outcome equation but also the potential instruments. This assumption has a desirable feature that most literature lacked: the potential instruments can be correlated with both the true part of the mismeasured regressor and the measurement error. Meanwhile, this setting has a more practical application than that with the constant or covariate-dependent probabilities of misclassification. For example, in the study of how an individual's health problem influences his or her labor supply, the self-reported health status is subject to misclassification and some objective health measures (like functional limitations or doctors' diagnoses) can serve as excluded variables for the misreported subjective measure. Moreover, objective health measures can provide information on health misclassification; both individuals who have relatively many functional limitations or doctors' diagnoses and those who have few functional limitations or doctors' diagnoses are unlikely to report the opposite health status. Thus, individuals for which the objective health measures take on sufficiently large or sufficiently small values have much lower possibility of misreporting and are less subject to misclassification. Including the objective health measures in an index, the true health, the misreported health and the measurement error are assumed to depend on this index. While the misreported health is observed, its specific model is unknown. Therefore, in a semiparametric model of the misreported health we estimate the index of health status. As this index takes on very large or very small values, we assume that individuals have much lower possibility of misreporting. As the index approaches its lower and upper support points, we assume that the probability of correct classification tends to one. To implement this intuition, we define a data dependent high probability set containing extreme index values. An estimator is proposed based on the observations from this high probability set. The set depends on the sample size with the property that the index values in this set become more extreme as the sample size increases. If this set is not sufficiently extreme, the bias in the estimator will be substantial as the set includes many "incorrect" observations. If this set becomes "too extreme", the bias becomes negligible but the variance of the estimator will be potentially significant due to having too few observations included. This high probability set is determined so as to optimize the bias-variance tradeoff.

This paper establishes the point identification without any prior knowledge on misclassification probabilities or higher order moment conditions of the error term. Moreover, it is not necessary to have equal misclassification probabilities, a condition often assumed in the literature. To study the large sample distribution, it is assumed that the conditional probabilities of the true regressor and its observed surrogate has the same order in the tails of the index. As in other pioneering literature (Andrews and Schafgans, 1998; Klein, Shen and Vella, 2015), I assume the distribution of the index, a linear combination of all exogenous variables, has thicker tails than that of the error term in the threshold-crossing model of the true binary regressor. This assumption is required to show that the bias and the variance of the estimator converge to zero sufficiently fast to obtain desirable large sample properties for the estimator.

The paper is organized as follows: Section 2.2 reviews related literature; Section 2.3 describes the model, explains the motivation underlying the approach and introduces the estimator; Section 2.4 discusses definitions and necessary assumptions for the establishment of the estimator's large sample distribution; Secton 2.5 presents theorems and the outline of their proof strategy; Section 2.6 provides evidence that the estimator performs very well in Monte Carlo simulations; and Section 2.7 draws conclusions. The Appendix contains all needed lemmas and their proofs.

2.2 Literature Review

With an array of real world survey data suffering from the issue of measurement error, models with mismeasured explanatory variables have been studied extensively. The simplest version of measurement error, the classical measurement error in linear regression models, assumes that the true unobservable, its mismeasured surrogate and the measurement error are continuous. Further, the measurement error is assumed to be uncorrelated with the true unobservable. The existence of such measurement error leads to an attenuation bias when estimating the impact of the mismeasured explanatory variable. A valid instrument suffices to address the attenuation bias as long as it satisfies exclusion restrictions and is uncorrelated with the measurement error.

The assumption that the true unobservable and the measurement error are uncorrelated has been relaxed. Chen, Hong, and Tamer (2005) assumed the measurement error was correlated with the true unobservable, which is very important in the case of misclassification. They used an auxiliary dataset to convey information on the relationship between the true variable and its mismeasured counterpart in the primary dataset, assuming the primary and auxiliary datasets had the same conditional densities of the true variable given its surrogate.

The correlation between the true explanatory variable and the measurement error arises automatically when the mismeasured regressor is binary. Conventionally, this type of measurement error is referred to as misclassification. The traditional assumption on the misclassification process is that the misclassification rate is constant; all the observations in a sample have the same probability of reporting a 1 when the true treatment is 0, and the same probability of reporting a 0 with the true treatment of 1. Based on this assumption, some papers rely on prior restrictions on the probabilities of misclassification or higher order (conditional) moment conditions of the regression error to achieve the partial or point identification. Aigner (1973) calculated the covariance of the measurement error and the true variable, first recognizing the negative correlation between these two random variables and thus deriving the least squares bias. To address this problem, he applied a modified least squares technique by incorporating prior knowledge of the misclassification process.

Bollinger (1996) identified bounds for the parameters in regression models with a mismeasured binary regressor by using the first and second moments of the observables. He also examined the bounds with different, stronger assumptions on the probabilities of misclassification. Based on this paper, van Hasselt and Bollinger (2012) continued to analyze the bounds for the regression coefficient of the misclassified regressor. They found that assumptions on probabilities of misclassification or homoscedasticity in regression error could shrink the identified set and therefore tighten the upper bound compared to Bollinger (1996). Moreover, if the assumptions on equal and constant misclassification (a 0 was misclassified to be a 1 with the same probability as a 1 was misclassified to be a 0) and homoscedasticity held simultaneously, the coefficient for the misclassified regressor was identified, rather than only partially identified to its lower and upper bounds. While such assumptions underlying the point identification are too strong, as admitted by the authors themselves, their work helps reveal the significant role the misclassification probabilities play in uncovering the potential set of parameters.

Chen et al. (2008a, 2008b) resolved the identification problem in nonparametric regression models with a misclassified binary regressor without additional auxiliary information like an instrumental variable. They assumed the zero conditional third moment of the regression error and recovered in closed form the conditional density of the dependent variable and the misclassification process for the nonparametric regression model (2008a). To generalize the results, they relaxed the assumption on the zero conditional third moment (2008b). Instead, the true regressor was assumed to provide no information on the second and third moments of the regression error. In other words, the expectations of the squared and cubic error conditioned on the true regressor were constant across all observations. The similar closed-form identification results were established immediately, potentially leading to a consistent estimator.

With the assumption on the constant misclassification rate, another strand of work contributing to the literature uses instrumental variables to identify the causal effect of a binary, misclassified regressor. Black et al. (2000) and Kane et al. (1999) both recognized the identification power when there were two mismeasured reports available as the surrogates for the true unobservable. Essentially, each of these two measures was able to serve for one another as an instrument (DiTraglia and García-Jimeno, 2017). These two important papers also demonstrated mathematically that any IV estimator provided an upper bound on the true treatment effect whereas the OLS estimator provided a lower bound as in Aigner (1973). Furthermore, they recovered a point estimate using a series of moment conditions, including the joint distribution of these two surrogates and the expectations of the dependent variable conditioned on the two measures. Such moments together identified coefficients, the probability distribution of the unobserved binary regressor and misclassification process of the two measures. Frazis and Loewenstein (2003) replaced one of the mismeasured variables in Black et al. (2000) and Kane et al. (1999) with an instrument for the other one. They used a set of moments to accommodate the instrument, which were equivalent to those used in Black et al. (2000) and Kane et al. (1999), and proposed a consistent GMM estimator. They also modified the technique in Hausman et al. (1998) to bound the probabilities of misclassification and incorporated these parameter restrictions into their GMM estimates. At the same time, the case where the misclassified regressor was endogenous was analyzed. In this case, the GMM estimation would be underidentified, although coefficients could be bounded.

All the papers described above, whether identified with instrumental variables or not, required that the probabilities of misclassification were constant across observations. This is a quite strong assumption; it implies that individuals with differing demographics, like education, age, income, etc., would misreport with identical probabilities. Mahajan (2006) relaxed this assumption in his nonlinear regression model, making the probabilities of misclassification depend on other covariates. Assuming the existence of an instrument that was uncorrelated with the measurement error, he used a series of conditional moments given the values of covariates and the instrument to identify the marginal effect of the true unobserved regressor and its misclassification rates. While Mahajan (2006) believed that these identification results could be readily applied to the case where the regressor of interest was not only misclassified but also endogenous in the outcome equation, DiTraglia and García-Jimeno (2017) pointed out that Mahajan's direct application was incorrect, since with the addition of endogeneity the instrument no longer indicated information on the endogeneous, mismeasured regressor. Then the identification proposed by Mahajan (2006) failed in the context of endogeneity. Also, DiTraglia and García-Jimeno (2017) showed that the instrument's second and third moment independence of the regression error point identified the treatment effect and the misclassification rates.

While the assumption about the constant probabilities of misclassification has been relaxed by these studies, no prior work considered the (conditional) dependence of instrumental variables and the measurement error. The assumption in previous studies on the lack of correlation between instrumental variables and the measurement error makes it difficult to find valid instrumental variables in practice, given the inherent correlation between the true unobserable and the measurement error. The more general assumption that allows the dependence of potential instrumental variables and the measurement error meets the theoretical and empirical necessity but poses technical challenges, as recognized in Mahajan (2006).

This paper contributes to the literature by studying a linear regression model and assuming that the misclassification process is dependent not only on covariates but also on potential instruments. Different from the previous misclassification literature, this paper employs the technique of identification through the extreme high probability set to estimate the causal effect of a binary treatment regressor when it is misclassified. Identification and estimation through the extreme high probability set is enlightened by the important work of Heckman (1990), Andrews and Schafgans (1998), and Klein, Shen and Vella (2015) in sample selection models. Heckman (1990) proposed the estimator for the intercept only using those observations for which the selection probabilities were close to 1, since as the selection probability approached 1, the conditional expectation of regression error for the selected was close to 0. Andrews and Schafgans (1998) used a smooth weighting function to mimic the process of choosing extreme high probability set in Heckman (1990). With assumptions on the known set, they obtained the large sample distribution for the estimator. Klein, Shen and Vella (2015) considered a sample selection model where the dependent variable was binary. Facing an unknown high probability set, they estimated the set and proved the large sample property for the marginal effect estimator that was defined based on the estimated high probability set, making the technique of identification at infinity technically and empirically feasible.

Besides relaxing the assumption of no correlation between the instrumental variable and the measurement error, this paper obtains the point identification without relying on any prior knowledge on probabilities of misclassification or very strong higher order moment conditions of the regression error, assumptions critical for identification in previous studies. Furthermore, the proposed technique does not require the assumption of having equal probabilities of misclassification.

2.3 Methodology

2.3.1 Model

In this paper I consider a linear regression model in which a binary regressor, Y_2^*, is subject to misclassification:

$$Y_{1i} = \alpha + X_{1i}\gamma + Y_{2i}^*\beta + \epsilon_i, \quad E(\epsilon_i|X_{1i}, Y_{2i}^*) = 0 \tag{2.1}$$

where $\{Y_{1i}, Y_{2i}^*, X_{1i}\}$ is for observation i in a random sample that is independent and identically distributed over i. Y_{1i}, Y_{2i}^* are scalars, and X_{1i} is a $1 \times k$ random vector, including all other exogenous covariates. Dropping the observational subscript, Y_2^* is a binary variable and its effect on the outcome Y_1 is of interest.

The misclassification issue in (2.1) arises because Y_2^* is rarely observed in practice. Instead, econometricians observe its surrogate, another binary variable $Y_2 = Y_2^* + \eta$, where η is the measurement error. In particular, mismeasurement only happens when the true variable is 0 but its surrogate is 1 ($\eta = 1$) or when the true variable is 1 but it is misreported to be 0 ($\eta = -1$). This type of measurement error is typically referred to as the misclassification. In the context of misclassification, the measurement error is negatively correlated with the true value of the regressor, given that when Y_2^* equals 0, the error η must be 0 or 1, and when Y_2^* equals 1, the error η must be 0 or -1. See Aigner (1973) for a statistical calculation of the covariance between the true variable and the measurement error. With observables, econometricians run the regression (2.1) as

$$\begin{aligned} Y_1 &= \alpha + X_1\gamma + Y_2\beta + (Y_2^* - Y_2)\beta + \epsilon \\ &= \alpha + X_1\gamma + Y_2\beta + (\epsilon - \eta\beta) \end{aligned} \tag{2.2}$$

In the classical measurement error model, the unobserved variable Y_2^* and its observed surrogate Y_2 are both continuous variables. Even if the measurement error η is independent of the unobserved true variable, it is well-established that the OLS estimate causes an attenuation bias. An IV estimator is consistent as long as the instrument is correlated with the true variable and uncorrelated with the regression error and the measurement error. However, when the mismeasured variable is binary, its true values will be negatively correlated with the measurement error, making the IV estimators inconsistent even if the instruments are correlated with the true variable and uncorrelated with the regression error. More explicitly, let X_2 be the potential instrument such that $cov(X_2, Y_2^*) \neq 0$ and $cov(X_2, \epsilon) = 0$. As $cov(Y_2^*, \eta) \neq 0$, it is not guaranteed that $cov(X_2, \eta) = 0$ even if X_2 is correlated with η only through the true variable Y_2^* (Frazis and Loewenstein, 2003). Therefore, the inherent correlation between the true part of the mismeasured variable and the measurement error makes IV estimators inconsistent.

While the instrumental variable strategy fails to address misclassification, the potential instrument does provide information on the unobserved true variable:

$$Y_2^* = I\{X_1\pi_1 + X_2\pi_2 > \mu\} \tag{2.3}$$

where X_1 is the covariate vector from the outcome equation, X_2 includes all the exclusion variables (or termed potential instruments),[1] and $I\{.\}$ is an indicator function. In the literature on misclassification, it is commonly assumed that the probabilities of misclassification are fixed across the population. Typically,

$$Pr(Y_2 = 1|Y_2^* = 0, X) = Pr(Y_2 = 1|Y_2^* = 0) = \alpha_0 \tag{2.4}$$

$$Pr(Y_2 = 0|Y_2^* = 1, X) = Pr(Y_2 = 0|Y_2^* = 1) = \alpha_1 \tag{2.5}$$

This is a very strong assumption; it implies that individuals with differing demographics, like education, age, income, etc., would misreport with the identical probabilities. Mahajan (2006) relaxed this by assuming that the probabilities of misclassification are functions of covariates from the outcome equation. As exclusion variables are potentially correlated with the misclassification error by its nature, I extend the assumption further to allow exclusion variables to indicate information not only on the true variable but also on the misclassification process. The probabilities of misclassification are

[1] Exclusion variables are akin to instrumental variables in terms of their independence of the outcome model conditional on Y_2^*. They are not directly named as instruments because they are allowed to be correlated with the measurement error and also because proposed estimator in this paper uses a function of exclusion variables rather than using these variables as they are, which will be explained later in more detail. In addition, any number of exclusion variables are sufficient. To include only one exclusion variable here makes it convenient for elaborating the estimation of the index in the general way.

functions of covariates as well as exclusion variables:

$$PL(X) = Pr(Y_2 = 1|Y_2^* = 0, X) \tag{2.6}$$

$$PR(X) = Pr(Y_2 = 0|Y_2^* = 1, X) \tag{2.7}$$

where X includes all the exogenous variables, X_1 and X_2.

To motivate this assumption, let Y_1 denote an individaul's labor supply and Y_2^* denote whether or not an individual has a health problem. By the model (2.1), we would like to examine how one's health declines affect his labor supply. In most empirical research, health status is self-reported, resulting in a subjective measure, Y_2 with errors. Some relatively objective measures are used to instrument for the subjective one, like functional limitations and doctors' diagnoses. Individuals who have more functional limitations or doctors' diagnoses are believed to have worse "true" health, and those who have fewer functional limitations or doctors' diagnoses usually have better "true" health. At the same time, it is reasonable to argue that those who have more functional limitations and doctors' diagnoses are much less likely to report themselves as healthy. For example, an individual diagnosed with diabetes, high blood pressure or cancer is unlikely to report his health as excellent. Individuals who have few or even no functional limitations or doctors' diagnoses would be unlikely to report their health status as unhealthy. In this way, the exclusion variables indicate information not only on the true unobservable but also on the misclassification process.

The relationship between the true variable's and its surrogate's distributions can be derived:

$$\begin{aligned} Pr(Y_2 = 1|X) &= Pr(Y_2 = 1|Y_2^* = 1, X)Pr(Y_2^* = 1|X) \\ &\quad + Pr(Y_2 = 1|Y_2^* = 0, X)Pr(Y_2^* = 0|X) \\ &= (1 - PR(X))P_2^*(X) + PL(X)(1 - P_2^*(X)) \end{aligned} \tag{2.8}$$

where $P_2^*(X) = Pr(Y_2^* = 1|X)$. The distribution of Y_2 depends on X and the equation (2.8) implies the relation between Y_2's and Y_2^*'s distributions. Define an index V as a linear combination of all the exogenous variables in X:

$$V = X_{11} + X_{12}\psi_{20} + X_{13}\psi_{30} + ... + X_{1k}\psi_{k0} + X_2\theta_0 \tag{2.9}$$

where $\{X_{11}, ..., X_{1k}\}$ are the k variables in the vector, X_1. I assume the model satisfies the following index restrictions:[2]

$$Pr(Y_2^* = 1|X) = Pr(Y_2^* = 1|V) = P_2^*(V) \tag{2.10}$$

$$Pr(Y_2 = 1|X) = Pr(Y_2 = 1|V) = P_2(V) \tag{2.11}$$

$$Pr(Y_2 = 1|Y_2^* = 0, X) = Pr(Y_2 = 1|Y_2^* = 0, V) = PL(V) \tag{2.12}$$

$$Pr(Y_2 = 0|Y_2^* = 1, X) = Pr(Y_2 = 0|Y_2^* = 1, V) = PR(V) \tag{2.13}$$

In the discrete choice model where the binary dependent variable was subject to misclassification, Hausman, Abrevaya and Scott-Morton (1998) showed that index parameters could be recovered consistently under the assumption that the probabilities of misclassification were constant. Then by equation (2.8), it is easily derived that the P_2 and P_2^* depend on the same index. The current paper assumes that the probabilities of misclassification are not constant and depend on the same index as P_2^* does. If P_2 depends on a perceived index which is some function of the actual index, it is natural to assume that P_2 and P_2^* still depend on the same index. Under these index restrictions, the relation between Y_2's and Y_2^*'s distributions becomes:

$$P_2(V) = (1 - PR(V))P_2^*(V) + PL(V)(1 - P_2^*(V)) \tag{2.14}$$

2.3.2 Model Simplification

With the index V, the model in (2.3) can be written:

$$Y_2^* = I\{Vb > \mu\} \tag{2.15}$$

$$Vb = X_1\pi_1 + X_2\pi_2 = (X_{11} + X_{12}\psi_{20} + X_{13}\psi_{30} + ... + X_{1k}\psi_{k0} + X_2\theta_0)b \tag{2.16}$$

Under the index restrictions in (2.10)-(2.13), the distribution of Y_2 depends on the same index, V, as that of Y_2^*. The index parameters can be recovered by estimating a binary response model of Y_2 in the process of maximizing the following guasi log-likelihood

[2] In a semiparametric model, it is the parameters of this normalized index that are identified, which suffices to obtain probabilities and marginal effects.

function:

$$L(\{\psi_m\}_{m=2}^k, \theta) \equiv \sum_i^N \{Y_{2i} \ln[\hat{P}_{2i}(V_i(\{\psi_m\}_{m=2}^k, \theta))]$$
$$+ (1 - Y_{2i}) \ln[1 - \hat{P}_{2i}(V_i(\{\psi_m\}_{m=2}^k, \theta))]\} \quad (2.17)$$

The parameters $\{\psi_m\}_{m=2}^k$ and θ are identified semiparametrically in a single-index model. The estimates of index parameters have been well established in semiparametric models, for example Klein and Spady (1993) and Ichimura (1993). The semiparametric model here is important because while the unobservable Y_2^* has a single-index model in (2.15), the specific model for the observable Y_2 is unknown.

With the index V recovered, I can simplify the model in (2.1) using the approach proposed by Robinson (1988) for partially linear models. To be specific, the outcome equation in (2.1) can be written as

$$Y_1 = \alpha + X_1\gamma + G(V) + newerror \quad (2.18)$$

where $G(V) = \mathbf{E}(Y_2^*|V)\beta = P_2^*(V)\beta$ and $newerror = (Y_2^* - G(V))\beta + \epsilon$. As Y_2^* is unobserved, the function $G(V)$ is unknown. Taking the expectation of every term in (2.18) conditional on the index V, then

$$\mathbf{E}(Y_1|V) = \alpha + \mathbf{E}(X_1|V)\gamma + G(V) \quad (2.19)$$

Making the difference between equations (2.18) and (2.19) on both sides, we can obtain a differenced model as

$$Y_1 - \mathbf{E}(Y_1) = (X_1 - \mathbf{E}(X_1|V))\gamma + newerror \quad (2.20)$$

Robinson (1988) showed that the parameter γ can be consistently estimated at the $\sqrt{N}$ rate. Subtracting the estimator of $X_1\gamma$ from both sides and still using Y_1 to denote the differenced outcome on the left hand side, the model in (2.1) can be simplified into the model as follows:

$$Y_{1i} = \alpha + Y_{2i}^*\beta + \epsilon_i, \quad E(\epsilon_i|Y_{2i}^*) = 0 \quad (2.21)$$

Without loss of generality, the rest of this paper focuses on the simplified model in (2.21).

It is desirable to use Robinson's technique to separate the estimation of the health impact from that of other covariates. As Bound (1991) explained, when other economic covariates were correlated with health in the labor supply equation, their coefficients would be also biased due to the biased estimate of the effect of health (resulted from measurement error). Also, such biased coefficients on other economic characteristics remained even if the biased impact of health could be addressed in some way. Therefore, to first estimate the coefficients on other covariates using Robinson's technique avoids the influence passed on by the mismeasured health status in estimation, making it possible to compare the relative importance of health and other economic characteristics in labor supply decisions.

2.3.3 Motivation for High Probability Set

The essential issue in misclassification is the failure to observe Y_2^* and the discrepancy between it and its surrogate Y_2. If in some regions Y_2 behaves in the same way as Y_2^* with a high probability, their discrepancy will disappear and the distribution of Y_2^* can be recovered in such regions. Equation (2.14) provides this opportunity. If $P_2^*(V)$ is similar to $P_2(V)$ in these regions, it is credible that $P_2(V)$ captures the variation of $P_2^*(V)$ there. Here, to be "similar" is in terms of having the same order in the extreme regions; this will be explained in more detail later in the paper.

Again, take the effect of health status on labor supply as an example. The index now denotes a linear function of functional limitations, doctors' diagnoses and other variables that influence health status. As individuals who have more functional limitations and doctors' diagnoses (a higher V) are more likely to have poor health and less likely to misreport their health status, $P_2(V)$ (the probability of reporting poor health given the index) will be similar to $P_2^*(V)$ (the probability of actually having poor health given the index) and both converge to 1 as the index V becomes sufficiently large. On the other extreme, as individuals who have few or no functional limitations or doctors' diagnoses (a lower V) are more likely to have good health and also less likely to misreport their health status, $P_2(V)$ will also be similar to $P_2^*(V)$ and both converge to 0 as the index V becomes sufficiently small. Thus, the index V indicates the regions where $P_2(V)$ looks like $P_2^*(V)$. I will refer to such regions as the high probability set, enlightened by the identification approach of Heckman (1990), Andrews and Schafgans (1998), and Klein, Shen and Vella (2015), which all estimated parameters of interest in sample selection models by using only those observations within their respective high probability sets.

In this paper the high probability set occurs when the index becomes sufficiently large or sufficiently small. The probability of $Y_2 = 1$ conditional on the index V, $P_2(V)$, is estimated semiparametrically. In this context, it is important to have a flexible model that allows Y_2's distribution to differ from Y_2^*'s distribution when not in the high probability set. In the high probability set, the probability of $Y_2^* = 1$ conditional on the index V, $P_2^*(V)$, can be recovered up to the order because the conditional probability $P_2(V)$ has the same order as the conditional probability $P_2^*(V)$ in the set. With N as the sample size, the high probability set is defined as

$$\{v : P_2(v) < N^{-a} \ or \ P_2(v) > 1 - N^{-a}\}, \ 0 < a < 1 \tag{2.22}$$

Observations are selected for which index values are located in the high probability set. As the value of index becomes sufficiently large, $P_2(v)$ converges to 1 and the cutoff point $(1 - N^{-a})$ moves toward 1 to select those index values, as the sample size N increases. The analogous rule applies to the other tail for sufficiently small index values. The high probability set parameter a determines which observations in the sample will be selected. To study it, I will introduce in later sections an infeasible estimator that is close to the feasible estimator. The optimal value of a will be determined to balance the order of the infeasible estimator's (squared) bias and variance. Based on it, I will define its sample counterpart as the optimal high probability set parameter for the feasible estimator. In the study of large sample properties for the feasible estimator, this sample counterpart will be proved to converge in probability to the optimal value of a.

2.3.4 The Estimator

With $\hat{P}_{2i}$ as an instrument for Y_{2i}, the estimator I propose is the IV estimator on a high probability set

$$\hat{\beta} = \frac{\sum_{i=1}^{N}(\hat{P}_{2i} - \bar{P}_2)Y_{1i}\hat{S}_i}{\sum_{i=1}^{N}(\hat{P}_{2i} - \bar{P}_2)Y_{2i}\hat{S}_i} \tag{2.23}$$

where $\hat{P}_{2i}$ and $\hat{S}_i$ abbreviate $\hat{P}_{2i}(\hat{V}_i)$ and $S_i(\hat{V}_i, \hat{a}_i, \hat{P}_{2i})$, respectively, and $\bar{P}_2$ is the weighted average of $\hat{P}_{2i}$, $\frac{\sum_{i=1}^{N}\hat{P}_{2i}\hat{S}_i}{\sum_{i=1}^{N}\hat{S}_i}$. The estimated conditional probability $\hat{P}_{2i}$ is given by Definition (D2) in next section, while the relevant definitions regarding $\hat{S}_i$ are from Definition (D3) and (D4). The S-function gives a high weight up to 1 to the observation i once the observation is "trapped" by the high probability set, while giving low weights of 0 to those observations excluded from the high probability set.

The estimator in (2.23) incorporates Newey's (1990) theory about optimal instruments. If Y_{2i}^* in the model (2.21) is not misclassified (now $Y_{2i} = Y_{2i}^*$) but endogenous, an optimal IV estimator will be

$$\tilde{\beta} = \frac{\sum_{i=1}^{N}(\hat{P}_{2i} - \bar{P}_2)Y_{1i}}{\sum_{i=1}^{N}(\hat{P}_{2i} - \bar{P}_2)Y_{2i}} \tag{2.24}$$

Without misclassification, (2.24) gives a consistent estimator for the parameter β. The estimator in (2.23) is also a type of IV estimator but differs from the conventional IV estimator by using a fraction of all observations. Based on this "IV-type" estimator, it appears possible to resolve the estimation in future research for models where the regressor of interest is not only misclassified but also endogenous. Essentially, the estimator $\hat{\beta}$ is attained by conducting the IV estimation only on the high probability set. Such observations are "effective" because their P_2 and P_2^* converge to 0 with the same order as the index becomes sufficiently small, and to 1 as the index becomes sufficiently large. On the set, the discrepancy of P_2 and P_2^* will vanish and the bias and variance of the estimator in (2.23) will converge to 0.

The selection of the high probability set presents a technical challenge. The effective sample size is the number of observations from the high probability set. As $N \to \infty$, a smaller fraction of all observations are used in the estimation of $\hat{\beta}$, because according to the definition in (2.22), a bigger N shrinks the size of high probability set. For a given value of N, as the value of a approaches 0, almost all of the observations are included in the estimation, resulting in a substantial bias in the presence of the discrepancy between P_2 and P_2^*. As the value of a approaches 1, only those observations for which the index is on the very extreme ends will be selected, resulting in a negligible bias but a potentially significant variance due to having too few observations included in the estimation. Therefore, the optimal a is data dependent and it is important to balance the bias (squared) and the variance of the estimator. In the next section, I provide assumptions and definitions required to study this bias-variance tradeoff.

2.4 Assumptions and Definitions

To establish consistency and asymptotic normality of the estimator in (2.23), I make the following assumptions:

- A1. The observations $\{Y_{1i}, Y_{2i}, X_i\}$ are i.i.d. over i, with $X_i \equiv [X_{1i}, X_{2i}]$.
- A2. $\mathbf{E}(\epsilon|Y_2^*, X) = 0$. $\mathbf{E}(\epsilon\epsilon'|Y_2^*, X) = \sigma^2 I$.
- A3. The index restrictions in (2.10)-(2.13) hold. There is at least a continuous variable in the vector X_1.
- A4. $\mathbf{E}(Y_{1i}|Y_{2i}^*, Y_{2i}, X_i) = \mathbf{E}(Y_{1i}|Y_{2i}^*, X_i)$.
- A5. The probability $P_{2i}(V_i)$ is a nondecreasing function of the index V_i, for $V_i < \bar{V_S}$ sufficiently small and for $V_i > \bar{V_L}$ sufficiently large. $P_{2i}(V_i) \to 0$ as $V_i \to -\infty$ and $P_{2i}(V_i) \to 1$ as $V_i \to +\infty$.
- A6. $P_{2i}(V_i)$ and $P_{2i}^*(V_i)$ converge to 0 (1) at the same rate as $V_i \to -\infty$ ($V_i \to +\infty$).
- A7. In the indicator function (2.3), let $G_v(.)$ and $F_\mu(.)$ denote the cumulative distribution functions of the index and the error μ_i, respectively. Since P_{2i}^* depends on the distribution of the error μ_i ($Y_{2i}^* = I\{V_i > \mu_i\}$, so $P_{2i}^* = F_\mu(V_i)$), and P_{2i}^* has the same order as P_{2i} on both ends, I don't distinguish the error μ_i for P_{2i}^*

from that for P_{2i}. Assume the tail conditions as:

For all $t < T$ sufficiently small,

$$G_v(t) > F_\mu(t) \tag{2.25}$$

For all $t > T$ sufficiently large,

$$1 - G_v(t) > 1 - F_\mu(t) \tag{2.26}$$

- A8. With $g_v(.)$ as the density for the index, V_i, $g_v(.)$ is increasing in the tail on the left, decreasing in the tail on the right. Further, $g_v(.) \geq N^{-\xi}$, where ξ is an arbitrarily small positive number. With $H(.) \equiv \frac{g_v(.)}{1-G_v(.)}$ as the hazard for the index V, I assume

$$\frac{1 - F_\mu(v_{er})}{1 - G_v(v_{er})} < H(v_{er})N^{-a_{0N}} \tag{2.27}$$

$$\frac{F_\mu(v_{el})}{1 - G_v(v_{el})} < H(v_{el})N^{-a_{0N}} \tag{2.28}$$

where v_{el} and v_{er} denote the left and right limit value of the index, respectively, such that $g_v(v) \geq N^{-\xi}$ for $v \geq v_{el}$ or $v \leq v_{er}$.

The first two assumptions are basic for many regression models. Assumption (A3) requires that the distributions of Y_2 and Y_2^* depend on the same index, which I have discussed in Secton 3.1. Assumption (A4) implies the non-differential measurement error, which will not provide extra information on the dependent variable given the true variable and other exogenous variables.

Assumption (A5) makes it possible to select "effective" observations through the high probability set in (2.22), given $P_{2i}(V_i)$'s monotonic feature for extreme index values. While (A5) assumes that $P_{2i}(V_i)$ converges to the same value as $P_{2i}^*(V_i)$ for the lower and upper support points of the index, Assumption (A6) steps further to guarantee that they converge in the same order in both tails. It suffices that $PL(V_i)$ converges to 0 as $V_i \to -\infty$ at a faster rate than $P_{2i}(V_i)$ converges to 0, while $PR(V_i)$ converges to 0 as $V_i \to +\infty$ at a faster rate than $P_{2i}(V_i)$ converges to 1. Meanwhile, $1 - PL(V_i) \nrightarrow 1$ as $V_i \to +\infty$ while $1 - PR(V_i) \nrightarrow 1$ as $V_i \to -\infty$. To check it, see the relation between Y_2's and Y_2^*'s distribution by equation (2.14) for the case where $V_i \to -\infty$. As $V_i \to +\infty$, check the rewritten relation equation as

$$1 - P_{2i}(V_i) = PR(V_i)P_{2i}^*(V_i) + (1 - PL(V_i))(1 - P_{2i}^*(V_i) \tag{2.29}$$

To state it informally, Assumption (A6) implies that $PL(V_i) + PR(V_i) < 1$ for very extreme values of the index. It exactly coincides with "Assumption 2- Restriction on the Extent of Misclassification" in Mahajan (2006) (pp.637) except that Mahajan (2006) assumed it on the general real line, $\mathbb{R}$, not only at infinity. The misclassification literature has discussed the similar assumption (e.g. Bollinger, 1996; Frazis and Loewenstein, 2003; van Hasselt and Bollinger, 2012; DiTraglia and García-Jimeno, 2017), making sure that the surrogate Y_{2i} is positively correlated with the true variable Y_{2i}^*. While the observed surrogate is contaminated by measurement error, it is at least better than an arbitrary guess.

The tail conditions specified in Assumption (A7) mean that the tails of the index distribution are fatter than those of the error term, μ_i, on both ends, similar to the assumption in Andrews and Schafgans (1998) and Klein, Shen and Vella (2015). Such tail conditions are needed in the establishment of the estimator's large sample properties. In particular, Assumption (A7) helps to bound the terms $\mathbf{E}(S_{Li})$ and $\mathbf{E}(S_{Ri})$ from below, which will contribute to the order analysis of the squared bias and the variance. See that in Lemma 1 in Appendix.

Finally, Assumption (A8) follows the assumption (A5b) in Klein, Shen and Vella (2015). The high probability set tends to select extreme observations and those "too extreme" observations must be removed by a trimming function. This assumption is necessary to guarantee the high probability set after trimming is not empty. Let v_l be the left threshold value of the index such that $P_2(v) \leq N^{-a}$ for $v \leq v_l$, and v_r be the right threshold value such that $P_2(v) \geq 1 - N^{-a}$ for $v \geq v_r$. Taking the left tail as an example, the nonempty high probability set implies that $v_{el} < v_l$. It holds when

$$F_\mu(v_{el}) < F_\mu(v_l) = N^{-a} \Leftrightarrow \frac{F_\mu(v_{el})}{1 - G_v(v_{el})} < \frac{N^{-a}}{1 - G_v(v_{el})} = H(v_{el})N^{-(a-\xi)} \tag{2.30}$$

where $F_\mu(v_l) = P_2(v_l) \leq N^{-a}$, since P_2^* has the same order as P_2 on the high probability set.

The relevant definitions used in the above assumptions are provided as follows:

- D1. **True and estimated index parameters**.

$$V = X_{11} + X_{12}\psi_{20} + X_{13}\psi_{30} + ... + X_{1k}\psi_{k0} + X_2\theta_0 \tag{2.31}$$

$$\hat{V} = X_{11} + X_{12}\hat{\psi}_2 + X_{13}\hat{\psi}_3 + ... + X_{1k}\hat{\psi}_k + X_2\hat{\theta} \tag{2.32}$$

  where $\{\psi_{m0}\}_{m=2}^k$ and θ_0 denote the true index parameters, and $\{\hat{\psi}_m\}_{m=2}^k$ and $\hat{\theta}$ are their corresponding estimators.

- D2. **Population conditional probability P_2 and its sample counterpart $\hat{P}_2$**. Here I define them at their realization t:

$$P_2(V = t) = Pr(Y_2 = 1|V = t) \tag{2.33}$$

$$\hat{P}_2(\hat{V} = t) = \frac{\sum_j \frac{1}{Nh} Y_{2j} K(\frac{t-V_j}{h})}{\sum_j \frac{1}{Nh} K(\frac{t-V_j}{h})} \tag{2.34}$$

  where K is a standard normal kernel with window $h = O(N^{-.2})$.

- D3. **The S-function**. $S(.)$ is a selection function by which those observations on the high probability set are selected in the estimation whereas those outside the high probability set are excluded. Dropping the subscripts, the definition of the selection function is

$$S(V, a, P_2) = S(V, x(a, P_2)) + S(V, y(a, P_2)) = S_L + S_R \tag{2.35}$$

$$S(V, x(a, P_2)) = \tau(V)C(x(a, P_2)) \tag{2.36}$$

$$S(V, y(a, P_2)) = \tau(V)C(y(a, P_2)) \tag{2.37}$$

$$C(z) = \begin{cases} 0 & z \leq 0 \\ 1 - exp\frac{-z^k}{b^k - z^k} & 0 < z < b \\ 1 & z \geq b \end{cases} \tag{2.38}$$

$$\tau(V) = \frac{1}{1 + \exp[N^{.2}(\frac{\mathbf{E}(g_v(V))N^{-.005}}{\ln(N)} - g_v(V))]} \tag{2.39}$$

$$x(a, P_2) = \ln \frac{1}{P_2} - \ln N^a \tag{2.40}$$

$$y(a, P_2) = \ln \frac{1}{1 - P_2} - \ln N^a \tag{2.41}$$

Here, $C(x(a, P_2))$ and $C(y(a, P_2))$ are the core of the selection function, an extension of the smooth selection function in Andrews and Schafgans (1998). It smoothly selects the high probability set, and the addition of a trimming function $\tau(V)$ helps to trim out those too extreme index values for which the density $g_v(V)$ goes to zero too fast. For notational convenience, let S_L denote $S(V, x(a, P_2))$ and S_R denote $S(V, y(a, P_2))$. I use the same smooth selection function form, S, as Klein, Shen and Vella (2015) except for the two extreme ends required to control here instead of only one end in their work. The sub-choosing function S_L mainly helps to select those observations for which indices are located on the left end in $\mathbb{R}$, while S_R does for the right end. $\hat{S}_L$ and $\hat{S}_R$ denote their sample counterparts, $S(\hat{V}, x(\hat{a}, \hat{P}_2))$ and $S(\hat{V}, y(\hat{a}, \hat{P}_2))$, respectively, and $\hat{S} = \hat{S}_L + \hat{S}_R$. Note that $\hat{S}$ is defined at the estimated optimal parameter $\hat{a}$ in (D4).

- D4. **True and estimated optimal parameters a_{0N} and $\hat{a}$.** While different values of the parameter a select high probability sets with different sizes, the optimal parameter a_{0N} trades off the squared bias and the variance of the infeasible estimator. With $A = \{a : 0 < a < .5\}$,[3] $\hat{S}_L(a) = S(\hat{V}, x(a, \hat{P}_2(\hat{V})))$ and $\hat{S}_R(a) = S(\hat{V}, x(a, \hat{P}_2(\hat{V})))$

$$a_{0N} = arg\min_{a \in A}(N^{2a-1}[\tfrac{\mathbf{E}(S_L^2)}{(\mathbf{E}(S_L))^2} + \tfrac{\mathbf{E}(S_R^2)}{(\mathbf{E}(S_R))^2}] - 1)^2 \tag{2.42}$$

$$\hat{a} = arg\min_{a \in A}(N^{2a-1}[\tfrac{\hat{\mathbf{E}}(\hat{S}_L^2(a))}{[\hat{\mathbf{E}}(\hat{S}_L(a))]^2} + \tfrac{\hat{\mathbf{E}}(\hat{S}_R^2(a))}{[\hat{\mathbf{E}}(\hat{S}_R(a))]^2}] - 1)^2 \tag{2.43}$$

where

$$\hat{\mathbf{E}}(\hat{S}_L(a)) = \tfrac{1}{N}\textstyle\sum_i \hat{S}_{Li}(a) \tag{2.44}$$

$$\hat{\mathbf{E}}(\hat{S}_R(a)) = \tfrac{1}{N}\textstyle\sum_i \hat{S}_{Ri}(a) \tag{2.45}$$

$$\hat{\mathbf{E}}(\hat{S}_L^2(a)) = \tfrac{1}{N}\textstyle\sum_i \hat{S}_{Li}^2(a) \tag{2.46}$$

$$\hat{\mathbf{E}}(\hat{S}_R^2(a)) = \tfrac{1}{N}\textstyle\sum_i \hat{S}_{Ri}^2(a) \tag{2.47}$$

[3]The range of a comes from the process of balancing the squared bias and the variance.

The optimal high probability set parameter should equate the orders of the squared bias and the variance of the estimator so that the estimator will converge to the true coefficient as fast as possible. In Appendix, Lemma 2 gives an upper bound on the order of the squared bias of the infeasible estimator explained below and Lemma 3 calculates the order of its variance. The optimal parameter a_{0N} is defined to equate these two orders in (2.42) and $\hat{a}$ is its sample counterpart.

2.5 Consistency and Normality

To analyze the large sample property for the estimator in (2.23), I rewrite it in an equivalent form as follows:

$$\hat{\beta} = \beta + \frac{\beta \frac{1}{N} \sum_{i=1}^{N} (Y_{2i}^{*} - Y_{2i})(\hat{P}_{2i} - \bar{P}_2)\hat{S}_i + \frac{1}{N} \sum_{i=1}^{N} \epsilon_i (\hat{P}_{2i} - \bar{P}_2)\hat{S}_i}{\frac{1}{N} \sum_{i=1}^{N} (\hat{P}_{2i} - \bar{P}_2) Y_{2i} \hat{S}_i} \tag{2.48}$$

Rather than studying the estimator directly, I will firstly study its infeasible substitute

$$\hat{\beta}^{*} = \beta + \frac{\beta \frac{1}{N} \sum_{i=1}^{N} (Y_{2i}^{*} - Y_{2i})(P_{2i} - \frac{\mathbf{E}(P_2 S(a_{0N}))}{\mathbf{E}(S(a_{0N}))}) S_i(a_{0N})}{\mathbf{E}(P_2 - \frac{\mathbf{E}(P_2 S(a_{0N}))}{\mathbf{E}(S(a_{0N}))}) Y_2 S(a_{0N})} + \frac{\frac{1}{N} \sum_{i=1}^{N} \epsilon_i (P_{2i} - \frac{\mathbf{E}(P_2 S(a_{0N}))}{\mathbf{E}(S(a_{0N}))}) S_i(a_{0N})}{\mathbf{E}(P_2 - \frac{\mathbf{E}(P_2 S(a_{0N}))}{\mathbf{E}(S(a_{0N}))}) Y_2 S(a_{0N})} \tag{2.49}$$

The infeasible estimator replaces $\hat{P}_{2i}$, $\hat{S}_i$ and $\bar{P}_2$ in (2.48) with P_{2i}, $S_i(a_{0N})$ and $\frac{\mathbf{E}(P_2 S(a_{0N}))}{\mathbf{E}(S(a_{0N}))}$, respectively. Its denominator becomes the population expectation rather than the sample average in (2.48). Note that while the the feasible estimator in (2.48) depends on the estimated high probability set parameter $\hat{a}$, the infeasible estimator is defined at the true optimal high probability set parameter a_{0N}. After examining the properties for the infeasible estimator, I will show that the feasible estimator in (2.23) (equivalent to (2.48)) is close to the infeasible substitute. In this way, I will get the consistency and asymptotic normality for the estimator in (2.23).

Theorem 1 Under Assumptions (A1-A8),

$$\hat{\beta}^{*} \xrightarrow{p} \beta \tag{2.50}$$

Theorem 2 Under Assumptions (A1-A8), with $Var(\hat{\beta}^*) = \frac{\mathbf{E}(S_L^2(a_{0N}))}{N[\mathbf{E}(S_L(a_{0N}))]^2} + \frac{\mathbf{E}(S_R^2(a_{0N}))}{N[\mathbf{E}(S_R(a_{0N}))]^2}$, $C_N = \sqrt{\frac{1}{Var(\hat{\beta}^*)}}$,

$$C_N(\hat{\beta}^* - \beta) \xrightarrow{d} Z^* \sim N(0,1) \tag{2.51}$$

Theorem 3 Under Assumptions (A1-A8),

$$C_N(\hat{\beta} - \hat{\beta}^*) \xrightarrow{p} 0 \tag{2.52}$$

Theorem 4 Under Assumptions (A1-A8),

$$C_N(\hat{\beta} - \beta) \xrightarrow{d} N(0,1) \tag{2.53}$$

2.6 Simulation Results

In this section, I study the performance of the estimator in (2.23) by Monte Carlo simulations for finite samples. The model is generated as follows:

$$Y_1 = 2 + 1.5X_1 + 3Y_2^* + \epsilon \tag{2.54}$$

$$Y_2^* = I\{3X_1 + 4X_2 + 5X_3 > \mu\} \tag{2.55}$$

$$V = 3X_1 + 4X_2 + 5X_3 \tag{2.56}$$

$$Y_2 = Y_2^* I\{U > PR\} + (1 - Y_2^*)I\{U < PL\} \tag{2.57}$$

where Y_2^* is the unobserved binary treatment. When $Y_2^* = 1$, the observed surrogate Y_2 is 0 with the probability of PR. When $Y_2^* = 0$, $Y_2 = 1$ with the probability of PL. The index V is the linear combination of X_1, X_2 and X_3. It is rescaled to have unity variance. U is a uniformly distributed random variable. The instruments and error term are generated:

$$[X_1, X_2, X_3]' \sim N(0, \Sigma) \tag{2.58}$$

$$\epsilon \sim N(0,1) \tag{2.59}$$

$$\mu \sim N(\bar{\mu}, .25) \tag{2.60}$$

where Σ is a non diagonal matrix, allowing the correlation between any two exogenous variable $X's$. Simulations are conducted when the value of $\bar{\mu}$ takes on -1, -.5, -.1, 0, .1,

.5, 1, respectively. The probabilities of misclassification PR, PL are functions of the index V:

$$PL = .9\Phi(5V) \tag{2.61}$$

$$PR = .9(1 - \Phi(5V)) \tag{2.62}$$

where $\Phi(.)$ is the standard normal CDF. The setting of PR, PL guarantees that PL converges to 0, PR not to 1 as the value of the index is sufficiently small and that PR converges to 0, PL not to 1 as the value of the index is sufficiently large, which is implied by Assumption (A6). It also makes sure that Y_2 and Y_2^* will be close to each other in the high probability set. In addition, the variance of the error term μ is .25 so that the distribution of the index has fatter tails than that of the error μ, satisfying tail conditions in Assumption (A7).[4]

For each value of $\bar{\mu}$, I run the simulation with 1000 observations for 1000 replications. In every replication the coefficient of Y_2^*, β, in the outcome equation is estimated. I compare the performance of the OLS estimator, IV estimator and the IV estimator based on the high probability set (hereafter referred to as IV estimator on HPS) proposed in this paper.

Table 2.1 shows the estimation results by the value of $\bar{\mu}$. In every simulation, I report the mean, standard deviation and root-mean-square error for OLS estimator, IV estimator and IV estimator on HPS, respectively. Overall, the IV estimator on HPS has a very good performance by having much smaller root-mean-square error than the other two counterparts. The OLS estimator is biased downward in all of the settings, which is consistent with the arguments in Aigner (1973). The standard IV estimator is also biased downward, but there is no definite answer to the direction of the bias in IV estimators, especially in my case where the instrument is allowed to be generally correlated with the measurement error outside of high probability set. A numerical calculation of the covariance for the instrument and the measurement error indicates it is positive, supporting an attenuation bias of the standard IV estimator.

---

[4]This design is allowed to increase the variance of the error term to make its tails relatively fatter while the index tails are held fixed. As expected, the thinner are the tails of the error term relative to those of the index, the better are the results of the IV estimation on the high probability set.

Specifically, in the models with μ taking the value of 0, -.5, -.1, .1 and .5, respectively, the IV estimator on HPS performs better than the other two estimators. It has a very small bias and only a slightly bigger standard deviation due to estimation only relying on a fraction of the full sample, and thus a much smaller root-mean-square error compared to its counterparts. In the other two models with $\mu = -1$ or $\mu = 1$, the IV estimator on HPS demonstrates a more significant advantange over the other estimators through the greatly reduced bias. With $\mu = -1$, the mean of the 1000 replications from OLS and IV techniques is just slightly above 1, whereas the mean from IV on HPS is about 2.9. A similar advantage occurs in the model with $\mu = 1$. Compared to the root-mean-square error 1.9 and 1.6 from the OLS and IV techniques, the root-mean-square error is only .2 in the IV estimator on HPS, implying the proposed estimator converges in probability to the true coefficient as fast as possible.

2.7 Conclusions

This paper studies the identification and estimation of models with a binary, misclassified treatment regressor, a problem faced in a broad range of applications in empirical studies. I assume that both the misclassification probabilities and the probability for the true binary variable depend on an index. Then the probability for the misclassified binary variable will be some unknown function of this index. With the index as a linear combination of covariates and potential instruments, I can recover the index semiparametrically. The measurement error is assumed to gradually vanish in the high probability set where the index becomes sufficiently large or small. I estimate the high probability set and establish an IV estimator based on it. The optimal high probability set is obtained by controlling a data-dependent parameter such that the squared bias and the variance of the estimator vanish at the same rate as the sample size increases. Under tail conditions, I study the large sample properties for the proposed estimator, showing that the estimator is consistent and asymptotically normal. I also conduct Monte Carlo simulations to test its validity in finite sample studies. The results show that the estimator proposed in this paper is valid, substantially improving the estimation compared to the OLS estimator and the classical IV estimator.

Besides misclassification, endogeneity is also prevalent in empirical studies. For example in the study of how the health status influences the labor supply, the true health is endogenous in the outcome equation, or the measurement error term is dependent on labor market outcomes (justification bias). Of course, the phenomenon that the endogenous regressor is misclassified exists not merely in health measures, making it appealing to address endogeneity as well as misclassification in many applications. As the current paper addresses the misclassification by relaxing the assumption of zero correlation between the potential instrument and the measurement error, it is promising to exploit the instrument to search for the high probability set that is free of misclassification and implement an IV strategy on that set. Then endogeneity and misclassification can be addressed simultaneously.

2.8 Appendix

Lemma 1. Under tail conditions in Assumption (A7),

$$\mathbf{E}(S_{Li}) > O(N^{-a}) \tag{2.63}$$

$$\mathbf{E}(S_{Ri}) > O(N^{-a}) \tag{2.64}$$

$$\mathbf{E}(S_{Li}^2) > O(N^{-a}) \tag{2.65}$$

$$\mathbf{E}(S_{Ri}^2) > O(N^{-a}) \tag{2.66}$$

Proof. To see they are bounded from below,[5]

$$\begin{aligned}\mathbf{E}(S_{Li}) &> Pr(x \geq b) \\ &= Pr[P_{2i} < N^{-a}\exp(-b)] \\ &= Pr[V_i < F_\mu^{-1}(N^{-a}\exp(-b))] \\ &= G_v[F_\mu^{-1}(N^{-a}\exp(-b))] \\ &> F_\mu[F_\mu^{-1}(N^{-a}\exp(-b))] \\ &= N^{-a}\exp(-b)\end{aligned} \tag{2.67}$$

[5]Since the trimming function only trims out a very small fraction of observations from the high probability set, it is reasonable to assume that the remainder constitutes the main body of the high probability set, and thus the orders of $\mathbf{E}(S_{Li})$ and $\mathbf{E}(S_{Ri})$ are mainly determined by the core $C(z)$.

and

$$\begin{aligned}
\mathbf{E}(S_{Ri}) &> Pr(y \geq b) \\
&= Pr[P_{2i} > 1 - N^{-a} \exp(-b)] \\
&= Pr[V_i > F_\mu^{-1}(1 - N^{-a} \exp(-b))] \\
&= 1 - G_v[F_\mu^{-1}(1 - N^{-a} \exp(-b))] \\
&> 1 - F_\mu[F_\mu^{-1}(1 - N^{-a} \exp(-b))] \\
&= N^{-a} \exp(-b)
\end{aligned} \tag{2.68}$$

where $P_2(v) = F_\mu(v)$ because P_2 has the same order as P_2^* on the high probability set. The proofs for $\mathbf{E}(S_{Li}^2)$ and $\mathbf{E}(S_{Ri}^2)$ follow in a similar way.

When the infeasible estimator $\hat{\beta}^*$ is defined at an arbitrary value of $a \in (0, .5)$, Lemmas 2 and 3 and thus Theorem 1 hold as follows.

Lemma 2. Under Assumptions (A1-A8)

$$Bias^2 = (\mathbf{E}(\hat{\beta}^*) - \beta)^2 \leq O(N^{-2a}) \tag{2.69}$$

Proof.

$$|Bias| = |\mathbf{E}(\hat{\beta}^*) - \beta| = \frac{|\beta \mathbf{E}(Y_2^* - Y_2)(P_2 - \frac{\mathbf{E}(P_2 S)}{\mathbf{E}(S)})S|}{\mathbf{E}(Y_2 P_2 S) - \frac{\mathbf{E}(P_2 S)}{\mathbf{E}(S)}\mathbf{E}(Y_2 S)} \tag{2.70}$$

We can analyze the order of the bias for three cases: (I) $O(\mathbf{E}(S_L)) = O(\mathbf{E}(S_R))$, (II) $O(\mathbf{E}(S_L)) > O(\mathbf{E}(S_R))$, and (III) $O(\mathbf{E}(S_L)) < O(\mathbf{E}(S_R))$. For case (I), $\frac{\mathbf{E}(P_2 S)}{\mathbf{E}(S)} \to \lambda \in (0, 1)$ as $N \to \infty$. The denominator in (2.70)

$$\begin{aligned}
\mathbf{E}(Y_2 P_2 S) - \frac{\mathbf{E}(P_2 S)}{\mathbf{E}(S)}\mathbf{E}(Y_2 S) &= \mathbf{E}(P_2^2 S_L) + \mathbf{E}(P_2^2 S_R) - \frac{\mathbf{E}(P_2 S)}{\mathbf{E}(S)}(\mathbf{E}(P_2 S_L) + \mathbf{E}(P_2 S_R)) \\
&= O(\mathbf{E}(S_R))
\end{aligned} \tag{2.71}$$

The numerator in (2.70)

$$
\begin{aligned}
|\mathbf{E}(Y_2^* - Y_2)(P_2 - \frac{\mathbf{E}(P_2S)}{\mathbf{E}(S)})S| &= |\mathbf{E}(\mathbf{E}[(Y_2^* - Y_2)(P_2 - \frac{\mathbf{E}(P_2S)}{\mathbf{E}(S)})S|X])| \\
&= |\mathbf{E}(P_2^* - P_2)(P_2 - \frac{\mathbf{E}(P_2S)}{\mathbf{E}(S)})S| \\
&\leq |\mathbf{E}(P_2^* - P_2)(P_2 - \frac{\mathbf{E}(P_2S)}{\mathbf{E}(S)})S_L| \\
&+ |\mathbf{E}(P_2^* - P_2)(P_2 - \frac{\mathbf{E}(P_2S)}{\mathbf{E}(S)})S_R| \qquad (2.72) \\
&\leq N^{-a}|\mathbf{E}(P_2 - \frac{\mathbf{E}(P_2S)}{\mathbf{E}(S)})S_L| + N^{-a}|\mathbf{E}(P_2 - \frac{\mathbf{E}(P_2S)}{\mathbf{E}(S)})S_R| \\
&= N^{-a}O(\mathbf{E}(S_L)) + N^{-a}O(\mathbf{E}(S_R)) \\
&= N^{-a}O(\mathbf{E}(S_R))
\end{aligned}
$$

Combining the orders of numerator and denominator,

$$
|Bias| \leq \frac{N^{-a}O(\mathbf{E}(S_R))}{O(\mathbf{E}(S_R))} = O(N^{-a}) \qquad (2.73)
$$

The analysis of orders for cases (II) and (III) leads to the same results.

Lemma 3. Under Assumptions (A1-A8)

$$
Var(\hat{\beta}^*) = \frac{\mathbf{E}(S_L^2)}{N(\mathbf{E}(S_L))^2} + \frac{\mathbf{E}(S_R^2)}{N(\mathbf{E}(S_R))^2} \qquad (2.74)
$$

Proof. Taking case (I) as an example,

$$
Var(\hat{\beta}^* - \beta) = \frac{Var[\beta(Y_2^* - Y_2)(P_2 - \frac{\mathbf{E}(P_2S)}{\mathbf{E}(S)})S + (P_2 - \frac{\mathbf{E}(P_2S)}{\mathbf{E}(S)})S\epsilon]}{N[\mathbf{E}(Y_2P_2S) - \frac{\mathbf{E}(P_2S)}{\mathbf{E}(S)}\mathbf{E}(Y_2S)]^2} \qquad (2.75)
$$

Analyzing the numerator,

$$
\begin{aligned}
&Var[\beta(Y_2^* - Y_2)(P_2 - \frac{\mathbf{E}(P_2S)}{\mathbf{E}(S)})S + (P_2 - \frac{\mathbf{E}(P_2S)}{\mathbf{E}(S)})S\epsilon] \\
&= \beta^2\mathbf{E}(Y_2^* - Y_2)^2(P_2 - \frac{\mathbf{E}(P_2S)}{\mathbf{E}(S)})^2S^2 \qquad (2.76) \\
&+ \mathbf{E}(P_2 - \frac{\mathbf{E}(P_2S)}{\mathbf{E}(S)})^2S^2\epsilon^2 - \beta^2[\mathbf{E}(Y_2* - Y_2)(P_2 - \frac{\mathbf{E}(P_2S)}{\mathbf{E}(S)})]^2
\end{aligned}
$$

The first term on the right hand side,

$$\begin{aligned}\mathbf{E}(Y_2^* - Y_2)^2(P_2 - \frac{\mathbf{E}(P_2S)}{\mathbf{E}(S)})^2S^2 &= \mathbf{E}\{(P_2 - \frac{\mathbf{E}(P_2S)}{\mathbf{E}(S)})^2S^2\mathbf{E}[(Y_2^* - Y_2)^2|X]\} \\ &= \mathbf{E}[P_RP_2^* + P_L(1-P_2^*)](P_2 - \frac{\mathbf{E}(P_2S)}{\mathbf{E}(S)})^2S_L^2 \\ &+ \mathbf{E}[P_RP_2^* + P_L(1-P_2^*)](P_2 - \frac{\mathbf{E}(P_2S)}{\mathbf{E}(S)})^2S_R^2 \\ &\leq N^{-a}\mathbf{E}(P_2 - \frac{\mathbf{E}(P_2S)}{\mathbf{E}(S)})^2S_L^2 + N^{-a}\mathbf{E}(P_2 - \frac{\mathbf{E}(P_2S)}{\mathbf{E}(S)})^2S_R^2 \\ &= N^{-a}\mathbf{E}(P_2 - \frac{\mathbf{E}(P_2S)}{\mathbf{E}(S)})^2S^2\end{aligned} \tag{2.77}$$

where the conditional expectation of $(Y_2^* - Y_2)^2$ given X is calculated by the sum of four cells ($Y_2^* = Y_2 = 0, Y_2^* = Y_2 = 1, Y_2^* = 1$ while $Y_2 = 0$, and $Y_2^* = 0$ while $Y_2 = 1$) multiplied by their respective conditional probabilities. The inequality comes from Assumption (A6).

The second term converges to zero slower than the first term, since

$$\begin{aligned}\mathbf{E}(P_2 - \frac{\mathbf{E}(P_2S)}{\mathbf{E}(S)})^2S^2\epsilon^2 &= \sigma^2\mathbf{E}(P_2 - \frac{\mathbf{E}(P_2S)}{\mathbf{E}(S)})^2S_L^2 + \sigma^2\mathbf{E}(P_2 - \frac{\mathbf{E}(P_2S)}{\mathbf{E}(S)})^2S_R^2 \\ &= O(\mathbf{E}(S_L^2)) + O(\mathbf{E}(S_R^2))\end{aligned} \tag{2.78}$$

For the third term, by the calculation of the Bias numerator,

$$[\mathbf{E}(Y_2{*} - Y_2)(P_2 - \frac{\mathbf{E}(P_2S)}{\mathbf{E}(S)})]^2 \leq O(N^{-2a}[\mathbf{E}(S_R)^2] \tag{2.79}$$

So the second term is the order contributor in the Variance numerator.

The order of the Variance denominator is

$$N[\mathbf{E}(Y_2P_2S) - \frac{\mathbf{E}(P_2S)}{\mathbf{E}(S)}\mathbf{E}(Y_2S)]^2 = O(N[\mathbf{E}(S_R)]^2) \tag{2.80}$$

Combining the numerator and variance of Variance,

$$\begin{aligned}Variance &= \frac{O(\mathbf{E}(S_L^2)) + O(\mathbf{E}(S_R^2))}{O(N[\mathbf{E}(S_R)]^2)} \\ &= \frac{\mathbf{E}(S_L^2)}{N(\mathbf{E}(S_L))^2} + \frac{\mathbf{E}(S_R^2)}{N(\mathbf{E}(S_R))^2}\end{aligned} \tag{2.81}$$

The last equality comes from the $O(\mathbf{E}(S_L)) = O(\mathbf{E}(S_R))$ in case (I). The same results hold for cases (II) and (III).

Theorem 1. Under Assumptions (A1-A8),

$$\hat{\beta}^* \xrightarrow{p} \beta \tag{2.82}$$

Proof. Lemma 2 shows $Bias^2 \to 0$ as $N \to \infty$. For the variance, since $\mathbf{E}(S_L^2) \leq \mathbf{E}(S_L)$, then $\frac{\mathbf{E}(S_L^2)}{(\mathbf{E}(S_L))^2} \leq \frac{1}{\mathbf{E}(S_L)}$. As Lemma 1 implies $\mathbf{E}(S_L) > O(N^{-a})$, $\frac{\mathbf{E}(S_L^2)}{N(\mathbf{E}(S_L))^2} \leq \frac{1}{N\mathbf{E}(S_L)} < \frac{1}{N^{1-a}}$. Because $0 < a < .5$, $\frac{\mathbf{E}(S_L^2)}{N(\mathbf{E}(S_L))^2} \to 0$ as $N \to \infty$. The similar argument applies to $\frac{\mathbf{E}(S_R^2)}{N(\mathbf{E}(S_R))^2}$. So $Variance \to 0$, as $N \to \infty$. Then consistency of $\hat{\beta}^*$ follows.

Theorem 2. Under Assumptions (A1-A8), with $Var(\hat{\beta}^*) = \frac{\mathbf{E}(S_L^2)}{N(\mathbf{E}(S_L))^2} + \frac{\mathbf{E}(S_R^2)}{N(\mathbf{E}(S_R))^2}$, $C_N = \sqrt{\frac{1}{Var(\hat{\beta}^*)}}$,

$$C_N(\hat{\beta}^* - \beta) \xrightarrow{d} Z^* \sim N(0,1) \tag{2.83}$$

Proof. The result follows from the Lindeberg condition as in Klein, Shen and Vella (2015).

From Lemmas 2 and 3, we have known the bias and variance of the infeasible estimator $\hat{\beta}^*$ that is defined at an arbitrary value of $a \in (0, .5)$. The optimal high probability set parameter for this infeasible estimator, a_{0N}, is determined in (2.42) to balance the orders of the squared bias and the variance. Since the optimal parameter a_{0N} has been determined, the infeasible estimator in the following lemmas and theorems is defined at a_{0N}. For notational simplicity, I abbreviate $S(a_{0N}), S_L(a_{0N})$ and $S_R(a_{0N})$ by S, S_L and S_R, respectively. To show the the feasible estimator is close to its infeasible substitute, I first prove the following lemmas.

Lemma 4. Under Assumptions (A1-A8),

$$C_N \frac{\frac{1}{N}\sum_{i=1}^N (Y_{2i}^* - Y_{2i})(P_{2i} - \frac{\mathbf{E}(P_2 S)}{\mathbf{E}(S)})S_i + \frac{1}{N}\sum_{i=1}^N \epsilon_i (P_{2i} - \frac{\mathbf{E}(P_2 S)}{\mathbf{E}(S)})S_i}{\mathbf{E}(P_2 - \frac{\mathbf{E}(P_2 S)}{\mathbf{E}(S)})Y_2 S} = O_p(1) \tag{2.84}$$

Proof. It is sufficient to show

$$T_1 \equiv C_N \frac{\frac{1}{N}\sum_{i=1}^{N}(Y_{2i}^* - Y_{2i})(P_{2i} - \frac{\mathbf{E}(P_2S)}{\mathbf{E}(S)})S_i}{\mathbf{E}(P_2 - \frac{\mathbf{E}(P_2S)}{\mathbf{E}(S)})Y_2S} = O_p(1) \tag{2.85}$$

and

$$T_2 \equiv C_N \frac{\frac{1}{N}\sum_{i=1}^{N}\epsilon_i(P_{2i} - \frac{\mathbf{E}(P_2S)}{\mathbf{E}(S)})S_i}{\mathbf{E}(P_2 - \frac{\mathbf{E}(P_2S)}{\mathbf{E}(S)})Y_2S} = O_p(1) \tag{2.86}$$

For T_1,

$$\begin{aligned}\mathbf{E}(T_1) &\leq C_N \frac{\mathbf{E}|(Y_2^* - Y_2)(P_2 - \frac{\mathbf{E}(P_2S)}{\mathbf{E}(S)})S}{\mathbf{E}(P_2 - \frac{\mathbf{E}(P_2S)}{\mathbf{E}(S)})Y_2S} \\ &= C_N N^{-a_{0N}} \frac{\mathbf{E}|P_2 - \frac{\mathbf{E}(P_2S)}{\mathbf{E}(S)}|S_L + \mathbf{E}|P_2 - \frac{\mathbf{E}(P_2S)}{\mathbf{E}(S)}|S_R}{\mathbf{E}(P_2 - \frac{\mathbf{E}(P_2S)}{\mathbf{E}(S)})Y_2S}\end{aligned} \tag{2.87}$$

According to the definition of a_{0N} in (2.42),

$$C_N = \frac{1}{\sqrt{\frac{\mathbf{E}(S_L^2)}{N(\mathbf{E}(S_L))^2} + \frac{\mathbf{E}(S_R^2)}{N(\mathbf{E}(S_R))^2}}} = N^{a_{0N}} \tag{2.88}$$

And $\frac{\mathbf{E}|P_2 - \frac{\mathbf{E}(P_2S)}{\mathbf{E}(S)}|S_L + \mathbf{E}|P_2 - \frac{\mathbf{E}(P_2S)}{\mathbf{E}(S)}|S_R}{\mathbf{E}(P_2 - \frac{\mathbf{E}(P_2S)}{\mathbf{E}(S)})Y_2S} = O(1)$. Therefore, $T_1 = O_p(1)$. For T_2, this term on the left hand side has zero mean and bounded variance. By Markov Inequality, it is bounded in probability.

Lemma 5. With Defition (D4), there exist $\delta > 0$ such that

$$\hat{a} - a_{0N} = o_p(N^{-\delta}) \tag{2.89}$$

Proof. See Lemma (6-7) in Klein, Shen and Vella (2015). Their lemmas considered that the estimated high probability set parameter $\hat{a}$ was close to the optimal high probability set parameter a_{0N}. That argument can be readily extended in this paper.

Theorem 3. Under Assumptions (A1-A8),

$$C_N(\hat{\beta} - \hat{\beta}^*) \xrightarrow{p} 0 \tag{2.90}$$

Proof. It is equivalent to show

$$C_N[(\hat{\beta} - \beta) - (\hat{\beta}^* - \beta)] \xrightarrow{p} 0 \tag{2.91}$$

where with $\bar{P}_2 \equiv \frac{\frac{1}{N}\sum_{i=1}^N \hat{P}_{2i}\hat{S}_i}{\frac{1}{N}\sum_{i=1}^N \hat{S}_i}$,

$$\hat{\beta} - \beta = \frac{\beta\frac{1}{N}\sum_{i=1}^N (Y_{2i}^* - Y_{2i})(\hat{P}_{2i} - \bar{P}_2)\hat{S}_i + \frac{1}{N}\sum_{i=1}^N \epsilon_i(\hat{P}_{2i} - \bar{P}_2)\hat{S}_i}{\frac{1}{N}\sum_{i=1}^N (\hat{P}_{2i} - \bar{P}_2)Y_{2i}\hat{S}_i} \tag{2.92}$$

$$\hat{\beta}^* - \beta = \frac{\beta\frac{1}{N}\sum_{i=1}^N (Y_{2i}^* - Y_{2i})(P_{2i} - \frac{\mathbf{E}(P_2S)}{\mathbf{E}(S)})S_i + \frac{1}{N}\sum_{i=1}^N \epsilon_i(P_{2i} - \frac{\mathbf{E}(P_2S)}{\mathbf{E}(S)})S_i}{\mathbf{E}(P_2 - \frac{\mathbf{E}(P_2S)}{\mathbf{E}(S)})Y_2S} \tag{2.93}$$

Let

$$\hat{A} = \frac{1}{N}\sum_{i=1}^N (Y_{2i}^* - Y_{2i})(\hat{P}_{2i} - \bar{P}_2)\hat{S}_i + \frac{1}{N}\sum_{i=1}^N \epsilon_i(\hat{P}_{2i} - \bar{P}_2)\hat{S}_i \tag{2.94}$$

$$A = \frac{1}{N}\sum_{i=1}^N (Y_{2i}^* - Y_{2i})(P_{2i} - \frac{\mathbf{E}(P_2S)}{\mathbf{E}(S)})S_i + \frac{1}{N}\sum_{i=1}^N \epsilon_i(P_{2i} - \frac{\mathbf{E}(P_2S)}{\mathbf{E}(S)})S_i \tag{2.95}$$

$$\hat{B} = \frac{1}{N}\sum_{i=1}^N (\hat{P}_{2i} - \bar{P}_2)Y_{2i}\hat{S}_i \tag{2.96}$$

$$B = \mathbf{E}(P_2 - \frac{\mathbf{E}(P_2S)}{\mathbf{E}(S)})Y_2S \tag{2.97}$$

Then the left-hand side in (2.91) equals

$$\begin{aligned} C_N(\frac{\hat{A}}{\hat{B}} - \frac{A}{B}) &= C_N(\frac{\hat{A} - A}{\hat{B}} + \frac{A}{\hat{B}} - \frac{A}{B}) \\ &= C_N(\frac{\hat{A} - A}{\hat{B}} - \frac{\hat{B} - B}{\hat{B}}\frac{A}{B}) \\ &= C_N\frac{\hat{A} - A}{B}\frac{B}{\hat{B}} - C_N\frac{A}{B}\frac{\hat{B} - B}{\hat{B}} \end{aligned} \tag{2.98}$$

To show the theorem, it suffices to show: (1) $C_N\frac{A}{B} = O_p(1)$, (2) $\frac{B}{\hat{B}} \xrightarrow{p} 1$, (3) $C_N\frac{\hat{A}-A}{B} \xrightarrow{p} 0$, and (4) $\frac{\hat{B}-B}{\hat{B}} \xrightarrow{p} 0$. The condition (4) implies (2) directly, then it suffices to show (1), (3) and (4). Lemma 4 shows condition (1). Employing Lemma 5 and arguments similar to these in Lemma 4, it can be shown that the conditions (3) and (4) hold.

Theorem 4. Under Assumptions (A1-A8),

$$C_N(\hat{\beta} - \beta) \xrightarrow{d} N(0,1) \tag{2.99}$$

Proof. The result follows from Theorems 1, 2 and 3.

Table 2.1: Estimation of Coefficient β

	True β		OLS	IV	IV on HPS
$\bar{\mu} = 0$	3	mean	2.1163	2.5306	2.9655
		std	0.0650	0.0710	0.0849
		rmse	0.8860	0.4747	0.0916
		mean	1.1382	1.3636	2.8823
$\bar{\mu} = -1$	3	std	0.0657	0.0865	0.1429
		rmse	1.8630	1.6386	0.1851
		mean	1.7986	2.1758	2.9478
$\bar{\mu} = -.5$	3	std	0.0681	0.0802	0.1003
		rmse	1.2033	0.8281	0.1131
		mean	2.1032	2.5164	2.9649
$\bar{\mu} = -.1$	3	std	0.0654	0.0701	0.0858
		rmse	0.8992	0.4887	0.0926
		mean	2.1030	2.5154	2.9651
$\bar{\mu} = .1$	3	std	0.0654	0.0724	0.0853
		rmse	0.8994	0.4900	0.0921
		mean	1.7968	2.1703	2.9459
$\bar{\mu} = .5$	3	std	0.0641	0.0788	0.0992
		rmse	1.2049	0.8334	0.1130
		mean	1.1344	1.3543	2.8791
$\bar{\mu} = 1$	3	std	0.0628	0.0861	0.1415
		rmse	1.8667	1.6479	0.1861

Chapter 3

Labor Supply Effects of Health Shocks: A New Approach on Misclassification of Health Measures

3.1 Introduction

Building on the seminal work of Grossman (1972), health has been regarded as part of human capital that plays an important role in determining an individual's labor market behavior. Health shocks will directly reduce the time that can be allocated between work and leisure and decrease wage rates by reducing productivity. Moreover, Gustman and Steinmeier (1986) pointed out that health shocks will change the marginal substitution rate of goods and leisure. Accordingly, it is necessary to study the labor supply effects of health shocks to assess individuals' economic circumstances, and also to assist policy makers in designing health-related policies to help particularly vulnerable groups.

Among the challenges in studying the labor supply effects of health shocks is the measurement error in health variables. Health economists desire a "perfect" health variable that accurately measures the true health status, and at the same time, captures the components of health that are determinants of work capacity. But such a "perfect" health variable is rarely available in real world data given the substantially varied definitions of health from individual to individual and the pervasive imperfect information among the population on own health status. In this context, the self-reported health measures serve as a comprehensive substitute for true health since they have been found to be more correlated with work capacity than other health measures (Blau et al., 2001). While such self-reported health measures provide information on overall health, they are subject to measurement error in a wide range of survey data. Butler et al. (1987) found sizable disparities between the reported arthritis symptoms and the recorded arthritis diagnoses. My analysis of the 2012 wave of the Health and Retirement Study (HRS) also suggests the existence of enormous inconsistencies between the measure of work-limiting health problems and the measure of self-reported health status. Failure to address the measurement error of health variables will lead to biased estimation and misleading recommendations for economic policy.

However, few papers recognize that the measurement error of many health variables is non-classical because these health measures are recorded as discrete variables in survey data. For example, in the binary case, if an individual truly has good health, the measurement error occurs only when he reports poor health, and vice versa. Such measurment error in binary health variables creates an econometric issue distinct from the classical measurement error. To address it, some previous studies use a series of objective health measures, like functional limitations or doctors' diagnoses, to instrument for self-assessed health measures, ignoring the possible correlation between the instruments and the measurement error as explained in Chapter 2. Another approach widely used in literature is to estimate a health index from objective health measures and then substitute this continuous index for the binary self-reported health variable in the labor supply equation, but this approach introduces interpretation and functional form challenges in practice.

This study uses the method proposed in Chapter 2 to handle the binary, misclassified health variable in the labor supply equation. Also extracting information on true health from objective health measures, this new method relies on such information to dynamically select observations that are free of misclassification. Based on these "correct" observations, an IV estimator is consistent. Using the 2012 wave of the Health and Retirement Study (HRS), this paper primarily examines the labor supply effects of health shocks for men and women aged 45-61. When examining the measure of self-reported health status, the results suggest that men (women) will reduce labor supply by 2,299 (1,929) hours per year when they rate their health as "Fair" or "Poor" and that OLS and traditional IV estimators suffer from considerable attenuation biases compared to the proposed estimator. When examining the measure of work-limiting health, the results suggest that the traditional IV strategy and the proposed new technique produce similar estimates, while the OLS estimate is biased towards zero. As a comparison, I also examine the sample aged 62-70 when the elderly seriously consider retirement. The estimation results suggest smaller reductions in hours of work than for the youger group, but the mean hours of work already lower.

The paper is organized as follows: Section 3.2 reviews literature; Section 3.3 explains how to apply the method proposed in Chapter 2; Section 3.4 describes the data and descriptive statistics of the sample; Section 3.5 discusses estimation results; and Section 3.6 draws conclusions.

3.2 Literature Review

The effect of health shocks on labor supply has been extensively studied for decades. Currie and Madrian (1999) gave a thorough review of literature, pointing out that most studies drew the same conclusion of the negative effect of health shocks on labor supply, but that there was no consensus on its magnitude due to differing health measures and identification methods used. Chirikos and Nestel (1985) used a ten-year record of respondents aged 45-64 from the National Longitudinal Surveys (NLS) to study how the history of health problems affected current wage rates and hours of work. Different from the literature that examined a variety of health measures from the survey data, they constructed a variable of health history from respondents' reported impairments, self-rated health status, and work-limiting health conditions. Including the estimated (log) wage in a tobit model of annual hours worked, they found that compared to the report of continuously good health, having had a health problem, regardless of current health status, lowered hours worked per year. The reduced annual hours of work ranged from 81 hours for white women to 388 hours for black women. Mitchell and Burkhauser (1990) addressed the wage and hours equations simultaneously using working-age men and women from the 1978 Survey of Disability and Work. Estimating a nonlinear simultaneous model for each gender with and without arthritis, they found that the condition of arthritis explained 42.1% of the difference in work hours between men with and without arthritis, 36.7% for women aged 18-44 and 51.0% for women aged 45-64.

As the aging of population has become a pressing issue in many developed countries, more attention has been focused on the labor market behavior of the elderly. Bound et al. (1999) used the first three waves of the Health and Retirement Study (HRS) to examine the association between health and labor market transitions for men and women who were 50-62 years old in the first wave. They only selected observations for which respondents were employed in Wave 2 and investigated whether between Wave 2 and Wave 3 these respondents stayed employed at the same job, changed to a different job, applied for some disability insurance or became unemployed without applying for any disability insurance. Since they argued that there likely was a difference between the effects of persistent poor health and of a recent health shock, they regressed the labor market transitions on current health as well as lagged health. The results indicated that workers who suffered a recent health shock were more likely to exit the labor force than those whose health deterioration occurred earlier. In addition, workers who remained in the market after a health shock tended to change to a different job as an alternative adaptation.

Disney et al. (2006) used a longitudinal sample of older adults aged 50-64 in 1991 from the British Household Panel Survey (BHPS) to study how adults make retirement decisions in the face of health shocks. They constructed a continuous health index year by year from a series of objective health indicators and defined the health shock for an individual as the deviation of his or her index value from the cohort mean. Estimating both fixed effects logit models and discrete time hazard models, they found that poor health predicted an increased probability of exiting the market.

García-Gómez et al. (2010) studied the transition into and out of the labor market using the first twelve waves of the British Household Panel Survey (BHPS). The authors selected a sample of individuals who were working in the first wave and a sample of individuals who were not working in the first wave. With the first sample, the authors used the discrete-time duration model to study the role that health limitations, self-assessed health measures and mental health status played in a worker's decision to become unemployed. With the second sample they examined the effect of health declines on a non-worker's re-entering the labor force. The results suggested that a health shock increased the probability of quitting the market and decreased the probability of re-entering the market. Surprisingly, a deterioration in some dimensions of mental health was associated with a rising likelihood of employment, possibly explained by a view that an individual who was stressed while unemployed would be likely to rejoin the labor force (Paul and Moser, 2006).

As Currie and Madrian (1999) reviewed, health is defined differently across various measures in survey data and the estimated effect of health on labor market outcomes is very sensitive to which measure of health is used. Subjective health measures are widely used in studies of labor market activities because they have been found to be more correlated with work capacity than objective measures. Blau et al. (2001) found that when including multiple health measures in labor supply equations, subjective measures that described the comprehensive health status usually had a larger explanatory power than objective indicators that reported only some narrow, concrete dimension of health. The subjectivity of such health measures, however, brings measurement error and may bias the estimation of the labor supply effects of health shocks. Individuals with the identical underlying health may have different thresholds to report poor health. Moreover, a majority of empirical studies suggest that the measurement error in subjective health variables is not random. In particular, individuals who supply less time to, or exit the labor market, are more likely to report worse health to justify their labor market withdrawal. Social Security disability benefits also provide financial incentives to report more severe conditions to meet the eligibility criteria.

To address measurement error in health variables, Stern (1989) proposed to use objective measures to instrument for subjective measures. Since then, this method has become one of the predominant strategies and has been widely used in empirical studies. The objective health measures commonly used in such studies include functional limitations, doctors' diagnoses or mortality information, which only imperfectly reflects the true health but is less likely to suffer from measurement error. See Charles (1999) for an empirical analysis. However, there are two potential flaws when using relatively objective measures to instrument for subjective measures. First, as many subjective health variables are dichotomous, the measurement error is typically non-classical. As it is explained in Chapter 2, the true health indicator is negatively correlated with the measurement error, making a traditional IV estimator biased. Second, Bound (1991) showed mathematically that the measurement error of health measures would distort the estimated effects of other economic factors that were correlated with health, for example education. Even if the measurement error of health measures was addressed, the estimates of the coefficients on these economic factors were still biased.

Another approach addressing the measurement error is to recover the latent health stock. Bound et al. (1999), Disney et al. (2006) and García-Gómez et al. (2010) estimated an underlying health index from a number of objective health indicators and substituted this index for the subjective health variable in the labor supply equation. This method has been extended beyond the study of labor market outcomes, for example, J*ü*rges (2007) examined the differences in self-reported health across countries by estimating the latent health index. While this method mitigates the estimation bias resulting from measurement error, the direct replacement of the discrete health measure by a continuous health index makes interpreting the results challenging. First, most health policies concern the change of individuals' health from "good" to "poor" instead of from "good" to "slightly less than good" or from "poor" to "slightly less than poor". Individuals who suffer from a slight health decline are likely to remain economically active in the labor market. Second, when constructing the continuous health index using a number of objective health indicators, different researchers may have different knowledge and beliefs of which objective indicators are critical as determinants of the true health. Once some important objective indicator was omitted, the evaluation of health might lose a significant dimension. In addition, based on the different objective indicators included, it becomes difficult to make a comparison among studies. Third, an assumption underlying the continuous index is the effect of the labor supply is constant across health status in linear regressions. When interpreting the effect of a health decline measured by the continuous index, the response in hours of work to the decline of health from "good" to "slightly less than good" is assumed the same as the response to the decline from "poor" to "slightly less than poor". However, individuals in good status probably respond differently from their counterparts in poor status when encountering similar health shocks. Therefore, it is crucial to directly address the measurement error in discrete health variables.

To summarize the literature, the essence of these two primary solutions is to extract reliable information on true health from relatively objective measures. While the first solution fully recognizes the discreteness of self-reported health variables, it fails to handle the inherent non-classical measurement error. Meanwhile, it always distorts the estimated coefficients on other economic factors that are correlated with health, making it difficult to compare the relative impacts on labor market decisions of health and other economic factors, for example education. The second solution constructs a continuous health index to circumvent the issue of non-classical measurement error, but ignoring the discreteness of many health measures introduces interpretation challenges in practice.

In contrast, this study uses the method proposed in Chapter 2 to handle the binary, misclassified health variable in the labor supply equation. Also extracting information on true health from objective health measures, this new method relies on such information to dynamically select observations that are free of misclassification. Based on these "correct" observations, an IV estimator is consistent in the estimation of the effect of health shocks on working hours.

3.3 Methodology

3.3.1 Model

In this paper I study the labor supply effect of health shocks using the following structural model:

$$Y_i = \alpha + X_i'\gamma + H_i^*\beta + \epsilon_i \tag{3.1}$$

where Y_i is the observation i's hours of work, H_i^* measures health status and X_i, a $k \times 1$ random vector, includes all other exogenous covariates, for example age, square of age, race, education, marital status and census region. The regression error ϵ_i has mean zero and variance σ^2. Observations are independent and identically distributed over i. The health measure H_i^* is dichotomous, 1 for "poor" and 0 for "good". Accordingly, the coefficient, β, reflects the effect of health shocks on hours worked.[1]

However, the true health status, H_i^*, is rarely observed in practice. Instead, the self-assessed health measures are collected in most survey data through asking the respondents to rate their own health status. While these self-assessed health measures have been shown to be more likely than the objective health measures to reflect one's work capacity, they are more susceptible to measurement error. As virtually all the questionnaire wording in survey data requires the respondents to answer "Yes or No", or at most to pick one from a few options that best describes their health status, these self-assessed health variables are binary or categorical. The measurement error of such health measures is distinct from that of the continuous health measures. Here, I focus on the binary health measures that are subject to measurement error. In particular, if the true health is 0, it can be only misreported, if at all, to be a 1 and if the true health is 1, it can be only misreported to be a 0. As a result, the measurement error is negatively correlated with the true unobserved health. Such measurement error is typically referred to as misclassification. Due to the inherent correlation between the true health, H_i^*, and the measurement error, a conventional IV technique may fail to obtain a consistent estimate. To be specific, when we use the objective health measures, for example functional limitations on daily life activities or doctors' diagnoses, to instrument for the subjective, self-assessed health measures, it is difficult to guarantee that these instruments are uncorrelated with the measurement error, since we know that the instruments are closely correlated with the true health status and that there is negative correlation between the true health and the measurement error.

While the instrumental variable strategy fails to address misclassification, these objective health measures do provide information on the unobserved true health. For example, if an individual reports limitations on many daily life functions and medical conditions diagnosed by doctors, he is likely to have poor health. If an individual

[1] There could be a mass of observations at zero hours that result in a potential sample selection issue. Here, I follow the argument of Angrist and Pischke (2008) about limited dependent variables to run a linear regression. In addition, I compare the marginal effects of all exogenous variables in the case where zero hours are accounted for to the marginal effects in the case where they are not. I find that these two cases present similar marginal effects. In future research, I will modify my theory to handle sample selection and more general, nonlinear models.

reports very few or no limitations or doctors' diagnoses, it is likely that his true health is good. Motivated by this idea, many previous studies agreed that the true, unobserved health depended on the objective health measures in some way. In accordance with the literature, I use the threshold-crossing model of the true health as follows:

$$H_i^* = I\{X_i'\pi_1 + Z_i\pi_2 > \mu\} \tag{3.2}$$

where X_i includes all the covariates from the structural model and Z_i is the objective health variable. The objective health variable is termed the excluded variable, because it does not directly affect the labor supply as long as the subjective health measure is included in the structural model. Identification requires only one excluded variable. To include only one excluded variable here makes it convenient for elaboration. When using real world data, usually many variables meet the exclusion restriction. In this paper I use a series of indicators of medical conditions diagnosed by doctors and a continuous variable, the number of one's functional limitations on daily life activities.

As the measurement error may be generally related to the excluded variable, this paper assumes non-constant misclassification probabilities. Mahajan (2006) firstly recognized that the assumption of constant misclassification probabilities was very strong and thus relaxed it by assuming that the probabilities of misclassification were functions of covariates from the structural equation. To go a step further, I assume that the probabilities of misclassification are functions not only of the covariates X_i but also of the objective health measure Z_i. This further relaxed assumption has a theoretical foundation. As we have discussed, there probably exists a general correlation between the measurement error and the objective health measures. So it is natural to extend the assumption by allowing the probabilities of misclassification to depend on X_i and Z_i. Moreover, this further assumption also has an applicable implication. Individuals who have more limitations on their daily life functions and medical conditions are much less likely to report themselves as healthy. Similarly, individuals who have fewer limitations or even no doctors' diagnoses would be unlikely to report their health status as unhealthy. In short, if an individual has extremely many or extremely few functional limitations and doctors' diagnoses, his or her probabilities of misreporting health status will be extremely low.

3.3.2 Health Index

In the threshold-crossing model of the true health status, the distribution of H_i^* depends on X_i and Z_i through their linear combination. This linear combination combines the demographics and an objective health indicator that are critical determinants of the true health. Define an index V_i:

$$V_i = X_{1i} + X_{2i}\psi_{20} + X_{3i}\psi_{30} + ... + X_{ki}\psi_{k0} + Z_i\theta_0 \tag{3.3}$$

where $\{X_{1i}, ..., X_{ki}\}$ are the k variables in the vector, X_i. I assume the distribution of H_i^* depends on the index, V_i, which is a normalized linear combination of X_i and Z_i. Then

$$P_i^* = Pr(H_i^* = 1|V_i) = Pr(V_i b > \mu_i|V_i) \tag{3.4}$$

$$V_i b = X_i\pi_1 + Z_i\pi_2 = (X_{1i} + X_{2i}\psi_{20} + X_{3i}\psi_{30} + ... + X_{ki}\psi_{k0} + Z_i\theta_0)b \tag{3.5}$$

where P_i^* is the probability of actually having poor health conditoned on the index V_i. The function of P_i^* relies on the unknown distribution of the error term μ_i. Since the index V_i contains many objective health indicators in practice, it is referred to as the health index hereafter.

The misclassification probabilities are assumed to depend on the covariates and the excluded variable. As the health index is a linear combination of covariates and objective health measures, the misclassification probabilities are functions of the index:[2]

$$PL(V_i) = Pr(H_i = 1|H_i^* = 0, V_i) \tag{3.6}$$

$$PR(V_i) = Pr(H_i = 0|H_i^* = 1, V_i) \tag{3.7}$$

where H_i is the self-assessed health measure, 1 for “poor” and 0 for “good”. Individuals with extreme values for the health index (e.g. perhaps due to having extremely many or extremely few functional limitations or doctors’ diagnoses) are less likely to

[2]With the misclassification functions being unknown, it is the normalized index V_i that can be recovered. For our purpose, identification of V_i is sufficient.

misreport their health status. Then as the health index approaches the extremely large or extremely small values, the misclassification probabilities will tend to be 0, that is, no misclassification.

Since both the true health and the misclassification probabilities depend on the health index, by the Law of Total Probability, the self-assessed health is also a function of the index:

$$P_i(V_i) = (1 - PR(V_i))P_i^*(V_i) + PL(V_i)(1 - P_i^*(V_i)) \tag{3.8}$$

where P_i is the probability of reporting poor health ($H_i = 1$) conditioned on the index V_i. While the specific model for the observable H_i is unknown, the parameters $\{\psi_{m0}\}_{m=2}^{k}$ and θ_0 in the index are identified semiparametrically in a single-index model. Therefore, the index V_i and the probability $P_i(V_i)$ are consistently estimated.

3.3.3 Estimator

Before proposing the estimator, it is necessary to simplify the structural model. With the index V_i recovered, I simplify the structural model using the approach proposed by Robinson (1988) for partially linear models. See Chapter 2 for details of the simplification. Essentially, the simplication process separates the estimation of coefficients on other economic covariates from the estimation of health effects on labor supply. Bound (1991) argued that the mismeasured health variables would distort the estimated coefficients on other economic covariates that were related to health. Furthermore, he showed that even if the measurement error of health variables was addressed, the distorted estimation of coefficients on other covariates still remained. To separate such twisting effects on labor supply, I use the Robinson's technique to first estimate the coefficients on other covariates, leaving the health variable alone in the simplified model to address. The Robinson's technique on other covariates does not affect the estimation of health effects on hours of work afterwards. At the same time, the later estimation of how the health shock influences labor supply will not impact the estimation of the coefficients on other covariates in the first step. In this way, we can individually achieve consistent estimates of coefficients on covariates and health variable, making it possible to compare the relative significance of health and other economic covariates in labor supply decisions. The simplified model is:

$$Y_{1i} = \alpha + H_i^* \beta + \epsilon_i \tag{3.9}$$

where Y_{1i} is interpreted as the differenced outcome by subtracting the estimator of $X_i'\gamma$ from the dependent variable Y_i in the structural model. Again, the coefficient β reflects the effect of health shocks on hours worked.

The observations with the extreme index values are important to address the misclassification. The proposed method in Chapter 2 essentially selects a set of the extreme observations, which is termed a high probability set. On this set the observations have no misclassification with a high probability. The estimator is essentially an instrumental variable strategy on the high probability set. The estimator of the coefficient on health shocks is

$$\tilde{\beta} = \frac{\sum_{i=1}^{N}(\hat{P}_i - \bar{P})Y_{1i}\hat{S}_i}{\sum_{i=1}^{N}(\hat{P}_i - \bar{P})H_i\hat{S}_i} \tag{3.10}$$

where $\hat{P}_i$ is the estimate of P_i from a semiparametric model as discussed in Chapter 2. The selection function $\hat{S}_i$ helps to assign different weights to the observations in the sample. See Chapter 2 for its mathematical definition. It assigns high weights (up to 1) to those observations with the extreme index values and low weights (down to 0) to those observations with modest index values. $\bar{P}$ denotes the weighted average of $\hat{P}_i$ on the high probability set, $\frac{\sum_{i=1}^{N}\hat{P}_i\hat{S}_i}{\sum_{i=1}^{N}\hat{S}_i}$.

As discussed in Chapter 2, the high probability set parameter $\hat{a}$ controls which observations are selected. If the selected index does not approach the sufficiently large or sufficiently small values, the bias of the estimator will be substantial because many misreported observations are included. If the value of the index is too extreme, the bias will become negligible but the variance will be significant due to having too few observations. Therefore, the high probability set parameter is determined in the dynamic balance of the (squared) bias and the variance.

It is noteworthy that there are two high probability set parameters, one for each tail. To be specific, there is a high probability set parameter $\hat{a}_1$ in the sub-choosing function $\hat{S}_L$ (see D3 in Chapter 2) that helps to select those observations with extremely small index values, while another parameter $\hat{a}_2$ in the sub-choosing function $\hat{S}_R$ selects observations with extremely large index values. The left tail of the index may be different from the right tail, so two different parameters are needed to control the potential different processes of selecting extreme observations. When the index is symmetrically distributed, these two paramaters will be identical.

3.4 Data

This study uses the wave in 2012 of the Health and Retirement Study (HRS) to analyze how an individual responds in labor supply to his own health shock. The Health and Retirement Study is a nationally representative survey of aging American households. The first cohort entered in 1992 and there were five other cohorts subsequently entering in 1993 (the AHEAD cohort), 1998 (the Children of Depression cohort and the War Baby cohort), 2004 (the Early Baby Boomer cohort), and 2010 (the Mid Baby Boomer cohort). Each respondent is interviewed biennially. In each interview wave, the HRS collects information on health status, employment history, wealth, income, social security, pension and demographics for respondents and their spouses (if any). The data contain an abundant set of health measures, including self-reported health status, work-limiting health problems, functional limitations, doctors' diagnoses, medical care utilization, body mass index and mental health scores. Such a rich set of measures helps to assess individuals' health status from different perspectives and makes it possible to implement the model proposed in Chapter 2.

This study mainly focuses on respondents aged between 45 and 61, because at age 62 individuals are able to collect Social Security retirement benefits and at 65 they are eligible to receive Medicare. Both provide individuals with sizable financial incentives to leave the labor force, leading them to be more likely to change their labor market behavior when they experience health shocks compared to those who have not reached these age thresholds. In addition, I exclude observations with missing reports on labor supply, health and some economic characteristics. As a result, there are 2,877 men and 3,958 women in the sample.

Table ?? defines the variables used in this study. The measure of labor supply is an individual's hours worked per year. The HRS contains the number of hours per week and the number of weeks per year a respondent devotes to his or her main and second job. For each job, I calculate the hours worked as a product of the number of hours per week and the number of weeks per year. The total hours worked are a sum of the hours of work from one's main and second jobs. Those who are described as not working have zero hours of work. The reported educational attainment has five categories: Less than high school, GED, High school, Some college, and College and above. Except for the reference category of Less than high school, there are four dummy variables in the regression model for the other four education levels, respectively. There are two self-assessed health measures used, work-limiting health problem ("Wl hlth" in Table ??) and self-reported health status[3]("Sr hlth" in Table ??). Their respective effects on labor supply will be examined individually. Since men and women may respond in different ways, I examine their responses separately.

3.4.1 Descriptive Statistics

Table 3.2 presents descriptive statistics for men and women, respectively. Panel A reports their basic demographics. The average ages of men and women are both about 56, and a majority of men and women are 51 years and older. There is no substantial gender difference in the distribution of educational attainment and female respondents have a slight edge over male respondents in completed higher education levels. While male respondents are 1.2 percentage points more likely to have the education level of less than high school than female respondents, 30 percent of females have some college education compared to 28.8 percent of males. Females are also slightly more likely to complete a bachelor's degree or higher. To account for the effect of other income sources that are unrelated to labor earnings, I calculate the non-labor income by subtracting

---

[3]Respondents are required to categorize their general health status as "Excellent", "Very Good", "Good", "Fair" and "Poor". Since different individuals may apply different definitions to "Excellent", "Very Good" and "Good", it is more common to translate this into a binary variable. In this paper I divide these five categories into two groups: 1 for "Fair" or "Poor" health and 0 for "Excellent", "Very Good" or "Good" health.

individuals' own labor income from their total household income. On average, females receive 4,000 dollars more of nonlabor income than males. There are more single women than single men in this sample, with 22.5 percent of women being divorced or widowed compared to only 13.5 percent of men in the same marital status.

Panel B compares the labor market behavior for the two genders. Generally speaking, males have a stronger attachment than females to the labor market. They participate in the labor force more actively than females and supply more hours of work. When investigating the males who are working, about 15 percent of them also work on a second job, compared to only 11 percent of females who supply labor in the market. In addition, male workers devote more hours per year than female workers not only to their first job but also to the second one, if any.

Panel C shows the men's and women's reports on their health status. Twenty-five percent of men report a health problem that limits their work. At the same time, there is a similar share (24%) of men reporting "Fair" or "Poor" health measured by self-reported health status. The proportions of women in unhealthy status measured by these two variables appear very close to each other, 26.7 percent for work-limiting health problems versus 27.5 percent for self-reported health status. But it is noteworthy that the group who reports work-limiting health problems is quite different from the group who reports "Fair" or "Poor" in self-reported health measure. Table 3.3 illustrates the reported discrepancy between these two health measures by gender. Among the 709 male respondents with some problem that limits their work, only 404 (57%) report "Fair" or "Poor" health, while there are 289 individuals in the group (693) with "Fair" or "Poor" health reporting no work-limiting problems. Such a large reporting differential between these two health measures also occurs in the sample of female respondents, providing evidence that self-assessed health varies greatly across variables used to measure health and that there may be considerable misclassification of these measures.

To instrument for the subjective health measures, I include the number of functional limitations[4] and several doctors' diagnoses. On average, males have 1.7 kinds of limitations on daily life functions and females have 2.3. Forty-nine percent of men and sixty percent of women report at least one type of limitation. Almost half of the sample suffers from high blood pressure and more than 18 percent from diabetes, indicating the high prevalence of chronic conditions among the aging population. Females are 10 percentage points more likely to experience psychological disorders than males, which aligns with the literature on gender difference in mental health (McManus et al., 2016).

3.5 Results

3.5.1 Self-reported Health Status

Table 3.4 presents estimation results for men aged 45-61 when the self-reported health status is examined. I apply the technique (termed IV on HPS) explained in the Methodology section to estimate the labor supply effect of health shocks. As a comparison, I provide the OLS and IV estimates.[5] When a man rates his health as "Fair" or "Poor", he will work 2,299 hours per year fewer than his counterparts who rate their health as "Good", "Very Good" or "Excellent". While the traditional IV estimation produces results that are not very far from the IV estimation on HPS, the OLS estimator demonstrates a substantial attenuation bias compared to the IV estimator based on high probability set; a man will reduce his labor supply by only 725 hours per year if his health is not good in the OLS regression.

Table 3.5 shows the estimation results for women aged between 45 and 61 when examining the self-reported health status. With respect to the estimated labor supply effect of health shocks, not only the OLS estimator but also the traditional IV estimator

---

[4]To increase the variation of the estimated health index, I use the number of functional limitations instead of a set of indicators for those limitations.

[5]Here, I regress an subjective health measure on covariates and instrumental variables in a semi-parametric model. The health index, as a linear combination of covariates and excluded variables, is estimated. The predicted expectation of the subjective health indicator conditional on this index is used as an optimal instrument for the subjective measure. See Newey (1990) for optimal instrumental variables.

demonstrate attenuation biases compared to the proposed IV estimator based on the high probability set. When estimating the effect of health using the IV technique on high probability set, the result shows that "Fair" or "Poor" health will reduce women's labor supply by 1,929 hours per year. As a comparison, the results from the other two estimators suggest that the working time devoted by women reduces by 602 hours in the OLS estimation and 1,544 hours in the traditional IV estimation.

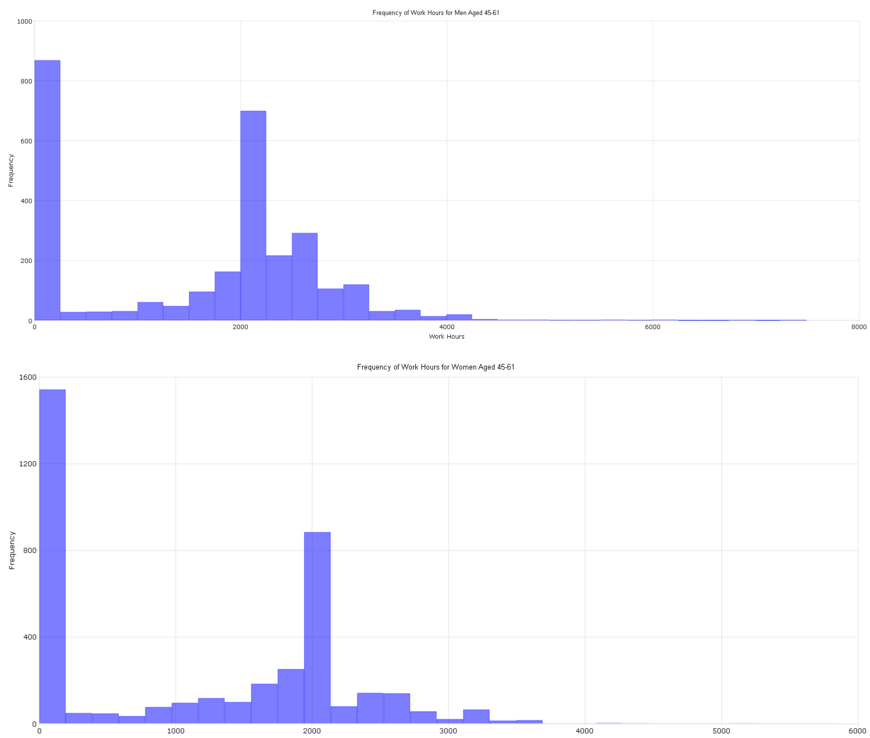

Figure 3.1: Frequency of Work Hours for Men (Upper) and Women (Lower)

Both men and women greatly reduce their labor supply in the face of health declines. In particular, both genders with "Fair" or "Poor" health work about 2,000 fewer hours per year than those who rate their health as "Excellent", "Very Good" or "Good". This amount of reduced working time approximately equals the total hours devoted by a full-time worker (On average, a full-time worker works 40 hours per week and 52 weeks per year). Figure 3.1 shows the histograms of men's and women's hours worked per year. Surprisingly, there are a sizable number of observations with working time above 2,000 hours per year, even up to 3,000 hours per year. It is plausible that a hardworker may decide to greatly reduce his or her hours of work or even exit the labor market when experiencing a health shock.

Generally speaking, the hours worked for males and females increase with age, and the increment rate decreases as people age. Whites supply more hours to the market than the non-whites. Neither males nor females respond in labor supply to the change in their non-labor income. While married men supply more hours to the labor market than single men, married women work slightly less (insignificant) than single women, probably implying the different incentives caused by marriage for males and females. It is noteworthy that the measurement error of health variables will bias the estimated coefficient on other covariates if they are correlated with the mismeasured health variables (Bound, 1991). For example, education is often acknowledged to be strongly related to health status[6] and thus may be correlated with the measurement error of health variables. For the OLS regression on self-reported health status, males with the education level of high school graduate supply 186 hours per year more than the males with educational attainment of less than high school (the reference category). Those with more education, some college or college and above, devote even more time to the labor market, 257 and 429 hours per year, respectively. In contrast, when implementing Robinson's technique for covariates before addressing the misclassification of health variables, those who have a GED degree, a high school diploma or some college education work fewer hours than those with the education level of less than high school.[7] Different from the sample of men, women who have more schooling, for instance high school graduate, some college, college and above, work more than their counterparts with less than high school education.

3.5.2 Work-limiting Health

Table 3.6 and Table 3.7 present the estimation results for men and women, respectively, when the work-limiting health is investigated. The IV estimate on HPS suggests that a

[6]Grossman (1972) believed that adults with higher education were able to produce health more efficiently. Fuchs (1982) proposed a theory that more education was a signal of lower discount rate. Adults with a lower discount rate would like to make a current investment and receive payoffs from the future. Those who had a lower discount rate tended to invest in education and health.

[7]A bachelor's degree or higher also slightly reduces one's hours of work compared to his counterparts with education below high school, but the result is statistically insignificant.

man will reduce his labor supply by 1,713 hours per year when he suffers from a work-related limitation in health. The traditional IV estimation produces similar results, while the OLS estimator is biased towards zero. For the sample of women, the IV estimate on HPS indicates that a woman will reduce her labor supply by 1,309 hours per year when she has a work-related health limitation, while she will reduce her labor supply by 1,052 and 1,278 hours per year by the OLS and traditional IV estimations, respectively.

When the work-limiting health is examined, the traditional IV estimate and the IV estimate on HPS yield close results. There are three possible explanations for it. The first possibility involves a hypothesis that the sample has endogeneity rather than measurement error. If it is true, a traditional IV technique will produce a consistent estimate compared to the biased OLS estimate. The IV estimate on HPS, which conducts an IV strategy for a fraction of observations, will also produce a similar consistent result except a larger standard error. The second possible explanation would be that there exists measurement error instead of endogeneity and that the potential instrument is valid. Since the valid instrument is uncorrelated with the measurement error, it handles the misclassification of work-limiting health, leading to the same pattern of results as the first explanation. In addition, a special data distribution can also explain this phenomenon; the difference between the true health and the misreported health may be symmetrically distributed around the mean of the index but with different signs. Then the expectation of the difference conditional on small index values will offset the expectation given large index values. As a result, the overall expectation of the difference between the true and the misreported health disappears. Future research needs to investigate the reason behind the close results between the traditional IV estimate and the proposed IV estimate on HPS. In addition, the estimated effect of work-limiting health is different from the effect of self-reported health, providing new empirical evidence that economic outcomes resulting from health declines are sensitive to the measures used, which is consistent with the review of Currie and Madrian (1999).

3.5.3 Retirement Age Group: Aged 62-70

After examining the working age population between 45 and 61 years old, I study individuals aged 62-70 as they may be considering retirement. With the financial incentives provided by the Social Security retirement benefits (starting from 62) and Medicare (starting from 65), individuals over 62 are more likely than those below 62 to choose retirement over staying in the market. While there may be more than one factor that influences one's labor supply decision at retirement age, it is interesting to study the role of health status.

Table 3.8 describes the characteristics of men and women in this age group. Compared to the working age group, an increased percentage of men and a slightly decreased percentage of women are married. Different from the similar distribution of educational attainment between men and women aged 45-61, older male respondents are 8.4 percentage points more likely to receive a bachelor's degree or higher than older female respondents, since more women (31%) than men (24%) stop schooling after a high school diploma. Above 62 years old, both men and women significantly reduce their labor force participation rate and hours of work. Only 42 percent of men and 34 percent of women in this age group are active in the labor market. For those who are working, the working time devoted to the main and second jobs also drops, except that elderly females work more hours on the second job than their counterparts aged 45-61, maybe a result from selection. Panel C indicates more individuals at retirement age are diagnosed with medical conditions than those at working age. While the older group is more likely to report work-limiting health problems than the younger group, the proportion of older individuals who rate their health as "Fair" or "Poor" is similar to the proportion of younger individuals reporting this status, regardless of gender. These two health measures may capture different dimensions of the true health status, or respondents may have different motivations when reporting these two measures. As such, it is necessary to study these two health measures in the labor supply equation.

Table 3.9 presents the estimation results of men and women aged 62-70. Compared to the sample aged 45-61, there is smaller reduction of hours worked when individuals experience health declines, except the sample of women when the work-limiting health is investigated. The results are reasonably expected because at retirement age both men and women decide to reduce hours of work or exit the labor force even if they are in good health. But the sample selection issue may be severe, as only 42 percent of males and 34 percent of females are economically active in the labor market. To test the existence of the sample selection issue and to accomodate it in the proposed IV technique on HPS, requires further theoretical developments.

3.6 Conclusions

This paper uses the method proposed in Chapter 2 to address the misclassification of health measures in studying the labor supply effects of health shocks. With objective health indicators providing exclusion restrictions for subjective health measures, I estimate the health index semiparametrically. Based on the high probability set where the index values are sufficiently large or sufficiently small, I implement the proposed IV estimator on this set using a sample from the 2012 wave of the Health and Retirement Study (HRS). The results suggest that both men and women aged 45-61 will reduce their labor supply substantially and that OLS and conventional IV methods demonstrate considerable attentuation biases when the self-reported health status is studied. When examining the measure of work-limiting health, the proposed technique and the conventional IV estimation yield similar results, while the OLS estimate is still biased towards zero. The greatly reduced work hours indicate that health problems occuring at middle age severely lower individuals' wage earnings and even force them to exit the labor market, which aligns with the literature on the relation between health and retirement. As individuals age, a larger proportion of men and women work zero hours in the labor market, leaving the sample selection bias a potential issue. Meanwhile, it is important to assess the financial status of the aging population through studying their labor market outcomes. Therefore, addressing the selection issue is a topic for future research.

Table 3.1: Definitions of variables in the structural model and exclusion variables

Variable	Definition
Structural Model: Y_i	
Hours worked	Total hours per year respondent works at the main and 2nd jobs.
Structural Model: X_i and H_i	
Age	Age of respondent.
Age squared	Square of age.
Nonlabor income	Total household nonlabor income, excluding wages and salaries earned by respondent.
Race	= 1 if white; = 0 otherwise.
GED	= 1 if having GED and 12 or fewer years of education; = 0 otherwise.
High school	= 1 if having high school diploma and 12 or fewer years of education; = 0 otherwise.
Some college	= 1 if having high school diploma or GED and 13 or more years of education, but less than bachelor degree; = 0 otherwise.
College above	= 1 if having college degree or greater; = 0 otherwise.
Married	= 1 if married (spouse present), married (spouse absent) or partnered; = 0 otherwise.
New England	= 1 if census division of residence is New England; = 0 otherwise.
Mid Atlantic	= 1 if census division of residence is Mid Atlantic; = 0 otherwise.
EN Central	= 1 if census division of residence is East North Central; = 0 otherwise.
WN Central	= 1 if census division of residence is West North Central; = 0 otherwise.
S Atlantic	= 1 if census division of residence is South Atlantic; = 0 otherwise.
ES Central	= 1 if census division of residence is East South Central; = 0 otherwise.
WS Central	= 1 if census division of residence is West South Central; = 0 otherwise.
Mountain	= 1 if census division of residence is Mountain; = 0 otherwise.
Wl hlth	= 1 if respondent has a health problem that limits the kind or amount of paid work; = 0 otherwise.
Sr hlth	= 1 if respondent reports “Poor” or “Fair” general health status; = 0 if respondent reports “Excellent”, “Very Good” or “Good” general health status.

Table 3.1 Continued

Variable	Definition
Exclusion Variables: Z_i	
Functional Limt #	Number of limitations on daily life activities, including “walking several blocks”, “sitting for about 2 hours”, “getting up from a chair after sitting for long periods”, “climbing several flights of stairs without resting, stooping/kneeling/crouching”, “lifting or carrying weights over 10 lbs”, “reaching arms above shoulder level” and “pushing or pulling large objects”.
Hibp	= 1 if respondent has been diagnosed with “high blood pressure or hypertension”; = 0 otherwise.
Diab	= 1 if respondent has been diagnosed with “diabetes or high blood sugar”; = 0 otherwise.
Cancr	= 1 if respondent has been diagnosed with “cancer or a malignant tumor of any kind except skin cancer”; = 0 otherwise.
Lung	= 1 if respondent has been diagnosed with “chronic lung disease except asthma such as chronic bronchitis or emphysema”; = 0 otherwise.
Heart	= 1 if respondent has been diagnosed with “heart attack, coronary heart disease, angina, congestive heart failure, or other heart problems”; = 0 otherwise.
Strok	= 1 if respondent has been diagnosed with “stroke or transient ischemic attack (TIA)”; = 0 otherwise.
Psych	= 1 if respondent has been diagnosed with “emotional, nervous, or psychiatric problems”; = 0 otherwise.

Table 3.2: Descriptive Statistics

Variable	Men	Women
Panel A: Basic Demographics		
Age	56.2 (3.3)	55.5 (3.7)
45-50 (%)	4.6	9.6
51-55 (%)	37.6	37.4
56-61 (%)	57.8	43.0
Non-white (%)	37.7	39.5
Nonlabor Income	45217.8 (85350.1)	49195.6 (93320.1)
Married (%)	76.1	65.7
Educational Attainment (%)		
Less Than High School	15.6	14.4
GED	6.0	5.8
High School	24.8	24.7
Some College	28.8	30.0
College Above	24.8	25.1
N	2877	3958
Panel B: Labor Supply		
Hours Worked	1566.9 (1183.4)	1177.6 (1075.5)
Working Group (%)	70.8	62.3
Working On 2nd Job/Working Group(%)	14.7	11.2
Hours On 1st Job Among 1st Job Workers	2141.1 (698.0)	1845.3 (685.6)
Hours On 2nd Job Among 2nd Job Workers	532.7 (518.2)	474.3 (486.8)
Panel C: Health Measures		
Wl hlth (%)	24.6	26.7
Sr hlth (%)	24.1	27.5
Functional Limt #	1.7 (2.4)	2.3 (2.6)
Hibp (%)	49.6	47.5
Diab (%)	19.2	18.5
Cancr (%)	5.1	9.0
Lung (%)	5.7	9.0
Heart (%)	14.2	12.3
Strok (%)	3.9	3.9
Psych (%)	14.4	23.9

Table 3.3: Discrepancy Between Two Subjective Health Measures

Men			
	Self-reported = 0	Self-reported = 1	Total
Work-limiting = 0	1879	289	2168
Work-limiting = 1	305	404	709
Total	2184	693	2877

Women			
	Self-reported = 0	Self-reported = 1	Total
Work-limiting = 0	2475	426	2901
Work-limiting = 1	394	663	1057
Total	2869	1089	3958

Table 3.4: Labor Supply Effect of Self-reported Health Status for Men Aged 45-61

	OLS	IV	IV on HPS
Health	-725.4	-2084.5	-2299.3
	(48.9)	(115.6)	(302.3)
Age	462.7	483.4	454.8
	(145.4)	(247.5)	(139.2)
Age Square	-4.5	-4.6	-4.4
	(1.3)	(2.2)	(1.3)
Nonlabor Income	0.0	-0.0	-0.0
	(0.0)	(0.0)	(0.0)
White	220.0	149.8	124.8
	(43.5)	(50.5)	(42.4)
Educational Attainment			
GED	-79.0	-244.2	-253.5
	(96.4)	(111.5)	(94.5)
High School	186.0	-99.3	-164.5
	(66.7)	(79.9)	(68.1)
Some College	257.0	-44.2	-132.6
	(65.7)	(80.1)	(68.3)
College & Above	428.8	37.5	-27.6
	(69.7)	(86.5)	(74.2)
Married	424.4	353.5	327.5
	(48.3)	(57.6)	(47.6)
N	2877	2877	2877

Figures in parentheses are standard errors.
Other covariates in every regression model include eight geographical dummy variables, indicating New England, Mid Atlantic, East North Central, West North Central, South Atlantic, East South Central, West South Central, and Mountain.

Table 3.5: Labor Supply Effect of Self-reported Health Status for Women Aged 45-61

	OLS	IV	IV on HPS
Health	-602.4	-1543.8	-1928.6
	(37.8)	(77.2)	(265.3)
Age	530.9	469.6	551.1
	(102.8)	(139.2)	(100.1)
Age Square	-5.1	-4.5	-5.2
	(0.9)	(1.3)	(0.9)
Nonlabor Income	-0.0	-0.0	-0.0
	(0.0)	(0.0)	(0.0)
White	123.2	104.6	58.3
	(34.3)	(38.0)	(33.9)
Educational Attainment			
GED	259.6	120.7	68.2
	(78.5)	(86.6)	(77.6)
High School	371.3	140.8	116.9
	(54.5)	(61.6)	(56.2)
Some College	472.5	199.0	168.6
	(53.3)	(61.6)	(55.8)
College & Above	595.5	242.8	225.2
	(56.8)	(67.7)	(61.1)
Married	-13.5	-69.0	-59.5
	(36.6)	(43.1)	(37.5)
N	3958	3958	3958

Figures in parentheses are standard errors.

Other covariates in every regression model include eight geographical dummy variables, indicating New England, Mid Atlantic, East North Central, West North Central, South Atlantic, East South Central, West South Central, and Mountain.

Table 3.6: Labor Supply Effect of Work-limiting Health for Men Aged 45-61

	OLS	IV	IV on HPS
Health	-1299.8	-1666.1	-1712.8
	(43.5)	(67.9)	(140.9)
Age	406.9	498.8	411.8
	(131.8)	(201.4)	(140.5)
Age Square	-3.9	-4.7	-4.0
	(1.2)	(1.8)	(1.3)
Nonlabor Income	0.0	0.0	0.0
	(0.0)	(0.000)	(0.0)
White	214.4	206.4	209.7
	(39.3)	(40.7)	(41.3)
Educational Attainment			
GED	107.5	141.4	167.5
	(87.3)	(90.3)	(93.8)
High School	267.0	237.0	254.9
	(59.8)	(61.6)	(62.5)
Some College	323.7	317.9	282.4
	(58.6)	(60.9)	(61.4)
College & Above	391.0	335.5	336.0
	(62.3)	(65.6)	(68.5)
Married	286.0	243.5	243.0
	(44.1)	(47.4)	(48.2)
N	2877	2877	2877

Figures in parentheses are standard errors.

Other covariates in every regression model include eight geographical dummy variables, indicating New England, Mid Atlantic, East North Central, West North Central, South Atlantic, East South Central, West South Central, and Mountain.

Table 3.7: Labor Supply Effect of Work-limiting Health for Women Aged 45-61

	OLS	IV	IV on HPS
Health	-1052.3	-1277.6	-1309.0
	(34.4)	(53.0)	(243.9)
Age	554.5	532.4	559.6
	(95.4)	(120.7)	(100.7)
Age Square	-5.2	-5.0	-5.3
	(0.9)	(1.1)	(0.9)
Nonlabor Income	-0.0	-0.0	-0.0
	(0.0)	(0.0)	(0.0)
White	139.6	153.1	115.0
	(31.8)	(32.7)	(33.4)
Educational Attainment			
GED	348.6	360.8	319.2
	(72.6)	(74.2)	(76.2)
High School	380.7	347.2	349.0
	(50.0)	(51.4)	(52.9)
Some College	523.7	499.3	512.5
	(48.4)	(49.9)	(50.7)
College & Above	584.7	534.7	541.8
	(51.5)	(54.3)	(54.9)
Married	-72.0	-99.4	-87.5
	(34.0)	(37.5)	(35.9)
N	3958	3958	3958

Figures in parentheses are standard errors.
Other covariates in every regression model include eight geographical dummy variables, indicating New England, Mid Atlantic, East North Central, West North Central, South Atlantic, East South Central, West South Central, and Mountain.

Table 3.8: Descriptive Statistics: Aged 62-70

Variable	Men	Women
Panel A: Basic Demographics		
Age	65.6	65.7
	(2.7)	(2.7)
Non-white (%)	24.0	25.4
Nonlabor Income	58700.8	45736.5
	(110775.5)	(87506.8)
Married (%)	82.0	63.2
Educational Attainment (%)		
Less Than High School	13.9	16.1
GED	4.8	4.9
High School	24.0	30.5
Some College	26.6	26.2
College Above	30.8	22.4
N	1631	2267
Panel B: Labor Supply		
Hours Worked	770.9	548.2
	(1065.4)	(898.5)
Working Group (%)	42.4	33.8
Working On 2nd Job/Working Group(%)	10.4	9.9
Hours On 1st Job Among 1st Job Workers	1774.4	1574.5
	(851.3)	(773.2)
Hours On 2nd Job Among 2nd Job Workers	482.4	522.8
	(463.1)	(618.3)
Panel C: Health Measures		
Wl hlth (%)	31.1	34.9
Sr hlth (%)	24.3	24.6
Functional Limt #	1.9	2.7
	(2.4)	(2.6)
Hibp (%)	64.1	61.6
Diab (%)	27.4	24.9
Cancr (%)	14.0	14.5
Lung (%)	8.8	11.3
Heart (%)	25.5	20.3
Strok (%)	6.7	6.1
Psych (%)	16.9	25.2

Table 3.9: Labor Supply Effect of Health Shocks for Men and Women Aged 62-70

	OLS	IV	IV on HPS	OLS	IV	IV on HPS
Men						
Wl hlth	-650.2 (53.3)	-991.4 (95.0)	-1079.0 (172.6)	–	–	–
Sr hlth	–	–	–	-407.1 (59.9)	-1120.3 (123.6)	-986.7 (190.4)
Women						
Wl hlth	-531.2 (37.2)	-649.2 (64.3)	-1886.2 (246.7)	–	–	–
Sr hlth	–	–	–	-358.8 (43.8)	-770.0 (84.8)	-1242.8 (286.0)

Figures in parentheses are standard errors.
Other covariates in every regression model include age, age squared, nonlabor income, race, educational attainment, married status and eight geographical dummy variables, indicating New England, Mid Atlantic, East North Central, West North Central, South Atlantic, East South Central, West South Central, and Mountain.

Chapter 4

Health Shocks and Household Labor Supply: Instantaneous and Adaptive Behavior of an Aging Workforce

4.1 Introduction

As individuals age, the risk of health problems increases. Johnson et al. (2005) found that for a sample of individuals who were aged 51 to 61 in 1992, 41 percent experienced some type of major medical problem between 1992 and 2002, and 34 percent had health problems that limited work. Such critical health events can negatively impact the economic circumstances of households. An individual's health decline may affect not only her own labor supply but also her spouse's labor supply. The knowledge of how individuals respond in labor supply to their partners' health conditions helps to assess the economic effect of poor health on households and therefore could inform the development of policies to insure households from economic loss due to illness.

However, the predicted effect of health shocks on spouse's labor supply is theoretically ambiguous. On one hand, poor health often reduces productivity in household production and requires time spent in care giving by the spouse. In this case, the spouse remains the primary home producer and caregiver, which pulls her away from the market. On the other hand, it has been well established that health declines reduce own labor supply, resulting in an income loss. To compensate for this loss, the spouse may need to earn more money and so increase her labor supply. Additionally, the increase of health care expenditures also promotes greater labor supplied to the market. Combining the pull imposed on spouses away from the market and the incentive of devoting more time to the market, the theoretical ambiguity makes empirical analysis crucial.

However, the empirical modeling of own and spousal health status on labor supply is complicated by measurement error and shared characteristics across spouses. As with a large body of previous work, this study uses self-reported work-limiting health conditions and self-rated health status as indicators of health. If respondents systematically report inaccurate health status to justify their labor force withdrawal, the models will suffer from endogeneity bias. In addition, if there exist shared characteristics across spouses, this will also bias the estimation. To address these problems, I use an instrumental variable strategy.

I use the Health and Retirement Study (HRS), a longitudinal dataset ranging from 1992 to 2012, to consider the impact of own and spousal health status on labor supply decisions. I include fixed effects into the model to control for the unobserved fixed characteristics. I also extend the analysis to consider how a spouse adjusts her labor supply over time in response to her partner's health shock. The research to date has focused on the instantaneous effect on an individual's labor supply soon after negative health shock to her spouse. I argue that this instantaneous effect captures only a part of the full picture. While some individuals respond immediately to their spouses' health shocks, those who are bound by labor contracts or who face costs caused by sudden labor supply changes probably postpone their responses. For those who would like to work more in the market after their spouses become ill, it may take some time to match up with vacant positions, and this too would not be reflected in the estimated instantaneous effect. The case will be more complex if an individual changes her mind over time after health problems initially affect her partner; for instance, an individual immediately reduces her labor supply due to the acute condition of her spouse, but then decides to raise her labor supply some years later when her spouse's health is expected to be stable or in recovery. In contrast, an individual probably maintains her labor supply at the beginning of her spouse's negative health spell, but decreases the work hours in the market as her spouse's health deteriorates. Few previous studies address these complicated situations. I will call this over time adjustment "adaptation" and develop an empirical model to test it by constructing variables in which the health indicators interact with the elapsed time since the first report of a work-limiting health condition.

This paper is organized as follows: the next section reviews the literature; Section 4.3 introduces and describes the sample used in this study; Section 4.4 explains methodology; Section 4.5 presents results; and Section 4.6 draws conclusions.

4.2 Literature Review

Much previous research has examined the effect of an individual's health status on her spouse's labor supply, but the results are mixed. Parsons (1976) uses the Productive Americans Survey (PAS) to examine the effect of work-limiting health conditions on own and spouse's time allocation. The time allocation contains productive hours in the labor market as well as in the home, each recorded in the PAS for husband and wife. For each spouse, market work hours and home work hours are regressed on own health status as well as spousal health status. Not surprisingly, in response to own health problems both spouses substantially reduce work hours in the market but maintain their home work hours. It is noteworthy that husbands and wives respond differently to spousal health shocks. A husband whose wife suffers a health related work limitation increases his work hours in the home rather than in the market, while a wife increases her work hours in the market rather than in the home when a health problem strikes her husband.

Levy (2002) examines the effect of a new diagnosis (cancer, diabetes, heart attack, chronic lung disease and stroke) on various economic outcomes of households and individuals, including labor supply. She uses the first four waves of the Health and Retirement Study (HRS) and separately examines insured and uninsured households. Levy estimates fixed-effect models and includes as explanatory variables interactions of insurance status both with demographic characteristics and the dummy for a new diagnosis. The results suggest no evidence of an added worker effect; individual labor supply is not affected by spousal health shocks. But Levy neglects the gender difference by using the full sample of combining males and females, with the effect on husbands potentially offset by the effect on wives.

Siegel (2006) uses the first wave (1992) of the Health and Retirement Study (HRS) to estimate the effect of the deterioration of husband's health on his wife's labor force participation and hours of work. There are two subjective health measures, the self-rated health status and work limitations, and two objective measures, the number of diagnoses of health conditions and the number of limitations of daily living activities and instrumental activities. The regressions suggest the dependence of the estimates on the health measures used; the more limitations of daily living activities and instrumental activities reported by the husband lead his wife to be more likely to work, while a woman whose husband reports his health as very good or good works fewer hours than a woman whose husband reports excellent health.

Parsons (1976), Levy (2002) and Siegel (2006) do not consider the bias of estimates caused by the systematic misreporting of health status to justify reduced labor supply. In contrast, Charles (1999) focuses on the endogeneity incurred by measurement error in the categorical health indicators, the self-reported work-limiting health condition and self-rated health status, when he determines how individuals react in labor supply to their spouses' health shocks. To account for measurement error bias, measures of functional limitations are used as instruments in a 2SLS regression. He also includes a fixed effect which he argues is correlated with own health and spouse's labor supply. He estimates this fixed-effect IV model using the first two waves of the Health and Retirement Study (HRS). The results indicate that both husbands and wives adjust labor supply significantly when their partners suffer health shocks but in different directions: husbands lower their work hours in the market when their wives' health declines, while wives work more in the face of their husbands' illnesses.

However, Charles' work is subject to criticism. As he explains, his fixed-effect IV model is estimated by 2SLS in which he regresses the endogenous health variables on the instruments, using the probit model in the first stage, and subsequently he replaces the endogenous health variables in the structural model in the second stage by the prediction from the first stage. It should be noted that this two-stage estimator is not an IV estimator. 2SLS can be very problematic if the first stage regression is not linear, like the probit model. So this two-stage estimator used by Charles is very sensitive to the probit model being exactly right and does not have the robustness properties of IV.

This study will improve upon Charles' model by proposing an optimal IV estimator. The optimal instrument strategy has the robustness and consistency properties of IV, which I will discuss in more detail below.

The above work all focuses on the examination of the instantaneous effect. The literature on how an individual adjusts over time in labor supply to their spouses' health declines is somewhat limited. Berger and Fleisher (1984) argue that an estimate using cross-sectional data that compares the response of wives whose husbands are healthy with those whose husbands have health problems will not reveal wives' response to a decline of their husbands' health over time. Past research using the first wave (1966) of the National Longitudinal Survey (NLS) found having a husband with a work-limiting health condition increased the probability a woman worked in the market and the number of hours she worked. However, Berger and Fleisher find that for wives whose husbands were reported to be healthy in 1966, the onset of a new spousal health problem of their husbands between 1966 to 1970 reduced their weeks worked. In contrast, wives whose husbands remained healthy during this period increased their work time engaged in the market. To explain the wives' adjustment over time, they regress the wives' weeks worked in 1970 on wives' weeks worked in 1966, husbands' work-limiting health problems during the period ranging from 1966 to 1970, and other controls. Among the control variables is the attractiveness of transfer income, which is the ratio of monthly public assistance in 1969 to the husband's hourly wage in 1966. Berger and Fleisher believe that the transfer income replaces the earnings loss related to husbands' health declines and therefore impacts wives' response to these declines. To capture the impact by the transfer income, a variable is constructed by interacting the husbands' work-limiting health problems with the attractiveness of the transfer income. As argued, the results suggest that a husband's negative health shock indeed impacts his wife's labor supply, but this impact depends on transfer payments: as the transfer payments are more attractive, wives decrease the work time devoted to the market. Even though the work by Berger and Fleisher is illuminating, they only examine the adjustment in labor supply of 1970 in response to the occurrence of health problems from 1966 to 1970, without tracing the history of labor supply since health problems were initially reported. I will measure the duration of an individual's health problem and study how the spouse adjusts her labor supply over this duration.

4.3 Data

I use the Health and Retirement Study (HRS) to analyze how an individual responds to her spouse's health shock. The HRS is a national panel survey consisting of households which have been interviewed biennially since 1992. It collects information spanning from 1992 to 2012 on employment history, health status, income, asset and demographics for individuals and their spouses. It includes six cohorts: the HRS cohort was first interviewed in 1992, the AHEAD cohort in 1993, the Children of Depression (CODA) cohort in 1998, the War Baby (WB) cohort also in 1998, the Early Baby Boomer (EBB) cohort in 2004 and the Mid Baby Boomer (MBB) cohort in 2010. The original HRS contains 37,319 respondents, each respondent being recorded biennially since her cohort entered the HRS.

This study is aimed at examining an individual's work decision when her partner suffers a health shock, so I focus on couples for whom there exists a long enough history to identify the impact of a health shock. I restrict my sample to observations from 1996 to 2012, eliminating observations from the first and second waves because some crucial variables in relation to health status, for example the functional limitations, are different for the first two waves compared to the subsequent nine waves. I restrict the sample to married couples[1] and eliminate single respondents and households in which there is a remarriage. Moreover, I only include waves in which both spouses are 45-70 years old and therefore old enough to become more vulnerable to health declines and likely to still be in the labor force. Additionally, I exclude waves with missing data on variables of interest, such as age, race, educational attainment, census region, household income, individual earnings, health measures and employment history. The above restrictions

[1] If either spouse in a household dies in a given wave, the subsequent observations for this household will be excluded from this sample. If one spouse reports herself to be divorced in a given wave, the corresponding observation for this household in this wave is excluded.

result in an unbalanced panel[2] of 5,823 couples with 21,779 observations for each spouse.

The key variables used in this study are described as follows:

The measure of labor supply is the hours worked per year. It is the sum of work hours in the major job as well as in the second job, each computed as the product of "hours of work per week" and "weeks worked per year." Individuals not working for pay are treated as having zero hours of work.

There are four measures of health shocks, two commonly considered to be more subjective and the other two more objective. The two more subjective health measures are work-limiting health conditions and self-rated health status. The variable of work-limiting health conditions has the value of 1 if respondents report a health problem that limits their work, and 0 otherwise. The self-rated health status reports the general health status for husbands and wives, categorized as "excellent", "very good", "good", "fair" and "poor". I recode this variable to a dummy variable: it equals 1 if respondents report their general health status as "fair" or "poor", and 0 if they report as "excellent", "very good" or "good". The number of individuals reporting "excellent" or "poor" in the original variable make up only a small proportion of the whole sample.[3] The two more objective health measures are functional limitations and doctor-diagnosed health problems. Functional limitations contain a set of variables recording data about difficulty with physical activities, like walking several blocks, climbing several flights of stairs without resting, and so on. Each variable equals 1 if respondents report some difficulty and 0 if not. Doctor-diagnosed health problems include high blood pressure, diabetes and the like. Respondents are asked if they have ever been diagnosed with any condition when they are first interviewed, and in subsequent waves, respondents have the opportunity to review the prior statement and report again if they have been diagnosed with any condition since the previous interview. This update is reflected in these variables. In each wave, if respondents have been diagnosed with any condition, the corresponding variable has the value of 1; otherwise, it equals 0.

[2]The sample excludes households with only one wave for the sake of a longitudinal dataset.

[3]There are not enough instrumental variables to instrument each category in the original variable, which I will discuss in more detail in methodology.

The duration of an individual's health shock is defined as the number of waves since the first report of her work-limiting health condition. Because the duration is unclear if an individual reported a work-limiting health condition in her first interview, this variable is created only for individuals who did not report any work-limiting health condition in their first interview.

I use "total household income" and "individual earnings" to derive household non-labor income. These two economic variables record information in the calendar year prior to the survey. The difference by subtracting both spouses' individual earnings from total household income is the household non-labor income. I convert household non-labor income to year 2000 dollars using the CPI-U.

The HRS includes other demographic variables, such as the data on age, race, educational attainment and census region for husbands and wives in each wave.

4.3.1 Descriptive Statistics

Table 4.1 presents basic descriptive findings. In this sample, about 88 percent of husbands and wives are white and there is a very small number of inter-racial marriages. The distribution of educational attainment indicates a gender difference in education for the sample cohort: husbands are 1.9 percentage points more likely to have less than high school degrees than wives, but 33 percent of husbands have college degrees and above compared to only 28 percent of wives. Twenty-six percent of husbands and thirty-one percent of wives have high school degrees. The distribution of husbands' educational attainment is comparatively even, while the distribution of wives' concentrates on the high school degree. Because my sample is unbalanced and households may enter or exit this sample in different waves, I describe variables that change across waves for two waves, the first wave when a household enters and the last wave up to which a household stays. The average age of husbands in the last wave is 62 and the average age of wives is just below 60. The average marital duration is more than 30 years in the last wave, implying a high commitment to marriage.

Table 4.1 reports descriptive statistics for two main health measures, work-limiting health conditions and self-rated health status. These two measures display health deterioration of husbands and wives with time. For example, in the first wave 17 percent of husbands report having work-limiting health conditions and by the last wave this percent rises to 24. Wives experience a similar deterioration of health. Self-rated health status also shows health declines for husbands and wives: in the first wave, 17 percent (16 percent) of husbands (wives) rate their health status as fair or poor rather than good, very good or excellent, and in the last wave 21 percent (19 percent) of husbands (wives) believe they have fair or poor health. In addition, a comparison of hours of work per year between the first and last waves indicates that both husbands and wives reduce their labor supply over the period of study.

4.4 Methodology

4.4.1 Basic Model

To examine how individuals react in labor supply to spouses' health shocks, I propose a model as follows:

$$L_{it} = \beta_0 + X'_{it}\beta + H^r_{it}\gamma_r + H^s_{it}\gamma_s + \alpha_i + \theta_t + \epsilon_{it} \tag{4.1}$$

where the dependent variable L_{it} is work hours of individual i at time t; H^r_{it}, H^s_{it} are own and spouse's health status, respectively, either work-limiting health conditions or self-rated health status; X_{it} are variables capturing determinants of labor supply other than health status, including own and spouse's ages and square of ages, own race and educational attainment, census region and household non-labor income; α_i and θ_t are the individual effects and time fixed effects, and ϵ_{it} is a normally-distributed error with zero mean. This equation includes the time fixed effects, θ_t, to take into account time period specific effects on labor supply such as the general economic climate.

The existence of measurement error complicates the empirical estimates of own and spousal health status on labor supply. The two health measures used in this model are both self-reported health and therefore more likely to be endogenous to work decisions. Bound (1991) casts doubt on the comparability of self-reported health across respondents because different people probably rate the same condition differently. He also argues that working age individuals prefer to report worse health to rationalize their reduction in labor supply. For example, the questionnaire wording of work-limiting health problems is "Now I want to ask how your health affects paid work activities. Do you have any impairment or health problem that limits the kind or amount of paid work you can do?" This questionnaire wording hints at the relationship between health problems and labor supply, and probably leads a respondent to misreport own or spouse's health status to justify his reduction of labor supply. To address this problem, I use two sets of more objective health measures simultaneously, functional limitations and doctor-diagnosed health problems, to instrument for work-limiting health conditions and self-rated health status.[4] Functional limitations and doctor-diagnoses reveal health status: if a respondent has a limitation to function in some physical activity in daily life or he has a diagnosis, it is highly possible that he reports his health as poor. Different from the two subjective health measures, each variable in functional limitations and doctors' diagnoses involves some specific physical limitation or diagnosis, which is very "narrow" and thus less subject to misreporting. And the questionnaire wording for these variables does not hint at the relationship between health problems and labor supply.

The self-reported health variables are regressed on all exogenous variables in equation (4.1) and functional limitations as well as doctor-diagnosed health problems as excluded restrictions:

$$H_{it}^{j} = I\{\pi_0 + X_{it}'\pi_1 + Z_{it}^{j} \geq \mu_{it}\},\ j = r, s \tag{4.2}$$

where H_{it}^{r}, H_{it}^{s} are individual i's and spouse's health status at time t, respectively; X_{it} include the same exogenous variables as in equation (4.1); Z_{it}^{r}, Z_{it}^{s} are the instruments for individual i and her spouse, respectively, including their respective functional limitations and doctor diagnosed health problems; and μ_{it} is a normally-distributed error

[4]This IV assumption may not hold, since such instruments are probably correlated with the measurement error of the discrete health variables, which is illustrated in earlier chapters. An extension of Chapter 2 to cover multiple mismeasured regressors in panel data regressions would make it possible to test this IV assumption. If the test indicated that the IV assumption fails, then an extension of the analysis based on high probability set in Chapter 2 could be employed. In future research, I plan to consider this extension.

with zero mean. Since the two health measures of $H_{it}^j, j = r, s$ are dummy variables, I estimate the probit regression for (4.2).

Newey (1990) proposes optimal instruments to minimize the covariance matrix of an IV estimator. The optimal instrument of an endogenous variable is the expectation of this variable conditioned on all regressors in the function of this endogenous variable. The optimal instrument strategy has the robustness and consistency properties of IV, and the variance of the estimator in this strategy would be smaller than the variance of other IV estimators. There are three steps to process the optimal instrument strategy:

- In step 1, I get optimal instruments for endogenous health measures.

$$P_{it}^j = \mathbf{E}(H_{it}^j | X_{it}, Z_{it}^j) = Pr(H_{it}^j = 1 | X_{it}, Z_{it}^j), \; j = r, s \tag{4.3}$$

  The optimal instrument is $P_{it}^j, j = r, s$. I predict the probability, $\hat{P}_{it}^j$, from the probit model in (4.2). The prediction is the estimated optimal instrument.

- In step 2, I regress the endogenous variable on optimal instruments. It is the first stage of 2SLS:

$$H_{it}^j = \delta_0 + X_{it}'\delta_1 + \hat{P}_{it}^r\delta_2 + \hat{P}_{it}^s\delta_3 + \omega_{it}, \; j = r, s \tag{4.4}$$

  From this regression, I get the predicted indicators for health, $\hat{H}_{it}^j, j = r, s$.

- In step 3, the predicted indicators for own and spouse's health from the step 2 substitute for the endogenous health variables in the structural model, as the second stage of 2SLS.

In step 3, I run a fexed-effect regression instead of an OLS regression, since there likely exists another endogenous problem: unobserved household characteristics. Charles (1999) discusses it taking an example of “laziness.” He argues laziness is a kind of fixed effect shared by a husband and a wife, since assortative mating makes a man and a woman with common characteristics more likely to get married. In this case, for example, laziness may prevent a woman from exercising and therefore worsen her health. At the same time, laziness negatively affects her husband's labor supply. So a wife's health will be endogenous in the function of her husband's labor supply. Even if taking no account of the process of assortative mating, Charles further argues that similar lifestyles within a couple still affect health status and labor supply of both spouses. Then an individual's health is still endogenous in the function of spouse's labor supply. Since my sample includes a majority of couples with long marital duration in which spouses are likely to live similar lifestyles, I add fixed effect in the structural model. So in the above step 3, I fit a fixed-effect model.

To clarify the necessity of fixed effects, a Hausman test for fixed effects is used to test whether the fixed effects are correlated with the explanatory variables. The Hausman test compares a random-effect optimal IV (REIV) model that includes random effects in step 3 to a fixed-effect optimal IV (FEIV) model that includes fixed effects in step 3. If fixed effects are uncorrelated with the explanatory variables in the "true" specification, the REIV model is consistent and efficient while the FEIV model is consistent but inefficient. However, a standard Hausman test has a very strong assumption: the REIV estimator must be the fully efficient estimator, requiring consistent estimation of the variance of REIV estimator, which is hardly ensured by the above three steps, particularly with the nonlinear specification in the first step. A bootstrapped version of Hausman test can address this problem: both REIV and FEIV are conducted with 400 bootstrap replications, and eventually the bootstrapped estimator of the variance used in the Hausman test is calculated. As a result, the bootstrapped version of Hausman test rejects the null hypothesis that α_i is uncorrelated with own and spouse's health status in equation (4.1). Therefore, it is necessary to include fixed effects in the structural model (4.1). In addition, I compare the OLS regression and the optimal IV regression without fixed effects to the optimal IV regression with fixed effects. The OLS regression, as a benchmark in the comparison, ignores endogeneity resulting from measurement error and potential fixed effects. The optimal IV estimator without fixed effects implements similar three steps as the optimal IV regression with fixed effects, except running an OLS regression in step 3 without taking into account the unobserved household characteristics. The estimation difference between the optimal IV regressions with and without fixed effects will provide empirical evidence that it is necessary to consider fixed effects.

It is crucial to analyze the validity of the optimal instrument I use in 2SLS. A valid instrument should be highly correlated with the endogenous variable, but not affect directly the dependent variable in the second stage of 2SLS. I use a Kleibergen-Paap LM test for under-identification and the Kleibergen-Paap Wald test for weak identification, since these two tests are consistent in this study where the error term clusters by individual instead of being independently and identically distributed. Given that $Z_{it}^{j}, j = r, s$ are the excluded restrictions, I conduct a t test for the individual significance of functional limitations and doctor-diagnosed health problems and the Wald test for their joint significance in step 1. Because there are two optimal instruments for two endogenous health variables, Stata cannot conduct a Sargan-Hansen test for over-identification even though there are actually more than two excluded restrictions, $Z_{it}^{j}, j = r, s$. To test over-identification, I split the $Z_{it}^{j}, j = r, s$ to two sets: one is the set of functional limitations and the other is the set of doctor-diagnosed health problems. Then I estimate the optimal instrument using only the set of functional limitations in step 1, and perform the above three steps by including the set of doctor diagnoses into the second stage of 2SLS. If the excluded restrictions do not directly affect labor supply, the coefficients on doctor diagnoses will be close to zero or insignificantly different from zero. So after estimating this 2SLS, I test the coefficients of doctor diagnoses individually and jointly. I also conduct the similar test by estimating optimal instruments using only doctor diagnoses and then including functional limitations in the second stage of 2SLS.

4.4.2 Adaptation Model

The estimate in the basic model characterizes the instantaneous effect of an individual's health shock on spouse's labor supply, failing to consider how a spouse adjusts her labor supply over time in response to her partner's health shock. Individuals who are bound by labor contracts or who face costs caused by sudden labor supply change probably postpone their responses to their own and spouse's health shocks. In addition, it may take time to match up with vacant positions for those who would like to work more or fewer hours after their spouses become ill. Therefore, I include the duration of the health shock in the labor supply model. More importantly, I argue that the effect of an individual's health problem on spouse's labor supply depends on the duration of health shock. For example, this labor immediately responds to a spouse's health shock but then increases hours of work as the spouse's health improves. In contrast, an individual could maintain his labor supply at the beginning of his spouse's health condition, but decrease work hours in the market as his spouse's health deteriorates. In other words, the effect of an individual's health shock on spouse's labor supply may vary with duration of health shock.

To examine this adjustment in labor supply over time, I propose a model as follows:

$$L_{it} = \beta_0 + X'_{it}\beta + H^r_{it}\gamma_r + H^s_{it}\gamma_s + H^r_{it} * D^r_{it}\varphi_r + H^s_{it} * D^s_{it}\varphi_s + \alpha_i + \theta_t + \epsilon_{it} \qquad (4.5)$$

where D^r_{it}, D^s_{it} are duration of individual i's and spouse's health shocks at time t; H^r_{it}, H^s_{it} are own and spouse's health status as in the basic model, respectively, either work-limiting health conditions or self-rated health status; $H^r_{it} * D^r_{it}, H^s_{it} * D^s_{it}$ are interactions of own and spouse's health status with their respective duration of health shocks; other symbols have the same definitions as in equation (4.1). The coefficients γ_r, γ_s measure the instantaneous effect of own and spouse's health shocks on an individual's worked hours, and the coefficients φ_r, φ_s of interaction terms identify how an individual adjusts in labor supply over time in response to her own and spouse's health problems.

Similar to the discussion in basic model, endogeneity resulting from measurement errors and unobserved household effects occurs in this model. So H^r_{it}, H^s_{it} and interaction terms $H^r_{it} * D^r_{it}, H^s_{it} * D^s_{it}$ are endogenous in the function of individual i's labor supply. The existence of interaction terms makes the model nonlinear and therefore complicates the estimation of the optimal IV estimator. Newey (1990) addresses the optimal IV estimation of nonlinear models. Based on his study, I conduct the optimal instrument strategy as the following three steps:

- In step 1, it is the same procedure as in basic model to get the optimal instruments $\hat{P}^j_{it}$ for health measures by the probit model in (4.2). Optimal instruments for interaction terms $H^j_{it} * D^j_{it}, j = r, s$ are naturally obtained by interacting $\hat{P}^j_{it}$ with duration $D^j_{it}, \hat{P}^j_{it} * D^j_{it}, j = r, s$.

- In step 2, I regress the endogenous variables on optimal instruments as the first stage of 2SLS:

$$H^j_{it} = \delta_0 + X'_{it}\delta_1 + \hat{P}^r_{it}\delta_2 + \hat{P}^s_{it}\delta_3 + \hat{P}^r_{it} * D^r_{it}\delta_4 + \hat{P}^s_{it} * D^s_{it}\delta_5 + \omega^1_{it} \tag{4.6}$$

$$H^j_{it} * D^j_{it} = \delta^*_0 + X'_{it}\delta^*_1 + \hat{P}^r_{it}\delta^*_2 + \hat{P}^s_{it}\delta^*_3 + \hat{P}^r_{it} * D^r_{it}\delta^*_4 + \hat{P}^s_{it} * D^s_{it}\delta^*_5 + \omega^2_{it} \tag{4.7}$$

  By OLS regressions, I can get predicted indicators for health and interaction terms, $\hat{H}^j_{it}, \widehat{H^j_{it} * D^j_{it}}, j = r, s$.

- In step 3, I substitute these predicted indicators for endogenous variables and estimate a fixed-effect regression as the second stage of 2SLS.

I use a Kleibergen-Paap LM test for under-identification and the Kleibergen-Paap Wald test for weak identification. Similar to the basic model, the adaptation model here is still just-identified, prohibiting the use of over-identification tests. To test the extra excluded restrictions $Z^j_{it}, j = r, s$, I apply the same methods used in basic model; I estimate optimal instrument using only the set of functional limitations in step 1 and perform the above three steps by including the set of doctor diagnoses into the second stage of 2SLS. Then I test the coefficients of doctor diagnoses individually and jointly. I also conduct the analogous test by estimating optimal instruments using only doctor diagnoses and then including functional limitations in the second stage of 2SLS.

4.5 Results

4.5.1 Basic Model

Table 4.2 presents the estimation results for husband's and wife's hours of work in the basic model, which are estimated using optimal instrument strategy with household effects. Regardless of which measure of health status is used, neither husbands nor wives significantly change their hours of work in response to spousal health problems. Consistent with previous literature, individuals always reduce their labor supply significantly when they themselves suffer health shocks: a husband works 1051 fewer hours per year when he has a work-limiting health condition, holding other factors constant, while a wife works 574 fewer hours per year when she has a work-limiting health condition. A husband reduces his hours of work by 1273 hours per year if he rates himself with fair or poor health, holding other factors constant, while a wife reduces her hours worked by 823 hours per year under her fair or poor health.

The P-value of Kleibergen-Paap LM statistic (prob> u = .000) shows that the estimates do not suffer from under-identification and the optimal instrument strategy is feasible. The null hypothesis that the estimator is weakly identified is rejected, because the Kleibergen-Paap Wald statistic is far more than 10, which is proposed by Staiger and Stock (1997) as a "rule of thumb". The over-identification test shows that the instruments do not affect directly labor supply, because the null hypothesis that a subset of excluded restrictions are jointly uncorrelated with labor supply after the left restrictions explain endogenous regressors is not rejected. Note this hypothesis is not rejected at 1% significance for the model of husband's hours worked when health is measured by self-rated health status.

The results of the first-stage in 2SLS are shown in Table 4.3. The results of the first-stage suggest that optimal instrument is significantly correlated with the endogenous measure of own health.

Table 4.4 shows the results of the probit model in (4.2). I regress the health measure on excluded restrictions, including functional limitations and doctor-diagnosed health problems. Coefficients of almost all excluded restrictions are highly significant, which implies the excluded restrictions used to explain endogenous measure of health are highly correlated with endogenous health variables.

For comparison, I also estimate OLS regressions for the structural model in equation (4.1). Table 4.5 presents the results for OLS regressions. It shows that husbands and wives respond differently to their spouses' health shocks: a wife's work-limiting health condition leads her husband to work 56 fewer hours per year, while a husband's work-limiting health condition leads his wife to work 80 more hours per year. A husband whose wife rates her health status as fair or poor rather than good, very good or excellent reduces his hours of work by 58 hours per year, while a wife whose husband rates his health status as fair or poor rather than good or better raises her hours worked by 20 hours per year and this estimated coefficient is insignificant. However, the results for OLS regressions may be subject to the attenuation bias caused by measurement errors in self-reported health measures.

Table 4.6 presents the results for optimal IV regressions without household effects and provides evidence of the attenuation bias in OLS regressions. Different from the OLS estimation results, the results in Table 4.6 show that wives increase a large number of hours worked in response to their husbands' health shocks, while there is no substantial effect on husbands' labor supply of their wives' health problems. For example, a husband's work-limiting health condition leads his wife to work 193 more hours per year, holding other factors constant, and if he rates his health as fair or poor rather than good or better, his wife increases her working hours by 295 hours per year. In contrast, a husband whose wife has a work-limiting health problem reduces work hours by 44 hours per year, and a husband increases a small number of work hours if his wife rates her health as fair or poor compared to good, very good or excellent. In addition, the effect of wife's health shock on husband's working hours is not significant, regardless of health measures.

To make sure the existence of household effects, I take the regression of husbands' working hours on work-limiting health conditions as an example and perform a bootstrapped version of Hausman test. This Hausman test rejects the null hypothesis that α_i is uncorrelated with own and spouse's health status in equation (4.1). Therefore, it is necessary and to include fixed household effects in the structural model (4.1). Compared to the optimal IV with household effects (Table 4.2), the optimal IV without household effects (Table 4.6) is biased upward because of failure to include household effects.

4.5.2 Adaptation Model

Table 4.7 presents the results of the structural model (4.5) with interaction terms. Similar to the results for the regression of the instantaneous effect in Table 4.2, individuals do not respond immediately to their spouse's health shocks. However, a husband would adjust in his work hours over time in response to his wife's health problem when self-rated health status is used. The effect of his wife's health problem on his hours worked depends on the duration of the wife's health problem. Holding other factors constant, as the duration of wife's health problem increases by one wave (two years), he reduces working hours by 165 hours. Wives appear not to adjust their labor supply over time in response to their husbands' health problems. In addition, husbands adjust their labor supply over time in response to their own health problems while wives do not.

To analyze the validity of instruments, I perform an under-identification test, weak identification test and over-identification test of endogenous regressors. The results reject null hypotheses of under-identification and weak identification. The over-identification test shows exogeneity of instruments, because the null hypothesis that a subset of excluded variables are jointly uncorrelated with labor supply is not rejected.

Table 4.8 shows the results for the first-stage regressions of 2SLS. Because both husband's and wife's labor supply models have four endogenous variables, each of the four endogenous variables is estimated in the first stage. The upper panel in Table 4.8 shows the results when work-limiting health conditions are used and the lower panel shows the results when self-rated health problems are used. The optimal instrument for own health is strongly correlated with the endogenous measure of own health, and the optimal instrument for the interaction of own health with its duration is also strongly correlated with the endogenous interaction term.

4.6 Conclusions

I use the HRS to examine the instantaneous and adaptive effects of individuals' health shocks on their spouses' labor supply. There are two measures of health status, work-limiting health conditions and self-rated health status. Regardless of which measure is used, both husbands and wives immediately respond to their own health shocks by reducing own working hours, but not to their spouses' health shocks. However, I argue that the effect of an individual's health problem on spouse's labor supply depends on the duration of that health problem. This is what I call the adaptive effect. Even though both husbands and wives do not respond immediately to their spouses' health shocks, husbands work 165 fewer hours as the duration of their wives' health problems increases by one wave (two years). This adaptive behavior occurs only when self-rated health status is used, which implies that the examined effect of individuals' health shocks on their spouses' hours of work depends on the measure of health. This research finding reveals that the impact on time allocation is greater for wives' poor health than for husbands'. The wives' health declines will gradually pull their husbands away from the market. Such households, therefore, are at greater risk for financial hardship.

The results have some implications for further research. First, given that the results suggest the effect of a wife's health problem on her husband's hours of work depends on the duration of wife's health problem and that this dependence is a linear function of duration, additional work is needed to consider semiparametric estimates. Then the assumption of linear dependence can be dropped and the structural model adapted to contain a regressor which is an unknown function of health status and its duration. If the "true" specification of the structural model indeed contains a function of health problems and its duration but the function is not the product of these two variables, the estimated linear dependence is probably misleading and the semiparametric estimates are reliable. Second, the method of over-identification test in this study deserves more attention. Because the model is just-identified, I test exogeneity of instruments by splitting the excluded restrictions into the set of functional limitations and doctor diagnoses, putting one set in the probit model to get the estimated optimal instrument and the other in the second-stage regression of 2SLS, then testing the joint significance of the latter set. I perform it again by switching the two sets. However, if the "true" specification of the probit model contains these two sets simultaneously, the estimate of optimal instrument may be problematic by excluding either set. More theoretical and empirical advances are needed to address this problem. Finally, in view of discussions in Chapter 2, the validity of objective health measures as instruments requires further examination. An extension of the proposed technique in Chapter 2 may make it possible to test whether the instruments are correlated with the measurement error or not. If the test indicated that the assumption of zero correlation between the instruments and the measurement error does not hold, an extension of Chapter 2 to accommodate multiple mismeasured regressors in panel data regressions would be a promising alternative. In future research, I plan to consider this extension.

Table 4.1: Descriptive Statistics

Variable	Husband		Wife	
	First wave	Last wave	First wave	Last wave
White (%)	87.7	–	88.4	–
Age	55.6	62.2	52.9	59.7
	(4.8)	(5.8)	(4.8)	(5.9)
Educational attainment (%)				
Less than high school	11.5	–	9.6	–
GED	5.0	–	4.5	–
High school	25.6	–	30.5	–
Some college	25.0	–	27.5	–
College and above	32.9	–	27.9	–
Marriage duration (for couples)	25.6	32.2	–	–
	(10.9)	(12.2)		
Household nonlabor income	32870.8	42583.6	–	–
(for couples, dollars in 2000)	(145499.5)	(411849.5)		
Health measures (%)				
Work-limiting health conditions	16.8	24.3	17.4	23.9
Self-rated health status	16.9	20.9	16.3	19.1
Hours worked per year	1836.3	1201.4	1295.2	921.2
	(1173.5)	(1229.4)	(1083.2)	(1035.3)
Number of households	5823	–	–	–
Number of observations	21779	21779	21779	21779

Figures in parentheses are standard errors.

Table 4.2: Optimal IV Estimates of the Effect of Health Shocks on Hours Worked (with Fixed Effects)

Variable	Work-limiting health conditions		Self-rated health status	
	Husband's hours	Wife's hours	Husband's hours	Wife's hours
Own health	-1050.8***	-574.1***	-1272.5***	-822.7***
	(75.4)	(75.3)	(107.2)	(116.6)
Spouse's health	-50.4	22.1	55.0	10.1
	(81.9)	(67.4)	(125.8)	(86.5)
Own age	126.2***	145.0***	142.2***	177.6***
	(40.2)	(33.7)	(43.4)	(34.5)
Own age squared	-1.2***	-1.4***	-1.2***	-1.6***
	(.3)	(.3)	(.3)	(.3)
Spouse's age	.9	69.6*	29.2	59.7
	(36.2)	(36.5)	(38.8)	(37.5)
Spouse's age squared	-.2	-.5*	-.3	-.4
	(.3)	(.3)	(.3)	(.3)
Household nonlabor income	-.0	-.0	-.0	.0
	(.0)	(.0)	(.0)	(.0)
F	136.9	59.2	116.4	53.7
Prob>F	.000	.000	.000	.000
Under-id test u (K-P LM statistic)	407.9	408.0	262.8	263.4
Prob>u	.000	.000	.000	.000
Weak id test w (K-P Wald statistic)	328.6	328.9	182.0	182.5
Over-id test: functional limitations in (4.2), doctors' diagnoses in (4.1)				
Wald ω^2 (14)	14.0	14.3	25.3	13.4
Prob>ω^2	.450	.429	.032	.498
Over-id: doctors' diagnoses in (4.2), functional limitations in (4.1)				
Wald ω^{*2} (16)	10.2	17.5	23.8	19.6
Prob>ω^{*2}	.854	.355	.094	.238
N	21779	21779	21779	21779

Figures in parentheses are standard errors.

*Indicates significance at the 10 percent level. ** indicates significance at the 5 percent level. *** Indicates significance at the 1 percent level.

Regressions include demographic controls for own race, own educational attainment, census region and interview wave.

Table 4.3: First-stage Regression of Optimal IV with Household Effects

Variable	Work-limiting health conditions: Husband's health	Work-limiting health conditions: Wife's health	Self-rated health status: Husband's health	Self-rated health status: Wife's health
Husband's optimal instrument	.688***	.019	.735***	.010
	(.023)	(.017)	(.031)	(.023)
Wife's optimal instrument	.017	.603***	.061***	.596***
	(.017)	(.024)	(.026)	(.031)
Husband's age	.012	-.007	.025*	-.008
	(.013)	(.013)	(.013)	(.013)
Husband's age squared	-.000	.000	-.000	.000
	(.000)	(.000)	(.000)	(.000)
Wife's age	-.018	-.008	.012	.022**
	(.012)	(.011)	(.012)	(.011)
Wife's age squared	.000	.000	-.000	-.000
	(.000)	(.000)	(.000)	(.000)
Household non-labor income	.000	-.000	.000	.000
	(.000)	(.000)	(.000)	(.000)
F	48.2	31.2	30.5	18.1
Prob >F	.000	.000	.000	.000
N	21779	21779	21779	21779

Figures in parentheses are standard errors.
*Indicates significance at the 10 percent level. ** indicates significance at the 5 percent level. *** Indicates significance at the 1 percent level.
Regressions include demographic controls for own race, own educational attainment, census region and interview wave.

Table 4.4: Probit Estimates of Optimal Instruments for Endogenous Health Measures

	Work-limiting health conditions		Self-rated health status	
Excluded variables (Z)	Husband's health	Wife's health	Husband's health	Wife's health
Walking several blocks	.638*** (.040)	.712*** (.037)	.540*** (.041)	.513*** (.039)
Sitting for 2 hours	.255*** (.040)	.258*** (.037)	.093** (.040)	.147*** (.037)
Getting up from chair	.108*** (.033)	.094*** (.034)	.147*** (.034)	.062* (.036)
Climbing several flights of stairs	.343*** (.035)	.262*** (.033)	.340*** (.034)	.343*** (.033)
Stooping	.352*** (.033)	.260*** (.033)	.128*** (.032)	.073** (.036)
Lifting weights over 10 lbs	.484*** (.050)	.500*** (.039)	.169*** (.049)	.355*** (.039)
Reaching arms above shoulder level	.298*** (.044)	.220*** (.044)	.130*** (.042)	.230*** (.042)
Pushing large objects	.527*** (.044)	.536*** (.037)	.350*** (.044)	.292*** (.039)
High blood pressure	.096*** (.032)	.060* (.033)	.206*** (.031)	.163*** (.034)
Diabetes	.214*** (.039)	.056 (.044)	.432*** (.038)	.460*** (.045)
Cancer	.181*** (.053)	.144*** (.050)	.341*** (.055)	.210*** (.053)
Chronic lung disease except asthma	.384*** (.058)	.261*** (.055)	.454*** (.053)	.344*** (.057)
Heart attack	.327*** (.038)	.214*** (.045)	.388*** (.037)	.278*** (.046)
Stroke or TIA	.529*** (.077)	.437*** (.095)	.329*** (.070)	.299*** (.093)
Emotional, nervous or psychiatric problems	.429*** (.051)	.378*** (.039)	.385*** (.049)	.334*** (.040)
Wald ω^2 (15)	3112.4	3642.5	2563.5	2599.4
Prob> ω^2	.000	.000	.000	.000
N	21779	21779	21779	21779

Figures in parentheses are standard errors.

*Indicates significance at the 10 percent level. ** indicates significance at the 5 percent level. *** Indicates significance at the 1 percent level.

Regressions include demographic controls for own and spouse's age and square of age, household non-labor income, own race, own educational attainment, census region and interview wave.

Table 4.5: OLS Estimates of the Effect of Health Shocks on Hours Worked

Variable	Work-limiting health conditions		Self-rated health status	
	Husband's hours	Wife's hours	Husband's hours	Wife's hours
Own health	-944.1***	-712.5***	-548.9***	-390.5***
	(22.9)	(20.2)	(25.9)	(25.0)
Spouse's health	-56.0**	79.6***	-58.2**	20.4
	(23.5)	(21.8)	(26.6)	(22.4)
Own age	235.5***	145.1***	235.7***	147.7***
	(32.0)	(26.3)	(33.6)	(27.2)
Own age squared	-2.6***	-1.7***	-2.7***	-1.7***
	(.3)	(.2)	(.3)	(.2)
Spouse's age	23.1	27.7	20.2	38.1
	(28.0)	(31.7)	(29.3)	(33.3)
Spouse's age	-.3	-.3	-.3	-.4
squared	(.2)	(.3)	(.3)	(.3)
Household non-	.0	-.0	.0	-.0
labor income	(.0)	(.0)	(.0)	(.0)
Own race	84.7***	-71.0***	66.8**	-88.9***
	(29.6)	(27.6)	(32.8)	(29.6)
Own educational attainment				
Less than high school	–	–	–	–
GED	-59.3	318.3***	-145.0**	257.6***
	(54.2)	(54.9)	(59.2)	(59.9)
High school	109.5***	242.5***	84.2**	240.5***
	(34.8)	(32.8)	(38.3)	(35.3)
Some college	154.7***	354.8***	119.7***	346.1***
	(37.1)	(34.9)	(40.3)	(37.8)
College and above	224.7***	384.2***	222.7***	393.4***
	(36.6)	(36.9)	(40.0)	(39.5)
Constant	-3425.1***	-2082.5**	-3240.1***	-2426.6***
	(1022.6)	(969.1)	(1073.5)	(1013.5)
F	288.8	167.7	200.0	107.7
Prob>F	.000	.000	.000	.000
R^2	.306	.231	.242	.181
N	21779	21779	21779	21779

Figures in parentheses are standard errors.
*Indicates significance at the 10 percent level. ** indicates significance at the 5 percent level. *** Indicates significance at the 1 percent level.
Regressions include controls for census region and interview wave.

Table 4.6: Optimal IV Estimates of the Effect of Health Shocks on Hours Worked (without Fixed Effects)

Variable	Work-limiting health conditions		Self-rated health status	
	Husband's hours	Wife's hours	Husband's hours	Wife's hours
Own health	-1206.6***	-833.7***	-1471.1***	-1122.4***
	(27.1)	(24.3)	(39.5)	(35.1)
Spouse's health	-43.5	193.1***	35.1	295.3***
	(28.7)	(26.4)	(40.4)	(34.8)
Own age	236.9***	146.7***	244.2***	174.0***
	(25.8)	(20.8)	(28.3)	(22.5)
Own age squared	-2.6***	-1.7***	-2.7***	-2.0***
	(.2)	(.2)	(.2)	(.2)
Spouse's age	23.0	24.0	11.2	14.2
	(22.5)	(24.7)	(24.6)	(27.0)
Spouse's age squared	-.3	-.3	-.2	-.2
	(.2)	(.2)	(.2)	(.2)
Household non-labor income	.0	-.0	.0	-.0
	(.0)	(.0)	(.0)	(.0)
Own race	79.3***	-70.3***	16.0	-132.6***
	(18.8)	(17.6)	(21.6)	(19.6)
Own educational attainment				
Less than high school	–	–	–	–
GED	-48.7	322.1***	-205.4***	182.7***
	(35.1)	(35.0)	(40.0)	(39.7)
High school	90.0***	233.3***	-55.0**	95.8***
	(22.3)	(20.6)	(26.5)	(24.5)
Some college	134.2***	347.7***	-41.2	188.8***
	(23.7)	(21.8)	(27.9)	(26.1)
College and above	182.7***	376.2***	-9.2	202.4***
	(23.4)	(23.3)	(28.4)	(27.9)
Constant	-3444.5***	-2009.6***	-2983.9***	-2228.2***
	(852.9)	(797.4)	(929.9)	(861.3)
Wald ω^2 (29)	10843.5	7062.2	8629.9	5962.4
Prob> ω^2	.000	.000	.000	.000
R^2	.299	.228	.163	.116
N	21779	21779	21779	21779

Figures in parentheses are standard errors.
*Indicates significance at the 10 percent level. ** indicates significance at the 5 percent level. *** Indicates significance at the 1 percent level.
Regressions include controls for census region and interview wave.

Table 4.7: Estimates of Adaptive Effect of Health Shocks on Hours Worked

Variable	Work-limiting health conditions		Self-rated health status	
	Husband's hours	Wife's hours	Husband's hours	Wife's hours
Husband's health	-1002.0*** (92.6)	-25.2 (75.3)	-1333.5*** (153.5)	-53.0 (111.7)
Duration of husband's health*husband's health	34.2 (35.8)	-21.1 (41.7)	63.9 (50.3)	-11.5 (55.3)
Wife's health	-73.0 (104.3)	-487.6*** (94.3)	140.6 (174.4)	-628.4*** (152.6)
Duration of wife's health*wife's health	-32.2 (37.9)	-25.8 (36.4)	-164.6** (65.9)	-72.6 (57.7)
F	108.6	45.9	92.7	43.0
Prob>F	.000	.000	.000	.000
Under-id test u (K-P LM statistic)	275.4	275.7	162.6	163.3
Prob>u	.000	.000	.000	.000
Weak id test w (K-P Wald statistic)	132.2	132.3	57.3	57.7
Over-id test: functional limitations in (4.2), doctors' diagnoses in (4.5)				
Wald ω^2 (14)	13.0	15.4	22.6	16.4
Prob> ω^2	.528	.354	.068	.292
Over-id test: doctors' diagnoses in (4.2), functional limitations in (4.5)				
Wald ω^{*2} (16)	9.7	21.4	18.5	22.3
Prob>ω^{*2}	.883	.165	.296	.134
N	15674	15674	15674	15674

Figures in parentheses are standard errors.
*Indicates significance at the 10 percent level. ** indicates significance at the 5 percent level. *** Indicates significance at the 1 percent level.
Regressions include demographic controls for own and spouse's age and square of age, household non-labor income, own race, own educational attainment, census region and interview wave.

Table 4.8: First-stage Regressions of Adaptation Models

Variable	Work-limiting health conditions			
	Husband's health	Wife's health	Husband's health*duration	Wife's health*duration
Opl inst for husband's health	.991*** (.035)	.047* (.025)	-.084** (.038)	.046 (.038)
Opl inst for wife's health	.012 (.031)	.935*** (.044)	-.032 (.038)	-.163*** (.039)
Opl inst for husband's health duration	-.056*** (.015)	-.008 (.016)	1.132*** (.058)	.011 (.022)
Opl inst for wife's health duration	.053** (.022)	-.001 (.021)	.027 (.040)	1.370*** (.071)
F	62.3	41.8	43.3	34.4
Prob>F	.000	.000	.000	.000
N	15674	15674	15674	15674
	Self-rated health status			
	Husband's health	Wife's health	Husband's health*duration	Wife's health *duration
Opl inst for husband's health	.768*** (.043)	.004 (.029)	-.038 (.036)	.038 (.039)
Opl inst for wife's health	.086** (.040)	.689*** (.048)	.016 (.035)	-.081* (.049)
Opl inst for husband's health duration	.003 (.023)	.009 (.016)	.967*** (.063)	.049* (.027)
Opl inst for wife's health duration	-.013 (.019)	.003 (.025)	.034 (.031)	.991*** (.102)
F	19.3	11.9	18.5	12.8
Prob>F	.000	.000	.000	.000
N	15674	15674	15674	15674

Figures in parentheses are standard errors.

*Indicates significance at the 10 percent level. ** indicates significance at the 5 percent level. *** Indicates significance at the 1 percent level.

Regressions include demographic controls for own and spouse's age and square of age, household non-labor income, own race, own educational attainment, census region and interview wave.

Bibliography

Aigner, D.J., 1973. Regression with a binary independent variable subject to errors of observation. Journal of Econometrics, 1(1), pp.49-59.

Andrews, D.W. and Schafgans, M.M., 1998. Semiparametric estimation of the intercept of a sample selection model. The Review of Economic Studies, 65(3), pp.497-517.

Angrist, J.D. and Pischke, J.S., 2008. Mostly harmless econometrics: An empiricist's companion. Princeton university press.

Berger, M.C. and Fleisher, B.M., 1984. Husband's health and wife's labor supply. Journal of Health Economics, 3(1), pp.63-75.

Black, D.A., Berger, M.C. and Scott, F.A., 2000. Bounding parameter estimates with nonclassical measurement error. Journal of the American Statistical Association, 95(451), pp.739-748.

Blau, D.M. and Gilleskie, D.B., 2001. The effect of health on employment transitions of older men. In Worker Wellbeing in a Changing Labor Market (pp. 35-65). Emerald Group Publishing Limited.

Bollinger, C.R., 1996. Bounding mean regressions when a binary regressor is mismeasured. Journal of Econometrics, 73(2), pp.387-399.

Bound, J., 1991. Self-Reported Versus Objective Measures of Health in Retirement Models. The Journal of Human Resources, 26 (1), pp.106–138.

Bound, J., Schoenbaum, M., Stinebrickner, T.R. and Waidmann, T., 1999. The dynamic effects of health on the labor force transitions of older workers. Labour Economics, 6(2), pp.179-202.

Butler, J.S., Burkhauser, R.V., Mitchell, J.M. and Pincus, T.P., 1987. Measurement error in self-reported health variables. The Review of Economics and Statistics, pp.644-650.

Charles, K.K., 1999. Sickness in the family: Health shocks and spousal labor supply. Ann Arbor, MI: Gerald R. Ford School of Public Policy, University of Michigan.

Chen, X., Hong, H. and Tamer, E., 2005. Measurement error models with auxiliary data. The Review of Economic Studies, 72(2), pp.343-366.

Chen, X., Hu, Y. and Lewbel, A., 2008a. Nonparametric identification of regression models containing a misclassified dichotomous regressor without instruments. Economics Letters, 100(3), pp.381-384.

Chen, X., Hu, Y. and Lewbel, A., 2008b. A note on the closed-form identification of regression models with a mismeasured binary regressor. Statistics & Probability Letters, 78(12), pp.1473-1479.

Chirikos, T.N. and Nestel, G., 1985. Further evidence on the economic effects of poor health. The Review of Economics and Statistics, pp.61-69.

Currie, J. and Madrian, B.C., 1999. Health, health insurance and the labor market. Handbook of labor economics, 3, pp.3309-3416.

Disney, R., Emmerson, C. and Wakefield, M., 2006. Ill health and retirement in Britain: A panel data-based analysis. Journal of health economics, 25(4), pp.621-649.

DiTraglia, F.J. and García-Jimeno, C., 2017. Mis-classified, binary, endogenous regressors: Identification and inference (No. w23814). National Bureau of Economic Research.

Frazis, H. and Loewenstein, M.A., 2003. Estimating linear regressions with mismeasured, possibly endogenous, binary explanatory variables. Journal of Econometrics, 117(1), pp.151-178.

Fuchs, V.R., 1982. Time Preference and Health: An Exploratory Study. Economic Aspects of Health, edited by Victor R. Fuchs, pp.93-120. Chicago: The University of Chicago Press.

García-Gómez, P., Jones, A.M. and Rice, N., 2010. Health effects on labour market exits and entries. Labour Economics, 17(1), pp.62-76.

Grossman, M., 1972. On the concept of health capital and the demand for health. Journal of Political economy, 80(2), pp.223-255.

Gustman, A.L. and Steinmeier, T.L., 1986. A Structural Retirement Model. Econometrica, 54 (3), pp.555–584.

Hausman, J.A., Abrevaya, J. and Scott-Morton, F.M., 1998. Misclassification of the dependent variable in a discrete-response setting. Journal of econometrics, 87(2), pp.239-269.

Heckman, J., 1990. Varieties of selection bias. The American Economic Review, 80(2), pp.313-318.

Ichimura, H., 1993. Semiparametric least squares (SLS) and weighted SLS estimation of single-index models. Journal of Econometrics, 58(1-2), pp.71-120.

Johnson, R.W., Mermin, G. and Uccello, C.E., 2005. When the nest egg cracks: Financial consequences of health problems, marital status changes, and job layoffs at older ages. Marital Status Changes, and Job Layoffs at Older Ages (December 2005).

Jürges, H., 2007. True health vs response styles: exploring cross-country differences in self-reported health. Health economics, 16(2), pp.163-178.

Kane, T.J., Rouse, C.E. and Staiger, D., 1999. Estimating returns to schooling when schooling is misreported (No. w7235). National Bureau of Economic Research.

Klein, R.W. and Spady, R.H., 1993. An efficient semiparametric estimator for binary response models. Econometrica: Journal of the Econometric Society, pp.387-421.

Klein, R., Shen, C. and Vella, F., 2015. Estimation of marginal effects in semiparametric selection models with binary outcomes. Journal of Econometrics, 185(1), pp.82-94.

Levy, H., 2002. The economic consequences of being uninsured. ERIU Working paper, 12.

Mahajan, A., 2006. Identification and estimation of regression models with misclassification. Econometrica, 74(3), pp.631-665.

McManus, S., Bebbington, P., Jenkins, R. and Brugha, T., 2016. Mental Health and Wellbeing in England: Adult Psychiatric Morbidity Survey 2014: a Survey Carried Out for NHS Digital by NatCen Social Research and the Department of Health Sciences, University of Leicester. NHS Digital.

Mitchell, J.M. and Burkhauser, R.V., 1990. Disentangling the effect of arthritis on earnings: a simultaneous estimate of wage rates and hours worked. Applied Economics, 22(10), pp.1291-1309.

Newey, W.K., 1990. Efficient instrumental variables estimation of nonlinear models. Econometrica: Journal of the Econometric Society, pp.809-837.

Parsons, D.O., 1976. Health, family structure, and labor supply.

Paul, K.I. and Moser, K., 2006. Incongruence as an explanation for the negative mental health effects of unemployment: Meta-analytic evidence. Journal of Occupational and Organizational Psychology, 79(4), pp.595-621.

Robinson, P.M., 1988. Root-N-consistent semiparametric regression. Econometrica: Journal of the Econometric Society, pp.931-954.

Siegel, M.J., 2006. Measuring the effect of husband's health on wife's labor supply. Health Economics, 15(6), pp.579-601.

Staiger, D. and Stock, J. H., 1997. Instrumental variables regression with weak instruments. Econometrica, 65(3), 557–586.

Stern, S., 1989. Measuring the effect of disability on labor force participation. Journal of human Resources, pp.361-395.

van Hasselt, M. and Bollinger, C.R., 2012. Binary misclassification and identification in regression models. Economics Letters, 115(1), pp.81-84.