Sensitivity Analysis for Instrumental Variables Regression With Overidentifying Restrictions

Dylan S. SMALL

Instrumental variables regression (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 instrumental variables (IVs); a valid IV is a variable that is associated with the treatment, 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 these IVs can be tested through 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 method using a study of food demand among rural households in the Philippines.

KEY WORDS: Causal inference; Econometrics; Structural equations models.

1. INTRODUCTION

A fundamental problem in making inferences about the causal effect of a treatment based on observational data is the potential presence of unmeasured confounding variables. Instrumental variables regression (IV regression) is a method for overcoming this problem. The method requires a valid instrumental variable (IV), which is a variable that is associated with the treatment, is independent of the unmeasured 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 unmeasured confounding variables and then uses this variation to estimate the causal effect of the treatment. An example is the study of Card (1995) on the causal effect of education earnings that 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 children in 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 unmeasured confounding variable (motivation), meaning that the proposed IV would be invalid. The validity of a proposed IV cannot be tested consistently (see Sec. 3.1). When critical assumptions for an analysis are 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 address uncertainty about partially unverifiable assumptions has long been recognized in causal inference (see Rosenbaum 2002, chap. 4, for a review). Sensitivity analyses for IV regression have been considered by Manski (1995), Angrist, Imbens, and Rubin (1996), and Rosenbaum (1999).

In this article 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 the work of Kane and Rouse (1993), who studied the causal effect of education on earnings as Card (1995) did, but, in addition to distance from nearest 4-year college, used tuition at state colleges of the state in which a person grew up as an IV. For IV regression with overidentifying restrictions, all of 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, then the use of multiple IVs can increase efficiency (see Sec. 2). Second, the use of multiple proposed IVs enables testing of the joint validity of all of the proposed IVs (to a limited extent) through the overidentifying restrictions test (ORT) (see Sec. 3).

For IV regression with overidentifying restrictions, it is common practice to report inferences that assume that all of 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 remains essential for addressing uncertainty about the validity of the proposed IVs when multiple proposed IVs are considered. We develop in this article 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.

The article 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, and 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, and in Section 6 we illustrate our sensitivity analysis method using a study of food demand. We provide conclusions and discussion in Section 7.

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 could in principle alter. For example, in the study of Kane and Rouse (1993) $y$ 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, $y_i = y_i(w)$. An additive, linear constant-effect causal model for the potential outcomes (as in Holland 1988) is

$$ y_i(w) = y_i(0) + \beta w^* .$$  \hspace{1cm} (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 $\beta$ is ordinary least squares (OLS) regression of $y^{obs}$ on $w^{obs}$. The OLS coefficient on $w$, $\beta_{OLS}$, has probability limit $\beta + \text{cov}(w^{obs}, y_i^{0}) / \text{var}(w^{obs})$. If $w^{obs}$ were randomly assigned, then $\text{cov}(w^{obs}, y_i^{0})$ would equal 0, and $\beta_{OLS}$ would be consistent. But in an observational study, often $\text{cov}(w^{obs}, y_i^{0}) \neq 0$ and $\beta_{OLS}$ is inconsistent. One strategy to address this problem is to attempt to collect data on all confounding variables $q$ and then regress $y^{obs}$ on $w^{obs}$ and $q$. If $w^{obs}$ is conditionally independent of $y^{0}$ given $q$ [i.e., the mechanism of assigning treatments $w^{obs}$ is ignorable (Rubin 1978)] and the regression function is specified correctly, then this strategy produces a consistent estimate of $\beta$. However, it often difficult to figure out and/or collect all confounding variables $q$.

IV regression is another strategy for estimating $\beta$ in (1). A vector of valid IVs $z_i$ is a vector of variables that satisfies the following assumptions (see Angrist et al. 1996; Tan 2006):

(A1) $z_i$ is associated with the observed treatment $w_i^{obs}$.

(A2) $z_i$ is independent of $\{y_i(w), w^* \in Y\}$ where $Y$ is the set of possible values of $w$. Note that under model (1), this is equivalent to $z_i$ being independent of $y_i^{0}$.

The basic idea of IV regression is to use $z$ to extract variation in $w^{obs}$ that is independent of the confounding variables and to use only this part of the variation in $w^{obs}$ to estimate the causal relationship between $w$ and $y$. Assumption (A1) is needed to be able to use $z$ to extract variation in $w^{obs}$ and (A2) is needed for the variation in $w^{obs}$ extracted from variation in $z$ 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) $z$ 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 $z$ are randomly assigned (or the mechanism of assigning the encouragement variables is ignorable) and encouragement in and of itself has no direct effect on the outcome, then $z$ is a vector of valid IVs (Holland 1988; Angrist et al. 1996). The model of Kane and Rouse (1993) can be viewed as an encouragement design in which it is assumed that the mechanism of assigning distance from nearest 4-year college and tuition at state colleges where a person grew up is ignorable, and that low levels of these variables encourage a person to attend college but have no direct effect on earnings.

For assumption (A2) to be plausible, it is often necessary to condition on a vector of covariates $x_i$ (Tan 2006). For example, Kane and Rouse (1993) conditioned 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 (and/or tuition at state colleges). Conditioning on $x_i$ also can increase the efficiency of the IV regression estimator. We call the variables in $x_i$ the included exogenous variables. We consider a linear model for $E(y_i^{(0)}(x, z_i), z_i) = \beta_{OLS} x_i + \lambda z_i + u_i, \quad E(u_i|x_i, z_i) = 0$. \hspace{1cm} (2)

We assume that $(w_i, x_i, z_i, u_i)$ are iid random vectors. This model has been considered by Holland (1988), among others. The model for the observed data is

$$ y_i = \beta w_i + \delta^T x_i + \lambda z_i + u_i, \quad E(u_i|x_i, z_i) = 0, \quad (w_i, x_i, z_i, u_i), i = 1, \ldots, N, \quad \text{are iid random vectors}. \hspace{1cm} (3)$$

For this model, we say that $z$ is a vector of valid IVs if it satisfies assumptions (A1') and (A2'), which are analogous to (A1) and (A2):

(A1') The partial $R^2$ for $z$ in the population regression of $w^{obs}$ on $x$ and $z$ is greater than 0.

(A2') The vector $\lambda$ in (3) equals 0.

(A1') says that $z$ is associated with $w^{obs}$ conditional on $x$. (A2') says that $z$ is uncorrelated with the composite $u$ of the confounding variables besides $x$. A sufficient condition for (A2') to hold is that $z$ is uncorrelated with any of the confounding variables besides those in $x$ and that $z$ itself has no direct effect on the outcome (Angrist et al. 1996).

We now consider inferences for $\beta$ under the assumption that $z$ is a vector of valid IVs. Under regularity conditions, $\beta$ 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 $\lambda = 0$ in (3), then $E^*(y|x, z) = \beta E^*(w|x, z) + \delta^T x$, where $E^*$ denotes linear projection. The TSLS estimator is obtained by regressing $w$ on $(x, z)$ using OLS to obtain $\hat{E}^*(w|x, z)$, then regressing $y$ on $\hat{E}^*(w|x, z)$ and $x$.
using OLS to estimate \( \beta \) and \( \delta \). The asymptotic distribution of \( \hat{\beta}_{TSLS} \) is
\[
\frac{\sqrt{N} (\hat{\beta}_{TSLS} - \beta)}{\hat{\sigma}_{\text{a}}^2} \xrightarrow{D} N(0,1) \quad (4)
\]
where \( \hat{w}_i \) is the predicted \( w_i \) from the OLS regression of \( w_{obs} \) on \( x \) and \( z \), \( \hat{E}^*(\hat{w} | x) \) is the estimated linear projection of \( \hat{w} \) onto \( x \), and \( \hat{\sigma}_{\text{a}}^2 = \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{\beta}_{TSLS}w_i - \hat{\beta}_{TSLS}x_i)^2 \). An asymptotically valid confidence interval (CI) can be obtained by inverting the Wald test based on (4). Note that the validity of these inferences for \( \beta \) does not require that \( E(y_i^0 | x_i, z_i) \) be a linear function of \((x,z)\). Let \( E^*(y_i^0 | x_i, z_i) = \delta^T z_i + \lambda^T x_i \). In addition, let \( u_i^0 = y_i^0 - E(y_i^0 | x_i, z_i) \) and assume that the \( u_i^0 \)'s are iid. Then under conditions (A1') and (A2'), the TSLS estimator is consistent, and the Wald CI is asymptotically valid (Davidson and MacKinnon 1993).

Using additional valid IVs can improve the efficiency of the TSLS estimator. For fixed \( x_1, \ldots, x_N \), the denominator of the left side of (4) is proportional to the partial \( R^2 \) for \( z \) from the OLS regression of \( w_{obs} \) on \( x \) and \( z \). Thus, using additional valid IVs that increase this partial \( R^2 \) increases the efficiency of \( \hat{\beta}_{TSLS} \).

In the rest of this article, we assume that the vector of proposed IVs \( z \) is not subject to the impact on inferences of (A2') being violated for \( z \). (A2') being violated means that \( \lambda \neq 0 \), which means that \( z \) is associated with a confounding variable(s) not in \( x \) and/or \( z \) has a direct effect on the outcome. If \( \lambda \neq 0 \), then \( (\hat{\beta}_{TSLS}, \hat{\delta}_{TSLS}) \) is not typically a consistent estimator of \((\beta, \delta)\). Let \( Y_N = (y_1, \ldots, y_N)^T \), \( W_N = (w_1, \ldots, w_N)^T \), \( X_N = (x_1, \ldots, x_N)^T \), \( Z_N = (z_1, \ldots, z_N)^T \), and \( D_N = [X_N, Z_N] \). The asymptotic bias of \( \hat{\beta}_{TSLS}, \hat{\delta}_{TSLS} \) is
\[
\lim_{N \to \infty} \left\{ (D_N(D_N^T D_N)^{-1} D_N^T W_N, X_N) \right\} \right. \times 
\left[ D_N(D_N^T D_N)^{-1} D_N^T W_N, X_N \right]^{-1} \times 
\left[ D_N(D_N^T D_N)^{-1} D_N^T W_N, X_N \right]^{-1} \times Z_N \lambda \quad (5)
\]
Before further considering the effects of \( \lambda \) not being equal to 0 on inferences, we consider a test of whether \( \lambda \) equals 0.

3. OVERIDENTIFYING RESTRICTIONS TEST AND ITS POWER

An ORT is a test of the assumption that \( \lambda = 0 \) in (3), which can be applied when \( \text{dim}(z) > 1 \). ORTs derive their name from the fact that when \( \text{dim}(z) > 1 \) and \( \lambda = 0 \), the parameter \( \beta \) in (3) is “overidentified,” meaning that any nonempty subset of the \( z \) variables could be used as the IVs to obtain a consistent estimate of \( \beta \). If \( \lambda = 0 \), then the different estimators of \( \beta \) that involve using different subsets of \( z \) as proposed IVs will all converge to the true \( \beta \). But if \( \lambda \neq 0 \), then the different estimators may converge to different limits. An ORT looks at the degree of agreement between the different estimates of \( \beta \) that involve using different subsets of \( z \) as proposed IVs. There are several versions of ORTs, but Newey (1985) showed that among ORTs based on a finite set of moment conditions, all such tests with maximal degrees of freedom \( \text{dim}(z) - 1 \) are asymptotically equivalent. We focus on one of these tests with maximal degrees of freedom, Sargan’s (1958) test, and refer to this as the ORT.

The test statistic for Sargan’s ORT is
\[
J_N = [(D_N)^T (Y_N - W_N \hat{\beta}_{TSLS} - X_N \hat{\delta}_{TSLS})] \times \left[ D_N(D_N^T D_N)^{-1} \right] \times [(D_N)^T (Y_N - W_N \hat{\beta}_{TSLS} - X_N \hat{\delta}_{TSLS})]. \quad (6)
\]
Under \( H_0: \lambda = 0 \), the distribution of \( J_N \) converges to \( \chi^2(\text{dim}(z) - 1) \) as \( N \to \infty \). A motivation for the test statistic \( J_N \) is that under \( H_0: \lambda = 0 \), we have \( \frac{1}{N} D_N^T (Y_N - W_N \hat{\beta}_{TSLS} - X_N \hat{\delta}_{TSLS}) \to 0 \) (because when \( \lambda = 0 \), \( E[D_N^T (Y_N - W_N \beta - X_N \delta)] = 0 \) and \( \hat{\beta}_{TSLS} \to \beta \), \( \hat{\delta}_{TSLS} \to \delta \)), but for many alternatives to \( H_0, \frac{1}{N} D_N^T (Y_N - W_N \hat{\beta}_{TSLS} - X_N \hat{\delta}_{TSLS}) \) converges in probability to \( c \neq 0 \), in which case \( J_N \) will tend to be higher under the alternative. In particular, if some but not all of the proposed IVs are valid, then (\( \hat{\beta}_{TSLS}, \hat{\delta}_{TSLS} \)) will typically converge in probability to a limit other than \((\beta, \delta)\) and, consequently, \( \frac{1}{N} D_N^T (Y_N - W_N \hat{\beta}_{TSLS} - X_N \hat{\delta}_{TSLS}) \) converges in probability to \( c \neq 0 \). Although the ORT is consistent (i.e., its power converges to 1) for many alternatives to \( \lambda = 0 \), such as those just discussed, the ORT is not consistent for some alternatives.

3.1 Inconsistency of the Overidentifying Restrictions Test

We consider model (3) plus a regression model for \( w_{obs} \),
\[
y_i = \beta w_i + \delta^T z_i + \lambda^T x_i + u_i, \quad (7)
\]
and also suppose that \((u_i, v_i)\) are iid bivariate normal. This assumption is standard in simultaneous equation system models, of which (7) is a special case (Hausman 1983). For this model, the ORT is not consistent for all alternatives (Kadane and Anderson 1977). We now characterize the alternatives for which the ORT is not consistent. Following Rothenberg (1971), we call the parameter vector \( \theta = (\beta, \lambda, \gamma, \delta, \upsilon, \sigma^2_u, \sigma^2_v, \sigma_{uv}) \) a structure. The structure \( \theta \) specifies a distribution for \((y_i, w_i)\) conditional on \((x_i, z_i)\). Two structures \( \theta_1 \) and \( \theta_2 \) are said to be observationally equivalent if they specify the same distribution for \((y_i, w_i)\) conditional on \((x_i, z_i)\). The ORT test statistic \( J_N \) is a function of the observable data \( (y_i, w_i, x_i, z_i: i = 1, \ldots, N) \), and thus \( J_N \) has the same distribution for observationally equivalent structures conditional on \((x_i, z_i: i = 1, \ldots, N) \) (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 (7), we obtain the “reduced form,”
\[
y_i = \beta u^T x_i + \delta^T z_i + \beta y^T z_i + \lambda^T x_i + \beta u_i + u_i
\]
\[
\equiv \rho^T x_i + \kappa^T z_i + e_i,
\]
\[
w_i = u^T x_i + y^T z_i + v_i,
\]
where \((e_i, v_i)\) is a bivariate normal distribution with mean 0. The distribution for \((y_i, w_i | x_i, z_i)\) depends only on the reduced-form parameter \( \pi = (\rho, \kappa, \upsilon, \sigma_u^2, \sigma_v^2, \sigma_{uv}) \). Also, any two structures that have reduced-form parameters \( \pi_1 \neq \pi_2 \) have different distributions for \((y_i, w_i) | x_i, z_i) \). Therefore, two structures
are observationally equivalent if and only if they have the same reduced-form parameter $\pi$.

The reduced-form parameter $\pi$ is a function of $h$ of the structural parameter $\theta = (\beta \nu + \delta, \beta \gamma + \lambda, v, y, \beta^2 \gamma^2 + 2\beta \sigma_{uv} + \sigma_u^2, \beta \sigma_v^2 + \sigma_v^2)$. For a reduced-form parameter $\pi^* = (\beta^*, \gamma^*, v^*, (\sigma^*_u)^2, (\sigma^*_v)^2, \lambda^*)$, the set of structures that have reduced-form parameter $h(\theta) = \pi^*$ is the following set, parameterized by $c$:

$$OE(\pi^*) = \{(\beta, \gamma, \delta, v, \sigma_u^2, \sigma_v^2, \lambda): \beta = c, \gamma = \gamma^*, \delta = \delta^* - c\gamma^*, v = v^*, \lambda = \lambda^* \}$$

The set of $\lambda$’s in $OE(\pi)$ for the true $\pi$ is

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

which is a line in the parameter space $\mathbb{R}^{\dim(x)}$ of $\lambda$. Thus we can identify $\lambda$ “up to a line.”

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

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

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

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

## 4. Sensitivity Analysis

For IV regression analysis under model (3), the critical assumption used to make inferences is that $\lambda = 0$. A sensitivity analysis considers what happens to inferences about $\beta$ if $\lambda$ is known to belong to a sensitivity zone, $A$, of values that a subject matter expert considers a priori plausible, rather than assuming that $\lambda = 0$. The output of our sensitivity analysis is a sensitivity interval (SI), an interval that has a high probability of containing $\beta$ as long as $\lambda \in A$. A SI is an analog 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 $\lambda \in A$ rather than $\lambda = 0$, then the evidence provided by the IV regression analysis that assumes that $\lambda = 0$ is strengthened. On the other hand, if inferences are significantly altered by assuming that $\lambda \in A$ rather than $\lambda = 0$, then the evidence provided by the IV regression analysis that assumes that $\lambda = 0$ is called into question. We allow the sensitivity zone for $\lambda$ to vary with the true $\beta$ (see Sec. 4.2 for reasons for this). We specify the sensitivity zone $A$ by $A = [A(\beta_0), \beta_0 \in \mathbb{R}]$, where $A(\beta_0)$ is the set of values for $\lambda$ that the subject matter expert considers would be a priori plausible before looking at the data if the true $\beta$ were to equal $\beta_0$. In Section 4.1 we present a method for constructing a SI for $\beta$ given a sensitivity zone $A$, and in Section 4.2 we provide a model for choosing the sensitivity zone $A$.

### 4.1 Constructing a Sensitivity Interval

A SI, like a CI, is a region of plausible values for $\beta$ given our assumptions and the data. A $(1-\alpha)$ SI given a sensitivity zone $A$ is a random interval that will contain the true $\beta$ with probability at least $(1-\alpha)$ under the assumption that $\lambda \in A(\beta)$. Our approach to forming a SI is to form a joint confidence region for $(\ell, \lambda)$ under the assumption that $\lambda \in A(\ell)$ and then to project this confidence region to form a CI for $\beta$. We form a joint $(1-\alpha)$ confidence region for $(\ell, \lambda)$ [under the assumption that $\lambda \in A(\ell)$] by inverting a level $\alpha$ test of $H_0: \beta = \beta_0$, $\lambda = \lambda_0$ for each $(\beta_0, \lambda_0)$ such that $\lambda_0 \in A(\beta_0)$. To test $H_0: \beta = \beta_0$, $\lambda = \lambda_0$, we use the fact that when $\lambda = \lambda_0$ and the transformed response variable $y_{k_0} = y - \lambda_0^* z$ is used in place of $y, z$ is a vector of valid IVs. A consistent estimator of $(\beta, \delta^T)^T$ under the assumption $\lambda = \lambda_0$ is obtained by applying TSLS estimation with $y_{k_0}$ as the response,

$$\left(\hat{\beta}_{TSLS, k_0}, \hat{\delta}_{TSLS, k_0}\right)^T = \left([D_N(D_N^T W_N)^{-1} D_N^T W_N X_N]^{-1} \times [D_N(D_N^T W_N)^{-1} D_N^T W_N X_N]^{-1} \times [D_N(D_N^T W_N)^{-1} D_N^T W_N X_N]^T [Y_N - Z_N \lambda]\right).$$

We can test $H_0: \lambda = \lambda_0$ using the following analog of (6):

$$J_N(\lambda_0) = \left([D_N][Y_N - Z_N \lambda_0 - W_N \hat{\beta}_{TSLS, k_0} - X_N \hat{\delta}_{TSLS, k_0}]^Tight) \times [\hat{\sigma}_{u, k_0}^{-2}]^{-1} \times \left[D_N[Y_N - Z_N \lambda_0 - W_N \hat{\beta}_{TSLS, k_0} - X_N \hat{\delta}_{TSLS, k_0}]ight]^{-1}$$

where $\hat{\sigma}_{u, k_0}^{-2} = \frac{1}{N} \sum_{i=1}^N (y_i - \lambda_0^* z_i - \hat{\beta}_{TSLS, k_0} w_i - \hat{\delta}_{TSLS, k_0} x_i)^2$. Under $H_0: \lambda = \lambda_0$, the statistic $J_N(\lambda_0)$ converges in distribution to $\chi^2(\dim(z) - 1)$ as $N \rightarrow \infty$. We can test $H_0: \beta = \beta_0|\lambda = \lambda_0$ using the test statistic

$$W(\beta_0, \lambda_0) = \frac{\hat{\beta}_{TSLS, k_0} - \beta_0}{\hat{\sigma}_{u, k_0} \sqrt{\sum_{i=1}^N (y_i - \hat{E}(\hat{w}|x_i))^2}}$$

under $H_0: \beta = \beta_0, \lambda = \lambda_0$, the statistic $W(\beta_0, \lambda_0)$ converges in distribution to $N(0, 1)$ as $N \rightarrow \infty$. For $\alpha_1 < \alpha$, let $T(\beta_0, \lambda_0) = 1$ if both (a) $J_N(\lambda_0) \leq \chi^2_{\alpha_1}(\dim(z) - 1)$ and (b) $W(\beta_0, \lambda_0))^2 \leq \chi^2_{\alpha_1}(\alpha_1 - \alpha_2)$ are satisfied, and $T(\beta_0, \lambda_0) = 0$ otherwise. The test that accepts $H_0: \beta = \beta_0, \lambda = \lambda_0$ if and only if $T(\beta_0, \lambda_0) = 1$ has asymptotic level $\alpha$ by the foregoing discussion and Bonferroni’s inequality. Let $C(\beta, \lambda)$ denote the region $\{(\beta_0, \lambda_0): T(\beta_0, \lambda_0) = 1\}$ and $\lambda_0 \in A(\beta_0)$; $C(\beta, \lambda)$ is a $(1-\alpha)$ confidence region for $(\ell, \lambda)$ given that $\lambda \in A(\beta)$. To place more
importance on β in constructing this confidence region, we generally make α1 ≪ α; in the empirical study in Section 6, we chose α1 = 0.05 for α = 0.05. Let \( S^*(A, Y_N, W_N, D_N) \) denote the projection of \( C(β, λ) \) into the β subspace of \( (β, λ) \), that is, \( S^*(A, Y_N, W_N, D_N) = \{ (β, λ) \in C(β, λ) \text{ for at least one } λ \in \mathbb{R}^{\dim(x)} \} \).

**Proposition 2.** \( S^*(A, Y_N, W_N, D_N) \) is a \((1-α)\) SL.

**Proof.** Suppose that the true value of λ belongs to \( A(λ) \) for the true value of β. Because \( C(β, λ) \) is a \((1-α)\) confidence region for \((β, λ)\) given that \( λ \in A(λ) \), we have \( P(β \in S^*(A, Y_N, W_N, D_N)) \geq 1-α \) and, consequently, \( P(β \in S^*(A, Y_N, W_N, D_N)) \geq P((β, λ) \in C(β, λ)) \geq 1-α \).

**4.2 Model for Choosing the Sensitivity Zone**

A crucial part of sensitivity analysis is choosing the sensitivity zone. This choice requires subject matter expertise, but methodology can help by providing an interpretable and manageable model for the subject matter expert to consider. Much of the insight from a sensitivity analysis often can 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 \( a - E(a|x) \) 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 Section 1, a might represent motivation. We assume that the following model holds in addition to (3): \( y_i = β w_i + f(x_i) + φ(a_i - E(a|x_i)) + b_i, \ E(b_i|x_i, z_i, a_i) = 0 \), where the function \( f(x_i) \) equals \( δ^T x_i + \sum_{j=1}^{\dim(x)} λ_j E(z_j|x_i) \). We assume that the \((E(z_j|x_i))\)’s are linear in some basis expansion for \( x_i \). Then \( f(x_i) = \xi^T g(x_i) \) for some vector \( \xi \) and basis functions \( g(x_i) \), and

\[
y_i = β w_i + \xi^T g(x_i) + φ(a_i - E(a|x_i)) + b_i.
\]

Under model (12), TSLS estimation using \( g(x_i) \) and \( a_i - E(a|x_i) \) as included exogenous variables and \( z \) as the vector of proposed IVs would provide consistent estimates of \( β, \xi, \) and \( φ \). For the numerical value of the parameter \( φ \) to be meaningful, the scale of the unobserved covariate \( a - E(a|x) \) must be restricted. We do this by assuming that \( \text{var}(a - E(a|x)) = 1 \).

Furthermore, we assume that a linear model holds for the relationship between \( a \) and \( z \): \( E(a|x, z) = σ^T x + ψ^T z \). Under models (3) and (12), \( λ \) equals \( ψ \) because (a) under (3), for all \( (x, z) \), we have \( E(y - β w|x, z) = δ^T x + λ^T z \) and (b) under (12), for all \( (x, z) \), we have \( E(y - β w|x, z) = ψ^T x + φ(a - E(a|x)) + E(z|x) = ψ^T x + φ(z - γ^T z) \).

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; Imbens 2003). Here \( a \) could represent a composite of several unobserved covariates rather than a single unobserved covariate. In particular, suppose that \( t_1, \ldots, t_m \) are unobserved covariates such that \( y_i = β w_i + f(x_i) + φ(t_1 - E(t_1|x_i)) + \cdots + φ(t_m - E(t_m|x_i)) + b_i, \ E(b_i|x_i, z_i, t_1, \ldots, t_m) = 0 \). Then we can set \( a_i = φ(t_1 + \cdots + φ(t_m) - E(t_m|x_i)) \).

We now discuss an approach to specifying a range of plausible values for \( λ = φ ψ \) given \( β = β_0 \) in terms of interpretable parameters. We consider \( φ \) and \( ψ \) separately. For specifying a plausible range of values for \( φ \), note that for any \( w^* \), we can rewrite (12) as \( y_i - β w_i + β w^* = β w^* + \xi^T g(x_i) + φ(a_i - E(a_i|x_i)) + b_i, \ E(b_i|x_i, z_i, a_i) = 0 \). The parameter \( φ \) measures the strength of the relationship between the unobserved covariate \( a - E(a|x) \) and the potential outcome \( y(w^*) = y - β(w^{obs} - w^*) \). Specifically,

\[
\frac{R^2_{unobs}}{R^2_{obs}} = \frac{φ^2}{\text{var}(y(β(w^{obs} - w^*)))} \left/ \frac{\text{var}(\xi^T g(x))}{\text{var}(y(β(w^{obs} - w^*)))} \right.
\]

is the proportion of variation in the potential outcomes \( y(w^*) \) explained by the regression on the unobserved covariate \( a - E(a|x) \) relative to the proportion of variation explained by the regression on the observed covariates \( x \). For a realistic \( w^* \), the proportion \( R^2_{unobs}/R^2_{obs} \) is an interpretable quantity. If \( R^2_{unobs}/R^2_{obs} = 0 \), then the proposed IVs \( z \) are valid; if \( 0 < R^2_{unobs}/R^2_{obs} < 1 \), then the proposed IVs \( z \) are invalid, but the unobserved covariate \( a - E(a|x) \) is a less strong predictor of the potential outcomes than the currently included vector of exogenous variables \( x \); and if \( R^2_{unobs}/R^2_{obs} > 1 \), then the proposed IVs are invalid, and \( a - E(a|x) \) is a stronger predictor of the potential outcomes than \( x \). A subject matter expert might want to provide a different range of plausible values for \( R^2_{unobs}/R^2_{obs} \) for \( φ \) that is positive for than when \( φ \) is negative. The relationship between \( R^2_{unobs}/R^2_{obs} \) and \( φ \) depends on \( \var(\xi^T g(x)) \). For \( β = β_0 \), we can consistently estimate \( φ \) by regressing \( y - β_0 w \) on \( g(x) \); denote this estimate by \( φ_0 \). For given values of sign(\( φ \)) and \( R^2_{unobs}/R^2_{obs} \), we estimate \( φ \) to be

\[
φ = \text{sign}(φ) \sqrt{R^2_{unobs}/R^2_{obs} \text{var}(\xi^T g(x))}. \tag{13}
\]

Note that our approach of working with \( R^2_{unobs}/R^2_{obs} \) specifies different ranges of \( φ \) for different values of \( β \). We view \( R^2_{unobs}/R^2_{obs} \) as a quantity for which experts can think about plausible values without considering 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 with 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 \( φ_{min} ≤ φ ≤ φ_{max} \), keeping in mind that \( φ \) is the change in the mean of the potential outcome \( y(w^{obs}) \) associated with one standard deviation change in the unobserved covariate \( a - E(a|x) \).

For specifying a range of plausible values for \( ψ \), we specify plausible values for the magnitude and direction of \( ψ \) using the norm \( ||ψ|| = \sqrt{\text{var}(ψ^T (z - E(z|x)))} \). For the magnitude of \( ψ \), note that under our sensitivity analysis model, \( E(a - E(a|x)|[z - E(z|x)]) = ψ^T (z - E(z|x)) \). Thus, using that \( \var(a - E(a|x)) = 1 \), we have that \( ||ψ||^2 = R^2 \) for the regression of \( a - E(a|x) \) on \( z - E(z|x) \); we denote this \( R^2 \) by \( R^2_{z-E(z|x)} \). Note that \( λ = φ ψ = \sqrt{R^2_{unobs}/R^2_{obs} \text{sign}(φ)} \times \sqrt{\text{var}(\xi_0^T g(x))} \sqrt{R^2_{z-E(z|x)} ||ψ||} \), so that equal values of
the product \( (R^2_{unobs}/R^2_{obs})R^2_{z−E(z|x)} \) produce equal values of \( \lambda \) if \( \phi(\alpha) \) and \( \Psi \| \Psi \| \) are kept fixed. For specifying the direction of \( \Psi \), we transform \( \Psi \) to \( \Psi^* \), where \( \Psi^* \) is the vector of coefficients for the regression of \( a - E(a|x) \) on the standardized values of \( z - E(z|x) \): \( \Psi^* = (\psi_1SD(z_1 - E(z_1|x)), \ldots, \psi_{dim(z)}SD(z_{dim(z)} - E(z_{dim(z)}|x))) \). Then we specify a range of plausible values for \( \{\psi_2/\psi_1^*, \ldots, \psi_{dim(z)}/\psi_1^*, \text{sign}(\psi_1^*)\} \). The sensitivity parameters \( \{\psi_2/\psi_1^*, \ldots, \psi_{dim(z)}/\psi_1^*, \text{sign}(\psi_1^*)\} \) specify the relative magnitude and sign of the coefficients in the regression of \( a - E(a|x) \) on the standardized values of \( z - E(z|x) \). For example, if \( \psi_2/\psi_1^* > 0 \), then \( z_1 - E(z_1|x) \) and \( z_2 - E(z_2|x) \) are either both positively or both negatively associated with the unobserved covariate \( a - E(a|x) \); if \( \psi_2/\psi_1^* < 0 \), then \( z_1 - E(z_1|x) \) and \( z_2 - E(z_2|x) \) are associated with \( a - E(a|x) \) in opposite directions. Also, if \( \psi_2/\psi_1^* > 1 \), then \( z_2 - E(z_2|x) \) has a higher absolute correlation with the unobserved covariate \( a - E(a|x) \) than \( z_1 - E(z_1|x) \), whereas if \( \psi_2/\psi_1^* < 1 \), then \( z_1 - E(z_1|x) \) has the higher absolute correlation. We assume that the range of plausible \( \{R^2_{z−E(z|x)}, \psi_2/\psi_1^*, \ldots, \psi_{dim(z)}/\psi_1^*, \text{sign}(\psi_1^*)\} \) does not depend on the true \( \beta \), because these sensitivity parameters reflect the association between \( a \) and \( z \).

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

\[
\hat{R}^2_{z−E(z|x)} = \text{var}(\psi^T(z - E(z|x)))
\]

\[
\hat{\psi}_2 = \psi_2SD(z_1 - E(z_1|x))
\]

\[
\hat{\psi}_1 = \psi_1SD(z_2 - E(z_2|x))
\]

\[
\ldots
\]

\[
\hat{\psi}_{dim(z)} = \psi_{dim(z)}SD(z_{dim(z)} - E(z_{dim(z)}|x))
\]

The steps of our sensitivity analysis procedure can be summarized as follows:

1. We specify the set of values of \( s = \{\phi(\alpha), R^2_{unobs}/R^2_{obs}, R^2_{z−E(z|x)}, \psi_2/\psi_1^*, \ldots, \psi_{dim(z)}/\psi_1^*, \text{sign}(\psi_1^*)\} \) that we consider plausible.

2. For a given \( \beta_0 \), we test \( H_0: \beta = \beta_0 \) as follows:
   a. For each \( s \) that we consider plausible from step 1, we estimate \( \phi \) and \( \Psi \) when \( \beta = \beta_0 \) using (13) and (14).
   b. We accept \( H_0: \beta = \beta_0 \) if \( \lambda_0: T(\beta_0, \lambda_0) = 1, \lambda_0 \in A(\beta_0) \) is not empty where \( T(\beta_0, \lambda_0) \) is the test of \( H_0: \beta = \beta_0, \lambda = \lambda_0 \) from Section 4.1.
   c. We test \( H_0: \beta = \beta_0 \) for each possible \( \beta_0 \) using step 2, and our SI is the set of \( \beta_0 \)'s that we accept.

We now illustrate how the SI can be computed for two proposed IVs. If one considers positive and negative associations of the unobserved covariate \( a - E(a|x) \) with the potential outcomes to be equally plausible, then a reasonable approach to specifying the sensitivity zone is to treat \( \phi = 0 \) and \( \phi < 0 \) symmetrically and to treat \( R^2_{unobs}/R^2_{obs}, R^2_{z−E(z|x)} \), and \( \psi_2/\psi_1^* \) as independent parameters that have intervals of plausible values, that is, the set of plausible values for \( \{\phi(\alpha), R^2_{unobs}/R^2_{obs}, R^2_{z−E(z|x)}, \psi_2/\psi_1^*, \text{sign}(\psi_1^*)\} \) is \( \{\phi(\alpha) = \pm 1\} \times \{0 \leq R^2_{unobs}/R^2_{obs} \leq \text{max}(R^2_{unobs}/R^2_{obs})\} \times \{0 \leq R^2_{z−E(z|x)} \leq \text{max}(R^2_{z−E(z|x)})\} \times \{d_1 \leq \psi_2/\psi_1^* \leq d_2\} \times \{\text{sign}(\psi_1^*) = \pm 1\} \). The sensitivity zone is characterized by the sensitivity parameters \( \{\phi(\alpha), R^2_{unobs}/R^2_{obs}, R^2_{z−E(z|x)}, \psi_2/\psi_1^*, \text{sign}(\psi_1^*)\} \). Note that \( \psi_2/\psi_1^* \) does not depend on the true \( \beta \), because these sensitivity parameters reflect the association between \( a \) and \( z \).

5. EXTENSIONS TO PANEL DATA AND HETEROGENEOUS TREATMENT EFFECTS

In model (3), we assume that \( (u_i, x_i, z_i, u_i) \) are iid random vectors. In this section we consider the following extension of model (3):

\[ y_{it} = \beta w_{it} + \delta^T x_{it} + \lambda^T z_{it} + u_{it}, \]

\[ i = 1, \ldots, N, \quad t = 1, \ldots, T, \]

\[ [W_i = (u_{i1}, \ldots, u_{iT})^T, X_i = (x_{i1}, \ldots, x_{iT})^T, Z_i = (z_{i1}, \ldots, z_{iT})^T, U_i = (u_{i1}, \ldots, u_{iT})^T, \]

\[ i = 1, \ldots, N, \quad \text{are independent but not necessarily iid random matrices} \]

\[ E(u_{it}|x_i, z_i) = 0. \]

For model (15), the TSLS estimator of \( \beta \) is consistent, but the asymptotic distribution is not (4) (White 1982).

One motivation for model (15) is when we have panel data with \( N \) units and \( T \) time periods and the additive linear constant effects models continues to hold, that is, \( y_{it}^{(w^*)} = \beta w^{(w^*)} + \delta^T x_{it} + \lambda^T z_{it} + u_{it} \). Model (15) allows for \( u_{i1}, \ldots, u_{iT} \) to be correlated. Model (15) also accommodates stratified cross-sectional or panel survey data and heteroscedasticity.

A second motivation for considering model (15) is to allow for heterogeneous treatment effects. Suppose that our model for potential outcomes is

\[ y_{i}^{(w^*)} = \beta_1 w_{i} + \delta^T x_{i} + \lambda^T z_{i} + u_{i}, \quad E(u_{i}|x_{i}, z_{i}) = 0. \] (16)

Unit i’s treatment effect is \( \beta_1 \). Let \( \hat{\beta} = \hat{E}(\beta_1) \) be the average treatment effect over the population. We can express the observed data from model (16) as

\[ y_{i} = \beta_1 w_{i} + \delta^T x_{i} + \lambda^T z_{i} + (\beta_1 - \beta) w_{i} + u_{i}, \]

\[ E(u_{i}|x_{i}, z_{i}) = 0. \] (17)
If we make the further assumption that

$$(\beta_b - \beta) \text{ is independent of } w_t | x_t, z_t,$$  

then (17) is equivalent to (15) with $T = 1$. Assumption (18) says that units do not select their treatment levels ($w_t$) based on the gains that they would experience from treatment ($\beta_b$) (Wooldridge 1997). If assumption (18) does not hold, then the TSLS estimator might not converge to $\beta$. Angrist and Imbens (1995) have discussed properties of TSLS estimates and the ORT when (18) does not hold.

For model (15), valid inferences when $\lambda = \lambda_0$ can be obtained using the generalized method of moments (GMM) with the moment conditions $E[(x_i, z_i, y_i) (y_{it} - \lambda_i' z_i - \beta w_{it} - \beta^T x_i)] = 0, t = 1, \ldots, T$ (Hansen 1982; Hall 2005). GMM can be viewed as a generalization of TSLS (Hall 2005). For forming a SI for $\beta$ under model (15), we use the same procedure as outlined in Section 4.1, except that we replace the test (10) with the GMM overidentifying restrictions test of Hansen (1982) and use the Wald test based on the GMM estimate of $\beta | \lambda = \lambda_0$ in place of the Wald test (11) based on TSLS. To choose the sensitivity zone under model (15), we use the method outlined in Section 4.2.

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. Here $y_{it}$ is the $i$th household’s food expenditure at time $t$, $w_{it}$ is the $i$th household’s log income at time $t$, and $x_{it}$ 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, where we assume that the income $w_{it}$ can be altered by a yearly lump sum payment (or charge) that households expect will continue permanently and that $y_{it}^{(w)}$ is measured shortly after such an alteration of income. The parameter $\beta$ represents the causal effect on a household’s short-run food expenditures of a one-unit increase in a household’s log income. Rather than focus directly on $\beta$, we focus on a 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 $1\%$ increase in income for households currently spending at the mean food expenditure level, and we denote it by $\eta$. The mean food expenditure of households is 31.14 pesos per capita per week, so that $\eta = 100\beta (\log 1.01) / 31.14 = .033\beta$.

Bouis and Haddad (1990) were concerned that regression of $y$ on $w^{obs}$ and $x$ would not provide an unbiased estimate of $\beta$ because of unmeasured 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 $w^{obs}$) are associated with their preferences (which are reflected in $y^{(w)}$) according to microeconomic theory (Bardhan and Udry 1999, chap. 2). To consistently estimate $\beta$, 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 that model (3) holds. Assumption (A1) for these proposed IVs appears to hold; the partial $R^2$ for $z$ from the regression of $w^{obs}$ on $x$ and $z$ is .27, and the test of the null hypothesis that the partial $R^2$ is 0 has $p$ value < .0001.

Using the proposed IVs $z_1$ and $z_2$, the TSLS estimate of the income elasticity of food demand $\eta$ is .70 with a 95% CI (assuming $\lambda = \lambda_0$ of (.60, .81). This CI and the SIs in the next section were computed using the method outlined in 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 unmeasured preferences might have influenced a household’s past choices on land acquisition and asset accumulation, which are reflected in a household’s present cultivated area per capita and worth of assets. The $p$ value for the ORT [Hansen’s (1982) version] is .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 for certain alternatives, 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 for the food demand study (using the model of Sec. 4.2), we consider a negative versus a positive association of the unobserved covariate $a - E(a|x)$ with the potential outcomes $y^{(w)}$ to be equally plausible. Thus we choose to use the method of specifying the sensitivity zone described at the end of Section 4.2. We estimate $E(z_1|x)$ and $E(z_2|x)$ 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 zones. To give the reader a sense of the economic meaning of the values of $\eta$ that are in the SIs, the following are estimates of the income elasticities of demand for various goods compiled by Nicholson (1995): medical services, .22; beer, .38; cigarettes, .50; electricity, .61; wine, .97; transatlantic air travel, 1.40; and automobiles, 3.00.

When examining Table 1, we first consider $\psi_1^2/\psi_1^2 = 1$, which corresponds to $z_1 - E(z_1|x)$ and $z_2 - E(z_2|x)$ being equally correlated with the unobserved covariate $a - E(a|x)$. For $\psi_2^2/\psi_1^2 = 1$ and $\max(R^2_{unobs}/R^2_{obs}) = .1$, which means that the unobserved covariate $a - E(a|x)$ has at most a modest association with the potential outcome $y^{(w)}$ relative to the association between $x$ and $y^{(w)}$, the SI for $R^2_{unobs}/R^2_{obs} = .5$ is (.52, .92), and the SI for $\max R^2_{E(x)} = 1$ is (.48, .98). The former SI is almost twice as wide as the CI that assumes that $\lambda = 0$, and the latter SI is more than twice as wide. For an unobserved covariate $a - E(a|x)$ with a potentially stronger but still moderate maximum association with $y^{(w)}$ relative to the association between $x$ and $y^{(w)}$ of $\max(R^2_{unobs}/R^2_{obs}) = .25$, the
Table 1. SIs for the income elasticity of food demand ($\eta$) for various settings of the sensitivity parameters
\[
\max(R^2_{z-E(z|x)}), \frac{\psi^*_2}{\psi^*_1}, \text{ and } \max(R^2_{unobs}/R^2_{obs})
\]

| $\max R^2_{z-E(z|x)}$ | $\frac{\psi^*_2}{\psi^*_1}$ | $.1$ | $.25$ | $.5$ | $.75$ | $1$ |
|------------------------|-----------------|------|------|------|------|------|
| $.5$                   | $1.000$          | $(.53, .91)$ | $(.48, .98)$ | $(.42, 1.07)$ | $(.36, 1.17)$ | $(.30, 1.28)$ |
| $.2$                   | $(.52, .92)$ | $(.47, 1.00)$ | $(.40, 1.13)$ | $(.33, 1.29)$ | $(.25, 1.48)$ |
| $.1$                   | $(.52, .92)$ | $(.47, 1.01)$ | $(.40, 1.15)$ | $(.33, 1.32)$ | $(.24, 1.55)$ |
| $.0001$                | $(.52, .91)$ | $(.48, .98)$ | $(.42, 1.09)$ | $(.36, 1.21)$ | $(.30, 1.33)$ |
| $- .5$                 | $(.55, .86)$ | $(.52, .90)$ | $(.47, .94)$ | $(.44, .98)$ | $(.41, 1.01)$ |
| $- 1$                  | $(.59, .81)$ | $(.58, .81)$ | $(.56, .81)$ | $(.55, .81)$ | $(.54, .81)$ |
| $- 2$                  | $(.56, .86)$ | $(.53, .90)$ | $(.49, .93)$ | $(.46, .96)$ | $(.43, .98)$ |
| $- 1,000$              | $(.53, .91)$ | $(.48, .97)$ | $(.42, 1.07)$ | $(.36, 1.16)$ | $(.29, 1.26)$ |
| $[- 5, 2]$             | $(.52, .92)$ | $(.47, 1.01)$ | $(.40, 1.15)$ | $(.32, 1.32)$ | $(.24, 1.55)$ |
| $[- 1,000, 1,000]$     | $(.52, .92)$ | $(.47, 1.01)$ | $(.40, 1.15)$ | $(.32, 1.32)$ | $(.24, 1.55)$ |
| $-1$                   | $(.50, .92)$ | $(.42, 1.07)$ | $(.30, 1.27)$ | $(.18, 1.54)$ | $(.06, 1.89)$ |
| $.5$                   | $(.48, .98)$ | $(.40, 1.14)$ | $(.25, 1.53)$ | $(.02, 2.27)$ | $(-.80, 3.37)$ |
| $.0001$                | $(.49, .96)$ | $(.42, 1.09)$ | $(.30, 1.34)$ | $(.17, 1.66)$ | $(-.03, 2.22)$ |
| $- .5$                 | $(.52, .89)$ | $(.47, .94)$ | $(.41, 1.01)$ | $(.35, 1.07)$ | $(.34, 1.12)$ |
| $- 1$                  | $(.58, .81)$ | $(.56, .81)$ | $(.54, .81)$ | $(.52, .81)$ | $(.50, .81)$ |
| $- 2$                  | $(.53, .89)$ | $(.49, .93)$ | $(.43, .98)$ | $(.38, 1.02)$ | $(.34, 1.05)$ |
| $- 1,000$              | $(.50, .95)$ | $(.42, 1.07)$ | $(.30, 1.26)$ | $(.17, 1.51)$ | $(.04, 1.86)$ |
| $[- 5, 2]$             | $(.48, .98)$ | $(.39, 1.15)$ | $(.24, 1.55)$ | $(-.07, 2.39)$ | $(-.21, 3.38)$ |
| $[- 1,000, 1,000]$     | $(.48, .98)$ | $(.39, 1.15)$ | $(.24, 1.55)$ | $(-.07, 2.39)$ | $(-.21, 3.38)$ |

SI for $\psi^*_2/\psi^*_1 = 1$, max $R^2_{z-E(z|x)} = .5$ is $(.47, 1.01)$ and the SI for $\psi^*_2/\psi^*_1 = 1$, max $R^2_{z-E(z|x)} = 1$ is $(.39, 1.15)$. The former is more than twice as much as that of $x$ and the latter more than three times as wide as the CI that assumes that $\lambda = \psi = 0$. For max($R^2_{unobs}/R^2_{obs}$) = 1 [i.e., the strength of association between $a = E(a|x)$ and $y^{(w)}$ can be as much as that of $x$ and $y^{(w)}$], the SI for max $R^2_{z-E(z|x)} = .5$ is $(.24, 1.55)$ and the SI for max $R^2_{z-E(z|x)} = 1$ is $(-2.11, 3.38)$. These intervals are wide in economic meaning; the lower ends are lower or close to lower than the income elasticity of the “necessity” of medical services, and the upper ends are higher than the income elasticity of the “luxury” of transatlantic air travel using the estimates from Nicholson (1995). For $\psi^*_2/\psi^*_1 = .5$ and $\psi^*_2/\psi^*_1 = 2$ [meaning that $z_1 - E(z_1|x)$ and $z_2 - E(z_2|x)$ are correlated with $a - E(a|x)$ 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 $\psi^*_2/\psi^*_1 = 1,000$, $0,0001$, and $- 1,000$ [meaning that one of $z_1 - E(z_1|x)$ or $z_2 - E(z_2|x)$ is much more strongly correlated with $a - E(a|x)$ than the other], the SIs are slightly shorter than for $\psi^*_2/\psi^*_1 = 1$. For $\psi^*_2/\psi^*_1 = -.5, -1$, and $- 2$ [meaning that $z_1 - E(z_1|x)$ and $z_2 - E(z_2|x)$ are correlated with $a - E(a|x)$ 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 $\psi^*_2/\psi^*_1 = 1,000$, $2, 1, .5, .0001$, or $- 1,000$. The last three rows of the top and bottom halves of Table 1 show SIs when the ranges for $\psi^*_2/\psi^*_1$ of $[.5, 2], [-2, - .5]$, and $[-1,000, 1,000]$ are considered.

We now discuss how $\psi^*_2/\psi^*_1$ affects the SI. Figure 1 shows the SI for fixed $\psi^*_2/\psi^*_1$ as $\psi^*_2/\psi^*_1$ varies from $-5$ to $5$ for max $R^2_{z-E(z|x)} = .5$ and max($R^2_{unobs}/R^2_{obs}$) = .5. The SI is widest for $\psi^*_2/\psi^*_1 = 1$ and narrowest for $\psi^*_2/\psi^*_1 = -1$. To illuminate how $\psi^*_2/\psi^*_1$ affects the SI, we consider what happens under model (7) as the sample size increases to infinity. Suppose that the line $\ell(\psi^*_2/\psi^*_1)$ on which $\psi^*_2/\psi^*_1$ specifies that $\lambda$ lies [i.e., $\ell(\psi^*_2/\psi^*_1) = (\hat{\lambda}_1, \hat{\lambda}_2)$: $\psi^*_2/\psi^*_1 = (\psi^*_2/\psi^*_1)^*\lambda_1 SD(z_1 - E(z_1|x))/SD(z_2 - E(z_2|x))$], equals the line $\ell$ on which $\lambda$ is...
identified to lie [see (8)]; this can happen only if $\ell$ crosses through $0$ and $\psi_2^*/\psi_1^* = \gamma_2 SD(z_2 - E(z_2[x]))/\gamma_1 SD(z_1 - E(z_1[x]))$. Then the SI converges to the set of $\beta^*$s for which $(\kappa - \beta \gamma) \in A(\beta)$. Suppose instead that $\ell'((\psi_2^*/\psi_1^*))$ does not equal $\ell$ and is not parallel to $\ell$. Let $\lambda^*$ be the point at which $\ell'((\psi_2^*/\psi_1^*))$ intersects $\ell$, and let $\lambda^*$ be the $\beta$ that satisfies $\lambda^* = (\kappa - \beta \gamma)$. If $(\kappa - \beta \gamma) \in A(\beta^*)$, then the SI for $\psi_2^*/\psi_1^*$ converges to $\beta^*$. If $(\kappa - \beta \gamma) \notin A(\beta^*)$, then the SI converges to the empty set. Moreover, if $\ell'((\psi_2^*/\psi_1^*))$ is parallel but not equal to $\ell$, then the SI converges to the empty set.

6.2 Comparison of SIs for One Proposed IV versus Two Proposed IVs

When there is only one proposed IV $z_1$, an SI still can be formed using a similar method to that outlined in Section 4. We specify a sensitivity zone $A(\beta_0)$ of plausible values of $\lambda$ given $\beta = \beta_0$, and reject $H_0: \beta = \beta_0$ if only if a level $\alpha$ test of $H_0: \beta = \beta_0|\lambda = \lambda_0$ based on (11) is rejected for all $\lambda \in A(\beta_0)$. One way of specifying a sensitivity zone is to specify the following range of plausible values for \( (\text{sign}(\phi), R_{\text{unobs}}^2/R_{\text{obs}}^2, R_{\text{z}=E(z|x)}^2) : [\text{sign}(\phi) = \pm 1] \times \{0 \leq R_{\text{unobs}}^2/R_{\text{obs}}^2 \leq \max(R_{\text{unobs}}^2/R_{\text{obs}}^2) \} \times \{0 \leq R_{\text{z}=E(z|x)}^2 \leq \max(R_{\text{z}=E(z|x)}) \} \).

The SI for $\beta$ with two or more proposed IVs uses both the restrictions on $\lambda$ from the specification of the sensitivity zone (a priori restrictions) and the information from the ORT to rule out certain $\lambda$ in the sensitivity zone as implausible (restrictions based on the data). The SI for $\beta$ with one proposed IV uses only the a priori restrictions on $\lambda$ from the specification of the sensitivity zone, because the ORT has no power for one proposed IV. Does using additional proposed IVs reduce the length of the SI compared with using just one proposed IV? This question is important for designing observational studies using IVs. Comparing using two proposed IVs to one proposed IV reveals two potential advantages of two proposed IVs: (a) Using two proposed IVs increases efficiency when $\lambda = 0$ compared with one proposed IV when the additional proposed IV helps predict $w_{\text{obs}}$; and (b) if the set of plausible $\psi_2^*/\psi_1^*$ is specified to be a single number, then two proposed IVs identify $\beta$ as long as the line $\ell$ on which $\lambda$ is identified to lie is not parallel or equal to the line $\ell'((\psi_2^*/\psi_1^*))$ (see the end of Sec. 6.1). When the set of plausible $\psi_2^*/\psi_1^*$ is specified to be an interval rather than a single number, $\beta$ is not point identified but is partially identified. A potential drawback of two proposed IVs is that uncertainty about both $\lambda_1$ and $\lambda_2$ 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 compared with two proposed IVs, the max($R_{\text{unobs}}^2/R_{\text{obs}}^2$) should stay the same [because the same unobserved covariate, $a = E(a|x)$, can be used for the one proposed IV as for two proposed IVs], but $R_{z_1=E(z_1|x)}^2$ (the $R^2$ from the regression of $a - E(a|x)$ on $z_1 - E(z_1[x])$) should be less than or equal to $R_{z_1=E(z_1|x)}^2$ [the $R^2$ from the regression of $a - E(a|x)$ on $z_1 - E(z_1[x])$ and $z_2 - E(z_2[x])$]. Comparing Tables 1 and 2 shows that if $R_{z_1=E(z_1|x)}^2 = .5$ and if $R_{z_1=E(z_1|x)}^2 = .25$ or $4$, then for the range $.25 \leq \psi_2^*/\psi_1^* \leq 2$, the SI based on worth of assets alone is at least as short as the SI based on both proposed IVs; for $R_{z_1=E(z_1|x)}^2 = .5$, the SI based on both proposed IVs is shorter. For the range $-2 \leq \psi_2^*/\psi_1^* \leq -.5$, the SI based on both proposed IVs is shorter for all of $R_{z_1=E(z_1|x)}^2 = .25, .4$, or $.5$. These results illustrate that even though using multiple proposed IVs allows for testing of their joint validity through the ORT, using multiple proposed IVs may or may not reduce the sensitivity to bias compared with using one proposed IV.

7. CONCLUSIONS AND DISCUSSION

We have developed a method of sensitivity analysis 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 that the proposed IVs are valid along with the $p$ value from the ORT. For example, in the food demand study considered herein, 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 have assumed throughout that (A1)’ holds, that is, the partial $R^2$ for $z$ from the population regression of $w_{\text{obs}}$ on $x$ and $z$ is greater than 0. When this partial $R^2$ is greater than 0 but is very small relative to the sample size, the proposed IVs are said to be “weak.” For weak IVs, the asymptotically justified inferences from the test statistics (10) and (11) can be inaccurate. For a setting of weak IVs, more accurate inferences can be obtained by using the test of Anderson and Rubin (1949) of

<table>
<thead>
<tr>
<th>max($R_{\text{unobs}}^2/R_{\text{obs}}^2$)</th>
<th>.1</th>
<th>.25</th>
<th>.5</th>
<th>.75</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cultivated area per capita as only IV</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.25</td>
<td>(.43, .88)</td>
<td>(.40, .94)</td>
<td>(.32, 1.00)</td>
<td>(.26, 1.10)</td>
<td>(.20, 1.19)</td>
</tr>
<tr>
<td>.4</td>
<td>(.41, .91)</td>
<td>(.34, .99)</td>
<td>(.27, 1.15)</td>
<td>(.16, 1.30)</td>
<td>(0, 1.49)</td>
</tr>
<tr>
<td>.5</td>
<td>(.40, .92)</td>
<td>(.32, 1.01)</td>
<td>(.20, 1.24)</td>
<td>(.03, 1.45)</td>
<td>(~.34, 1.77)</td>
</tr>
<tr>
<td>Worth of assets as only IV</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.25</td>
<td>(.56, .92)</td>
<td>(.52, .98)</td>
<td>(.46, 1.07)</td>
<td>(.44, 1.16)</td>
<td>(.38, 1.24)</td>
</tr>
<tr>
<td>.4</td>
<td>(.55, .95)</td>
<td>(.49, 1.02)</td>
<td>(.43, 1.17)</td>
<td>(.36, 1.32)</td>
<td>(.28, 1.49)</td>
</tr>
<tr>
<td>.5</td>
<td>(.55, .96)</td>
<td>(.48, 1.09)</td>
<td>(.40, 1.25)</td>
<td>(.32, 1.45)</td>
<td>(.15, 1.72)</td>
</tr>
</tbody>
</table>
$H_0: \beta = \beta_0, \lambda = \lambda_0$ or other tests described by Stock, Wright, and Yogo (2002). However, for the food demand study, the IVs are not weak so that using inferences based on (10) and (11) is reasonable. Stock et al. considered two IVs to be weak if the $F$-statistic for testing that the coefficients on $z$ in the population regression of $w^{obs}$ on $x$ and $z$ are both 0 is $<11.59$ (see Table 1 of Stock et al. 2002); this $F$ statistic for the food demand study is 306.3, much larger than 11.59.

[Received September 2005. Revised March 2007.]

REFERENCES


