Sensitivity Analysis for Instrumental Variables Regression With Overidentifying Restrictions [1]

Dylan S. Small

University of Pennsylvania

July 27, 2006

[1]Dylan S. Small is Assistant Professor, Department of Statistics, The Wharton School of the University of Pennsylvania, Philadelphia, PA 19104-6340 (E-mail: *dsmall@wharton.upenn.edu*). The author is grateful to Robert Stine and Tom Ten Have for valuable comments on a draft and to Joshua Angrist, Ed George and Paul Rosenbaum for insightful discussion. The author also thanks the associate editor and three referees for helpful comments

Abstract

Instrumental variables (IV) regression is a method for making causal inferences about the effect of a treatment based on an observational study in which there are unmeasured confounding variables. The method requires one or more valid IVs; a valid IV is a variable that is independent of unmeasured confounding variables and has no direct effect on the outcome. Often there is uncertainty about the validity of the proposed IVs. When a researcher proposes more than one IV, the validity of the IVs can be tested via the "overidentifying restrictions test." Although the overidentifying restrictions test does provide some information, the test has no power versus certain alternatives and can have low power versus many alternatives due to its omnibus nature. To fully address uncertainty about the validity of the proposed IVs, we argue that a sensitivity analysis is needed. A sensitivity analysis examines the impact of plausible amounts of invalidity of the proposed IVs on inferences for the parameters of interest. We develop a method of sensitivity analysis for IV regression with overidentifying restrictions that makes full use of the information provided by the overidentifying restrictions test, but provides more information than the test by exploring sensitivity to violations of the validity of the proposed IVs in directions for which the test has low power. Our sensitivity analysis uses interpretable parameters that can be discussed with subject matter experts. We illustrate our methods using a study of food demand among rural households in the Philippines.

Keywords: causal inference; econometrics; structural equations models

1. Introduction

A central problem in making inferences about the causal effect of a treatment based on an observational study is the potential presence of unobserved confounding variables. Instrumental variables (IV) regression is a method for overcoming this problem. The method requires a valid IV, which is a variable that is independent of the unobserved confounding variables and has no direct effect on the outcome. IV regression uses the IV to extract variation in the treatment that is unrelated to the unobserved confounding variables and then uses this variation to estimate the causal effect of the treatment. An example of IV regression is Card (1995), which studies the causal effect of education on earnings and uses the distance a person lived when growing up from the nearest 4-year college as an IV.

In many applications of IV regression, researchers are uncertain about the validity of the proposed IV. For example, Card (1995) was concerned that families that place a strong emphasis on education are more likely to choose to live near a college and that the children of these families are more likely to be motivated to achieve labor market success. Under this scenario, the proposed IV (distance from nearest 4-year college) would be associated with an unobserved confounding variable (motivation), meaning that the proposed IV would be invalid. The validity of a proposed IV cannot be consistently tested (see Section 3.1). When a critical assumption for an analysis is partially unverifiable, a *sensitivity analysis* is useful. A sensitivity analysis examines the impact of plausible (in the view of subject matter experts) violations of the assumptions. The value of doing a sensitivity analysis to account for uncertainty about partially unverifiable assumptions has long been recognized in causal inference. Rosenbaum (2002, Ch.4) provides a review. Sensitivity analyses for IV regression have been considered by Manski (1995), Angrist, Imbens and Rubin (1996), Rosenbaum (1999) and Hogan, Lancaster, Roy and Alderson (2003).

In this paper, we develop a method of sensitivity analysis for IV regression when there is more than one proposed IV, a setting not considered by previous studies of IV regression sensitivity analysis. IV regression with more than one proposed IV is called *IV regression with overidentifying restrictions* because only one valid IV is needed to identify the causal effect of treatment so more than one IV "overidentifies" the causal effect. IV regression with overidentifying restrictions is often used in economics. An example is Kane and Rouse (1993), which studies the causal effect of education on earnings as Card (1995) did, but, in addition to distance from nearest 4-year college, uses tuition at state colleges of the state in which a person grew up as an IV. For IV regression with overidentifying restrictions, all the proposed IVs must be valid for the inferences to be correct. Although this is a more stringent requirement than one proposed IV being valid, there are two benefits to considering multiple proposed IVs (Davidson and MacKinnon, 1993). First, if all of the proposed IVs are valid, the use of multiple IVs can increase efficiency (see Section 2). Second, the use of multiple proposed IVs enables the joint validity of all the proposed IVs to be tested (to a limited extent) via the overidentifying restrictions test (ORT) (see Section 3).

For IV regression with overidentifying restrictions, it is common practice to report inferences that assume all the proposed IVs are valid along with the p-value for the ORT. However, the results of the ORT are hard to interpret because the test may have low power for many alternatives and is not even consistent for some alternatives. Consequently, a sensitivity analysis is still essential for addressing uncertainty about the validity of the proposed IVs when multiple proposed IVs are considered. We develop a method of sensitivity analysis that uses the information provided by the ORT, but provides more information than the test by also exploring the sensitivity of inferences to violations of assumptions in directions for which the test has low power.

Our paper is organized as follows. In Section 2, we describe a model for IV regression and inferences for it. In Section 3, we describe the ORT and its power. In Section 4, we develop our method of sensitivity analysis. In Section 5, we discuss extensions of our method to IV regression with heterogeneous treatment effects and/or panel data. In Section 6, we illustrate our sensitivity analysis method using a study of food demand. In Section 7, we provide discussion.

2. Instrumental Variables Regression Model

In this section, we describe an additive, linear, constant effect causal model and explain how valid IVs enable identification of the model. For defining causal effects, we use the potential outcomes approach (Neyman, 1923; Rubin, 1974). Let y denote an outcome and w denote a treatment variable that an intervention can in principle alter. For example, in Kane and Rouse's 1993 study, y is earnings, w is amount of education and the intervention that could alter w is for an individual to choose to acquire more or less education. Let $y_i^{(w*)}$ denote the outcome that would be observed for unit i if unit i's level of w was set equal to w^*. We assume that the potential outcomes for unit i depend only on the level of w set for unit i and not on the levels of w set for other units – this is called the stable unit treatment value assumption by Rubin (1986). Let $y_i^{obs} := y_i$ and $w_i^{obs} := w_i$ denote the observed values of y and w for unit i. Each unit has a vector of potential outcomes, one for each possible level of w, but we observe only one potential outcome, $y_i = y_i^{(w_i)}$. An additive, linear constant effects causal model for the potential outcomes as in Holland (1988) is

$$y_i^{(w^*)} = y_i^{(0)} + \beta w^*. \tag{1}$$

Our parameter of interest is $\beta = y_i^{(w^*+1)} - y_i^{(w^*)}$, the causal effect of increasing w by one unit. One way to estimate β is ordinary least squares (OLS) regression of y^{obs} on w^{obs}. The OLS coefficient on w, $\hat{\beta}_{OLS}$, has probability limit $\beta + Cov(w_i^{obs}, y_i^{(0)})/Var(w_i^{obs})$. If w_i^{obs} were randomly assigned, then $Cov(w_i^{obs}, y_i^{(0)})$ would equal 0 and $\hat{\beta}_{OLS}$ would be consistent. But in an observational study, often $Cov(w_i^{obs}, y_i^{(0)}) \neq 0$ and $\hat{\beta}_{OLS}$ is inconsistent. One strategy to address this problem is to attempt to collect data on all confounding variables $\mathbf{q}$ and then to regress y^{obs} on w^{obs} and $\mathbf{q}$. If w^{obs} is conditionally independent of $y^{(0)}$ given $\mathbf{q}$ (i.e., w^{obs} is ignorable) and the regression function is specified correctly, this strategy produces a consistent estimate of β. However, it often difficult to figure out and/or collect all confounding variables $\mathbf{q}$.

IV regression is another strategy for estimating β in (1). A vector of valid IVs $\mathbf{z}_i$ is a vector of variables that satisfies (see Angrist et al., 1996 and Tan, 2005):

(A1) $\mathbf{z}_i$ are covariates whose value would not be changed by an intervention that changed w for unit i from w_i^{obs} to $w^* \neq w_i^{obs}$;

(A2) $\mathbf{z}_i$ is associated with the observed treatment w_i^{obs}; and

(A3) $\mathbf{z}_i$ is independent of $\{y_i^{(w^*)}, w^* \in \mathcal{W}\}$ where $\mathcal{W}$ is the set of possible values of w; note that under model (1), this is equivalent to $\mathbf{z}_i$ being independent of $y_i^{(0)}$.

The basic idea of IV regression is to use $\mathbf{z}$ to extract variation in w^{obs} that is uncorrelated with the confounding variables and to only use this part of the variation in w^{obs} to estimate the causal relationship between w and y. Assumption (A1) is needed for the potential outcomes $y_i^{(w^*)}$ to be well defined (Rubin, 1986). (A2) is needed to be able to use $\mathbf{z}$ to extract variation in w^{obs}. (A3) is needed for the variation in w extracted from variation in $\mathbf{z}$ to be independent of the confounding variables.

An example of the usefulness of IVs is the encouragement design (Holland, 1988). An encouragement design is used when we want to estimate the causal effect of a treatment w that we cannot control, but we can control (or observe from a natural experiment) variable(s) $\mathbf{z}$ which, depending on their level, encourage or do not encourage a unit to have a high level of w^{obs}. If the levels of the encouragement variables $\mathbf{z}$ are randomly assigned (or ignorable) and encouragement, in and of itself, has no direct effect on the outcome, then $\mathbf{z}$ is a vector of valid IVs (Holland, 1988; Angrist et al., 1996). Kane and Rouse's study can be viewed as an encouragement design in which it is being assumed that distance from nearest 4-year college and tuition at state colleges where a person grew up are ignorable, and low levels of these variables encourage a person to attend college, but have no direct effect on earnings.

In order for assumption (A3) to be plausible, it is often necessary to condition on a vector of covariates $\mathbf{x}_i$ (Tan, 2005). For example, Kane and Rouse condition on region, city-size and family background because these variables may be associated with both potential earnings outcomes $y_i^{(w^*)}$ and distance from nearest 4-year college (or tuition at state colleges). Conditioning on $\mathbf{x}_i$ can also increase the efficiency of the IV regression estimator. We call the variables in $\mathbf{x}_i$ the included exogenous variables. We consider a linear model for $E(y_i^{(0)}|\mathbf{x}_i, \mathbf{z}_i)$:

$$y_i^{(w^*)} = \beta w^* + \boldsymbol{\delta}^T \mathbf{x}_i + \boldsymbol{\lambda}^T \mathbf{z}_i + u_i, \; E(u_i|\mathbf{x}_i, \mathbf{z}_i) = 0. \tag{2}$$

We assume that $(w_i, \mathbf{x}_i, \mathbf{z}_i, u_i)$ are iid random vectors. This model has been considered by Holland (1988) among others. The model for the observed data is

$$\begin{aligned} y_i = \beta w_i + \boldsymbol{\delta}^T \mathbf{x}_i + \boldsymbol{\lambda}^T \mathbf{z}_i + u_i, \qquad & E(u_i|\mathbf{x}_i, \mathbf{z}_i) = 0 \\ (w_i, \mathbf{x}_i, \mathbf{z}_i, u_i), i = 1, \ldots, N \quad & \text{are iid random vectors} \end{aligned} \tag{3}$$

For this model, we restate assumptions (A2) and (A3) for $\mathbf{z}$ being a vector of valid IVs as

(A2') Letting $E^*(w^{obs}|\mathbf{x}, \mathbf{z}) = \boldsymbol{\nu}^T \mathbf{x} + \boldsymbol{\gamma}^T \mathbf{z}$, where E^* denotes linear projection, the vector $\boldsymbol{\gamma}$ does not equal $\mathbf{0}$.

(A3') The vector $\boldsymbol{\lambda}$ in (3) equals $\mathbf{0}$.

(A2') says that $\mathbf{z}$ is associated with w^{obs} conditional on $\mathbf{x}$. (A3') says that $\mathbf{z}$ is independent of the unobserved confounding variables conditional on $\mathbf{x}$.

We now consider inferences for β under the assumption that $\mathbf{z}$ is a vector of valid IVs. Under regularity conditions, β can be consistently estimated by the two stage least squares (TSLS) method (White, 1982; Hausman, 1983). The TSLS estimator is based on the fact that if $\boldsymbol{\lambda} = \mathbf{0}$ in (3), then $E^*(y|\mathbf{x}, \mathbf{z}) = \beta E^*(w|\mathbf{z}) + \boldsymbol{\delta}^T \mathbf{x}$; the TSLS estimator is obtained by regressing w on $(\mathbf{x}, \mathbf{z})$ using OLS to obtain $\hat{E}^*(w|\mathbf{x}, \mathbf{z})$ and then regressing y on $\hat{E}^*(w|\mathbf{x}, \mathbf{z})$ and $\mathbf{x}$ using OLS to estimate β and $\boldsymbol{\delta}$. The asymptotic distribution of $\hat{\beta}_{TSLS}$ is

$$\frac{\sqrt{N}(\hat{\beta}_{TSLS} - \beta)}{\hat{\sigma}_u^2 \sum_{i=1}^N (\hat{w}_i - \hat{E}^*(\hat{w}|\mathbf{x}_i))^2} \xrightarrow{D} N(0, 1), \tag{4}$$

where $\hat{w}$ is the predicted w from the OLS regression of w^{obs} on $\mathbf{x}$ and $\mathbf{z}$, $\hat{E}^*(\hat{w}|\mathbf{x})$ is the estimated linear projection of $\hat{w}$ onto $\mathbf{x}$ and $\hat{\sigma}_u^2 = (1/N)\sum_{i=1}^N (y_i - \hat{\beta}_{TSLS} w_i - \hat{\boldsymbol{\delta}}_{TSLS}\mathbf{x}_i)^2$. For fixed $\mathbf{x}_1, \ldots, \mathbf{x}_N$, the denominator of the left hand side of (4) is proportional to the partial R^2 for $\mathbf{z}$ from the OLS regression of w on $\mathbf{x}$ and $\mathbf{z}$. Thus, using additional valid IVs that increase the R^2 for the OLS regression of w on $\mathbf{x}$ and $\mathbf{z}$ (i.e., the components of $\boldsymbol{\gamma}$ in (A2') for the additional IVs do not equal 0) increases efficiency.

One method of forming a confidence interval (CI) for β is to invert a Wald test for β based on the TSLS estimate. We focus in this paper on an alternative CI proposed by Anderson and Rubin (1949). We note that if $\beta = \beta_0$ and model (3) holds with $\boldsymbol{\lambda} = \mathbf{0}$, then $y_i - \beta_0 w_i = \boldsymbol{\delta}^T \mathbf{x}_i + u_i, E(u_i|\mathbf{x}_i, \mathbf{z}_i) = 0$. The Anderson-Rubin (AR) test for $H_0 : \beta = \beta_0$ is to test whether $\boldsymbol{\tau} = \mathbf{0}$ in the regression

$$E(y_i - w_i\beta_0|\mathbf{x}_i, \mathbf{z}_i) = \boldsymbol{\chi}^T \mathbf{x}_i + \boldsymbol{\tau}^T \mathbf{z}_i. \tag{5}$$

The AR CI for β is the inversion of this test. An advantage of the AR CI for β compared to the Wald CI based on the TSLS estimator is that the validity of the AR CI is robust to the vector $\boldsymbol{\gamma}$ in (A2') being of small magnitude (Kleibergen, 2005).

The validity of the above inferences for β do not require $E(y_i^{(0)}|\mathbf{x}_i, \mathbf{z}_i)$ to be a linear function of $(\mathbf{x}, \mathbf{z})$. Let $E^*(y_i^{(0)}|\mathbf{x}_i, \mathbf{z}_i) = \boldsymbol{\delta}^T \mathbf{x}_i + \boldsymbol{\lambda}^T \mathbf{z}_i$. Also, let $u_i' = y_i^{(0)} - E^*(y_i^{(0)}|\mathbf{x}_i, \mathbf{z}_i)$ and assume the u_i' are iid. Then under conditions (A1), (A2') and (A3'), the TSLS estimator is consistent and the AR CI is valid (Davidson and MacKinnon, 1993).

If $\boldsymbol{\lambda} \neq \mathbf{0}$, then $(\hat{\beta}_{TSLS}, \hat{\boldsymbol{\delta}}_{TSLS})$ is not a consistent estimator of $(\beta, \boldsymbol{\delta})$. Let $\mathbf{Y}_N = (y_1, \ldots, y_N)^T$, $\mathbf{W}_N = (w_1, \ldots, w_N)^T$, $\mathbf{X}_N = (\mathbf{x}_1, \ldots, \mathbf{x}_N)^T$, $\mathbf{Z}_N = (\mathbf{z}_1, \ldots, \mathbf{z}_N)^T$ and $\mathbf{D}_N = [\mathbf{X}_N, \mathbf{Z}_N]$. The asymptotic bias of $\hat{\beta}_{TSLS}, \hat{\boldsymbol{\delta}}^T_{TSLS}$ is $\lim_{N\to\infty}\{([\mathbf{D}_N(\mathbf{D}_N^T\mathbf{D}_N)^{-1}\mathbf{D}_N^T \mathbf{W}_N, \mathbf{X}_N]^T\ [\mathbf{D}_N(\mathbf{D}_N^T\mathbf{D}_N)^{-1}\mathbf{D}_N^T\mathbf{W}_N, \mathbf{X}_N])^{-1}[\mathbf{D}_N(\mathbf{D}_N^T\mathbf{D}_N)^{-1}\mathbf{D}_N^T\mathbf{W}_N, \mathbf{X}_N]^T\mathbf{Z}_N\boldsymbol{\lambda}\}$.

Besides causal inference, IVs are also useful for analysis of simultaneous equations systems (Hausman, 1983) and measurement error (Fuller, 1987).

3. Overidentifying restrictions test and its power

The ORT is a test of the assumption that $\boldsymbol{\lambda} = \mathbf{0}$ in (3) which can be applied when

$\dim(\mathbf{z}) > 1$. The ORT derives its name from the fact that when $\dim(\mathbf{z}) > 1$ and $\boldsymbol{\lambda} = \mathbf{0}$, the parameter β in (3) is "overidentified," meaning that any nonempty subset of the $\mathbf{z}$ variables could be used as the IVs to obtain a consistent estimate of β. If $\boldsymbol{\lambda} = \mathbf{0}$, then the different estimators of β that involve using different subsets of $\mathbf{z}$ as proposed IVs will all converge to the true β. But if $\boldsymbol{\lambda} \neq \mathbf{0}$, then the different estimators may converge to different limits. The ORT looks at the degree of agreement between the different estimates of β that involve using different subsets of $\mathbf{z}$ as proposed IVs. Several versions of the ORT have been developed, e.g., Anderson and Rubin (1949) and Sargan (1958). Newey (1985) shows that among ORTs based on a finite set of moment conditions, all such tests with maximal degrees of freedom ($\dim(\mathbf{z}) - 1$) are asymptotically equivalent. We shall focus on one of these tests with maximal degrees of freedom, Sargan's test, and refer to this as *the* ORT.

To motivate the ORT, we augment (3) with a regression model for w,

$$\begin{aligned} y_i &= \beta w_i + \boldsymbol{\delta}^T \mathbf{x}_i + \boldsymbol{\lambda}^T \mathbf{z}_i + u_i \\ w_i &= \boldsymbol{\nu}^T \mathbf{x}_i + \boldsymbol{\gamma}^T \mathbf{z}_i + v_i, \quad E((u_i, v_i)|\mathbf{x}_i, \mathbf{z}_i) = \mathbf{0} \end{aligned} \tag{6}$$

Substituting the model for w_i into the model for y_i in (6), we have

$$\begin{aligned} y_i &= (\beta \boldsymbol{\nu}^T + \boldsymbol{\delta}^T)\mathbf{x}_i + (\beta \boldsymbol{\gamma}^T + \boldsymbol{\lambda}^T)\mathbf{z}_i + \beta v_i + u_i \\ &\equiv \boldsymbol{\rho}^T \mathbf{x}_i + (\beta \boldsymbol{\gamma}^T + \mathbf{P}_{\boldsymbol{\gamma}} \boldsymbol{\lambda}) P_{\boldsymbol{\gamma}} \mathbf{z}_i + (\mathbf{P}_{\boldsymbol{\gamma}^\perp} \boldsymbol{\lambda})^T \mathbf{P}_{\boldsymbol{\gamma}^\perp} \mathbf{z}_i + e_i, \end{aligned} \tag{7}$$

where $E(e_i|\mathbf{x}_i, \mathbf{z}_i) = 0$ and $\mathbf{P}_{\boldsymbol{\gamma}}\mathbf{a}$ and $\mathbf{P}_{\boldsymbol{\gamma}^\perp}\mathbf{a}$ denote the projection of the vector $\mathbf{a}$ in the direction $\boldsymbol{\gamma}$ and into the space orthogonal to $\boldsymbol{\gamma}$ in $\mathbb{R}^{\dim(\mathbf{z})}$ respectively. Note that when $\boldsymbol{\lambda} = \mathbf{0}$, we have $\mathbf{P}_{\boldsymbol{\gamma}^\perp}\boldsymbol{\lambda} = \mathbf{0}$. The ORT is a test of $H_0 : \mathbf{P}_{\boldsymbol{\gamma}^\perp}\boldsymbol{\lambda} = \mathbf{0}$ in the regression model (7). In particular, the ORT test statistic

$$\begin{aligned} J_N = &[(\mathbf{Z}_N)^T(\mathbf{Y}_N - \mathbf{W}_N \hat{\beta}_{TSLS} - \mathbf{X}_N \hat{\boldsymbol{\delta}}_{TSLS})]^T [\hat{\sigma}_u^2 (\mathbf{Z}_N^T \mathbf{Z}_N)]^{-1} \\ &[(\mathbf{Z}_N)^T(\mathbf{Y}_N - \mathbf{W}_N \hat{\beta}_{TSLS} - \mathbf{X}_N \hat{\boldsymbol{\delta}}_{TSLS})] \end{aligned} \tag{8}$$

is a generalized score test (Boos, 1992) of $H_0 : \mathbf{P}_{\boldsymbol{\gamma}^\perp}\boldsymbol{\lambda} = \mathbf{0}$ using the estimating equations $E[(\mathbf{x}, \mathbf{z})^T(y - \boldsymbol{\rho}^T\mathbf{x} - (\beta\boldsymbol{\gamma}^T + \mathbf{P}_{\boldsymbol{\gamma}}\boldsymbol{\lambda})P_{\boldsymbol{\gamma}}\mathbf{z} - (\mathbf{P}_{\boldsymbol{\gamma}^\perp}\boldsymbol{\lambda})^T\mathbf{P}_{\boldsymbol{\gamma}^\perp}\mathbf{z})] = \mathbf{0}$, $E[(\mathbf{x}, \mathbf{z})(w-$

$\boldsymbol{\nu}^T\mathbf{x}-\boldsymbol{\gamma}^T\mathbf{z})] = \mathbf{0}$. Under $H_0 : \boldsymbol{\lambda} = \mathbf{0}$, the distribution of J_N converges to $\chi^2(\dim(\mathbf{z})-1)$ as $N \to \infty$. This null distribution does not depend on the model for w in (6) being correct because if we let $E^*(w|\mathbf{x}_i, \mathbf{z}_i) = \boldsymbol{\nu}^T\mathbf{x}_i + \boldsymbol{\lambda}^T\mathbf{x}_i$, then the estimating equations on which the generalized score statistic J_N is based are still valid.

The ORT is incorporated into the AR test of $\beta = \beta_0$ (Kleibergen, 2005). Under model (6) (see also (7)), we can write

$$\begin{aligned} y - \beta_0 w &= (\beta-\beta_0)w + \boldsymbol{\delta}^T\mathbf{x}_i + \boldsymbol{\lambda}^T\mathbf{z}_i + u_i \\ &= [(\beta-\beta_0)\boldsymbol{\nu}^T + \boldsymbol{\delta}^T]\mathbf{x}_i + [(\beta-\beta_0)\boldsymbol{\gamma}^T + \boldsymbol{\lambda}^T]\mathbf{z}_i + \beta v_i + u_i \\ &= [(\beta-\beta_0)\boldsymbol{\nu}^T + \boldsymbol{\delta}^T]\mathbf{x}_i + [(\beta-\beta_0)\boldsymbol{\gamma} + \mathbf{P}_{\boldsymbol{\gamma}}\boldsymbol{\lambda}]^T\mathbf{P}_{\boldsymbol{\gamma}}\mathbf{z}_i + (\mathbf{P}_{\boldsymbol{\gamma}^\perp}\boldsymbol{\lambda})^T\mathbf{P}_{\boldsymbol{\gamma}^\perp}\mathbf{z}_i + e_i. \end{aligned}$$

Thus, under model (6), the AR test of $H_0 : \beta = \beta_0$ actually tests $H_0 : (\beta-\beta_0)\boldsymbol{\gamma} + \mathbf{P}_{\boldsymbol{\gamma}}\boldsymbol{\lambda} = \mathbf{0}, \mathbf{P}_{\boldsymbol{\gamma}^\perp}\boldsymbol{\lambda} = \mathbf{0}$. Kleibergen (2005) shows that the AR test statistic is the sum of a test statistic for $H_{0A} : (\beta-\beta_0)\boldsymbol{\gamma}^T + \mathbf{P}_{\boldsymbol{\gamma}}\boldsymbol{\lambda} = \mathbf{0}$ and a test statistic for $H_{0B} : \mathbf{P}_{\boldsymbol{\gamma}^\perp}\boldsymbol{\lambda} = \mathbf{0}|\beta = \beta_0$. The latter test statistic is similar to J_N except that it estimates σ_u^2 under the assumption $\beta = \beta_0$ rather than $\beta = \hat{\beta}_{TSLS}$. Thus, the AR test of $H_0 : \beta = \beta_0$ incorporates the ORT and is better regarded as a test of $H_0 : \beta = \beta_0, \boldsymbol{\lambda} = \mathbf{0}$.

3.1 Inconsistency of the Overidentifying Restrictions Test

For some alternatives to $\boldsymbol{\lambda} = \mathbf{0}$, the ORT is not consistent (i.e., its power does not converge to one). Specifically, consider model (6) and also suppose that (u_i, v_i) are iid bivariate normal (this assumption is standard in simultaneous equation system models, of which (6) is a special case (Hausman, 1983)). For this model, the ORT is not consistent (Kadane and Anderson 1977). We now characterize the alternatives for which the ORT is not consistent. Following Rothenberg (1971), we call the parameter vector $\boldsymbol{\theta} = (\beta, \boldsymbol{\lambda}, \boldsymbol{\gamma}, \boldsymbol{\delta}, \boldsymbol{\upsilon}, \sigma_u^2, \sigma_v^2, \sigma_{uv})$ a *structure*. The structure $\boldsymbol{\theta}$ specifies a distribution for (y_i, w_i) conditional on $(\mathbf{x}_i, \mathbf{z}_i)$. Two structures $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ are said to be *observationally equivalent* if they specify the same distribution for (y_i, w_i) conditional on $(\mathbf{x}_i, \mathbf{z}_i)$. The ORT test statistic J_N is a function of the observable data

$\{y_i, w_i, \mathbf{x}_i, \mathbf{z}_i : i = 1, \ldots, N\}$ and therefore J_N has the same distribution for observationally equivalent structures conditional on $\{\mathbf{x}_i, \mathbf{z}_i : i = 1, \ldots, N\}$ (Kadane and Anderson 1977).

We now characterize the equivalence classes of observationally equivalent structures. By substituting the model for w into the model for y in (6), we obtain the "reduced form":

$$\begin{aligned} y_i &= \beta \boldsymbol{\upsilon}^T \mathbf{x}_i + \boldsymbol{\delta}^T \mathbf{x}_i + \beta \boldsymbol{\gamma}^T \mathbf{z}_i + \boldsymbol{\lambda}^T \mathbf{z}_i + \beta v_i + u_i \\ &\equiv \boldsymbol{\rho}^T \mathbf{x}_i + \boldsymbol{\kappa}^T \mathbf{z}_i + e_i, \\ w_i &= \boldsymbol{\upsilon}^T \mathbf{x}_i + \boldsymbol{\gamma}^T \mathbf{z}_i + v_i, \end{aligned}$$

where (e_i, v_i) has a bivariate normal distribution with mean $\mathbf{0}$. The distribution for $(y_i, w_i)|\mathbf{x}_i, \mathbf{z}_i$ depends only on the reduced form parameter $\boldsymbol{\pi} = (\boldsymbol{\rho}, \boldsymbol{\kappa}, \boldsymbol{\upsilon}, \boldsymbol{\gamma}, \sigma_e^2, \sigma_v^2, \sigma_{ev})$. Also, any two structures which have reduced form parameters $\boldsymbol{\pi}_1 \neq \boldsymbol{\pi}_2$ have different distributions for $(y_i, w_i)|\mathbf{x}_i, \mathbf{z}_i$. Therefore, two structures are observationally equivalent if and only if they have the same reduced form parameter $\boldsymbol{\pi}$.

The reduced form parameter $\boldsymbol{\pi}$ is a function h of the structural parameter $\boldsymbol{\theta}$, $\boldsymbol{\pi} = h(\boldsymbol{\theta}) = (\beta\boldsymbol{\upsilon} + \boldsymbol{\delta}, \beta\boldsymbol{\gamma} + \boldsymbol{\lambda}, \boldsymbol{\upsilon}, \boldsymbol{\gamma}, \beta^2\sigma_v^2 + 2\beta\sigma_{uv} + \sigma_u^2, \sigma_v^2, \beta\sigma_v^2 + \sigma_{uv})$. For a reduced form parameter $\boldsymbol{\pi}^* = (\boldsymbol{\rho}^*, \boldsymbol{\kappa}^*, \boldsymbol{\upsilon}^*, \boldsymbol{\gamma}^*, (\sigma_e^*)^2, (\sigma_v^*)^2, \sigma_{ev}^*)$, the set of structures which have reduced form parameter $h(\boldsymbol{\theta}) = \boldsymbol{\pi}^*$ is the following set parameterized by c,

$$\begin{aligned} OE(\boldsymbol{\pi}^*) = \{(\beta, \boldsymbol{\lambda}, \boldsymbol{\gamma}, \boldsymbol{\delta}, \boldsymbol{\upsilon}, \sigma_u^2, \sigma_v^2, \sigma_{uv}) : \beta = c, \boldsymbol{\lambda} = \boldsymbol{\kappa}^* - c\boldsymbol{\gamma}^*, \boldsymbol{\gamma} = \boldsymbol{\gamma}^*, \boldsymbol{\delta} = \boldsymbol{\rho}^* - c\boldsymbol{\upsilon}^*, \\ \boldsymbol{\upsilon} = \boldsymbol{\upsilon}^*, \sigma_u^2 = (\sigma_e^*)^2 + c^2(\sigma_v^*)^2 - 2c\sigma_{ev}^*, \sigma_v^2 = (\sigma_v^*)^2, \sigma_{uv} = \sigma_{ev}^* - c(\sigma_v^*)^2, c \in \mathbb{R}\}. \end{aligned}$$

The set of $\boldsymbol{\lambda}$'s in $OE(\boldsymbol{\pi})$ for the true $\boldsymbol{\pi}$ is

$$\ell = \{\boldsymbol{\lambda} : \boldsymbol{\lambda} = \boldsymbol{\kappa} - c\boldsymbol{\gamma}, c \in \mathbb{R}\}, \tag{9}$$

which is a line in the parameter space $\mathbb{R}^{\dim(\mathbf{z})}$ of $\boldsymbol{\lambda}$. Thus, we can identify $\boldsymbol{\lambda}$ "up to a line."

The line ℓ crosses through $\mathbf{0}$ if and only if $\boldsymbol{\lambda} = c\boldsymbol{\gamma}$ for some constant c. Combining this fact and 1) the fact that for each $\boldsymbol{\lambda}$ on ℓ, $OE(\boldsymbol{\pi})$ contains a point with that value of $\boldsymbol{\lambda}$ and 2) the fact that J_N has the same distribution for observationally equivalent structures conditional on $\{(\mathbf{x}_i, \mathbf{z}_i) : i = 1, \ldots, N\}$, we have

Proposition 1. *If the structure* $\boldsymbol{\theta} = (\beta, \boldsymbol{\lambda}, \boldsymbol{\gamma}, \boldsymbol{\delta}, \boldsymbol{\upsilon}, \sigma_u^2, \sigma_v^2, \sigma_{uv})$ *has the property that* $\boldsymbol{\lambda} = c\boldsymbol{\gamma}$ *for some constant* $c \neq 0$ *where* $\boldsymbol{\gamma} \neq \mathbf{0}$, *then the ORT is not consistent for the structure* $\boldsymbol{\theta}$.

The null hypothesis of the ORT can be viewed as H_0 : the line ℓ on which $\boldsymbol{\lambda}$ is identified to lie crosses through $\mathbf{0}$, rather than $H_0 : \boldsymbol{\lambda} = \mathbf{0}$. Note that if $\dim(\mathbf{z}) = 1$, all structures satisfy the former null hypothesis so the ORT has no power.

Because the ORT is not consistent for certain alteratives to $\boldsymbol{\lambda} = \mathbf{0}$ and potentially has low power for many alternatives, it is important to consider which of the values of $\boldsymbol{\lambda}$ that the ORT has low power against are plausible and how much would it alter inferences about β if such plausible values of $\boldsymbol{\lambda}$ were true.

4. Sensitivity Analysis

For IV regression analysis under model (3), the critical assumption used to make inferences is that $\boldsymbol{\lambda} = \mathbf{0}$. A sensitivity analysis considers what happens to inferences about β if $\boldsymbol{\lambda}$ is known to belong to a sensitivity region $\mathbf{A}$ of values that a subject matter expert considers a priori plausible, rather than assuming $\boldsymbol{\lambda} = \mathbf{0}$. The output of our sensitivity analysis is a sensitivity interval (SI), an interval that has a high probability of containing β as long as $\boldsymbol{\lambda} \in \mathbf{A}$. A SI is an analogue of a CI for a setting in which a parameter of interest is not point identified. If inferences are not significantly altered by assuming that $\boldsymbol{\lambda} \in \mathbf{A}$ rather than $\boldsymbol{\lambda} = \mathbf{0}$, the evidence provided by the IV regression analysis that assumes $\boldsymbol{\lambda} = \mathbf{0}$ is strengthened. On the other hand, if inferences are significantly altered by assuming that $\boldsymbol{\lambda} \in \mathbf{A}$ rather than $\boldsymbol{\lambda} = \mathbf{0}$, the evidence provided by the IV regression analysis that assumes $\boldsymbol{\lambda} = \mathbf{0}$ is called into question. We allow the sensitivity region for $\boldsymbol{\lambda}$ to vary with the true β (see Section 4.2 for reasons). We specify the sensitivity region $\mathbf{A}$ by $\mathbf{A} = \{\mathbf{A}(\beta_0), \beta_0 \in \mathbb{R}\}$, where $\mathbf{A}(\beta_0)$ is the set of values for $\boldsymbol{\lambda}$ that the subject matter expert considers would be a priori plausible before looking at the data if the true β were to equal β_0. We present in Section 4.1 a method for constructing a SI for β given a sensitivity region $\mathbf{A}$ and in Section 4.2 a model for choosing the sensitivity region $\mathbf{A}$.

4.1 Method for Constructing Sensitivity Interval

A SI, like a CI, is a region of plausible values for β given our assumptions and the data. A $(1-\alpha)$ SI given a sensitivity region $\mathbf{A}$ is a random interval that will contain the true β with probability at least $(1-\alpha)$ under the assumption that $\boldsymbol{\lambda} \in \mathbf{A}(\beta)$. Our approach to forming a SI is to form a joint confidence region for $(\beta, \boldsymbol{\lambda})$ under the assumption that $\boldsymbol{\lambda} \in \mathbf{A}(\beta)$ and then to project this confidence region to form a CI for β. We form a joint $(1-\alpha)$ confidence region for $(\beta, \boldsymbol{\lambda})$ (under the assumption that $\boldsymbol{\lambda} \in \mathbf{A}(\beta)$) by inverting a level α test of $H_0 : \beta = \beta_0, \boldsymbol{\lambda} = \boldsymbol{\lambda}_0$ for each $(\beta_0, \boldsymbol{\lambda}_0)$ such that $\boldsymbol{\lambda}_0 \in \mathbf{A}(\beta_0)$. To test $H_0 : \beta = \beta_0, \boldsymbol{\lambda} = \boldsymbol{\lambda}_0$, we generalize the AR test by testing $H_0 : \boldsymbol{\tau} = \boldsymbol{\lambda}_0$ in the regression model (5). Let $C(\beta, \boldsymbol{\lambda})$ denote the $(1-\alpha)$ joint confidence region for $(\beta, \boldsymbol{\lambda})$ formed using this procedure. Note that it will often be the case that $\{\boldsymbol{\lambda} : (\beta_0, \boldsymbol{\lambda}) \in C(\beta, \boldsymbol{\lambda})\}$ is a strict subset of $\mathbf{A}(\beta_0)$, i.e., the data makes certain values of $\boldsymbol{\lambda}$ implausible if $\beta = \beta_0$ that the subject matter expert considered would be a priori plausible if β were to equal β_0; this is related to the discussion in Section 3 of how the AR test incorporates the ORT. Let $S^*(\mathbf{A}, \mathbf{Y}_N, \mathbf{W}_N, \mathbf{D}_N)$ denote the projection of $C(\beta, \boldsymbol{\lambda})$ into the β subspace of $(\beta, \boldsymbol{\lambda})$ i.e., $S^*(\mathbf{A}, \mathbf{Y}_N, \mathbf{W}_N, \mathbf{D}_N) = \{\beta : (\beta, \boldsymbol{\lambda}) \in C(\beta, \boldsymbol{\lambda})$ for at least one $\boldsymbol{\lambda} \in \mathbb{R}^{\dim(\mathbf{z})}\}$.

Proposition 2. *$S^*(\mathbf{A}, \mathbf{Y}_N, \mathbf{W}_N, \mathbf{D}_N)$ is a $(1-\alpha)$ SI.*

Proof. Suppose the true value of $\boldsymbol{\lambda}$ belongs to $\mathbf{A}(\beta)$ for the true value of β. Because $C(\beta, \boldsymbol{\lambda})$ is a $(1-\alpha)$ confidence region for $(\beta, \boldsymbol{\lambda})$ given that $\boldsymbol{\lambda} \in \mathbf{A}(\beta)$, we have $P((\beta, \boldsymbol{\lambda}) \in C(\beta, \boldsymbol{\lambda})) \geq 1-\alpha$ and consequently $P(\beta \in S^*(\mathbf{A}, \mathbf{Y}_N, \mathbf{W}_N, \mathbf{D}_N)) \geq P((\beta, \boldsymbol{\lambda}) \in C(\beta, \boldsymbol{\lambda})) \geq 1-\alpha$. □

Note that $S^*(\mathbf{A}, \mathbf{Y}_N, \mathbf{W}_N, \mathbf{D}_N)$ might not be an interval. Following Dufour (1997), we obtain an interval by taking $S(\mathbf{A}, \mathbf{Y}_N, \mathbf{W}_N, \mathbf{D}_N) = [\inf\{\beta : \beta \in S^*(\mathbf{A}, \mathbf{Y}_N, \mathbf{W}_N, \mathbf{D}_N)\},$

$\sup\{\beta : \beta \in S^*(\mathbf{A}, \mathbf{Y}_N, \mathbf{W}_N, \mathbf{D}_N)\}]$.

4.2 Model for Choosing Sensitivity Region

A crucial part of sensitivity analysis is choosing the sensitivity region. This choice requires subject matter expertise but methodology can help by providing an interpretable and manageable model for the subject matter expert to think about. Much of the insight from a sensitivity analysis can often be obtained from a relatively simple parametric model (Imbens 2003).

For our model for sensitivity analysis, we assume that there exists an unobserved covariate a such that if $a - E(a|\mathbf{x})$ were added to the analysis as an included exogenous variable, then the IV regression analysis would provide a consistent estimate of β. For example, in Card's study mentioned in the introduction, a might represent motivation. We assume the following model holds in addition to (3): $y_i = \beta w_i + f(\mathbf{x}_i) + \phi(a_i - E(a|\mathbf{x}_i)) + b_i$, $E(b_i|\mathbf{x}_i, \mathbf{z}_i, a_i) = 0$. The function $f(\mathbf{x}_i)$ equals $\boldsymbol{\delta}^T\mathbf{x}_i + \sum_{j=1}^{\dim(\mathbf{z})} \lambda_j E(z_{ij}|\mathbf{x}_i)$. We assume that the $E(z_{ij}|\mathbf{x}_i)$ are linear in some basis expansion for $\mathbf{x}$. Then $f(\mathbf{x}_i) = \boldsymbol{\xi}^T\mathbf{g}(\mathbf{x}_i)$ for some vector $\boldsymbol{\xi}$ and basis functions $\mathbf{g}(\mathbf{x}_i)$ and

$$y_i = \beta w_i + \boldsymbol{\xi}^T\mathbf{g}(\mathbf{x}_i) + \phi(a_i - E(a_i|\mathbf{x}_i)) + b_i, \quad E(b_i|\mathbf{x}_i, \mathbf{z}_i, a_i) = 0. \tag{10}$$

Under model (10), TSLS estimation using $\mathbf{g}(\mathbf{x}_i)$ and $a_i - E(a_i|\mathbf{x}_i)$ as included exogenous variables and $\mathbf{z}$ as the vector of proposed IVs would provide consistent estimates of β, $\boldsymbol{\xi}$ and ϕ. In order for the numerical value of the parameter ϕ to be meaningful, the scale of the unobserved covariate $a - E(a|\mathbf{x})$ must be restricted. We do this by assuming that $Var(a - E(a|\mathbf{x})) = 1$. Furthermore, we assume that a linear model holds for the relationship between a and $(\mathbf{x}, \mathbf{z})$: $E(a|\mathbf{x}, \mathbf{z}) = \boldsymbol{\varpi}^T\mathbf{x} + \boldsymbol{\psi}^T\mathbf{z}$. Under models (3) and (10), $\boldsymbol{\lambda}$ equals $\phi\boldsymbol{\psi}$ because (i) under (3), for all ($\mathbf{x}$,$\mathbf{z}$), we have $E(y - \beta w|\mathbf{x}, \mathbf{z}) = \boldsymbol{\delta}^T\mathbf{x} + \boldsymbol{\lambda}^T\mathbf{z}$ and (ii) under (10), for all $(\mathbf{x}, \mathbf{z})$, we have $E(y - \beta w|\mathbf{x}, \mathbf{z}) = E(\boldsymbol{\xi}^T\mathbf{g}(\mathbf{x}) + \phi(a - E(a|\mathbf{x})) + b|\mathbf{x}, \mathbf{z}) = \boldsymbol{\xi}^T\mathbf{g}(\mathbf{x}) + \phi(\boldsymbol{\psi}^T\mathbf{z} - \boldsymbol{\psi}^T E(\mathbf{z}|\mathbf{x}))$.

The idea of assuming that there exists an unobserved covariate a such that if a were added to the analysis, then the analysis would provide consistent estimates is in the spirit of much work on sensitivity analysis in causal inference, e.g., Rosenbaum (1986) and Imbens (2003). a could represent a composite of several unobserved covariates rather than a single unobserved covariate. In particular, suppose $t_{i1}, \ldots, t_{im}$ are unobserved covariates such that $y_i = \beta w_i + f(\mathbf{x}_i) + \phi_1(t_{i1} - E(t_1|\mathbf{x}_i)) + \cdots + \phi_m(t_{im} - E(t_m|\mathbf{x}_i)) + b_i,\ E(b_i|\mathbf{x}_i, \mathbf{z}_i, t_{i1}, \ldots, t_{im}) = 0$. Then we can set $a_i = (\phi_1 t_{i1} + \cdots + \phi_m t_{im})/SD\{\phi_1(t_1 - E(t_1|\mathbf{x})) + \cdots + \phi_m(t_m - E(t_m|\mathbf{x}))\}$.

We now discuss an approach to specifying a range of plausible values for $\boldsymbol{\lambda} = \phi\boldsymbol{\psi}$ given $\beta = \beta_0$ in terms of interpretable parameters. We consider ϕ and $\boldsymbol{\psi}$ separately. For specifying a plausible range of values for ϕ, it is useful to rewrite (10) as $y_i - \beta w_i + \beta w^* = \beta w^* + \boldsymbol{\xi}^T \mathbf{g}(\mathbf{x}_i) + \phi(a_i - E(a_i|\mathbf{x}_i)) + b_i,\ E(b_i|\mathbf{x}_i, \mathbf{z}_i, a_i) = 0$ for any w^*. The parameter ϕ measures the strength of the relationship between the unobserved covariate $a - E(a|\mathbf{x})$ and the potential outcome $y^{(w^*)} = y - \beta(w^{obs} - w^*)$. Specifically,

$$\frac{R^2_{unobs}}{R^2_{obs}} = \frac{\frac{\phi^2}{Var(y - \beta(w^{obs} - w^*))}}{\frac{Var(\boldsymbol{\xi}^T \mathbf{g}(\mathbf{x}))}{Var(y - \beta(w^{obs} - w^*))}}$$

is the proportion of variation in the potential outcomes $y^{(w^*)}$ explained by the regression on the unobserved covariate $a - E(a|\mathbf{x})$ relative to the proportion of variation explained by the regression on the observed covariates $\mathbf{x}$. For a realistic w^*, the proportion R^2_{unobs}/R^2_{obs} is an interpretable quantity. If $R^2_{unobs}/R^2_{obs} = 0$, the proposed IVs $\mathbf{z}$ are valid; if $0 < R^2_{unobs}/R^2_{obs} < 1$, the proposed IVs $\mathbf{z}$ are invalid but the unobserved covariate $a - E(a|\mathbf{x})$ is a less strong predictor of the potential outcomes than the currently included vector of exogenous variables $\mathbf{x}$; and if $R^2_{unobs}/R^2_{obs} > 1$, the proposed IVs are invalid and $a - E(a|\mathbf{x})$ is a stronger predictor of the potential outcomes than $\mathbf{x}$. A subject matter expert might want to provide a different range of plausible values for R^2_{unobs}/R^2_{obs} for when ϕ is positive compared to for ϕ negative. The relationship between R^2_{unobs}/R^2_{obs} and ϕ depends on $Var(\boldsymbol{\xi}^T \mathbf{g}(\mathbf{x}))$. For $\beta = \beta_0$, we can consistently estimate $\boldsymbol{\xi}$ by regressing $y - \beta_0 w$ on $\mathbf{g}(\mathbf{x})$; denote this estimate by $\hat{\boldsymbol{\xi}}_{\beta_0}$. For given values of sign(ϕ) and R^2_{unobs}/R^2_{obs}, we estimate ϕ to be

$$\hat{\phi} = \text{sign}(\phi)\sqrt{(R^2_{unobs}/R^2_{obs})\hat{Var}(\hat{\boldsymbol{\xi}}^T_{\beta_0}\mathbf{g}(\mathbf{x}))}. \tag{11}$$

Accordingly, if the expert's range for R^2_{unobs}/R^2_{obs} for $\phi > 0$ is $0 \leq R^2_{unobs}/R^2_{obs} \leq \max(R^2_{unobs}/R^2_{obs})^+$ and for $\phi < 0$ is $0 \leq R^2_{unobs}/R^2_{obs} \leq \max(R^2_{unobs}/R^2_{obs})^-$, we take our sensitivity analysis range for ϕ when $\beta = \beta_0$ to be

$$-\sqrt{\hat{Var}(\hat{\boldsymbol{\xi}}^T_{\beta_0}\mathbf{g}(\mathbf{x}))}\sqrt{\max(R^2_{unobs}/R^2_{obs})^-} \leq \phi \leq \sqrt{\hat{Var}(\hat{\boldsymbol{\xi}}^T_{\beta_0}\mathbf{g}(\mathbf{x}))}\sqrt{\max(R^2_{unobs}/R^2_{obs})^+}.$$

Note that our approach specifies different ranges of ϕ for different values of β. We view R^2_{unobs}/R^2_{obs} as a quantity that experts can think about plausible values for without thinking about the true value of β. Because the relationship between R^2_{unobs}/R^2_{obs} and ϕ depends on β, we allow for different ranges of ϕ for different β's. Our approach of calibrating the effect of an unobserved covariate by comparing it to the effect of observed covariates has been used in sensitivity analysis for ignorable treatment assignment by Rosenbaum (1986) and Imbens (2003) among others. Another approach to choosing a range for ϕ is to directly specify a range $\phi_{\min} \leq \phi \leq \phi_{\max}$, keeping in mind that ϕ is the change in the mean of the potential outcome $y^{(w*)}$ that is associated with a one standard deviation change in the unobserved covariate $a - E(a|\mathbf{x})$.

For specifying a range of plausible values for $\boldsymbol{\psi}$, we specify plausible values for the magnitude and direction of $\boldsymbol{\psi}$ using the norm $\|\boldsymbol{\psi}\| = \sqrt{Var\{\boldsymbol{\psi}^T(\mathbf{z} - E(\mathbf{z}|\mathbf{x}))\}}$. For the magnitude of $\boldsymbol{\psi}$, note that under our sensitivity analysis model, $E(\{a - E(a|\mathbf{x})\}|\{\mathbf{z} - E(\mathbf{z}|\mathbf{x})\}) = \boldsymbol{\psi}^T(\mathbf{z} - E(\mathbf{z}|\mathbf{x}))$. Thus, using that $Var(a - E(a|\mathbf{x})) = 1$, $\|\boldsymbol{\psi}\|^2$ is the R^2 for the regression of $a - E(a|\mathbf{x})$ on $\mathbf{z} - E(\mathbf{z}|\mathbf{x})$; we denote this R^2 by $R^2_{\mathbf{z}-E(\mathbf{z}|\mathbf{x})}$. Note that $\boldsymbol{\lambda} = \phi\boldsymbol{\psi} = \sqrt{R^2_{unobs}/R^2_{obs}}\text{sign}(\phi)\sqrt{Var(\boldsymbol{\xi}^T_{\beta_0}\mathbf{g}(\mathbf{x}))}\sqrt{R^2_{\mathbf{z}-E(\mathbf{x}|\mathbf{x})}}(\boldsymbol{\psi}/\|\boldsymbol{\psi})$, so that equal values of the product $(R^2_{unobs}/R^2_{obs})R^2_{\mathbf{z}-E(\mathbf{z}|\mathbf{x})}$ produce equal values of $\boldsymbol{\lambda}$ if $\text{sign}(\phi)$ and $\boldsymbol{\psi}/\|\boldsymbol{\psi}\|$ are kept fixed. For specifying the direction of $\boldsymbol{\psi}$, we transform $\boldsymbol{\psi}$ to $\boldsymbol{\psi}^*$ where $\boldsymbol{\psi}^*$ is the vector of coefficients for the regression of $a - E(a|\mathbf{x})$ on the standardized values of $\mathbf{z} - E(\mathbf{z}|\mathbf{x})$: $\boldsymbol{\psi}^* = \{\psi_1 SD(z_1 - E(z_1|\mathbf{x})), \ldots, \psi_{\dim(\mathbf{z})}SD(z_{\dim(\mathbf{z})} - E(z_{\dim(\mathbf{z})}|\mathbf{x}))\}$. Then we specify a range of plausible values for $\{\psi^*_2/\psi^*_1, \ldots, \psi^*_{\dim(\mathbf{z})}/\psi^*_1,$

$\mathrm{sign}(\psi_1^*)\}$. The sensitivity parameters $\{\psi_2^*/\psi_1^*, \ldots, \psi_{\dim(\mathbf{z})}^*/\psi_1^*, \mathrm{sign}(\psi_1^*)\}$ specify the relative magnitude and sign of the coefficients in the regression of $a - E(a|\mathbf{x})$ on the standardized values of $\mathbf{z} - E(\mathbf{z}|\mathbf{x})$. For example, if $\psi_2^*/\psi_1^* > 0$, then $z_1 - E(z_1|\mathbf{x})$ and $z_2 - E(z_2|\mathbf{x})$ are either both positively or both negatively associated with the unobserved covariate $a - E(a|\mathbf{x})$; if $\psi_2^*/\psi_1^* < 0$, then $z_1 - E(z_1|\mathbf{x})$ and $z_2 - E(z_2|\mathbf{x})$ are associated with $a - E(a|\mathbf{x})$ in opposite directions. Also, if $|\psi_2^*/\psi_1^*| > 1$, then $z_2 - E(z_2|\mathbf{x})$ has a higher absolute correlation with the unobserved covariate $a - E(a|\mathbf{x})$ than $z_1 - E(z_1|\mathbf{x})$, while if $|\psi_2^*/\psi_1^*| < 1$, then $z_1 - E(z_1|\mathbf{x})$ has the higher absolute correlation. Note that we might want to consider different ranges of plausible $\{(R^2_{\mathbf{z}-E(\mathbf{z}|\mathbf{x})}), \psi_2^*/\psi_1^*, \ldots, \psi_{\dim(\mathbf{z})}^*/\psi_1^*, \mathrm{sign}(\psi_1^*)\}$ for $\phi > 0$ and $\phi < 0$. However, we assume that the range of plausible $\{R^2_{\mathbf{z}-E(\mathbf{z}|\mathbf{x})}, \psi_2^*/\psi_1^*, \ldots, \psi_{\dim(\mathbf{z})}^*/\psi_1^*, \mathrm{sign}(\psi_1^*)\}$ does not depend on the magnitude of ϕ or on β_0.

For a given set of sensitivity parameters $\{R^2_{\mathbf{z}-E(\mathbf{z}|\mathbf{x})}, \psi_2^*/\psi_1^*, \ldots, \psi_{\dim(\mathbf{z})}^*/\psi_1^*, \mathrm{sign}(\psi_1^*)\}$, we estimate the associated sensitivity parameters $(\psi_1, \ldots, \psi_{\dim(\mathbf{z})})$ by:

$$
\begin{aligned}
R^2_{\mathbf{z}-E(\mathbf{z}|\mathbf{x})} &= \hat{Var}(\hat{\boldsymbol{\psi}}^T(\mathbf{z} - \hat{E}(\mathbf{z}|\mathbf{x}))) \\
\hat{\psi}_2/\hat{\psi}_1 &= (\psi_2^* \hat{SD}\{z_1 - \hat{E}(z_1|\mathbf{x})\})/(\psi_1^* \hat{SD}\{z_2 - \hat{E}(z_2|\mathbf{x})\}), \\
&\ldots \\
\hat{\psi}_{\dim(\mathbf{z})}/\hat{\psi}_1 &= (\psi_{\dim(\mathbf{z})}^* \hat{SD}\{z_1 - \hat{E}(z_1|\mathbf{x})\})/(\psi_1^* \hat{SD}\{z_{\dim(\mathbf{z})} - \hat{E}(z_{\dim(\mathbf{z})}|\mathbf{x})\}). \quad (12)
\end{aligned}
$$

Here is a summary of our sensitivity analysis procedure. We specify the sensitivity parameters $\max(R^2_{unobs}/R^2_{obs})^+, \max(R^2_{unobs}/R^2_{obs})^-$ and ranges of $\{R^2_{\mathbf{z}-E(\mathbf{z}|\mathbf{x})}, \psi_2^*/\psi_1^*, \ldots, \psi_{\dim(\mathbf{z})}^*/\psi_1^*, \mathrm{sign}(\psi_1^*)\}$ that we consider plausible when $\phi > 0$ and $\phi < 0$. This specifies a range of plausible $\mathbf{s} = \{R^2_{unobs}/R^2_{obs}, \mathrm{sign}(\phi),\ R^2_{\mathbf{z}-E(\mathbf{z}|\mathbf{x})}, \psi_2^*/\psi_1^*, \ldots, \psi_{\dim(\mathbf{z})}^*/\psi_1^*, \mathrm{sign}(\psi_1^*)\}$ given $\beta = \beta_0$. The combination of an $\mathbf{s}$ and the population distribution of $(y, \mathbf{x}, \mathbf{z})$ determines ϕ and $\boldsymbol{\psi}$ given $\beta = \beta_0$. Because the distribution of $(y, \mathbf{x}, \mathbf{z})$ is unknown, we estimate ϕ and $\boldsymbol{\psi}$ (which determine $\boldsymbol{\lambda} = \phi\boldsymbol{\psi}$) using (11) and (12). We then take $\mathbf{A}(\beta_0)$ to be the set of these estimates of $\boldsymbol{\lambda}$ for the range of $\mathbf{s}$ that we consider plausible given $\beta = \beta_0$. Based on $\mathbf{A} = \{\mathbf{A}(\beta_0), \beta_0 \in \mathbb{R}\}$, we compute the SI

$S(\mathbf{A}, \mathbf{Y}_N, \mathbf{W}_N, \mathbf{D}_N)$ using the method of Section 4.1. An R function for computing $S(\mathbf{A}, \mathbf{Y}_N, \mathbf{W}_N, \mathbf{D}_N)$ that employs a grid search is available from the author.

5. Extensions to Panel Data and Heterogeneous Treatment Effects

In model (3), we assume that $(w_i, \mathbf{x}_i, \mathbf{z}_i, u_i)$ are iid random vectors. In this section, we consider the following extension of model (3):

$$y_{it} = w_{it}\beta + \boldsymbol{\delta}^T\mathbf{x}_{it} + \boldsymbol{\lambda}^T\mathbf{z}_{it} + u_{it}, i = 1, \ldots, N, \quad t = 1, \ldots, T,$$
$$(\mathbf{W}_i = (w_{i1}, \ldots, w_{iT})^T, \mathbf{X}_i = (\mathbf{x}_{i1}^T, \ldots, \mathbf{x}_{iT}^T), \mathbf{Z}_i = (\mathbf{z}_{i1}^T, \ldots, \mathbf{z}_{iT})^T, \mathbf{U}_i = (u_{i1}, \ldots, u_{iT})^T)$$
$$\text{are independent but not necessarily iid random matrices and } E(u_i|\mathbf{x}_i, \mathbf{z}_i) = 0. \quad (13)$$

For model (13), the TSLS estimator of β is consistent but the asymptotic distribution is not (4) (White, 1982).

One motivation for model (13) is when we have panel data with N units and T time periods and the additive linear constant effects models continues to hold, i.e., $y_{it}^{(w^*)} = \beta w^* + \boldsymbol{\delta}^T\mathbf{x}_{it} + \boldsymbol{\lambda}^T\mathbf{z}_{it} + u_{it}$; model (13) allows for $u_{i1}, \ldots, u_{iT}$ to be correlated. Model (13) also accommodates stratified cross-sectional survey data and heteroskedasticity.

A second motivation for considering model (13) is to allow for heterogeneous treatment effects. Suppose our model for potential outcomes is

$$y_i^{(w^*)} = \beta_i w^* + \boldsymbol{\delta}^T\mathbf{x}_i + \boldsymbol{\lambda}^T\mathbf{z}_i + u_i, \quad E(u_i|\mathbf{x}_i, \mathbf{z}_i) = 0. \quad (14)$$

Unit i's treatment effect is β_i. Let $\beta = E(\beta_i)$ be the average treatment effect over the population. We can express the observed data from model (14) as

$$y_i = \beta w_i + \boldsymbol{\delta}^T\mathbf{x}_i + \boldsymbol{\lambda}^T\mathbf{z}_i + (\beta_i - \beta)w_i + u_i, \quad E(u_i|\mathbf{x}_i, \mathbf{z}_i) = 0. \quad (15)$$

If we make the further assumption that

$$(\beta_i - \beta) \text{ is independent of } w_i|\mathbf{x}_i, \mathbf{z}_i, \quad (16)$$

then (15) is equivalent to (13) with $T = 1$. Assumption (16) says that units do not select their treatment levels (w_i) based on the gains they would experience from treatment (β_i) (Wooldridge, 1997). If assumption (16) does not hold, then the TSLS estimator might not converge to β. Angrist and Imbens (1995) discuss properties of TSLS estimates and the ORT when (16) does not hold.

For model (13), our procedure for choosing the sensitivity region can be carried out using the method of Section 4.2. For constructing the SI, we use the same procedure as in Section 4.1 except that in forming the confidence region for $\boldsymbol{\tau}$, we use the following (Huber, 1967) estimate of the covariance matrix of $(\hat{\boldsymbol{\chi}}_{OLS}, \hat{\boldsymbol{\tau}}_{OLS})$: $\hat{V}_{robust} = (\mathbf{D}_N^T\mathbf{D}_N)^{-1}(\sum_{i=1}^N[\mathbf{X}_i, \mathbf{Z}_i]^T\hat{\mathbf{f}}_i\hat{\mathbf{f}}_i^T[\mathbf{X}_i, \mathbf{Z}_i])(\mathbf{D}_N^T\mathbf{D}_N)^{-1}$, where $\hat{\mathbf{f}}_i = (y_{i1} - w_i\beta_0 - \hat{\boldsymbol{\chi}}_{OLS}^T\mathbf{x}_{i1} - \hat{\boldsymbol{\tau}}_{OLS}^T\mathbf{z}_{i1}, \ldots, y_{iT} - w_i\beta_0 - \hat{\boldsymbol{\chi}}_{OLS}^T\mathbf{x}_{iT} - \hat{\boldsymbol{\tau}}_{OLS}^T\mathbf{z}_{iT})$ are the residuals for unit i from OLS estimation of (5) and $\mathbf{D}_N$ is the $(TN) \times (\dim(\mathbf{x}) + \dim(\mathbf{z}))$ matrix formed by stacking $[\mathbf{X}_1, \mathbf{Z}_1], \ldots, [\mathbf{X}_N, \mathbf{Z}_N]$. Under regularity conditions, $\sqrt{NT}\hat{V}_{robust}$ is a consistent estimator of the asymptotic variance of $(\hat{\boldsymbol{\chi}}_{OLS}, \hat{\boldsymbol{\tau}}_{OLS})$ for $N \to \infty$, T fixed in the presence of heteroskedasticity and arbitrary patterns of correlation among $u_{i1}, \ldots, u_{iT}$ within units (White, 1984, pp. 134-142). For model (13), the analogue of the ORT test statistic J_N is the test statistic of Hansen (1982).

6. Illustrative Example: Demand for Food

As an illustrative example, we consider an IV regression model proposed by Bouis and Haddad (1990) for modeling the causal effect of income changes on food expenditure in a study of Philippine farm households. In the study, 406 households, obtained by a stratified random sample, were interviewed at four time points. y_{it} is the ith household's food expenditure at time t, w_{it} is the ith household's log income at time t and $\mathbf{x}_{it}$ consists of mother's education, father's education, mother's age, father's age, mother's nutritional knowledge, price of corn, price of rice, population density of the municipality, number of household members expressed in adult equivalents and dummy variables for the round of the survey. We consider model (1) for this data. The parameter β represents the causal effect on a household's short run food expenditures of a one unit increase in a household's log income, where the income increase arises through a yearly lump sum payment that households expect will continue permanently. We assume that over the short run period between the time of the income increase and the measurement of food expenditures, a household does not alter its farm's production level. Rather than focus directly on β, we focus on the following more interpretable quantity: the income elasticity of food demand at the mean level of food expenditure. This is the percent change in food expenditure caused by a one percent increase in income for households currently spending at the mean food expenditure level and we denote it by η. The mean food expenditure of households is 31.14 pesos per capita per week so that $\eta = 100\beta(\log 1.01)/31.14 = 0.032\beta$.

Bouis and Haddad were concerned that regression of y on w and $\mathbf{x}$ would not provide an unbiased estimate of β because of unobserved confounding variables. In particular, because farm households make production and consumption decisions simultaneously and there are multiple incomplete markets in the study area, the households' production decisions (which affect their log income w) are associated with their preferences (which are partially unobserved) according to microeconomic theory (Bardhan and Udry, 1999, Ch. 2). To consistently estimate β, Bouis and Haddad proposed two IVs, cultivated area per capita (z_1) and worth of assets (z_2). Bouis and Haddad's reasoning behind proposing these variables as IVs is that "land availability is assumed to be a constraint in the short run, and therefore exogenous to the household decisionmaking process." Following Bouis and Haddad, we assume model (3) holds. Assumption (A2') for these proposed IVs appears to hold – the test of $H_0 : \boldsymbol{\gamma} = \mathbf{0}$ has a p-value of < 0.01.

Using the proposed IVs z_1 and z_2, the TSLS estimate of the income elasticity of food demand η is 0.65 with a 95% CI (assuming $\boldsymbol{\lambda} = \mathbf{0}$) of $(0.49, 0.85)$. This CI (and the SIs below) are computed using the method of Section 5 to account for the stratified random sampling design, the repeated measurements on households and the possibility of heterogeneous treatment effects. One concern about the validity of the proposed IVs is that a household's unobserved preferences might have influenced a household's past choices on land acquisition and asset accumulation, which are reflected in a househould's present cultivated area per capita and worth of assets. The p-value for the ORT (Hansen's 1982 version) is 0.12, indicating that there is no evidence to reject the joint validity of the proposed IVs. But, as discussed in Section 3, the ORT is not consistent and a sensitivity analysis is useful to clarify the extent to which inferences vary over plausible violations of the validity of the proposed IVs.

6.1 Sensitivity Analysis for the Food Demand Study

For our sensitivity analysis, we always set $\max(R^2_{unobs}/R^2_{obs})^+ = \max(R^2_{unobs}/R^2_{obs})^- \equiv \max(R^2_{unobs}/R^2_{obs})$ and consider ranges of plausible values for $\{R^2_{\mathbf{z}-E(\mathbf{z}|\mathbf{x})}, \psi_2^*/\psi_1^*, \text{sign}(\psi_1^*)\}$ that are the same for $\phi > 0$ and $\phi < 0$ and that have the form: $\{0 \leq R^2_{\mathbf{z}-E(\mathbf{z}|\mathbf{x})} \leq \max R^2_{\mathbf{z}-E(\mathbf{z}|\mathbf{x})}\} \times$ {range of ψ_2^*/ψ_1^*}$\times$ $\{\text{sign}(\psi_1^*) = 1, \text{sign}(\psi_1^*) = -1\}$. For the range of ψ_2^*/ψ_1^*, we consider either a point or an interval. We estimate $E(z_1|\mathbf{x})$ and $E(z_2|\mathbf{x})$ by considering quadratic response surfaces and using the variable selection method of Zheng and Loh (1995) with $h_n(k) = k \log n$. Table 1 reports SIs for various sensitivity regions. To give the reader a sense of the economic meaning of the values of η that are in the SIs, the following are estimates of the income elasticities of demand for various goods compiled by Nicholson (1995): medical services, 0.22; beer, 0.38; cigarettes, 0.50; electricity, 0.61; wine, 0.97; transatlantic air travel, 1.40; and automobiles, 3.00.

In examining Table 1, we first consider $\psi_2^*/\psi_1^* = 1$, which corresponds to $z_1 - E(z_1|\mathbf{x})$ and $z_2 - E(z_2|\mathbf{x})$ being equally correlated with the unobserved covariate $a - E(a|\mathbf{x})$. For $\psi_2^*/\psi_1^* = 1$ and $\max(R^2_{unobs}/R^2_{obs}) = 0.1$, which means the unobserved covariate $a - E(a|\mathbf{x})$ has at most a modest association with the potential outcome $y^{(w^*)}$ relative to the association between $\mathbf{x}$ and $y^{(w^*)}$, the SI for $\max R^2_{\mathbf{z}-E(\mathbf{z}|\mathbf{x})} = 0.5$ is $(0.33, 1.10)$ and the SI for $\max R^2_{\mathbf{z}-E(\mathbf{z}|\mathbf{x})} = 1$ is $(0.25, 1.28)$. These SIs are both more than twice as wide as the CI that assumes $\boldsymbol{\lambda} = \mathbf{0}$. For an unobserved covariate $a - E(a|\mathbf{x})$ with a potentially stronger but still moderate association with

$y^{(w^*)}$ relative to the association between $\mathbf{x}$ and $y^{(w^*)}$ of $\max(R^2_{unobs}/R^2_{obs}) = 0.25$, the SI for $\psi_2^*/\psi_1^* = 1$, $\max R^2_{\mathbf{z}-E(\mathbf{z}|\mathbf{x})} = 0.5$ is $(0.21, 1.38)$ and the SI for $\psi_2^*/\psi_1^* = 1$, $\max R^2_{\mathbf{z}-E(\mathbf{z}|\mathbf{x})} = 1$ is $(-0.03, 2.53)$. These intervals are wide in economic meaning – the lower ends are lower than the income elasticity of the "necessity" of medical services and the upper ends are close to or higher than the income elasticity of the "luxury" of transatlantic air travel using the estimates from Nicholson (1995). For larger values of $\max(R^2_{unobs}/R^2_{obs})$ of 0.5, 0.75 and 1, the SIs become even wider, blowing up to $(-\infty, \infty)$ for $\max(R^2_{unobs}/R^2_{obs}) = 0.75$ or 1 for both $R^2_{\mathbf{z}-E(\mathbf{z}|\mathbf{x})} = 0.5$ and 1. For $\psi_2^*/\psi_1^* = 0.5$ and $\psi_2^*/\psi_1^* = 2$ (meaning that $z_1 - E(z_1|\mathbf{x})$ and $z_2 - E(z_2|\mathbf{x})$ are correlated with $a - E(a|\mathbf{x})$ in the same direction and the magnitude of the correlation of one is twice that of the other), the SIs are similar to those for $\psi_2^*/\psi_1^* = 1$. For $\psi_2^*/\psi_1^* = 1000$, 0.0001 and -1000 (meaning that one of $z_1 - E(z_1|\mathbf{x})$ or $z_2 - E(z_2|\mathbf{x})$ is much more strongly correlated with $a - E(a|\mathbf{x})$ than the other), the SIs are slightly shorter than for $\psi_2^*/\psi_1^* = 1$ for $\max(R^2_{unobs}/R^2_{obs}) = 0.1$ and considerably shorter for $\max(R^2_{unobs}/R^2_{obs}) = 0.25, 0.5, 0.75$ or 1. For $\psi_2^*/\psi_1^* = -0.5$, -1 and -2 (meaning that $z_1 - E(z_1|\mathbf{x})$ and $z_2 - E(z_2|\mathbf{x})$ are correlated with $a - E(a|\mathbf{x})$ in opposite directions and the magnitude of the correlations is not necessarily the same but is of the same order of magnitude), the SIs are considerably shorter than for $\psi_2^*/\psi_1^* = 1000$, 2, 1, 0.5, 0.0001 or -1000. The last three rows of the top half and bottom half of Table 1 show SIs when the ranges for ψ_2^*/ψ_1^* of $[-.5, 2]$, $[-2, -.5]$ and $[-1000, 1000]$ are considered.

We now discuss how ψ_2^*/ψ_1^* affects the SI. Figure 1 shows the SI for fixed ψ_2^*/ψ_1^* as ψ_2^*/ψ_1^* varies from -5 to 5 for $\max R^2_{\mathbf{z}-E(\mathbf{z}|\mathbf{x})} = 0.5$ and $\max(R^2_{unobs}/R^2_{obs}) = 0.25$. The SI is widest for $\psi_2^*/\psi_1^* = 1$ and narrowest for $\psi_2^*/\psi_1^* = -1$. To illustrate the role of ψ_2^*/ψ_1^* in the SI, Figure 2 shows how the tests of $H_0 : \eta = 1$ and $H_0 : \eta = 1.5$ are constructed. First, consider the range $-\infty < \psi_2^*/\psi_1^* < \infty$. The sensitivity region is an ellipse. For this sensitivity region, we accept $H_0 : \eta = 1$ because the sensitivity region intersects the ellipse of $\boldsymbol{\lambda}$'s that are plausible given $\eta = 1$ (i.e., the confidence region for $\boldsymbol{\tau}$ in the regression (5) with $\beta_0 = 1/.032$); we reject $\eta = 1.5$ because there is no corresponding intersection. For a fixed ψ_2^*/ψ_1^*, the sensitivity region is the line segment that is the intersection of the line which ψ_2^*/ψ_1 specifies $\boldsymbol{\lambda}$ lies on, $\ell'(\psi_2^*/\psi_1^*) = \{(\lambda_1, \lambda_2) : \lambda_2 SD(z_2 - E(z_2|\mathbf{x})) = (\psi_2^*/\psi_1^*)\lambda_1 SD(z_1 - E(z_1|\mathbf{x}))\}$, and the sensitivity region for $-\infty < \psi_2^*/\psi_1^* < \infty$. We accept $H_0 : \eta = \eta_0$ if this line segment intersects the ellipse of $\boldsymbol{\lambda}$'s that are plausible given $H_0 : \eta = \eta_0$. For example, for $\max(R^2_{\mathbf{z}-E(\mathbf{z}|\mathbf{x})}) = 0.5$ and $\max(R^2_{unobs}/R^2_{obs}) = 0.25$, we accept $H_0 : \eta = 1$ for $\psi_2^*/\psi_1^* = 1$ but reject it for $\psi_2^*/\psi_1^* = -1$ as Figure 2 shows.

To illuminate how ψ_2^*/ψ_1^* affects the SI, we consider what happens under model (6) as the sample size increases to infinity. Suppose the line $\ell'(\psi_2^*/\psi_1^*)$ which ψ_2^*/ψ_1^* specifies $\boldsymbol{\lambda}$ lies on equals the line ℓ that $\boldsymbol{\lambda}$ is identified to lie on (see (9)); this can only happen if ℓ crosses through $\mathbf{0}$ and $\psi_2^*/\psi_1^* = [\gamma_2 SD(z_2 - E(z_2|\mathbf{x}))]/[\gamma_1 SD(z_1 - E(z_1|\mathbf{x}))]$. Then the SI converges to the set of β's for which $(\boldsymbol{\kappa} - \beta\boldsymbol{\gamma}) \in \mathbf{A}(\beta)$. Suppose instead $\ell'(\psi_2^*/\psi_1^*)$ does not equal ℓ. Let $\boldsymbol{\lambda}^*$ be the point at which $\ell'(\psi_2^*/\psi_1^*)$ intersects ℓ and let β^* be the point that satisfies $\boldsymbol{\lambda}^* = (\boldsymbol{\kappa} - \beta^*\boldsymbol{\gamma})$. If $(\boldsymbol{\kappa} - \beta^*\boldsymbol{\gamma}) \in \mathbf{A}(\beta^*)$, then the SI for ψ_2^*/ψ_1^* converges to β^*. If $(\boldsymbol{\kappa} - \beta^*\boldsymbol{\gamma}) \notin \mathbf{A}(\beta^*)$ or $\ell'(\psi_2^*/\psi_1^*)$ is parallel but not equal to ℓ, then the SI converges to the empty set.

6.2 Comparison of SIs for One Proposed IV vs. Two Proposed IVs

This section addresses whether using additional proposed IVs reduces the length of the SI compared to using just one proposed IV, a question which is important for designing observational studies using IVs. For one IV, a SI can be computed as in Section 4 except that ψ is one-dimensional. Suppose that ψ is restricted to be ≥ 0. Then the sensitivity region can be characterized by the sensitivity parameters $\max(R^2_{unobs}/R^2_{obs})^+$, $\max(R^2_{unobs}/R^2_{obs})^-$ and $\max(R^2_{z-E(z|\mathbf{x})})$; to check whether $\beta = \beta_0$ is in the SI, we check whether the line segment $-\sqrt{R^2_{z,E(z|\mathbf{x})}/\hat{Var}(z - \hat{E}(z|\mathbf{x}))}$ $\sqrt{\hat{Var}(\hat{\boldsymbol{\xi}})_{\beta_0}\mathbf{g}(\mathbf{x}) \max(R^2_{unobs}/R^2_{obs})^-} \leq \lambda \leq \sqrt{R^2_{z-E(z|x)}/\hat{Var}(z - \hat{E}(z|\mathbf{x}))}$ $\sqrt{\hat{Var}(\hat{\boldsymbol{\xi}}_{\beta_0}\mathbf{g}(\mathbf{x})) \max(R^2_{unobs}/R^2_{obs})^+}$ intersects the CI of λ's given $\beta = \beta_0$ from the

max $R^2_{\mathbf{z}-E(\mathbf{z}\|\mathbf{x})}$	ψ_2^*/ψ_1^*	$\max(R^2_{unobs}/R^2_{obs})^+ = \max(R^2_{unobs}/R^2_{obs})^-$				
		0.1	0.25	0.5	0.75	1
0.5	1000	$(0.37, 1.05)$	$(0.32, 1.21)$	$(0.29, 1.37)$	$(0.29, 1.38)$	$(0.29, 1.38)$
	2	$(0.34, 1.09)$	$(0.24, 1.36)$	$(0.08, 2.25)$	$(-0.12, 48.57)$	$(-\infty, \infty)$
	1	$(0.33, 1.10)$	$(0.21, 1.38)$	$(-0.03, 2.53)$	$(-\infty, \infty)$	$(-\infty, \infty)$
	0.5	$(0.34, 1.08)$	$(0.22, 1.34)$	$(-0.03, 2.14)$	$(-\infty, \infty)$	$(-\infty, \infty)$
	0.0001	$(0.37, 1.03)$	$(0.30, 1.17)$	$(0.29, 1.35)$	$(0.29, 1.44)$	$(0.29, 1.44)$
	-0.5	$(0.46, 0.91)$	$(0.46, 0.91)$	$(0.46, 0.91)$	$(0.46, 0.91)$	$(0.46, 0.91)$
	-1	$(0.49, 0.85)$	$(0.49, 0.85)$	$(0.49, 0.85)$	$(0.49, 0.85)$	$(0.49, 0.85)$
	-2	$(0.45, 0.92)$	$(0.45, 0.92)$	$(0.45, 0.92)$	$(0.45, 0.92)$	$(0.45, 0.92)$
	-1000	$(0.37, 1.05)$	$(0.32, 1.21)$	$(0.29, 1.37)$	$(0.29, 1.37)$	$(0.29, 1.37)$
	$[0.5, 2]$	$(0.33, 1.10)$	$(0.21, 1.38)$	$(-0.05, 2.53)$	$(-\infty, \infty)$	(∞, ∞)
	$[-2, -0.5]$	(0.45,0.92)	(0.45,0.92)	(0.45,0.92)	(0.45,0.92)	(0.45,0.92)
	$[-1000, 1000]$	$(0.33, 1.10)$	$(0.21, 1.38)$	$(-0.05, 2.53)$	$(-\infty, \infty)$	$(-\infty, \infty)$
1	1000	$(0.33, 1.16)$	$(0.29, 1.37)$	$(0.29, 1.38)$	$(0.29, 1.38)$	$(0.29, 1.38)$
	2	$(0.27, 1.27)$	$(0.08, 2.25)$	$(-\infty, \infty)$	$(-\infty, \infty)$	$(-\infty, \infty)$
	1	$(0.25, 1.28)$	$(-0.03, 2.53)$	$(-\infty, \infty)$	$(-\infty, \infty)$	$(-\infty, \infty)$
	0.5	$(0.26, 1.25)$	$(-0.03, 2.14)$	$(-\infty, \infty)$	$(-\infty, \infty)$	$(-\infty, \infty)$
	0.0001	$(0.32, 1.13)$	$(0.29, 1.35)$	$(0.29, 1.44)$	$(0.29, 1.44)$	$(0.29, 1.44)$
	-0.5	$(0.46, 0.91)$	$(0.46, 0.91)$	$(0.46, 0.91)$	$(0.46, 0.91)$	$(0.46, 0.91)$
	-1	$(0.49, 0.85)$	$(0.49, 0.85)$	$(0.49, 0.85)$	$(0.49, 0.85)$	$(0.49, 0.85)$
	-2	$(0.45, 0.92)$	$(0.45, 0.92)$	$(0.45, 0.92)$	$(0.45, 0.92)$	$(0.45, 0.92)$
	-1000	$(0.33, 1.16)$	$(0.29, 1.37)$	$(0.29, 1.37)$	$(0.29, 1.37)$	$(0.29, 1.37)$
	$[0.5, 2]$	$(0.25, 1.28)$	$(-0.05, 2.53)$	(∞, ∞)	$(-\infty, \infty)$	$(-\infty, \infty)$
	$[-2, -0.5]$	(0.45,0.92)	(0.45,0.92)	(0.45,0.92)	(0.45,0.92)	(0.45,0.92)
	$[-1000, 1000]$	$(0.25, 1.28)$	$(-0.05, 2.53)$	(∞, ∞)	$(-\infty, \infty)$	$(-\infty, \infty)$

Table 1: SIs for the income elasticity of food demand (η) for various settings of the four sensitivity parameters $\max\left(R^2_{\text{unobs}}/R^2_{\text{obs}}\right)^+$, $\max\left(R^2_{\text{unobs}}/R^2_{\text{obs}}\right)^-$, ψ_2^*/ψ_1^* and $\max R^2_{\mathbf{z}-E(\mathbf{z}|\mathbf{x})}$.

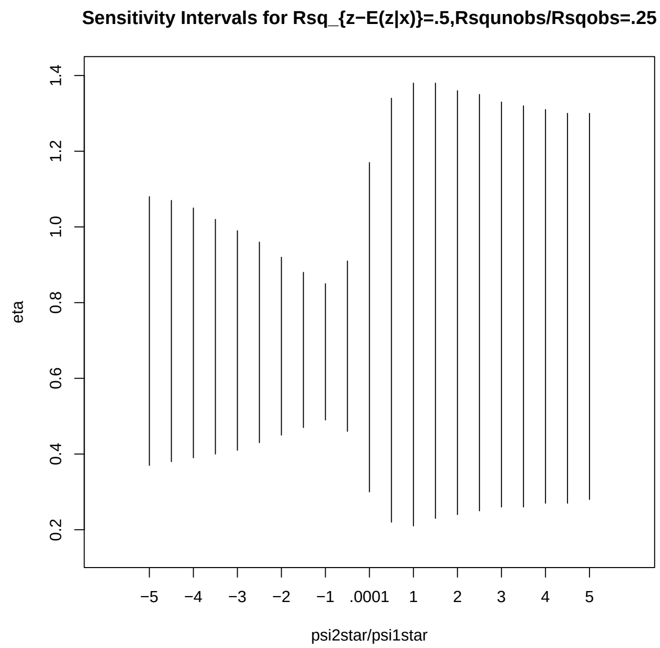

Figure 1: The SIs as ψ_2^*/ψ_1^* varies from -5 to 5 for $\max R^2_{\mathbf{z}-E(\mathbf{z}|\mathbf{x})} = 0.5$ and $\max(R^2_{unobs}/R^2_{obs}) = 0.25$.

regression (5). In comparing using two proposed IVs to one proposed IV, two potential advantages of two proposed IVs are: (1) using two proposed IVs increases efficiency when $\boldsymbol{\lambda} = \mathbf{0}$ compared to one proposed IV when the components of $\boldsymbol{\gamma}$ in (A2') are both $\neq 0$; (2) if ψ_2^*/ψ_1^* is specified to be a single number, then using the two proposed IVs identifies β as long as the line ℓ on which $\boldsymbol{\lambda}$ is identified to lie is not parallel to the line $\ell'(\psi_1^*/\psi_2^*)$ (see end of Section 6.1). However, when ψ_2^*/ψ_1^* is specified to be in a range rather than a single number, then β is not identified, and a potential drawback of two proposed IVs is that uncertainty about both λ_1 and λ_2 needs to be incorporated into the SI.

Table 2 presents SIs for using only one IV. Note that when forming a SI for one proposed IV as compared to two proposed IVs, the $\max(R^2_{unobs}/R^2_{obs})$ should stay the same (since the same unobserved covariate $a - E(a|\mathbf{x})$ can be used for the one proposed IV as for two proposed IVs) but $R^2_{z_1-E(z_1|\mathbf{x})}$ (the R^2 from the regression of $a - E(a|\mathbf{x})$ on $z_1 - E(z_1|\mathbf{x})$) should be less than or equal to $R^2_{\mathbf{z}-E(\mathbf{z}|\mathbf{x})}$ (the R^2 from

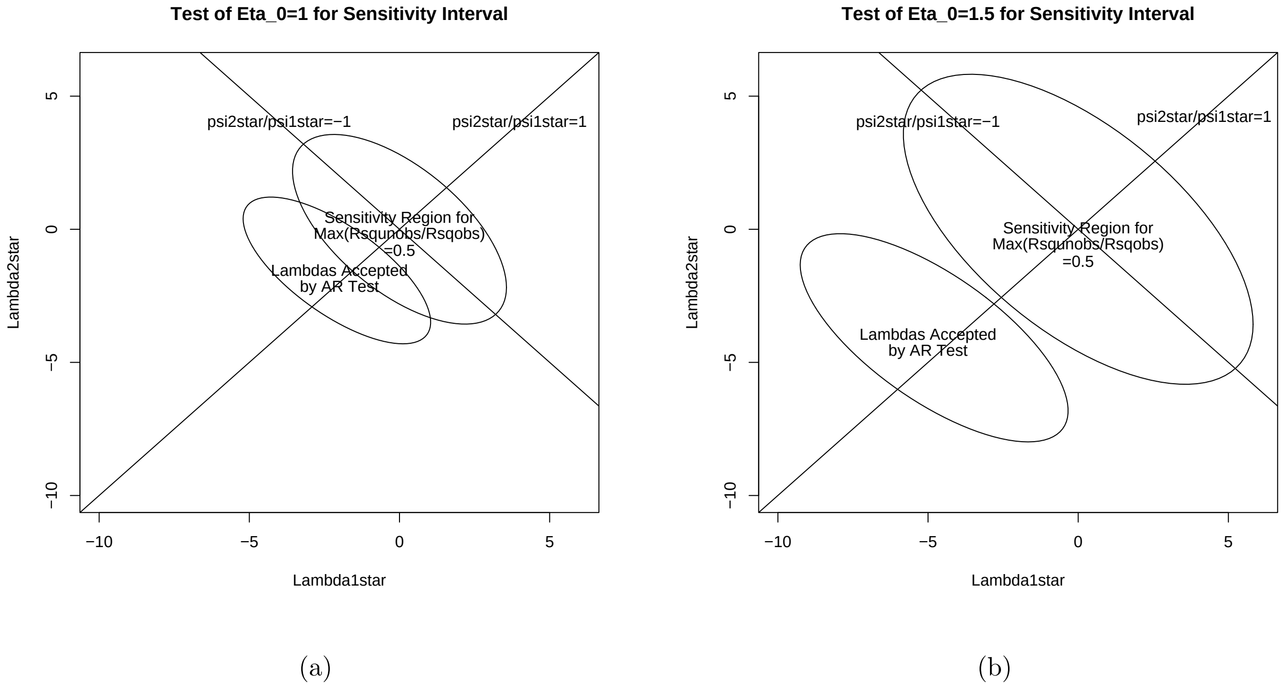

Figure 2: The figures depict the test of $H_0 : \eta = \eta_0$ for $\max(R^2_{unobs}/R^2_{obs})^+ = \max(R^2_{unobs}/R^2_{obs})^- = 0.25$, $R^2_{\mathbf{z}-E(\mathbf{z}|\mathbf{x})} = 0.5$ and various ranges of ψ_2^*/ψ_1^*. Figure (a) shows the test for $\eta_0 = 1$ and (b) shows $\eta_0 = 1.5$. The $\boldsymbol{\lambda}$'s in the plots are standardized in the following way: the x axis is $\lambda_1^* = \lambda_1 \times \hat{SD}(z_1 - \hat{E}(z_1|\mathbf{x}))$ and the y axis is $\lambda_2^* = \lambda_2 \times \hat{SD}(z_2 - \hat{E}(z_2|\mathbf{x}))$.

the regression of $a - E(a|\mathbf{x})$ on $z_1 - E(z_1|\mathbf{x})$ and $z_2 - E(z_2|\mathbf{x})$). Comparing Table 1 to Table 2 shows that if $R^2_{\mathbf{z}-E(\mathbf{z}|\mathbf{x})} = .5$ and if $R^2_{z_1-E(z_1|\mathbf{x})} = .25$ or $.4$, then for the range $.5 \leq \psi_2^*/\psi_1^* \leq 2$, the SI based on cultivated area per capita alone is shorter than the SI based on both proposed IVs; the same conclusion holds when worth of assets is the only proposed IV. However, for the range $-2 \leq \psi_2^*/\psi_1^* \leq -0.5$, using both proposed IVs produces shorter SIs. These results illustrate that even though using multiple proposed IVs allows for their joint validity to be tested via the ORT, using multiple proposed IVs may or may not reduce sensitivity to bias compared to using one proposed IV.

7. Conclusions and Discussion

$\max R^2_{z-E(z\|\mathbf{x})}$	$\max(R^2_{unobs}/R^2_{obs})^+ = \max(R^2_{unobs}/R^2_{obs})^-$				
	0.1	0.25	0.5	0.75	1
Cultivated area per capita as only IV					
0.25	$(0.40, 1.01)$	$(0.33, 1.18)$	$(0.22, 1.50)$	$(0.11, 2.11)$	$(-0.07, 4.71)$
0.4	$(0.37, 1.08)$	$(0.27, 1.35)$	$(0.08, 2.33)$	$(-\infty, \infty)$	$(-\infty, \infty)$
0.5	$(0.35, 1.12)$	$(0.22, 1.50)$	$(-0.07, 4.71)$	$(-\infty, \infty)$	$(-\infty, \infty)$
Worth of assets as only IV					
0.25	$(0.40, 0.99)$	$(0.32, 1.13)$	$(0.19, 1.39)$	$(0.05, 1.84)$	$(-0.21, 3.38)$
0.4	$(0.36, 1.04)$	$(0.24, 1.27)$	$(0.01, 1.99)$	$(-\infty, \infty)$	$(-\infty, \infty)$
0.5	$(0.34, 1.08)$	$(0.19, 1.39)$	$(-0.21, 3.38)$	$(-\infty, \infty)$	$(-\infty, \infty)$

Table 2: SIs for the income elasticity of food demand (η) for using only one IV for various settings of sensitivity parameters $\max\left(R^2_{\text{unobs}}/R^2_{\text{obs}}\right)^+$, $\max\left(R^2_{\text{unobs}}/R^2_{\text{obs}}\right)^-$ and $\max R^2_{z-E(z|\mathbf{x})}$.

We have developed a method of sensitivity analysis method for IV regression with overidentifying restrictions that enables a subject matter expert to combine his or her knowledge about the potential invalidities of the proposed IVs with the information provided by the data. The sensitivity analysis is more informative than the usual practice of reporting a CI that assumes the proposed IVs are valid along with the p-value from the ORT. For example, in the food demand study considered, the ORT does not reject both proposed IVs being valid, but the sensitivity analysis shows that moderate violations of the validity of the proposed IVs that are not rejected as implausible by the data would alter inferences substantively.

We now discuss two topics related to our paper for which future research would be useful. We have formed our SI by using the AR test (5) to form a joint confidence region for $(\beta, \boldsymbol{\lambda})$ given $\boldsymbol{\lambda} \in \mathbf{A}(\beta)$ and then projected the joint confidence region to form a CI for β. Investigation of other methods of forming joint confidence regions for $(\beta, \boldsymbol{\lambda})$ would be useful. For example, Kleibergen (2005) proposes a class of tests for $H_0 : \beta = \beta_0, \boldsymbol{\lambda} = \boldsymbol{\lambda}_0$ based on decomposing the AR test statistic into a part that tests $\beta = \beta_0$ and another part that tests $\boldsymbol{\lambda} = \boldsymbol{\lambda}_0 | \beta = \beta_0$. Another approach is to test $\boldsymbol{\lambda} = \boldsymbol{\lambda}_0$ using an analogue of the ORT and to test $\beta = \beta_0 | \boldsymbol{\lambda} = \boldsymbol{\lambda}_0$ using a Wald test based on the TSLS estimate, and then combine these tests using Bonferroni.

We have developed a frequentist approach to sensitivity analysis for IV regression. The SIs we construct provide "worst case" inferences about β over the sensitivity region for $\boldsymbol{\lambda}$. A less conservative approach is to put a distribution on $\boldsymbol{\lambda}$ over the sensitivity region as in a Bayesian approach.

References

Anderson, T.W. and Rubin, H. (1949), "Estimation of the Parameters of Coefficients of a Single Equation in a Simultaneous System and Their Asymptotic Expansions," *Econometrica*, 41, 683-714.

Angrist, J. and Imbens, G. (1995), "Estimation of Average Causal Effects," *Journal of the American Statistical Association*, 90, 431-442.

Angrist, J.D., Imbens, G.W. and Rubin, D.R. (1996), "Identification of Causal Effects Using Instrumental Variables," *Journal of the American Statistical Association*, 91, 444-472.

Bardhan, P. and Udry, C. (1999), *Development Microeconomics*, Oxford Press.

Boos, D.D. (1992), "On Generalized Score Tests," *The American Statistician*, 46, 327-333.

Bouis, H.E. and Haddad, L.J. (1990), *Effects of Agricultural Commercialization on Land Tenure, Household Resource Allocation, and Nutrition in the Philippines*, Research Report 79, IFPRI, Boulder, CO: Lynne Rienner Publishers.

Card, D. (1995), "Using Geographic Variation in College Proximity to Estimate the Return to Schooling," In *Aspects of Labour Market Behaviour*, University of Toronto Press.

Davidson, R. and MacKinnon, J. (1993), *Estimation and Inference in Econometrics*, New York: Oxford University Press.

Dufour, J.-M. (1997), "Some impossibility theorems in econometrics with applications to structural and dynamic models," *Econometrica*, 6, 1365-1387.

Fuller, W.A. (1987), *Measurement Error Models*, New York: Wiley.

Hansen, L.P. (1982), "Large sample properties of generalized method of moments estimators," *Econometrica*, 50, 1029-1054.

Hausman, J.A. (1983), "Specification and Estimation of Simultaneous Equation Models," In *Handbook of Econometrics, Volume 1*, North-Holland.

Hogan, J.W., Lancaster, T., Roy, J. and Alderson, D. (2003), "Evaluating Sensitivity to Exclusion Restrictions in Semiparametric Instrumental Variables Estimators," presented as Invited Paper at Joint Statistical Meetings August 2003.

Holland, P.W. (1988), "Causal Inference, Path Analysis and Recursive Structural Equation Models," *Sociological Methodology*, 18, 449-484.

Huber, P.J. (1967), "The behavior of maximum likelihood estimates under nonstandard conditions," *Proc. Fifth Berkeley Symposium in Mathematical Statistics*, 1, 221-233.

Imbens, G.W. (2003), "Sensitivity to Exogeneity Assumptions in Program Evaluation," *The American Economic Review*, 93, 126-132.

Kadane, J.B. and Anderson, T.W. (1977), "A Comment on the Test of Overidentifying Restrictions," *Econometrica*, 45, 1027-1032.

Kane, T. and Rouse, C. (1993), "Labor market returns to two-and- four-year colleges: is a credit a credit and do degrees matter?", NBER Working Paper #4268.

Kleibergen, F. (2005), "Generalizing Weak Instrument Robust IV Statistics Towards Multiple Parameters, Restricted Covariance Matrices and Identification Statistics," *Journal of Econometrics*, in press.

Manski, C. (1995), *Identification Problems in the Social Sciences*, Harvard Press.

Newey, W.K. (1985), "Generalized Method of Moments Specification Testing," *Journal of Econometrics*, 29, 229-256.

Neyman, J. (1923), "On the Application of Probability Theory to Agricultural Experiments," trans. D. Dabrowska, *Statistical Science*, 1990, 5, 463-480.

Nicholson, W. (1995), *Microeconomic Theory: Basic Principles and Extensions*, sixth edition., New York: Harcourt, Brace & Company.

Rothenberg, T.J. (1971), "Identification in Parametric Models," *Econometrica*, 39, 577-591.

Rosenbaum, P.R. (1986), "Dropping Out of High School in the United States: an Observational Study," *Journal of Educational and Behavioral Statistics*, 11, 207-224.

Rosenbaum, P.R. (1999), "Using Combined Quantile Averages in Matched Observational Studies," *Applied Statistics*, 48, 63-78.

Rosenbaum, P.R. (2002), *Observational Studies*, second edition, New York: Springer-Verlag.

Rubin, D.B. (1974), "Estimating Causal Effects of Treatments in Randomized and Non-randomized Studies," *Journal of Educational Psychology*, 66, 688-701.

Rubin, D.B. (1986), "Statistics and Causal Inference: Comment: Which Ifs Have Causal Answers," *Journal of the American Statistical Association*, 81, 961-962.

Sargan, J.D. (1958), "The Estimation of Economic Relationships Using Instrumental Variables," *Econometrica*, 26, 393-415.

Tan, Z. (2005), "Regression and Weighting Methods for Causal Inference Using Instrumental Variables," *Journal of the American Statistical Association*, in press.

White, H. (1982), "Instrumental Variables Regression with Independent Observations," *Econometrica*, 50, 483-500.

White, H. (1984), *Asymptotic Theory for Econometricians*, Academic Press.

Wooldridge, J.M. (1997), "On Two Stage Least Squares Estimation of the Average Treatment Effect in a Random Coefficient Model," *Economics Letters*, 56, 129-133.

Zheng, X. and Loh, W.-Y. (1995), "Consistent Variable Selection in Linear Models," *Journal of the American Statistical Association*, 90, 151-156.