The Influence Function of Semiparametric Estimators[*]

Hidehiko Ichimura   Whitney K. Newey

University of Tokyo   MIT

July 2015
Revised March 2021

Abstract

There are many economic parameters that depend on nonparametric first steps. Examples include games, dynamic discrete choice, average exact consumer surplus, and treatment effects. Often estimators of these parameters are asymptotically equivalent to a sample average of an object referred to as the influence function. The influence function is useful in local policy analysis, in evaluating local sensitivity of estimators, and constructing debiased machine learning estimators. We show that the influence function is a Gateaux derivative with respect to a smooth deviation evaluated at a point mass. This result generalizes the classic Von Mises (1947) and Hampel (1974) calculation to estimators that depend on smooth nonparametric first steps. We give explicit influence functions for first steps that satisfy exogenous or endogenous orthogonality conditions. We use these results to generalize the omitted variable bias formula for regression to policy analysis for and sensitivity to structural changes. We apply this analysis and find no sensitivity to endogeneity of average equivalent variation estimates in a gasoline demand application.

JEL Classification: C13, C14, C20, C26, C36
Keywords: Influence function, semiparametric estimation, NPIV.

1 Introduction

There are many estimators of economic parameters that depend on nonparametric first steps. Examples include games, dynamic discrete choice, average consumer surplus, and treatment effects. Often these estimators are asymptotically equivalent to a sample average. The object being averaged is referred to as the influence function.

The influence function has several important uses. It can be used for quantifying local policy effects. For example, Firpo, Fortin, and Lemieux (2009) use influence functions to quantify local policy effects of changes in explanatory variables on quantiles or other characteristics of a distribution. We give local policy effects of structural changes. The influence function can also be used to measure sensitivity of estimators to misspecification. Its use for qualitative sensitivity measures is where the influence function gets its name in the robust estimation literature, see Hampel (1974). The expected GMM influence function under a misspecified distribution is the GMM sensitivity measure given in Andrews, Gentkow, and Shapiro (2017). We quantify sensitivity for objects that depend on solutions to orthogonality conditions. We use this quantification to generalize the classic omitted variables bias formula for regression coefficients to many other objects. We apply these results to estimate sensitivity of equivalent variation bounds to endogeneity of gasoline demand.

Another important use of the influence function is construction of orthogonal moment functions where first step estimation has no first order effect on moments. Orthogonal moment functions reduce bias in GMM from model selection and regularization of the first step and enable machine learning for high dimensional first steps, as in Chernozhukov et al. (2018) and Chernozhukov et al. (2020). The influence function formulae given here are used in Chernozhukov et al. (2020) to derive orthogonal moment functions. The influence function can also be used to compare asymptotic efficiency of estimators and find efficient ones. Efficient estimation is important in many econometric settings where weak assumptions are made to make models empirically plausible. Knowing the form of the influence function also facilitates asymptotic theory by showing in advance the conclusion of an asymptotic expansion.

Newey (1994) showed that the influence function of an estimator could be obtained from the probability limit (plim) of the estimator. A functional equation was given that can be solved for the influence function without an asymptotic, large sample expansion. Hahn (1998) and Hirano, Imbens, and Ridder (2003) applied this approach to derive the influence function of important treatment effect estimators. A primary purpose of this paper is to give a simpler way of calculating the influence function and to illustrate its usefulness for applied researchers. We show that the influence function can be calculated from a derivative of the plim with

respect to a scalar mixture of the true distribution with another distribution. This calculation extends the classic Von Mises (1947), Hampel (1974), and Huber (1981) Gateaux derivative calculation to objects that exist only for continuous distributions. We also illustrate how this Gateaux derivative can be used to facilitate empirical research on local policy analysis, quantify sensitivity of estimators, and construct orthogonal moment functions.

The functional equation in Newey (1994) has been solved to obtain influence functions in many important settings. Newey (1994) did so for estimators that depend on a first step least squares projection or a probability density function (pdf). Bajari, Hong, Krainer, and Nekipelov (2010) and Bajari, Chernozhukov, Hong, and Nekipelov (2009) did so for game models and Hahn and Ridder (2013, 2016) did so for nonparametric generated regressors. We use the Gateaux derivative calculation to derive influence functions for first steps that solve orthogonality conditions, both exogenous and endogenous. These calculations provide explicit influence function formulae for a variety of estimators in addition to those already in the literature. The calculations also illustrate the simplicity and usefulness of the formulae here in making the influence function more widely available for empirical research involving local policy analysis, estimator sensitivity, or orthogonal moment functions.

We estimate sensitivity to endogeneity of bounds on average equivalent variation for gasoline demand. This application is motivated by the difficulty of simultaneously allowing for price endogeneity and general preferences in demand analysis. Hausman and Newey (2016) gave nonparametric estimators of bounds on average equivalent variation with general preferences that are independent of prices and income. For scalar heterogeneity and endogenous prices Blundell, Horowitz, and Parey (2017) estimate the gasoline demand function via nonparametric quantile instrumental variables, which is computationally difficult and only allows scalar heterogeneity. The bound sensitivity we give is much simpler and allows general heterogeneity. We find that for gasoline demand the average equivalent variation bounds are not very sensitive to endogeneity and that the sensitivity is not statistically significant.

A distinctive feature of our approach is that the influence function is obtained directly from the moment conditions defining the estimator without solving an integral equation or going through a probabilistic calculations in the form of asymptotic arguments. In this sense, our result allows us to study semiparametric estimators analogously to the estimators obtained based on the parametric maximum likelihood or the generalized method of moment conditions. Using asymptotic arguments Robinson (1988), Powell, Stock, and Stoker (1989), Goldstein and Messer (1992), Ichimura (1993), Klein and Spady (1993), and Chaudari, Doksum, and Samarov (1997) gave influence function formulae for important semiparametric estimators. Newey (1994) gave general explicit influence function formulae where a first step is an infinite dimensional

regressions or pdf. Ai and Chen (2007, 2012), Ichimura and Lee (2010), Ackerberg et al. (2014), Chen and Liao (2015), and Chen and Pouzo (2015) gave interesting and useful characterizations of influence functions for estimators with first steps that solve conditional moment restrictions or that are maximizers of an objective function. The results of this paper are complementary to this previous work in providing explicit formulae for influence functions for estimators that solve orthogonality conditions. Such explicit formula are useful for policy and sensitivity analysis and for construction of orthogonal moment functions.

The influence function of an estimator may be different than the efficient influence function for the parameter of a semiparametric model considered e.g. by Bickel et al. (1993). These do coincide in models where a parameter is exactly identified; see Chen and Santos (2015). One can think of the object derived here as the efficient influence function for the parameter that is defined as the plim of an estimator for a general, unrestricted distribution. This parameter is exactly identified in the model with the unrestricted distribution so the efficient influence function coincides with the influence function of the estimator. This is the approach taken by Newey (1994) to finding the influence function of an estimator. We simplify this approach in a way that makes it more applicable to empirical research.

Validity of the influence function calculation given here depends on distributional variation that is a smooth approximation to a distribution that puts all probability on a point, i.e. is a point mass. After the first version of this paper appeared on arXiv, Luedtke, Carone, and van der Laan (2015) and Carone, Luedtke, and van der Laan (2016) used such deviations in estimation. This construction is useful in that setting, but we emphasize that we have a different goal here; to calculate the influence function of any semiparametric estimator.

Muhkin (2019) used the influence function to derive local effects of changing one object of interest on another object of interest. Also, the local effects are integrated to obtain global effects. This work also shows the usefulness of the influence functional calculation given here.

Summarizing, the contributions of this paper are to i) give a simpler way of calculating the influence function; ii) derive explicit influence function formulae for functions satisfying exogenous and endogenous orthogonality conditions; iii) give local policy effects and sensitivity to structural changes and illustrate their use in empirical research; and iv) show absence of local sensitivity to endogeneity of equivalent variation in a gasoline demand application.

In Section 2 we give the Gateaux derivative formula for the influence function and describe several important uses of this formula. Section 3 gives the influence function for exogenous orthogonality conditions and uses that to derive local policy effects and sensitivity for structural change. It is shown that these formula generalize the classic omitted variables bias formula. Section 4 gives the influence function for endogenous orthogonality conditions. Section 5 dis-

cusses extensions and conclusions. Online Appendices give regularity conditions for validity of the influence function calculation, characterize the influence function for minimum distance estimators, extend the explicit influence function formulae to misspecified orthogonality conditions, and give regularity conditions for endogenous orthogonality conditions.

2 The Influence Function and Its Uses

The estimators and objects in this paper are allowed to depend on a first step nonparametric estimator. We refer to these estimators as semiparametric. We denote such an estimator by $\hat{\theta}$, which is a function of the data $W_1, ..., W_n$ where n is the number of observations. Throughout the paper we will assume that the data observations W_i are i.i.d. with some cumulative distribution function (CDF) F_0. We let θ_0 denote the probability limit of $\hat{\theta}$ when F_0 is the distribution of W_i.

In this paper we focus on asymptotically linear estimators that satisfy

$$\sqrt{n}(\hat{\theta} - \theta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi(W_i) + o_p(1), \ E[\psi(W)] = 0, \ E[\psi(W)^T \psi(W)] < \infty. \tag{2.1}$$

The asymptotic variance of $\hat{\theta}$ is then $E[\psi(W)\psi(W)^T]$. The function $\psi(w)$ is referred to as the influence function, following terminology of Hampel (1974). It gives the influence of a single observation in the leading term of the expansion in equation (2.1). It also quantifies the effect of a small change in the distribution of W on the probability limit of $\hat{\theta}$ as we further explain below. Very many root-n consistent semiparametric estimators are asymptotically linear under sufficient regularity conditions, including M-estimators, Z-estimators, estimators based on U-statistics, and many others; see Bickel, Klaasen, Ritov, and Wellner (1993) and Van der Vaart (1998).

Newey (1994) showed how the influence function of an estimator can be obtained without deriving the stochastic expansion in equation (2.1). Let F be any distribution that is unrestricted except for regularity conditions and $\theta(F)$ denote the probability limit of $\hat{\theta}$ when F is the CDF of W. Here $\theta(F)$ can be thought of as the probability limit of $\hat{\theta}$ under general misspecification where F is only required to satisfy some regularity conditions (like some random variables being continuously distributed and/or existence of certain moments) but is otherwise unrestricted. Also, let $\{F_\beta\}$ be any parametric family of distributions passing through F_0 with $F_\beta = F_0$ when $\beta = 0$ and satisfying certain regularity conditions with score (derivative of log-likelihood) $S_\beta(w)$ at $\beta = 0$. Then by Van der Vaart (1991) it follows that the influence function satisfies

$$\frac{\partial \theta(F_\beta)}{\partial \beta} = E[\psi(W) S_\beta(W)], \tag{2.2}$$

[4]

when the estimator $\hat{\theta}$ is locally regular in the sense discussed in Van der Vaart (1991). This is a functional equation from which $\psi(w)$ may be obtained by varying $\{F_\beta\}$ and the associated score $S_\beta(w)$. In several important settings the influence function has been obtained by solving this functional equation without the stochastic expansion in equation (2.1). Newey (1994) did this for first step regression and density estimation. Hahn (1998) obtained the influence function for the regression estimator of the average treatment effect and Hirano, Imbens, and Ridder (2003) for inverse propensity score weighted estimators. Hahn and Ridder (2013, 2016) did so for first step generated regressors and control functions and Bajari et al. (2009, 2010) for estimating game models.

A main purpose of this paper is to give a simpler, more direct way of calculating the influence than solving equation (2.2). Let H denote a CDF such that $\theta(F_\tau)$ exists for $F_\tau = (1-\tau)F_0 + \tau H$ where τ is a scalar with $0 \leq \tau < C$ for $0 < C < 1$. Equation (2.1) and regularity conditions discussed in Appendix A imply that

$$\frac{d\theta(F_\tau)}{d\tau} = \int \psi(w)H(dw), \ E[\psi(W)] = 0, \ E[\psi(W)^2] < \infty, \tag{2.3}$$

where throughout the paper $d/d\tau$ denotes a derivative from the right at $\tau = 0$. This equation suggests a direct way to calculate the influence function:

STEP I: Calculate $d\theta(F_\tau)/d\tau$ for any H such that the derivative exists;

STEP II: Evaluate the derivative formula at $H = \Delta_w$, where Δ_w is the CDF with $\Pr(W = w) = 1$, to obtain $\psi(w) = \int \psi(\tilde{w})\Delta_w(d\tilde{w})$ as a function of w.

Equation (2.3) does not justify Step II because the derivative need not exist when $H = \Delta_w$. In particular $\theta(F_\tau)$ may not be well defined when $\theta(F)$ depends on a pdf or conditional expectation because of the discrete component Δ_w of $F_\tau = (1-\tau)F_0 + \tau\Delta_w$. The nonexistence of a pdf of F_τ of $(1-\tau)F_0 + \tau\Delta_w$ at any $\tau > 0$ can make $\theta(F_\tau)$ undefined. Nevertheless Step II is justified as a limit as H approaches Δ_w, similar to Lebesgue (1904) differentiation in analysis; e.g. see Wheeden and Zygmund (1977). A precise justification for Step II is given in Appendix A.

The calculation in Steps I and II generalizes the classic Hampel (1974) formula, $\psi(w) = d\theta((1-\tau)F_0 + \tau\Delta_w)/d\tau$, to cases where existence of $\theta(F)$ requires some components of W be continuously distributed. Such cases are very important for semiparametric estimators where $\theta(F)$ can depend on limits of nonparametric estimators of densities, conditional expectations, or other objects whose existence requires W have continuously distributed components. Steps I and II provide a simpler and more direct way of obtaining $\psi(w)$ than solving the integral equation (2.2).

[5]

The influence function does not exist when $\theta(F)$ does not satisfy the Stein (1956) necessary conditions for existence of a root-n consistence estimator. In that case Steps I and II will fail. To illustrate, suppose W is continuously distributed with pdf $f_0(w)$ and $\theta(F) = f(\bar{w})$ is the pdf of W at some fixed $\bar{w}$. In that case

$$\frac{d\theta(F_\tau)}{d\tau} = h(\bar{w}) - f_0(\bar{w}). \tag{2.4}$$

Because $h(\bar{w})$ is the pdf of $H(w)$ at the point $\bar{w}$ it cannot be represented as the expectation over H of a function with finite second moment. In general Steps I and II will fail whenever equation (2.3) is not satisfied. As in equation (2.4), this failure will often be evident in the calculation of $d\theta(F_\tau)/d\tau$.

Equation (2.3) motivates the use of the influence function in empirical work. The Gateaux derivative $d\theta(F_\tau)/d\tau$ is the local effect of changing the distribution F on the object $\theta(F)$. If we broaden the interpretation of $\theta(F)$ to include economic objects of interest, such as a feature of the distribution of outcome variables, then $d\theta(F_\tau)/d\tau$ can be thought of as a local policy effect of changing the distribution of the data. Equation (2.3) then can be used to obtain the local policy effect from the influence function, as did Firpo, Fortin, and Lemeiux (2009) for the policy effect of changing the distribution of regressors. When $\theta(F)$ is the probability limit of an estimator $\hat{\theta}$ we can think of $d\theta(F_\tau)/d\tau$ as the local sensitivity of that estimator to changes in F, which gives local effects of misspecification. The GMM sensitivity analysis of Andrews, Gentzkow, and Shapiro (2017) has precisely the form of equation (2.3), as will be discussed in Section 2.2. In addition, when $\theta(F)$ is the true expectation of an identifying moment function evaluated at the limit of a first step estimator, the influence function can be used to create orthogonal moments that have zero Gateaux derivative with respect to the first step. As discussed in Chernozukov et al. (2018), this use of the influence function is helpful for debiased machine learning of objects of interest.

In the remainder of this Section we describe more fully these important uses of the influence function that are of direct interest to empirical researchers. Here we show how this paper can be applied to obtain novel policy effects of structural change, local sensitivity measures and Hausman tests, and orthogonal moment functions.

2.1 Local Policy Analysis of Structural Changes

In many settings $\theta(F)$ may be an economic quantity of interest. Changes in F can sometimes be thought of as changes in a policy. From equation (2.3) we see that $\int \psi(w)H(dw)$ is the derivative of $\theta(F)$ as F_τ changes away from F_0 in the direction $H - F_0$. If H is thought of as resulting from a change in policy, then $\int \psi(w)H(dw)$ will be the derivative of the economic

quantity of interest with respect to that policy change, i.e. a local policy effect.

Firpo, Fortin, and Lemeiux (2009) derive such effects where $\theta(F)$ is specified as some feature of the marginal distribution of an outcome variable Y and the change in policy is a change in the distribution of explanatory variables X. Because $\theta(F)$ depends only on the marginal distribution of Y, the influence function of $\theta(F)$ will be $\psi(y)$ that depends only on y. For example if $\theta(F)$ is the pth quantile of F, satisfying $F_Y(\theta(F)) = p$, then $\psi(y) = [1(y < \theta_0) - p]/f_{Y0}(\theta_0)$, where $f_{Y0}(y)$ is the true marginal pdf of Y. Because the distribution of X is different in H but nothing else is different than in F_0, the conditional distribution of Y given X will be the same for H as it is for F_0. Then by iterated expectations the local policy effect is

$$\frac{d\theta(F_\tau)}{d\tau} = \int \psi(y)H(dw) = E_H[\psi(Y)] = E_H[E[\psi(Y)|X]].$$

Firpo, Fortin, and Lemeiux (2009) analyze such policy effects for quantiles of Y, other objects $\theta(F)$ of interest, and for a variety of alternative policy shifts in the distribution of X as represented by H.

One can also specify the policy effect of a structural change where the conditional distribution of Y given X changes and the marginal distribution of X remains unchanged. The local policy effect of a structural change is

$$\frac{d\theta(F_\tau)}{d\tau} = E_H[\psi(Y)] = E[E_H[\psi(Y)|X]].$$

Here we see the that local effect of a structural change in the direction $H - F_0$ is captured by the conditional expectation $E_H[\psi(Y)|X]$ of the influence function $\psi(Y)$ for the distribution H.

Other local policy effects can be considered by specifying $\theta(F)$ to be something other than a feature of the distribution of a random variable Y. One example of such a $\theta(F)$ is a bound on average equivalent variation from Hausman and Newey (2016). The Gateaux derivative formula in equation (2.3) can be used to derive local policy effects of structural changes on this and many other objects. In Section 3 we do so for $\theta(F)$ that depends on conditional location.

Specification and estimation of global policy effects using quantile regressions was developed by Machado and Mata (2005), Albrecht, Björklund, Vroman (2003), and Melly (2005). Estimators of global effects based distribution regression were developed by Chernozhukov, Fernandez-Val, and Melly (2013). Local policy effects are useful for evaluating small policies. Also, Muhkin (2019) shows that global policy effects can be obtained from integrating local effects, making local effects of interest even for evaluation of global effects.

2.2   Local Sensitivity and Local Hausman Tests

Quantifying local sensitivity of an estimator to misspecification, or more generally to a change in distribution of the data, is another important use of the influence function. Equation (2.3)

gives the Gateaux derivative of the probability limit $\theta(F)$ in the direction $H - F_0$. If the distribution H allows for misspecification then $\int \psi(w)H(dw)$ measures sensitivity of $\theta(F)$ to local misspecification. More generally if H is a different distribution than that of the data then $\int \psi(w)H(dw)$ measures the sensitivity of $\theta(F)$ to a distribution shift. Qualitative and quantitative sensitivity measures can be constructed based on $\psi(w)$. A qualitative sensitivity characteristic is boundedness of $\psi(w)$, which guarantees that $d\theta(F_\tau)/d\tau$ is bounded over all possible H. This is the classic robustness characteristic of Hampel (1974) and Huber (1981) that is defined by boundedness of the influence function.

Quantitative measures of estimator sensitivity can also be based on $\psi(w)$. Conley, Hansen, and Rossi (2012) and Andrews, Gentzkow, and Shapiro (2017) give measures of sensitivity of IV and GMM estimators, respectively, to moment misspecification. The sensitivity measure for GMM is exactly $\int \psi(w)H(dw)$ for the GMM influence function. To explain, suppose that there is a vector function $g(w, \theta)$ of a data observation w and parameter vector θ satisfying a moment condition $E[g(W, \theta_0)] = 0$. A GMM estimator is obtained as $\hat{\theta} = \arg\min_\theta \hat{g}(\theta)' \hat{\Psi} \hat{g}(\theta)$ where $\hat{g}(\theta) = \sum_{i=1}^n g(W_i, \theta)/n$ are sample moments and $\hat{\Psi}$ is a positive semi-definite weighting matrix. It is well known that the influence function for GMM under correct specification (i.e. $E[g(W, \theta_0)] = 0$) is

$$\psi(w) = -(G'\Psi G)^{-1}G'\Psi g(w, \theta_0), \;\; G = \left.\frac{\partial}{\partial \theta}E[g(W, \theta)]\right|_{\theta=\theta_0}, \;\; \Psi = \text{plim}(\hat{\Psi}).$$

Therefore for GMM the local sensitivity will be

$$\frac{d\theta(F_\tau)}{d\tau} = \int \psi(w)H(dw) = -(G'\Psi G)^{-1}G'\Psi \int g(w, \theta_0)H(dw).$$

This is the local sensitivity formula given in Andrews, Gentzkow, and Shapiro (2017). When the dimension of $g(w, \theta)$ is bigger than that of θ this formula imposes correct specification of the moments, i.e. $E[g(W, \theta_0)] = 0$. Imbens (1997) gives the influence function for GMM allowing for misspecification and Muhkin (2019) describes its use for sensitivity analysis.

Equation (2.3) gives the local sensitivity of any estimator to a change of F in the direction $H - F_0$. In Section 3 we derive local sensitivity of a functional of conditional location and illustrate its use in estimating sensitivity of average equivalent variation bounds to endogeneity of gasoline prices.

Local sensitivity can be used to construct local Hausman specification tests for any object of interest with an influence function. A first order expansion gives

$$\theta(H) - \theta(F_0) = \theta(F_1) - \theta(F_0) \approx \left(\frac{d\theta(F_\tau)}{d\tau}(\tau - 0)\right)_{\tau=1} = \int \psi(w)H(dw). \tag{2.5}$$

Thus we see that $\int \psi(w)H(dw)$ is a first order approximation to the effect of changing the distribution F on the probability limit $\theta(F)$ of the estimator $\hat{\theta}$ corresponding to Hausman's

(1978) idea of checking sensitivity of an estimator of an object of interest to model assumptions. A estimator of $d\theta(F_\tau)/d\tau$ can be formed from an estimator of the influence function $\psi(w)$ and an alternative H by substituting the estimated influence function in equation (2.3) and integrating over H. Standard errors can be constructed using asymptotic theory or the bootstrap and an asymptotic t-statistic formed in the usual way. From equation (2.5) we see that such a t-statistic is a local Hausman test of the effect of misspecification in the direction H.. This approach can give local Hausman specification tests for any estimator with an influence function in any direction H. In Section 3 we illustrate such tests by testing for a significant effect of endogeneity of price on average equivalent variation for gasoline demand. It is beyond the scope of this paper to develop the general asymptotic theory of such tests. We discuss these tests here to illustrate the usefulness of the influence function in empirical work.

The covariance between the influence functions of two different estimators was suggested by Gentzkow and Shapiro (2015) and Andrews, Gentzkow, and Shapiro (2017) as a measure of sensitivity of one estimator with respect to another. Muhkin (2019) gives a geometric interpretation of this covariance as a directional derivative of one functional with respect another. As Muhkin (2019) shows, the covariance between two influence functions is the Gateaux derivative of $\theta(F)$ with respect to a departure from F_0 in a direction G that corresponds to a change in the other functional. In this way the influence functions for two different estimators are useful for constructing measures of sensitivity. For brevity we omit further specifics but note that this is an active and important research topic that is potentially useful for empirical work, where influence functions are key ingredients.

2.3 Orthogonal Moment Functions

Another important use of influence functions is in the construction of orthogonal moment functions for GMM with a nonparametric first step. Orthogonal moment functions are those where the expected moment functions have zero derivative with respect to the first step. GMM with orthogonal moment functions does not suffer from the large model selection and regularization biases of some estimators based on nonorthogonal moment functions. Avoiding such biases can be particularly important for machine learning first steps, as discussed in Chernozhukov et al. (2018) and shown in Chernozhukov et al. (2020).

To describe orthogonal moment functions consider a vector of functions $g(w, \gamma, \theta)$ where γ is a (possibly) nonparametric first step with true value γ_0, θ is the parameter vector of interest, and the moment condition $E[g(W, \gamma_0, \theta_0)] = 0$ is satisfied. This moment condition can be thought of as an identifying moment for θ_0, with γ_0 obtained from a first step. In general the first order effect of γ on $E[g(W, \gamma, \theta_0)]$ may be nonzero, leading to bias in a GMM estimator

[9]

based on sample moments $\hat{g}(\theta) = \sum_{i=1}^{n} g(W_i, \hat{\gamma}, \theta)/n$, where $\hat{\gamma}$ is a first step estimator of γ_0 that is plugged in. As shown in Chernozhukov et al. (2020), orthogonal moment functions can be constructed by adding to the identifying moments the influence function $\phi(w, \gamma_0, \alpha_0, \theta)$ of $E[g(W, \gamma(F), \theta)]$, where α are additional unknown functions on which ϕ may depend and $\gamma(F)$ is the probability limit of the first step estimator $\hat{\gamma}$ when F is the true distribution of W. This $\phi(w, \gamma, \alpha, \theta)$ can be calculated by Steps I and II applied to equation (2.3) for $E[g(W, \gamma(F), \theta)]$, i.e

$$\frac{dE[g(W, \gamma(F_\tau), \theta)]}{d\tau} = \int \phi(w, \gamma_0, \alpha_0, \theta) H(dw), \ E[\phi(W, \gamma_0, \alpha_0, \theta)] = 0. \tag{2.6}$$

Orthogonal moment functions can then be constructed as

$$\psi(w, \gamma, \alpha, \theta) = g(w, \gamma, \theta) + \phi(w, \gamma, \alpha, \theta).$$

The influence function $\phi(w, \gamma_0, \alpha_0, \theta)$ of $E[g(W, \gamma(F), \theta)]$ is an "adjustment term," analyzed in Newey (1994), that accounts for the presence of the first step $\hat{\gamma}$ in the moment functions. Adding this adjustment term to the original, identifying moment functions $g(w, \gamma, \theta)$ makes orthogonal moments. Calculating the adjustment term from Steps I and II is simpler than obtaining ϕ from the functional equation in Newey (1994). This simplicity facilitates the construction of orthogonal moment functions. We illustrate by calculating the adjustment term ϕ for solutions to exogenous orthogonality conditions in Section 3 and endogenous orthogonality conditions in Section 4. In Chernozhukov et al. (2020) the adjustment term for quantile orthogonality conditions is used to obtain debiased machine learning estimators for functionals solutions to quantile conditions.

Local policy analysis, sensitivity measures, and constructing orthogonal moment functions are three uses of the influence function that are of direct interest for empirical research. The results of this paper are useful in providing a simpler method of calculating the influence function that can then be used to construct local policy effects of structural changes, local sensitivity analysis and local Hausman tests for any estimator with an influence function, and orthogonal moment functions that can be used in debiased machine learning. In the next Section we illustrate by deriving the influence function for conditional location effects, constructing sensitivity measures for estimators of such effects, and applying them to average equivalent variation bounds.

Another important use of the influence function is in asymptotic efficiency comparisons, where it is convenient to bypass the stochastic expansion in equation (2.1). Knowing the influence function is also useful for showing that the asymptotic expansion in equation (2.1) is satisfied, because the influence function implies the precise form of the remainder. For brevity we omit further discussions of these uses of the influence function.

[10]

3 Exogenous Orthogonality Conditions

Many interesting economic and causal effects depend on a function that solves an orthogonality condition and depends only on instrumental variables. Such functions include high dimensional or additive specifications of orthogonality conditions for quantiles or expectiles. Effects of interest include bounds on average equivalent variation and average derivatives. In this Section we derive the influence function for such effects using Step I and Step II. We quantify local policy effects and local sensitivity for these effects. In addition we give an application to sensitivity of bounds on average equivalent variation to endogeneity in gasoline demand.

3.1 Functions Satisfying Exogenous Orthogonality Conditions

The unknown functions we consider depend on a vector of regressors X that may be infinite dimensional. We will denote a possible unknown function by γ with $\gamma(x)$ being its realization at $X = x$. We will impose the restriction that γ is in a set of functions Γ that is linear and closed in mean square, meaning that every γ in Γ has finite second moment and that if $\gamma_k \in \Gamma$ for each positive integer k and $E[\{\gamma_k(X) - \gamma(X)\}^2] \longrightarrow 0$ then $\gamma \in \Gamma$. We give examples of Γ in the second paragraph to follow.

We specify that the plim $\gamma_0 = \gamma(F_0)$ of a nonparametric estimator for the true distribution F_0 satisfies an orthogonality condition where a residual $\rho(W, \gamma)$ with finite second moment is orthogonal in the population to all $b \in \Gamma$. That is we specify that γ_0 satisfies

$$E[b(X)\rho(W, \gamma_0)] = 0 \text{ for all } b \in \Gamma. \tag{3.7}$$

This is like an instrumental variables orthogonality condition where the function γ depends only on the same variables X that the instrumental variables $b(X)$ depend on. This dependence of the functions γ and instrumental variables b on the same X is the "exogenous" referred to in the title of this Section. In the next Section we consider orthogonality conditions where γ may depend on different variables than X, corresponding to instrumental variables settings where there is endogeneity.

If Γ is specified to be all functions of Γ with finite second moment then equation (3.7) will be a conditional moment restriction $E[\rho(W, \gamma_0)|X] = 0$. We also allow Γ to be a smaller set. For example, a set of functions of interest for high dimensional estimation are those that are linear combinations of a sequence of functions $(b_1(X), b_2(X), ...)$ each having finite second moment. A corresponding Γ would be limits in mean square of linear combinations $\sum_{j=1}^{\infty} \beta_j b_j(X)$ where $\beta_j \neq 0$ for only a finite number of integers j. Another example is a set of functions that are additive in distinct components of X. For $X = (X_1, X_2)$ this Γ is the mean square closure of all

functions $\gamma(X) = \gamma_1(X_1) + \gamma_2(X_2)$ that are additive in in X_1 and X_2 with finite second moment. The high dimensional, additive, and unrestricted specifications of Γ are each of interest.

A leading example of the residual function is $\rho(W, \gamma) = Y - \gamma(X)$ for an outcome variable Y having finite second moment. In this example the orthogonality condition of equation (3.7) specifies that γ_0 is the least squares projection of Y on the set of functions Γ, i.e. $\gamma_0 = \arg\min_{\gamma \in \Gamma} E[\{Y - \gamma(X)\}^2]$. In this example γ_0 is the conditional expectation if Γ is all functions of X with finite second moment, or is the least squares projection of Y on the closure of linear combinations of $(b_1(X), b_2(X), ...)$, or is the least squares projection on the closure of additive functions. Newey (1994) gives the influence function for functionals of such γ_0.

There are other important examples of the residual function.

QUANTILE: In this case there is an outcome variable Y and the residual function is

$$\rho(W, \gamma) = 1(Y < \gamma(X)) - p,$$

where $0 < p < 1$. This $\rho(W, \gamma)$ is the negative of the derivative with respect to u of the "check function" $q_p(u) = |u| \{p1(u > 0) + (1 - p)1(u < 0)\}$ evaluated at $u = Y - \gamma(X)$; see Koenker and Bassett (1978). By convexity of $q_p(Y - \gamma(X))$ in γ,

$$\gamma_0 = \arg\min_{\gamma \in \Gamma} E[q_p(Y - \gamma(X))].$$

Here $\gamma_0(X)$ will the pth conditional quantile of Y when Γ is unrestricted. For other specifications of Γ the γ_0 will be minimum of the expected check function over Γ.

EXPECTILE: In this case the residual function is

$$\rho(W, \gamma) = [-p - (1 - 2p)1(Y < \gamma(X))](Y - \gamma(X)).$$

This $\rho(W, \gamma)$ is the negative of the derivative with respect to u of the asymmetric squared residual function $\bar{q}_p(u) = (u^2/2)\{p1(u > 0) + (1 - p)1(u < 0)\}$ evaluated at $u = Y - \gamma(X)$, as in Newey and Powell (1987). By convexity of $\bar{q}_p(Y - \gamma(X))$ in γ

$$\gamma_0 = \arg\min_{\gamma \in \Gamma} E[\bar{q}_p(Y - \gamma(X))].$$

Here $\gamma_0(X)$ will the pth conditional expectile of Y given X when Γ is unrestricted. For other specifications of Γ the γ_0 will be minimum of the asymmetric squared residual function over Γ.

BINARY CHOICE: In this case there is a binary outcome variable $Y \in \{0, 1\}$, a known CDF $\Lambda(a)$ with derivative (pdf) $\Lambda_a(a)$, and the residual is

$$\rho(W, \gamma) = \frac{-\Lambda_a(\gamma(X))}{\Lambda(\gamma(X))[1 - \Lambda(\gamma(X))]}\{Y - \Lambda(\gamma(X))\}.$$

[12]

This $\rho(W, \gamma)$ is $\partial Q(W, a)/\partial a$ at $a = \gamma(X)$ for the negative binary pseudo-likelihood

$$Q(W, a) = -Y \ln \Lambda(a) - (1 - Y) \ln[1 - \Lambda(a)].$$

When $\ln(\Lambda_a(a))$ is concave this $Q(W, a)$ will be convex in a, see Pratt (1981). For example the logit CDF $\Lambda(a) = e^a/[1 + e^a]$ has this property with $\Lambda_a(a)/\{\Lambda(a)[1 - \Lambda(a)]\} = 1$. The γ_0 will satisfy

$$\gamma_0 = \arg\min_{\gamma \in \Gamma} E[Q(W, \gamma(X))].$$

Here $\gamma_0(X)$ will be $\Lambda^{-1}(\Pr(Y = 1|X))$ when Γ is unrestricted. For other specifications of Γ the γ_0 will minimize the expected value of the negative log-likelihood $-E[Y \ln(\Lambda(\gamma(X))) + (1 - Y) \ln\{1 - \Lambda(\gamma(X))\}]$ over Γ.

These cases of the residual function have the common feature that $\rho(W, \gamma) = dQ(W, a)/da|_{a=\gamma(X)}$ where $Q(W, a)$ is a convex function. In all such cases equation (3.7) will be the necessary and sufficient first order condition for

$$\gamma_0 = \arg\min_{\gamma \in \Gamma} E[Q(W, \gamma)],$$

when the argmin exists and some regularity conditions are satisfied. We focus on the orthogonality condition because it is potentially more general.

3.2 The Influence Function

We derive the influence function of objects of the form

$$\theta(F) = E_F[m(W, \gamma(F))], \quad E_F[b(X)\rho(W, \gamma(F))] = 0 \text{ for all } b \in \Gamma. \tag{3.8}$$

Here the object of interest is the expectation of the function $m(W, \gamma)$ at γ_0. One example of this $\theta(F)$ is a bound on average equivalent variation discussed in Section 3.4 to follow. Other examples will be discussed later in this Section.

The influence function of $\theta(F)$ will be the sum of two terms. To explain let $F_\tau = (1 - \tau)F_0 + \tau H = F_0 + \tau(H - F_0)$, $0 < \tau < 1$, denote a convex combination of the true CDF F_0 with another CDF H as discussed in Section 2 and let $\gamma_\tau = \gamma(F_\tau)$ and $E_\tau[\cdot] = E_{F_\tau}[\cdot]$. By the chain rule of calculus,

$$\begin{aligned}
\frac{\partial}{\partial \tau}\theta(F_\tau) &= \frac{\partial}{\partial \tau}E_\tau[m(W, \gamma_0)] + \frac{\partial}{\partial \tau}E[m(W, \gamma_\tau)] \\
&= \int m(w, \gamma_0)\{H - F_0\}(dw) + \frac{\partial}{\partial \tau}E[m(W, \gamma_\tau)] \\
&= \int [m(w, \gamma_0) - \theta_0]H(dw) + \frac{\partial}{\partial \tau}E[m(W, \gamma_\tau)].
\end{aligned}$$

[13]

We see in this equation that influence function of $\theta(F)$ will be the sum of $m(w, \gamma) - \theta$ and a term $\phi(w, \gamma, \alpha)$ satisfying

$$\frac{\partial}{\partial \tau} E[m(W, \gamma_\tau)] = \int \phi(w, \gamma_0, \alpha_0) H(dw), \tag{3.9}$$

with

$$\frac{\partial}{\partial \tau} \theta(F_\tau) = \int \psi(w, \gamma_0, \alpha_0, \theta_0) H(dw), \quad \psi(w, \gamma, \alpha, \theta) = m(W, \gamma) - \theta + \phi(w, \gamma, \alpha)$$

The first term $m(w, \gamma) - \theta$ accounts for the unknown distribution F that averages over W in $m(W, \gamma_0) - \theta_0$. The second term $\phi(w, \gamma, \alpha)$ accounts for estimation of the unknown γ_0 satisfying the orthogonality condition of equation (3.7). This $\phi(w, \gamma, \alpha)$ is the adjustment term from Newey (1994) that accounts for a nonparametric estimator of γ_0 satisfying equation (3.7). We focus here on the derivation of $\phi(w, \gamma, \alpha)$.

To derive $\phi(w, \gamma, \alpha)$ we assume that $\gamma_\tau = \gamma(F_\tau)$ satisfies the orthogonality condition in equation (3.8) for each τ so that for all $b \in \Gamma$

$$E_\tau[b(X)\rho(W, \gamma_\tau)] \equiv 0 \text{ for all } b \in \mathcal{B},$$

identically in τ. We are implicitly assuming here that $\mathcal{B}$ does not depend on τ which will hold for the F_τ of Appendix A. Differentiating this identity with respect to τ and applying the chain rule of calculus, so that the derivative is the sum of derivatives with respect to τ in $E_\tau[b(X)\rho(W, \gamma_0)]$ and $E[b(X)\rho(W, \gamma_\tau)]$, gives

$$\begin{aligned} 0 &= \frac{\partial}{\partial \tau} E_\tau[b(X)\rho(W, \gamma_0)] + \frac{\partial}{\partial \tau} E[b(X)\rho(W, \gamma_\tau)] \\ &= \int b(x)\rho(w, \gamma_0) H(dw) + \frac{\partial}{\partial \tau} E[b(X)\rho(W, \gamma_\tau)], \text{ for all } b \in \Gamma. \end{aligned}$$

Solving gives

$$-\frac{\partial}{\partial \tau} E[b(X)\rho(W, \gamma_\tau)] = \int b(x)\rho(w, \gamma_0) H(dw), \text{for all } b \in \Gamma.$$

The object being integrated on the right provides a candidate for adjustment term $\phi(w, \gamma, \alpha)$. This equation will give us equation (3.9) if there is $\alpha_0 \in \Gamma$ with

$$\frac{\partial}{\partial \tau} E[m(W, \gamma_\tau)] = -\frac{\partial}{\partial \tau} E[\alpha_0(X)\rho(W, \gamma_\tau)]. \tag{3.10}$$

Such an $\alpha_0(X)$ will exist under the following two conditions.

ASSUMPTION 1: *There exists $v_m(X)$ such that $\partial E[m(W, \gamma_\tau)]/\partial \tau = \partial E[v_m(X)\gamma_\tau(X)]/\partial \tau$ and $E[v_m(X)^2] < \infty$.*

[14]

This condition is like equation (4.4) of Newey (1994) in requiring that $\partial E[m(W, \gamma_\tau)]/\partial \tau$ can be represented as the derivative of an expected product of a function $v_m(X)$ with $\gamma_\tau(X)$ where $v_m(X)$ has finite second moment. Assumption 1 can be shown to be a necessary condition for $\theta(F)$ to have a finite semiparametric variance bound. This condition is satisfied for important effects as further discussed below.

ASSUMPTION 2: There is $v_\rho(X) > 0$ that is bounded and bounded away from zero such that $\partial E[b(X)\rho(W, \gamma_\tau)]/\partial \tau = \partial E[b(X)v_\rho(X)\gamma_\tau(X)]/\partial \tau$ for every $b \in \Gamma$.

Here $v_\rho(X)$ will be the derivative of $E[\rho(W, a)|X]$ with respect to the scalar a evaluated at $a = \gamma_0(X)$ by $E[b(X)\rho(W, \gamma_\tau)] = E[b(X)E[\rho(W, \gamma_\tau)|X]]$. It allows for $\rho(W, \gamma)$ to not be continuous as long as $E[\rho(W, a)|X]$ is differentiable in a. That $v_\rho(X) > 0$ is a sign normalization while $v_\rho(X)$ being bounded and bounded away from zero is important for the results.

Under Assumptions 1 and 2 equation (3.10) becomes

$$\frac{\partial}{\partial \tau} E[v_m(X)\gamma_\tau(X)] = -\frac{\partial}{\partial \tau} E[\alpha_0(X)v_\rho(X)\gamma_\tau(X)].$$

Since $\gamma_\tau \in \Gamma$ this condition will be satisfied if for all $\gamma \in \Gamma$,

$$E[v_m(X)\gamma(X)] = -E[\alpha_0(X)v_\rho(X)\gamma(X)].$$

Subtracting $E[v_m(X)\gamma(X)]$ from both sides gives

$$\begin{aligned} 0 &= -E[v_m(X)\gamma(X)] - E[\alpha_0(X)v_\rho(X)\gamma(X)] \\ &= E[v_\rho(X)\{\frac{-v_m(X)}{v_\rho(X)} - \alpha_0(X)\}\gamma(X)] \text{ for all } \gamma \in \Gamma, \end{aligned}$$

where the second equality follows by multiplying and dividing by $v_\rho(X)$ in $-E[v_m(X)\gamma(X)]$. This is the orthogonality condition that is necessary and sufficient for $\alpha_0(X)$ to be the weighted least squares projection of $-v_m(X)/v_\rho(X)$ on Γ for weight $v_\rho(X)$. Thus we have:

PROPOSITION 1: *If Assumptions 1 and 2 are satisfied then*

$$\phi(w, \gamma, \alpha) = \alpha(x)\rho(w, \gamma), \ \alpha_0(x) = \arg\min_{\alpha \in \Gamma} E[v_\rho(X)\{-v_m(X)/v_\rho(X) - \alpha(X)\}^2].$$

This formula generalizes Proposition 4 of Newey (1994) where $\phi(w, \gamma, \alpha)$ was given for least squares projections where $\rho(W, \gamma) = Y - \gamma(W)$. Here we give $\phi(w, \gamma, \alpha)$ for any exogenous orthogonality conditions where Assumptions 1 and 2 are satisfied, including quantiles, expectiles, and binary choice.

[15]

EXAMPLE 1: *Quantile Functional;* For $\rho(W, \gamma) = 1(Y < \gamma(X)) - p$

$$v_\rho(X) = \frac{\partial \Pr(Y < \gamma_0(X) + a|X)]}{\partial a} = f_{Y|X}(\gamma_0(X)|X)$$

where $f_{Y|X}(y|X)$ is the pdf of Y conditional on X. The adjustment term is

$$\phi(w, \gamma, \alpha) = \alpha(X)[1(y < \gamma(x)) - p],$$

where α_0 is given in Proposition 1. The formula for α_0 depends on the functional $m(W, \gamma)$ through the derivative term $v_m(W)$ and is given by

$$\alpha_0(x) = \arg\min_{\alpha \in \Gamma} E[f_{Y|X}(\gamma_0(X)|X)\{-v_m(X)/f_{Y|X}(\gamma_0(X)|X) - \alpha(X)\}^2].$$

For instance consider a weighted average derivative functional where $m(W, \gamma) = w(x)\partial\gamma(x)/\partial x_1$. Integration by parts gives

$$
\begin{aligned}
E[m(W, \gamma)] &= \int w(x)\frac{\partial\gamma(x)}{\partial x}f_0(x)dx = -\int \frac{\partial\{w(x)f_0(x)\}}{\partial x_1}\gamma(x)dx = E[v_m(X)\gamma(X)], \\
v_m(X) &= -\frac{1}{f_0(X)}\frac{\partial\{w(X)f_0(X)\}}{\partial x_1}.
\end{aligned}
$$

When Γ is unrestricted Proposition 1 given $\alpha_0(X) = -v_m(X)/f_{Y|X}(\gamma_0(X)|X)$ and the adjustment term coincides with that of Chauduri, Doksum, and Tsybakov (1997). Ackerberg et al. (2014) also gave an expression for the adjustment term for functionals other than the weighted average derivative with $v_m(X)$ replaced by a functional derivative of $E[m(W, \gamma)]$. When Γ is restricted then $\alpha_0(X)$ being the weighted projection of $-v_m(X)/f_{Y|X}(\gamma_0(X)|X)$ on Γ with weight $f_{Y|X}(\gamma_0(X)|X)$. Proposition 1 generalizes the previous results to allow restrictions on γ.

EXAMPLE 2: *Expectile Functional;* For a conditional expectile $\rho(W, \gamma) = -[p1(Y > \gamma(X)) + (1 - p)1(Y < \gamma(X))][Y - \gamma(X)]$, so that

$$v_\rho(X) = p\Pr(Y > \gamma_0(X)|X) + (1 - p)\Pr(Y < \gamma_0(X)|X),$$

which is bounded and bounded away from zero. The adjustment term is

$$\phi(w, \gamma, \alpha) = -\alpha(X)[p1(Y > \gamma(X)) + (1 - p)1(Y < \gamma(X))][Y - \gamma(X)],$$

where $\alpha_0(X)$ is given in Proposition 1. The formula for α_0 depends on the functional $m(W, \gamma)$ through the derivative term $v_m(W)$ and is given. When Γ is unrestricted and $m(W, \gamma) = w(x)\partial\gamma(x)/\partial x_1$ then $v_m(X)$ will be as in Example 1 and $\alpha_0(X) = -v_m(X)/v_\rho(X)$. We are not aware of previous results on the adjustment term for functions that minimize the expectile objective function.

[16]

Examples 1 and 2 illustrate how the term $v_m(X)$ is determined by the functional of interest while $v_\rho(X)$ is determined by the residual $\rho(W, \gamma)$. Proposition 1 shows how these aspects are combined to determine the $\alpha_0(X) \in \mathcal{B}$ that multiplies the residual $\rho(W, \gamma)$ to form the adjustment term. From equation (3.10) we see that this $\alpha_0(X)$ is precisely the function that makes the effect of γ_τ on $E[m(W, \gamma_\tau)]$ equal to the effect of γ_τ on $-E[\alpha_0(X)\rho(W, \gamma_\tau)]$. Proposition 1 shows that this $\alpha_0(X)$ is a projection of $-v_m(X)/v_\rho(X)$ on Γ weighted by $v_\rho(X)$

The explicit formula in Proposition 1 is useful for quantifying local policy effects and local sensitivity of semiparametric estimators, as we will illustrate in the remainder of this Section. Proposition 1 also illustrates how the influence function can be obtained with calculus, under natural differentiability conditions like Assumptions 1 and 2. They key steps in deriving Proposition 1 are to use the first order condition for $\gamma(F)$ to derive candidates for the influence function and to show that equation (3.9) is satisfied for one of those candidates.

3.3   Generalizing the Omitted Variable Bias Formula

The influence function for exogenous orthogonality conditions can be used to quantify local sensitivity to distributional changes of any object with an influence function. We consider structural changes where the distribution of X remains the same but the distribution of the outcome variable Y given X is different. We focus on the case where $m(w, \gamma)$ depends only on x, which covers many examples of interest and leads to simple, intuitive formulas. We consider H where the marginal distribution of X is the same as for F_0 but $\rho(W, \gamma_0)$ may not be orthogonal to Γ. Because $E_H[m(W, \gamma_0)] = E[m(X, \gamma_0)] = \theta_0$ the local sensitivity to such H is given by the following result:

PROPOSITION 2: *If Assumptions 1 and 2 are satisfied, $m(W, \gamma_0)$ depends only on X, and H has the same marginal distribution of X as F_0 then*

$$\frac{d\theta(F_\tau)}{d\tau} = E_H[\alpha_0(X)\rho(W, \gamma_0)]. \tag{3.11}$$

Here we see that the local sensitivity is the expected product of $\alpha_0(X)$ with the conditional mean of the residual $\rho(W, \gamma_0)$ under the alternative distribution H. This local sensitivity formula generalizes the classic omitted variable bias formula to the local bias of any object that depends on the solution to an exogenous orthogonality condition, as we now demonstrate.

EXAMPLE 3: *Omitted Variable Bias Formula;* Here we show that the classic omitted variable bias formula is a special case of Proposition 2. Consider the conditional mean $\gamma_0(X) = E[Y|X]$

where X has finite support and let D be the indicator function of one of the possible discrete outcomes of X. Then there is Z, θ_0, and γ_0 such that

$$E[Y|D, Z] = \gamma_0(X) = D\theta_0 + Z'\gamma_0.$$

Take the object of interest to be θ_0. Let $\tilde{D} = D - E[D|Z]$ be the residual from the population least squares regression of D on Z. Then the coefficient θ_0 is a functional of $\gamma_0(X)$ given by

$$\theta_0 = E[\alpha_0(X)\gamma_0(X)], \ \alpha_0(X) = \frac{\tilde{D}}{E[\tilde{D}^2]}.$$

Let $\varepsilon := Y - \gamma_0(X) = \rho(W, \gamma_0)$. The sensitivity is then

$$\frac{d\theta(F_\tau)}{d\tau} = E[\alpha_0(X)E_H[Y - \gamma_0(X)|X]] = \frac{E[\tilde{D}E_H[\varepsilon|X]]}{E[\tilde{D}^2]}.$$

If there is an omitted variable $\tilde{Z}$ under H so that the distribution ε is the same as $Y - \gamma_0(X) - \tilde{Z}$ then

$$\frac{d\theta(F_\tau)}{d\tau} = E[\alpha_0(X)E_H[\varepsilon|X]] = \frac{E[\tilde{D}E_H[\tilde{Z}|X]]}{E[\tilde{D}^2]}.$$

This formula is the classic omitted variables bias formula.

Example 4 shows that Proposition 2 generalizes the omitted variables bias formula for one coefficient of a linear regression to any object that depends on a solution to an exogenous orthogonality condition. We will illustrate that generalization by estimating the local sensitivity of a bound on average equivalent variation to endogeneity of the price in a gasoline demand application.

An estimator of the local sensitivity can be obtained from an estimator $\hat{\alpha}(x)$ of the term $\alpha_0(x)$ in the influence function and from a specification $\hat{H}$ of the joint distribution of X and $\rho(W, \gamma_0)$ under misspecification as

$$\widehat{\frac{d\theta(F_\tau)}{d\tau}} = \int [\hat{\alpha}(x)\rho(w, \hat{\gamma})] \, \hat{H}(dw).$$

Construction of a local Hausman test based on this object would require an estimator of the asymptotic variance of the sensitivity $\widehat{d\theta(F_\tau)}/d\tau$. It is beyond the scope of this to derive the asymptotic variance of the sensitivity and construct a consistent estimator of that asymptotic variance, although a bootstrap variance estimator could be used and should prove valid. We will illustrate in the gasoline demand example how this could be done.

An important part of $\widehat{d\theta(F_\tau)}/d\tau$ is an estimator $\hat{\alpha}(x)$ of the $\alpha_0(x)$ that appears in the adjustment term of Proposition 1. One way to construct $\hat{\alpha}(x)$ is to use equation (3.10) and a

series approximation using a dictionary of functions $b(x) = (b_1(x), ..., b_p(x))'$ with $b_j \in \mathcal{B}$ for each j. Equation (3.10) with $\gamma_\tau = \gamma_0 + \tau b_j$ gives

$$\frac{\partial}{\partial \tau} E[m(W, \gamma_0 + \tau b_j)] = -E[\alpha_0(X)\rho_\gamma(W, \gamma_0)b_j(X)], \ (j = 1, ..., p), \tag{3.12}$$

where $\rho_\gamma(W, \gamma) = \partial\rho(W, \gamma + a)/\partial a$ for a constant a at $a = 0$. These are moment conditions that can be used to estimate $\alpha_0(X)$ as a linear combination of basis functions. The idea is to replace expectations with sample averages, γ_0 with an estimator $\hat{\gamma}$, and $\alpha_0(X)$ with a linear combination $\pi'b(X)$ and then solve for an estimator of π. Let

$$\hat{M} = (\hat{M}_1, ..., \hat{M}_p)', \ \hat{M}_j = \frac{\partial}{\partial \tau} \frac{1}{n} \sum_{i=1}^{n} m(W_i, \hat{\gamma} + \tau b_j),$$

$$\hat{G} = \frac{1}{n} \sum_{i=1}^{n} \rho_\gamma(W_i, \hat{\gamma})b(X_i)b(X_i)'.$$

Then a version of equation (3.12) that replaces with sample moments, $\hat{\gamma}$, and $\pi'b(X)$ in place of $\alpha_0(X)$ is $\hat{M} = -\hat{G}\pi$. Solving for π gives

$$\hat{\alpha}(x) = \hat{\pi}'b(x), \ \hat{\pi} = -\hat{G}^{-1}\hat{M}. \tag{3.13}$$

For a quantile orthogonality conditions where $\rho(W, \gamma)$ is not continuous one can use kernel weighting to construct $\hat{G}$ as in Example 2 of Chernozhukov et al. (2020).

For regression where $\rho(W, \gamma) = Y - \gamma(X)$ this $\hat{\alpha}(x)$ is the same as in Equation (6.2) from Newey (1994). For other choices of $\rho(W, \gamma)$ this $\hat{\alpha}(x)$ could be derived from series expansions given in Ai and Chen (2007), Ackerberg, Chen, and Hahn (2012), and Ackerberg et al. (2014) for conditional moment restrictions and Chen and Liao (2015) more generally. Such interesting estimators of the adjustment term would be particularly useful when its form is not known. Here we rely on the explicit moment condition for $\alpha_0(X)$ in equation (3.12) that is a special case of the Chernozhukov et al. (2020) use of orthogonal moment functions to estimate functions on which they depend.

3.4 Sensitivity of Average Equivalent Variation for Gasoline Demand

One object that depends on a conditional expectation is the Hausman and Newey (2016) bound on average equivalent variation (AEV) for heterogenous demand. This bound allows for completely general heterogeneity where the demand function for each person can be unique to that person. The bound does depend on preferences being independent of observed price and income, a strong exogeneity restriction. Here we test the effect of dropping that exogeneity restriction on AEV using the local sensitivity results we have obtained.

[19]

An important motivation for this test is the difficulty of allowing for endogeneity with general heterogeneity. Endogeneity can be allowed for using control functions, as in Hausman and Newey (2016), but existence of control functions imposes strong restrictions as in Blundell and Matzkin (2014). Blundell, Horowitz, and Parey (2017) allow for endogeneity where there is an instrument for price but restrict heterogeneity to be scalar where bounds on AEV are not known. Here we take a different approach to allowing for endogeneity, where we test for sensitivity to bounds on AEV to endogeneity.

To describe and carry out this test we first describe the AEV bound and apply Proposition 1 to derive its influence function.

EXAMPLE 4: *Average Equivalent Variation Bound;* Here Y is the share of income spent on a commodity and $X = (P_1, Z)$, where P_1 is the price of the commodity and Z includes income Z_1, prices of other goods, and other observable variables affecting utility. Let $\check{p}_1 < \bar{p}_1$ be lower and upper prices over which the price of the commodity can change, κ a bound on the income effect, and $\omega(z)$ some weight function. The object of interest is

$$\theta_0 = E\left[\omega(Z)\int_{\check{p}_1}^{\bar{p}_1}\left(\frac{Z_1}{u}\right)\gamma_0(u, Z)\exp(-\kappa[u - \check{p}_1])du\right], \qquad (3.14)$$

where u is a variable of integration. If individual heterogeneity in consumer preferences is independent of X and κ is a lower (upper) bound on the derivative of consumption with respect to income across all individuals, then θ_0 is an upper (lower) bound on the weighted average over consumers and over the distribution of Z of equivalent variation for a change in the price of the first good from $\check{p}_1$ to $\bar{p}_1$.

This object is a special case of that considered in Proposition 1 where $v(u) = u^2/2$, $\gamma_0(X) = E[Y|X]$, and $m(w, \gamma)$ depends only on x and is given by

$$m(x, \gamma) = \omega(z)\int_{\check{p}_1}^{\bar{p}_1}(z_1/u)\gamma(u, z)\exp(-\kappa[u - \check{p}_1])du.$$

From the form of $E[m(X, \gamma)]$ and multiplying and dividing by the conditional pdf $f(p_1|z)$ we find

$$\alpha_0(x) = f(p_1|z)^{-1}\omega(z)1(\check{p}_1 < p_1 < \bar{p}_1)(z_1/p_1)\exp(-\kappa[p_1 - \check{p}_1]).$$

where $f(p_1|z)$ is the conditional pdf of P_1 given Z.

We apply Example 4 to test sensitivity of a bound on AEV to endogeneity of price using gasoline demand data in Hausman and Newey (2016, 2017) and Blundell, Horowitz, and Parey (2017). We use the estimator $\hat{\alpha}(x)$ given in equation (3.13) for several choices of basis functions. For an estimate of $\varepsilon = Y - \gamma(X)$ that allows for endogeneity we use a linear instrumental variable

estimator where the share equation has a constant, ln(price), and ln(income) with the Blundell, Horowitz, and Parey (2017) price instrument that is the distance from the Gulf of Mexico. We take $\hat{\varepsilon}_i$, $(i = 1, ..., n)$ to be the residuals from the linear instrumental variables estimation and the sensitivity estimator to be

$$\frac{\widehat{d\theta(F_\tau)}}{d\tau} = \frac{1}{n} \sum_{i=1}^{n} \hat{\alpha}(X_i)\hat{\varepsilon}_i.$$

This sensitivity estimate will depart from zero when $\hat{\alpha}(X_i)$, which depends on the price variable, is correlated with the instrumental variables residuals $\hat{\varepsilon}_i$. In this application we use the delta method and standard calculations to obtain standard a standard error for the sensitivity estimator.

We use gasoline demand data from the 2001 U.S. National Household Transportation Survey (NHTS). This survey is conducted every 5-8 years by the Federal Highway Administration. The survey is designed to be a nationally representative cross section which captures 24-hour travel behavior of randomly-selected households. Data collected includes detailed trip data and household characteristics such as income, age, and number of drivers. We restrict our estimation sample to households with either one or two gasoline-powered cars, vans, SUVs and pickup trucks. We exclude Alaska and Hawaii. We use daily gasoline consumption, monthly state gasoline prices, and annual household income. The data we use consists of 8,908 observations. Note that the mean price of gasoline was \$1.33 per gallon with the mean number of drivers in a household equal to 2.04.

We specify the weight function in the measure of AEV to be $\omega(Z) = 1$ and consider a price change from the mean of price in the data to a price that is 10 percent higher. We set $\kappa = 0$ so that the sensitivity will be for a lower bound on AEV when gasoline is a normal good (the income effect is positive) for all consumers. For the basis function $b(x)$ used to estimate $\hat{\alpha}(x)$ we consider bivariate linear, quadratic, and cubic function in ln(price) and ln(income). Because their presence had little effect on AEV estimates in Hausman and Newey (2016, 2017) we do not use covariates here. We do use simulation to estimate the integral that appears in $m(x, \gamma)$ in the bound. For u_i uniformly distributed on $[\check{p}_1, \bar{p}_1]$ the $\hat{\alpha}(x)$ is given by

$$\hat{\alpha}(x) = \hat{\pi}'b(x), \ \ \hat{\pi} = [\sum_{i=1}^{n} b(X_i)b(X_i)']^{-1}(\bar{p}_1 - \check{p}_1) \sum_{i=1}^{n} \left(\frac{Z_{1i}}{u_i}\right) b(u_i, Z_i),$$

where $x = (p_1, z')'$ and z_1 is income.

Table 1 reports the sensitivity estimates and their standard errors for linear, quadratic, and cubic specifications of $b(x)$.

Table 1: AEV Sensitivity to Endogeneity

	Sensitivity	AEV Bound
Linear	1.44	25.08
	(.554)	(1.37)
Quadratic	.487	33.93
	(.640)	(1.05)
Cubic	-1.20	32.27
	(.946)	(.805)

We find statistically significant evidence of sensitivity to endogeneity for the linear specification of demand but not for the quadratic or cubic. We also find that the sensitivity estimates are quite small for all three specifications. This absence of sensitivity of the AEV bound to endogeneity suggests there is little need in this application to allow for price endogeneity in the estimation of a lower bound on AEV.

4 Endogenous Orthogonality Conditions

There are many interesting economic and causal effects that depend on functions satisfying endogenous orthogonality conditions where the function of interest depends on variables that are not instruments. Such solutions to orthogonality conditions come from first order conditions to economic choice problems or define causal functions of interest. Objects of interest that depend on such functions include policy and sensitivity effects like those of Sections 2 and 3.

In this Section we derive the influence function for effects that depend on the probability limit of a nonparametric instrumental variables (NPIV) estimator like those in Newey and Powell (2003), Newey (1991), and Ai and Chen (2003). We consider an estimator $\hat{\gamma}$ with a probability limit $\gamma_0 = \gamma(F_0)$ that is the unique solution to orthogonality conditions

$$E[b(X)\rho(W, \gamma)] = 0, \ b \in \mathcal{B}, \ \gamma \in \Gamma. \tag{4.15}$$

Here $\mathcal{B}$ is a linear set of possible instrumental variables $b(X)$ and γ is restricted to a linear set Γ similar to Section 3.1. We depart from Section 3.1 in allowing the unknown function γ to depend on variables Z that are different than the instruments X. This set up generalizes the conditional moment restrictions environment of Newey and Powell (1989, 2003), Newey (1991), and Ai and Chen (2003) to orthogonality conditions with linear restrictions on γ.

Restrictions on the structural functions and on the instrumental variables are of interest to empirical researchers for at least two reasons. First imposing correct restrictions on the structural function can improve efficiency of the estimator and mitigate the well known ill-posed inverse problem for NPIV that can lead to imprecise estimators. For example imposing partially linear or additive structure on γ can make estimators more precise. Second imposing restrictions

on the instrumental variables can help reduce the well known Nagar (1959) instrumental variable bias. Such biases are known to be important in empirical applications such as Angrist and Kreuger (1991). By allowing such restrictions we provide the researcher with more flexibility to choose a model that can lead to good inference properties for policy or sensitivity analysis with endogeneity. We leave to future work the application of the results of this Section to policy and sensitivity analysis. We focus here on showing how Steps I and II can be used to derive influence functions in complicated and important settings which is a primary purpose of this paper.

4.1 The Estimator

We will derive influence functions for $\hat{\gamma}$ that is a first step NPIV estimator based on the orthogonality conditions in equation (4.15). Let $b^K(x) = (b_1(x), ..., b_K(x))'$ be the first K elements of a sequence of instrumental variables. We assume that $b^K(X)$ spans $\mathcal{B}$ as K grows meaning that any element of $\mathcal{B}$ can be approximated arbitrarily well by a linear combination of $b^K(X)$ for K large enough. The NPIV estimator we consider is

$$\hat{\gamma} = \arg \min_{\gamma \in \Gamma_n} \hat{Q}(\gamma), \tag{4.16}$$

$$\hat{Q}(\gamma) = \frac{1}{n} \sum_{i.=1}^n \rho(W_i, \gamma) b^K(X_i)^T \left(\sum_{i=1}^n b^K(X_i) b^K(X_i)^T \right)^- \sum_{i=1}^n b^K(X_i) \rho(W_i, \gamma),$$

where Γ_n is a subset of Γ and A^- denotes a generalized inverse of a matrix A. For example, Γ_n could be the set of linear combinations of L functions $p_1(z), ..., p_L(z)$ where $p_\ell(\cdot) \in \Gamma$ for each ℓ. We assume that a minimum exists with probability approaching one, as could be guaranteed in some settings using Chen and Pouzo (2015). This $\hat{\gamma}$ has the form of NPIV given in Newey and Powell (1989, 2003), Newey (1991), Ai and Chen (2003), and Darolles, Florens, and Renault (2011). We differ from this prior work in allowing the instrumental variables to be restricted to the set $\mathcal{B}$.

The influence function for the object of interest will depend on the plim γ_τ of $\hat{\gamma}$ when the distribution of W is $F_\tau = (1-\tau)F_0 + \tau H$. Since $\hat{\gamma}$ minimizes the sample objective function $\hat{Q}(\gamma)$ the usual extremum estimator theory (e.g. Amemiya, 1985), will imply that γ_τ is the minimum of the plim $Q_\tau(\gamma)$ of $\hat{Q}(\gamma)$ when the distribution of W is F_τ. To describe $Q_\tau(\gamma)$ assume that $\mathcal{B}$ does not depend on τ, which can be shown to hold under regularity conditions on H. Let $\pi_\tau(a(W)|X)$ denote the linear projection of $a(W)$ on $\mathcal{B}$ when W has CDF F_τ, satisfying

$$\pi_\tau(a(W)|X) \in \mathcal{B}, \ E_\tau[\{a(W) - \pi_\tau(a(W)|X)\}b(X)] = 0 \text{ for all } b(X) \in \mathcal{B} \tag{4.17}$$

[23]

Then it follows exactly as in Newey (1991) that for $K \longrightarrow \infty$ and $K/n \longrightarrow 0$,

$$\text{plim}(\hat{Q}(\gamma)) = Q_\tau(\gamma) := E_\tau[\{\pi_\tau(\rho(W, \gamma)|X)\}^2]. \tag{4.18}$$

Intuitively, from standard regression results we see that $\hat{Q}(\gamma)$ is the sample average of squares of predicted values from the least squares regression of $\rho(W_i, \gamma)$ on $b^K(X_i)$, $(i = 1, ..., n)$. Then by the law of large numbers, consistency of a sample regression for a population regression, and the growth of K it will follow that plim of $\hat{Q}(\gamma)$ will be expected value of the square of the predicted value from the population regression of $\rho(W, \gamma)$ on $\mathcal{B}$, giving equation (4.18). It then follows by extremum estimator theory and from Γ_n assumed to approximate Γ that

$$\text{plim}(\hat{\gamma}) = \gamma_\tau := \arg\min_{\gamma \in \Gamma} Q_\tau(\gamma).$$

We will assume that γ_τ is unique, which could be shown to hold under more primitive conditions in Chen and Pouzo (2015).

As in Section 3.2 the focus of this Section is deriving the adjustment term $\phi(w, \gamma, \alpha)$ that satisfies $\partial E[m(W, \gamma_\tau)]/\partial \tau = \int \phi(w, \gamma_0, \alpha_0) H(dw)$. The first order condition for γ_τ has a key role in deriving the adjustment term. To describe the first order condition let $\Delta \in \Gamma$ denote a possible deviation of γ away from γ_τ. Assume that there is $v_{\rho\tau}(W)$ such that

$$\frac{\partial \pi_\tau(\rho(W, \gamma_\tau + \zeta\Delta)|X)}{\partial \zeta} = \pi_\tau(v_{\rho\tau}(W)\Delta(Z)|X).$$

The calculus of variations, first order condition for the minimization of $Q(\gamma_\tau + \zeta\Delta)/2$ at $\zeta = 0$ is

$$\begin{aligned}
0 &= \left. \frac{d}{d\zeta} E_\tau[\{\pi_\tau(\rho(W, \gamma_\tau + \zeta\Delta)|X)\}^2]/2 \right|_{\zeta=0} \tag{4.19} \\
&= E_\tau[\pi_\tau(\rho(W, \gamma_\tau)|X)\frac{\partial \pi_\tau(\rho(W, \gamma_\tau + \zeta\Delta)|X)}{\partial \zeta}] \\
&= E_\tau[\pi_\tau(\rho(W, \gamma_\tau)|X)\pi_\tau(v_{\rho\tau}(W)\Delta(Z)|X)] \text{ for all } \Delta \in \Gamma,
\end{aligned}$$

identically in τ. This first order condition has a form analogous to two-stage least squares, being orthogonality of the residual $\rho(W, \gamma_\tau)$ with instruments obtained by projecting the derivative of the residual on the set of instrumental variables. We use this first order condition and the orthogonality condition in equation (4.18) to characterize the adjustment term.

4.2 The Adjustment Term

Similarly to Section 3 the influence function of $\theta(F) = E_F[m(W, \gamma(F))]$ will be the sum of $m(W, \gamma_0) - \theta_0$ and the adjustment term. We focus on derivation of the adjustment term here. To characterize the adjustment term we proceed analogously to Section 3.2 by differentiating

the first order condition with respect to τ and applying the chain rule. For notational simplicity let $\pi(A(W)|X)$ denote the projection of $A(W)$ on $\mathcal{B}$ for $\tau = 0$. We carry out these calculations for the case where $\pi(\rho(W, \gamma_0)|X) = 0$, where either the orthogonality conditions are correctly specified or γ_0 is exactly identified so that the plim of $\hat{\gamma}$ solves the orthogonality conditions (see Chen and Santos, 2015, for exact identification). In Appendix B we derive the adjustment term under misspecification where $\pi(\rho(W, \gamma_0)|X) \neq 0$.

Differentiating the identity of equation (4.19) with respect to τ, using the third equality and $\pi(\rho(W, \gamma_0)|X) = 0$, gives

$$0 = \frac{\partial}{\partial \tau} E[\pi_\tau(\rho(W, \gamma_\tau)|X)\pi(v_\rho(W)\Delta(Z)|X)] \text{ for all } \Delta \in \Gamma, \quad (4.20)$$

where $v_\rho(W) = v_{\rho 0}(W)$. Define the set $\mathcal{A}$ to be the mean square closure of the set of $\pi(v_\rho(W)\Delta(Z)|X)$ for $\Delta \in \Gamma$, i.e.

$$\mathcal{A} = \{\alpha(X) : \text{for all } \varepsilon > 0 \text{ there is } \Delta(Z) \in \Gamma \text{ with } E[\{\alpha(X) - \pi(v_\rho(W)\Delta(Z)|X)\}^2] < \varepsilon\}, \quad (4.21)$$

Then the first order condition in equation (4.19) becomes

$$0 = \frac{\partial}{\partial \tau} E[\pi_\tau(\rho(W, \gamma_\tau)|X)\alpha(X)] \text{ for all } \alpha \in \mathcal{A}.$$

Next we use the orthogonality condition (4.17) for the projection. Because $\mathcal{A}$ is a subset of $\mathcal{B}$ it follows that

$$E_\tau[\rho(W, \gamma_\tau)\alpha(X)] = E_\tau[\pi_\tau(\rho(W, \gamma_\tau)|X)\alpha(X)] \text{ for all } \alpha \in \mathcal{A}$$

identically in τ. Differentiating both sides with respect to τ and applying the chain rule gives

$$\frac{\partial}{\partial \tau} E_\tau[\rho(W, \gamma_\tau)\alpha(X)] = \frac{\partial}{\partial \tau} E_\tau[\pi(\rho(W, \gamma_0)|X)\alpha(X)] + \frac{\partial}{\partial \tau} E[\pi_\tau(\rho(W, \gamma_\tau)|X)\alpha(X)] = 0,$$

by $\pi(\rho(W, \gamma_0)|X) = 0$ and equation (4.20). Applying the chain rule to the left-and side and solving then gives

$$-\frac{\partial}{\partial \tau} E[\rho(W, \gamma_\tau)\alpha(X)] = \int \alpha(x)\rho(w, \gamma_0)H(dw) \text{ for all } \alpha \in \mathcal{A}. \quad (4.22)$$

Similarly to Section 3.1 the object being integrated on the right provides a candidate for adjustment term $\phi(w, \gamma, \alpha)$. To find $\alpha_0(X)$ such that equation (3.10) is satisfied we impose the following conditions.

ASSUMPTION 3: *There exists* $v_m(Z)$ *such that*

$$\frac{\partial}{\partial \tau} E[m(W, \gamma_\tau)] = \frac{\partial}{\partial \tau} E[v_m(Z)\gamma_\tau(Z)], \ E[v_m(X)^2] < \infty.$$

[25]

This condition is analogous to Assumption 1 in specifying an expected product form for $dE[m(W, \gamma_\tau)]/d\tau$, and similarly will be required for existence of the adjustment term.

ASSUMPTION 4: *There exists* $v_\rho(W)$ *such that for all* $b \in B$,

$$\frac{\partial}{\partial \tau} E[\rho(W, \gamma_\tau) b(X)] = \frac{\partial}{\partial \tau} E[v_\rho(W) \gamma_\tau(Z) b(X)].$$

This condition is similar to Assumption 2 in specifying a derivative condition involving the residual $\rho(W, \gamma)$ as a function of γ.

Unlike Section 3 the differentiability conditions in Assumptions 3 and 4 are not sufficient to show that the adjustment term has the form $\alpha(x)\rho(w, \gamma)$ for some $\alpha_0(x)$. The presence of endogeneity, where γ depends on variables different than the instrumental variables X, creates the need for a link between $v_m(Z)$, functions of X, and $v_\rho(W)$. The following condition establishes the needed link. Let $\Pi(d(W)|Z) = \arg\min_{\gamma \in \Gamma} E[\{d(W) - \gamma(Z)\}^2]$ denote the least squares projection of a function $d(W)$ on Γ.

ASSUMPTION 5: *There is* $b_m(X) \in \mathcal{B}$ *such that*

$$\Pi(v_m(Z)|Z) = -\Pi(v_\rho(W) b_m(X)|Z).$$

This condition requires that the projection of $v_m(Z)$ on Γ must be equal to the projection of $-v_\rho(W) b_m(X)$ on Γ for some instrumental variable $b_m(X)$. This condition is restrictive in a way that is related to the Severini and Tripathi (2012) necessary conditions for root-n consistent estimation as discussed in Example 6 to follow.

Assumptions 3-5 imply that the adjustment term will have the form $\alpha(X)\rho(W, \gamma)$ where $\alpha_0(X)$ is the least squares projection of $b_m(X)$ on $\mathcal{A}$. To see this note that by $\gamma_\tau \in \Gamma$ and Assumption 5,

$$
\begin{aligned}
E[v_m(Z)\gamma_\tau(Z)] &= E[\Pi(v_m(Z)|Z)\gamma_\tau(Z)] = -E[\Pi(v_\rho(W) b_m(X)|Z)\gamma_\tau(Z)] \\
&= -E[v_\rho(W) b_m(X)\gamma_\tau(Z)] = -E[b_m(X)\pi(v_\rho(W)\gamma_\tau(Z)|X)] \\
&= -E[\alpha_0(X)\pi(v_\rho(W)\gamma_\tau(Z)|X)] = -E[\alpha_0(X)v_\rho(W)\gamma_\tau(Z)],
\end{aligned}
$$

for all τ where the fifth equality follows by $\pi(v_\rho(W)\gamma_\tau(Z)|X) \in \mathcal{A}$. Then by Assumption 3 and 4 and differentiating we have

$$
\begin{aligned}
\frac{d}{d\tau} E[m(W, \gamma_\tau)] &= \frac{d}{d\tau} E[v_m(Z)\gamma_\tau(Z)] = -\frac{d}{d\tau} E[\alpha_0(X)v_\rho(W)\gamma_\tau(Z)] \\
&= -\frac{d}{d\tau} E[\alpha_0(X)\rho(W, \gamma_\tau)] = \int \alpha_0(x)\rho(w, \gamma_0) H(dw).
\end{aligned}
$$

[26]

where the last equality follows from equation (4.22). This equation shows the following result:

PROPOSITION 3: *If Assumptions 3-5 are satisfied and $\pi(\rho_0(W, \gamma_0)|X) = 0$ then the adjustment term is*

$$\phi(w, \gamma, \alpha) = \alpha(x)\rho(w, \gamma),$$

where $\alpha_0(X)$ is the least squares projection of $b_m(X)$ on A satisfying

$$\alpha_0(X) = \arg\min_{\alpha \in \mathcal{A}} E[\{b_m(X) - \alpha(X)\}^2].$$

The derivation of Proposition 3 is more complicated than Proposition 1 because of endogeneity and the link condition in Assumption 5. The function $\alpha_0(X)$ quantifies how the instrumental variables affect the adjustment term. It is constrained to be an element of $\mathcal{A}$ because NP2SLS projects functions of Z on the set of instrumental variables $\mathcal{B}$, just as parametric two-stage least square does. When multiple sets of orthogonality conditions are available, e.g. as could be the case if $E[\rho(W, \gamma_0)|X] = 0$, $\alpha_0(X)$ can vary with $\mathcal{B}$. This effect of the choice of $\mathcal{B}$ on the influence function is analogous to parametric instrumental variables estimation, where the influence function can vary with the choice of linear combination of instrumental variables.

EXAMPLE 5: *Additive Structural Functions and Instruments;* We consider NPIV where $\gamma(Z) = \gamma_1(Z_1) + \gamma(Z_2)$ is restricted to be additive in distinct components Z_1 and Z_2 of $Z = (Z_1, Z_2)$. Such a restriction can reduce the severity of the ill-posed inverse problem. The instrumental variables $b(X) = b_1(X_1) + b_2(X_2)$ are also restricted to be additive in distinct components of X_1 and X_2 of X. Such a restriction can identify the additive components $\gamma_1(Z_1)$ and $\gamma_2(Z_2)$ while while limiting the number of instrumental variables to reduce the Nagar (1959) bias of instrumental variables estimators. Here Γ and $\mathcal{B}$ are mean square closures of sets of additive functions. It will be convenient here to just refer to additive functions rather the mean square closures of sets of functions, though not every function in the closure need be additive.

One thing of note about the adjustment term here is that $\alpha_0(X)$ is in $\mathcal{B}$ and so it is an additive function of X_1 and X_2. The form of $\alpha_0(X)$ will be determined by the form of $v_m(Z)$ and $v_\rho(W)$ and the link condition of Assumption 5. Here $\Pi(A(W)|Z)$ is the projection on (the mean square closure of) additive functions. Also the elements of $\mathcal{B}$ are (in the closure of) additive functions. Suppose that the residual is linear with

$$\rho(W, \gamma) = Y - \gamma_1(Z_1) - \gamma_2(Z_2).$$

Then $v_\rho(W) = -1$ so that Assumption 5 is existence of $b_m \in \mathcal{B}$ with

$$\Pi(v_m(Z)|Z) = \Pi(b_m(X)|Z).$$

[27]

This requires that the projection of $v_m(Z)$ on additive functions of Z_1 and Z_2 must be equal to the projection of an additive function of X_1 and X_2 on additive functions of Z_1 and Z_2. For example if Z_1 is a scalar and $m(w, \gamma) = \omega(z_1)\partial\gamma_1(z_1)/\partial z_1$ then as in Example 1,

$$v_m(Z) = -\frac{1}{f_0(Z)}\frac{\partial\{w(Z_1)f_0(Z)\}}{\partial z_1} = -\frac{\partial w(Z_1)}{\partial z_1} - w(Z_1)\frac{\partial f_0(Z)/\partial z_1}{f_0(Z)}.$$

Here it would suffice for Assumption 5 that there $b^I(X_1)$ and $b^{II}(X) = b_1^{II}(X_1) + b_2^{II}(X_2)$ such that

$$-\frac{\partial w(Z_1)}{\partial z_1} = E[b^{II}(X_1)|Z_1], \ \Pi(w(Z_1)\frac{\partial f_0(Z)/\partial z_1}{f_0(Z)}|Z) = \Pi(b^{II}(X)|Z) \qquad (4.23)$$

For quantile orthogonality conditions where $\rho(W, \gamma) = 1(Y < \gamma(Z)) - p$, it follows similarly to Section 3 that

$$v_\rho(W) = f(\gamma_0(Z)|Z, X),$$

where $f(Y|Z, X)$ is the pdf of Y conditional on Z and X. Assumption 5 is then existence of $b_m \in \mathcal{B}$ with

$$\Pi(v_m(Z)|Z) = \Pi(f(\gamma_0(Z)|Z, X)b_m(X)|Z).$$

This requires that the projection of $v_m(Z)$ on additive functions of Z_1 and Z_2 must be equal to the projection of a weighted additive function of X_1 and X_2 on additive functions of Z_1 and Z_2.

This condition also imposes smoothness restrictions on $v_m(Z)$, as further discussed in Example 7 to follow.

The function $\alpha_0(X)$ quantifies how the instrumental variables affect the influence function. It is constrained to be an element of $\mathcal{A}$ because NP2SLS projects functions of Z on the set of instrumental variables $\mathcal{B}$, just as parametric 2SLS does. When multiple sets of orthogonality conditions are available, e.g. as could be the case if $E[\rho(W, \gamma_0)|X] = 0$, $\alpha_0(X)$ can vary with $\mathcal{B}$. This effect of the choice of $\mathcal{B}$ on the influence function is analogous to parametric IV estimation, where the influence function can vary with the choice of linear combination of instrumental variables. This example differs from previous results for NP2SLS by allowing orthogonality conditions in the choice of the instrumental variables set B.

To help relate Proposition 3 to prior work we consider a simple example of an object of interest for conditional moment restrictions.

EXAMPLE 6: *Linear Function of a Linear Structural Equation;* A relatively simple example has $m(W, \gamma) = v_m(Z)\gamma(Z)$ for a $v_m(Z)$ with $E[v_m(Z)^2] < \infty$, $\rho(w, \gamma) = y - \gamma(z)$, and Γ and $\mathcal{B}$ are unrestricted, so that the orthogonality condition of equation (4.15) is

$$Y = \gamma_0(Z) + \varepsilon, \ E[\varepsilon|X] = 0.$$

[28]

This is a linear NPIV equation. Assumptions 3 and 4 are satisfied with $v_m(Z)$ as given in this example and $v_\rho(W) = -1$. Then Assumption 5 is existence of $b_m(X)$ such that

$$v_m(Z) = E[b_m(X)|Z]. \tag{4.24}$$

Also $\mathcal{A}$ is the mean square closure of $E[\Delta(Z)|X]$ over all $\Delta(Z)$ with finite second moment and $\alpha_0(X)$ is the projection of $b_m(X)$ on $\mathcal{A}$. The adjustment term is then

$$\phi(W, \gamma_0, \alpha_0) = \alpha_0(X)\{Y - \gamma_0(Z)\}. \tag{4.25}$$

It is interesting to note that existence of a solution $b_m(X)$ to equation (4.24) is the necessary condition of Severini and Tripathi (2012) for existence of a root-n consistent estimator of $\theta_0 = E[v_m(Z)\gamma_0(Z)]$. This condition is restrictive in imposing that coefficients in a singular value expansion of $b_m(X)$ must decline at certain rates. This example shows the precise relationship of that necessary condition to the $\alpha_0(X)$ in the adjustment term. The $\alpha_0(X)$ is the projection of $b_m(X)$ on $\mathcal{A}$.

The formula for the adjustment term given here is related to a prior influence function formula given in Ai and Chen (2007, p. 40). In the notation here the Ai and Chen (2007) formula is

$$\phi(W, \gamma_0, \alpha_0) = E[v^*(Z)|X]\{Y - \gamma_0(Z)\}, \tag{4.26}$$

where $v^*(Z)$ is a Riesz representer in an extended Hilbert space described in Ai and Chen (2003, 2007). Equations (4.25) and (4.26) coincide for $\alpha_0(X) = E[v^*(Z)|X]$. Equation (4.25) is more explicit in giving the precise relationship between $\alpha_0(X)$ and the $b_m(X)$ of the Severini and Tripathi (2012) necessary condition. Also Proposition 3 allows orthogonality conditions that are more general than conditional moment restrictions. Interesting and useful Hilbert space characterizations of the adjustment term in Proposition 3 could be obtained as in Chen and Liao (2015) and/or Chen and Pouzo (2015) by extending their results for conditional moment restrictions to orthogonality conditions. The more explicit formula in Proposition 3 may prove useful for policy and sensitivity analysis and the construction of orthogonal moment functions.

The NPIV objective function in equation (4.16) can be modified to allow a weighted second moment matrix in the middle as in Ai and Chen (2003) where $\sum_{i=1}^{n} b^K(X_i)b^K(X_i)^T$ is replaced by $\sum_{i=1}^{n} \omega(X_i)b^K(X_i)b^K(X_i)^T$ for $\omega(X_i) > 0$. Such a modification with $\omega(X_i) = Var(\rho(W_i, \gamma)|X_i)$ would lead to improved asymptotic efficiency of $\hat{\theta}$ if γ were a finite dimensional parameter vector and $m(W, \gamma)$ did not depend on W. Proposition 3 can be modified in a straightforward way to allow for the presence of such a $\omega(X_i)$ by replacing $\rho(W, \gamma)$ with $\omega(X)^{-1}\rho(W, \gamma)$ and $E_\tau[\cdot]$ with the weighted expectation $E_\tau[\omega(X)(\cdot)]$, including in the projection π. Further details are beyond the scope of this paper.

[29]

5   Extensions and Conclusions

It is straightforward to extend the results we have given to objects that depend on multiple nonparametric estimators. As discussed in Newey (1994) such objects will have a separate adjustment term for each nonparametric estimator and the overall adjustment term will be the sum of the separate adjustment terms. Also, each separate adjustment term can be computed from varying one nonparametric estimator while holding the others fixed at their limit. It is also straightforward to extend the results to objects of interest that maximize objective functions other than that for GMM. This extension is described in Appendix B.

This paper gives explicit influence function formulae for first steps that satisfy exogenous or endogenous orthogonality conditions. It is shown how such formulae are useful for characterizing local policy effects of structural changes, quantifying sensitivity of semiparametric estimators, and constructing orthogonal moment functions. Those results are used to generalize the omitted variable bias formula for regression to obtain the local effect of misspecification on policies and estimators that depend on solutions to exogenous orthogonality conditions. This analysis is applied to a gasoline demand data set where we find no evidence that average equivalent variation bounds are sensitive to endogeneity.

6   References

Ackerberg, D., X. Chen, J. Hahn (2012): "A Practical Asymptotic Variance Estimator for Two-Step Semiparametric Estimators," *The Review of Economics and Statistics* 94, 481-498.

Ackerberg, D., X. Chen, J. Hahn and Z. Liao (2014): "Asymptotic Efficiency of Semiparametric Two-step GMM" *Review of Economic Studies* 81, 919-943.

Ai, C. and X. Chen (2003): "Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions," *Econometrica* 71, 1795–1843.

Ai, C. and X. Chen (2007): "Estimation of Possibly Misspecified Semiparametric Conditional Moment Restriction Models with Different Conditioning Variables," *Journal of Econometrics* 141, 5–43.

Ai, C. and X. Chen (2012): "The Semiparametric Efficiency Bound for Models of Sequential Moment Restrictions Containing Unknown Functions," *Journal of Econometrics* 170, 442–457.

Albrecht, J., A. Bjorklund, and S. Vroman (2003): "Is There a Glass Ceiling in Sweden?" *Journal of Labor Economics* 21, 145-177.

Amemiya, T. (1985): *Advanced Econometrics*, Cambridge, Harvard University Press.

Andrews, D. W. K. (1994): "Asymptotics for Semiparametric Econometric Models via Stochastic Equicontinuity," *Econometrica* 62, 43–72.

Andrews, D.W.K. (2011): "Examples of L2-Complete and Boundedly-Complete Distributions," Cowles Foundation Discussion Paper No. 1801, Yale University.

Andrews, I., M. Gentzkow, and J.M. Shapiro (2017): "Measuring the Sensitivity of Parameter Estimates to Estimation of Moments," *Quarterly Journal of Economics,* 132, 1553-1592.

Angrist, J.D. and A.B. Krueger (1991): "Does Compulsory School Attendance Affect Schooling and Earnings?" *The Quarterly Journal of Economics* 106, 979-1014.

Bajari, P., V. Chernozhukov, H. Hong, and D. Nekipelov (2009): "Nonparametric and Semiparametric Analysis of a Dynamic Discrete Game," working paper, Stanford.

Bajari, P., H. Hong, J. Krainer, and D. Nekipelov (2010): "Estimating Static Models of Strategic Interactions," *Journal of Business and Economic Statistics* 28, 469-482.

Bickel, P, C. Klaasen, Y. Ritov, and J. Wellner (1993): *Efficient and Adaptive Estimation for Semiparametric Models,* Washington, Johns Hopkins.

Blundell, R. and R. Matzkin (2014): "Control Functions In Nonseparable Simultaneous Equations Models," *Quantitative Economics* 5, 271–295.

Blundell, R., J. Horowitz, and M. Parey (2017): Nonparametric Estimation of a Nonseparable Demand Function Under the Slutzky Inequality Restriction," *The Review of Economics and Statistics* 99: 291–304.

Carone, M., A.R. Luedtke, M.J. van der Laan (2016): "Toward computerized efficient estimation in infinite dimensional models," arXiv: 1608.08717v1 31Aug2016.

Chaudhuri, P.K. Doksum, and A. Samarov (1997): "On Average Derivative Quantile Regression," *The Annals of Statistics* 25, 715-744.

Chen, X., O. Linton, and I. van Keilegom, (2003): "Estimation of Semiparametric Models When the Criterion Function is not Smooth," *Econometrica* 71, 1591–1608.

Chen, X., and Z. Liao (2015): "Sieve Semiparametric GMM Under Weak Dependence," *Journal of Econometrics* 189, 163-186.

Chen, X. and D. Pouzo (2015): "Sieve Wald and QLR Inferences on Semi/Nonparametric Conditional Moment Models," *Econometrica* 83, 1013-1079.

Chen, X. and A. Santos (2015): "Overidentification in Regular Models," *Econometrica* 86, 1771-1817.

Chernozhukov, V., I. Fernandez-Val, and B. Melly (2013): "Inference on Counterfactual Distributions," *Econometrica* 81, 2205-2268.

Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins (2018): "Double/Debiased Machine Learning for Treatment and Structural Parameters," *Econometrics Journal* 21, C1–C68.

Chernozhukov, V., J.-C. Escanciano, H. Ichimura, W. Newey, and J. Robins (2020): "Locally

Robust Semiparametric Estimation," https://arxiv.org/pdf/1608.00033.pdf.

Conley, T.J., C.B. Hansen, and P.E. Rossi (2012): "Plausibly Exogenous," *The Review of Economics and Statistics* 94, 260-272.

Darolles, S., Y. Fan, J. P. Florens, and E. Renault (2011): "Nonparametric Instrumental Regression," *Econometrica* 79, 1541–1565.

Firpo, S., N.M. Fortin, and T. Lemieux (2009): "Unconditional Quantile Regressions," *Econometrica* 77, 953–973.

Gentzkow, M., and J.M. Shapiro (2015): "Measuring the Sensitivity of Parameter Estimates to Sample Statistics," working paper.

Goldstein, L. and K. Messer (1992): "Optimal Plug-in Estimators for Nonparametric Functional Estimation," *Annals of Statistics* 20, 1306–1328.

Hahn, J., (1998): "On the Role of Propensity Score in Efficient Semiparametric Estimation of Average Treatment Effects" *Econometrica* 66, 315-332.

Hahn, J. and G. Ridder (2013): "The Asymptotic Variance of Semi-parametric Estimators with Generated Regressors," *Econometrica* 81, 315-340.

Hahn, J. and G. Ridder (2016): "Three-stage Semi-Parametric Inference: Control Variables and Differentiability," working paper.

Hansen, L.P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica* 50, 1029-1054.

Hampel, F. R. (1974): "The Influence Curve and Its Role In Robust Estimation," *Journal of the American Statistical Association* 69, 383–393.

Hausman, J.A. (1978): "Specification Tests in Econometrics," *Econometrica* 46, 1251-1271.

Hausman, J. A. and W. K. Newey (2016): "Individual Heterogeneity and Average Welfare," *Econometrica* 84, 1225-1248.

Hausman, J.A. and W.K. Newey (2017): "Nonparametric Welfare Analysis," *Annual Review of Economics* 9, 521–546.

Hirano, K., G.W. Imbens, G. Ridder (2003): "Efficient Estimation of Average Treatment Effects Using the Propensity Score," *Econometrica* 71, 1161-1189.

Huber, P. (1981): *Robust Statistics,* New York: John Wiley and Sons.

Ichimura, H. (1993): "Semiparametric Least Squares (SLS) and Weighted SLS Estimation of Single-index Models," *Journal of Econometrics* 58, 71-120.

Ichimura, H. and S. Lee (2010): "Characterization of the asymptotic distribution of semiparametric M-estimators," *Journal of Econometrics* 159, 252–266.

Imbens, G.W. (1997): "One-Step Estimators for Over-Identified Generalized Method of Moments Models," *The Review of Economic Studies* 64, 359–383.

Klein, R.W. and R.H. Spady (1993): "An Efficient Semiparametric Estimator for Binary Response Models," *Econometrica* 61, 387-421.

Koenker, R. and G. Bassett (1978): "Regression Quantiles," *Econometrica* 46, 33-50.

Kress, R. (1989): *Linear Integral Equations,* New York: Springer-Verlag.

Lebesgue, H. (1904): *Leçons sur l'Intégration et la Recherche des Fonctions Primitives*, Paris: Gauthier-Villars.

Luedtke, A.R., M. Carone, M.J. van der Laan (2015): "Toward computerized efficient estimation in infinite dimensional models," arXiv: 1608.08717.

Machada, J.A.F and J. Mata (2005): "Counterfactual Decomposition of Changes in Wage Distributions Using Quantile Regression," *Journal of Applied Econometrics* 20, 445-465.

Melly, B. (2005): "Decomposition of Differences in Distribution Using Quantile Regression," *Labour Economics* 12, 577-590.

Mammen, E., C. Rothe, M. Schienle (2012): "Nonparametric Regression with Nonparametrically Generated Covariates," *Annals of Statistics* 40, 1132–1170.

Mukhin, Y. (2019): "Sensitivity of Regular Estimators," arXiv 1805.08883.

Nagar, A.L. (1959): "The Bias and Moment Matrix of the General k-Class Estimators of the Parameters in Simultaneous Equations," *Econometrica* 27, 575-595.

Newey, W.K. (1991): "Uniform Convergence in Probability and Stochastic Equicontinuity," *Econometrica* 59, 1161-1167.

Newey, W.K. (1994): "The Asymptotic Variance of Semiparametric Estimators," *Econometrica* 62, 1349–1382.

Newey, W.K. and J.L. Powell (1987): "Asymmetric Least Squares Estimation and Testing," *Econometrica* 55, 819-847.

Newey, W.K., and J.L. Powell (1989): "Instrumental Variable Estimation of Nonparametric Models," presented at Econometric Society winter meetings, 1988.

Newey, W.K., and J.L. Powell (2003): "Instrumental Variable Estimation of Nonparametric Models," *Econometrica* 71, 1565-1578.

Powell, J.L., J.H. Stock, and T.M. Stoker (1989): "Semiparametric Estimation of Index Coefficients," *Econometrica* 57, 1403-1430.

Pratt, J.W. (1981): "Concavity of the Log Likelihood," *Journal of the American Statistical Association* 76, 103-106.

Ritov, Y. and P.J. Bickel (1990): "Achieving Information Bounds in Non and Semiparametric Models," *Annals of Statistics* 18, 925-938.

Robinson, P.M. (1988): "Root-N-Consistent Semiparametric Regression," *Econometrica* 56, 931-954.

Severini, T. and G. Tripathi (2012): "Efficiency Bounds for Estimating Linear Functionals of Nonparametric Regression Models with Endogenous Regressors," *Journal of Econometrics* 170, 491-498.

Sherman, R. (1993): "The Limiting Distribution of the Maximum Rank Correlation Estimator," *Econometrica* 61, 123-137.

Stein, C. (1956): "Efficient Nonparametric Testing and Estimation," in *Proceedings of the Third Berkeley Symposium in Mathematical Statistics and Probability, vol. 1,* 187-196, Berkeley, University of California Press.

Van der Vaart, A.W. (1991): "On Differentiable Functionals," *Annals of Statistics* 19, 178–204.

Van der Vaart, A. W. and J. A. Wellner (1996): *Weak Convergence and Empirical Processes*, New York: Springer-Verlag.

Von Mises, R. (1947): "On the Asymptotic Distribution of Differentiable Statistical Functions," *Annals of Mathematical Statistics* 18, 309-348.

Wheeden, R.L. and A. Zygmund (1977): *Measure and Integral – An introduction to Real Analysis*, Marcel Dekker.

7 Online Appendices

The following three appendices are for online.

7.1 A: Validity of the Influence Function Calculation

In this Appendix we show validity of Steps I and II of the influence function calculation. Step I requires differentiability of $\theta(F_\tau)$ and the formula in equation (2.3). Step II) requires that evaluating the derivative at a point mass gives the influence function. We justify Step II) as a limit as H approaches a point mass similarly to Lebesgue differentiation from analysis. Lebesgue differentiation shows that the limit of an integral of a function over an interval divided by the length of the interval converges almost surely to the value of the function at a point as the interval collapses on that point. We give regularity conditions and classes of continuous, smooth probability distributions where expectation of the influence function converges to its value at a point as the probability distribution collapses on the point.

The fundamental starting point for the influence function calculation is that the estimator is asymptotically linear with an influence function, i.e. that it satisfies equation (2.1). We take a modern, high level approach to regularity conditions in assuming that the estimator is locally regular for a set alternative distributions H that can approximate a point mass.

[34]

DEFINITION A1: $\hat{\theta}$ is locally regular for F_τ if there is a fixed random variable Y such that for any $\tau_n = O(1/\sqrt{n})$ and $W_1, ..., W_n$ i.i.d. with distribution F_{τ_n},

$$\sqrt{n}[\hat{\theta} - \theta(F_{\tau_n})] \xrightarrow{d} Y.$$

This local regularity condition is familiar from the efficient estimation literature. Local regularity of $\hat{\theta}$ is not a primitive condition but it is plausible when F_0 satisfies conditions for existence of $\theta(F)$ and H is well behaved relative to F_0. For example F_0 could satisfy regularity conditions like some random variables being continuously distributed and expectations of certain functions existing and H could be a uniformly bounded, very smooth deviation from F_0. In such settings it is plausible many estimators $\hat{\theta}$ would be locally regular. We construct such H in this Appendix so that local regularity is plausibly satisfied for many semiparametric estimators $\hat{\theta}$.

We consider a sequence $(H_w^j)_{j=1}^\infty$ taking the form

$$H_w^j(\tilde{w}) = E[1(W \leq \tilde{w})\delta_w^j(W)], \tag{7.27}$$

where for each j the random variable $\delta_w^j(W)$ is bounded with $E[\delta_w^j(W)] = 1$. In $H_w^j(\tilde{w})$ the variable $\tilde{w}$ represents a possible value of the random variable W. As we will discuss this $H_w^j(\tilde{w})$ will have the needed properties when $\delta_w^j(W)$ is chosen appropriately. In particular the support of $H_w^j(\tilde{w})$ will approach $\{w\}$ as that the support of $\delta_w^j(w)$ does. Throughout we will assume that w is a vector of real numbers of fixed dimension r. We impose the following properties:

ASSUMPTION A1: F_0 is absolutely continuous with respect to a measure μ on $\mathbb{R}^r$ with pdf $f_0(w)$, $\delta_w^j(W)$ is not constant, bounded, and $E[\delta_w^j(W)] = 1$.

By $\delta_w^j(W)$ bounded $F_\tau^j = (1 - \tau)F_0 + \tau H_w^j$ will be a CDF for small enough τ with pdf with respect to μ given by

$$f_\tau(\tilde{w}) = f_0(w)[1 - \tau + \tau\delta_w^j(\tilde{w})] = f_0(\tilde{w})[1 + \tau S(\tilde{w})], S(\tilde{w}) = \delta_w^j(\tilde{w}) - 1, \tag{7.28}$$

where we suppress the j superscript and w subscript on $f_\tau(\tilde{w})$ and $S(\tilde{w})$ for notational convenience. Note that by $S(\tilde{w})$ bounded there is C such that for small enough τ,

$$(1 - \tau)f_0/C \leq f_\tau \leq Cf_0, \tag{7.29}$$

so that f_τ and f_0 will be absolutely continuous with respect to each other. Thus, variables that are continuously distributed under F_0 will also be continuously distributed under F_τ^j. Also objects that have expectation close to zero for F_0 will also have expectation close to zero under F_τ^j and vice versa. If $\theta(F)$ being well defined depends on existence of derivatives of the pdf

for F then that restriction can be imposed by choosing $\delta_w^j(\tilde{w})$ so its derivatives exist. In these ways we can choose $\delta_w^j(w)$ so that $f_\tau(\tilde{w})$ satisfies the restrictions needed for $\theta(F_\tau^j)$ to be well defined.

We assume that the sequence $(\delta_w^j)_{j=1}^\infty$ satisfies a condition leading to

$$\lim_{j\longrightarrow\infty} \int \psi(\tilde{w}) H_w^j(d\tilde{w}) \longrightarrow \psi(w), \qquad (7.30)$$

thus justifying Step II of the influence function calculation. Define a function $a(\tilde{w})$ is μ to be almost surely continuous at w if for any $\varepsilon > 0$ there is a neighborhood N of w and a subset N_μ of N such that $\mu(N_\mu) = \mu(N)$ and $|a(\tilde{w}) - a(w)| < \varepsilon$ for all $\tilde{w} \in N_\mu$.

ASSUMPTION A2: *If $a(\tilde{w})$ is μ almost surely continuous at w and $E[a(W)^2] < \infty$ then $\lim_{j\longrightarrow\infty} E[a(W)\delta_w^j(W)] = a(w)$.*

This Assumption will be sufficient for equation (7.30). There are a variety of ways that $\delta_w^j(W)$ can be chosen so that Assumption 2 will be satisfied. The basic idea is to consder w where $f_0(\tilde{w})$ is bounded away from zero on a neighborhood of w and choose $\delta_w^j(\tilde{w}) = g_w^j(\tilde{w})/f_0(\tilde{w})$ where $g_w^j(\tilde{w})$ is a bounded pdf and the support of $g_w^j(\tilde{w})$ to converge to $\{w\}$. In the Appendix D we will choose $g_w^j(\tilde{w})$ in a way that is helpful for endogenous orthogonality conditions. Another choice of $g_w^j(\tilde{w})$ that will lead to equation (7.30) in many cases can be based on a nonnegative kernel $K(u)$ with bounded support S, as in the following result.

LEMMA A1: *If i) $K(u) \geq 0$, $\int K(u)du = 1$, and $K(u)$ has bounded support S; ii) there is a neighborhood N of z and $C > 0$ such that $f_0(w) \geq C$ almost surely μ for $w \in N$; iii) $\mu(w + \sigma S) > 0$ for all $\sigma > 0$; then for any $(\sigma(j))_{j=1}^\infty$ with $\sigma(j) > 0$, $\sigma(j) \longrightarrow 0$, and $w + \sigma(j)S \subseteq N$ for all $\sigma(j)$, Assumptions 1 and 2 are satisfied for*

$$\delta_w^j(W) = f_0(W)^{-1}[\int 1(w + \sigma(j)S)\sigma(j)^{-r}K\left(\frac{\tilde{w} - w}{\sigma(j)}\right)\mu(d\tilde{w})]^{-1}\sigma(j)^{-r}K\left(\frac{W - w}{\sigma(j)}\right).$$

Proof: Note that

$$\int 1(W + \sigma(j)S)\sigma(j)^{-r}K\left(\frac{\tilde{w} - W}{\sigma(j)}\right)\mu(d\tilde{w}) > 0$$

by iii). Also, $K\left((W - w)/\sigma(j)\right)$ is nonzero only on a subset of N so that $\delta_w^j(W)$ is bounded by i and ii). In addition $E[\delta_w^j(W)] = 1$ by construction.

Suppose $a(W)$ has finite second moment and is continuous at w a.s. μ. Then for any $\varepsilon > 0$ there is j_ε large enough such that for $j \geq j_\varepsilon$,

$$a(w) - \varepsilon \leq a(W) \leq a(w) + \varepsilon$$

[36]

a.s. μ for $W \in w + \sigma(j)S$. Since $\delta_w^j(W)$ is nonnegative and nonzero only on $W \in w + \sigma(j)S$ we have

$$a(w) - \varepsilon = E[\{a(w) - \varepsilon\}\delta_w^j(W)] \leq E[a(W)\delta_w^j(W)] \leq E[\{a(w) + \varepsilon\}\delta_w^j(W)] = a(w) + \varepsilon,$$

for all $j \geq j_\varepsilon$. The conclusion follows by ε being any positive number. $Q.E.D.$

The choice of $\delta_w^j(W)$ in Lemma A1 is simply a device to help the limit of the Gateaux derivative exist under as general conditions as possible. The limit, and hence the influence function, does not depend on the kernel. Also, we could replace the continuity of $a(\tilde{w})$ at w in Assumption 1 with other conditions that are sufficient for equation (7.30) on a set of w with probability one under F_0. Equation (7.30) is analogous to the Lebesgue differentiation theorem that is known to hold under quite general conditions on $a(\tilde{w})$. For example, for the $\delta_w^j(w)$ of Lemma A1 equation (7.30) can be shown to hold for any measurable $a(\tilde{w})$ if μ is the sum of Lebesgue measure and a measure with a finite number of atoms. We use the continuity condition of Assumption A1 because it is relatively simple to state and because many influence functions will be μ almost sure continuous on a set of w that has probability one.

The next result shows that the influence function formula (??) is valid for H_w^j as specified in equation (7.27).

THEOREM A2: *If Assumptions A1 and A2 are satisfied, $\hat{\theta}$ is asymptotically linear with influence function $\psi(\tilde{w})$, $\hat{\theta}$ is locally regular for $F_\tau^j(\tilde{w}) = (1 - \tau)F_0(\tilde{w}) + \tau H_w^j(\tilde{w})$ for each integer j and $H_w^j(\tilde{w}) = E[1(W \leq \tilde{w})\delta^j(W)]$, and $\psi(\tilde{w})$ is μ almost surely continuous at w, then $d\theta(F_\tau^j)/d\tau$ exists, $d\theta(F_\tau^j)/d\tau = \int \psi(\tilde{w})H_w^j(d\tilde{w})$, and equation (7.30) is satisfied.*

Proof: By $S(\tilde{w}) = \delta_w^j(\tilde{w}) - 1$ bounded there is an open set T containing zero such that for all $\tau \in T$, $1 + \tau S(\tilde{w})$ is positive, bounded away from zero, and $f_\tau(\tilde{w})^{1/2} = f_0(\tilde{w})^{1/2}[1 + \tau S(\tilde{w})]^{1/2}$ is continuously differentiable in τ with

$$s_\tau(\tilde{w}) = \frac{d}{d\tau}f_0(\tilde{w})^{1/2}[1 + \tau S(\tilde{w})]^{1/2} = \frac{1}{2}\frac{f_0(\tilde{w})^{1/2}S(\tilde{w})}{[1 + \tau S(\tilde{w})]^{1/2}} \leq Cf_0(\tilde{w})^{1/2}S(\tilde{w}).$$

By $S(\tilde{w})$ bounded, $\int \left[Cf_0(\tilde{w})^{1/2}S(\tilde{w})\right]^2 d\mu < \infty$. Then by the dominated convergence theorem $f_0(\tilde{w})^{1/2}[1 + \tau S(\tilde{w})]^{1/2}$ is mean-square differentiable and $I(\tau) = \int s_\tau(\tilde{w})^2 d\mu$ is continuous in τ on a neighborhood of zero. By Assumption 1 $S(W)$ is not zero so that $I(\tau) > 0$. Then by Theorem 7.2 and Example 6.5 of Van der Vaart (1998) it follows that for any $\tau_n = O(1/\sqrt{n})$ a vector of n observations $(W_1, ..., W_n)$ that is i.i.d. with pdf $f_{\tau_n}(\tilde{w})$ is contiguous to $(W_1, ..., W_n)$ that is i.i.d. with pdf $f_0(\tilde{w})$. Therefore,

$$\sqrt{n}(\hat{\theta} - \theta_0) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \psi(W_i) + o_p(1)$$

[37]

holds when $(W_1, ..., W_n)$ are i.i.d. with pdf $f_{\tau_n}(\tilde{w})$.

Next define $\mu_w^j = E[\psi(W)S(W)] = E[\psi(W)\delta_w^j(W)]$. Then by $E[\psi(W)] = 0$,

$$E_\tau[\psi(W)] = \tau\mu_w^j.$$

Suppose $(W_1, ..., W_n)$ are i.i.d. with pdf $f_{\tau_n}(\tilde{w})$. Let $\theta(\tau) = \theta((1-\tau)F_0 + \tau G_w^j)$, $\theta_n = \theta(\tau_n)$, and $\breve{\psi}_n(W) = \psi(W) - \tau_n\mu_w^j$. Adding and subtracting terms,

$$\sqrt{n}\left(\hat{\theta} - \theta_n\right) = \sqrt{n}(\hat{\theta} - \theta_0) - \sqrt{n}(\theta_n - \theta_0) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \psi(W_i) + o_p(1) - \sqrt{n}(\theta_n - \theta_0)$$

$$= \frac{1}{\sqrt{n}}\sum_{i=1}^n \breve{\psi}_n(W_i) + o_p(1) + \sqrt{n}\tau_n\mu_w^j - \sqrt{n}(\theta_n - \theta_0).$$

Note that $E_{\tau_n}[\breve{\psi}_n(W)] = 0$. Also, by τ_n bounded,

$$E_\tau[1(\left\|\breve{\psi}_n(W)\right\| \geq M)\left\|\breve{\psi}_n(W)\right\|^2] \leq CE[1(\left\|\breve{\psi}_n(W)\right\| \geq M)\left\|\breve{\psi}_n(W)\right\|^2]$$

$$\leq CE[1(\left\|\breve{\psi}_n(W)\right\| \geq M)(\|\psi(W)\|^2 + C)]$$

$$\leq CE[1(\|\psi(W)\| \geq M - C)(\|\psi(W)\|^2 + C)] \longrightarrow 0,$$

as $M \longrightarrow \infty$, so the Lindbergh-Feller condition for a central limit theorem is satisfied. Furthermore, it follows by similar calculations that $E_{\tau_n}[\breve{\psi}_n(W)\breve{\psi}_n(W)^T] \longrightarrow V$. Therefore, by the Lindbergh-Feller central limit theorem, $\sum_{i=1}^n \breve{\psi}_n(W_i)/\sqrt{n} \xrightarrow{d} N(0, V)$. By local regularity $\sqrt{n}(\hat{\theta} - \theta_n) \xrightarrow{d} N(0, V)$ implying that

$$\sqrt{n}\tau_n\mu_w^j - \sqrt{n}(\theta_n - \theta_0) \longrightarrow 0. \tag{7.31}$$

Next, we follow the proof of Theorem 2.1 of Van der Vaart (1991). The above argument shows that local regularity implies that eq. (7.31) holds for all $\tau_n = O(1/\sqrt{n})$. Consider any sequence $r_m \longrightarrow 0$. Let n_m be the subsequence such that

$$(1 + n_m)^{-1/2} < r_m \leq n_m^{-1/2}.$$

Let $\tau_n = r_m$ for $n = n_m$ and $\tau_n = n^{-1/2}$ for $n \notin \{n_1, n_2, ...\}$. By construction, $\tau_n = O(1/\sqrt{n})$, so that eq (7.31) holds. Therefore it also holds along the subsequence n_m, so that

$$\sqrt{n_m}r_m\left\{\mu_z^j - \frac{\theta(r_m) - \theta_0}{r_m}\right\} = \sqrt{n_m}r_m\mu_z^j - \sqrt{n_m}[\theta(r_m) - \theta_0] \longrightarrow 0.$$

By construction $\sqrt{n_m}r_m$ is bounded away from zero, so that $\mu_z^h - [\theta(r_m) - \theta_0]/r_m \longrightarrow 0$. Since r_m is any sequence converging to zero it follows that $\theta(\tau)$ is differentiable at $\tau = 0$ with derivative μ_z^j. The conclusion then follows by Assumption 2. Q.E.D.

[38]

Let H_w^∞ be the CDF with $\Pr(W = w) = 1$. Theorem A2 gives sufficient conditions for equation (7.30) which is

$$\psi(w) = \int \psi(\tilde{w}) H_w^\infty(d\tilde{w}) = \lim_{H_w^j \longrightarrow H_w^\infty} \int \psi(\tilde{w}) H_w^j(d\tilde{w}),$$

where the first equality holds by definition of H_w^∞. The second equality states that $\psi(w)$ is the Lebesgue derivative of $\int \psi(\tilde{w}) H(d\tilde{w})$ based on the regularity conditions of Assumptions A1 and A2 and the sequences of functions detailed there. This Lebesgue differentiation conclusion justifies Step II of the Gateaux derivative calculation as simply evaluating the Lebesgue derivative at a point. This evaluation with be valid with probability one under Assumptions A1 and A2.

We emphasize that the purpose of Theorem 2 is quite different than the results of Bickel, Klaasen, Ritov, and Wellner (1993), Van der Vaart (1991) and other important contributions to the semiparametric efficiency literature. Here $\theta(F)$ is not a parameter of some semiparametric model. Instead $\theta(F)$ is associated with an *estimator* $\hat{\theta}$, being the probability limit of that estimator when F is a distribution that is unrestricted except for regularity conditions, as in Newey (1994). Our goal is to use $\theta(F)$ to calculate the influence function of $\hat{\theta}$ under the assumption that $\hat{\theta}$ is asymptotically linear. The purpose of Theorem A2 is to justify Steps I and II as a way to do that calculation. In contrast, the goal of the semiparametric efficiency literature is to find the efficient influence function for a parameter of interest when F belongs to a family of distributions.

To highlight this contrast, note that the Gateaux derivative limit calculation can be applied to obtain the influence function under misspecification while efficient influence function calculations generally impose correct specification. Indeed, the definition of $\theta(F)$ *requires* that misspecification be allowed for, because $\theta(F)$ is limit of the estimator θ under all distributions F that are unrestricted except for regularity condition. Of course correct specification may lead to simplifications in the form of the influence function. Such simplifications will be incorporated automatically when the Gateaux derivative limit is taken at an F_0 that satisfies model restrictions.

Theorem 2 is like Van der Vaart (1991, Theorem 2.1) in having differentiability of $\theta(F_\tau)$ as a conclusion. It differs in restricting the paths to have the form $(1 - \tau)F_0 + \tau H_w^j$. Such a restriction on the paths actually weakens the local regularity hypothesis because θ only has to be locally regular for a particular kind of path rather than the general class of paths in Van der Vaart (1991). We note that this result allows for the distribution of W to have discrete components because the dominating measure μ may have atoms.

The weak nature of the local regularity condition highlights the strength of the asymptotic linearity hypothesis. Primitive conditions for asymptotic linearity can be quite strong and com-

plicated. For example, it is known that asymptotic linearity of estimators with a nonparametric first step often requires some degree of smoothness in the functions being estimated, see Ritov and Bickel (1990). Our purpose here is to bypass those conditions in order to justify the Gateaux derivative formula for the influence function. The formula for the influence function can then be used in all the important ways outlined in Section 2.

It is also common to bypass regularity conditions when calculating the influence function or asymptotic variance of parametric estimators. There are well known formulae that allow us to do this, such as Hansen (1982) for GMM estimators. The Gateaux derivative limit provides such a formula for semiparametric estimators. It provides an influence function formula that will be valid "under sufficient regularity conditions" analogously to the GMM formula for parametric estimators.

7.2   Appendix B: The Influence Function of Semiparametric M Estimators

In this Appendix we give the general structure of the influence function for a semiparametric M-estimator and show that the adjustment term is zero for any first step that maximizes the same objective function as the does the parameter of interest. These results generalize A maximization (M) estimator satisfies

$$\hat{\theta} = \arg\max_{\theta \in B} \hat{Q}(\theta),$$

for a function $\hat{Q}(\theta)$ that depends on the data and parameters. M estimators have long been studied. A more general type that is useful when $\hat{Q}(\theta)$ is not continuous has $\hat{Q}(\theta) \geq \sup_{\theta \in B} \hat{Q}(\theta) - \hat{R}$, where the remainder $\hat{R}$ is small in large samples. The plin $\theta(F)$ of $\hat{\theta}$ will be the maximizer of the probability limit of $\hat{Q}(\theta)$ under standard regularity conditions. Thus, the influence function will depend only on the limit of the objective function and so is not affected by whether $\hat{\theta}$ is an approximate or exact maximizer of $\hat{Q}(\theta)$. The way we give of calculating the influence function will work for many estimators of this form, including those maximizing U-processes as considered by Sherman (1993).

We can use the Gateaux derivative to characterize the influence function for semiparametric M-estimators. Let $Q_\tau(\theta)$ denote the limit of the objective function $\hat{Q}(\theta)$ when the CDF of W_i is F_τ. Then under standard regularity conditions the plim of $\hat{\theta}$ is

$$\theta_\tau = \arg\max_{\theta \in \Theta} Q_\tau(\theta).$$

Suppose that $Q_\tau(\theta)$ is twice continuously differentiable in θ and θ_τ is in the interior of the parameter set. Then θ_τ satisfies the first order conditions $dQ_\tau(\theta_\tau)/d\theta = 0$. By the implicit

function theorem, for $\Lambda = \partial^2 Q(\theta_0)/\partial\theta\partial\theta'$ we have

$$\frac{d\theta_\tau}{d\tau} = -\Lambda^{-1} \frac{\partial^2 Q_\tau(\theta_0)}{\partial\tau\partial\theta}\bigg|_{\tau=0} = -\Lambda^{-1} \frac{\partial}{\partial\tau}\left\{\frac{\partial Q_\tau(\theta_0)}{\partial\theta}\right\}.$$

Comparing this equation with equation (2.3) we see that the influence function $\psi(w)$ of a semiparametric M estimator can be calculated by evaluating the derivative with respect to τ of $dQ_\tau(\theta_0)/d\theta$ at the distribution H_w^∞ with $W = w$ and premultiplying by $-\Lambda^{-1}$. For $\xi(W)$ such that $dQ_\tau(\theta_0)/d\theta = \int \xi(w)H(dw)$ the influence function of $\hat\theta$ will be

$$\psi(W) = -\Lambda^{-1}\xi(W).$$

This formula generalizes that of Newey (1994) for semiparametric GMM to M-estimation.

For M-estimators, certain nonparametric components of $\hat Q(\theta)$ can be ignored in deriving the influence function. The ignorable components are those that have been "concentrated out," meaning they have a limit that maximizes the limit of $\hat Q(\theta)$. In such cases the dependence of these functions on θ captures the whole asymptotic effect of their estimation. To show this result, suppose that there is a function γ that depends on θ and possibly other variables and a function $\tilde Q_\tau(\theta, \gamma, F)$ such that $Q_\tau(\theta) = \tilde Q_\tau(\theta, \gamma_\tau)$ where

$$\gamma_\tau = \arg\max_\gamma \tilde Q_\tau(\theta, \gamma).$$

Here $\tilde Q_\tau(\theta, \gamma_\tau)$ is the limit of $\hat Q(\theta)$ and γ_τ the limit of a nonparametric estimator on which $\hat Q(\theta)$ depends, when W has CDF F_τ. Since γ_τ maximizes over all γ it must maximize over $\tilde\tau$ as the function $\gamma_{\tilde\tau}$ varies. The first order condition for maximization over $\tilde\tau$ is

$$\frac{d\tilde Q_\tau(\theta, \gamma_{\tilde\tau})}{d\tilde\tau}\bigg|_{\tilde\tau=\tau} = 0.$$

This equation holds identically in θ, so that we can differentiate both sides of the equality with respect to θ, evaluate at $\theta = \theta_0$ and $\tau = 0$, and interchange the order of differentiation to obtain

$$\frac{\partial^2 \tilde Q(\theta_0, \gamma_\tau)}{\partial\tau\partial\theta} = 0.$$

Then it follows by the chain rule that

$$\frac{\partial^2 \tilde Q_\tau(\theta_0, \gamma_\tau)}{\partial\tau\partial\theta} = \frac{\partial^2 \tilde Q_\tau(\theta_0, \gamma_0)}{\partial\tau\partial\theta} + \frac{\partial^2 \tilde Q(\theta_0, \gamma_\tau)}{\partial\tau\partial\theta} = \frac{\partial^2 \tilde Q_\tau(\theta_0, \gamma_0)}{\partial\tau\partial\theta}. \tag{7.32}$$

That is, the influence function can be obtained by treating the limit γ_τ as if it were equal to the true value γ_0.

Equation (7.32) generalizes Proposition 2 of Newey (1994) and Theorem 3.4 of Ichimura and Lee (2010) to objective functions that are not necessarily a sample average of a function of

θ and γ. There are many important estimators included in this generalization. One of those is NPIV where the residual includes both parametric and nonparametric components. The result implies that estimation of the function of the nonparametric component γ can be ignored in calculating the influence function of θ. Another interesting estimator is partially linear regression with generated regressors. There the estimation of the nonparametric component can also be ignored in deriving the influence function, just as in Robinson (1988), though the presence of generated regressors will often affect the influence function, as in Hahn and Ridder (2013, 2016) and Mammen, Rothe, and Schienle (2012).

7.3 Endogenous Orthogonality Conditions with Misspecification.

In this Appendix we derive the adjustment term for endogenous orthgonality conditions under overidentification and misspecification where

$$\bar{\pi}(X) = \pi(\rho(W, \gamma_0)|X) \neq 0.$$

To derive the influence function note that the first order conditions from equation (4.19) give

$$
\begin{aligned}
0 &= E_\tau[\pi_\tau(\rho(W, \gamma_\tau)|X)\pi_\tau(v_{\rho\tau}(W)\Delta(Z)|X)] \\
&= E_\tau[\pi_\tau(\rho(W, \gamma_\tau)|X)v_{\rho\tau}(W)\Delta(Z)] \text{ for all } \Delta \in \Gamma,
\end{aligned}
$$

identically in τ. Define $\alpha(X, \Delta) := \pi(v_\rho(W)\Delta(Z)|X)$ for $\Delta \in \Gamma$. Differentiating the previous identity with respect to τ gives for all $\Delta \in \Gamma$

$$0 = \frac{\partial}{\partial \tau}E[\pi_\tau(\rho(W, \gamma_\tau)|X)\alpha(X, \Delta)] + \int \phi_1(w, \Delta)H(dw) + T_{v_\rho}(\Delta),$$

$$\phi_1(w, \Delta) \quad : \quad = \bar{\pi}(X)v_\rho(W)\Delta(Z) - E[\bar{\pi}(X)v_\rho(W)\Delta(Z)], \ T_{v_\rho}(\Delta) := \frac{\partial}{\partial \tau}E[\bar{\pi}(X)v_{\rho\tau}(W)\Delta(W)].$$

where $v_\rho(W) = v_{\rho 0}(W)$. Solving gives

$$\frac{\partial}{\partial \tau}E[\pi_\tau(\rho(W, \gamma_\tau)|X)\alpha(X, \Delta)] = -\int \phi_1(w, \Delta)H(dw) - T_{v_\rho}(\Delta) \tag{7.33}$$

for all $\Delta \in \Gamma$.

Next we use the orthogonality condition (4.17) for the projection. Because $\mathcal{A}$ is a subset of $\mathcal{B}$ it follows that

$$E_\tau[\rho(W, \gamma_\tau)\alpha(X, \Delta)] = E_\tau[\pi_\tau(\rho(W, \gamma_\tau)|X)\alpha(X, \Delta)] \text{ for all } \Delta \in \Gamma,$$

identically in τ. Differentiating both sides of this identify with respect to τ and applying the

chain rule gives

$$
\begin{aligned}
\frac{\partial}{\partial\tau}E_\tau[\rho(W,\gamma_\tau)\alpha(X,\Delta)] &= \frac{\partial}{\partial\tau}E_\tau[\bar{\pi}(X)\alpha(X,\Delta)] + \frac{\partial}{\partial\tau}E[\pi_\tau(\rho(W,\gamma_\tau)|X)\alpha(X,\Delta)] \\
&= \frac{\partial}{\partial\tau}E_\tau[\bar{\pi}(X)\alpha(X,\Delta)] - \int \phi_1(w,\Delta)H(dw) - T_{v_\rho}(\Delta) \\
&= -\int \phi_\Gamma(w,\Delta)H(dw) - T_{v_\rho}(\Delta), \\
\phi_\Gamma(w,\Delta) &= \bar{\pi}(X)\{v_\rho(X)\Delta(Z) - \alpha(X,\Delta)\},
\end{aligned}
$$

for all $\Delta \in \Gamma$ where the second equality follows by equation (7.33) and the third equality equality follows by $E[\bar{\pi}(X)\alpha(X,\Delta)] = E[\bar{\pi}(X)v_\rho(W)\Delta(Z)]$. Applying the chain rule to the left-and side and solving then gives

$$
\begin{aligned}
-\frac{\partial}{\partial\tau}E[\rho(W,\gamma_\tau)\alpha(X,\Delta)] &= \frac{\partial}{\partial\tau}E_\tau[\rho(W,\gamma_0)\alpha(X,\Delta)] + \int \phi_\Gamma(w,\Delta)H(dw) + T_{v_\rho}(\Delta) \quad (7.34) \\
&= \int \{\rho(w,\gamma_0)\alpha(x,\Delta) + \phi_\Gamma(w,\Delta)\}H(dw) + T_{v_\rho}(\Delta),
\end{aligned}
$$

for all $\Delta \in \Gamma$, where the last equality follows by the first order condition at $\tau = 0$ that implies $E[\rho(W,\gamma_0)\alpha(X,\Delta)] = 0$ for all Δ. Suppose that there exists b_m such that the projection of b_m on $\mathcal{A}$ is $\alpha(X,\Delta_m)$ for some $\Delta_m \in \Gamma$ and

$$
\Pi(v_m(Z)|Z) = -\Pi(v_\rho(W)b_m(X)|Z).
$$

Then by $\gamma_\tau(Z) \in \Gamma$,

$$
\begin{aligned}
E[v_m(Z)\gamma_\tau(Z)] &= E[\Pi(v_m(Z)|Z)\gamma_\tau(Z)] = -E[\Pi(v_\rho(W)b_m(X)|Z)\gamma_\tau(Z)] \quad (7.35) \\
&= -E[v_\rho(W)b_m(X)\gamma_\tau(Z)] = -E[b_m(X)\pi(v_\rho(W)\gamma_\tau(Z)|X)] \\
&= -E[\alpha(X,\Delta_m)\pi(v_\rho(W)\gamma_\tau(Z)|X)] = -E[\alpha(X,\Delta_m)v_\rho(W)\gamma_\tau(Z)].
\end{aligned}
$$

Then differentiating gives

$$
\begin{aligned}
\frac{\partial}{\partial\tau}E[m(W,\gamma_\tau)] &= \frac{\partial}{\partial\tau}E[v_m(Z)\gamma_\tau(Z)] = -\frac{\partial}{\partial\tau}E[\alpha(X,\Delta_m)v_\rho(W)\gamma_\tau(Z)] \\
&= -\frac{\partial}{\partial\tau}E[\alpha(X,\Delta_m)\rho(W,\gamma_\tau)] \\
&= \int \{\rho(w,\gamma_0)\alpha(x,\Delta_m) + \phi_\Gamma(w,\Delta_m)\}H(dw) + T_{v_\rho}(\Delta_m)
\end{aligned}
$$

where the first equality follows by Assumption 3, the second equality by equation (7.35), the third equality by Assumption 4, and the fourth equality by equation (7.34). Combining this last equation with the conditions on which it depends gives the following result:

PROPOSITION C1: *If i) Assumptions 3-4 are satisfied; ii) there exists $b_m(X)$ and $\Delta_m \in \Gamma$ such that $\alpha(X,\Delta_m)$ is the projection of $b_m(X)$ on $\mathcal{A}$ and $\Pi(v_m(Z)|Z) = \Pi(v_\rho(W)b_m(X)|Z)$; and*

[43]

iii) there is $\phi_\rho(w)$ such that $\partial E[\bar{\pi}(X)v_{\rho\tau}(W)\Delta_m(W)]/\partial\tau = \int \phi_\rho(w)H(dw)$ then the adjustment term is

$$\phi(w, \gamma, \alpha) = \alpha(x, \Delta_m)\rho(w, \gamma) + \bar{\pi}(x)\{v_\rho(x)\Delta_m(z) - \pi(v_\rho(X)\Delta_m(Z)|X = x)\} + \phi_\rho(w).$$

This expression for the influence function contains the term $\phi_\rho(w)$ which is the influence function of $E[\bar{\pi}(X)v_{\rho\tau}(W)\Delta_m(Z)]$. This $\phi_\rho(w)$ need not exist. In particular for quantile orthogonality conditions where $v_{\rho\tau}(W)$ depends on the conditional pdf of Y given Z and X evaluated at the point $Y = \gamma_0(Z)$ it seems that this $\phi_\rho(w)$ generally does not exist. In that case the NPIV estimator will not root-n consistent under misspecification. This problem does not appear to be present for expectiles, where $E[\bar{\pi}(X)v_{\rho\tau}(W)\Delta_m(Z)]$ can be shown to have an influence function.

Ai and Chen (2007, p. 40) gave an influence function for a function of the solution to a conditional moment restriction under misspecification. In this case the expression given in Proposition 3 is analogous to that in Ai and Chen (2007). Proposition C1 generalizes that expression to orthogonality conditions.