Labor supply effects of ill health: a weighted instrumental variable approach to misclassification of health measures

Ning Li
Department of Economics, Salisbury University

Abstract
This paper examines the impact of ill health on labor supply, addressing a key challenge in economic research: the misreporting of self-assessed health status. To overcome limitations in prior empirical methods for handling misclassified health, we propose a new approach that identifies the true effect of ill health using observations less likely to be misreported. Rather than assuming fixed misclassification rates, we develop a semi-parametrically estimated health index that extracts more accurate health information from available data. By leveraging this health index to inform misreporting probabilities, we employ a weighted IV estimator and optimize the weighting scheme to balance the tradeoff between squared bias and variance. We demonstrate the superior performance of our model via both simulations and real-world data. Using data from the 2012 Health and Retirement Study (HRS), our findings indeed suggest that conventional methods, including OLS and standard IV techniques, significantly underestimate the negative impact of ill health on labor supply. Our approach reveals a much larger reduction in labor market participation due to ill health, highlighting the economic vulnerability of individuals with health limitations. These findings have important policy implications for designing social safety nets and employment policies to better support workers facing health challenges.

Keywords: Ill Health, Labor Supply, Misclassification, Health Index

JEL Classification: C14, I10, J22

1. Introduction

Building on the seminal work of Grossman (1972), health has been regarded as a component of human capital that plays a crucial role in shaping an individual's labor market behavior. Extensive evidence highlights the impact of poor health status, including increased days lost to illness, reduced productivity, and changes in the marginal rate of substitution between goods and leisure (Bubonya et al., 2017; Gustman and Steinmeier, 1986; Stansfeld et al., 1995; Zhang et al., 2015). With the adoption of various health, labor force, and social support policies (e.g., paid family and medical leave laws), the relationship between health and labor market dynamics has received growing attention, particularly in light of the COVID-19 pandemic, which introduced unprecedented complexities and challenges. For instance, Sheiner and Salwati (2022) point out the severity of a recent surge in disability during the pandemic, estimating that 1.5 million more Americans had a disability between January and September 2022 compared to January 2020. This public health crisis has the potential to significantly alter the labor market dynamics, particularly on the demand side, and will likely place financial strain on the Social Security benefits system as well as related public assistance programs. As such, accurately evaluating the labor supply effects of ill health is crucial for assessing individuals' economic circumstances and aiding policymakers in designing health-related policies to support particularly vulnerable groups.

Among the challenges in studying the labor supply effects of ill health is the measurement error in health variables (e.g., disability). 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 (or from survey to survey) and the pervasive imperfect information among the population on own health status (Bound, 1991; Bound et al. 1999; Gosling and Saloniki, 2014; Liu and Millimet, 2020). In this context, subjective, self-assessed health measures serve as a comprehensive substitute for true health and are therefore widely used in empirical studies, since they have been found to be more correlated with work capacity than other health measures (Blau and Gilleskie, 2001). While such subjective health measures provide information on true health, they are subject to measurement error in a wide range of survey data. Butler et al. (1987) find sizable disparities between the reported arthritis symptoms and the recorded arthritis diagnoses. Brachet (2008) also notes substantial misreporting of maternal smoking status that is reported by respondents themselves in survey data. In line with their findings, the analysis of the 2012 wave of the Health and Retirement Study (HRS) in this paper likewise suggests the existence of enormous inconsistencies between two subjective health measures, the measure of work-limiting health problems and the measure of self-reported health status, which probably implies measurement errors in subjective health variables. Failure to address the measurement error of health variables may lead to biased estimation and misleading recommendations for economic policy.

It is noteworthy that the measurement error of many health variables is non-classical because these health variables are recorded as discrete 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 measurement error in binary health variables, which is typically referred to as misclassification, will be by its nature negatively correlated with the unobserved true health (Aigner, 1973), creating 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 misclassified, subjective health measures. This approach ignores the possible inherent correlation between the instruments and the measurement error, rendering conventional IV estimators biased. Another approach widely used in the literature is to estimate a health index from objective health measures and then substitute this continuous index for the subjective health variable in the labor supply equation, but this approach introduces interpretation and functional form challenges in practice.

This study confronts aforementioned challenges that have not been fully addressed in previous research and proposes a new method to handle the misclassified subjective health regressor in the labor supply equation. Under a greatly relaxed assumption on misclassification rates that are allowed to vary depending on a health index as a linear combination of demographic characteristics and objective health measures, this paper estimates the health index semi-parametrically without imposing any restrictions on the parametric modeling of reported health or misclassification process. Rather than directly substituting this index for subjective health regressors in the labor supply equation, this paper employs it to discern observations in the sample that are free of misclassification. A high probability set is utilized to capture these correctly reported observations. With an estimated high probability set, an IV estimator is proposed to estimate the labor supply effect of ill health by assigning positive weights to observations in the high probability set and zero weights to observations outside the set. This paper optimizes the estimation of the high probability set by balancing the tradeoff between squared bias and variance of the proposed estimator. The results of Monte Carlo simulations suggest that the proposed IV estimator on the high probability set substantially reduces misclassification biases and root-mean-square errors compared to OLS and conventional IV estimators. Further, the proposed estimator is robust to severe and asymmetric misclassifications, whereas OLS and conventional IV estimators increasingly underestimate the coefficient as misclassification rates rise.

Using the 2012 wave of the Health and Retirement Study (HRS), this paper examines the labor supply effects of ill health for men and women aged 45-61. When examining the measure of self-reported health status, the results suggest that women will reduce labor supply by 1,900 hours per year when they rate their health as "Fair" or "Poor" and that OLS and traditional IV estimators demonstrate 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 technique produce similar estimates, while the OLS estimate is biased towards zero.

Beyond this study, the proposed weighted IV estimator can be broadly applied to point identification and estimation in treatment models with misclassification across economic and social science studies. For example, evaluating the efficacy and safety of a newly developed medication may require addressing the challenge that patients do not take the medication as prescribed. Patients' out-of-pocket expenditures on the medication or their health insurance coverage may serve as instruments for implementing the weighted IV approach. In the field of marketing, firms seek to collect clients' feedback (e.g., customer satisfaction), but clients' true attitudes are often misclassified, leading to misleading business decisions. Identifying clients who genuinely provide feedback will help firms capture accurate market signals from the demand side and develop strategic plans accordingly. In the domain of political science, voting and election surveys often suffer from misclassification in pre-election polling. This study enables the distinction between individuals who accurately report their political affiliation and those who do not, making it feasible to investigate the relationship between political affiliation and relevant outcomes.

The paper is organized as follows: Section 2 reviews the relevant literature; Section 3 explains the methodology and its theoretical foundation; Section 4 presents the Monte Carlo simulation results; Section 5 describes the data and provides descriptive statistics of the sample; Section 6 discusses the estimation results; and Section 7 draws conclusions.

2. Literature Review

The effect of ill health on labor supply has been extensively studied for decades (Bound et al., 1999; Chirikos and Nestel, 1985; Disney et al., 2006; García-Gómez et al., 2010; Gosling and Saloniki, 2014; Leroux et al., 2012; Mitchell and Burkhauser, 1990). Currie and Madrian (1999) gives a thorough review of literature, pointing out that most studies draw the same conclusion of the negative effect of health problems on labor supply, but that there is no consensus on its magnitude due to differing health measures and identification methods used.

As generally recognized in the literature, 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. Self-assessed subjective health measures are widely used in studying labor market activities because they are found to be more correlated with work capacity than objective measures. Blau and Gilleskie (2001) find that when including multiple health measures in labor supply equations, subjective measures that describe the comprehensive health status usually have a larger explanatory power than objective indicators that report only some narrow, concrete dimension of health. While subjective health variables have such major advantages, their subjectivity introduces measurement error in various manners. Individuals with the identical underlying health may have different thresholds to report poor health (Baker et al., 2004). 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 economic inactivity in the labor market (Bound et al. 1999). Such a justification motivation can be heightened especially by financial incentives. Maestas et al. (2021) find that unemployment due to the Great Recession promotes the application for Social Security disability benefits, which provide applicants with financial incentives to report more severe conditions to meet the eligibility criteria. Whether random or not, measurement error may bias the estimated relationship between ill health and labor supply.

To address measurement error in health variables, Stern (1989) proposes using objective measures to instrument for subjective measures. Since then, this method has become one of the predominant strategies widely used in empirical studies. The objective health measures commonly used in such studies include functional limitations, doctors' diagnoses, or mortality rate, which only imperfectly reflects the true health. Advantageously, they are argued to be less subject to measurement error, as each objective health measure rates a narrow dimension of health with a relatively concrete "yes or no" threshold. Also, the questionnaire wording regarding such objective measures is less likely to motivate respondents to rationalize their labor force withdrawal by misreporting their health. See Charles (1999) for an empirical example.

However, there are two potential flaws when using relatively objective measures to instrument for subjective measures. First, because many subjective health variables are dichotomous, their measurement error is non-classical. In particular, the measurement error is negatively correlated with the unobserved true health indicator (Aigner, 1973), a phenomenon typically referred to as misclassification. Due to the inherent correlation between true health and measurement error, conventional IV techniques may fail to produce a consistent estimate. The validity of the conventional IV approach requires that the instruments be correlated with true health but uncorrelated with measurement error. However, the inherent correlation between true health and measurement error makes it difficult to ensure that these instruments remain uncorrelated with the measurement error. Further, testing the validity of instruments required by the conventional IV technique is hardly feasible, as true health and measurement error are unobserved, posing a significant challenge to justifying the conventional IV approach. Second, Bound (1991) demonstrates mathematically that measurement error in health measures distorts the estimated effects of other economic factors correlated with health, such as education. Even if the measurement error in health variables is addressed in some way, the estimated coefficients of these economic factors remain biased.

Another approach to addressing measurement error is to recover the latent health status. Bound et al. (1999), Disney et al. (2006) and García-Gómez et al. (2010) construct an underlying health index from a number of objective health indicators and substitute this index or its variant for the subjective health variable when modeling labor market behavior. This method has been extended beyond the study of labor market outcomes, for example Jürges (2007) examines 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 with a continuous health index makes interpreting the results challenging and relies heavily on the parametric assumptions for modeling both health and measurement error. First, individuals who suffer from a slight decline in health, as indicated by the continuous index, are likely to remain economically active in the labor market. There is no clear way to determine the extent to which declines in the health index can be considered an illness that influences individuals' labor market decisions. 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 true health.

Once research omits an important objective indicator, 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 across studies. Third, estimation of the health index in prior studies assumes parametric modeling of health and specific distribution of measurement error, which introduces functional form challenges.

To summarize the literature, the essence of these two predominant approaches is to extract reliable information on true health from relatively objective measures. While the first approach fully recognizes the discreteness of subjective health variables, it fails to handle the inherent non-classical measurement error. Meanwhile, it tends to distort the estimated coefficients on other economic factors correlated with health, making it difficult to compare the relative impacts on labor market decisions of health and other economic factors, such as education. The second approach constructs a continuous health index to circumvent the issue of non-classical measurement error, but substituting this continuous health index for a misclassified, subjective health regressor introduces interpretation and functional form challenges in practice.

This paper contributes to the literature in five important respects. First, it addresses the issues arising from ignoring the inherent non-classical nature of measurement error in health variables or replacing reported health regressors in the labor supply model with a constructed health index. Second, it relaxes the assumptions on the misclassification process. Specifically, the rate of misreporting health status is not assumed to be constant and can vary with demographic characteristics and objective health measures. Estimation under this relaxed assumption not only complements the existing misclassification literature but also significantly expands the methodological toolkit for relevant microeconomic analyses. Third, a latent index of true health is estimated semi-parametrically without imposing any functional form restrictions on the reported health model or distributional assumptions on the misclassification process. Fourth, rather than directly substituting the estimated health index for subjective health regressors in labor supply equations, as done in previous studies, this paper uses this index to identify observations that are free from misclassification. Fifth, it relies only on these correctly classified observations to estimate the labor supply effect of ill health. The selection of observations is data-driven, balancing the tradeoff between squared bias and variance in the proposed estimator.

3. Methodology

3.1. Model

To study the labor supply effect of ill health, this paper uses the following structural model:[1]

$$Y_i = \alpha + X_i'\gamma + H_i^*\beta + \varepsilon_i \tag{1}$$

where Y_i is the observation i's hours of work, H_i^* measures true 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 division. 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, equal to 1 for ill health and 0 for good health. Accordingly, the coefficient, β, reflects the effect of ill health on hours worked.

[1] The structural model is a cross-sectional regression that would be contaminated by the misclassification of the health regressor. While a regression model with panel data can offer advantages for causal inference, for example accounting for time-invariant unobserved heterogeneity, the development of its econometric theories remains limited, particularly in obtaining point estimates. This limitation likely arises from the complexity of addressing both misclassification and heterogeneity simultaneously. The challenge intensifies when misclassification rates are not constant and allowed to vary across individuals. The work of Denteh and Kédagni (2022) is within the very limited literature on misclassification in regression models with panel data, where they estimate bounds on treatment effects in the presence of a misclassified treatment regressor, rather than yielding point estimates. Our work in the present paper aims to point estimate the labor supply equation when misclassification rates of the health regressor (treatment) vary across individuals.

However, the true health status, H_i^*, is rarely observed in practice. Instead, 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 variables is distinct from that of continuous health measures. Here, we focus on the binary health measures that are subject to measurement error.[2] In particular, if the true health is 0, it can be only misreported, if at all, to be 1 and if the true health is 1, it can be only misreported to be 0. Due to the special measurement error, conventional IV techniques may lead to biased estimates as discussed in the Literature Review section.

Notwithstanding the issues associated with conventional IV techniques, the objective health measures do provide information on the unobserved true health. For example, if an individual reports limitation on many daily life functions and numerous medical conditions diagnosed by doctors, they are likely to have poor true health. If an individual reports very few limitations or no doctors' diagnoses, it is likely that their true health is good. Motivated by this idea, many previous studies agree that the unobserved true health depends on the objective health measures in some way. In accord with the literature, we use the threshold-crossing model of the true health as follows:

$$H_i^* = I\{X_i'\pi_1 + Z_i\pi_2 > \mu_i\} \tag{2}$$

where X_i includes all the covariates from the structural model (1) and Z_i is the objective health variable. The objective health variable is termed the excluded variable (like instruments) because

[2] See Hu (2008) for models with a misclassified explanatory variable, where the true explanatory variable and its surrogate are categorized into more than two values.

it can affect the labor supply only through the subjective health measure and thus satisfy the exclusion restriction. The objective health, Z_i, can be a scalar or a vector. Only one excluded variable in (2) suffices to achieve identification and makes it convenient for elaboration. The estimation strategy proposed below remains valid when Z_i captures more than one excluded variable. Moreover, the proposed strategy enables us to estimate the linear combination of X_i and Z_i semi-parametrically without requiring knowledge of the distribution of the error term μ_i. Such greater flexibility distinguishes this study from previous work that parametrically recovers a latent true health index by assuming specific distributions of μ_i. The estimation of the linear combination is discussed in more detail below.

Different from previous research, this paper assumes non-constant misclassification probabilities. Mahajan (2006) is the first to recognize that assuming constant misclassification probabilities is a strong assumption[3] and thus relax it by allowing these probabilities to be functions of covariates in the structural equation. Building on this, we assume that misclassification probabilities depend not only on the covariates X_i in the structural equation but also on the objective health measure Z_i.

To the best of our knowledge, this study is the first to estimate a labor supply equation in which the health variable is subject to misreporting and misclassification rates depend on both included demographic characteristics and excluded objective health measures. This further relaxation has a theoretical basis. As discussed in the Literature Review section, conventional IV estimation struggles to address the misclassified health regressor due to the lack of a guarantee that objective health measures are uncorrelated with measurement error. Given the potential

[3] For example, it is very strong to assume that individuals with different ages, education levels, or incomes have the same probability of misreporting their health status.

correlation between measurement error and objective health measures (excluded variables), we extend the assumption by allowing misclassification probabilities to depend on both X_i and Z_i. Moreover, this assumption also has practical implications. Individuals who have more functional limitations or medical conditions are far less likely to report themselves as healthy, while those with minimal limitations or no medical diagnoses are unlikely to report poor health. In short, individuals with either a very high or very low number of functional limitations and medical diagnoses have an extremely low probability of misreporting their health status.

3.2. Health Index

This study utilizes an estimated health index to extract information about true health from objective health measures and to identify observations that are less likely to be misreported. Unlike the composite health index commonly used in the literature, which heavily relies on researchers' knowledge and assumptions about which variables should be included and how they should be aggregated, the health index in this paper is estimated within a regression model of subjective health variable under a relatively general assumption: true health and misclassification rates depend on demographic characteristics and objective health measures through this health index. More importantly, rather than directly substituting this health index for the misreported health variable in the labor supply equation, the goal of estimating the index is to identify observations that are less subject to misclassification. This approach eliminates interpretational challenges that arise from directly regressing on the health index. Additionally, leveraging the health index in this way arguably removes the influence of researchers' subjective judgement in constructing the index. Below we discuss the semiparametric estimation of the health index.

Normalize the linear combination in (2) to obtain an index V_i:

$$V_i = X_{1i} + X_{2i}\varphi_{20} + X_{3i}\varphi_{30} + \cdots + X_{ki}\varphi_{k0} + Z_i\theta_0 \quad (3)$$

where $X_{1i}, \cdots, X_{ki}$ are the k variables in the vector, X_i. The distribution of H_i^* is assumed to depend on X_i and Z_i through the index V_i as follows:

$$P_i^*(V_i) = Pr(H_i^* = 1|V_i) = Pr(V_i b > \mu_i|V_i) \quad (4)$$

$$V_i b = X_i\pi_1 + Z_i\pi_2 = (X_{1i} + X_{2i}\varphi_{20} + X_{3i}\varphi_{30} + \cdots + X_{ki}\varphi_{k0} + Z_i\theta_0)b \quad (5)$$

where P_i^* represents the probability of actually having ill health, conditioned on the index V_i. Since the index V_i contains many objective health indicators in practice, it is referred to as the **health index**. The misclassification rates are assumed to be functions of the health index:[4]

$$PL(V_i) = Pr(H_i = 1|H_i^* = 0, V_i) \quad (6)$$

$$PR(V_i) = Pr(H_i = 0|H_i^* = 1, V_i) \quad (7)$$

where H_i is the subjective, self-assessed health measure, taking the value 1 for ill health and 0 for good health. Individuals with extreme values of the health index (e.g., having either a very high or very low number of functional limitations and doctors' diagnoses) are less likely to misreport their health status. In other words, as the health index approaches extremely large or small values, the misclassification rates tend to zero, meaning no misclassification.

By the Law of Total Probability, the self-assessed health is also a function of the index:

$$P_i(V_i) = \left(1 - PR(V_i)\right)P_i^*(V_i) + PL(V_i)(1 - P_i^*(V_i)) \quad (8)$$

where P_i is the probability of reporting ill health ($H_i = 1$) given the index V_i.

While the specific model for the observable H_i is unknown, the parameters $\{\varphi_{m0}\}_{m=2}^{k}$ and θ_0 in the index are identified and estimated semi-parametrically in a single-index model.[5] With

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

[5] While the specific model of H_i is unknown, it is feasible to estimate a binary response model of H_i by maximizing the following quasi log-likelihood function:

these index parameters, V_i and $P_i(V_i)$ are consistently estimated. Their identification and estimation are particularly powerful, as no restrictions are imposed on the functional form governing how the index determines the probability of actually having ill health ($P_i^*(V_i)$), the misclassification process ($PL(V_i)$ and $PR(V_i)$), and the probability of reporting ill health ($P_i(V_i)$).

3.3. High Probability Set

To select observations with extreme values of the health index, we define a high probability set that contains observations with a high likelihood of having no misclassification, as follows:

$$\{V_i: P_i(V_i) < N^{-a} \; or \; P_i(V_i) > 1 - N^{-a}\}, 0 < a < 1 \tag{9}$$

where N is the sample size and a is the high probability set parameter. In the high probability set, P_i is assumed to be a monotonic function of the health index V_i. Thus, as the sample size N increases, the set tends to select observations with extreme index values. The more extreme an index value is, the higher the probability the observation correctly reports its true health status, and therefore, a higher weight should be assigned to this observation in estimation. See Andrews and Schafgans (1998), Heckman (1990), and Klein et al. (2015) for seminal theoretical developments on achieving identification and estimation in sample selection models by leveraging extreme observations. The important parameter a is not chosen arbitrarily. The determination of its optimal value will be discussed in detail below.

$$L(\{\varphi_m\}_{m=2}^{k}, \theta) \equiv \sum_{i}^{N} \left\{H_i \ln\left[\hat{P}_i\left(V_i(\{\varphi_m\}_{m=2}^{k}, \theta)\right)\right] + (1 - H_i)\ln\left[1 - \hat{P}_i\left(V_i(\{\varphi_m\}_{m=2}^{k}, \theta)\right)\right]\right\}$$

where $\hat{P}_i$ is a kernel estimator of P_i

$$\hat{P}_i(\hat{V}_i = t) = \frac{\sum_j \frac{1}{Nh} H_i K\left(\frac{t - V_j}{h}\right)}{\sum_j \frac{1}{Nh} K\left(\frac{t - V_j}{h}\right)}$$

where K is a standard normal kernel with window $h = O(N^{-.2})$. For the detailed establishment of semiparametric model estimation, see Klein and Spady (1993) and Ichimura (1993).

The construction of the high probability set is inspired by Shen (2013), who examines health expenditures, which are only observed for patients who seek medical care services. To address sample selection issues in her empirical analysis, Shen proposes a similar high probability set to "trap" observations where the probability of seeking medical care is extremely close to 1. By using only these trapped observations, Shen consistently estimates the health expenditure model, as these observations do not suffer from selection bias. In a different context, this paper estimates the labor supply effect of ill health, also relying on the high probability set, as observations in this set do not suffer from misclassification.

3.4. Estimator

Without loss of generality, we propose the estimator for the following simplified model:[6]

$$Y_i = \alpha + H_i^* \beta + \varepsilon_i \tag{10}$$

Again, the coefficient β reflects the effect of ill health on hours worked. The proposed estimator for the coefficient on health is

$$\hat{\beta} = \frac{\sum_{i=1}^{N} (\hat{P}_i - \bar{P}) Y_i \hat{S}_i}{\sum_{i=1}^{N} (\hat{P}_i - \bar{P}) H_i \hat{S}_i} \tag{11}$$

where $\hat{P}_i$ is a kernel estimator of P_i in a semiparametric model, as explained in footnote 5, $\hat{S}_i$ is a weighting function that assigns positive weights to observations in the high probability set and zero weights to those outside the set. See **Appendix B** for its detailed definition. $\bar{P}$ denotes the weighted average of $\hat{P}_i$ in the high probability set, given by:

[6] With the index V_i recovered, we simplify the structural model (1) using the approach proposed by Robinson (1988) for partially linear models. The simplification separates the estimation of the coefficient on the health regressor from that on other economic covariates, overcoming the problems raised by Bound (1991) that the estimated coefficients on other covariates will be distorted by the mismeasured health regressor if these covariates are correlated with health. See **Appendix A** for detailed steps of simplification.

$$\frac{\sum_{i=1}^{N} \hat{P}_i \hat{S}_i}{\sum_{i=1}^{N} \hat{S}_i} \quad (12)$$

The estimator $\hat{\beta}$ essentially represents an instrumental variable strategy applied to the high probability set.[7]

For the unknown high probability set, estimating it and determining the appropriate weighting scheme presents a technical challenge. The high probability set parameter a, which is an argument in the weighting function $\hat{S}_i$, controls which observations are selected by the high probability set. If the selected index does not approach sufficiently large or sufficiently small values ($a \to 0$),[8] the bias of the estimator will be substantial because many misreported observations are included. Conversely, if the index value is too extreme ($a \to 1$),[9] the bias will become negligible, but the variance will be significant due to having too few observations. Therefore, the optimal high probability set parameter, a, is determined by balancing the (squared) bias and variance in the proposed estimator. Key assumptions and results related to this balance will be discussed below.

In practice, two high probability set parameters are used in this study, one for each tail. Specifically, the parameter a_1 selects observations with extremely small index values (left tail), while the parameter a_2 selects observations with extremely large index values (right tail). The left

---

[7] For comparison, the traditional IV estimator can be seen as assigning equal weight to each observation in the sample. For example, for the model with only one regressor

$$Y_i = \alpha + H_i^* \beta + \varepsilon_i$$

the traditional IV estimator is

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

where $\hat{P}_i$ is the kernel estimator of the instrument P_i, and $\bar{P}$ is the sample mean of $\hat{P}_i$.

[8] With $a \to 0$, the high probability set approximates the following set,

$$\{V_i : P_i(V_i) < 1 \; or \; P_i(V_i) > 0\}$$

which captures almost all the observations in the sample.

[9] With $a \to 1$, the high probability set approximates the following set,

$$\{V_i : P_i(V_i) \approx 0 \; or \; P_i(V_i) \approx 1\}$$

which includes too extreme observations only.

tail of the index may differ from the right tail in distribution, so two different parameters are needed to control the potential differences in the selection process from each tail. When the index is symmetrically distributed, these two parameters will be identical.

3.5. Key Assumptions and Results

To construct the high probability set and implement the proposed weighted IV technique, we make the following assumptions:

Assumption 1 (Non-Differential Measurement Error).

$$\mathbb{E}(Y_i|H_i^*, H_i, X_i, Z_i) = \mathbb{E}(Y_i|H_i^*, X_i, Z_i). \tag{13}$$

Assumption 1 is conventionally referred to as **non-differential measurement error**—a type of measurement error that provides no additional information about the outcome variable beyond what is given by the true health status and excluded variables. This assumption is standard and has been extensively utilized in the measurement error literature (Chen, et al., 2008; DiTraglia and Garcia-Jimeno, 2019; Kasahara and Shimotsu, 2022; Mahajan, 2006; Schennach, 2022).

Assumption 2 (Misclassification Restriction at Infinity). $P_i(V_i)$ and $P_i^*(V_i)$ converge to 0 (1) at the same rate as $V_i \to -\infty$ ($V_i \to +\infty$).

Assumption 2 implies that the probabilities of actually having ill health and reporting ill health approach 0 or 1 at the same rate as the health index becomes sufficiently small or sufficiently large, respectively. A sufficient condition for this assumption is: $PL(V_i)$ converges to 0 as $V_i \to -\infty$ at a faster rate than $P_i(V_i)$ converges to 0, while $PR(V_i)$ converges to 0 as $V_i \to +\infty$ at a faster rate than $P_i(V_i)$ converges to 1. Meanwhile, $1 - PL(V_i) \nrightarrow 1$ as $V_i \to +\infty$ and $1 - PR(V_i) \nrightarrow 1$ as $V_i \to -\infty$. To see the validity of this condition, check equation (8) when $V_i \to -\infty$ and the following equation when $V_i \to +\infty$:

$$1 - P_i(V_i) = PR(V_i)P_i^*(V_i) + (1 - PL(V_i))(1 - P_i^*(V_i)) \quad (14)$$

Assumption 2 implies that $PL(V_i) + PR(V_i) < 1$ for extreme values of the index. Interestingly, this restriction on misclassification aligns with "Assumption 2 – Restriction on the Extent of Misclassification" in Mahajan (2006, p.637), except that Mahajan imposes this condition on the entire real line $\mathbb{R}$ rather than only at infinity. Similar assumptions have also been discussed in the misclassification literature (see Bollinger, 1996; DiTraglia and Garcia-Jimeno, 2019; Frazis and Loewenstein, 2003; van Hasselt and Bollinger, 2012) and in the influential review by Hu (2008). This restriction on misclassification ensures that the surrogate H_i is positively correlated with the true health status H_i^*. While the observed surrogate is contaminated by measurement error, it remains more informative than an arbitrary guess.

Assumption 3 (Tail Conditions). In the distribution of H_i^* as illustrated in (4), let $G_v(.)$ and $F_\mu(.)$ denote the cumulative distribution functions of the index V_i and the error μ_i, respectively. Assume for all $t < T$ sufficiently small,

$$G_v(t) > F_\mu(t) \quad (15)$$

and for all $t > T$ sufficiently large,

$$1 - G_v(t) > 1 - F_\mu(t) \quad (16)$$

The tail conditions specified in Assumption 3 require that the tails of the index distribution are heavier than those of the error term μ_i in (2). This assumption provides a lower bound for the terms $\mathbb{E}(S_{Li})$ and $\mathbb{E}(S_{Ri})$, contributing to the order analysis of the squared bias and variance. See below for a detailed order analysis. The concept of heavier tails in the index distribution has been explored by Andrews and Schafgans (1998) and Klein et al. (2015) to establish the large sample properties of their estimators. Intuitively, if the index distribution has excessively thin tails, it becomes difficult to select enough observations from the extreme ends, thereby weakening the implementation of the proposed weighting scheme.

Result 1. Under Assumptions 2 and 3,

$$\mathbb{E}(S_{Li}) \geq O(N^{-a}) \tag{17}$$

$$\mathbb{E}(S_{Ri}) \geq O(N^{-a}) \tag{18}$$

$$\mathbb{E}(S_{Li}^2) \geq O(N^{-a}) \tag{19}$$

$$\mathbb{E}(S_{Ri}^2) \geq O(N^{-a}) \tag{20}$$

Proof. See **Appendix C**.

Result 2. Under Assumptions 1-3,

$$|Bias| \leq O(N^{-a}) \tag{21}$$

$$Var(\hat{\beta}) = \frac{\mathbb{E}\left\{S[\hat{V}, x(a, \hat{P})]^2\right\}}{N\left\{\mathbb{E}\left[S\left(\hat{V}, x(a, \hat{P})\right)\right]\right\}^2} + \frac{\mathbb{E}\left\{S[\hat{V}, y(a, \hat{P})]^2\right\}}{N\left\{\mathbb{E}\left[S\left(\hat{V}, y(a, \hat{P})\right)\right]\right\}^2} \tag{22}$$

Proof. See **Appendix D**.

As shown in Result 2, the bias and variance of the proposed estimator depend on the high probability set parameter a. The optimal parameter a is estimated to ensure that squared bias and variance converge to zero at the same rate.

4. Monte Carlo Simulations

We conduct Monte Carlo simulations to evaluate the performance of the proposed estimator in finite sample studies. The data are generated as follows:

$$Y_i = 2 + 1.5X_{1i} + 3H_i^* + \varepsilon_i \tag{23}$$

$$H_i^* = I\{V_i > \mu_i\} \tag{24}$$

$$V_i = 3X_{1i} + 4X_{2i} + 5X_{3i} \tag{25}$$

$$H_i = H_i^* I\{U_i > PR_i\} + (1 - H_i^*) I\{U_i < PL_i\} \tag{26}$$

where H_i^* is a dichotomous variable representing the unobserved true health status, which determines the outcome of interest, Y_i. H_i^* depends on an index V_i through a threshold-crossing model. The index V_i is a linear combination of included and excluded variables, X_{1i}, X_{2i}, and X_{3i}, which are drawn from a multivariate normal distribution:

$$X' = [X_{1i}, X_{2i}, X_{3i}]' \sim N(0, \Sigma) \tag{27}$$

where $\sum$ is a non-diagonal matrix that allows for correlations between any two exogenous variables in X. The index is rescaled to have unity variance. The error terms are normally distributed as follows:

$$\varepsilon_i \sim N(0,1) \tag{28}$$

$$\mu_i \sim N(\bar{\mu}, .25) \tag{29}$$

where the mean $\bar{\mu}$ takes on various values, each corresponding to a specific simulation model. The tails of μ_i are thinner than those of the index V_i, satisfying the tail conditions in Assumption 3. The random variable U_i follows a standard uniform distribution. It generates random numbers such that when $H_i^* = 1$, its surrogate H_i is observed to be 0 with probability PR_i, and when $H_i^* = 0$, H_i equals 1 with probability PL_i. The misclassification rates, PR_i and PL_i, are functions of the index V_i:

$$PL_i = \alpha_L \Phi(5V_i) \tag{30}$$

$$PR_i = \alpha_R [1 - \Phi(5V_i)] \tag{31}$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function. The constants α_L and α_R control the degree of misreporting in the true health variable; larger values indicate more severe misclassification.

We set the values of $\bar{\mu}$, α_L, and α_R in two scenarios: (1) $\bar{\mu}$ takes the values 0, -1, -0.5, 0.5, or 1, while $\alpha_L = \alpha_R = 0.9$, and (2) with $\bar{\mu} = 0$, various values are assigned to α_L and α_R to model severe, moderate, and no symmetric misclassification as well as asymmetric misclassification. For each model, we run Monte Carlo simulations with 1,000 observations for 1,000 replications. In each replication, the coefficient on H_i^* in the outcome equation is estimated using the proposed technique. For comparison, we also estimate this coefficient using OLS and conventional IV regressions. We report the mean, standard deviation, and root-mean-square error (RMSE) of the estimates from 1,000 replications for OLS, conventional IV, and the proposed IV on HPS regressions separately.

Table 1 presents the simulation results for five model setups in the first scenario, where $\bar{\mu}$ takes values 0, -1, -0.5, 0.5, or 1, while $\alpha_L = \alpha_R = 0.9$. For each model setup, we report the mean, standard deviation, and RMSE based on 1,000 replications for OLS, conventional IV, and the proposed IV on HPS techniques. Additionally, for the IV estimator on HPS, we report the mean and standard deviation of the optimal high probability set parameter from 1,000 replications. Overall, the proposed IV estimator on HPS performs well, exhibiting significantly smaller RMSEs compared to OLS and conventional IV estimators. The OLS and conventional IV estimators tend to underestimate the coefficients on H^*, particularly in models where $\bar{\mu}$ takes values of -1 or 1. For instance, when $\bar{\mu} = 1$, the OLS and conventional IV estimates account for only 37.8% and 44.9% of the true coefficient, respectively, leading to substantial biases. Moreover, the RMSEs for OLS and conventional IV estimators are 1.866 and 1.654, respectively, whereas the proposed IV estimator on HPS has an RMSE of 0.470, indicating that it converges in probability to the true coefficient at a much faster rate. Similar underestimation results are observed when $\bar{\mu} = -1$. Such underestimations align with the findings of Aigner (1973) and Bound (1991). In contrast, the IV estimator on HPS significantly reduces bias while only slightly increasing the standard deviation due to the use of a fraction of the sample. In addition, the method proposed in this paper successfully attains unbiased estimates of the coefficient on X_1 by employing the technique of Robinson (1998), as discussed in Appendix A, validating the separation between estimating coefficients on health regressor and other covariates.

Table 2 presents the simulation results for the second scenario, where we examine severe, moderate, and no symmetric misclassification as well as asymmetric misclassification while keeping $\bar{\mu} = 0$. When no misclassification is present, OLS, conventional IV, and the proposed IV on HPS techniques produce nearly identical unbiased estimates. However, as misclassification severity increases, the proposed IV estimator on HPS demonstrates an obvious advantage over the other two estimators. It remains robust to misclassification severity, while OLS and conventional IV increasingly underestimate the coefficient as misclassification rates rise. Furthermore, the proposed IV estimator on HPS is also robust to asymmetric misclassification, featuring significantly reduced bias and much smaller RMSEs compared to the other two estimators.

5. Data

This study uses the 2012 wave of the Health and Retirement Study (HRS) to analyze the labor supply effect of one's ill health. The Health and Retirement Study is a nationally representative survey of aging American households starting from 1992. It 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, self-reported work-limiting health problems, functional limitations, and doctors' diagnoses to name a few. Such a rich set of measures helps to assess various aspects of individuals' health status and makes it possible to implement the observation selection as illustrated in the Methodology section. Since the initial HRS cohort was first interviewed in 1992, new cohorts have been introduced subsequently in 1993, 1998, 2004, 2010, and 2016. Among the waves in the HRS, the 2012 wave is chosen with two considerations in mind: First, this study aims to analyze the labor supply decision of individuals aged 45-61 when they experience ill health. Given an increasing proportion of early cohorts go out of this specific age range with time, the use of a recent wave (2012) after a new cohort (2010 cohort) entered the survey guarantees enough observations in this age group and has meaningful implications that relate most to the current economy. Second, the Great Recession in 2007-2009 caused an elevated rate of unemployment, which peaked in 2010. Analysis of 2012 data helps understand people's labor market behavior during recession recovery.

This study focuses on respondents aged between 45 and 61 because at age 62 individuals are able to collect Social Security retirement benefits, which 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 problems compared to those who have not reached this age threshold. Dobkin et al. (2018) and Li (2023) both disclose the role of formal insurance that Social Security retirement benefits have played in mitigating negative effects of health shocks on employment or earnings; the negative effects are found to be much insured by Social Security retirement income once individuals reach its age threshold. In addition, we exclude observations with missing data on labor supply, health measures, age, race, education, income, marital status, and census division. As a result, there are 2,995 men and 4,089 women in the sample.

Table 3 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 job and second job. For each job, we 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 for pay are designated having zero hours of work. There are two self-assessed health measures used, work-limiting health problem and self-reported health status.[10] Their respective effects on labor supply will be examined individually. The excluded variables include the number of functional limitations and seven indicators of doctors' diagnoses. Since ill health may affect the hours worked of men and women differently, we examine the labor supply effect for the genders separately.

Table 4 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. To account for the effect of other income sources that are unrelated to labor earnings, we calculate the non-labor income by subtracting individuals' own labor income from their total household income. On average, females have 4,600 dollars more of nonlabor income than males. There are more single women than single men in this sample, with 22.1 percent of women being divorced or widowed and 8.0 percent never married, compared to only 13.3 percent of divorced or widowed men and 7.8 percent of never married men.

Panel B presents the labor market behavior for the genders. Males have a stronger attachment to the labor market than females. They have a higher labor force participation rate (70.8%) than females (62.4%) and supply more hours of work. Conditional on working for pay, 15.8 percent of

[10] 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 we divide these five categories into two groups: 1 for "Fair" or "Poor" health and 0 for "Excellent," "Very Good," or "Good" health.

men also work on a second job, compared to only 13.4 percent of women. In addition, male workers devote more hours per year than female workers not only to their main job but also to the second job.

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 (24.3%) 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, 26.9 percent for work-limiting health problems versus 27.7 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 5** illustrates the reported discrepancy between these two health measures by gender. Among the 738 male respondents with a health problem that limits their work, only 424 (57.5%) report “Fair” or “Poor” health, while in the group with “Fair” or “Poor” health (727) there are 303 individuals (41.7%) reporting no work-limiting problems. The female sample also demonstrates a similar large reporting inconsistency between these two health measures. Such substantial measurement discrepancies add evidence that different self-assessed measures of health may evaluate different dimensions of the true health and thus introduce considerable misclassification when serving as a surrogate of the unobserved true health status in the labor supply equation.

To instrument for the subjective health measures, we include the number of functional limitations[11] and several doctors' diagnoses. On average, males have 1.7 types of limitations on daily life functions and females have 2.3. Forty-eight 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

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

and more than 19 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).

6. Results

Table 6 and **Table 7** present estimation results for men and women aged 45-61, respectively, when the self-reported health status is examined. We apply the instrumental strategy on the high probability set (termed IV on HPS hereafter) explained in the Methodology section to estimate the labor supply effect of ill health. As a comparison, we provide the OLS and IV estimates.[12] When a man rates his health as “Fair” or “Poor,” he will work 2,027 fewer hours per year 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 734 hours per year if his health is “Fair” or “Poor.” In the estimation for women, not only the OLS estimator but also the traditional IV estimator demonstrate attenuation biases compared to the proposed IV estimator based on the high probability set. The IV estimate on HPS suggests that “Fair” or “Poor” health reduces women's labor supply by 1,900 hours per year. As a comparison, the results of the other two estimates suggest that the working time devoted by women reduces by 615 hours in the OLS estimation and 1,577 hours in the traditional IV estimation.

[12] We employ an optimal instrumental variable strategy. The predicted expectation of the subjective health indicator conditional on the health index is used as an optimal instrument for the subjective health regressor. See Newey (1990) for optimal instrumental variables.

Both men and women greatly reduce their labor supply in the face of health declines. In particular, the 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.” It may reflect a change in the extensive margin of labor supply. Given the average annual hours worked of 2,144 for men and 1,844 for women who are working for pay as exhibited in **Table 4**, the estimated reduction of about 2,000 in yearly working hours may indicate a high likelihood of exiting the labor market when experiencing health problems. It provides supporting evidence of people’s withdrawal from the labor force due to their health declines that has been documented in the literature (Bound et al., 1999; Disney et al., 2006; García-Gómez et al., 2010). Meanwhile, it could also be driven by the change in the intensive margin of labor supply. When analyzing the sample distribution of men’s and women’s yearly hours worked, we surprisingly find that there are a sizable number of respondents working more than 2,000 hours per year, even up to 3,000 hours per year. The greatly reduced working hours may also result from those hard workers who previously worked more hours than average workers when they were having good health and then switched to working less following their health deteriorations.

Table 8 and **Table 9** present the estimation results for men and women, respectively, when the work-limiting health is used. The IV estimate on HPS suggests that a man will reduce his labor supply by 1,659 hours per year when he suffers from a work-related limitation in health. The traditional IV estimation produces similar results, while the OLS estimate is biased towards zero. For the sample of women, the IV estimate on HPS indicates that a woman will reduce her labor supply by 1,323 hours per year when she has a work-related health limitation, while she will reduce her labor supply by 1,059 and 1,301 hours per year by the OLS and traditional IV estimations, respectively.

The traditional IV estimate and the IV estimate on HPS yield close results. There are two 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. 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 which measure of health is used, which is consistent with the review of Currie and Madrian (1999) and findings of Siegel (2016) and Li (2023).

7. Conclusion

The U.S. has been among leading nations with the highest percentage of GDP devoted to health care sector (17.3% in 2022).[13] Nevertheless, Americans do not necessarily access to higher quality health care or have better health outcomes. Many health metrics, for example life expectancy, of the U.S. are inferior to those of other OECD countries. Ill health can significantly impair people's

[13] See the website for Centers for Medicare & Medicaid Services: https://www.cms.gov/data-research/statistics-trends-and-reports/national-health-expenditure-data/historical#:~:text=U.S.%20health%20care%20spending%20grew,spending%20accounted%20for%2017.3%20percent.

enjoyment of good health per se, productivity in labor market, as well as wellbeing. Studying the labor supply effect of ill health adds an important piece to the full picture of evaluating economic and welfare consequences of health declines. Knowing such information helps assess economic values of improved health outcomes and better assist policy makers in allocating health resources in a more cost-effective manner.

However, studying labor supply effects of ill health presents empirical challenges, one of which is due to pervasive measurement error in binary health variables commonly encountered in survey data. To solve the problems associated with approaches to misclassified health measures that are predominantly used in the literature, this study extracts information on unobserved true health through a health index as a linear combination of demographic characteristics and objective health measures. Allowing for general correlation between exogenous variables (included and excluded) and misclassification process, this paper assumes that both true health and misclassification rates depend on this health index. Based on such a generic assumption, this paper estimates the health index semi-parametrically without imposing any restrictions on the functional form of reported health model or the distribution of misclassification process. This paper defines a high probability set to select observations with extreme values of health index; as index values become extreme, misclassification rates of corresponding observations are argued to approach zero. With an estimated high probability set, the labor supply effect of ill health is estimated by implementing an IV strategy that assigns positive weights to observations on the high probability set and zero weights to observations off the set. This paper optimizes the estimation of the high probability set by balancing the tradeoff between the squared bias and variance of the proposed estimator. The results of Monte Carlo simulations suggest that the proposed IV estimator on the high probability set demonstrates significant advantages over OLS and conventional IV estimators.

It shows substantially reduced biases and root-mean-square errors, as well as robustness to severe and asymmetric misclassification rates compared to the latter two estimators.

This study uses the 2012 wave of the HRS to examine the labor supply effect of ill health. The results suggest that OLS and conventional IV methods considerably underestimate the amount of reduced working hours when people experience illnesses. In particular, the proposed strategy estimates that women who report themselves as having “Fair” or “Poor” health reduce their labor supply by 1,900 hours per year, while the estimated reduction in yearly hours worked is only 615 and 1,577 in OLS and conventional IV estimations, respectively. The greatly reduced work hours may reflect changes both in the extensive and intensive margin of labor supply. The findings suggest that ill health places individuals in a position of increased vulnerability.

This study has several limitations. First, this is a cross-sectional study. Under the assumption on the absence of time-invariant unobserved heterogeneity, the proposed estimation strategy effectively resolves misclassified health treatment variables. Nevertheless, the possible existence of heterogeneity and the complexity of handling misclassification in panel data warrants further development of theoretical econometrics in this area. Second, this study assumes that misclassification process is exogenous to labor supply. While this assumption holds in many empirical analyses, including those beyond labor economics, the endogeneity of misreporting remains a technical issue, as people may misreport their health status based on economic outcomes of interest. Since the proposed strategy is essentially an IV estimation based on observations that are free of misclassification, it is promising to readily extend this strategy to settings where both misclassification and endogeneity co-exist. Yet such extensions require careful examination in theory and practice. Third, this study employs a linear labor supply model. Given the prevalence of zero hours worked in aging populations and the potential heterogeneous treatment effects, further research is needed to extend the proposed strategy to nonlinear regression models.

References

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

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

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.

Baker, M., Stabile, M. and Deri, C., 2004. What do self-reported, objective, measures of health measure?. *Journal of human Resources*, *39*(4), pp.1067-1093.

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, John, 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.

Brachet, T., 2008. Maternal smoking, misclassification, and infant health. *Misclassification, and Infant Health (September 16, 2008)*.

Bubonya, M., Cobb-Clark, D.A. and Wooden, M., 2017. Mental health and productivity at work: Does what you do matter?. *Labour economics*, *46*, pp.150-165.

Butler, J.S., Burkhauser, R.V., Mitchell, J.M. and Pincus, T.P., 1987. Measurement error in self-reported health variables. *Review of Economics and Statistics*, 69(4), 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., Hu, Y. and Lewbel, A., 2008. Nonparametric identification of regression models containing a misclassified dichotomous regressor without instruments. *Economics Letters*, *100*(3), pp.381-384.

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.

Denteh, A. and Kédagni, D., 2022. Misclassification in Difference-in-differences Models. *arXiv preprint arXiv:2207.11890*.

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 Garcia-Jimeno, C., 2019. Identifying the effect of a mis-classified, binary, endogenous regressor. *Journal of Econometrics*, *209*(2), pp.376-390.

Dobkin, C., Finkelstein, A., Kluender, R. and Notowidigdo, M.J., 2018. The economic consequences of hospital admissions. *American Economic Review*, *108*(2), pp.308-352.

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.

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.

Gosling, A. and Saloniki, E.C., 2014. Correction of misclassification error in disability rates. *Health economics*, *23*(9), pp.1084-1097.

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*, vol. 54, no. 3, pp. 555-584.

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

Hu, Y., 2008. Identification and estimation of nonlinear models with misclassification error using instrumental variables: A general solution. *Journal of Econometrics*, *144*(1), pp.27-61.

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

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

Kasahara, H. and Shimotsu, K., 2022. Identification of Regression Models with a Misclassified and Endogenous Binary Regressor. *Econometric Theory*, *38*(6), pp.1117-1139.

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.

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.

Leroux, J., Rizzo, J.A. and Sickles, R., 2012. The role of self-reporting bias in health, mental health and labor force participation: a descriptive analysis. *Empirical Economics*, *43*, pp.525-536.

Li, N., 2023. Health and household labor supply: instantaneous and adaptive behavior of an aging workforce. *Review of Economics of the Household*, *21*(4), pp.1359-1378.

Liu D, Millimet DL., 2021. Bounding the joint distribution of disability and employment with misclassification. *Health economics*, 30(7), pp.1628-1647. doi: 10.1002/hec.4265.

Maestas, N., Mullen, K.J. and Strand, A., 2021. The effect of economic conditions on the disability insurance program: Evidence from the great recession. *Journal of Public Economics*, *199*, p.104410.

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.

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.

Schennach, S., 2022. Measurement systems. *Journal of Economic Literature*, *60*(4), pp.1223-1263.

Shen, C., 2013. Determinants of health care decisions: Insurance, utilization, and expenditures. *Review of Economics and Statistics*, *95*(1), pp.142-153.
Sheiner, L. and Salwati, N., 2022. How Muchis Long COVID Reducing Labor Force Participation? Not Much (So Far).
Siegel, M. J. (2006). Measuring the effect of husband's health on wife's labor supply. *Health Economics*, *15*(6), 579-601.
Stansfeld, S., Feeney, A., Head, J., Canner, R., North, F. and Marmot, M., 1995. Sickness absence for psychiatric illness: the Whitehall II Study. *Social science & medicine*, *40*(2), pp.189-197.
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.
Zhang, W., Sun, H., Woodcock, S. and Anis, A., 2015. Illness related wage and productivity losses: Valuing ‘presenteeism’. *Social Science & Medicine*, *147*, pp.62-71.

Appendices

Appendix A: Model Simplification

Before proposing the estimator, it is necessary to simplify the structural model (1). With the index V_i recovered, we simplify the structural model using the approach proposed by Robinson (1988) for partially linear models. Rewrite the outcome equation in (1) as follows:

$$Y_i = \alpha + X_i'\gamma + G(V_i) + newerror \tag{A1}$$

where $G(V_i) = \mathbb{E}(H_i^*|V_i)\beta = P_i^*(V_i)\beta$ and $newerror = \left(H_i^* - G(V_i)\right)\beta + \varepsilon_i$. As H_i^* is unobserved, the function $G(V_i)$ is unknown. Taking the expectation of every term in (A1) conditional on the index V_i, then

$$\mathbb{E}(Y_i|V_i) = \alpha + \mathbb{E}(X_i|V_i)'\gamma + G(V_i) \tag{A2}$$

Making the difference between equations (A1) and (A2) on both sides, we can obtain a differenced model as

$$Y_i - \mathbb{E}(Y_i|V_i) = \left(X_i - \mathbb{E}(X_i|V_i)\right)'\gamma + newerror \tag{A3}$$

Regarding this differenced model, Robinson (1988) shows that the OLS estimator of the parameter γ is consistent at the $\sqrt{N}$ rate. Substracting the estimator of $X_i'\gamma$ from both sides and still using Y_i to denote the differenced outcome on the left-hand side, the model in (1) can be simplified into the model as follows:

$$Y_i = \alpha + H_i^*\beta + \varepsilon_i \tag{A4}$$

Again, the coefficient β reflects the effect of ill health on hours worked.

Essentially, this simplification process separates the estimation of coefficients on other economic covariates from the estimation of health effects on labor supply. Bound (1991) argues that the mismeasured health variables will distort the estimated coefficients on other economic covariates that are related to health. Furthermore, he mathematically shows that even if the measurement error of health variables is addressed, the distorted estimation of coefficients on other covariates remains. To separate such twisting effects on labor supply, we 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 ill health 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.

Appendix B: Definition of $\hat{S}$

Dropping the subscripts, the mathematical definition of $\hat{S}$ is

$$\hat{S} = S(\hat{V}, a, \hat{P}) = S\left(\hat{V}, x(a, \hat{P})\right) + S\left(\hat{V}, y(a, \hat{P})\right) \tag{A5}$$

$$S\left(\hat{V}, x(a, \hat{P})\right) = \tau(\hat{V})C\left(x(a, \hat{P})\right) \tag{A6}$$

$$S\left(\hat{V}, y(a, \hat{P})\right) = \tau(\hat{V})C\left(y(a, \hat{P})\right) \tag{A7}$$

where

$$C(z) = \begin{cases} 0, & z \leq 0 \\ 1 - exp\frac{-z^k}{b^k - z^k}, & 0 < z < b \\ 1, & z \geq b \end{cases} \tag{A8}$$

$$\tau(V) = \frac{1}{1 + exp\left[N^{.2}\left(\frac{\mathbb{E}(g_v(V))N^{-.005}}{\ln N} - g_v(V)\right)\right]} \tag{A9}$$

$$x(a, P) = \ln\tfrac{1}{P} - \ln N^a \tag{A10}$$

$$y(a, P) = \ln\tfrac{1}{1-P} - \ln N^a \tag{A11}$$

where b, k are constants and N the sample size. The construction of the selection function $\hat{S}$ follows the definition of the selection function used in Shen (2013) and Klein et al. (2015) that select extreme observations from one tail. To also select observations from the other tail, the present study extends their definition analogously on the other extreme end. Here, $C\left(x(a, \hat{P})\right)$ and $C\left(y(a, \hat{P})\right)$ are the core of the selection function $\hat{S}$, an extension of the smooth selection function in Andrews and Schafgans (1998). As an individual has an extremely small health index value $\left(\hat{P} < N^{-a} \Rightarrow x(a, \hat{P}) > 0\right)$, $C\left(x(a, \hat{P})\right)$ will assign a positive weight to this observation. When he or she has an extremely large health index value $\left(\hat{P} > 1 - N^{-a} \Rightarrow y(a, \hat{P}) > 0\right)$, $C\left(y(a, \hat{P})\right)$ will assign a positive weight to this observation too. As such, the observations on the high probability set are positively weighted, while those observations excluded from the high probability set are zero weighted. The addition of the trimming function $\tau(\hat{V})$ helps to trim out those too extreme index values for which the index density $g_v(\hat{V})$ goes to zero too fast. In this way, the selection function $\hat{S}$ 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.

Appendix C: Proof of Result 1

Proof of Result 1. Since the trimming function only trims out a very small fraction of observations from the high probability set, the remainder constitutes the main body of the high probability set, and thus the orders of $\mathbb{E}(S_{Li})$ and $\mathbb{E}(S_{Ri})$ are mainly determined by the core $C(z)$.

$$\begin{aligned}
&\mathbb{E}(S_{Li}) > Prob(x \geq b) \\
&= Prob[P_i(V_i) < N^{-a}e^{-b}] \\
&= Prob\left[V_i < F_\mu^{-1}(N^{-a}e^{-b})\right] \\
&= G_v\left[F_\mu^{-1}(N^{-a}e^{-b})\right] \\
&= F_\mu\left[F_\mu^{-1}(N^{-a}e^{-b})\right] \\
&= N^{-a}e^{-b}
\end{aligned} \tag{A12}$$

And

$$\begin{aligned}
&\mathbb{E}(S_{Ri}) > Prob(y \geq b) \\
&= Prob[P_i(V_i) > 1 - N^{-a}e^{-b}] \\
&= Prob\left[V_i > F_\mu^{-1}(1 - N^{-a}e^{-b})\right] \\
&= 1 - G_v\left[F_\mu^{-1}(1 - N^{-a}e^{-b})\right] \\
&> 1 - F_\mu\left[F_\mu^{-1}(1 - N^{-a}e^{-b})\right] \\
&= N^{-a}e^{-b}
\end{aligned} \tag{A13}$$

where the second equality uses Assumption 2 that $P_i(V_i)$ and $P_i^*(V_i)$ converge to 0 (1) at the same rate on the high probability set. The proofs for $\mathbb{E}(S_{Li}^2)$ and $\mathbb{E}(S_{Ri}^2)$ follow in a similar method.

Appendix D: Proof of Result 2

Proof of Result 2. The proposed estimator in (11) can be rewritten as follows:

$$\hat{\beta}-\beta=+\frac{\beta\frac{1}{N}\sum_{i=1}^{N}(H_i^*-H_i)(\hat{P}_i-\bar{P}_i)S_i+\frac{1}{N}\sum_{i=1}^{N}\varepsilon_i\left((\hat{P}_i-\bar{P}_i)S_i\right)}{\frac{1}{N}\sum_{i=1}^{N}(\hat{P}_i-\bar{P}_i)H_iS_i} \tag{A14}$$

As for any i.i.d observations $\{M_i\}$, $\frac{1}{N}\sum_{i=1}^{N}M_i$ converges to its expectation $\mathbb{E}(M_i)$ at a $\sqrt{N}$ rate, it is interchangeable to study the bias and variance of $\hat{\beta}$ while replacing sample averages with their expectations. Dropping the subscript,

$$|Bias|=\left|\mathbb{E}(\hat{\beta}-\beta)\right|\leq\left|\frac{\beta\mathbb{E}\left[(H^*-H)\left(P-\frac{\mathbb{E}(PS)}{\mathbb{E}(S)}\right)S\right]}{\mathbb{E}\left[\left(P-\frac{\mathbb{E}(PS)}{\mathbb{E}(S)}\right)HS\right]}\right|+\left|\frac{\mathbb{E}\left[\varepsilon\left(P-\frac{\mathbb{E}(PS)}{\mathbb{E}(S)}\right)S\right]}{\mathbb{E}\left[\left(P-\frac{\mathbb{E}(PS)}{\mathbb{E}(S)}\right)HS\right]}\right|$$

$$=\left|\frac{\beta\mathbb{E}\left[(H^*-H)\left(P-\frac{\mathbb{E}(PS)}{\mathbb{E}(S)}\right)S\right]}{\mathbb{E}\left[\left(P-\frac{\mathbb{E}(PS)}{\mathbb{E}(S)}\right)HS\right]}\right| \tag{A15}$$

The second equality uses the assumption of mean zero of the error term ε. We analyze the order of the bias for three cases, depending on the relative rates by which $\mathbb{E}(S_L)$ and $\mathbb{E}(S_R)$ converge to zero: (I) $O\left(\mathbb{E}(S_L)\right)=O\left(\mathbb{E}(S_R)\right)$, (II) $O\left(\mathbb{E}(S_L)\right)>O\left(\mathbb{E}(S_R)\right)$, and (III) $O\left(\mathbb{E}(S_L)\right)<O\left(\mathbb{E}(S_R)\right)$.

For case (I), $\frac{\mathbb{E}(PS)}{\mathbb{E}(S)}\rightarrow\lambda\in(0,1)$ as $N\rightarrow\infty$. To see it,

$$\frac{\mathbb{E}(PS)}{\mathbb{E}(S)}=\frac{\mathbb{E}(PS_L)+\mathbb{E}(PS_R)}{\mathbb{E}(S_L)+\mathbb{E}(S_R)}$$

$$=\frac{O\left(\mathbb{E}(S_R)\right)}{O\left(\mathbb{E}(S_L)\right)+O\left(\mathbb{E}(S_R)\right)}$$

$$=\lambda \tag{A16}$$

where $\lambda\in(0,1)$. The first equality uses the definition of S, and the second uses Result 1 as well as the fact that $P<N^{-a}$ on the left tail and $P>1-N^{-a}$ on the right tail, so that $O\left(\mathbb{E}(PS_L)\right)<O\left(\mathbb{E}(PS_R)\right)=O(\mathbb{E}(S_R))$. Through similar calculations, under Case (II) $\frac{\mathbb{E}(PS)}{\mathbb{E}(S)}\rightarrow 0$ as $N\rightarrow\infty$, and

under Case (III) $\frac{\mathbb{E}(PS)}{\mathbb{E}(S)} \to \lambda \in (0,1)$ as $N \to \infty$. These immediate results can be applied to analyses of Cases (II) and (III).

The order of the numerator in (A15)

$$\begin{aligned}
\left|\mathbb{E}\left[(H^* - H)\left(P - \frac{\mathbb{E}(PS)}{\mathbb{E}(S)}\right)S\right]\right| &= \left|\mathbb{E}\left\{\mathbb{E}\left[(H^* - H)\left(P - \frac{\mathbb{E}(PS)}{\mathbb{E}(S)}\right)S|X\right]\right\}\right| \\
&= \left|\mathbb{E}(P^* - P)\left(P - \frac{\mathbb{E}(PS)}{\mathbb{E}(S)}\right)S\right| \\
&\le \left|\mathbb{E}(P^* - P)\left(P - \frac{\mathbb{E}(PS)}{\mathbb{E}(S)}\right)S_L\right| + \left|\mathbb{E}(P^* - P)\left(P - \frac{\mathbb{E}(PS)}{\mathbb{E}(S)}\right)S_R\right| \\
&\le N^{-a}\left|\mathbb{E}\left(P - \frac{\mathbb{E}(PS)}{\mathbb{E}(S)}\right)S_L\right| + N^{-a}\left|\mathbb{E}\left(P - \frac{\mathbb{E}(PS)}{\mathbb{E}(S)}\right)S_R\right| \\
&\le N^{-a}O\big(\mathbb{E}(S_L)\big) + N^{-a}O\big(\mathbb{E}(S_R)\big) \\
&= N^{-a}O\big(\mathbb{E}(S_R)\big)
\end{aligned} \tag{A17}$$

The first equality uses the law of iterated expectation, and the second uses the definition of P^* and P. The third inequality uses the definition of the selection function S, the fourth holds because of the construction of high probability sets and Assumption 2, and the fifth uses the fact that $\frac{\mathbb{E}(PS)}{\mathbb{E}(S)}$ converges to a constant in Case (I).

The denominator in (A15) is

$$\begin{aligned}
\left|\mathbb{E}(PHS) - \frac{\mathbb{E}(PS)}{\mathbb{E}(S)}\mathbb{E}(HS)\right| &= \left|\mathbb{E}[PS\mathbb{E}(H|X)] - \frac{\mathbb{E}(PS)}{\mathbb{E}(S)}\mathbb{E}[S\mathbb{E}(H|X)]\right| \\
&= \left|\mathbb{E}(P^2S) - \frac{\mathbb{E}(PS)}{\mathbb{E}(S)}\mathbb{E}(PS)\right| \\
&= \left|\mathbb{E}(P^2S_L) + \mathbb{E}(P^2S_R) - \frac{\mathbb{E}(PS)}{\mathbb{E}(S)}[\mathbb{E}(PS_L) + \mathbb{E}(PS_R)]\right| \\
&= O\big(\mathbb{E}(S_R)\big)
\end{aligned} \tag{A18}$$

The first equality uses the law of iterated expectation, the second holds by the definition of P, the last uses Result 1 as well as the fact that $P < N^{-a}$ on the left tail and $P > 1 - N^{-a}$ on the right tail, so that $O(\mathbb{E}(P^2 S_L)) < O(\mathbb{E}(P^2 S_R)) = O(\mathbb{E}(S_R))$ and $O(\mathbb{E}(PS_L)) < O(\mathbb{E}(PS_R)) = O(\mathbb{E}(S_R))$.

With the numerator and denominator above,

$$|Bias| \leq \frac{N^{-a} O(\mathbb{E}(S_R))}{O(\mathbb{E}(S_R))} = O(N^{-a}) \tag{A19}$$

Analyses of cases (II) and (III) lead to the same bias results.

Turning to the variance of the proposed estimator for case (I),

$$Var(\hat{\beta} - \beta) = \frac{Var\left[\beta(H^* - H)\left(P - \frac{\mathbb{E}(PS)}{\mathbb{E}(S)}\right)S + \left(P - \frac{\mathbb{E}(PS)}{\mathbb{E}(S)}\right)S\varepsilon\right]}{N\left[(PHS) - \frac{\mathbb{E}(PS)}{\mathbb{E}(S)}\mathbb{E}(HS)\right]^2} \tag{A20}$$

Analyzing the numerator in (A20),

$$\begin{aligned} &Var\left[\beta(H^* - H)\left(P - \frac{\mathbb{E}(PS)}{\mathbb{E}(S)}\right)S + \left(P - \frac{\mathbb{E}(PS)}{\mathbb{E}(S)}\right)S\varepsilon\right] \\ &= \mathbb{E}\left[\beta(H^* - H)\left(P - \frac{\mathbb{E}(PS)}{\mathbb{E}(S)}\right)S + \left(P - \frac{\mathbb{E}(PS)}{\mathbb{E}(S)}\right)S\varepsilon\right]^2 \\ &- \left\{\mathbb{E}\left[\left[\beta(H^* - H)\left(P - \frac{\mathbb{E}(PS)}{\mathbb{E}(S)}\right)S + \left(P - \frac{\mathbb{E}(PS)}{\mathbb{E}(S)}\right)S\varepsilon\right]\right]\right\}^2 \\ &= \beta^2\mathbb{E}(H^* - H)^2\left(P - \frac{\mathbb{E}(PS)}{\mathbb{E}(S)}\right)^2 S^2 + \mathbb{E}\left(P - \frac{\mathbb{E}(PS)}{\mathbb{E}(S)}\right)^2 S^2\varepsilon^2 \\ &- \beta^2\left[\mathbb{E}(H^* - H)\left(P - \frac{\mathbb{E}(PS)}{\mathbb{E}(S)}\right)\right]^2 \end{aligned} \tag{A21}$$

The first term on the right-hand side is:

$$\mathbb{E}(H^* - H)^2 \left(P - \frac{\mathbb{E}(PS)}{\mathbb{E}(S)}\right)^2 S^2 = \mathbb{E}\left\{\left(P - \frac{\mathbb{E}(PS)}{\mathbb{E}(S)}\right)^2 S^2 \mathbb{E}[(H^* - H)^2 | X]\right\}$$

$$= \mathbb{E}[P_R P^* + P_L(1 - P^*)] \left(P - \frac{\mathbb{E}(PS)}{\mathbb{E}(S)}\right)^2 S_L^2$$

$$+ \mathbb{E}[P_R P^* + P_L(1 - P^*)] \left(P - \frac{\mathbb{E}(PS)}{\mathbb{E}(S)}\right)^2 S_R^2$$

$$\leq N^{-a}\mathbb{E}\left(P - \frac{\mathbb{E}(PS)}{\mathbb{E}(S)}\right)^2 S_L^2 + N^{-a}\mathbb{E}\left(P - \frac{\mathbb{E}(PS)}{\mathbb{E}(S)}\right)^2 S_R^2 \qquad \text{(A22)}$$

The first equality uses the law of iterated expectation. For the second equality, the conditional expectation of $(H^* - H)^2$ given X is calculated by the sum of four cells ($H^* = H = 0, H^* = H = 1$, $H^* = 1$ while $H = 1$, and $H^* = 0$ while $H = 1$) multiplied by their respective conditional probabilities:

$$\mathbb{E}[(H^* - H)^2 | X] = \Pr(H^* = 1, H = 0 | X) + \Pr(H^* = 0, H = 1 | X)$$

$$= \Pr(H = 0 | H^* = 1, X) \Pr(H^* = 1 | X) + \Pr(H = 1 | H^* = 0, X) \Pr(H^* = 0 | X)$$

$$= P_R P^* + P_L(1 - P^*) \qquad \text{(A23)}$$

The inequality in (A22) uses Assumption 2.

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

$$\mathbb{E}\left(P - \frac{\mathbb{E}(PS)}{\mathbb{E}(S)}\right)^2 S^2\varepsilon^2 = \mathbb{E}\left[\left(P - \frac{\mathbb{E}(PS)}{\mathbb{E}(S)}\right)^2 S^2 \mathbb{E}(\varepsilon^2 | X)\right]$$

$$= \sigma^2\mathbb{E}\left(P - \frac{\mathbb{E}(PS)}{\mathbb{E}(S)}\right)^2 S_L^2 + \sigma^2\mathbb{E}\left(P - \frac{\mathbb{E}(PS)}{\mathbb{E}(S)}\right)^2 S_R^2$$

$$= O(\mathbb{E}S_L^2) + O(\mathbb{E}S_R^2) \qquad \text{(A24)}$$

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

$$\left[\mathbb{E}(H^* - H)\left(P - \frac{\mathbb{E}(PS)}{\mathbb{E}(S)}\right)\right]^2 \leq O(N^{-2a}[\mathbb{E}(S_R)]^2) \tag{A25}$$

As $\mathbb{E}(S_R^2) \geq [\mathbb{E}(S_R)]^2$, the second term determines the order of the variance numerator.

The denominator of the variance in (A20) is:

$$N\left[(PHS) - \frac{\mathbb{E}(PS)}{\mathbb{E}(S)}\mathbb{E}(HS)\right]^2 = O(N[\mathbb{E}(S_R)]^2) \tag{A26}$$

With the numerator and denominator of the variance above,

$$Var(\hat{\beta}) = \frac{O\left(\mathbb{E}S_L^2\right) + O\left(\mathbb{E}S_R^2\right)}{O(N[\mathbb{E}(S_R)]^2)}$$

$$= \frac{\mathbb{E}(S_L^2)}{N[\mathbb{E}(S_L)]^2} + \frac{\mathbb{E}(S_R^2)}{N[\mathbb{E}(S_R)]^2} \tag{A27}$$

The last equality comes from $O\left(\mathbb{E}(S_L)\right) = O\left(\mathbb{E}(S_R)\right)$ in case (I). The proof for cases (II) and (III) follow the same reasoning as that of Case (I).

References for Appendices

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.

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

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.

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

Shen, C., 2013. Determinants of health care decisions: Insurance, utilization, and expenditures. *Review of Economics and Statistics*, *95*(1), pp.142-153.

Table 1 Simulation Results for the First Scenario.

True parameters: Intercept (Intcp) = 2, coefficient on X_1 = 1.5, and coefficient on H^* = 3.

	OLS			IV			IV on HPS			
	Intcp	X_1	H^*	Intcp	X_1	H^*	Intcp	X_1	H^*	HPS Param
$\bar{\mu} = 0$										
Mean	2.441	1.559	2.117	2.237	1.522	2.524	2.083	1.501	2.833	0.377
Sd	0.038	0.027	0.053	0.042	0.035	0.058	0.039	0.031	0.057	0.002
Rmse	0.442	0.065	0.885	0.240	0.041	0.480	0.092	0.031	0.176	--
$\bar{\mu} = -1$										
Mean	3.838	1.569	1.137	3.780	1.542	1.353	2.456	1.502	2.546	0.352
Sd	0.046	0.026	0.053	0.057	0.033	0.071	0.122	0.027	0.126	0.007
Rmse	1.838	0.074	1.864	1.781	0.053	1.649	0.472	0.027	0.471	--
$\bar{\mu} = -0.5$										
Mean	3.087	1.565	1.799	2.929	1.530	2.166	2.273	1.501	2.719	0.357
Sd	0.046	0.027	0.055	0.053	0.035	0.065	0.068	0.029	0.077	0.004
Rmse	1.088	0.070	1.203	0.931	0.046	0.836	0.281	0.029	0.291	--
$\bar{\mu} = +0.5$										
Mean	2.114	1.566	1.797	1.904	1.531	2.164	2.008	1.501	2.719	0.372
Sd	0.029	0.027	0.053	0.034	0.035	0.066	0.034	0.029	0.077	0.003
Rmse	0.118	0.071	1.204	0.102	0.046	0.839	0.035	0.029	0.292	--
$\bar{\mu} = +1$										
Mean	2.024	1.568	1.135	1.867	1.542	1.347	1.996	1.501	2.548	0.352
Sd	0.026	0.027	0.051	0.032	0.033	0.070	0.033	0.028	0.128	0.007
Rmse	0.035	0.073	1.866	0.137	0.053	1.654	0.033	0.028	0.470	--
N	3000	3000	3000	3000	3000	3000	3000	3000	3000	3000
Replication	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000

Notes: Intcp, intercept; Sd, standard deviation; Rmse, root-mean-square error; HPS Param, high probability set parameter.

Table 2 Simulation Results for the Second Scenario.

True parameters: Intercept (Intcp) = 2, coefficient on X_1 = 1.5, and coefficient on H^* = 3.

	OLS			IV			IV on HPS			
	Intcp	X_1	H^*	Intcp	X_1	H^*	Intcp	X_1	H^*	HPS Param
$\boldsymbol{No\ misclassification}$: $\boldsymbol{PL = 0, PR = 0}$										
Mean	1.999	1.501	3.000	1.999	1.501	3.000	2.000	1.501	3.000	0.347
Sd	0.026	0.019	0.037	0.033	0.023	0.053	0.044	0.023	0.065	0.003
Rmse	0.026	0.019	0.037	0.033	0.023	0.053	0.044	0.023	0.065	--
$\boldsymbol{Moderate\ misclassification}$: $\boldsymbol{PL = 0.5 * \Phi(5V), PR = 0.5 * [1 - \Phi(5V)]}$										
Mean	2.244	1.531	2.510	2.131	1.511	2.736	2.020	1.501	2.959	0.361
Sd	0.035	0.024	0.048	0.038	0.031	0.056	0.042	0.028	0.060	0.003
Rmse	0.247	0.039	0.493	0.136	0.033	0.270	0.046	0.028	0.073	--
$\boldsymbol{Severe\ misclassification}$: $\boldsymbol{PL = 0.9 * \Phi(5V), PR = 0.9 * [1 - \Phi(5V)]}$										
Mean	2.441	1.559	2.117	2.237	1.522	2.524	2.083	1.501	2.833	0.377
Sd	0.038	0.027	0.053	0.042	0.035	0.058	0.039	0.031	0.057	0.002
Rmse	0.442	0.065	0.885	0.240	0.041	0.480	0.092	0.031	0.176	--
$\boldsymbol{Asymmetric\ misclassification}$: $\boldsymbol{PL = 0.7 * \Phi(5V), PR = 0.2 * [1 - \Phi(5V)]}$										
Mean	2.108	1.526	2.576	2.014	1.512	2.735	2.004	1.501	2.945	0.3597
Sd	0.032	0.024	0.046	0.037	0.031	0.056	0.043	0.027	0.061	0.003
Rmse	0.113	0.035	0.426	0.039	0.033	0.271	0.044	0.027	0.082	--
$\boldsymbol{Asymmetric\ misclassification}$: $\boldsymbol{PL = 0.2 * \Phi(5V), PR = 0.7 * [1 - \Phi(5V)]}$										
Mean	2.314	1.527	2.576	2.248	1.513	2.735	2.050	1.502	2.945	0.360
Sd	0.035	0.023	0.048	0.039	0.030	0.056	0.040	0.027	0.061	0.003
Rmse	0.316	0.035	0.427	0.251	0.033	0.271	0.064	0.027	0.082	--
N	3000	3000	3000	3000	3000	3000	3000	3000	3000	3000
Replication	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000

Notes: Intcp, intercept; Sd, standard deviation; Rmse, root-mean-square error; HPS Param, high probability set parameter.

Table 3 Variable Description.

Variable	Description of Variable
Structural Model: Y_i	
$Hours\ worked$	Total hours respondent works per year at the main and 2nd job.
Structural Model: X_i and H_i	
Age	Age in years.
$Age\ squared$	Square of age.
$Nonlabor\ income$	Total household nonlabor income in thousands of dollars, excluding wages and salaries earned by respondent.
$Race$	= 1 if white; = 0 otherwise.
$Married$	= 1 if married or partnered; = 0 otherwise.
Educational Attainment	
$Less\ than\ high\ school$	Reference.
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's degree; = 0 otherwise
$College\ above$	= 1 if having college degree or greater; = 0 otherwise.
Census Division	
$New\ England$	= 1 if census division of residence is New England; = 0 otherwise.
$Mid\ Atlantic$	= 1 if census division of residence is Middle Atlantic; = 0 otherwise.
$East\ North\ Central$	= 1 if census division of residence is East North Central; = 0 otherwise.
$West\ North\ Central$	= 1 if census division of residence is West North Central; = 0 otherwise.
$South\ Atlantic$	= 1 if census division of residence is South Atlantic; = 0 otherwise.
$East\ South\ Central$	= 1 if census division of residence is East South Central; = 0 otherwise.
$West\ South\ Central$	= 1 if census division of residence is West South Central; = 0 otherwise.
$Mountain$	= 1 if census division of residence is Mountain; = 0 otherwise.
$Pacific\ or\ Not\ US$	Reference.
$Work\ limiting\ health\ conditions$	= 1 if respondent has a health problem that limits the kind or amount of paid work; = 0 otherwise.
$Self-reported\ health\ status$	= 1 if respondent reports "Fair" or "Poor" general health status; = 0 if respondent reports "Excellent," "Very good," or "Good" general health status.
Excluded Variables: Z_i	
$Functional\ limitations\ \#$	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."
$High\ blood\ pressure$	= 1 if respondent has been diagnosed with "high blood pressure or hypertension;" = 0 otherwise.
$Diabetes$	= 1 if respondent has been diagnosed with "diabetes or high blood sugar;" = 0 otherwise.
$Cancer$	= 1 if respondent has been diagnosed with "cancer or a malignant tumor of any kind except skin cancer;" = 0 otherwise.
$Lung\ problems$	= 1 if respondent has been diagnosed with "chronic lung disease except asthma such as chronic bronchitis or emphysema;" = 0 otherwise.
$Heart\ problems$	= 1 if respondent has been diagnosed with "heart attack, coronary heart disease, angina, congestive heart failure, or other heart problems;" = 0 otherwise.

$Stroke$	= 1 if respondent has been diagnosed with "stroke or transient ischemic attack (TIA);" = 0 otherwise.
$Psychiatric\ problems$	= 1 if respondent has been diagnosed with "emotional, nervous, or psychiatric problems;" = 0 otherwise.

Table 4 Descriptive Statistics.

Variable	Men	Women
Panel A: Basic Demographics		
Age	56.1	55.5
	(3.3)	(3.7)
$45-50$ (%)	4.8	9.9
$51-55$ (%)	37.6	37.5
$56-61$ (%)	57.6	52.6
$White$ (%)	62.0	60.6
$Nonlabor\ income$	47.6	52.2
$(in\ thousands\ of\ dollars)$	(90.3)	(96.9)
$Married$ (%)	76.6	66.3
$Less\ than\ high\ school$ (%)	15.8	14.3
GED (%)	6.2	6.1
$High\ school$ (%)	25.1	24.8
$Some\ college$ (%)	28.5	29.9
$College\ and\ above$ (%)	24.4	24.8
Panel B: Labor Supply		
$Hours\ worked$	1567.0	1181.2
	(1185.1)	(1075.5)
$Working\ for\ pay$ (%)	70.8	62.4
$Working\ on\ 2nd\ job\ conditional\ on\ working\ for\ pay$ (%)	15.8	13.4
$Hours\ worked\ on\ 1st\ job\ conditional\ on\ working\ for\ pay$	2144.1	1844.2
	(699.9)	(686.6)
	(N = 2113)	(N = 2545)
$Hours\ worked\ on\ 2nd\ job\ conditional\ on\ working\ on\ 2nd\ job$	530.3	459.7
	(518.9)	(483.7)
	(N = 307)	(N = 297)
Panel C: Health Measures		
$Work\ limiting\ health\ conditions$ (%)	24.6	26.9
$Self-reported\ health\ status$ (%)	24.3	27.7
$Functional\ limitations$ #	1.7	2.3
	(2.4)	(2.6)
$High\ blook\ pressure$ (%)	49.7	47.6
$Diabetes$ (%)	19.9	19.1
$Cancer$ (%)	5.2	9.0
$Lung\ problems$ (%)	5.6	9.0
$Heart\ problems$ (%)	14.6	12.4
$Stroke$ (%)	3.9	4.0
$Psychiatric\ problems$ (%)	14.4	24.2
N	2995	4089

Notes: The sample is from the HRS 2012 wave. Figures in parentheses are standard errors.

Table 5 Frequencies of Reports on Two Subjective Health Measures.

Men

	Good, very good, or excellent self − reported health	*Fair or poor self − reported health*	*Total*
Work − limiting health condition (no)	1954	303	2257
Work − limiting health condition (yes)	314	424	738
Total	2268	727	2995

Women

	Good, very good, or excellent self − reported health	*Fair or poor self − reported health*	*Total*
Work − limiting health condition (no)	2544	444	2988
Work − limiting health condition (yes)	413	688	1101
Total	2957	1132	4089

Notes: The sample is from the HRS 2012 wave. Figures in each cell indicate the number of observations in corresponding categories.

Table 6 Labor Supply Effects of Self-reported Health Status for Men.

	OLS	IV	IV on HPS
$Fair\ or\ poor$	-734.4	-2014.7	-2026.6
$self-reported\ health$	(-828.2, -640.6)	(-2232.6, -1796.7)	(-2588.5, -1464.8)
Age	498.3	498.4	475.3
	(221.3, 775.3)	(27.9, 968.9)	(209.8, 740.8)
Age^2	-4.9	-4.7	-4.6
	(-7.4, -2.3)	(-8.9, -.5)	(-7.0, -2.2)
$Nonlabor\ income$	.5	-1.9	-2.0
$(in\ thousands\ of\ dollars)$	(.0, .9)	(-3.0, -.7)	(-3.0, -1.0)
$White$	231.3	186.5	173.8
	(148.1, 314.5)	(90.2, 282.8)	(92.8, 254.8)
$Married$	423.8	395.2	375.4
	(330.6, 517.0)	(285.0, 505.4)	(282.6, 468.2)
$Less\ than\ high\ school$	Reference	Reference	Reference
GED	-67.8	-202.4	-225.0
	(-250.6, 115.1)	(-413.3, 8.6)	(-405.0, -45.1)
$High\ school$	178.4	-65.9	-123.5
	(51.2, 305.5)	(-216.9, 85.2)	(-252.5, 5.5)
$Some\ college$	250.3	-7.8	-78.1
	(124.7, 376.0)	(-159.6, 144.0)	(-208.1, 51.9)
$College\ and\ above$	428.1	95.1	44.6
	(294.7, 561.5)	(-69.7, 260.0)	(-97.1, 186.4)
N	2995	2995	2995

Notes:

1. Figures in parentheses are 95% confidence interval.
2. The OLS, IV, and IV on HPS regressions control for eight dummy variables of census division of residence described in Table 1, including New England, Middle Atlantic, East North Central, West North Central, South Atlantic, East South Central, West South Central, and Mountain (Pacific/Not US as the reference group).

Table 7 Labor Supply Effects of Self-reported Health Status for Women.

	OLS	IV	IV on HPS
$Fair\ or\ poor$	-615.2	-1577.1	-1900.4
$self-reported\ health$	(-687.7, -542.7)	(-1726.0, -1428.1)	(-2322.0, -1478.7)
Age	508.1	439.0	506.3
	(312.8, 703.4)	(169.8, 708.3)	(316.9, 695.6)
Age^2	-4.9	-4.2	-4.8
	(-6.6, -3.1)	(-6.6, -1.7)	(-6.5, -3.0)
$Nonlabor\ income$	-.9	-2.2	-2.0
$(in\ thousands\ of\ dollars)$	(-1.3, -.6)	(-3.0, -1.3)	(-2.6, -1.3)
$White$	119.9	98.3	53.7
	(53.9, 186.0)	(24.6, 172.0)	(-11.5, 118.9)
$Married$	-3.9	-33.7	-30.4
	(-74.6, 66.8)	(-117.6, 50.2)	(-103.6, 42.7)
$Less\ than\ high\ school$	Reference	Reference	Reference
GED	229.9	100.4	62.6
	(81.2, 378.7)	(-64.4, 265.2)	(-84.1, 209.3)
$High\ school$	377.3	149.6	155.3
	(272.3, 482.2)	(30.5, 268.6)	(48.1, 262.5)
$Some\ college$	465.5	206.7	203.1
	(363.1, 568.0)	(87.8, 325.5)	(96.6, 309.7)
$College\ and\ above$	609.0	272.0	271.5
	(499.2, 718.7)	(140.8, 403.1)	(154.7, 388.2)
N	4089	4089	4089

Notes:

1. Figures in parentheses are 95% confidence interval.
2. The OLS, IV, and IV on HPS regressions control for eight dummy variables of census division of residence described in Table 1, including New England, Middle Atlantic, East North Central, West North Central, South Atlantic, East South Central, West South Central, and Mountain (Pacific/Not US as the reference group).

Table 8 Labor Supply Effects of Work-limiting Health Condition for Men.

	OLS	IV	IV on HPS
$Work - limiting$	-1288.9	-1666.6	-1659.0
$health\ condition\ (yes)$	(-1372.6, -1205.2)	(-1797.0, -1536.3)	(-1888.4, -1429.6)
Age	422.8	471.8	421.0
	(170.8, 674.8)	(84.4, 859.2)	(156.5, 685.5)
Age^2	-4.1	-4.5	-4.1
	(-6.4, -1.8)	(-7.9, -1.0)	(-6.5, -1.7)
$Nonlabor\ income$	.6	-.6	.7
$(in\ thousands\ of\ dollars)$	(.1, 1.0)	(-1.6, .3)	(.3, 1.1)
$White$	227.4	231.4	205.5
	(151.8, 303.0)	(152.6, 310.2)	(126.2, 284.7)
$Married$	284.4	262.6	248.6
	(198.9, 369.9)	(170.9, 354.3)	(156.1, 341.1)
$Less\ than\ high\ school$	Reference	Reference	Reference
GED	120.5	166.9	163.1
	(-45.7, 286.6)	(-6.0, 339.7)	(-14.0, 340.3)
$High\ school$	274.8	260.4	245.6
	(160.4, 389.1)	(141.8, 379.1)	(127.0, 364.1)
$Some\ college$	321.0	318.9	298.7
	(208.4, 433.7)	(201.1, 436.6)	(181.4, 416.0)
$College\ and\ above$	410.0	368.3	351.1
	(290.4, 529.6)	(241.0, 495.7)	(222.1, 480.0)
N	2995	2995	2995

Notes:

1. Figures in parentheses are 95% confidence intervals.
2. The OLS, IV, and IV on HPS regressions control for eight dummy variables of census division of residence described in Table 1, including New England, Middle Atlantic, East North Central, West North Central, South Atlantic, East South Central, West South Central, and Mountain (Pacific/Not US as the reference group).

Table 9 Labor Supply Effects of Work-limiting Health Condition for Women.

	OLS	IV	IV on HPS
$Work - limiting$	-1059.2	-1300.8	-1323.1
$health\ condition\ (yes)$	(-1125.2, -993.3)	(-1402.2, -1199.3)	(-1548.9, -1097.2)
Age	533.4	525.4	495.2
	(352.4, 714.4)	(293.6, 757.2)	(304.8, 685.7)
Age^2	-5.0	-4.9	-4.7
	(-6.7, -3.4)	(-7.0, -2.8)	(-6.4, -2.9)
$Nonlabor\ income$	-.9	-1.2	-.9
$(in\ thousands\ of\ dollars)$	(-1.2, -.6)	(-2.0, -.5)	(-1.2, -.6)
$White$	135.4	146.7	115.3
	(74.3, 196.6)	(83.6, 209.7)	(51.1, 179.4)
$Married$	-58.9	-68.3	-70.5
	(-124.5, 6.7)	(-140.8, 4.2)	(-139.7, -1.4)
$Less\ than\ high\ school$	Reference	Reference	Reference
GED	323.2	344.3	287.2
	(185.7, 460.6)	(204.0, 484.6)	(142.9, 431.5)
$High\ school$	383.5	351.0	344.2
	(287.3, 479.7)	(251.8, 450.1)	(242.4, 446.0)
$Some\ college$	523.2	509.6	516.3
	(430.1, 616.3)	(413.3, 605.9)	(418.7, 613.9)
$College\ and\ above$	599.2	559.1	565.6
	(499.7, 698.7)	(453.7, 664.4)	(460.1, 671.2)
N	4089	4089	4089

Notes:

1. Figures in parentheses are 95% confidence intervals.
2. The OLS, IV, and IV on HPS regressions control for eight dummy variables of census division of residence described in Table 1, including New England, Middle Atlantic, East North Central, West North Central, South Atlantic, East South Central, West South Central, and Mountain (Pacific/Not US as the reference group).