

Disability and Employment: Reevaluating the Evidence in Light of Reporting Errors

Brent KREIDER and John V. PEPPER

Measurement error in health and disability status has been widely accepted as a central problem in social science research. Long-standing debates about the prevalence of disability, the role of health in labor market outcomes, and the influence of federal disability policy on declining employment rates have all emphasized issues regarding the reliability of self-reported disability. In addition to random error, inaccuracy in survey datasets may be produced by a host of economic, social, and psychological factors that can lead respondents to misreport work capacity. We develop a nonparametric foundation for assessing how assumptions on the reporting error process affect inferences on the employment gap between the disabled and nondisabled. Rather than imposing the strong assumptions required to obtain point identification, we derive sets of bounds that formalize the identifying power of primitive nonparametric assumptions that appear to share broad consensus in the literature. Within this framework, we introduce a finite-sample correction for the analog estimator of the monotone instrumental variable (MIV) bound. Our empirical results suggest that conclusions derived from conventional latent variable reporting error models may be driven largely by ad hoc distributional and functional form restrictions. We also find that under relatively weak nonparametric assumptions, nonworkers appear to systematically overreport disability.

KEY WORDS: Corrupt sampling; Finite-sample bias correction; Measurement error; Monotone instrumental variable; Nonparametric bounds.

1. INTRODUCTION

Measuring employment rates among disabled persons has been a matter of intense concern for policy analysts, especially since the passage of the Americans With Disabilities Act (ADA) in 1990. Most studies rely on self-reported health information to analyze relationships between employment and disability. Burkhauser, Daly, Houtenville, and Nigras (2002), for example, made extensive use of survey questions of the general form: "Does a health impairment limit the kind or amount of work you can perform?" Evidence from these studies suggests that employment rates between nondisabled and disabled persons have widened substantially since the introduction of the ADA. Yet reporting errors in disability status contaminate estimates of conditional employment rates. Citing "grave concerns about the accuracy and reliability of widely disseminated information about employment rates among people with disabilities," the National Council on Disability (NCD) (2002) warned that disability measurement error "could lead to ineffective or even dangerous public policy decisions."

In this article we develop a nonparametric foundation for assessing how different assumptions on the reporting error process affect inferences on the employment gap between disabled and nondisabled persons. Measurement error in health status has been accepted as a central problem in social science research (e.g., Institute of Medicine 2002; U.S. General Accounting Office 1997). More than 20 years ago, Anderson and Burkhauser (1984) characterized the measurement of work capacity in survey datasets as "the major unsettled issue in the empirical literature on the labor supply of older workers," and

the debates have only intensified over time. Prominent debates about the prevalence of disability, the role of health in labor market decisions, and the influence of Social Security Disability Insurance (SSDI) policy on declining labor force participation rates have all emphasized issues regarding the reliability of self-reported disability information. Bound (1991) provided an illuminating analyses of the econometric issues surrounding disability reporting errors.

In particular, there is widespread concern about the accuracy of self-reported disability status in survey datasets. Whereas most studies treat self-reports of work limitation as fully accurate, the literature encompasses a wide range of views on reporting errors. Some researchers contend that disability reporting is largely reliable (e.g., Stern 1989; Dwyer and Mitchell 1999; Benítez-Silva, Buchinsky, Chan, Cheidvasser, and Rust 2004), whereas others contend that strong economic and psychological incentives to misreport disability, coupled with potential difficulties with interpreting the survey questions, make self-reports nearly devoid of content (e.g., Myers 1982; Bove 1993; Hale 2001). The psychology literature discusses the potential medical role of "negative affectivity" in respondents' self-assessments of disability status (see, e.g., Watson and Clark 1984). The unknown reliability of proxy or imputed responses raises further concerns (Lee, Mathiowetz, and Tourangeau 2004).

Others take middle-ground positions by formally treating self-reports as reliable for members of certain subpopulations but not others. For example, many researchers have emphasized that eligibility for disability transfers is specifically tied to diminished work capacity. Bound and Burkhauser (1999, p. 3446) suggested the possibility that "those who apply for SSDI and especially those who are awarded benefits tend to exaggerate the extent of their work limitations." More generally, many have suggested that the threshold for claiming disability may be lower for those who find themselves out of the labor force, either voluntarily or involuntarily (e.g., Kerkhofs and Lindeboom 1995; O'Donnell 1998; Kreider 1999, 2000).

Brent Kreider is Associate Professor, Department of Economics, Iowa State University, Ames, IA 50011 (E-mail: bkreider@iastate.edu). John V. Pepper is Associate Professor, Department of Economics, University of Virginia, Charlottesville, VA 22904 (E-mail: jvpepper@virginia.edu). This research was supported in part by the Joint Center for Poverty Research, the Sandell Grant Program, the W. E. Upjohn Institute for Employment Research, and the Bankard Fund for Political Economy. The authors received valuable comments from two anonymous referees, Christopher Bollinger, John Bound, Helle Bunzel, Charles Manski, Francesca Molinari, Shakeeb Khan, Steven Stern, Justin Tobias, and seminar participants at Cornell, Pennsylvania, Michigan, Iowa State, Syracuse, and Virginia. They also benefited from discussion at meetings of the Econometric Society, Allied Social Science Association, International Health Economics Association, Northwestern/University of Chicago Joint Center for Poverty Research, Midwest Economic Association, and Southern Economic Association.

© 2007 American Statistical Association
Journal of the American Statistical Association
June 2007, Vol. 102, No. 478, Applications and Case Studies
DOI 10.1198/016214506000000997

Departing from the existing disability and employment literature, we do not focus on providing point estimates of the employment gap between the disabled and nondisabled. Instead, we derive analytic bounds that allow us to assess the identifying power of different assumptions on the disability reporting error process within a unifying methodological framework. We estimate conditional employment probabilities using information on respondents in the Health and Retirement Study (HRS) and the Survey of Income and Program Participation (SIPP). After describing the data in Section 2, we formalize the identification problem created by arbitrary misreporting in Section 3. New methodological results allow us to assess the sensitivity of the identification problem to variation in the nature and degree of corruption in a regressor—namely, disability status. Our approach is similar in spirit to that of Horowitz and Manski (1995), who assessed the problem of identifying a marginal distribution in corrupt data. We extend their approach to allow for corruption of a binary regressor in a conditional distribution. In this setting, we show how the classical assumption of exogenous or “nondifferential” measurement error considered by Aigner (1973) and Bollinger (1996) can be used to tighten the upper bound on the employment gap.

Section 4 introduces the notion of partial verification of reports within particular observed subgroups (e.g., workers or disability beneficiaries). By allowing for some classification errors within partially verified subgroups, we depart from both the parametric disability literature (e.g., Kreider 1999; McGarry 2004) and nonparametric bounds literature (e.g., Horowitz and Manski 1998; Dominitz and Sherman 2004), which assume fully accurate reporting within verified subgroups. Section 5 considers the identifying power of monotonicity restrictions that link employment and disability to certain covariates, such as age or the likelihood of being approved for disability benefits. Within this framework, we introduce a nonparametric method for correcting the finite-sample bias of the analog estimator of Manski and Pepper’s (2000) monotone instrumental variables (MIV) bound. Under relatively weak assumptions, our results support contentions in the literature that nonworkers systematically overreport disability. Section 6 concludes.

2. THE DATA

Our main analysis uses data from the 1992–1993 HRS and the 1996 SIPP. Providing detailed information about health and

disability, work history, and participation in government transfer programs using a panel design, the HRS and SIPP are perhaps the two most important data sources for studying the effects of health status and public policy on work outcomes. In Section 5 we further check the robustness of our results using the publicly released 5% extract from the 2000 Decennial Census of Population.

The HRS is a nationally representative panel survey of households whose heads were nearing retirement age (age 51–61 years) in 1992–1993. We use self-reported health and labor force participation information from all 12,503 respondents age 40 and older. We also record other characteristics, such as gender, occupation, race, years of schooling, marital status, the receipt of government assistance for a disability, whether the responses came from a proxy respondent, and various functional limitations and physician-diagnosed health conditions. As part of our identification strategy, some of our analysis incorporates reported health and employment information from the second wave, which was conducted two years after the first wave. We also record whether the respondent died before the second wave was conducted.

The SIPP is a nationally representative longitudinal survey covering the U.S. civilian noninstitutionalized population. We use data from the first wave of the 1996 panel, a nationally representative sample of 36,800 households. Because respondents older than 69 were not asked about work limitations, we restrict the SIPP sample to the 29,807 individuals age 40–69.

Table 1 displays selected means and standard deviations. In the HRS, 21.9% of the sample responded that an impairment limits or precludes paid work, and 66.3% report currently working for pay. The corresponding values in the SIPP data are 18.8% and 69.5%. The differences between the two surveys primarily reflect differences in the surveyed age distributions (see the last column in Table 1).

Table 2 presents labor force participation rates by self-assessed work limitation and age. In the HRS, the employment rates are .294 among those reporting to be disabled and .766 for those reporting to be nondisabled. Thus the difference in employment rates by reported disability status—the reported employment gap—is $-.472$. The corresponding reported employment gap in the SIPP is $-.482$.

3. THE IDENTIFICATION PROBLEM

To infer the employment gap between disabled and nondisabled persons, we consider what self-reports reveal about true

Table 1. Means and Standard Deviations

	HRS (n = 12,503)		SIPP (n = 29,807)		
	Mean	Standard deviation	Mean	Standard deviation	Weighted mean*
Work-limited (self-reported)	.216	.411	.188	.391	.218
Disability precludes work	.094	.291	.112	.315	.135
Yes to either of the above (X = 1)	.219	.414	.188	.391	.218
Labor force participant (L = 1)	.663	.473	.695	.460	.686
Current receipt of disability income	.101	.301	.051	.220	.066
Age	56.0	5.26	54.9	8.62	55.9
Years of schooling	12.0	3.27	12.8	3.24	12.6
High school graduate	.707	.455	.827	.377	.805
College graduate	.175	.380	.228	.368	.211
Nonwhite race	.280	.449	.162	.382	.162

*Weighted to match the HRS age distribution.

Table 2. Conditional Employment Probabilities by Self-Reported Disability Status

Age, years	All		Work limitation (self-reported)		No work limitation (self-reported)	
	HRS (n = 12,503)	SIPP (n = 29,807)	HRS (n = 2,742)	SIPP (n = 5,597)	HRS (n = 9,761)	SIPP (n = 24,210)
40–49	.766	.838	.487	.440	.819	.896
50–54	.737	.785	.341	.368	.828	.878
55–59	.662	.670	.286	.284	.779	.792
60–64	.555	.462	.245	.185	.662	.570
65–69	.316	.257	.119	.136	.409	.305
70+	.224		.087		.287	
All	.663	.695	.294	.304	.766	.786
Weighted*		.680		.290		.789

*Weighted to match the HRS age distribution.

disability as measured by current social norms or the particular research question of interest. Clearly, survey designers have an expectation that respondents can place questions about work limitation in a reasonable social context. Some respondents may use thresholds different than those implied by the social norms, but the data do not reveal these respondents.

To evaluate the implications of invalid response in corrupt data, we introduce notation that distinguishes between self-reports and accurate reports. Let $L = 1$ indicate that the respondent is employed, and $L = 0$ otherwise. Similarly, let $X = 1$ indicate that the respondent reports being limited in the ability to work, and let $W = 1$ indicate that the individual is truly limited in the ability to work relative to social norms (or other specified criteria). Finally, let Z indicate whether a respondent provides accurate information, with $Z = 1$ if $W = X$ and $Z = 0$ otherwise. We are interested in learning how the employment rate varies by true disability status,

$$\beta = P(L = 1|W = 1) - P(L = 1|W = 0). \quad (1)$$

The data reveal $P(L = 1|X)$ but not $P(L = 1|W)$; therefore, β is not identified by the sampling process. To see this, we can decompose the employment rate among truly disabled persons as

$$\begin{aligned} P(L = 1|W = 1) &= \frac{P(L = 1, W = 1)}{P(W = 1)} \\ &= (P(L = 1, X = 1) + P(L = 1, X = 0, Z = 0) \\ &\quad - P(L = 1, X = 1, Z = 0)) \\ &\quad / (P(X = 1) + P(X = 0, Z = 0) - P(X = 1, Z = 0)). \quad (2) \end{aligned}$$

The data identify the fraction who self-report disability, $P(X = 1)$, and the joint probability of being employed and claiming to be disabled, $P(L = 1, X = 1)$, but they do not reveal the distribution of accurate reporters. Some unknown fraction of respondents, $P(X = 1, Z = 0)$, inaccurately report being disabled (false-positives), whereas others, $P(X = 0, Z = 0)$, inaccurately report being nondisabled (false-negatives). In the absence of restrictions on misreporting, the data are uninformative; we know only that the conditional employment rate lies between 0 and 1.

3.1 Nondifferential Classification Errors

The classical prescription used to address these identification problems is to assume that the reporting error process is exoge-

nous. In particular, suppose that reporting errors are independent of the employment outcome conditional on true disability status,

$$P(X = 1|W) = P(X = 1|W, L). \quad (3)$$

This type of “nondifferential” classification error has been studied by Aigner (1973) and Bollinger (1996). When the independence assumption (3) holds, Bollinger’s theorem 1 applied to a binary outcome can be used to show that β is bounded away from 0 (in this case, from above) by the reported employment gap, $P(L = 1|X = 1) - P(L = 1|X = 0)$ (< 0). This independence assumption clearly confers strong identifying power. Using the HRS data, for example, β is estimated to be less than $-.472$, reflecting well-known attenuation bias associated with random measurement error.

Although the nondifferential measurement error assumption is powerful, Bound, Brown, and Mathiowetz (2001, p. 3725) noted that the assumption is strong and often implausible. In our context, the assumption requires that, conditional on true disability status, unemployed respondents are no more likely to report being disabled than employed respondents. This assumption effectively rules out, for example, the possibilities that labor market outcomes affect respondents’ perceptions of their disability status or that employment outcomes may be associated with perceived disability status in addition to true disability status. We proceed under the premise that the assumption of nondifferential errors (3) may not hold in this application.

3.2 Lower-Bound Accurate Reporting Rate

To characterize the identification problem in the absence of the nondifferential classification errors assumption, it is useful to consider what can be learned with a known lower bound on the fraction of respondents that accurately report disability status. In particular, suppose that

$$P(Z = 1) \geq \nu, \quad (4)$$

where ν is a known lower bound on the accurate reporting rate. Horowitz and Manski (1995) applied this degree assumption when assessing the problem of identifying a marginal distribution in corrupt data.

By varying the value of ν , we can effectively consider the wide range of views characterizing the debate on inaccurate reporting. Those willing to assume that all reports are accurate can set $\nu = 1$, in which case the sampling process identifies the

conditional employment rates. Those believing that all reports are potentially inaccurate can set $v = 0$, in which case the sampling process is uninformative. Middle-ground positions can be evaluated by setting v between 0 and 1.

The lower bound in (4) implies restrictions on the unknown joint distributions in (2). In particular, if the degree of misreporting is no greater than some known fraction, $1 - v$, then the following sharp “degree bounds” apply (see the App. for a proof).

Proposition 1. Let $P(Z = 1) \geq v$. Then $P(L = 1|W = 1)$ is bounded sharply as follows:

$$\frac{P(L = 1, X = 1) - \delta}{P(X = 1) - 2\delta + (1 - v)} \leq P(L = 1|W = 1) \leq \frac{P(L = 1, X = 1) + \gamma}{P(X = 1) + 2\gamma - (1 - v)}, \quad (5)$$

where

$$\delta = \begin{cases} \min\{(1 - v), P(L = 1, X = 1)\} & \text{if } P(L = 1, X = 1) - P(L = 0, X = 1) - (1 - v) \leq 0 \\ \max\{0, (1 - v) - P(L = 0, X = 0)\} & \text{otherwise} \end{cases}$$

and

$$\gamma = \begin{cases} \min\{(1 - v), P(L = 1, X = 0)\} & \text{if } P(L = 1, X = 1) - P(L = 0, X = 1) + (1 - v) \leq 0 \\ \max\{0, (1 - v) - P(L = 0, X = 1)\} & \text{otherwise.} \end{cases}$$

To estimate the bounds in Proposition 1, we replace the population probabilities with sample analogs. Bounds for $P(L = 1|W = 0)$ are obtained by replacing $X = 1$ with $X = 0$ and vice versa. An upper (lower) bound on β can be found by subtracting the lower (upper) bound on $P(L = 1|W = 0)$ from the upper (lower) bound on $P(L = 1|W = 1)$. Although these bounds on β are intuitive and simple to compute, they are not sharp. In the Appendix we show how the constraint $P(Z = 1) \geq v$ places further restrictions on β and formalize sharp bounds.

Note that when the lower-bound fraction of accurate reporters is relatively small, the bounds on the conditional employment rates are uninformative. For example, when the degree of misreporting can exceed the fraction of respondents reporting to be disabled workers, $(1 - v) \geq P(L = 1, X = 1)$, the lower bound on $P(L = 1|W = 1)$ is 0. After all, despite self-reports to the contrary, all of these respondents may be nondisabled. Similarly, the upper bound is 1 when $(1 - v) \geq P(L = 0, X = 1)$.

The striking feature of the estimates from the HRS sample is that these bounds are uninformative across a wide range of values for v . When $v = 0$, the employment gap can lie anywhere between -1 and 1 . The HRS data remain uninformative unless it is known that the accurate reporting rate exceeds .41, and the lower bound remains at -1 unless v exceeds .82. The sign of β is identified as negative (i.e., the data reveal that disabled persons are less likely to work than nondisabled persons) only if at least 88% of the respondents are known to provide accurate reports. Results are similar for the SIPP data. Under weak assumptions on the degree of accurate reporting, the data provide only modest information on the true conditional employment rates.

4. NONPARAMETRIC PARTIAL VERIFICATION MODEL

Concerns about misreporting focus primarily on financial and social incentives for certain types of respondents to exaggerate the extent of lost work capacity. First, eligibility into some government assistance programs (e.g., SSDI) is contingent on being sufficiently work-impaired. In addition to monthly cash benefits, Supplemental Security Income (SSI) beneficiaries are immediately eligible for Medicaid benefits, and SSDI beneficiaries become eligible for Medicare benefits after a 2-year waiting period. Second, some people may feel social pressure to participate in the labor force until normal retirement age unless their ability to work is impaired (see Bound 1991). Those who find themselves out of work (or prefer not to work) may feel more compelled to claim that a functional limitation (e.g., difficulty climbing stairs) interferes with the ability to work.

Short of assuming that all respondents provide accurate self-reports, several studies have identified the true disability rate by combining distributional restrictions with assumptions that certain types of respondents provide accurate reports. The existing literature provides a number of restrictions (see, e.g., Bound and Burkhauser 1999). Kreider (1999) and McGarry (2004), for example, assumed that workers provide fully accurate responses, remaining agnostic about the reports from nonworkers. In the spirit of this literature, we evaluate what can be learned about the conditional employment rates given prior information on the degree of misreporting within four observed subgroups: (a) disability beneficiaries (10% in the HRS), (b) respondents who claimed no disability in the second wave of the survey despite being out of the labor force (27%), (c) respondents who were gainfully employed (66%), and (d) respondents who claimed no work limitation in the current wave (78%). For the HRS and SIPP, 94% and 93% of the respondents satisfied at least one of these criteria. Although members of these groups may face little incentive to misreport, we allow for the possibility of some reporting errors within verified groups. Note that given the high thresholds and restrictive screening processes used in government disability programs, verifying a work limitation among beneficiaries is not tantamount to assuming that the limitation is sufficiently severe to warrant eligibility into the program.

Formally, let $Y = 1$ indicate that a respondent belongs to a “verified” subgroup, with $Y = 0$ otherwise. At least some fraction v_y of the self-reports in such groups are assumed accurate: $P(Z = 1|Y = 1) \geq v_y$. No other restrictions are imposed on the error process within verified groups, and no prior information exists for the error process in the unverified groups. Under these assumptions, we derive the following proposition (see the App. for a proof).

Proposition 2. Let $P(Z = 1|Y = 1) \geq v_y$. Then $P(L = 1|W = 1)$ is bounded sharply as follows:

$$\frac{P(L = 1, X = 1, Y = 1) - \delta}{P(X = 1, Y = 1) + P(L = 0, Y = 0) - 2\delta + (1 - v_y)P(Y = 1)} \leq P(L = 1|W = 1) \leq \frac{P(L = 1, X = 1, Y = 1) + P(L = 1, Y = 0) + \gamma}{P(X = 1, Y = 1) + P(L = 1, Y = 0) + 2\gamma - (1 - v_y)P(Y = 1)},$$

where

$$\delta = \begin{cases} \min\{(1 - v_y)P(Y = 1), P(L = 1, X = 1)\} \\ \text{if } \alpha \leq 0 \\ \max\{0, (1 - v_y)P(Y = 1) - P(L = 0, X = 0, Y = 1)\} \\ \text{otherwise;} \end{cases}$$

$$\gamma = \begin{cases} \min\{(1 - v_y)P(Y = 1), P(L = 1, X = 0)\} \\ \text{if } \alpha' \leq 0 \\ \max\{0, (1 - v_y)P(Y = 1) - P(L = 0, X = 1, Y = 1)\} \\ \text{otherwise;} \end{cases}$$

$$\alpha = P(L = 1, X = 1, Y = 1) - P(L = 0, X = 1, Y = 1) - P(L = 0, Y = 0) - (1 - v_y)P(Y = 1);$$

and

$$\alpha' = P(L = 1, X = 1, Y = 1) - P(L = 0, X = 1, Y = 1) + P(L = 1, Y = 0) + (1 - v_y)P(Y = 1).$$

As before, bounds for $P(L = 1|W = 0)$ are obtained by replacing $X = 1$ with $X = 0$, and vice versa, and bounds on β can be computed by subtracting the appropriate bound on $P(L = 1|W = 0)$ from the appropriate bound on $P(L = 1|W = 1)$. Given the verification in Proposition 2, these bounds on β are sharp. Using similar verification assumptions, Kreider and Pepper (2006) derived sharp bounds on the marginal distribution $P(W = 1)$. Building on Propositions 1 and 2, Kreider and Hill (2006) considered the case of a continuous outcome.

Under the assumption of fully accurate reporting within verified subgroups, $v_y = 1$, the bounds of Proposition 2 simplify. In particular, it follows that if $v_y = 1$, then we have the following result:

Corollary 1. If $Y = 1 \Rightarrow Z = 1$ (full verification), then

$$\frac{P(L = 1, X = 1, Y = 1)}{P(X = 1, Y = 1) + P(L = 0, Y = 0)} \leq P(L = 1|W = 1) \leq \frac{P(L = 1, X = 1, Y = 1) + P(L = 1, Y = 0)}{P(X = 1, Y = 1) + P(L = 1, Y = 0)}. \quad (6)$$

The width of these bounds depends on the joint distribution of the observed random variables, $\{L, X, Y\}$. An alternative

derivation of the result in this corollary is provided in the bound for regressor censoring of Horowitz and Manski (1998).

Empirical results for the HRS and SIPP data are presented in Table 3. Columns A and C provide results for the degree bounds (Prop. 1) for selected values of v , and columns B and D present results for the verification bounds with $v_y = 1$. Bootstrapped 90% confidence intervals around the point estimates of the bounds are computed based on the bias-corrected percentile method (Efron and Tibshirani 1993) using 1,000 pseudosamples.

As noted earlier, the estimated Proposition 1 degree bounds are uninformative across a wide range of values for v . In contrast, the Proposition 2 verification bounds are always informative for $P(Y = 1) > 0$. Nonetheless, the sign of β remains unidentified unless responses for all four groups are verified. In that case, β is estimated to lie within $[-.472, -.298]$ for the HRS and within $[-.482, -.255]$ for the SIPP. For both datasets, these bounds are 39 points narrower than the corresponding degree bounds.

Intuitively, the bounds widen if respondents in verified subgroups may misreport. Nevertheless, for a sufficiently large v_y , partial verification always improves on the Proposition 1 bounds in (5). Consider, for example, the Proposition 1 bound where $v = .10$, the fraction of disability beneficiaries. If we assume partial verification of beneficiaries alone, then the HRS upper bound is improved if even 27% of beneficiaries are known to provide valid responses.

The existing empirical literature assumes fully accurate self-reports within verified groups and imposes strong structure on the nature of reporting errors within remaining groups (e.g., independence between errors and outcomes). When we relax the usual distributional and functional form restrictions and isolate the identifying power of the verification assumptions, there remains much uncertainty about true conditional employment rates unless nearly all respondents are known to provide accurate disability reports.

5. MONOTONICITY RESTRICTIONS

We next formalize the notion that the employment rate may be known to vary monotonically with certain covariates, such as age or the likelihood of being approved for federal disability insurance benefits. Suppose, for example, that the conditional

Table 3. Estimated Bounds and 90% Confidence Intervals for the Employment Gap Under the Lower-Bound Accurate Reporting Rate and Partial Verification Assumptions

Verified group	HRS (n = 12,503)				SIPP (n = 29,807)			
	v	(A) Degree bounds	(B) Partial verification		v	(C) Degree bounds	(D) Partial verification	
Beneficiaries	.101	[-1.000, 1.000] [-1.000, 1.000]*	[-.976, .846] [-.980, .854]		.051	[-1.000, 1.000] [-1.000, 1.000]	[-.984, .938] [-.986, .940]	
Wave 2 verification	.267	[-1.000, 1.000] [-1.000, 1.000]	[-.793, .703] [-.802, .714]		.175	[-1.000, 1.000] [-1.000, 1.000]	[-.829, .916] [-.833, .920]	
Workers	.663	[-1.000, .448] [-1.000, .458]	[-.839, .361] [-.849, .367]		.695	[-1.000, .396] [-1.000, .403]	[-.842, .323] [-.848, .328]	
Claim no disability	.781	[-1.000, .387] [-1.000, .396]	[-.784, .361] [-.790, .368]		.812	[-1.000, .344] [-1.000, .349]	[-.800, .323] [-.804, .328]	
All of the above	.938	[-.819, -.259] [-.839, -.247]	[-.472, -.298] [-.488, -.277]		.933	[-.856, -.240] [-.862, -.232]	[-.482, -.255] [-.493, -.241]	

*Bootstrapped 90% confidence intervals (Efron and Tibshirani 1993).

employment rate is nonincreasing with age,

$$\begin{aligned}
 &age_1 \leq age_0 \leq age_2 \\
 &\implies P(L = 1|W, age_2) \leq P(L = 1|W, age_0) \\
 &\leq P(L = 1|W, age_1), \quad (7)
 \end{aligned}$$

for all $age_1 \leq age_0$ and all $age_0 \leq age_2$.

With corrupt data, the conditional probabilities in (7) are not identified. However, we can bound these probabilities using the methods described earlier. Let $LB(age)$ and $UB(age)$ be the known lower and upper bounds, given the available information on $P(L = 1|W, age)$. Then the MIV restriction formalized by Manski and Pepper (2000, prop. 1) implies that

$$\begin{aligned}
 \sup_{age_2 \geq age_0} LB(age_2) &\leq P(L = 1|W, age_0) \\
 &\leq \inf_{age_1 \leq age_0} UB(age_1). \quad (8)
 \end{aligned}$$

No other restrictions are implied by the MIV assumption. These MIV models are not nested in the usual parametric models (e.g., probit models that impose different assumptions such as homogeneity), or vice versa.

The MIV bound on the conditional employment rate is obtained using the law of total probability. If the conditional employment rate is weakly decreasing with the MIV age, then

$$\begin{aligned}
 LB_{MIV} &\equiv \sum_{age_0} P(age = age_0) \left\{ \sup_{age_2 \geq age_0} LB(age_2) \right\} \\
 &\leq P(L = 1|W) \\
 &\leq \sum_{age_0} P(age = age_0) \left\{ \inf_{age_1 \leq age_0} UB(age_1) \right\} \\
 &\equiv UB_{MIV}. \quad (9)
 \end{aligned}$$

The MIV assumption alone has no identifying power, so we combine this assumption with the previous verification assumptions. In this setting, the MIV can have identifying power if either the verification probability or an observed conditional employment rate is not monotonic with age.

5.1 Finite-Sample Bias

Estimation of the MIV bounds is complicated by the fact that one must impose the monotonicity restrictions in (8) over collections of various estimates. In finite samples, estimators that take sups and infs are systematically biased. In this setting, moreover, this bias is of particular concern in that it leads the estimated bounds to be too narrow, rather than too wide, in finite samples. The sup of the lower bound estimates is biased upward, and the inf of the upper bound estimates is biased downward. To date, no one has developed a correction for the finite-sample bias of the MIV estimator, however.

To address this concern, we introduce a modified MIV estimator that directly measures and accounts for this finite-sample bias using the nonparametric bootstrap correction (see Efron and Tibshirani 1993). To illustrate the basic idea, let T_n be a consistent analog estimator of some unknown parameter θ such that the bias of this estimator is $b_n = E(T_n) - \theta$. Using the bootstrap distribution of T_n , this bias can be estimated as $\hat{b} = E^*(T_n) - T_n$, where $E^*(\cdot)$ is the expectation operator with respect to the bootstrap distribution. A bootstrap bias-corrected estimator then follows as $T_n^c = T_n - \hat{b} = 2T_n - E^*(T_n)$. This bootstrap bias correction has been found to effectively reduce finite-sample bias (in Monte Carlo simulations) and be asymptotically efficient at higher orders in various settings (see, e.g., Parr 1983; Efron and Tibshirani 1993; Hahn, Kuersteiner, and Newey 2002; Ramalho 2005).

In our setting, the finite bias is simulated from the bootstrap distributions of the estimated Proposition 2 bounds for each age group. To estimate these bounds using the HRS, we divide the sample into 25 age groups containing 500 respondents per group (503 in the oldest group). For the SIPP sample, each age represents its own MIV group, with cell sizes ranging from to 642 to 1,692 (mean, 994). Then for each cell, the verification bounds—which are functions of various nonparametrically estimable probabilities—are estimated, and the MIV restrictions in (8) are applied. Figure 1 displays the lower bound estimate

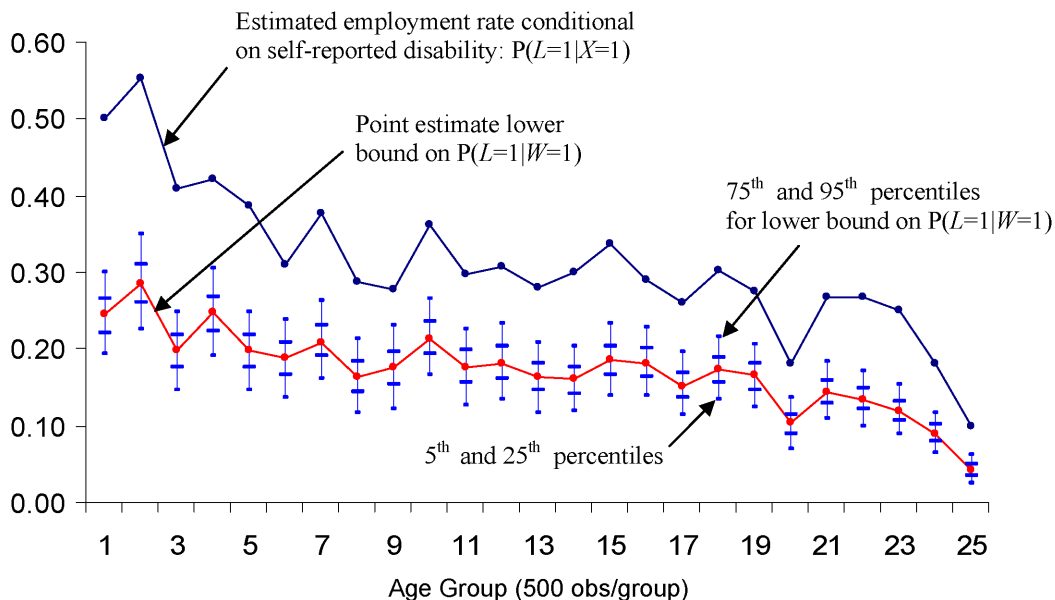


Figure 1. Bootstrapped Age-Specific Histograms for Lower Bounds on $P(L = 1|W = 1)$ in the HRS When Disability Status Is Verified for Workers.

and bootstrap distribution of $P(L = 1|W = 1)$ found using the HRS sample under the assumption that workers' responses are valid. The bias of the MIV estimator is estimated from these bootstrap sampling distributions.

To clarify the mechanics of our approach, let the parameter of interest, θ , be the equation (9) lower bound on $P(L = 1|W = 1)$ (other cases are analogous); let $LB_n(j)$ be the estimated Proposition 2 lower bound on $P(L = 1|W = 1, age = j)$ for each age group $j = 1, \dots, J$ (see, e.g., Fig. 1); and let T_n be the MIV lower-bound estimate across all age groups. In particular, $T_n = \sum_j P_n(j) \{\sup_{j' \geq j} LB_n(j')\}$, where $P_n(j)$ is the fraction of respondents in age group j . The bias b_n is estimated using the bootstrap sampling distribution of $LB_n(j)$. The first step is to randomly draw with replacement from the empirical distribution to obtain K independent pseudosamples of the original data. Then, using these samples, compute a set of K lower-bound MIV estimates of $P(L = 1|W = 1)$. Let $T_n^k, k = 1, \dots, K$, be the K lower-bound bootstrap estimates, and let $E^*(T_n) = \frac{1}{K} \sum_{k=1}^K T_n^k$ be the expected lower bound from the bootstrap distribution. Finally, compute the estimated bias, \hat{b} , and the bias-corrected MIV estimator, $T_n^c = 2T_n - E^*(T_n)$.

5.2 Empirical Results

Table 4 presents bias-corrected MIV bounds, confidence intervals, and estimated finite-sample biases for the HRS and SIPP samples. For each of our verification groups taken in isolation, the improvements in the MIV bounds compared with the verification bounds are generally modest. When only workers are verified (Fig. 1), for example, the MIV estimate of the lower bound for β using the HRS is $-.824$, a small improvement compared with the analogous verification bound of $-.839$. The improvement from the age MIV is somewhat larger for the SIPP, with the lower bound improving by 5 percentage points, from $-.842$ to $-.794$. Also note that in the case where workers are verified, the finite-sample bias plays only a modest role in both the HRS and SIPP samples. In the HRS, the bias is estimated to be 1.4 percentage points. Reflecting the larger sample

sizes, the bias in the SIPP is estimated to be 1.0 percentage point.

A more striking result emerges for the case when respondents within all four verification groups are assumed to provide accurate reports. In this case the estimated MIV bounds for the true employment gap estimated from the HRS do not contain the self-reported employment gap, $-.472$, and the 90% confidence intervals do not overlap. A similar result holds for the SIPP as the confidence interval lower bound, $-.441$, exceeds the self-reported value of β , $-.482$. This finding was also confirmed using the publicly released 5% extract of 3,806,011 individuals age 40–69 from the 1990 decennial Census (even without longitudinal information that might verify some self-reports based on responses from a subsequent wave). About 15.2% of the Census respondents reported being limited in the ability to work. The 90% confidence interval lower bound for β is $-.411$, which exceeds the self-reported value of $-.474$. With age cell sizes averaging nearly 130,000 observations in the Census, the estimated finite-sample bias for the standard MIV lower-bound estimator is negligible, $<.001$. In contrast, the estimated bias is 2.9 percentage points in the HRS and 2.5 percentage points in the SIPP.

Thus, if employment weakly decreases with age, then these findings suggest that conventional models that presume valid self-reports are likely to be misspecified. Because the unverified group comprises nonworkers who claim to be disabled, these findings support concerns in the literature that members of this group may systematically overreport disability. Also note that this finding is inconsistent with the nondifferential independence assumption, $P(X = 1|W) = P(X = 1|W, L)$, discussed in Section 3.1.

To further assess the sensitivity of this finding, we applied two other MIV assumptions in the HRS sample. First, we treated age as an MIV in disability instead of in employment. Second, instead of age, a natural MIV that exploits information from various individual characteristics in the HRS data can be constructed as the outcome of a respondent's disability insurance application decision. In particular, we let the categorical variable A equal 0 if the respondent has not applied for disability benefits, 1 if a disability application was rejected, 2 if an application was accepted after appeal, and 3 if an application was accepted immediately. Using A as the dependent variable, we constructed an MIV as fitted values from an ordered probit model of the application outcome. The specification includes indicators for a large set of physician-diagnosed health conditions and activities of daily living limitations, an indicator for subsequent mortality (died before wave 2), an indicator for ideal body mass, age, education, race, gender, marital status, veteran status, and asset level. (Details from this regression are available on request.) We define the ideal range to be 20–25 kg/m², following Fahey, Insel, and Roth (1997). In both of these cases, we find that the lower-bound MIV estimator exceeds the self-reported employment gap. For example, given full verification within the previously discussed subgroups, the 90% confidence interval for β narrows to $[-.443, -.289]$ after the disability application MIV is imposed (500 observations per cell).

Consistent with the literature, we have maintained the assumption that all verified respondents provide accurate reports of disability status. Although these verified subgroups may not

Table 4. Estimated Bounds and 90% Confidence Intervals for the Employment Gap Under the Age MIV Assumption

Verified group	HRS (n = 12,503)	SIPP (n = 29,807)
Beneficiaries	$[-.971, .831]^a$	$[-.974, .915]$
	$[-.977, .842]^b$	$[-.980, .921]$
	$+0.13 \text{ } -.022^c$	$+0.007 \text{ } -.007$
Wave 2 verification	$[-.762, .672]$	$[-.805, .881]$
	$[-.775, .686]$	$[-.815, .887]$
	$+0.021 \text{ } -.027$	$+0.012 \text{ } -.012$
Workers	$[-.824, .357]$	$[-.794, .322]$
	$[-.838, .366]$	$[-.807, .327]$
	$+0.014 \text{ } -.008$	$+0.010 \text{ } -.004$
Claim no disability	$[-.767, .357]$	$[-.785, .322]$
	$[-.780, .367]$	$[-.791, .327]$
	$+0.006 \text{ } -.008$	$+0.004 \text{ } -.004$
All of the above	$[-.431, -.308]$	$[-.438, -.255]$
	$[-.449, -.281]$	$[-.441, -.241]$
	$+0.029 \text{ } -.033$	$+0.025 \text{ } -.023$

NOTE: For the HRS, each age group is constructed with 500 observations per age cell (503 for the oldest group); for the SIPP, each age represents its own MIV group, with cell sizes ranging from 642 to 1,692 (mean, 994).

^aMIV point estimates, corrected for finite-sample bias.

^bBootstrapped 90% confidence intervals.

^cEstimated finite-sample bias.

have economic or social incentives to systematically misreport, some inaccurate responses may still exist; respondents may have difficulties in answering subjective questions, valid reports can be miscoded, and so forth. Proxy reports among verified groups may be of particular concern. Conceptual difficulties in answering questions about disability status may be compounded for respondents answering on behalf of others. Nonetheless, although proxy respondents may have less information about the extent of an impairment or its changing dynamics, they also may have less incentive to misreport. Lee et al. (2004) compared estimates of the number of disabled persons by respondent type in an environment in which self-response versus proxy was randomized. Among their primary findings, self-respondents and proxy respondents were equally likely to report disability during the initial interview, but proxy respondents were less likely to report disability in the second wave of the survey. The type of proxy mattered; spouses tended to give more consistent responses. This consistency could signify less misreporting among spouse proxies or that misreporting among individuals tends to spill over to the spouse's report. In our HRS sample, <5% of the responses came from proxy respondents; of those cases, the vast majority (>90%) were spouses. In our SIPP sample, nearly 30% of the responses came from proxies (of undocumented type).

We examined the sensitivity of our results to varying degrees of misreporting among proxies within the four verified groups. Specifically, we let $P(Z = 1|Y = 1, proxy = 1) \geq v'_y$. When $v'_y = 1$, all proxy reports within the verified groups are known to be accurate. When $v'_y = 0$, all proxy reports may be inaccurate. For the HRS, the 90% confidence interval for β expands from $[-.449, -.281]$ when $v'_y = 1$ to $[-.470, -.255]$ when $v'_y = 0$, a 4.7-percentage point increase in the widths of the bounds. The confidence interval for the SIPP expands from $[-.441, -.241]$ to $[-.583, -.105]$, a 28-percentage point increase in the width. Our earlier conclusion that the 90% confidence interval does not contain the self-reported value of β still holds in the HRS even if all of the proxy reports may be inaccurate. The conclusion still holds in the SIPP if v'_y exceeds .75.

More generally, using Proposition 2, we can allow for the possibility of reporting errors from other sources within verified groups. For $P(Z = 1|Y = 1) \geq v_y$ and arbitrary reporting errors, the 90% confidence interval does not contain the self-reported value of β if v_y exceeds .95 in the HRS and .92 in the SIPP. These critical values fall substantially, however, if invalid response among the verified can be treated as random error attributable to difficulties in answering subjective questions, coding mistakes, and so forth. In particular, suppose

$$P(W = 1|Y = 1) = P(W = 1|Y = 1, Z). \quad (10)$$

When all four subgroups are verified, the bias-corrected MIV bounds do not contain the self-report of the employment gap, β , as long as invalid response within each observed verified subgroup does not exceed about 15% in the HRS ($v_y = .85$) or about 30% in the SIPP ($v_y = .70$). Using the disability application index MIV, the confidence interval does not contain the self-reported value of β unless more than about 25% ($v_y = .75$) of respondents in the verified groups may misreport.

In summary, evidence that some types of respondents systematically overreport disability is replicated across different data and MIV assumptions and is robust to departures from the assumption of fully accurate reporting within verified groups.

6. CONCLUSION

Concerns over the validity of self-reported disability measures have been central in the many debates about the labor market outcomes of older persons. Given arbitrary errors in disability reporting, there is a critical and long-standing gap in our knowledge about how different data and assumptions affect inferences. The Institute of Medicine (2002) has highlighted the lack of information on reporting errors and called for more research into the nature and consequences of these errors. The usual approach has been to identify parameters of interest by imposing strong distributional and functional form assumptions on the nature of misreporting. Most studies assume fully accurate reporting, but others have modeled the nature of misreporting in the context of conventional latent variable models.

We have developed and applied a unifying nonparametric methodology that allows us to assess the power of different assumptions about the error process in self-reported measures of work limitation when inferring the employment gap between the nondisabled and disabled. We began by extending Horowitz and Manski's (1995) univariate setting to the case of a corrupt variable in a conditional distribution. We then examined the identifying power of "partial verification" and monotonicity restrictions on reporting errors. Within this framework, we introduced a method for correcting the finite-sample bias in Manski and Pepper's (2000) MIV estimator.

Although our approach cannot resolve decades of uncertainty tied to disability reporting errors, our analysis takes an important step in formalizing the identification problem and highlighting the identifying power of various primitive assumptions. Much of our analysis reveals the uncertainty created by arbitrary reporting errors. When we isolate the identifying power of popular verification assumptions without the usual distributional restrictions, much uncertainty about the true conditional employment rates often remains. This important negative result supports concerns that conclusions derived from conventional latent variable models may be driven largely by ad hoc parametric restrictions. Moreover, some of the estimated bounds under the MIV restrictions do not include the employment gap based on self-reported data, thus casting doubt on the validity of treating self-reports as fully accurate.

APPENDIX: PROOFS OF PROPOSITIONS 1 AND 2

Proof of Proposition 1 (Degree Bounds): $P(Z = 1) \geq v$

Decompose the conditional probability in (2) as

$$P(L = 1|W = 1) = \frac{P(L = 1, X = 1) + a - b}{(P(X = 1) + P(L = 0, X = 0, Z = 0) - P(L = 0, X = 1, Z = 0) + a - b)},$$

where $a \equiv P(L = 1, X = 0, Z = 0)$ with $0 \leq a \leq \min[(1 - v), P(L = 1, X = 0)]$ and $b \equiv P(L = 1, X = 1, Z = 0)$ with $0 \leq b \leq \min[(1 - v), P(L = 1, X = 1)]$. Then, for conjectured values of a and b , it follows that

$$\frac{P(L = 1, X = 1) - b}{P(X = 1) - b + \min\{(1 - v) - b, P(L = 0, X = 0)\}} \leq P(L = 1|W = 1) \leq \frac{P(L = 1, X = 1) + a}{P(X = 1) + a - \min\{(1 - v) - a, P(L = 0, X = 1)\}}. \quad (A.1)$$

These bounds are identified by finding the values of $\{a, b\}$ that maximize the upper bound and minimize the lower bound. First, note that these extrema are realized only if $(1 - v) - b \leq P(L = 0, X = 0)$ and $(1 - v) - a \leq P(L = 0, X = 1)$, in which case (A.1) simplifies to

$$\frac{P(L = 1, X = 1) - b}{P(X = 1) - 2b + (1 - v)} \leq P(L = 1|W = 1) \leq \frac{P(L = 1, X = 1) + a}{P(X = 1) + 2a - (1 - v)}. \quad (A.2)$$

Differentiating this bound with respect to a and b reveals that the lower bound is minimized when $b = \delta$ and the upper bound is maximized when $a = \gamma$. Proposition 1 follows.

Proof of Proposition 2 (Partial Verification): $P(Z = 1|Y = 1) \geq v_Y$

The employment rate among disabled persons is given by $P(L = 1|W = 1) = \frac{P(L=1, W=1)}{P(W=1)}$. Decompose the numerator as

$$P(L = 1, W = 1) = P(L = 1, X = 1, Y = 1) + P(L = 1, W = 1, Y = 0) + P(L = 1, X = 0, Y = 1, Z = 0) - P(L = 1, X = 1, Y = 1, Z = 0),$$

and decompose the denominator as

$$P(W = 1) = P(X = 1, Y = 1) + P(L = 1, W = 1, Y = 0) + P(L = 0, W = 1, Y = 0) + P(L = 1, X = 0, Y = 1, Z = 0) + P(L = 0, X = 0, Y = 1, Z = 0) - P(L = 1, X = 1, Y = 1, Z = 0) - P(L = 0, X = 1, Y = 1, Z = 0).$$

Let $b = P(L = 1, X = 1, Y = 1, Z = 0)$, where $0 \leq b \leq \min[(1 - v_Y) \times P(y = 1), P(L = 1, X = 1, Y = 1)]$, and let $a = P(L = 1, X = 0, Y = 1, Z = 0)$, where $0 \leq a \leq \min[(1 - v_Y)P(Y = 1), P(L = 1, X = 0, Y = 1)]$. Then, for conjectured values of a and b , it follows that

$$\begin{aligned} & (P(L = 1, X = 1, Y = 1) - b) \\ & \bigg/ (P(X = 1, Y = 1) + P(L = 0, Y = 0) - b \\ & \quad + \min\{(1 - v_Y)P(Y = 1) - b, P(L = 0, X = 0, Y = 1)\}) \\ & \leq P(L = 1|W = 1) \\ & \leq (P(L = 1, X = 1, Y = 1) + P(L = 1, Y = 0) + a) \\ & \bigg/ (P(X = 1) + P(L = 1, Y = 0) + a \\ & \quad - \min\{(1 - v_Y)P(Y = 1) - a, P(L = 0, X = 1, Y = 1)\}). \end{aligned} \quad (A.3)$$

Because a and b are unknown parameters, these bounds are not identified. Bounds are identified by finding the values of $\{a, b\}$ that maximize the upper bound and minimize the lower bound. First, note that these extrema are realized only if $(1 - v_Y)P(Y = 1) - b \leq P(L = 0, X = 0, Y = 1)$ and $(1 - v_Y)P(Y = 1) - a \leq P(L = 0, X = 1, Y = 1)$, in which case (A.3) simplifies to

$$\frac{P(L = 1, X = 1, Y = 1) - b}{P(X = 1, Y = 1) + P(L = 0, Y = 0) - 2b + (1 - v_Y)P(Y = 1)} \leq P(L = 1|W = 1) \leq \frac{P(L = 1, X = 1, Y = 1) + P(L = 1, Y = 0) + a}{P(X = 1) + P(L = 1, Y = 0) + 2a - (1 - v_Y)P(Y = 1)}. \quad (A.4)$$

Differentiating this bound with respect to a and b reveals that the lower bound is minimized when $b = \delta$ and the upper bound is maximized when $a = \gamma$. Proposition 2 follows.

Sharp Degree Bounds on β

Suppose that one has prior information on the maximum degree of inaccurate responses, $P(Z = 1) \geq v$. Using the same logic as in Proposition 1, we can also bound $P(L = 1|W = 0)$,

$$\frac{P(L = 1, X = 0) - a}{P(X = 0) - 2a + (1 - v)} \leq P(L = 1|W = 0) \leq \frac{P(L = 1, X = 0) + b}{P(X = 1) + 2b - (1 - v)}. \quad (A.5)$$

Combining (A.1) and (A.5), we have the following result:

Proposition A.1. Let $P(Z = 1) \geq v$; then

$$\begin{aligned} & \inf_{b \in (0, \min\{(1-v), P(L=1, X=1)\})} \left[\frac{P(L = 1, X = 1) - b}{P(X = 1) - 2b + (1 - v)} \right. \\ & \quad \left. - \frac{P(L = 1, X = 0) + b}{P(X = 0) + 2b - (1 - v)} \right] \\ & \leq \beta \\ & \leq \sup_{a \in (0, \min\{(1-v), P(L=1, X=0)\})} \left[\frac{P(L = 1, X = 1) + a}{P(X = 1) + 2a - (1 - v)} \right. \\ & \quad \left. - \frac{P(L = 1, X = 0) - a}{P(X = 0) - 2a + (1 - v)} \right]. \end{aligned} \quad (A.6)$$

Over part of the range for v , these bounds differ from the naive bounds obtained directly from Proposition 1. Consider, for example, the lower bound in Proposition A.1. If the value of the unknown parameter b that minimizes the first expression [i.e., the lower bound on $P(L = 1|W = 1)$] differs from the value of b that maximizes the second expression [i.e., the upper bound on $P(L = 1|W = 0)$], then the two bounds on β will differ, and the Proposition A.1 bounds will be tighter. The two bounds will be identical when the lower bound on $P(L = 1|W = 1)$ and the upper bound on $P(L = 1|W = 0)$ are realized at same value of the unknown parameter b .

[Received November 2004. Revised December 2005.]

REFERENCES

Aigner, D. J. (1973), "Regression With a Binary Independent Variable Subject to Errors of Observation," *Journal of Econometrics*, 2, 49–59.

Anderson, K., and Burkhauser, R. (1984), "The Importance of the Measure of Health in Empirical Estimates of the Labor Supply of Older Men," *Economics Letters*, 16, 375–380.

Benítez-Silva, H., Buchinsky, M., Chan, H. M., Cheidvasser, S., and Rust, J. (2004), "How Large Is the Bias in Self-Reported Disability?" *Journal of Applied Econometrics*, 19, 649–670.

Bollinger, C. (1996), "Bounding Mean Regressions When a Binary Regressor Is Mismeasured," *Journal of Econometrics*, 73, 387–399.

Bound, J. (1991), "Self-Reported versus Objective Measures of Health in Retirement Models," *Journal of Human Resources*, 26, 106–138.

Bound, J., Brown, C., and Mathiowetz, N. (2001), "Measurement Error in Survey Data," in *Handbook of Econometrics*, eds. J. Heckman and E. Leamer, Amsterdam: Elsevier Science, pp. 3705–3843.

Bound, J., and Burkhauser, R. (1999), "Economic Analysis of Transfer Programs Targeted on People With Disabilities," in *Handbook of Labor Economics*, Vol. 3C, eds. O. Ashenfelter and D. Card, Amsterdam: Elsevier Science, pp. 3417–3528.

Bowe, F. (1993), "Statistics, Politics, and Employment of People With Disabilities," *Journal of Disability Policy Studies*, 4, 84–91.

Burkhauser, R., Daly, M., Houtenville, A., and Nigras, N. (2002), "Self-Reported Work Limitation Data: What They Can and Cannot Tell Us," *Demography*, 39, 541–555.

Dominitz, J., and Sherman, R. (2004), "Sharp Bounds Under Contaminated or Corrupted Sampling With Verification, With an Application to Environmental Pollutant Data," *Journal of Agricultural, Biological and Environmental Statistics*, 9, 319–338.

Dwyer, D., and Mitchell, O. (1999), "Health Problems as Determinants of Retirement: Are Self-Rated Measures Endogenous?" *Journal of Health Economics*, 18, 173–193.

- Efron, B., and Tibshirani, R. (1993), *An Introduction to the Bootstrap*, New York: Chapman & Hall.
- Fahey, T., Insel, P., and Roth, W. (1997), *Fit and Well: Core Concepts and Labs in Physical Fitness and Wellness* (2nd ed.), Mountain View, CA: Mayfield, p. 137.
- Hahn, J., Kuersteiner, G., and Newey, W. (2002), "Higher-Order Properties of the Bootstrap and Jackknife Bias-Corrected Maximum Likelihood Estimators," working paper, Massachusetts Institute of Technology, Department of Economics.
- Hale, T. (2001), "The Lack of a Disability Measure in Today's Current Population Survey," *Monthly Labor Review*, 124, 38–40.
- Horowitz, J., and Manski, C. F. (1995), "Identification and Robustness With Contaminated and Corrupted Data," *Econometrica*, 63, 281–302.
- _____ (1998), "Censoring of Outcomes and Regressors Due to Survey Nonresponse: Identification and Estimation Using Weights and Imputations," *Journal of Econometrics*, 84, 37–58.
- Institute of Medicine (2002), *The Dynamics of Disability: Measuring and Monitoring Disability for Social Security Programs*, eds. G. S. Wunderlich, D. P. Rice, and N. L. Amado, Washington, DC: National Research Council, National Academy Press.
- Kerkhofs, M., and Lindeboom, M. (1995), "Subjective Health Measures and State-Dependent Reporting Errors," *Health Economics*, 4, 221–235.
- Kreider, B. (1999), "Latent Work Disability and Reporting Bias," *Journal of Human Resources*, 34, 734–769.
- _____ (2000), "Behavioral Responses to Changes in Federal Disability Policy: The Role of Measured Disability on Inferences," in *Research in Human Capital and Development, Vol. XIII: The Economics of Disability*, eds. D. Salkever and A. Sorkin, Stamford, CT: JAI Press, pp. 81–107.
- Kreider, B., and Hill, S. (2006), "Partially Identifying Treatment Effects With an Application to Covering the Uninsured," working paper, Iowa State University, Department of Economics.
- Kreider, B., and Pepper, J. (2006), "Inferring Disability Status From Corrupt Data," working paper, Iowa State University, Department of Economics.
- Lee, S., Mathiowetz, N., and Tourangeau, R. (2004), "Perceptions of Disability: The Effect of Self- and Proxy-Response," *Journal of Official Statistics*, 20, 671–686.
- Manski, C. F., and Pepper, J. (2000), "Monotone Instrumental Variables, With an Application to the Returns to Schooling," *Econometrica*, 68, 997–1010.
- McGarry, K. (2004), "Health and Retirement: Do Changes in Health Affect Retirement Expectations?" *Journal of Human Resources*, 39, 624–648.
- Myers, R. (1982), "Why Do People Retire From Work Early?" *Social Security Bulletin*, 45, 10–14.
- National Council on Disability (2002), *National Disability Policy: A Progress Report, December 2000–December 2001*, Washington, DC: Author.
- O'Donnell, O. (1998), "The Effect of Disability on Employment Allowing for Work Incapacity," Working Paper 98-13, University of Kent at Canterbury, Department of Economics.
- Parr, W. (1983), "A Note on the Jackknife, the Bootstrap and the Delta Method Estimators of Bias and Variance," *Biometrika*, 70, 719–722.
- Ramalho, J. (2005), "Small-Sample Bias of Alternative Estimation Methods for Moment Condition Models: Monte Carlo Evidence for Covariance Structures," *Studies in Nonlinear Dynamics and Econometrics*, 9, article 3, 1–18.
- Stern, S. (1989), "Measuring the Effect of Disability on Labor Force Participation," *Journal of Human Resources*, 24, 361–395.
- U.S. General Accounting Office (1997), "Social Security Disability: SSA Must Hold Itself Accountable for Continued Improvement in Decision-Making," HEHS-97-102, August 12.
- Watson, D., and Clark, L. A. (1984), "Negative Affectivity: The Disposition to Experience Negative Emotional States," *Psychological Bulletin*, 96, 465–490.