

Instrumental Variable Estimation of Nonlinear Errors-in-Variables Models

Susanne M. Schennach
(Very preliminary and incomplete.)

First draft: October 2003
This version: May 11, 2004

Abstract

In linear specifications, the bias due to the presence of measurement error in a regressor can be entirely avoided when either repeated measurements or instruments are available for the mismeasured regressor. The situation is more complex in nonlinear settings. While identification and root n consistent estimation of general nonlinear specifications have recently been proven in the presence of repeated measurements, similar results relying on instruments have so far only been available for polynomial specifications and absolutely integrable regression functions.

This paper addresses two unresolved issues. First, it is shown that instruments indeed allow for the fully nonparametric identification of general nonlinear regression models in the presence of measurement error. Second, when the regression function is parametrically specified, a root n consistent and asymptotically normal estimator is provided. The starting point of the proposed approach is a system of two functional equations that relate conditional expectations of observed variables to the regression function of interest, as first proposed by Hausman, Ichimura, Newey and Powell (1991) for polynomial specifications. Both the proof of nonparametric identification and the construction of the estimator rely on a representation of these functional equations in terms of Fourier transforms. The proposed estimation procedure takes the form of a generalized method of moment estimator with a plugged-in nonparametric kernel density estimate. As a result, standard techniques borrowed from the semiparametrics literature can be used to establish the estimator's asymptotic properties.

1 Introduction

Estimators based on instrumental variables (IV) have long been used to estimate linear regressions models of the form

$$y = \theta x + \varepsilon \tag{1}$$

where y is the dependent variable, θ is the parameter vector of interest and where the error term ε is potentially correlated with the explanatory variable x . This correlation between ε and x could arise either from endogeneity or measurement error in the regressors. Indeed, if the observed regressor x and the true regressors x^* are related through $x = x^* + \Delta x$, where Δx is a zero mean measurement error that is uncorrelated with x^* , the true model $y = \theta x^* + \Delta y$ is related to the observed Model (1) by

$$y = \theta x^* + \Delta y = \theta x - \theta \Delta x + \Delta y = \theta x + \varepsilon \tag{2}$$

where the disturbance term $\varepsilon = -\theta \Delta x + \Delta y$ is correlated with x , which prompts the need for IV estimation. When the specification is nonlinear

$$y = g(x, \theta) + \varepsilon, \tag{3}$$

IV estimation admits a straightforward extension when the correlation between ε and x is due to endogeneity, but not when it is due to measurement error. The simple additive separation between the observed regressor in the measurement error illustrated in Equation (2) is no longer possible. This problem, first pointed out by Amemiya (1985), has prompted a long search for a solution. Hausman, Ichimura, Newey, and Powell (1991) have provided an asymptotically normal and root n consistent estimator that requires no distributional assumptions regarding the model’s variables in the special case of polynomial specifications. Subsequently, Newey (2001) has shown that with distributional assumptions, root n consistent and asymptotically normal estimation is possible for general functional forms and that without distributional assumptions, consistent estimation is possible, assuming that the model is identified. Under the assumption that the regression function is absolutely integrable, Wang and Hsiao (2003) provide a root n consistent estimator for general functional forms. They also show identification for models having $N_x + 1$ parameters or less, where N_x is the dimension of the mismeasured regressor. However, so far, a general proof that instruments enable the identification of a nonlinear specification with measurement error has remained elusive and existing root n consistent and asymptotically normal estimators of such models are only applicable under rather restrictive assumptions. The difficulty of identifying and estimating nonlinear measurement error models using instruments sharply contrasts with the analogous problem employing repeated measures, where the polynomial case (Hausman, Ichimura, Newey, and Powell (1991)), the identification and consistent estimation in the absence of distributional assumptions (Hausman, Newey, and Powell (1995), Li (2002)), and the root n consistent and asymptotically normal estimation (Schennach (2004a)) have now been fully solved.

The present paper fills the gaps in this ongoing search for a solution to the measurement error problem in nonlinear IV estimation. First, we show nonparametric identification of the regression function without assuming its absolute integrability and without distributional assumptions. In the case where $g(x, \theta)$ is parametric, and in the absence of distributional assumptions, we provide a root n consistent and asymptotically normal estimator. The properties of the proposed estimator are investigated through Monte Carlo simulations.

2 Review of the Theory of Generalized Functions

The concept of “generalized functions” is central to the present paper, because most results will rely on Fourier transforms, which often do not exist within the set of ordinary functions. This section thus recalls the definitions and known results that are relevant to our problem. The formal proof of these results can be found, for instance, in Lighthill (1962). We focus on the case of scalar-valued generalized functions of a scalar variable.

In order to define generalized functions, we first need the following definition.

Definition 1 *Let \mathcal{T} be the set of all functions $s : \mathbb{R} \mapsto \mathbb{R}$ that (i) are everywhere differentiable any number*

of times and (ii) are such that¹ $\left| \frac{d^k s(t)}{dt^k} \right| = O(|t|^{-m})$ as $|t| \rightarrow \infty$ for all $k, m \in \mathbb{N}^+$. Functions in \mathcal{T} are called “test” functions.

Intuitively, functions in \mathcal{T} are both extremely smooth and have extremely thin tails.

Definition 2 A generalized function² b is a sequence of functions b_k in \mathcal{T} such that $\lim_{k \rightarrow \infty} \int b_k(t) s(t) dt$ exists for all $s \in \mathcal{T}$. Let \mathcal{G} denote the set of all generalized functions. (We take the convention that integrals without explicit bounds extend over \mathbb{R} .)

Note that the limit of the sequence $b_k(t)$ may not be part of \mathcal{T} , which is precisely what makes the concept of generalized functions more general than a function. The value of the integral $\int b(t) s(t) dt$ for a given $s \in \mathcal{T}$ is then defined as $\lim_{k \rightarrow \infty} \int b_k(t) s(t) dt$. Perhaps the best known example of a generalized function is Dirac’s delta function $\delta(t)$, defined, for instance, by the sequence

$$b_k(t) = \sqrt{\frac{k}{2\pi}} \exp\left(-\frac{kt^2}{2}\right). \quad (4)$$

Another important example of a generalized function is the j -th derivative of the delta function, denoted by $\delta^{(j)}(t)$ and defined by the sequence $d^j b_k(t) / dt^j$, where $b_k(t)$ is as in Equation (4). The generalized function $\delta^{(j)}(t)$ has the property that $\delta^{(0)}(t) \equiv \delta(t)$ and

$$\int \delta^{(j)}(t) s(t) dt = (-1)^j \frac{d^j s(t)}{dt^j}$$

for $j \in \mathbb{N}$.

Definition 3 Two generalized functions $a(t)$ and $b(t)$ are said to be equal if their associated sequences $a_k(t)$ and $b_k(t)$, respectively, are such that $\lim_{k \rightarrow \infty} \int a_k(t) s(t) dt = \lim_{k \rightarrow \infty} \int b_k(t) s(t) dt$ for all $s \in \mathcal{T}$.

Note that this definition does not require that $a_k(t) = b_k(t)$ for all k and hence, a given generalized function can be defined in terms of more than one sequence. The set of generalized functions is closed under addition, subtraction and differentiation. The product of a generalized function with an ordinary function is guaranteed to be a generalized function if all of the ordinary function’s derivatives exist and diverge no faster than a power of t as $|t| \rightarrow \infty$. However, the product of two generalized functions may not be a generalized function.

Ordinary functions can be viewed as particular cases of generalized functions. For instance, if we let \mathcal{I} be the set of all ordinary functions $c(t)$ such that $\int (1+t^2)^{-m} |c(t)| dt$ is finite for some $m \in \mathbb{N}$, then all ordinary functions in \mathcal{I} are also generalized functions. A generalized function $b(t)$ is said to equal to an ordinary function $c(t)$ in an interval I if for all $s \in \mathcal{T}$ that are supported on I , we have $\int b(t) s(t) dt = \int c(t) s(t) dt$. In the case of Dirac’s delta function, $\delta(t)$ is equal to the 0 function over any interval that does not contain

¹By convention $d^k s(t) / dt^k = s(t)$ for $k = 0$.

²Generalized functions can also be defined as bounded linear functional on \mathcal{T} , but this definition is less convenient for our purposes.

0. However, $\delta(t)$ is not equal to any ordinary function over any interval that includes 0. This concept is important because it will allow us to treat generalized functions as ordinary functions, as long as we stay away from their “singular” points.

Perhaps the most important result for our purpose is that the Fourier transform of a generalized function is a generalized function. As a particular case of this result, the Fourier transform of any function in \mathcal{I} is a generalized function. Hence, in general, the Fourier transform of an ordinary function will not necessarily be an ordinary function, but rather a generalized function.

An important property of any generalized function $b(t)$ is that it can always be decomposed as (see Lighthill (1962), Gel’fand and Shilov (1964))

$$b(t) = b_o(t) + b_s(t) \tag{5}$$

where $b_o(t)$ is an ordinary function while $b_s(t)$ is purely singular, consisting solely of a linear combination of delta functions derivatives of a finite order, as described in more detail in Section 4.2. Moreover, the product of a generalized function $b(t)$ with an ordinary function $a_o(t)$ can be decomposed as

$$b(t) a_o(t) = b_o(t) a_o(t) + b_s(t) a_o(t)$$

where $b_o(t) a_o(t)$ is an ordinary function and where $b_s(t) a_o(t)$ is purely singular, as implied by Lemma 5. Of course, $b(t) a_o(t)$ will only be well-defined if $a_o(t)$ admits a sufficient number of continuous derivatives at the points where the delta functions derivatives contained in $b(t)$ are located.

While this review focuses on so-called *tempered* distributions, there exist more general classes of generalized functions. For instance, as described in Gel’fand and Shilov (1964), the set \mathcal{T} can be limited to compactly supported infinitely differentiable functions, which expands the set of generalized functions for which the limit $\lim_{k \rightarrow \infty} \int a_k(t) s(t) dt$ exists for any $s \in \mathcal{T}$. However, for simplicity, we will focus on functions $a(t)$ whose Fourier transforms $\alpha(\tau)$ are tempered distributions, therefore limiting ourselves to functions $a(t)$ that do not diverge faster than any power of t as $|t| \rightarrow \infty$.

3 Identification

For simplicity, let $y, x, x^*, \Delta y, \Delta x, \Delta x^*$ be scalar random variables and w be a random vector. We consider the model:

$$\begin{aligned} y &= g(x^*) + \Delta y & E[\Delta y | w, \Delta x^*] &= 0 \\ x &= x^* + \Delta x & E[\Delta x | w, \Delta x^*, \Delta y] &= 0 \\ x^* &= X(w) + \Delta x^* & \Delta x^* \text{ independent from } w \text{ and }^3 E[\Delta x^*] &= 0 \end{aligned} \tag{6}$$

where $g(x^*)$ is the function to be determined, while $X(w)$ is an unknown function of a random vector w of instruments. The variables x, y, w are observable while the variables $x^*, \Delta x, \Delta y, \Delta x^*$ are not. The assumptions made are the same as in Newey (2001), except that we allow for nonparametric $g(x^*)$ and $X(w)$. Our task consists in identifying $g(x^*)$ based solely on the knowledge of the observed variables.

³The assumption that that $E[\Delta x^*] = 0$ results in no loss of generality since this can always be achieved by allowing for a constant shift in the function $X(w, \alpha)$.

Since

$$x = x^* + \Delta x = X(w) + \Delta x^* + \Delta x \quad (7)$$

where $E[\Delta x^* + \Delta x|w] = 0$, the function $X(w)$ can be determined from a standard nonparametric least-square projection of x on w (both of which are observable) and is therefore identified. Hence, for the purpose of establishing identification, we define the observed scalar random variable

$$z = X(w). \quad (8)$$

Model (6) can then be rewritten as

$$\begin{aligned} y &= g(x^*) + \Delta y & E[\Delta y|z, u] &= 0 \\ x &= x^* + \Delta x & E[\Delta x|z, u, \Delta y] &= 0 \\ x^* &= z - u. & u &\text{ independent from } z \text{ and } E[u] = 0 \end{aligned} \quad (9)$$

where, for convenience, we have set $u = -\Delta x^*$.

Newey (2001) suggests that the function $g(x^*)$ may be identified from the knowledge of the conditional expectations $E[y|z]$ and $E[xy|z]$ through the equalities:

$$E[y|z] = \int g(z - u) dF(u) \quad (10)$$

$$E[xy|z] = \int (z - u) g(z - u) dF(u) \quad (11)$$

where $F(u)$ denotes the cdf of u and where the integral extend over the whole real line. The heuristic argument supporting this suggestion is the fact that this model is characterized by two unknown functions $g(x^*)$ and $F(u)$ and we have two functional equations available. Moreover, in the special case of a polynomial $g(x^*)$, it is known from Hausman, Ichimura, Newey, and Powell (1991) that the knowledge of the conditional expectations $E[y|z]$ and $E[xy|z]$ is sufficient to identify $g(x^*)$. However, the proof of identification of this model in the general case has so far been missing.

Equations (10) and (11) take on a particularly simple representation in terms of their Fourier transforms.

Lemma 1 *Let $\mathbf{i} = \sqrt{-1}$ and define the following Fourier transforms*

$$\varepsilon_y(\zeta) = \int E[y|z] e^{\mathbf{i}\zeta z} dz \quad (12)$$

$$\varepsilon_{xy}(\zeta) = \int E[xy|z] e^{\mathbf{i}\zeta z} dz \quad (13)$$

$$\gamma(\zeta) = \int g(x^*) e^{\mathbf{i}\zeta x^*} dx^* \quad (14)$$

$$\phi(\zeta) = \int e^{\mathbf{i}\zeta u} dF(u). \quad (15)$$

Whenever $\gamma(\zeta)$, $\varepsilon_y(\zeta)$ and $\varepsilon_{xy}(\zeta)$ are well-defined generalized functions, then Equations (10) and (11) are equivalent to

$$\varepsilon_y(\zeta) = \gamma(\zeta) \phi(\zeta) \quad (16)$$

$$\mathbf{i}\varepsilon_{xy}(\zeta) = \dot{\gamma}(\zeta) \phi(\zeta) \quad (17)$$

where derivatives are denoted by dots and where Definition 3 provides the meaning of an equality between generalized functions.

Proof. Since $\gamma(\zeta)$, $\varepsilon_y(\zeta)$, $\varepsilon_{xy}(\zeta)$ and $\phi(\zeta)$ are well-defined generalized functions, the interchange of the order of integration and the interchange of derivative and integration operations performed below are allowed.⁴

$$\begin{aligned}
\varepsilon_y(\zeta) &= \int \int g(z-u) dF(u) e^{i\zeta z} dz \\
&= \int \int g(z-u) e^{i\zeta z} dz dF(u) \\
&= \int \int g(x^*) e^{i\zeta(x^*+u)} d(x^*) dF(u) \\
&= \int g(x^*) e^{i\zeta x^*} d(x^*) \int e^{i\zeta u} dF(u) \\
&= \gamma(\zeta) \phi(\zeta)
\end{aligned}$$

$$\begin{aligned}
\varepsilon_{xy}(\zeta) &= \int \int (z-u) g(z-u) dF(u) e^{i\zeta z} dz \\
&= \int x^* g(x^*) e^{i\zeta x^*} d(x^*) \int e^{i\zeta u} dF(u) \\
&= \left(-\mathbf{i} \frac{\partial}{\partial \zeta} \int g(x^*) e^{i\zeta x^*} d(x^*) \right) \int e^{i\zeta u} dF(u) \\
&= \left(-\mathbf{i} \frac{\partial}{\partial \zeta} \gamma(\zeta) \right) \phi(\zeta) \\
&\equiv -\mathbf{i} \dot{\gamma}(\zeta) \phi(\zeta)
\end{aligned}$$

■

The Fourier transform of a probability distribution, such as $\phi(\zeta)$, is called a characteristic function and can be shown to be a well-behaved function, namely, its complex modulus is bounded and it is uniformly continuous (Loève (1977)). These properties are consequences of the fact that probability distributions are absolutely integrable by definition. However, functions that are not necessarily absolutely integrable, such as $g(x^*)$, $E[y|z]$, and $E[xy|z]$, may have generalized functions as their Fourier transforms instead of ordinary functions. The requirement, in Lemma 1, that the Fourier transforms $\gamma(\zeta)$, $\varepsilon_y(\zeta)$, and $\varepsilon_{xy}(\zeta)$ exist within the set of generalized functions is very weak; aside from excluding pathological cases such as nonmeasurable functions or functions that are unbounded or undefined over an interval,⁵ it limits the rate at which $g(x^*)$ can diverge as $|x^*| \rightarrow \infty$ to some finite power of x^* . The only commonly used function that does not satisfy this requirement is the exponential. However, as shown in Schennach (2004b), with an exponential specification, $g(x^*)$ is actually not identified from Equations (10) and (11), so allowing for exponentials would bring no additional benefits.

⁴Formally, this is justified as follows. Every generalized function is defined via an inner product with test functions. After a sufficient number of integration by parts, this inner product can be written as the integral of an absolutely integrable function, thus permitting the use of Fubini's Theorem.

⁵Lemma 1 does require $E[y|z]$ and $E[xy|z]$ to be everywhere defined, thus necessitating that the density of z be supported on \mathbb{R} .

Wang and Hsiao (2003) avoid the complications arising from singular Fourier transforms by assuming that $g(x^*)$ and $x^*g(x^*)$ are absolutely integrable (which implies that $E[y|z]$, $E[xy|z]$ are as well). The following Lemma and Theorem establish that, without making this absolute integrability assumption, it is nevertheless possible to show nonparametric identification of $g(x^*)$ in Model (9).

Lemma 2 *Using the assumptions of Lemma 1, Equations (16) and (17) are equivalent to*

$$\varepsilon_y(\zeta) = \gamma(\zeta)\phi(\zeta) \quad (18)$$

$$\mathbf{i}\varepsilon_{(z-x)y}(\zeta) = \gamma(\zeta)\dot{\phi}(\zeta) \quad (19)$$

where $\varepsilon_{(z-x)y}(\zeta) = \int E[(z-x)y|z]e^{i\zeta z}dz$ and where Definition 3 provides the meaning of an equality between generalized functions.

Proof. This can be shown by differentiating each side of Equation (16) with respect to ζ :

$$\begin{aligned} \frac{\partial}{\partial \zeta}\varepsilon_y(\zeta) &= \frac{\partial}{\partial \zeta} \int E[y|z]e^{i\zeta z}dz \\ &= \mathbf{i} \int E[zy|z]e^{i\zeta z}dz \\ &\equiv \mathbf{i}\varepsilon_{zy}(\zeta) \end{aligned}$$

$$\frac{\partial}{\partial \zeta}(\gamma(\zeta)\phi(\zeta)) = \dot{\gamma}(\zeta)\phi(\zeta) + \gamma(\zeta)\dot{\phi}(\zeta)$$

and we obtain:

$$\mathbf{i}\varepsilon_{zy}(\zeta) = \dot{\gamma}(\zeta)\phi(\zeta) + \gamma(\zeta)\dot{\phi}(\zeta). \quad (20)$$

Now, calculating $\mathbf{i}\varepsilon_{zy}(\zeta) - \mathbf{i}\varepsilon_{xy}(\zeta)$, we obtain $\mathbf{i}\varepsilon_{(z-x)y}(\zeta) = \gamma(\zeta)\dot{\phi}(\zeta)$, which is Equation (19). Note that, although the differentiation operation causes a loss of information (as derivatives are unaffected by constant shifts), the whole system of two equations does not suffer from this loss because we keep the original equation $\varepsilon_y(\zeta) = \gamma(\zeta)\phi(\zeta)$ as part of the system. ■

We then need a few conditions to state our identification result.

Assumption 1 $\phi(\zeta) \neq 0$ for all $\zeta \in \mathbb{R}$.

Requiring the characteristic function $\phi(\zeta)$ of the disturbance u to be nonvanishing everywhere is a standard assumption in the deconvolution literature (Carroll, Ruppert, and Stefanski (1995), Fan (1991), Fan and Truong (1993), Li and Vuong (1998), Li (2002), Horowitz and Markatou (1996), Schennach (2004a)). When $\phi(\zeta) = 0$ over some set, $\gamma(\zeta)$ can take any value over the interior of that set without changing the observables $\varepsilon_y(\zeta)$ and $\varepsilon_{xy}(\zeta)$ and it is therefore impossible to fully recover $\gamma(\zeta)$.

Assumption 2 *There exists a positive finite or infinite constant $\bar{\zeta}$ such that (i) $\gamma(\zeta) \neq 0$ almost everywhere in $[-\bar{\zeta}, \bar{\zeta}]$ and (ii) $\gamma(\zeta) = 0$ for all $|\zeta| > \bar{\zeta}$.*⁶

⁶There are no constraints on the behavior of $\gamma(\zeta)$ at $\zeta = \pm\bar{\zeta}$. Also note that if $\gamma(\zeta)$ contains delta function derivatives at some point ξ , $\gamma(\zeta)$ is not equal to the zero function at $\zeta = \xi$ and therefore $\gamma(\xi) \neq 0$.

While Assumption 2 requires that $\gamma(\zeta)$ vanish beyond some frequency $\bar{\zeta}$, it allows $\bar{\zeta}$ to be infinite, so that the case $\gamma(\zeta) \neq 0$ almost everywhere in \mathbb{R} is included as a particular case. It is important to note that the constant $\bar{\zeta}$ does not need to be known. Assumption 2 is fairly weak, as it basically excludes specifications whose Fourier transform vanishes on a finite interval. Such functions exist, but are not commonly used in nonlinear specifications. The asymmetry in the assumptions regarding $\phi(\zeta)$ and $\gamma(\zeta)$ comes from the fact that our main focus is on identifying $\gamma(\zeta)$ and not $\phi(\zeta)$. If we wanted to identify $\phi(\zeta)$ we would need to impose that $\gamma(\zeta) \neq 0$ almost everywhere in \mathbb{R} . Assumptions 1 and 2 can probably be relaxed when parametric constraints on $g(x^*)$ are imposed, since it may then be sufficient to identify $\gamma(\zeta)$ for some, but not necessarily all, ζ .

Assumption 3 $\gamma(\zeta)$, $\varepsilon_y(\zeta)$ and $\varepsilon_{(z-x)y}(\zeta)$ are well-defined generalized functions.

This Assumption is satisfied when the tails of $g(x^*)$ grow no faster than some finite power k of x^* and when the absolute moments of u up to the order $k+1$ exist. These conditions ensure that the singularities in $\gamma(\zeta)$ are no worse than delta function derivatives of a finite order and that $\phi(\zeta)$ admits enough derivatives so that the products $\varepsilon_y(\zeta) = \gamma(\zeta)\phi(\zeta)$ and $\varepsilon_{(z-x)y}(\zeta) = \gamma(\zeta)\dot{\phi}(\zeta)$ are well-defined (by Lemma 5).

We can now state our identification result.

Theorem 1 Under Assumptions 1-3, if $E[|u|] < \infty$, then $g(x^*)$ in Model (9) is nonparametrically identified. Moreover, if $\bar{\zeta} > 0$ in Assumption 2,

$$g(x^*) = \frac{1}{2\pi} \int \gamma(\zeta) e^{-i\zeta x^*} d\zeta \quad (21)$$

where⁷

$$\gamma(\zeta) = \begin{cases} 0 & \text{if } \varepsilon_y(\zeta) = 0 \\ \varepsilon_y(\zeta) \exp\left(-\int_0^\zeta \frac{i\varepsilon_{(z-x)y,o}(\xi)}{\varepsilon_{y,o}(\xi)} d\xi\right) & \text{otherwise} \end{cases}, \quad (22)$$

and where $\varepsilon_{y,o}(\xi)$ and $\varepsilon_{(z-x)y,o}(\xi)$ denote the ordinary function components of $\varepsilon_y(\xi)$ and $\varepsilon_{(z-x)y}(\xi)$, respectively.

Proof. It is only possible to have $\bar{\zeta} = 0$ when $g(x^*)$ is a polynomial, a case which has already been shown to be identified (Hausman, Ichimura, Newey, and Powell (1991)). Hence, we focus on the case where $\bar{\zeta} > 0$.

For $|\zeta| > \bar{\zeta}$, the fact that $\gamma(\zeta) = 0$ can be directly inferred from Equation (18) and the fact that $\varepsilon_y(\zeta) = 0$, since $|\phi(\zeta)| > 0$, as stated in the first part of Equation (22).

We next focus on $|\zeta| \leq \bar{\zeta}$. As indicated in Section 2, any generalized function (such as $\gamma(\zeta)$) can be decomposed as the sum of an ordinary function, denoted by an ‘‘o’’ subscript (e.g. $\gamma_o(\zeta)$), and a purely singular component, denoted by an ‘‘s’’ subscript (e.g. $\gamma_s(\zeta)$), which consists of a linear combination of delta function derivatives. Decomposing $\varepsilon_y(\zeta)$ and $\varepsilon_{xy}(\zeta)$ in a similar fashion and substituting these decompositions into

⁷When the ratio $i\varepsilon_{(z-x)y,o}(\xi)/\varepsilon_{y,o}(\xi)$ takes the forms $0/0$ or ∞/∞ , we take the convention that $i\varepsilon_{(z-x)y,o}(\xi)/\varepsilon_{y,o}(\xi) \equiv \lim_{\xi^* \rightarrow \xi} i\varepsilon_{(z-x)y,o}(\xi^*)/\varepsilon_{y,o}(\xi^*)$, a limit that is shown to always exist in the proof of the theorem. Also, by convention, the statement $\varepsilon_y(\zeta) = 0$ is false when $\varepsilon_y(\zeta)$ contains a delta function derivative at ζ .

Equations (18) and (19) yields

$$\varepsilon_{y,o}(\zeta) + \varepsilon_{y,s}(\zeta) = (\gamma_o(\zeta) + \gamma_s(\zeta))\phi(\zeta) \quad (23)$$

$$\mathbf{i}\varepsilon_{(z-x)y,o}(\zeta) + \mathbf{i}\varepsilon_{(z-x)y,s}(\zeta) = (\gamma_o(\zeta) + \gamma_s(\zeta))\dot{\phi}(\zeta). \quad (24)$$

Since the product of an ordinary function with an ordinary function is an ordinary function, while the product of a purely singular component with an ordinary function is purely singular (as indicated in Section 2), Equations (23) and (24) imply that

$$\varepsilon_{y,o}(\zeta) = \gamma_o(\zeta)\phi(\zeta) \quad (25)$$

$$\mathbf{i}\varepsilon_{(z-x)y,o}(\zeta) = \gamma_o(\zeta)\dot{\phi}(\zeta). \quad (26)$$

Since all quantities are now ordinary functions, Equations (25) and (26) can be manipulated according to the usual rules of multiplication and division. Under the assumption that $\phi(\zeta) \neq 0$, and for any ζ such that $\gamma_o(\zeta) \neq 0$, we can divide each side of Equation (26) by the corresponding side of Equation (25) to obtain

$$\frac{\dot{\phi}(\zeta)}{\phi(\zeta)} = \frac{\mathbf{i}\varepsilon_{(z-x)y,o}(\zeta)}{\varepsilon_{y,o}(\zeta)}. \quad (27)$$

This equation holds almost everywhere in $[-\bar{\zeta}, \bar{\zeta}]$, since the assumption that $\gamma(\zeta) \neq 0$ almost everywhere in $[-\bar{\zeta}, \bar{\zeta}]$ also implies that $\gamma_o(\zeta) \neq 0$ almost everywhere in $[-\bar{\zeta}, \bar{\zeta}]$. By Lemma 4 (in the Appendix) and the assumption that $E[|u|] < \infty$, both $\dot{\phi}(\zeta)$ and $\phi(\zeta)$ are continuous. Since $\phi(\zeta) \neq 0$ for all $\zeta \in \mathbb{R}$ by assumption, the ratio $\dot{\phi}(\zeta)/\phi(\zeta)$ is continuous everywhere. Since Equation (27) holds almost everywhere in $[-\bar{\zeta}, \bar{\zeta}]$ and since $\dot{\phi}(\zeta)/\phi(\zeta)$ is continuous, the ratio $\mathbf{i}\varepsilon_{(z-x)y,o}(\xi)/\varepsilon_{y,o}(\xi)$ contains no essential singularity and its value can be defined everywhere in $[-\bar{\zeta}, \bar{\zeta}]$ by taking limits (that is, we take the convention that $\mathbf{i}\varepsilon_{(z-x)y,o}(\xi)/\varepsilon_{y,o}(\xi)$ is a shorthand notation for $\lim_{\xi^* \rightarrow \xi} \mathbf{i}\varepsilon_{(z-x)y,o}(\xi^*)/\varepsilon_{y,o}(\xi^*)$). With this convention, Equation (27) holds for all $\zeta \in [-\bar{\zeta}, \bar{\zeta}]$.

Integrating each side of Equation (27) with respect to ζ , yields

$$\ln \phi(\zeta) - \ln \phi(0) = \int_0^\zeta \frac{\mathbf{i}\varepsilon_{(z-x)y,o}(\xi)}{\varepsilon_{y,o}(\xi)} d\xi$$

for $|\zeta| < \bar{\zeta}$. Making use of the boundary condition $\phi(0) = \int e^{i0u} dF(u) = \int dF(u) = 1$, and taking exponentials on each side, we obtain⁸

$$\phi(\zeta) = \exp\left(\int_0^\zeta \frac{\mathbf{i}\varepsilon_{(z-x)y,o}(\xi)}{\varepsilon_{y,o}(\xi)} d\xi\right), \quad (28)$$

which provides the value of $\phi(\zeta)$ for $|\zeta| \leq \bar{\zeta}$ in terms of observable quantities.

Next, multiplying each side of Equation (16) by $(\phi(\zeta))^{-1}$ establishes that

$$\gamma(\zeta) = \frac{\varepsilon_y(\zeta)}{\phi(\zeta)} \quad (29)$$

⁸Although, Equation (28) is reminiscent of an identity due to Kotlarski (see Rao (1992), p. 21), it differs substantially in that it involves the Fourier transforms of conditional expectations rather than probability densities.

where $\phi(\zeta)$ is known from Equation (28). This operation is justified because (i) $\phi(\zeta) \neq 0$ by assumption, (ii) multiplication of a generalized function by the ordinary function $((\phi(\zeta))^{-1})$ is allowed, provided that the ordinary function admits a sufficient number of continuous derivatives, which is the case here, since the result of this operation, $\gamma(\zeta)$, is a well-defined generalized function, by assumption. Substituting Equation (28) into Equation (29) yields

$$\gamma(\zeta) = \varepsilon_y(\zeta) \exp\left(-\int_0^\zeta \frac{\mathbf{i}\varepsilon_{(z-x)y,o}(\xi)}{\varepsilon_{y,o}(\xi)} d\xi\right), \quad (30)$$

which is the second part of Equation (22). Finally, as indicated in Equation (21), $g(x^*)$ is simply given by the inverse Fourier transform of $\gamma(\zeta)$, by definition. \blacksquare

Interestingly, while $\gamma(\zeta)$ is identified for all ζ , $\phi(\zeta)$ is only identified for $|\zeta| \leq \bar{\zeta}$.

4 Semiparametric Estimation

Although $g(x^*)$ and $X(w)$ in Model (9) is actually nonparametrically identified, we focus on the case where $g(x^*)$ and $X(w)$ are parametrically specified. Accordingly, we denote the regression function by $g(x^*, \theta)$, its Fourier transform by $\gamma(\zeta, \theta)$ and let $\dot{\gamma}(\zeta, \theta) = \partial\gamma(\zeta, \theta)/\partial\zeta$, where $\theta \in \mathbb{R}^{N_\theta}$ is to be determined. Similarly, the unknown function entering the instrumental equation is written as $X(w, \alpha)$ where $\alpha \in \mathbb{R}^{N_\alpha}$ is to be determined. Note that the distribution of the disturbance u remains nonparametric, making this a semiparametric estimation problem. The appeal of this specific case is the possibility of obtaining root n consistency. The proposed estimator will be based on moment conditions of the form

$$E\left[y \frac{r_y(z, \theta)}{p(z)}\right] + E\left[xy \frac{r_{xy}(z, \theta)}{p(z)}\right] = 0 \quad (31)$$

$$E\left[1y \frac{r_{1y}(z, \theta)}{p(z)}\right] = 1 \quad (32)$$

where $p(z)$ is the density of z and where the functions $r_{1y}(z, \theta)$, $r_y(z, \theta)$, and $r_{xy}(z, \theta)$ are known functions of $\gamma(\zeta, \theta)$ to be subsequently defined.⁹ Note that, in Equation (32), the prefactor $1y$ (obviously equal to y) is a mnemonic device, so all the above expectations have the form

$$E\left[\check{y} \frac{r_{\check{y}}(z, \theta)}{p(z)}\right] \quad (33)$$

for $\check{y} = y, xy, 1y$.

Clearly, regularity conditions will be needed to ensure that these expectations exist and can be root n consistently estimated, despite the presence of a division by the potentially vanishing density $p(z)$. Also, $p(z)$ will need to be nonparametrically estimated, and the resulting estimator falls into the class of GMM estimator with a plug-in nonparametric first-step estimate (as considered, for instance, by Newey (1994)). In this section, we will construct the functions $r_{\check{y}}(z, \theta)$ for $\check{y} = y, xy, 1y$ that will enable the determination of θ via Equations (31) and (32).

⁹Note that the function $r_{1y}(z, \theta)$ differs from $r_y(z, \theta)$ and does *not* denote the first element of $r_y(z, \theta)$.

4.1 Heuristic treatment of the absolutely integrable case

To provide some intuition regarding the form of the moment conditions, we start by providing suitable functions $r_{\check{y}}(z, \theta)$ for $\check{y} = y, xy, 1y$ in the simple case where both $g(x^*, \theta)$ and $x^*g(x^*, \theta)$ are absolutely integrable with respect to x^* . We will subsequently relax this assumption. The general idea is to solve Equations (16) and (17) for $\gamma(\zeta, \theta)$, the Fourier transform of the function $g(x^*, \theta)$ of interest.

Multiplying each side of Equation (17) by $-\mathbf{i}\gamma(\zeta, \theta)$ yields:

$$\varepsilon_{xy}(\zeta)\gamma(\zeta, \theta) = -\mathbf{i}\dot{\gamma}(\zeta, \theta)\gamma(\zeta, \theta)\phi(\zeta). \quad (34)$$

Note that by the absolute integrability of $g(x^*, \theta)$ and $x^*g(x^*, \theta)$, all quantities are ordinary functions and multiplication between them is allowed. By Equation (16) the quantity $\gamma(\zeta, \theta)\phi(\zeta)$ can be identified to $\varepsilon_y(\zeta)$ and we obtain:

$$\varepsilon_{xy}(\zeta)\gamma(\zeta, \theta) = -\mathbf{i}\varepsilon_y(\zeta)\dot{\gamma}(\zeta, \theta). \quad (35)$$

An interesting feature of Equation (35) is that the characteristic function $\phi(\zeta)$ of the error term u has been entirely removed from the problem and we are left with a single functional equation in one unknown function $\gamma(\zeta, \theta)$. The quantities $\varepsilon_{xy}(\zeta)$ and $\varepsilon_y(\zeta)$ are Fourier transforms of conditional expectations involving observable variables and it should therefore be possible to estimate them. The functional forms of $\gamma(\zeta, \theta)$ and $\dot{\gamma}(\zeta, \theta)$ are known from the assumed functional form of $g(x^*, \theta)$, and Equation (35) thus provides a way to estimate the true value of θ , denoted θ^* .

Equation (35) effectively provides us with an infinite number of restrictions, as it must hold for all $\zeta \in \mathbb{R}$. Since $g(x^*, \theta)$ is parametric, we can reduce Equation (35) to a finite system of equations without losing the information Equation (35) provides regarding θ . We can thus replace Equation (35) by a finite system of equations defined by

$$\int \varepsilon_y(\zeta)\mathbf{i}\dot{\gamma}(\zeta, \theta)\omega(\zeta)d\zeta + \int \varepsilon_{xy}(\zeta)\gamma(\zeta, \theta)\omega(\zeta)d\zeta = 0 \quad (36)$$

for some vector of weighting functions $\omega(\zeta)$ chosen so that basic rank conditions hold in order to avoid colinearity among the equations. The vector of weighting functions $\omega(\zeta)$ is helpful because it simplifies both the estimation problem and the proof of root n consistency. Next, if we define

$$\rho_y(\zeta, \theta) = \mathbf{i}\dot{\gamma}(-\zeta, \theta)\omega(-\zeta), \quad (37)$$

$$\rho_{xy}(\zeta, \theta) = \gamma(-\zeta, \theta)\omega(-\zeta) \quad (38)$$

Equation (36) can be written as

$$\int \varepsilon_y(\zeta)\rho_y(-\zeta, \theta)d\zeta + \int \varepsilon_{xy}(\zeta)\rho_{xy}(-\zeta, \theta)d\zeta = 0. \quad (39)$$

and, by Parseval's identity, this equality can be expressed as

$$\int E[y|z]r_y(z, \theta)dz + \int E[xy|z]r_{xy}(z, \theta)dz = 0 \quad (40)$$

where $r_y(z, \theta)$ and $r_{xy}(z, \theta)$ denote the inverse Fourier transform of $\rho_y(\zeta, \theta)$ and $\rho_{xy}(\zeta, \theta)$, respectively. Now, by multiplying and dividing the integrands by $p(z)$, the density of z , Equation (40) becomes

$$E \left[E[y|z] \frac{r_y(z, \theta)}{p(z)} \right] + E \left[E[xy|z] \frac{r_{xy}(z, \theta)}{p(z)} \right] = 0, \quad (41)$$

which is equivalent to Equation (31), by iterated expectations. Obviously, the vector of weighting functions $\omega(\zeta)$ has to be chosen so that $r_y(z, \theta)$ and $r_{xy}(z, \theta)$ are such that all expectations in Equation (41) exist and can be root n consistently estimated.

Note that Equation (35) does not actually identify the scale of $\gamma(\zeta, \theta)$ since multiplying $\gamma(\zeta, \theta)$ (and therefore $\hat{\gamma}(\zeta, \theta)$) by a constant maintains the validity of Equation (35). This is why the additional Equation (32) is needed. By using the knowledge that $\phi(0) = 1$, since a proper distribution must integrate to 1, Equation (16) evaluated at $\zeta = 0$ give us an avenue to estimate the scale, since

$$\varepsilon_y(0) = \gamma(0, \theta) \phi(0) = \gamma(0, \theta). \quad (42)$$

Provided¹⁰ that $\gamma(0, \theta) \neq 0$, this Equation can be trivially rewritten as

$$\int \varepsilon_y(\zeta) \delta(\zeta) \frac{1}{\gamma(0, \theta)} d\zeta = 1, \quad (43)$$

where the delta function $\delta(\zeta)$ merely extracts the value of $\varepsilon_y(\zeta)$ at $\zeta = 0$. Define $r_{1y}(z, \theta)$ to be the inverse Fourier transform of $\delta(\zeta) \frac{1}{\gamma(0, \theta)}$, namely,

$$r_{1y}(z, \theta) = \frac{1}{2\pi} \frac{1}{\gamma(0, \theta)}. \quad (44)$$

Using Parseval's identity and performing the same manipulations as before, Equation (43) can then be written as

$$E \left[E[y|z] \frac{r_{1y}(z, \theta)}{p(z)} \right] = 1, \quad (45)$$

which is equivalent to Equation (32). Of course, additional regularity conditions will be needed to ensure that all expectations in Equations (41) and (45) exist and can be root n consistently estimated.

4.2 General case

While the previous section has justified the form of Equations (31) and (32) when absolute integrability assumptions have been made, it will now be shown that the same basic form of moment conditions apply more generally with a suitable choice of the functions $r_{1y}(z, \theta)$, $r_y(z, \theta)$ and $r_{xy}(z, \theta)$. The main difference resides in the fact that $\gamma(\zeta, \theta)$ is no longer an ordinary function but a generalized function.

Any tempered distribution can be decomposed as a sum of an ordinary function and a finite linear combination of derivatives of delta functions $\delta^{(k)}(\zeta)$ (Lighthill (1962)). Accordingly, we assume the following.

¹⁰We make the assumption that $\gamma(0, \theta) \neq 0$ to simplify the discussion at this point. The more general case discussed in the next section does not require this assumption.

Assumption 4 $\gamma(\zeta, \theta)$ admits the decomposition¹¹

$$\gamma(\zeta, \theta) = \gamma_o(\zeta, \theta) + 2\pi \sum_{k=0}^{\bar{k}} \gamma_k(\theta) (-\mathbf{i})^k \delta^{(k)}(\zeta) \quad (46)$$

where $\gamma_o(\zeta, \theta)$ is an ordinary function, $\bar{k} \in \mathbb{N}$, and the $\gamma_k(\theta)$ for $k = 0, \dots, \bar{k}$ are θ -dependent scalar parameters. Without loss of generality, $\gamma_{\bar{k}}(\theta) \neq 0$.

Since the functional form of $g(x^*, \theta)$ is known, this decomposition can be performed exactly via an analytic calculation¹² of the Fourier transform of $g(x^*, \theta)$. Equation (46) assumes that all singularities are centered at $\zeta = 0$. While it is straightforward to extend our treatment to allow for singularities at other locations, thus allowing for sines and cosines in the specification, we do not explore this eventuality here. Singularities in $\gamma(\zeta, \theta)$ located away from the origin are only possible if the tails of $g(x^*, \theta)$ have an oscillating behavior as $|x^*| \rightarrow \infty$. Model specifications having this property are not commonly used in practical applications. The benefit of a simplified notation therefore outweighs the slight loss in generality. Clearly, $\dot{\gamma}(\zeta, \theta)$ also admits a similar decomposition of the form¹³

$$\dot{\gamma}(\zeta, \theta) = \dot{\gamma}_o(\zeta, \theta) + 2\pi \sum_{k=-1}^{\bar{k}} \gamma_k(\theta) (-\mathbf{i})^k \delta^{(k+1)}(\zeta). \quad (47)$$

By Equations (16) and (17), $\varepsilon_y(\zeta)$ and $\varepsilon_{xy}(\zeta)$ admit a similar decomposition

$$\varepsilon_y(\zeta) = \varepsilon_{y,o}(\zeta) + 2\pi \sum_{k=0}^{\bar{k}} \varepsilon_{y,k} (-\mathbf{i})^k \delta^{(k)}(\zeta) \quad (48)$$

$$\mathbf{i}\varepsilon_{xy}(\zeta) = \mathbf{i}\varepsilon_{xy,o}(\zeta) + 2\pi \mathbf{i} \sum_{k=-1}^{\bar{k}} \varepsilon_{xy,k} (-\mathbf{i})^{k+1} \delta^{(k+1)}(\zeta). \quad (49)$$

where $\varepsilon_{y,o}(\zeta)$ and $\varepsilon_{xy,o}(\zeta)$ are ordinary functions and $\varepsilon_{y,k}$ for $k = 0, \dots, \bar{k}$ and $\varepsilon_{xy,k}$ for $k = -1, \dots, \bar{k}$ are scalar parameters. We know the maximum order \bar{k} of the singularities in $\varepsilon_y(\zeta)$ since it has to correspond to the one of $\gamma(\zeta, \theta)$, by Equations (16) and the fact that $\phi(\zeta)$ is an ordinary function. By a similar reasoning, the maximum order of the singularities in $\varepsilon_{xy}(\zeta)$ corresponds to the one of $\dot{\gamma}(\zeta, \theta)$ by Equation (17). However, $\varepsilon_y(\zeta)$ and $\varepsilon_{xy}(\zeta)$ are quantities that need to be estimated and achieving the above decomposition with noisy estimates will require a special treatment. Nevertheless, the existence of such a decomposition makes it, in principle, possible to rewrite our basic estimating equations in a manner that distinguishes the ordinary and singular components of each generalized function.

¹¹The factor $(-\mathbf{i})^k$ is included so that the coefficients $\gamma_k(\theta)$ are real-valued.

¹²There exist numerous symbolic computational tools which can calculate Fourier transforms that include generalized functions, such as Maple or Mathematica. Alternatively, Table I in Lighthill (1962) provides numerous Fourier transforms.

¹³Note that if $\gamma_o(\zeta, \theta)$ has a step discontinuity at $\zeta = 0$, $\partial\gamma_o(\zeta, \theta)/\partial\zeta$ will contain a delta function. Hence, we define $\dot{\gamma}_o(\zeta, \theta)$ to be the ordinary part of $\partial\gamma_o(\zeta, \theta)/\partial\zeta$ and $\gamma_{-1}(\zeta)$ contains the magnitude of the step in $\gamma_o(\zeta, \theta)$. The term $\gamma_{-1}(\theta)\delta(\zeta)$ will actually never be needed in the estimation procedure we propose.

Lemma 3 Under Assumption 4, Equations (16) and (17), are equivalent to the following system of equations

$$\varepsilon_{y,o}(\zeta) = \gamma_o(\zeta, \theta) \phi(\zeta) \quad (50)$$

$$\mathbf{i}\varepsilon_{xy,o}(\zeta) = \dot{\gamma}_o(\zeta, \theta) \phi(\zeta) \quad (51)$$

$$\Sigma_y = \Gamma_y(\theta) \Phi \quad (52)$$

$$\mathbf{i}\Sigma_{xy} = \Gamma_{xy}(\theta) \Phi \quad (53)$$

where the $(\bar{k} + 1) \times 1$ vectors Φ , Σ_y , Σ_{xy} are given by¹⁴

$$\Phi = \left(\phi(0), -\mathbf{i} \frac{d\phi(0)}{d\zeta}, \dots, (-\mathbf{i})^{\bar{k}} \frac{d^{\bar{k}}\phi(0)}{d\zeta^{\bar{k}}} \right)' \quad (54)$$

$$\Sigma_y = (\varepsilon_{y,0}, \dots, \varepsilon_{y,\bar{k}})' \quad (55)$$

$$\Sigma_{xy} = (\varepsilon_{xy,0}, \dots, \varepsilon_{xy,\bar{k}})' \quad (56)$$

and where the elements of the $(\bar{k} + 1) \times (\bar{k} + 1)$ matrices $\Gamma_y(\theta)$ and $\Gamma_{xy}(\theta)$ are given by

$$\Gamma_{y,j+1 \ k+1}(\theta) = \binom{k+j}{j} \gamma_{k+j}(\theta) \mathbf{1}(k+j \leq \bar{k}) \text{ for } j, k = 0, \dots, \bar{k} \quad (57)$$

$$\Gamma_{xy,j+1 \ k+1}(\theta) = \binom{k+j+1}{j+1} \dot{\gamma}_{k+j}(\theta) \mathbf{1}(k+j \leq \bar{k}) \text{ for } j, k = 0, \dots, \bar{k}. \quad (58)$$

and where the functions $\gamma_o(\zeta, \theta)$ and $\dot{\gamma}_o(\zeta, \theta)$, and the scalars $\varepsilon_{y,k}$, $\varepsilon_{xy,k}$, are defined via Equations (46) through (49).

This result is shown by substituting Equations (46) through (49) into Equations (16) and (17) and by equating the coefficients of the singularities of the same order (see Appendix). We can use this result to devise an estimation procedure that allows for singular $\gamma(\zeta, \theta)$.

As Equations (50) and (51) involve ordinary functions and have the same form as Equations (16) and (17), we can follow the derivation presented in Section 4.1. After introducing a vector of weighting functions $\omega(\zeta)$, we can use an equation of the form

$$\int \varepsilon_{y,o}(\zeta) \dot{\gamma}_o(\zeta, \theta) \omega(\zeta) d\zeta = \mathbf{i} \int \varepsilon_{xy,o}(\zeta) \gamma_o(\zeta, \theta) \omega(\zeta) d\zeta \quad (59)$$

to obtain a finite system of equations. In order for Equation (59) to be useful, it must be possible to separate the ‘‘ordinary function’’ component $\varepsilon_{y,o}(\zeta)$ and $\varepsilon_{xy,o}(\zeta)$ from the quantities $\varepsilon_y(\zeta)$ and $\varepsilon_{xy}(\zeta)$ that are actually observed. Fortunately, there is a simple way to achieve this. Since the location and the order of the singularities are known from the functional form of $\gamma(\zeta, \theta)$, it is straightforward to choose weighting functions $\omega(\zeta)$ that converge to zero sufficiently fast as ζ approaches the singularities so that we have $\int \varepsilon_y(\zeta) \dot{\gamma}_o(\zeta, \theta) \omega(\zeta) d\zeta = \int \varepsilon_{y,o}(\zeta) \dot{\gamma}_o(\zeta, \theta) \omega(\zeta) d\zeta$ and $\int \varepsilon_{xy}(\zeta) \gamma_o(\zeta, \theta) \omega(\zeta) d\zeta = \int \varepsilon_{xy,o}(\zeta) \gamma_o(\zeta, \theta) \omega(\zeta) d\zeta$. More specifically, as $|\zeta| \rightarrow 0$, $\gamma_o(\zeta, \theta) \omega(\zeta)$ must behave as ζ^j , $j \geq \bar{k} + 1$ while $\dot{\gamma}_o(\zeta, \theta) \omega(\zeta)$ must behave as

¹⁴Note that the vector Σ_{xy} does not contain the element $\varepsilon_{xy,-1}$ because it brings to no additional information for the purpose of identifying θ .

ζ^j , $j \geq \bar{k} + 2$, where \bar{k} is as in Equations (48) and (49). These constraints achieve the desired result because $\int \delta^{(k)}(\zeta) \zeta^j d\zeta = (-1)^k \left[d^k (\zeta^j) / d\zeta^k \right]_{\zeta=0} = 0$, since $k \leq \bar{k} + 1 < j$.

Equation (59) thus provides us with moment conditions of the form

$$E \left[y \frac{q_y(z, \theta)}{p(z)} \right] + E \left[xy \frac{q_{xy}(z, \theta)}{p(z)} \right] = 0 \quad (60)$$

where the vector $q_y(z, \theta)$ denotes the inverse Fourier transform of $\dot{\gamma}_o(\zeta, \theta) \omega(\zeta)$ while $q_{xy}(z, \theta)$ denotes the inverse Fourier transform¹⁵ of $\gamma_o(\zeta, \theta) \omega(\zeta)$. Since $p(z) \rightarrow 0$ as $|z| \rightarrow \infty$, it is essential that $\omega(\zeta)$ be chosen so that $q_y(z, \theta)$ and $q_{xy}(z, \theta)$ decay sufficiently rapidly as $|z| \rightarrow 0$ for the expectations in Equation (60) to exist. As will be discussed in detail in Section 5.3.1, selecting $\omega(\zeta)$ to be very smooth will typically achieve this goal.

Since Equation (60) remains valid after a multiplication of $\gamma_o(\zeta, \theta)$ by a constant (as both $q_y(z, \theta)$ and $q_{xy}(z, \theta)$ are then multiplied by the same constant), an additional equation may be needed to determine the scale of $\gamma_o(\zeta, \theta)$. The knowledge that $1 = \phi(0) = \lim_{\zeta \rightarrow 0} \varepsilon_{y,o}(\zeta) / \gamma_o(\zeta, \theta)$ provides the necessary information to obtain the scale of $\gamma_o(\zeta, \theta)$. This constraint can be written in the form of an inner product

$$\int \frac{\varepsilon_y(\zeta)}{\gamma_o(\zeta, \theta)} \varpi(\zeta) d\zeta = 1, \quad (61)$$

for some function $\varpi(\zeta)$ satisfying the following assumption.

Assumption 5 *The exists a function $\varpi(\zeta)$ such that (i) $\int \phi(\zeta) \varpi(\zeta) d\zeta = \phi(0)$ and (ii) $\varpi(\zeta) / \gamma_o(\zeta, \theta) = O(\zeta^{\bar{k}+1})$ as $\zeta \rightarrow 0$.*

The first condition states that $\varpi(\zeta)$ indeed extracts the value of $\phi(\zeta) = \varepsilon_{y,o}(\zeta) / \gamma_o(\zeta, \theta)$ at $\zeta = 0$, while the second ensures that the singularities in $\varepsilon_y(\zeta)$ have no effect on the results. More primitive conditions implying the existence of such a $\varpi(\zeta)$, as well as a method to construct it, will be given in Section 5.3.3. As before, Equation (61) can be written as an expectation:

$$E \left[y \frac{q_{1y}(z, \theta)}{p(z)} \right] = 1 \quad (62)$$

where $q_{1y}(z, \theta)$ is the inverse Fourier transform of $\varpi(\zeta) / \gamma_o(\zeta, \theta)$. Note that, unlike the absolutely integrable case, it is not reasonable to take $\varpi(\zeta)$ to be a delta function because its inverse Fourier transform is a constant, which does not decay as z goes to infinity, and therefore cannot compensate for divergence due to the division by the density $p(z)$. We will describe how to construct a nondegenerate $\varpi(\zeta)$ in a subsequent section.

In general, Equation (60) (and Equation (62)) do not constrain all the degrees of freedom of θ , indicating that some (or all¹⁶) of the information needed to identify θ is actually contained in the singularities in $\varepsilon_y(\zeta)$

¹⁵Note that even though $\gamma_o(\zeta, \theta)$ is an ordinary function, it is possible that $\lim_{\zeta \rightarrow 0} \gamma_o(\zeta, \theta) = \infty$, in which case $\omega(\zeta)$ has to be chosen so that $\gamma_o(\zeta, \theta) \omega(\zeta)$ remains bounded.

¹⁶In the extreme case where $\gamma(\zeta, \theta)$ is purely singular (i.e. $\gamma_o(\zeta, \theta) = 0$), as in the case of polynomials, Equation (60) provides no information.

and $\varepsilon_{xy}(\zeta)$. In this case, Equations (52) and (53) are needed as well. These equations form a linear system of equations for Φ and it is straightforward to solve for Φ in each equation to obtain

$$(\Gamma_y(\theta))^{-1} \Sigma_y = (\Gamma_{xy}(\theta))^{-1} \Sigma_{xy}. \quad (63)$$

The required inverses exist because the matrices $\Gamma_y(\theta)$ and $\Gamma_{xy}(\theta)$ have a triangular structure with nonzero elements on the diagonal.¹⁷ The first element of Φ is known to be equal to $\phi(0) = 1$ (since a distribution must integrate to 1), thus providing the scale of the singular part, and Equation (63) can then be cast into a more informative form:

$$S_{-1}(\Gamma_y(\theta))^{-1} \Sigma_y = S_{-1}(\Gamma_{xy}(\theta))^{-1} \Sigma_{xy} \quad (64)$$

$$S_1(\Gamma_y(\theta))^{-1} \Sigma_y = 1, \quad (65)$$

where S_1 is a $1 \times (\bar{k} + 1)$ selection matrix extracting the first element of a $(\bar{k} + 1)$ dimensional vector while S_{-1} is a $\bar{k} \times (\bar{k} + 1)$ selection matrix extracting the \bar{k} remaining elements of that vector. The matrices $\Gamma_y(\theta)$ and $\Gamma_{xy}(\theta)$ are known from Lemma 3 and the functional form of $\gamma(\zeta, \theta)$.

It is interesting to note the connection between our procedure and the one proposed by Hausman, Ichimura, Newey, and Powell (1991) for polynomial specifications. The matrices $\Gamma_y(\theta)$ and $\Gamma_{xy}(\theta)$ entering the definition of the moment conditions in Lemma 3 have a triangular form and their inversion can therefore be performed via a recursive calculation. Not surprisingly, these recursion relations are identical to the ones of Hausman, Ichimura, Newey, and Powell (1991). In the case of a polynomial specification, the elements of the vectors Σ_y and Σ_{xy} are directly related to the polynomial coefficients of the regression of y on z and of the regression of xy on z . However, for a general nonpolynomial specification, a procedure must be devised to extract these coefficients from $\varepsilon_y(\zeta)$ and $\varepsilon_{xy}(\zeta)$, as we will do next.

Recall, from Equations (48) and (49), that the vectors Σ_y and Σ_{xy} contain the “magnitudes” $\varepsilon_{y,k}$ and $\varepsilon_{xy,k}$ of the singularities present in $\varepsilon_y(\zeta)$ and $\varepsilon_{xy}(\zeta)$. We need a way to extract estimates of these magnitudes from estimates of $\varepsilon_y(\zeta)$ and $\varepsilon_{xy}(\zeta)$, which will be achieved by finding functions $\nu_{y,j}(\zeta, \theta)$ and $\nu_{xy,j}(\zeta, \theta)$ that have the following property.

Assumption 6 *There exists functions $\nu_{y,j}(\zeta, \theta)$ and $\nu_{xy,j}(\zeta, \theta)$ for $j = 0, \dots, \bar{k}$ such that,*

$$\varepsilon_{y,j} = \int \varepsilon_y(\zeta) \nu_{y,j}(\zeta, \theta^*) d\zeta \quad (66)$$

$$\varepsilon_{xy,j} = \int \varepsilon_{xy}(\zeta) \nu_{xy,j}(\zeta, \theta^*) d\zeta. \quad (67)$$

where θ^* denote the true value of θ .

It is fairly simple to find such functions when the ordinary part $\gamma_o(\zeta, \theta)$ vanishes in a neighborhood $[-\eta, \eta]$ of the singularities at $\zeta = 0$, such as in the case when $g(x^*, \theta)$ is a polynomial. In this case, any

¹⁷The determinant of a triangular matrix is equal to the product of the diagonal elements. Due to our convention of indices, the “diagonal” elements have indices $(\bar{k} + 1, 1), (\bar{k}, 2), \dots, (1, \bar{k} + 1)$ instead of the usual $(1, 1), (2, 2), \dots, (\bar{k} + 1, \bar{k} + 1)$ but reordering the rows does not change the magnitude of the determinant.

function $\nu_{y,j}(\zeta, \theta)$ supported on a set contained in $]-\eta, \eta[$ and behaving as $(-\mathbf{i})^j \zeta^j / (j!2\pi)$ as $\zeta \rightarrow 0$ will satisfy Equation (66) since $\frac{(-\mathbf{i})^j}{j!2\pi} \int 2\pi \varepsilon_{y,k} (-\mathbf{i})^k \delta^{(k)}(\zeta) \zeta^j d\zeta = \varepsilon_{y,k} \frac{\mathbf{i}^{k-j}}{j!} \left[\frac{d^k}{d\zeta^k} \zeta^j \right]_{\zeta=0} = \varepsilon_{y,k} \mathbf{i}^{k-j} \mathbf{1} (j=k) = \varepsilon_{y,k} \mathbf{1} (j=k)$ (and similarly for $\nu_{xy,j}(\zeta, \theta)$). In general, when $\varepsilon_{y,o}(\zeta)$ does not vanish in a neighborhood of the singularities, $\nu_{y,j}(\zeta, \theta)$ and $\nu_{xy,j}(\zeta, \theta)$ must be chosen specifically to ensure that the unwanted contributions of the ordinary parts $\varepsilon_{y,o}(\zeta)$ and $\varepsilon_{xy,o}(\zeta)$ to the integrals vanish. A procedure to achieve this under primitive regularity conditions will be given in Section 5.3.2.

4.3 Summary

We can now combine the results derived so far by “stacking” Equations (60), (62), (64) and (65) (with Σ_y and Σ_{xy} expressed via Equations (66) and (67), respectively). The resulting system of moment conditions has the form of Equations (31) and (32) and provides the information needed for the estimation of θ .

Theorem 2 *Under Assumptions 4 and 6, if Equation 60 (and (62)) hold for some $q_y(z, \theta)$, $q_{xy}(z, \theta)$ and $q_{1y}(z, \theta)$, then the solution θ to the following system of equations (if it exists and is unique) gives the true value θ^* :*

$$E \left[y \frac{r_y(z, \theta)}{p(z)} \right] + E \left[xy \frac{r_{xy}(z, \theta)}{p(z)} \right] = 0 \quad (68)$$

$$E \left[1y \frac{r_{1y}(z, \theta)}{p(z)} \right] = 1 \quad (69)$$

where $p(z)$ denotes the density of z and

$$r_y(z, \theta) = (q'_y(z, \theta), s'_y(z, \theta))' \quad (70)$$

$$r_{xy}(z, \theta) = (q'_{xy}(z, \theta), s'_{xy}(z, \theta))' \quad (71)$$

$$r_{1y}(z, \theta) = (q_{1y}(z, \theta), s_{1y}(z, \theta))'. \quad (72)$$

$$s_y(z, \theta) = S_{-1}(\Gamma_y(\theta))^{-1} V_{y,\cdot}(z, \theta) \quad (73)$$

$$s_{xy}(z, \theta) = -S_{-1}(\Gamma_{xy}(\theta))^{-1} V_{xy,\cdot}(z, \theta) \quad (74)$$

$$s_{1y}(z, \theta) = S_1(\Gamma_y(\theta))^{-1} V_{y,\cdot}(z, \theta) \quad (75)$$

$$V_{y,\cdot}(z, \theta) = (V_{y,0}(z, \theta), \dots, V_{y,\bar{k}}(z, \theta))'$$

$$V_{xy,\cdot}(z, \theta) = (V_{xy,0}(z, \theta), \dots, V_{xy,\bar{k}}(z, \theta))'$$

where $V_{y,j}(z, \theta)$ and $V_{xy,j}(z, \theta)$ denote the inverse Fourier transform of $\nu_{y,j}(\zeta, \theta)$ (from Equations (66)) and $\nu_{xy,j}(\zeta, \theta)$ (from Equation (67)), respectively, for $j = 0, \dots, \bar{k}$, where S_1 and S_{-1} are as in Equations (64) and (65) and where $\Gamma_y(\theta)$ and $\Gamma_{xy}(\theta)$ are given by Equation (57) and (58), respectively.

Note that the vectors $q_y(z, \theta)$, $q_{xy}(z, \theta)$ may be reduced to an “empty” vector if the Fourier transform $\gamma(\zeta, \theta)$ has no ordinary function component. Conversely, if $\gamma(\zeta, \theta)$ is a pure ordinary function, then the

vectors $s_y(z, \theta)$ and $s_{xy}(z, \theta)$ are “empty” and this case reduces to the derivation of Section 4.1. In addition, the functions $q_{1y}(z, \theta)$ and $s_{1y}(z, \theta)$, which deal with the scale of the ordinary and the singular parts, respectively, may not be simultaneously needed. The vector $r_{1y}(\zeta, \theta)$ may therefore contain 2 or 1 elements, depending whether it is possible to change the scale of the ordinary and singular part independently or not. It may even be “empty” if the model prevents any change in the scale (e.g., as in a logit model).

Obtaining the asymptotic properties of our estimator thus reduces to analyzing the asymptotic properties of a GMM estimator with a nonparametric first step estimating the density $p(z)$. The estimator defined in this section relies on a suitable choice of the functions $\omega(\zeta)$, $\varpi(\zeta)$, $\mu_{y,j}(\zeta)$ and $\mu_{xy,j}(\zeta)$ which determine the moment functions $r_y(z, \theta)$, $r_{xy}(z, \theta)$ and $r_{1y}(z, \theta)$. In addition to the constraints on these functions that we have described in this section, it will be important to check that the resulting moment functions satisfy the standard regularity conditions of a GMM estimator. This will be the topic of the next section.

5 Asymptotic properties

5.1 Definition of the estimator

The practical implementation of the GMM estimator defined through Equations (68) and (69) requires the following steps. Let (x_j, y_j, w_j) for $j = 1, \dots, n$ be a given sample. First, the variable z_j needs to be constructed from the instruments w_j (see Equation (8)). To this effect, parameter α in Model (6) is estimated using standard (nonlinear) least-squares on the specification

$$x_j = X(w_j, \alpha) + (\Delta x_j^* + \Delta x_j) \quad (76)$$

where $E[(\Delta x_j^* + \Delta x_j) | w_j] = 0$ by the assumptions of Model (6). The resulting $\hat{\alpha}$ is used to define the variable \hat{z}_j as

$$\hat{z}_j = X(w_j, \hat{\alpha}). \quad (77)$$

The variable \hat{z}_j estimates the true $z_j = X(w_j, \alpha^*)$, where α^* denotes the true value of α . Let $p(\cdot | \alpha)$ denote the density of the quantity $X(w_j, \alpha)$ for a given α and let $p(z) = p(z | \alpha^*)$. Next, a nonparametric kernel density estimate of $p(\cdot | \hat{\alpha})$ at point \hat{z}_j can be obtained from

$$\hat{p}(\hat{z}_j | \hat{\alpha}) = (nh)^{-1} \sum_{i=1, i \neq j}^n K((\hat{z}_i - \hat{z}_j) / h)$$

for some kernel¹⁸ $K(\cdot)$ and some bandwidth sequence $h \rightarrow 0$ as $n \rightarrow \infty$.

Finally, $\hat{\theta}$ is defined as the solution to $\hat{Q}(\theta, \hat{\alpha}) = 0$, where

$$\hat{Q}(\theta, \alpha) \equiv n^{-1} \sum_{j=1}^n \left(\frac{Y(x_j, y_j, w_j, \theta, \alpha)}{\hat{p}(X(w_j, \alpha) | \alpha)} - \mathbf{e} \right) \mathbf{1}(\hat{p}(X(w_j, \alpha) | \alpha) \geq \tau) \quad (78)$$

$$Y(\tilde{x}, \tilde{y}, \tilde{w}, \theta, \alpha) = \begin{bmatrix} \tilde{y} r_y(X(\tilde{w}, \alpha), \theta) + \tilde{x} \tilde{y} r_{xy}(X(\tilde{w}, \alpha), \theta) \\ 1 \tilde{y} r_{1y}(X(\tilde{w}, \alpha), \theta) \end{bmatrix} \quad (79)$$

$$\mathbf{e} = \underbrace{(0, \dots, 0)'}_{N_\theta - N_s} \underbrace{(1, \dots, 1)'}_{N_s} \quad (80)$$

¹⁸The kernel $K(\cdot)$ has nothing to do with the reproducing kernel $\kappa(\cdot)$ introduced earlier.

where $1(\cdot)$ is the indicator function, equal to 1 when the event \cdot occurs and τ is some trimming threshold such that $\tau \rightarrow 0$ as $n \rightarrow \infty$ designed to keep divisions by zero under control.¹⁹ The scalar N_s is the dimension of the range of $r_{1y}(z, \theta)$ (from Theorem 2) and can therefore be 0, 1, or 2. The true value of θ , denoted θ^* , is the solution to $Q(\theta, \alpha^*) = 0$, where

$$Q(\theta, \alpha) = E[Q(x, y, w, \theta, \alpha)] \quad (81)$$

$$Q(\tilde{x}, \tilde{y}, \tilde{w}, \theta, \alpha) = \frac{Y(\tilde{x}, \tilde{y}, \tilde{w}, \theta, \alpha)}{p(X(\tilde{w}, \alpha) | \alpha)} - \mathbf{e}. \quad (82)$$

5.2 Asymptotic normality and root n consistency

A few standard regularity conditions are needed to establish the asymptotics of the estimator $\hat{\theta}$. Some of the regularity conditions are restrictions on the functions $r_{\tilde{y}}(z, \theta)$, for $\tilde{y} = y, xy, 1y$ described in the previous sections. Since these functions are specified by the researcher, we will also provide guidance regarding how to construct functions that satisfy these restrictions.

Assumption 7 (y_j, x_j, w_j) is an iid sequence of random variables distributed as (y, x, w) .

While we make the iid assumption to simplify the exposition, generic results on semiparametric estimators found in Andrews (1995) could be used to relax it.

Assumption 8 (i) Let $\mathcal{C} \in \mathbb{R}^{N_\alpha}$ be a compact set such that $\alpha^* = \arg \min_{\alpha \in \mathcal{C}} E[(x - X(w, \alpha))^2]$ is unique and lies in the interior of \mathcal{C} .

(ii) $E[\sup_{\alpha \in \mathcal{C}} X^2(w, \alpha)] < \infty$ and $E[x^2] < \infty$,

(iii) $X(w, \alpha)$ is continuous in α for $\alpha \in \mathcal{C}$,

(iv) $X(w, \alpha)$ is continuously differentiable in α for $\alpha \in \mathcal{A}$, a neighborhood of α^* ,

(v) $E\left[\sup_{\alpha \in \mathcal{A}} \left\| \frac{\partial X(w, \alpha)}{\partial \alpha} \right\|^2\right] < \infty$,

(vi) $E\left[\frac{\partial X(w, \alpha^*)}{\partial \alpha} \frac{\partial X(w, \alpha^*)}{\partial \alpha'}\right]$ is nonsingular, and

(vii) $E\left[(x - X(w, \alpha^*))^2 \left\| \frac{\partial X(w, \alpha^*)}{\partial \alpha} \right\|^2\right] < \infty$.

Assumption 8 collects all the standard regularity conditions traditionally used to show asymptotic normality and root n consistency of the first-step estimator $\hat{\alpha}$ in iid settings.

Assumption 9 There exists a unique θ^* in the interior of some compact set $\Theta \subset \mathbb{R}^{N_\theta}$ such that $Q(\theta^*, \alpha^*) = 0$, for $Q(\theta^*, \alpha^*)$ as in Equation (81).

Assumption 9 is basically implied by the identification results given in the previous section. All that is added beyond what we have already shown is the assumption of the existence of a compact set Θ that contains

¹⁹The trimming is not introduced to ensure that expectations such as $E[yr_y(z, \theta)/p(z)]$ or $E[(yr_y(z, \theta)/p(z))^2]$ exist but rather to show that remainder terms are asymptotically negligible. If $E[(yr_y(z, \theta)/p(z))^2]$, for instance, did not exist, no trimming scheme would restore the root n consistent estimation of the moment $E[yr_y(z, \theta)/p(z)]$.

only one of the potentially multiple solutions to our equations. This assumption also indirectly imposes that the choice of the functions $r_{\check{y}}(z, \theta^*)$ for $\check{y} = y, xy, 1y$ does not inadvertently delete the information that permits identification.

Assumption 10 *The functions $r_{\check{y}}(z, \theta)$ for $\check{y} = y, xy, 1y$ are real-valued.*

The assumption is notationally and practically convenient, although not strictly necessary. It is automatically satisfied when the weighting functions $\mu_j(\zeta)$ or $\mu_{y,j}(\zeta), \mu_{xy,j}(\zeta)$, the elements of $\omega(\zeta)$, $\mu_y(\zeta)$ and $\mu_y(\zeta)$ are symmetric, e.g. $\omega(\zeta) = \omega^\dagger(-\zeta)$, where \dagger denotes complex conjugation.

Assumption 11 *The functions $r_{\check{y}}(z, \theta)$ for $\check{y} = y, xy, 1y$ are continuously differentiable in θ for $\theta \in \Theta$ and all $z \in \mathbb{R}$.*

Assumption 12 $E \left[|\check{y}| \sup_{\alpha \in \mathcal{A}} (p(X(w, \alpha) | \alpha))^{-1} \sup_{\theta \in \Theta} \|r_{\check{y}}(X(w, \alpha), \theta)\| \right] < \infty$ for $\check{y} = y, xy, 1y$.

Assumption 13 $E \left[|\check{y}| \sup_{\alpha \in \mathcal{A}} (p(X(w, \alpha) | \alpha))^{-1} \sup_{\theta \in \mathcal{N}} \|r_{\check{y}, \theta}(X(w, \alpha), \theta)\| \right] < \infty$, for some neighborhood $\mathcal{N} \subset \Theta$ of θ^* and where $r_{\check{y}, \theta}(z, \theta^*) = \partial r_{\check{y}}(z, \theta^*) / \partial \theta'$ for $\check{y} = y, xy, 1y$.

Assumptions 11, 12 and 13 impose conventional continuity and dominance conditions that imply uniform convergence in probability of the quantities that define the estimator and its limiting distribution.

Assumption 14 $E [\psi_\theta(x, y, w) \psi'_\theta(x, y, w)]$ exists, where

$$\psi_\theta(\tilde{x}, \tilde{y}, \tilde{w}) = \begin{bmatrix} (\tilde{y} - E[y|\tilde{z}]) \frac{r_y(\tilde{z}, \theta^*)}{p(\tilde{z}|\alpha^*)} + (\tilde{x}\tilde{y} - E[xy|\tilde{z}]) \frac{r_{xy}(\tilde{z}, \theta^*)}{p(\tilde{z}|\alpha^*)} \\ (\tilde{y} - E[y|\tilde{z}]) \frac{r_{1y}(\tilde{z}, \theta^*)}{p(\tilde{z}|\alpha^*)} \end{bmatrix} \quad (83)$$

where $\tilde{z} = X(\tilde{w}, \alpha^*)$.

Assumption 14 ensures that the asymptotic variance of the estimator exists for α^* fixed, which is essential to obtain root n consistency.

Assumption 15 *The matrix*

$$\Delta = \begin{bmatrix} \int E[y|z] r_{y, \theta}(z, \theta^*) dz + \int E[xy|z] r_{xy, \theta}(z, \theta^*) dz \\ \int E[y|z] r_{1y, \theta}(z, \theta^*) dz \end{bmatrix}$$

is nonsingular.

Assumption 15 is a rank condition that avoids colinearity in the moment conditions at the true value θ^* . It also imposes that the number of moment constraints is equal to the dimension of θ (so that Δ is square). It is obviously possible to relax this just-identified constraint, but for simplicity, we do not consider this here.

Assumption 16 *The kernel function $K(z)$ satisfies (i) $\int K(z) dz = 1$, (ii) $K(z) = K(-z)$ (iii) $\int K(z) z^j dz = 0$ for $j = 1, \dots, N_K - 1$ (iv) $\int |K(z)| |z|^{N_K} dz < \infty$ for some $N_K \in \mathbb{N}$ (v) $K(0) < \infty$ and (vi) $dK(z)/dz$ exists.*

Assumption 16 defines a standard bias-reducing kernel of order N_k .

Assumption 17 *The Fourier transform of $p(z|\alpha)$, denoted $\pi_\alpha(\zeta)$, satisfies $\sup_{\alpha \in \mathcal{A}} \int |\zeta|^{N_k} |\pi_\alpha(\zeta)| d\zeta < \infty$.*

Assumption 17 is slightly stronger than requiring the N_k -th derivative of $p(z|\alpha)$ with respect to z to be continuous (uniformly in z and α) and is slightly weaker than imposing that the $(N_k + 2)$ -th derivative of $p(z|\alpha)$ be absolutely integrable uniformly in α . It is used to show uniform convergence in probability of the kernel density estimate.

Assumption 18 *(i) $n^{1/2}h^2\tau^2 \rightarrow \infty$ (ii) $n^{1/2}h^{N_k}\tau^{-1} \rightarrow 0$ (iii) $\tau \rightarrow 0$ (iv) $h \rightarrow 0$ as $n \rightarrow \infty$.*

Assumption 18 imposes constraints on the rates at which h and τ can go to zero as $n \rightarrow \infty$.

Assumption 19 $E \left[|\check{y}| \sup_{\alpha \in \mathcal{A}} (p(X(w, \alpha)|\alpha))^{-1} \sup_{\theta \in \Theta} \|r_{\check{y}}(X(w, \alpha), \theta)\| 1(p(z) \leq \tau) \right] = o(n^{-1/2})$, for $\check{y} = y, xy, 1y$.

Assumption 19 ensures that the bias introduced by trimming is asymptotically negligible. Following standard practice (e.g. Hardle and Stoker (1989), Assumption 8), this assumption is stated in a relatively high-level form.

Assumption 20 $Q(\theta, \alpha)$ and $\frac{\partial}{\partial \theta'} Q(\theta, \alpha)$ are continuous in α for all $\alpha \in \mathcal{A}$, uniformly in θ for $\theta \in \Theta$, where $Q(\theta, \alpha)$ is given by Equation (81).

Assumption 21 $E \left[\sup_{\alpha \in \mathcal{A}} \left\| \frac{\partial Q(x, y, w, \theta^*, \alpha)}{\partial \alpha'} \right\| \right] < \infty$ where $Q(x, y, w, \theta^*, \alpha)$ is given by Equation (82).

These two last Assumptions ensure that root n consistency of $\hat{\theta}$ is possible despite the statistical noise in the first step estimator $\hat{\alpha}$.

Theorem 3 *Under Assumptions 7 through 19, $n^{1/2}(\hat{\theta} - \theta^*) \xrightarrow{d} \mathbb{N}(0, \Delta^{-1}\Omega\Delta^{-1})$, where Δ is given in Assumption 15 and $\Omega = E[\Psi(x, y, w)\Psi'(x, y, w)]$, where*

$$\Psi(\tilde{x}, \tilde{y}, \tilde{w}) = \psi_\theta(\tilde{x}, \tilde{y}, \tilde{w}) - \frac{\partial Q(\theta^*, \alpha^*)}{\partial \alpha'} \left(E \left[\frac{\partial X(w, \alpha^*)}{\partial \alpha} \frac{\partial X(w, \alpha^*)}{\partial \alpha'} \right] \right)^{-1} \frac{\partial X(\tilde{w}, \alpha^*)}{\partial \alpha} (\tilde{x} - X(\tilde{w}, \alpha^*)) \quad (84)$$

Note that the term subtracted from $\psi_\theta(\tilde{x}, \tilde{y}, \tilde{w})$ in Equation (83) is the correction term for the first-step estimation of α .

5.3 Construction of the moment conditions

Most of the regularity conditions described in the previous section require the existence of moments of the general form $E \left[x^l y \frac{\tilde{r}(z)}{\tilde{p}(z)} \right]$ for $l = 0, 1$, where $\tilde{p}(z)$ is some function closely related to the density $p(z|\alpha)$ while $\tilde{r}(z)$ is a function directly related to $r_y(z, \theta)$, $r_{xy}(z, \theta)$ or $r_{1y}(z, \theta)$. Since $p(z|\alpha) \rightarrow 0$ as $|z| \rightarrow \infty$, it is essential that the numerator $\tilde{r}(z)$ decays sufficiently rapidly as $|z| \rightarrow \infty$ to ensure the existence of the expectations stated in the regularity conditions given in the previous section. Since the functions $r_y(z, \theta)$,

$r_{xy}(z, \theta)$ or $r_{1y}(z, \theta)$ are most naturally constructed from an inverse Fourier transform operation, achieving sufficiently thin tails may not be entirely obvious. We thus describe a methodology to guide the choice of the user-specified *weighting functions* $\omega(\zeta)$, $\nu_{y,j}(\zeta, \theta)$, $\nu_{xy,j}(\zeta, \theta)$ or $\varpi(\zeta)$ introduced in Section 4.2, that enter the definitions of $r_{\check{y}}(z, \theta)$ for $\check{y} = y, xy, 1y$.

While the fact that our estimator involves a choice of various functions may appear unusual at first, the reader is reminded that such a choice often arises in instrumental variable estimation. Indeed, whenever the researcher wishes to impose a conditional mean or an independence restriction, an infinite set of moment conditions would, technically, be needed.²⁰ In practice, researchers typically *choose* a finite set of instruments and perhaps various nonlinear functions of them based on considerations of convenience, sensitivity to outliers, variance reduction, weak instrument bias, etc. Our choice of the weighting functions is conceptually analogous to the choice of which nonlinear functions of a given set of instruments are to be used in conventional instrumental variable estimation, when the disturbances are assumed to satisfy conditional mean restrictions.

In principle, it should be possible to construct weighting functions that are specifically designed to minimize the asymptotic variance of the estimator. For conciseness, we do not explore this issue in the present paper, but it would constitute an interesting topic for future work. We limit ourselves to providing weighting functions enabling root n consistent estimation, although perhaps not efficient estimation. It should be noted that no other previous work on measurement error models with instruments has considered the issue of efficiency either (Hausman, Ichimura, Newey, and Powell (1991), Newey (2001), Wang and Hsiao (2003)).

5.3.1 Choice of $\omega(\zeta)$

The vector of weighting functions $\omega(\zeta)$ was introduced in Section 4.2 to handle the ordinary function component of $\gamma(\zeta, \theta)$, denoted $\gamma_o(\zeta, \theta)$. Our goal is to find a $\omega(\zeta)$ such that the inverse Fourier transforms of the functions $\omega(\zeta)\gamma_o(\zeta, \theta)$ and $\omega(\zeta)\dot{\gamma}_o(\zeta, \theta)$ are rapidly decaying in z .

The basic idea is to rely on the well-known fact that a function's rate of decay as its argument goes to infinity is governed by the smoothness of its Fourier transform. Formally, if $\frac{d^k \sigma(\zeta)}{d\zeta^k}$ is absolutely integrable, then its inverse Fourier transform $s(z)$ is $o(|z|^{-k})$ as $|z| \rightarrow \infty$.²¹ Lemma 7 in the Appendix refines this result, providing sufficient conditions for exponentially decaying tails ($O(\exp(-c|z|^k))$) for $c, k \in \mathbb{R}^+$.

While we have no control over the smoothness of $\gamma(\zeta, \theta)$, since it is given by the specification of the model, we can choose the weighting function $\omega(\zeta)$ to be as smooth as possible. Moreover, when $\gamma_o(\zeta, \theta)$ fails to be smooth at various points, it is possible to pick $\omega(\zeta)$ such that it vanishes where $\gamma_o(\zeta, \theta)$ is not smooth, thus ensuring that the products $\omega(\zeta)\gamma_o(\zeta, \theta)$ and $\omega(\zeta)\dot{\gamma}_o(\zeta, \theta)$ are smooth, and thus resulting in functions $r_{\check{y}}(z, \theta)$ that are rapidly decaying, as desired.

²⁰Of course, it is well-known that there exists a finite set of instruments that can achieve the semiparametric efficiency bound. Nevertheless, each of these optimal instruments is a nonparametric functional of the data generating process, and infinite dimensional nuisance parameters still cannot be avoided.

²¹This result follows from Theorem 17 in Lighthill (1962), after noting that the Fourier transform of $(\mathbf{i}z)^k s(z)$ is $d^k \sigma(\zeta)/d\zeta^k$.

As discussed in Section 4.2, the function $\omega(\zeta)$ must also have the property that it “deletes” singularities in $\varepsilon_y(\zeta)$ and $\varepsilon_{xy}(\zeta)$. For simplicity, we consider the simple case where $\gamma_o(\zeta, \theta)$ is continuously differentiable at $\zeta = 0$, so that the rates at which $\omega(\zeta)\gamma_o(\zeta, \theta) \rightarrow 0$ and $\omega(\zeta)\dot{\gamma}_o(\zeta, \theta) \rightarrow 0$ as $\zeta \rightarrow 0$, are the same as the rate at which $\omega(\zeta) \rightarrow 0$. The elements $\omega_j(\zeta)$ of the weighting function vector $\omega(\zeta)$ can then be selected as follows. Given the highest order of the singularities \bar{k} , from Equations (48) and (49), set

$$\omega_j(\zeta) = \zeta^{\bar{k}+2} \exp(-C_j \zeta^2) \quad (85)$$

where the C_j are some constants. This choice of $\omega(\zeta)$ satisfies two criteria: (i) the $\omega_j(\zeta)$ are very smooth, making it likely that the $E[\tilde{y}r_{\tilde{y}}(z, \theta)/p(z)]$ (and all the related expectations needed in the regularity conditions of Section 5.2) exist for sufficiently small C_j and (ii) each $\omega_j(\zeta)$ behaves as $\zeta^{\bar{k}+2}$ as $|\zeta| \rightarrow 0$, thus ensuring that singularities up to order $\bar{k} + 1$ do not contribute to the inner product between $\varepsilon_y(\zeta)$ and $\dot{\gamma}_o(\zeta, \theta)\omega(\zeta)$ as well as between $\varepsilon_{xy}(\zeta)$ and $\gamma_o(\zeta, \theta)\omega(\zeta)$.

5.3.2 Choice of $\nu_{y,j}(\zeta, \theta)$ and $\nu_{xy,j}(\zeta, \theta)$

The weighting functions $\nu_{y,j}(\zeta, \theta)$, or $\nu_{xy,j}(\zeta, \theta)$ were introduced in Section 4.2 to handle the singular terms in $\gamma(\zeta, \theta)$, denoted by $2\pi \sum_{k=0}^{\bar{k}} \gamma_k(\theta) (-\mathbf{i})^k \delta^{(k)}(\zeta)$ in Equation (46). These weighting functions must satisfy three requirements: (i) they must extract the magnitude of the singularity of order j from $\varepsilon_y(\zeta)$ and $\varepsilon_{xy}(\zeta)$, (ii) they must be insensitive to the ordinary function terms of $\varepsilon_y(\zeta)$ and $\varepsilon_{xy}(\zeta)$ and (iii) their inverse Fourier transforms must be rapidly decaying so that the expectations entering the regularity conditions of Section 5.2 are finite. Let us address each requirement in turn.

First, let us assume that we have at our disposal some families of functions $\mu_{y,j}(\zeta)$ for $j = 0, \dots, \bar{k}$ and $\mu_{xy,j}(\zeta)$ for $j = 0, \dots, \bar{k} + 1$ that are known to be orthogonal to $\phi(\zeta)$. The following theorem then provides a recipe to form the appropriate linear combination among them so as to obtain functions that extract the magnitude of each singularity, thus satisfying conditions (i) and (ii) above.

Theorem 4 *If there exist functions $\mu_{y,j}(\zeta)$ for $j = 0, \dots, \bar{k}$ and $\mu_{xy,j}(\zeta)$ for $j = 0, \dots, \bar{k} + 1$ satisfying $\int \mu_{y,j}(\zeta) \phi(\zeta) d\zeta = 0$ and if the matrices M_y and M_{xy} , whose elements are given by*

$$M_{y,jk} = 2\pi (\mathbf{i})^k \left[\frac{d^k}{d\zeta^k} \left(\frac{\mu_{y,j}(\zeta)}{\gamma_o(\zeta, \theta)} \right) \right]_{\zeta=0} \quad \text{for } j, k = 0, \dots, \bar{k} \quad (86)$$

$$M_{xy,jk} = 2\pi (\mathbf{i})^k \left[\frac{d^k}{d\zeta^k} \left(\frac{\mu_{xy,j}(\zeta)}{\dot{\gamma}_o(\zeta, \theta)} \right) \right]_{\zeta=0} \quad \text{for } j, k = 0, \dots, \bar{k} + 1, \quad (87)$$

exist and are both nonsingular, then Assumption 6 holds with $\nu_{y,j}(\zeta, \theta)$ and $\nu_{xy,j}(\zeta, \theta)$ given by

$$\nu_{y,j}(\zeta, \theta) = \sum_{k=0}^{\bar{k}} (M_y^{-1})_{jk} \frac{\mu_{y,k}(\zeta)}{\gamma_o(\zeta, \theta)} \quad (88)$$

$$\nu_{xy,j}(\zeta, \theta) = \sum_{k=0}^{\bar{k}+1} (M_{xy}^{-1})_{j+1,k} \frac{\mu_{xy,k}(\zeta)}{\hat{\gamma}_o(\zeta, \theta)}. \quad (89)$$

Proof. We show the result for $\nu_{y,j}(\zeta, \theta)$ only since the proof is similar for $\nu_{xy,j}(\zeta, \theta)$. By substituting Equation (48) into Equation (66), we obtain

$$\begin{aligned} & \int \varepsilon_y(\zeta) \nu_{y,j}(\zeta, \theta^*) d\zeta \\ &= \int \left(\varepsilon_{y,o}(\zeta) + 2\pi \sum_{k=0}^{\bar{k}} \varepsilon_{y,k} (-\mathbf{i})^k \delta^{(k)}(\zeta) \right) \sum_{l=0}^{\bar{k}} (M_y^{-1})_{jl} \frac{\mu_{y,l}(\zeta)}{\gamma_o(\zeta, \theta^*)} d\zeta \\ &\equiv P_1 + P_2 \end{aligned}$$

where, by assumption,

$$\begin{aligned} P_1 &= \sum_{l=0}^{\bar{k}} (M_y^{-1})_{jl} \int \varepsilon_{y,o}(\zeta) \frac{\mu_{y,l}(\zeta)}{\gamma_o(\zeta, \theta^*)} d\zeta \\ &= \sum_{l=0}^{\bar{k}} (M_y^{-1})_{jl} \int \phi(\zeta) \mu_{y,l}(\zeta) d\zeta = 0. \end{aligned}$$

On the other hand,

$$\begin{aligned} P_2 &= 2\pi \sum_{k=0}^{\bar{k}} \varepsilon_{y,k} (-\mathbf{i})^k \sum_{l=0}^{\bar{k}} (M_y^{-1})_{jl} \int \delta^{(k)}(\zeta) \frac{\mu_{y,l}(\zeta)}{\gamma_o(\zeta, \theta^*)} d\zeta \\ &= 2\pi \sum_{k=0}^{\bar{k}} \varepsilon_{y,k} (-\mathbf{i})^k \sum_{l=0}^{\bar{k}} (M_y^{-1})_{jl} (-1)^k \left[\frac{d^k}{d\zeta^k} \left(\frac{\mu_{y,l}(\zeta)}{\gamma_o(\zeta, \theta^*)} \right) \right]_{\zeta=0} \\ &= 2\pi \sum_{k=0}^{\bar{k}} \varepsilon_{y,k} \sum_{l=0}^{\bar{k}} (M_y^{-1})_{jl} (\mathbf{i})^k \left[\frac{d^k}{d\zeta^k} \left(\frac{\mu_{y,l}(\zeta)}{\gamma_o(\zeta, \theta^*)} \right) \right]_{\zeta=0} \\ &= \sum_{k=0}^{\bar{k}} \varepsilon_{y,k} \sum_{l=0}^{\bar{k}} (M_y^{-1})_{jl} M_{lk} \\ &= \sum_{k=0}^{\bar{k}} \varepsilon_{y,k} \mathbf{1}(j=k) = \varepsilon_{y,j}. \end{aligned}$$

■

We formulate the hypotheses in terms of orthogonality to $\phi(\zeta)$ rather than, say, $\varepsilon_{y,o}(\zeta)$ and $\varepsilon_{xy,o}(\zeta)$ because much more is known regarding the behavior of characteristic functions (such as $\phi(\zeta)$) than regarding general Fourier transforms. Orthogonality to $\phi(\zeta)$ can be expressed in terms of a variety of more primitive conditions. We consider two alternative cases, namely, (i) the distribution of the disturbance u has compact support and (ii) the moment generating function of the disturbance u exists at least over an interval.

In both cases, the following function provides a convenient building block to construct the functions $\mu_{y,j}(\zeta)$ and $\mu_{xy,j}(\zeta)$:

$$\sigma(\zeta) = \exp(-\cos^{-2}(\zeta\pi/2)) 1(|\zeta| \leq 1). \quad (90)$$

This function is compactly supported and infinitely many times differentiable (including at $|\zeta| = 1$). It is a refinement over the well-known function $\exp\left(-\left(1-\zeta^2\right)^{-1}\right) 1(|\zeta| \leq 1)$ that improves the rate of decay of the inverse Fourier transform of $\sigma(\zeta)$ to $\exp(-c|z|)$ for some $c > 0$ instead of merely faster than $|z|^{-k}$ for any $k \in \mathbb{N}$, as shown in Theorem 6 in the Appendix. Note that a similar result holds in the opposite direction, that is, the Fourier transform of

$$\tilde{s}(z) = \exp(-\cos^{-2}(z\pi/2)) 1(|z| \leq 1), \quad (91)$$

also decays as $\exp(-c|z|)$ for some $c > 0$.

In the case where the distribution of the disturbance u has compact support \mathbb{S} , we consider a family of functions $U_{y,j}(u)$ for $j = 0, \dots, \bar{k}$ and $U_{xy,j}(u)$ for $j = 0, \dots, \bar{k} + 1$ whose supports do not overlap \mathbb{S} . Setting $\mu_{\check{y},j}(\zeta)$ to be the Fourier transform of $U_{\check{y},j}(u)$ for $\check{y} = y, xy$, Parseval's identity then implies that $\int \mu_{\check{y},j}(\zeta) \phi(\zeta) d\zeta = 2\pi \int U_{\check{y},j}(u) dF(u) = 0$, as required by the orthogonality assumption of Theorem 4. Natural candidates for the functions $U_{\check{y},j}(u)$ are of the form

$$U_{\check{y},j}(u) = \tilde{s}((u - u_j)/\eta_j)$$

for some constants u_j, η_j chosen so that the supports of the $U_{\check{y},j}(u)$ do not overlap the support of the distribution of the disturbance u and where $\tilde{s}(\cdot)$ is as defined in Equation (91). On the one hand, the compact support of $\tilde{s}(\cdot)$ makes it straightforward to ensure that supports of the $U_{\check{y},j}(u)$ do not overlap the one of $F(u)$. On the other hand, the smoothness of $\tilde{s}(\cdot)$ ensures that the $\mu_{\check{y},j}(\zeta)$ are rapidly decaying functions of ζ so that the ratios $\mu_{y,k}(\zeta)/\gamma_o(\zeta, \theta)$ and $\mu_{xy,k}(\zeta)/\dot{\gamma}_o(\zeta, \theta)$ entering the definition of $\nu_{y,j}(\zeta, \theta)$ in Theorem 4 do not diverge as $|\zeta| \rightarrow 0$ even if $\gamma_o(\zeta, \theta) \rightarrow 0$ or $\dot{\gamma}_o(\zeta, \theta) \rightarrow 0$ as $|\zeta| \rightarrow \infty$.

Under the weaker assumption that the moment generating function of u exists over an interval, functions $\mu_{\check{y},j}(\zeta)$ for $\check{y} = y, xy$ that satisfy the orthogonality assumption of Theorem 4 can also be constructed. The existence of the moment generating function of the distribution of a disturbance has been previously assumed in other works on nonlinear measurement error problems (e.g. Hausman, Newey, and Powell (1995)), and sometimes even stronger constraints are imposed (e.g. Newey (2001) uses assumptions implying that the nonparametric quantity $\phi(\cdot)$ belongs to a known compact set). The following theorem proves helpful to devise suitable functions $\mu_{\check{y},j}(\zeta)$.

Theorem 5 *Let $\lambda(\zeta)$ be (i) infinitely many times differentiable (ii) supported on $[-\eta, \eta]$ for $\eta > 0$, (iii) such that $\int \lambda(\zeta) d\zeta = C \in \mathbb{R}$ and (iv) $\int |\lambda(\zeta)| d\zeta < \infty$. If the moment generating function of the distribution of u exists over $[-\eta - \epsilon, \eta + \epsilon]$, then $\phi(\zeta)$, the characteristic function of u , is such that $\int \mu(\zeta) \phi(\zeta) d\zeta = C\phi(0)$ for*

$$\mu(\zeta) = \sum_{k=0}^{\infty} \frac{1}{k!} \frac{d^k}{d\zeta^k} \left(\zeta^k \lambda(\zeta) \right). \quad (92)$$

Proof. If the moment generating function of the distribution of u exists over an interval $[-\eta - \epsilon, \eta + \epsilon]$, then the characteristic function $\phi(\zeta)$ will be analytic in a strip $|\text{Im } \zeta| \leq \eta + \epsilon$ in the complex plane (Lukacs (1970), Theorem 7.1.1 and Corollary 7.1.1). It follows that the Taylor series of $\phi(0)$, expanded around some $\zeta \in [-\eta, \eta]$,

$$\phi(0) = \sum_{k=0}^{\infty} \frac{(-\zeta)^k}{k!} \frac{d^k \phi(\zeta)}{d\zeta^k},$$

is convergent. After multiplying by $\lambda(\zeta)$ and integrating over ζ , we have

$$\int_{-\eta}^{\eta} \lambda(\zeta) \phi(0) d\zeta = \int_{-\eta}^{\eta} \sum_{k=0}^{\infty} \frac{(-\zeta)^k}{k!} \frac{d^k \phi(\zeta)}{d\zeta^k} \lambda(\zeta) d\zeta,$$

where the left-hand side is $C\phi(0)$ since $\int_{-\eta}^{\eta} \lambda(\zeta) d\zeta = C$. The integral and summation can be interchanged by Fubini's Theorem since the Taylor series of $\phi(\zeta)$ is absolutely summable for $\zeta \in [-\eta, \eta]$ ($\sum_{k=0}^{\infty} \left| \frac{(-\zeta)^k}{k!} \frac{d^k \phi(\zeta)}{d\zeta^k} \right| \leq \sum_{k=0}^{\infty} \frac{\eta^k}{k!} E[|u|^k] = E[\exp(\eta|u|)] \leq E[\exp(\eta u)] + E[\exp(-\eta u)]$) and $\int |\lambda(\zeta)| d\zeta < \infty$ by assumption. After integrating the right-hand side by parts, we obtain

$$C\phi(0) = \sum_{k=0}^{\infty} \int_{-\eta}^{\eta} \frac{1}{k!} \left(\frac{d^k}{d\zeta^k} (\zeta^k \lambda(\zeta)) \right) \phi(\zeta) d\zeta,$$

or $\int \mu(\zeta) \phi(\zeta) d\zeta = C\phi(0)$, where

$$\mu(\zeta) = \sum_{k=0}^{\infty} \frac{1}{k!} \frac{d^k}{d\zeta^k} (\zeta^k \lambda(\zeta)).$$

■

A natural candidate for $\lambda(\zeta)$ in the above theorem is a translated and scaled version of the function²² $\zeta^j \sigma(\zeta) - 2(2\zeta)^j \sigma(2\zeta)$ for some $j \in \mathbb{N}$, where $\sigma(\zeta)$ is defined in Equation (90). Indeed, these functions are compactly supported,²³ infinitely many times differentiable, absolutely integrable and integrate to zero, implying that $C = 0$ and thus that $\int \mu(\zeta) \phi(\zeta) d\zeta = 0$ in Theorem 5. Therefore, for any $\lambda(\zeta)$ of such form, the resulting function $\mu(\zeta)$ can be used as a valid choice of $\mu_{\check{y},j}(\zeta)$.

Note that in the particular, but relatively common, case where the moment generating function of u exists over the whole real line, the function $\lambda(\zeta)$ does not need to have compact support. In such a case, well behaved $\mu_{\check{y},j}(\zeta)$ can be easily obtained by selecting $\lambda(\zeta)$ to be a linear combination of Gaussians multiplied by a polynomial with coefficients such that $\int \lambda(\zeta) d\zeta = 0$.

5.3.3 Choice of $\varpi(\zeta)$

A weighting function $\varpi(\zeta)$ is sometimes needed to extract the scale of $\gamma_o(\zeta, \theta)$, as described in Assumption 5 in Section 4.2. The most important requirement it must fulfill is $\int \varpi(\zeta) \phi(\zeta) d\zeta = \phi(0)$. The obvious choice $\varpi(\zeta) = \delta(\zeta)$ is unfortunately very nonsmooth and does not yield a function $r_{1y}(\zeta, \theta)$ that has a sufficiently

²²The prefactor ζ^j is included so that the $\mu_{\check{y},j}(\zeta)$ exhibit a variety of behaviors in the neighborhood of the origin for different j , thus making it likely that the matrices $M_{\check{y},jk}$ are invertible. Writing $\lambda(\zeta)$ as a difference between two functions that differ only by their scale is a simple way to obtain a function integrating to zero.

²³After suitable translation and scaling, their compact supports are contained inside the interval over which the moment generating function exists.

thin tail. As in the previous section, we will therefore devise smoother choices of $\varpi(\zeta)$ in two independent cases (i) when the distribution of u has compact support \mathbb{S} and (ii) when the moment generating function of the distribution of u exists over some interval.

In the case of compact support \mathbb{S} , we note that any function $W(z)$ that is equal to 1 over \mathbb{S} will be such that

$$\int W(z) dF(u) = \int dF(u) = \phi(0) \quad (93)$$

so that setting $\varpi(\zeta)$ to be the Fourier transform of $W(z)$ satisfies $\int \varpi(\zeta) \phi(\zeta) d\zeta = \phi(0)$, by Parseval's identity. We also need to ensure that $\varpi(\zeta)$ is such that $\varpi(\zeta)/\gamma_o(\zeta, \theta) = O(\zeta^{\bar{k}+1})$ as $\zeta \rightarrow 0$, as stated in Assumption 5. If $(\gamma_o(\zeta, \theta))^{-1}$ is continuous at $\zeta = 0$, this is equivalent to requiring that $\varpi(\zeta) = O(\zeta^{\bar{k}+1})$, which can be achieved by picking a $W(z)$ such that $\int W(z) z^j dz = 0$ for $j = 0, \dots, \bar{k}$, by the moment theorem.

When the moment generating function of the distribution of u exists over some interval, Theorem 5 with $C = 1$ and

$$\lambda(\zeta) = \frac{\zeta^{\bar{k}+1} \sigma(\zeta)}{\int \xi^{\bar{k}+1} \sigma(\xi) d\xi}$$

where $\sigma(\zeta)$ is defined in Equation (90), provides a function $\mu(\zeta)$ which can be used as a valid $\varpi(\zeta)$. Indeed, the resulting $\varpi(\zeta)$ satisfies $\int \varpi(\zeta) \phi(\zeta) d\zeta = \phi(0)$ and behaves as $\zeta^{\bar{k}+1}$ as $\zeta \rightarrow 0$, which implies that $\varpi(\zeta)/\gamma_o(\zeta, \theta) = O(\zeta^{\bar{k}+1})$, if $(\gamma_o(\zeta, \theta))^{-1}$ is continuous at $\zeta = 0$. In the special case where the moment generating function of u exists over the whole real line, a natural choice for $\lambda(\zeta)$ is a linear combination of Gaussians multiplied by a polynomial with coefficients such that $\int \lambda(\zeta) d\zeta = 1$.

5.4 Computational aspects

The implementation of the estimator is considerably simplified by the fact that all the relatively abstract operations requiring Fourier transforms involve nonrandom quantities. The end result of these operations is a vector of nonlinear functions whose expectations are to be evaluated from the observed data.

The first step in the implementation of the estimator is the calculation of the Fourier transform $\gamma(\zeta, \theta)$ of $g(x^*, \theta)$. Symbolic mathematical packages such as Maple and Mathematica are often able to carry out such transforms automatically, even when the answers involve delta function derivatives. When an analytic expression for $\gamma(\zeta, \theta)$ is not available, the following hybrid analytical and numerical approach can be used. The idea is to write $g(x^*, \theta)$ as

$$g(x^*, \theta) = (g(x^*, \theta) - T(x^*, \theta)) + T(x^*, \theta)$$

where $T(x^*, \theta)$ represents the asymptotic behavior of $g(x^*, \theta)$ for large $|x^*|$ and where $(g(x^*, \theta) - T(x^*, \theta))$ is absolutely integrable (with respect to x^*). If the tail $T(x^*, \theta)$ follows a simple behavior such as a linear combination of functions of the form $(x^*)^{k_1} (\ln(x^*))^{k_2}$, then its Fourier transforms $\Theta(\zeta, \theta)$ can be found in standard Fourier transforms Tables (such as Table I in Lighthill (1962)). Typically, $\Theta(\zeta, \theta)$ will contain both a sum of delta function derivatives, which will provide the values of $\gamma_j(\theta)$ in Equations (57) and (58), as

well as an ordinary function part $\Theta_o(\zeta, \theta)$. The Fourier transform of the remaining absolutely integrable contribution $(g(x^*, \theta) - T(x^*, \theta))$ can then be obtained numerically via

$$\gamma(\zeta, \theta) - \Theta(\zeta, \theta) = \lim_{\substack{t^* \rightarrow \infty, \\ b \rightarrow 0}} \sum_{t=-t^*}^{t^*} (g(tb, \theta) - T(tb, \theta)) e^{i\zeta tb}.$$

All the ordinary function contributions, $\gamma_o(\zeta, \theta) = \Theta_o(\zeta, \theta) + \gamma(\zeta, \theta) - \Theta(\zeta, \theta)$, are then added and their value over a grid $\mathbb{G} = \{\zeta \in \mathbb{R} : \zeta = tb, t = -t^*, \dots, 0, \dots, t^*\}$ is stored, while making sure that the grid is sufficiently fine ($b \rightarrow 0$) and extended ($t^* \rightarrow \infty$) to provide an accurate numerical approximation to $\gamma_o(\zeta, \theta)$.

6 Monte Carlo Simulations

We consider three different specifications, namely, a polynomial, a rational fraction and a logit model. In all cases, the mismeasured regressor x is generated from

$$\begin{aligned} x &= x^* + \Delta x \\ x^* &= z - u \end{aligned}$$

with z, u and Δx drawn from the following distributions

$$\begin{aligned} z &\sim N(0, 1), \\ u &\sim N(0, 1/4), \\ \Delta x &\sim N(0, 1/4). \end{aligned} \tag{94}$$

Note that the ratio of the standard deviation of the measurement error Δx to the standard deviation of the true regressor x^* is $(1/2) / \sqrt{(1 + 1/4)} \approx 0.45$, so that the measurement error is fairly large. In addition the R^2 of the equation $x = z - u + \Delta x$ is $2/3$, indicating that the “strength” of the instrument is of a magnitude that is fairly typical for applications. The distribution of z is deliberately chosen to be a normal in order to explore the proposed estimator in a worst-case scenario where the requirements on the rate at which the $r_{\hat{y}}(z, \theta)$ must decay are the most stringent in order to compensate for the division by the thin-tailed density of z in the moment conditions.

The dependent variable y is generated from

$$y = g(x^*, \theta) + \Delta y, \tag{95}$$

where the functional form of $g(x^*, \theta)$ and the distribution of Δy differ for each model.

For the kernel density estimation of the density of z , an infinite order kernel is used, which has the desirable property that the estimation bias decays faster than any power of the bandwidth h as $h \rightarrow 0$. The specific kernel $K(z)$ used is the inverse Fourier transform of

$$\kappa(\zeta) = \left(\int_{-\infty}^{\infty} \sigma\left(\frac{\xi + 2}{1.9}\right) d\xi \right)^{-1} \int_{-\infty}^{\zeta} \left(\sigma\left(\frac{\xi + 2}{1.9}\right) - \sigma\left(\frac{\xi - 2}{1.9}\right) \right) d\xi \tag{96}$$

where $\sigma(\zeta)$ is given by Equation (90). The prefactor ensures that $\kappa(0) = 1$ and therefore that $\int K(z) dz = 1$, as should be the case for a valid kernel. It is the fact that $\kappa(\zeta)$ is constant over $[-0.1, 0.1]$ which makes $K(z)$ an infinite order kernel. The function $\kappa(\zeta)$ inherits the smoothness of the function $\sigma(\zeta)$, thus ensuring that $K(z)$ is rapidly decaying.

The “optimal” bandwidth parameter h and trimming parameter τ are chosen so as to minimize the GMM objective function associated with the proposed estimator evaluated at θ^* . In our simulation study, this is achieved by scanning values of h from 0.5 to 1.5 in multiplicative increments of 1.1 and values of τ from 0.005 to 0.05 in multiplicative increments of 1.5. The GMM objective function for the given level of smoothing and trimming is then evaluated for 50 replicated samples of 1000 observations and averaged. The “optimal” bandwidth and trimming parameters are found to be:

$$\begin{aligned} h &= 0.585 \\ \tau &= 0.026. \end{aligned}$$

The “optimal” values obtained for all three models considered are the same, within the accuracy implied by the spacings between the consecutive values of h or τ scanned. This is perhaps not surprising since the distribution of z to be nonparametrically estimated is common across all the models.

The finite sample properties of the proposed estimator (for the given values of h and τ) are studied by drawing 5000 samples of 1000 independent observations. As a point of comparison, we also calculate the standard instrumental variable estimator using $\partial g(z, \theta) / \partial \theta$ as a vector of instruments and x as the regressor in addition to a standard nonlinear least squares estimator using x as the regressor, although both of these estimators are clearly biased in the presence of measurement error.

Let $\hat{\theta}_k$ denote any element of $\hat{\theta}$, the parameter vector estimated by any one the three estimators, and let θ_k^* denote any element of θ^* , the true value of the parameter vector. The three estimators are compared on the basis of their bias

$$\text{Bias} = E \left[\hat{\theta}_k \right] - \theta_k^*,$$

their standard deviation

$$\text{Std. Dev.} = \left(E \left[\left(\hat{\theta}_k - E \left[\hat{\theta}_k \right] \right)^2 \right] \right)^{1/2},$$

their root mean square error

$$\text{RMSE} = \left(E \left[\left(\hat{\theta}_k - \theta_k^* \right)^2 \right] \right)^{1/2}$$

and their overall root mean square error

$$\text{RMSE}_{\text{all}} = \left(\text{tr} E \left[\left(\hat{\theta} - \theta^* \right) \left(\hat{\theta} - \theta^* \right)' \right] \right)^{1/2}.$$

Note that the last quantity is a convenient summary measure of the overall performance of an estimator.

Although our estimator is based on moment conditions which have zero expectation at the true value of the parameter vector, it is perfectly normal that it could be biased in a finite sample. First, the moment

conditions used for estimation are nonlinear in θ , and it is well-known that, in this context, just identified GMM exhibits a bias of order n^{-1} , where n is sample size (see, for instance, Newey and Smith (2003)). Second, the implementation of the estimator relies on kernel smoothing and trimming, two techniques which introduce their own bias. Simulations prove to be a helpful tool to verify that the potential presence of such biases does not overcome the benefits of the elimination of the measurement error-induced bias.

We now describe the specifics of each simulation.

6.1 Polynomial Model

This model is defined by

$$\begin{aligned} g(x^*, \theta) &= \theta_1 + \theta_2 x^* + \theta_3 (x^*)^2 + \theta_4 (x^*)^3 \\ \Delta y &\sim N(0, 1/4) \end{aligned}$$

where

$$\theta_1 = 1, \quad \theta_2 = 1, \quad \theta_3 = 0, \quad \theta_4 = -0.5.$$

The Fourier transform of a polynomial contains no ordinary function component and therefore the weighting functions $\omega(\zeta)$ and $\varpi(\zeta)$ do not need to be introduced. The weighting functions $\nu_{y,j}(\zeta, \theta)$ and $\nu_{xy,j}(\zeta, \theta)$ for $j = 0, \dots, \bar{k}$ (where $\bar{k} = 3$) are chosen to be of the form

$$\nu_{y,j}(\zeta, \theta) = \sum_{l=0}^{\bar{k}} a_{y,jl} (\mathbf{i}\zeta)^l \exp\left(-\frac{1}{2} \left(\frac{\zeta}{(1.1)\pi/2}\right)^2\right) \quad (97)$$

$$\nu_{xy,j}(\zeta, \theta) = \sum_{l=0}^{\bar{k}+1} a_{xy,jl} (\mathbf{i}\zeta)^l \exp\left(-\frac{1}{2} \left(\frac{\zeta}{(1.1)\pi/2}\right)^2\right) \quad (98)$$

where the constant coefficients $a_{y,jl}$ and $a_{xy,jl}$ are chosen so that Assumption 6 holds. This is achieved by substituting Equations (97) and (98) into

$$\frac{\partial^k \nu_{y,j}(0, \theta)}{\partial \zeta^k} = 1 (j = k) \text{ for } j, k = 0, \dots, \bar{k} \quad (99)$$

$$\frac{\partial^k \nu_{xy,j}(0, \theta)}{\partial \zeta^k} = 1 (j + 1 = k) \text{ for } j = 0, \dots, \bar{k} \text{ and } k = 0, \dots, \bar{k} + 1 \quad (100)$$

and by solving for $a_{y,jl}$ and $a_{xy,jl}$ in the resulting system of linear equations.

Table 1, compares the performance of the proposed estimator relative to IV and OLS. Although the bias of the proposed estimator is slightly larger than the one of IV for three of the coefficients (θ_1, θ_3 and θ_4), the bias of IV for the remaining coefficient (θ_2) is overwhelmingly large, making the overall performance of IV poor. This is best illustrated by substituting the expected values²⁴ of the coefficients obtained from each estimator into the polynomial specification and by overlapping the graph of each resulting polynomial over the “true” model specification. As seen in Figure 1a), the proposed estimator is much closer to the true specification than any of the other estimators. While the reduction in bias achieved with our estimator comes

²⁴That is, their average over the replications.

	Bias				Std. Dev.				RMSE				
	θ_1	θ_2	θ_3	θ_4	θ_1	θ_2	θ_3	θ_4	θ_1	θ_2	θ_3	θ_4	all
present	-0.052	-0.066	-0.017	0.053	0.166	0.185	0.239	0.045	0.174	0.197	0.239	0.070	0.362
IV	0.001	0.423	0.001	-0.014	0.127	0.301	0.105	0.082	0.127	0.519	0.105	0.083	0.551
OLS	0.000	-0.430	0.001	0.211	0.068	0.129	0.061	0.039	0.068	0.449	0.061	0.215	0.506

Table 1: Simulations results for a polynomial specification.

at the expense of increased standard errors for some coefficients, the overall RMSE (the column labeled by “all” in Table 1) is still lower for the proposed estimator than for the other two estimators.

6.2 Rational fraction

The second example is a specification of the form

$$g(x^*, \theta) = \theta_1 + \theta_2 x^* + \frac{\theta_3}{(1 + (x^*)^2)^2}$$

$$\Delta y \sim N(0, 1/4)$$

where

$$\theta_1 = 1, \quad \theta_2 = 1, \quad \theta_3 = 2.$$

The Fourier transform of $g(x^*, \theta)$ in this case contains both an ordinary and a singular component:

$$\gamma(\zeta, \theta) = \theta_1 2\pi \delta(\zeta) + \theta_2 2\pi \mathbf{i} \delta^{(1)}(\zeta) + \theta_3 \frac{\pi}{2} (1 + |\zeta|) e^{-|\zeta|}. \quad (101)$$

We clearly need to specify the weighting functions $\nu_{y,j}(\zeta, \theta)$ and $\nu_{xy,j}(\zeta, \theta)$ in order to obtain the polynomial coefficients θ_1 and θ_2 . To this effect, we employ Theorem 4 with $\mu_{y,j}(\zeta)$ and $\mu_{xy,j}(\zeta)$ obtained from Theorem 5 with

$$\lambda(\zeta) = (\mathbf{i}\zeta)^j \exp\left(-\frac{1}{2} \left(\frac{\zeta}{(2.1)\pi/2}\right)^2\right) - 2(\mathbf{i}2\zeta)^j \exp\left(-\frac{1}{2} \left(\frac{2\zeta}{(2.1)\pi/2}\right)^2\right).$$

Note that since the distribution of u is a normal, whose moment generating function exists over the whole real line, we are allowed to select $\lambda(\zeta)$ to be supported on \mathbb{R} .

The ordinary part in Equation (101) depends on a single parameter and, consequently, only the scale of the ordinary part needs to be determined. As a result, the vector of weighting function $\omega(\zeta)$ is not needed — only $\varpi(\zeta)$ is. Theorem 5 is then used to obtain $\varpi(\zeta)$ with

$$\lambda(\zeta) = (\mathbf{i}\zeta)^3 \exp\left(-\frac{1}{2} \left(\frac{\zeta}{(1.6)\pi/2}\right)^2\right) \times \left(\int (\mathbf{i}\xi)^3 \exp\left(-\frac{1}{2} \left(\frac{\xi}{(1.6)\pi/2}\right)^2\right) d\xi\right)^{-1}.$$

The prefactor $(\mathbf{i}\zeta)^3$ ensures that the singular parts do not affect the estimation of the ordinary part.

Table 2 summarizes the results of the simulations for the rational fraction model and clearly illustrates the bias-correcting power of the proposed estimator. While the IV estimator exhibits a fortuitously low bias on the θ_2 parameter, it clearly fails to produce unbiased estimates of the coefficient on the nonlinear term (θ_3). As is seen in Figure 1b), the proposed estimator provides a nearly unbiased estimate of the height of the

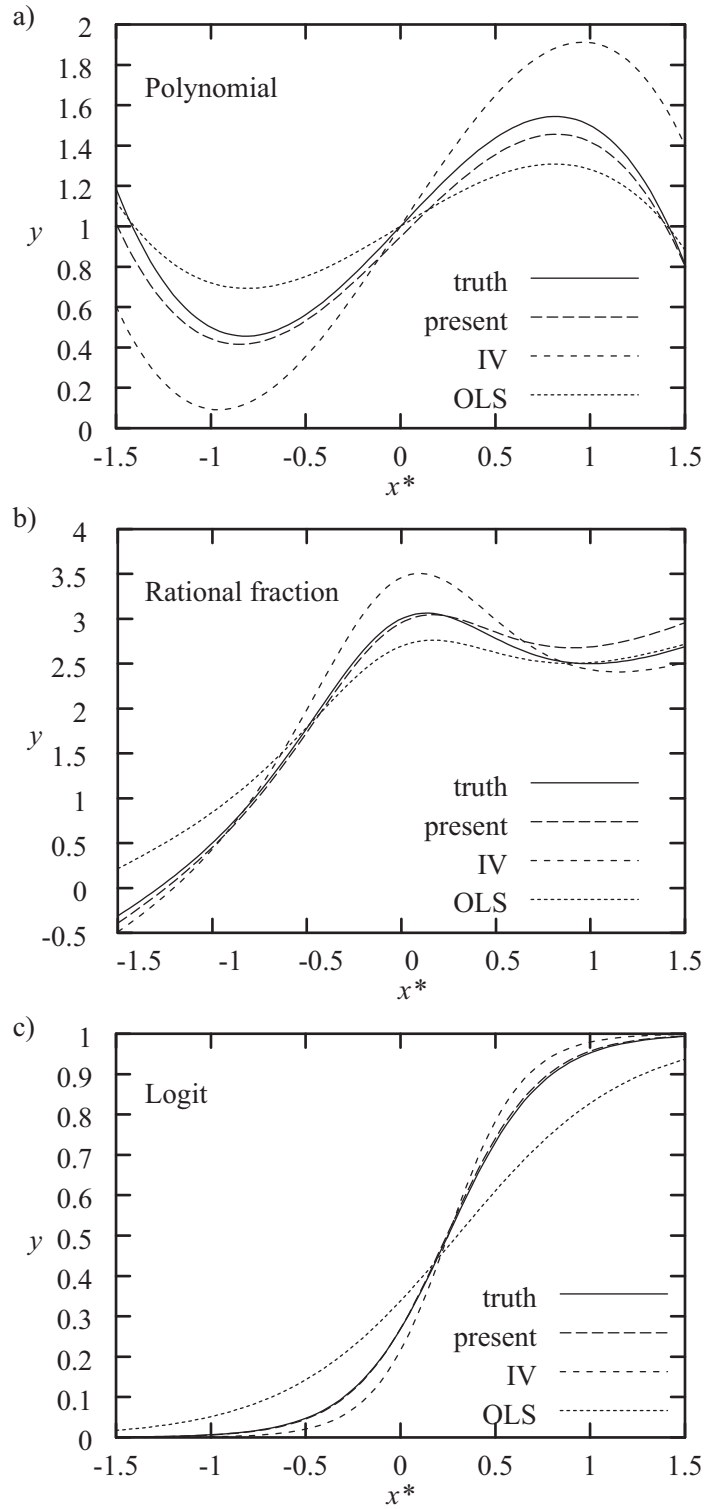


Figure 1: Graphical representation of the bias of each estimator studied. Note that for the logit model in c), the curve for the standard IV estimator excludes the 75% of the replications that do not yield a finite estimate of θ_2 . The actual performance of IV is therefore far worse than indicated by the graph.

	Bias			Std. Dev.			RMSE			
	θ_1	θ_2	θ_3	θ_1	θ_2	θ_3	θ_1	θ_2	θ_3	all
present	0.107	0.117	-0.150	0.146	0.139	0.328	0.181	0.182	0.361	0.443
IV	-0.244	0.001	0.704	0.084	0.028	0.191	0.258	0.028	0.729	0.774
OLS	0.338	-0.166	-0.643	0.046	0.022	0.085	0.341	0.167	0.649	0.752

Table 2: Simulation results for the rational fraction specification.

nonlinear component of the specification, unlike IV, which overestimates it, and OLS, which underestimates it. The proposed estimator has, overall, a bias of only about 10% for this model. Since our estimator typically exhibits larger standard error than both IV and OLS, it is instructive to verify whether it still comes out ahead when both bias and variance are taken into account. Indeed, the overall RMSE clearly points towards the proposed estimator as the best alternative.

6.3 Logit

The logit model can be written as a regression model with the following specification

$$g(x^*, \theta) = \frac{\exp(\theta_1 + \theta_2 x^*)}{1 + \exp(\theta_1 + \theta_2 x^*)} \quad (102)$$

where the disturbance is the form

$$\Delta y = \begin{cases} 1 - g(x^*, \theta) & \text{with probability } g(x^*, \theta) \\ -g(x^*, \theta) & \text{with probability } 1 - g(x^*, \theta) \end{cases} .$$

and where we set

$$\theta_1 = -1, \quad \theta_2 = 4.$$

The singular part of the Fourier transform of $g(x^*, \theta)$ given in Equation (102) contains a single delta function $\pi\delta(\zeta)$. Since this term does not depend on θ , it provides no information to estimate the model and we therefore only need to consider the ordinary part. In addition, the scale of the logistic function is entirely determined by the constraint that a logistic must tends to 1 as $x^* \rightarrow \infty$ and to 0 as $x^* \rightarrow -\infty$ (for $\theta_2 > 0$), so there is no need to estimate the scale. As a result, logit falls into the class of models where the only weighting function needed is $\omega(\zeta)$. The two elements of $\omega(\zeta)$ are chosen to be

$$\omega_j(\zeta) = (\mathbf{i}\zeta)^{j+2} \exp\left(-\frac{1}{2} \left(\frac{\zeta}{(1.5)\pi/2}\right)^2\right) \quad (103)$$

for $j = 1, 2$. Note that the prefactor $(\mathbf{i}\zeta)^{j+2}$ in Equation (103) is chosen to ensure that $\gamma(\zeta, \theta)\omega(\zeta)$ and $\dot{\gamma}(\zeta, \theta)\omega(\zeta)$ are well-behaved. Indeed, the ordinary part $\gamma_o(\zeta, \theta)$ behaves as ζ^{-1} as $\zeta \rightarrow 0$ (and thus $\dot{\gamma}_o(\zeta, \theta)$ behaves as ζ^{-2}) and the above choice of $\omega(\zeta)$ guarantees that its product with $\gamma_o(\zeta, \theta)$ or $\dot{\gamma}_o(\zeta, \theta)$ is bounded.

The results shown in Table 3 and the graph of Figure 1c) clearly indicate that the proposed estimator is nearly unbiased, unlike IV and OLS. Once again, despite its relatively large standard errors relative to IV and OLS, our estimator still outperforms IV and OLS in terms of overall RMSE (see last column). It

	Bias		Std. Dev.		RMSE		
	θ_1	θ_2	θ_1	θ_2	θ_1	θ_2	all
present	0.002	0.095	0.648	0.685	0.648	0.692	0.948
IV	-0.291	1.151	0.242	0.660	0.378	1.326	1.379
OLS	0.329	-1.759	0.104	0.165	0.345	1.767	1.800

Table 3: Simulation results for the logit model.

should also be noted that, for the logit model, the IV estimator using $\partial g(z, \theta) / \partial \theta$ as instruments exhibits the undesirable tendency to give a $\hat{\theta}_2$ that diverges to infinity about 75% of the time. The results for the IV estimator given in Table 3 and Figure 1c) are averages over only the replications of that did converge to a finite value. The actual performance of IV is therefore far worse than reported in the table.

7 Application

Section to be written.

8 Conclusion

This paper addresses two unresolved issues. First, it is shown that instruments indeed permit nonparametric identification of general nonlinear regression models in the presence of measurement error. Second, when the regression function is parametrically specified, a root n consistent and asymptotically normal estimator is provided. The starting point of the proposed approach is a system of two functional equations that relate conditional expectations of observed variables to the regression function of interest, as first proposed by Hausman, Ichimura, Newey, and Powell (1991) for polynomial specifications. Both the proof of nonparametric identification and the construction of the estimator rely on a representation of these functional equations in terms of Fourier transforms. The proposed estimation procedure takes the form of a generalized method of moment estimator with plugged-in nonparametric kernel density estimate. As a result, standard techniques borrowed from the semiparametrics literature could be used to establish its asymptotic properties.

The approach taken in this paper encompasses the approaches of both Wang and Hsiao (2003) and Hausman, Ichimura, Newey, and Powell (1991). When the regression function satisfies some integrability assumptions (as in Wang and Hsiao (2003)), all the Fourier transforms entering the definition of the estimator become ordinary functions and the derivation of Section 4.1 provides the moment conditions needed for estimation. When the regression function is a polynomial (as in Hausman, Ichimura, Newey, and Powell (1991)), its Fourier transform is a linear combination of delta function derivatives and Section 4.2 then provides the moment conditions (in the special case where the Fourier transforms are purely singular). In addition, the proposed methodology also covers functions that are not absolutely integrable and not necessarily polynomial.

References

- AMEMIYA, Y. (1985): “Instrumental Variable Estimator for the Nonlinear Errors-in-Variables Model,” *Journal of Econometrics*, 28, 273–289.
- ANDREWS, D. W. K. (1995): “Nonparametric Kernel Estimation for Semiparametric Models,” *Econometric Theory*, 11, 560–596.
- CARROLL, R. J., D. RUPPERT, AND L. A. STEFANSKI (1995): *Measurement Error in Nonlinear Models*. New York: Chapman & Hall.
- FAN, J. (1991): “On the Optimal Rates of Convergence for Nonparametric Deconvolution Problems,” *Annals of Statistics*, 19(3), 1257–1272.
- FAN, J., AND Y. K. TRUONG (1993): “Nonparametric Regression with Errors in Variables,” *Annals of Statistics*, 21(4), 1900–1925.
- GEL’FAND, I. M., AND G. E. SHILOV (1964): *Generalized Functions*. Academic Press, New York.
- HARDLE, W., AND T. STOKER (1989): “Investigating Smooth Multiple Regression by the Method of Average Derivatives,” *Journal of the American Statistical Association*, 84, 986–995.
- HAUSMAN, J., H. ICHIMURA, W. NEWEY, AND J. POWELL (1991): “Measurement Errors in Polynomial Regression Models,” *Journal of Econometrics*, 50, 271–295.
- HAUSMAN, J., W. NEWEY, AND J. POWELL (1995): “Nonlinear Errors in Variables. Estimation of Some Engel Curves,” *Journal of Econometrics*, 65, 205–233.
- HOROWITZ, J. L., AND M. MARKATOU (1996): “Semiparametric Estimation of Regression Models for Panel Data,” *Review of Economic Studies*, 63, 145.
- LI, T. (2002): “Robust and consistent estimation of nonlinear errors-in-variables models,” *Journal of Econometrics*, 110, 1–26.
- LI, T., AND Q. VUONG (1998): “Nonparametric Estimation of the Measurement Error Model Using Multiple Indicators,” *Journal of Multivariate Analysis*, 65, 139–165.
- LIGHTHILL, M. J. (1962): *Introduction to Fourier Analysis and Generalized Function*. London: Cambridge University Press.
- LOÈVE, M. (1977): *Probability Theory I*. New York: Springer.
- LUKACS, E. (1970): *Characteristic Functions*. Griffin, London, second edn.
- NEWEY, W. (1994): “The Asymptotic Variance of Semiparametric Estimators,” *Econometrica*, 62, 1349–1382.

- (2001): “Flexible Simulated Moment Estimation of Nonlinear Errors-in-Variables Models,” *Review of Economics and Statistics*, 83, 616–627.
- NEWKEY, W., AND D. MCFADDEN (1994): “Large Sample Estimation and Hypothesis Testing,” in *Handbook of Econometrics*, ed. by R. F. Engel, and D. L. McFadden, vol. IV. Elsevier Science.
- NEWKEY, W., AND R. J. SMITH (2003): “Higher-Order Properties of GMM and Generalized Empirical Likelihood Estimators,” *Econometrica*, forthcoming.
- PAGAN, A., AND A. ULLAH (1999): *Nonparametric Econometrics*. Cambridge University Press, Cambridge, UK.
- POWELL, J., J. STOCK, AND T. STOKER (1989): “Semiparametric Estimation of Index Coefficients,” *Econometrica*, 57, 1403–1430.
- RAO, P. (1992): *Identifiability in Stochastic Models*. New York: Academic Press.
- SCHENNACH, S. M. (2004a): “Estimation of Nonlinear Models with Measurement Error,” *Econometrica*, 72, 33–75, <http://home.uchicago.edu/~smschenn/trica.pdf>.
- (2004b): “Exponential specifications and measurement error,” *Economics Letters*, forthcoming, <http://home.uchicago.edu/~smschenn/expiv.pdf>.
- WANG, L., AND C. HSIAO (2003): “Identification and Estimation of Semiparametric Nonlinear Errors-in-Variables Models,” Working Paper, University of Southern California.

A Proofs

While the following Lemma resembles a well-known result regarding characteristic functions (see Loève (1977), following Property 13.1), it generalizes it to apply to Fourier transforms of any absolutely integrable function.

Lemma 4 *If $s(z)$ is absolutely integrable, then its Fourier transform $\sigma(\zeta)$ is continuous. In particular, if $\int |z|^k |s(z)| dz < \infty$ for some $k \in \mathbb{N}$, then $d^k \sigma(\zeta) / d\zeta^k$ is continuous.*

Proof. First note that $\sigma_j(\zeta) = \int e^{i\zeta z} s(z) 1(|z| \leq j) dz$ is continuous in ζ for every j :

$$\begin{aligned}
|\sigma_j(\zeta) - \sigma_j(\xi)| &= \left| \int (e^{i\zeta z} - e^{i\xi z}) s(z) 1(|z| \leq j) dz \right| \\
&= \left| \int e^{i(\zeta+\xi)z/2} \left(e^{i(\zeta-\xi)z/2} - e^{-i(\zeta-\xi)z/2} \right) s(z) 1(|z| \leq j) dz \right| \\
&\leq \int |e^{i(\zeta+\xi)z/2}| |2 \sin((\zeta - \xi)z/2)| |s(z)| 1(|z| \leq j) dz \\
&= \int |2 \sin((\zeta - \xi)z/2)| |s(z)| 1(|z| \leq j) dz \\
&\leq \int |(\zeta - \xi)z| |s(z)| 1(|z| \leq j) dz \\
&\leq |\zeta - \xi| j \int |s(z)| 1(|z| \leq j) dz \\
&\leq |\zeta - \xi| j \int |s(z)| dz.
\end{aligned}$$

Next, observe that $\sigma_j(\zeta)$ converges to $\sigma(\zeta)$ uniformly in ζ :

$$\begin{aligned}
|\sigma(\zeta) - \sigma_j(\zeta)| &= \left| \int e^{i\zeta z} s(z) (1 - 1(|z| \leq j)) dz \right| \\
&\leq \int |e^{i\zeta z}| |s(z)| (1 - 1(|z| \leq j)) dz \\
&= \int_{|z| \geq j} |s(z)| dz \\
&\rightarrow 0 \text{ as } j \rightarrow \infty.
\end{aligned}$$

Since $\sigma_j(\zeta)$ is a sequence of continuous functions converging uniformly to $\sigma(\zeta)$, the limiting function $\sigma(\zeta)$ must be continuous. The second assertion then follows from the fact that the Fourier transform of $(\mathbf{i}z)^k s(z)$ is $d^k \sigma(\zeta) / d\zeta^k$. ■

Definition 4 *For some function $\psi(\zeta)$, let $\frac{d^{-1}}{d\zeta^{-1}} \psi(\zeta) \equiv \int_a^\zeta \psi(\xi) d\xi$ for some arbitrary constant a . For $k \geq 1$, define $\frac{d^{-k-1}}{d\zeta^{-k-1}} \psi(\zeta) \equiv \frac{d^{-1}}{d\zeta^{-1}} \frac{d^{-k}}{d\zeta^{-k}} \psi(\zeta)$, by recursion.*

Lemma 5 *If $\phi(\zeta)$ is k times continuously differentiable at $\zeta = 0$, then $\delta^{(k)}(\zeta) \phi(\zeta) = (-1)^k \sum_{j=0}^k \binom{k}{j} \frac{d^{k-j} \phi(0)}{d\zeta^{k-j}} \delta^{(j)}(\zeta)$.*

Proof. Let ψ be some test function in \mathcal{T} as given in Definition 1. By k repeated integration by parts, we have,

$$\int \left(\delta^{(k)}(\zeta) \phi(\zeta) \right) \psi(\zeta) d\zeta = (-1)^k \int \left(\frac{d^{-k}}{d\zeta^{-k}} \delta^{(k)}(\zeta) \right) \frac{d^k}{d\zeta^k} (\phi(\zeta) \psi(\zeta)) d\zeta,$$

after noting that the boundary terms vanish due to the thin tails of $\psi(\zeta)$ and all of its derivatives. Next,

$$\begin{aligned}
\int \left(\delta^{(k)}(\zeta) \phi(\zeta) \right) \psi(\zeta) d\zeta &= (-1)^k \int \delta(\zeta) \left(\frac{d^k}{d\zeta^k} (\phi(\zeta) \psi(\zeta)) \right) d\zeta \\
&= (-1)^k \int \delta(\zeta) \sum_{j=0}^k \binom{k}{j} \frac{d^{k-j} \phi(\zeta)}{d\zeta^{k-j}} \frac{d^j \psi(\zeta)}{d\zeta^j} d\zeta \\
&= (-1)^k \sum_{j=0}^k \binom{k}{j} \frac{d^{k-j} \phi(0)}{d\zeta^{k-j}} \frac{d^j \psi(0)}{d\zeta^j} \\
&= (-1)^k \sum_{j=0}^k \binom{k}{j} \frac{d^{k-j} \phi(0)}{d\zeta^{k-j}} \int \delta^{(j)}(\zeta) \psi(\zeta) d\zeta \\
&= \int \left((-1)^k \sum_{j=0}^k \binom{k}{j} \frac{d^{k-j} \phi(0)}{d\zeta^{k-j}} \delta^{(j)}(\zeta) \right) \psi(\zeta) d\zeta.
\end{aligned}$$

■

Proof of Lemma 3. Substituting Equations (46) through (49) into Equations (16) and (17), we obtain

$$\begin{aligned}
\varepsilon_{y,o}(\zeta) + 2\pi \sum_{k=0}^{\bar{k}} \varepsilon_{y,k} (-\mathbf{i})^k \delta^{(k)}(\zeta) &= \left(\gamma_o(\zeta, \theta) + 2\pi \sum_{k=0}^{\bar{k}} \gamma_k(\theta) (-\mathbf{i})^k \delta^{(k)}(\zeta) \right) \phi(\zeta) \\
\mathbf{i}\varepsilon_{xy,o}(\zeta) + 2\pi \mathbf{i} \sum_{k=-1}^{\bar{k}} \varepsilon_{xy,k} (-\mathbf{i})^{k+1} \delta^{(k+1)}(\zeta) &= \left(\dot{\gamma}_o(\zeta, \theta) + 2\pi \sum_{k=0}^{\bar{k}} \gamma_k(\theta) (-\mathbf{i})^{k+1} \delta^{(k+1)}(\zeta) \right) \phi(\zeta).
\end{aligned}$$

Equating the ordinary functions part of each expression yields

$$\begin{aligned}
\varepsilon_{y,o}(\zeta) &= \gamma_o(\zeta, \theta) \phi(\zeta) \\
\mathbf{i}\varepsilon_{xy,o}(\zeta) &= \dot{\gamma}_o(\zeta, \theta) \phi(\zeta),
\end{aligned}$$

while equating the singular parts yields

$$\begin{aligned}
\sum_{k=0}^{\bar{k}} \varepsilon_{y,k} (-\mathbf{i})^k \delta^{(k)}(\zeta) &= \sum_{k=0}^{\bar{k}} \gamma_k(\theta) (-\mathbf{i})^k \delta^{(k)}(\zeta) \phi(\zeta) \\
\sum_{k=-1}^{\bar{k}} \mathbf{i}\varepsilon_{xy,k} (-\mathbf{i})^{k+1} \delta^{(k+1)}(\zeta) &= \sum_{k=0}^{\bar{k}} \gamma_k(\theta) (-\mathbf{i})^{k+1} \delta^{(k+1)}(\zeta) \phi(\zeta).
\end{aligned}$$

By Lemma 5, we have

$$\begin{aligned}
\sum_{k=0}^{\bar{k}} \varepsilon_{y,k} (-\mathbf{i})^k \delta^{(k)}(\zeta) &= \sum_{k=0}^{\bar{k}} \gamma_k(\theta) (-\mathbf{i})^k \sum_{j=0}^k \binom{k}{j} \phi^{(k-j)}(0) \delta^{(j)}(\zeta) \\
\sum_{k=-1}^{\bar{k}} \mathbf{i}\varepsilon_{xy,k} (-\mathbf{i})^{k+1} \delta^{(k+1)}(\zeta) &= \sum_{k=0}^{\bar{k}} \gamma_k(\theta) (-\mathbf{i})^{k+1} \sum_{j=0}^{k+1} \binom{k+1}{j} \phi^{(k+1-j)}(0) \delta^{(j)}(\zeta)
\end{aligned}$$

Simple manipulations then give

$$\begin{aligned}
\sum_{k=0}^{\bar{k}} \varepsilon_{y,k} (-\mathbf{i})^k \delta^{(k)}(\zeta) &= \sum_{k=0}^{\bar{k}} \gamma_k(\theta) (-\mathbf{i})^k \sum_{j=0}^{\bar{k}} \binom{k}{j} \mathbf{1}(j \leq k) \phi^{(k-j)}(0) \delta^{(j)}(\zeta) \\
\sum_{k=-1}^{\bar{k}} \mathbf{i}\varepsilon_{xy,k} (-\mathbf{i})^{k+1} \delta^{(k+1)}(\zeta) &= \sum_{k=0}^{\bar{k}} \gamma_k(\theta) (-\mathbf{i})^{k+1} \sum_{j=0}^{\bar{k}+1} \binom{k+1}{j} \mathbf{1}(j \leq k+1) \phi^{(k+1-j)}(0) \delta^{(j)}(\zeta)
\end{aligned}$$

$$\begin{aligned}
\sum_{j=0}^{\bar{k}} \varepsilon_{y,j} (-\mathbf{i})^j \delta^{(j)}(\zeta) &= \sum_{j=0}^{\bar{k}} \sum_{k=0}^{\bar{k}} \binom{k}{j} \gamma_k(\theta) (-\mathbf{i})^k \mathbf{1}(j \leq k) \phi^{(k-j)}(0) \delta^{(j)}(\zeta) \\
\sum_{j=-1}^{\bar{k}} \mathbf{i} \varepsilon_{xy,j} (-\mathbf{i})^{j+1} \delta^{(j+1)}(\zeta) &= \sum_{j=0}^{\bar{k}+1} \sum_{k=0}^{\bar{k}} \binom{k+1}{j} \gamma_k(\theta) (-\mathbf{i})^{k+1} \mathbf{1}(j \leq k+1) \phi^{(k+1-j)}(0) \delta^{(j)}(\zeta) \\
&= \sum_{j=-1}^{\bar{k}} \sum_{k=0}^{\bar{k}} \binom{k+1}{j+1} \gamma_k(\theta) (-\mathbf{i})^{k+1} \mathbf{1}(j \leq k) \phi^{(k-j)}(0) \delta^{(j+1)}(\zeta).
\end{aligned}$$

Equating the coefficients of the delta function derivatives of the same order gives

$$\begin{aligned}
\varepsilon_{y,j} (-\mathbf{i})^j &= \sum_{k=0}^{\bar{k}} \binom{k}{j} \gamma_k(\theta) (-\mathbf{i})^k \mathbf{1}(j \leq k) \phi^{(k-j)}(0) \\
\mathbf{i} \varepsilon_{xy,j} (-\mathbf{i})^{j+1} &= \sum_{k=0}^{\bar{k}} \binom{k+1}{j+1} \gamma_k(\theta) (-\mathbf{i})^{k+1} \mathbf{1}(j \leq k) \phi^{(k-j)}(0) \\
\varepsilon_{y,j} (-\mathbf{i})^j &= \sum_{l=-j}^{\bar{k}-j} \binom{j+l}{j} \gamma_{j+l}(\theta) (-\mathbf{i})^{j+l} \mathbf{1}(j \leq j+l) \phi^{(j+l-j)}(0) \\
&= \sum_{l=-j}^{\bar{k}-j} \binom{j+l}{j} \gamma_{j+l}(\theta) (-\mathbf{i})^{j+l} \mathbf{1}(0 \leq l) \phi^{(l)}(0) \\
&= \sum_{l=0}^{\bar{k}-j} \binom{j+l}{j} \gamma_{j+l}(\theta) (-\mathbf{i})^{j+l} \phi^{(l)}(0) \\
&= \sum_{l=0}^{\bar{k}} \binom{j+l}{j} \gamma_{j+l}(\theta) (-\mathbf{i})^{j+l} \mathbf{1}(l \leq \bar{k}-j) \phi^{(l)}(0) \\
&= \sum_{k=0}^{\bar{k}} \binom{k+j}{j} \gamma_{k+j}(\theta) (-\mathbf{i})^{k+j} \mathbf{1}(k \leq \bar{k}-j) \phi^{(k)}(0) \\
\mathbf{i} \varepsilon_{xy,j} (-\mathbf{i})^j &= \sum_{l=-j}^{\bar{k}-j} \binom{j+l+1}{j+1} \gamma_{j+l}(\theta) (-\mathbf{i})^{j+l} \mathbf{1}(j \leq j+l) \phi^{(j+l-j)}(0) \\
&= \sum_{l=0}^{\bar{k}-j} \binom{j+l+1}{j+1} \gamma_{j+l}(\theta) (-\mathbf{i})^{j+l} \phi^{(l)}(0) \\
&= \sum_{l=0}^{\bar{k}+1} \binom{j+l+1}{j+1} \gamma_{j+l}(\theta) (-\mathbf{i})^{j+l} \mathbf{1}(l \leq \bar{k}-j) \phi^{(l)}(0) \\
&= \sum_{k=0}^{\bar{k}} \binom{k+j+1}{j+1} \gamma_{k+j}(\theta) (-\mathbf{i})^{k+j} \mathbf{1}(k \leq \bar{k}-j) \phi^{(k)}(0) \text{ for } j \geq 0. \\
\varepsilon_{y,j} &= \sum_{k=0}^{\bar{k}} \binom{k+j}{j} \gamma_{k+j}(\theta) \mathbf{1}(k \leq \bar{k}-j) (-\mathbf{i})^k \phi^{(k)}(0) \\
\mathbf{i} \varepsilon_{xy,j} &= \sum_{k=0}^{\bar{k}} \binom{k+j+1}{j+1} \gamma_{k+j}(\theta) \mathbf{1}(k \leq \bar{k}-j) (-\mathbf{i})^k \phi^{(k)}(0).
\end{aligned}$$

■

While the following Lemma may seem familiar, we were not able to find this result at the required level of generality in the existing literature (Theorem 1 and 3 in Andrews (1995) and Theorem 2.8 in Pagan and Ullah (1999) come very close, however).

Lemma 6 *Under Assumptions 7, 16 and 17*

$$\sup_{\alpha \in \mathcal{A}} \sup_{z \in \mathbb{R}} |\tilde{p}(z|\alpha) - p(z|\alpha)| = O_p\left(n^{-1/2}h^{-1}\right) + O\left(h^{N_k}\right)$$

where $\tilde{p}(z|\alpha) = (nh)^{-1} \sum_{j=1}^n K((z_j - z)/h)$ and $p(z|\alpha)$ is the density of $z = X(w, \alpha)$ for a given function $X(w, \alpha)$ of some random vector w . The same result holds with $\tilde{p}(z|\alpha)$ replaced by $\hat{p}(z|\alpha) = (nh)^{-1} \sum_{j=1}^n K((z_j - z)/h) 1(z_i \neq z)$.

Proof. This proof is based in part on the proof of Theorem 2.8 in Pagan and Ullah (1999). Note that $\sup_{\alpha \in \mathcal{A}} \sup_{z \in \mathbb{R}} |\tilde{p}(z|\alpha) - p(z|\alpha)| \leq R + B$, where

$$\begin{aligned} R &= \sup_{\alpha \in \mathcal{A}} \sup_{z \in \mathbb{R}} |\tilde{p}(z|\alpha) - E[\tilde{p}(z|\alpha)]|, \\ B &= \sup_{\alpha \in \mathcal{A}} \sup_{z \in \mathbb{R}} |E[\tilde{p}(z|\alpha)] - p(z|\alpha)|. \end{aligned}$$

By the convolution Theorem,

$$\begin{aligned} R &= \sup_{\alpha \in \mathcal{A}} \sup_{z \in \mathbb{R}} \left| \int \kappa(h\zeta) n^{-1} \sum_{j=1}^n (e^{i\zeta z_j} - E[e^{i\zeta z_j}]) e^{-i\zeta z} d\zeta \right| \\ &\leq \sup_{\alpha \in \mathcal{A}} \sup_{z \in \mathbb{R}} \int |\kappa(h\zeta)| \left| n^{-1} \sum_{j=1}^n (e^{i\zeta z_j} - E[e^{i\zeta z_j}]) \right| d\zeta \\ &= \sup_{\alpha \in \mathcal{A}} \int |\kappa(h\zeta)| \left| n^{-1} \sum_{j=1}^n (e^{i\zeta z_j} - E[e^{i\zeta z_j}]) \right| d\zeta \end{aligned}$$

where $\kappa(\zeta)$ denotes the Fourier transform of $K(z)$. We then have

$$\begin{aligned} E[R] &\leq \sup_{\alpha \in \mathcal{A}} \int |\kappa(h\zeta)| E \left[\left| n^{-1} \sum_{j=1}^n (e^{i\zeta z_j} - E[e^{i\zeta z_j}]) \right| \right] d\zeta \\ &\leq \sup_{\alpha \in \mathcal{A}} \int |\kappa(h\zeta)| \left(E \left[\left| n^{-1} \sum_{j=1}^n (e^{i\zeta z_j} - E[e^{i\zeta z_j}]) \right|^2 \right] \right)^{1/2} d\zeta \\ &= \sup_{\alpha \in \mathcal{A}} \int |\kappa(h\zeta)| \left(n^{-1} E \left[(e^{i\zeta z_j} - E[e^{i\zeta z_j}]) (e^{-i\zeta z_j} - E[e^{-i\zeta z_j}]) \right] \right)^{1/2} d\zeta \\ &= \sup_{\alpha \in \mathcal{A}} n^{-1/2} \int |\kappa(h\zeta)| \left(E \left[(e^{i\zeta z_j} - E[e^{i\zeta z_j}]) (e^{-i\zeta z_j} - E[e^{-i\zeta z_j}]) \right] \right)^{1/2} d\zeta \\ &\leq n^{-1/2} 2^{1/2} \int |\kappa(h\zeta)| d\zeta \\ &= n^{-1/2} h^{-1} 2^{1/2} \int |\kappa(\zeta)| d\zeta \\ &= O\left(n^{-1/2} h^{-1}\right) \end{aligned}$$

and $R = O_p(n^{-1/2}h^{-1})$ by Markov's inequality.

By the convolution Theorem once again,

$$B = \sup_{\alpha \in \mathcal{A}} \sup_{z \in \mathbb{R}} \left| \int (1 - \kappa(h\zeta)) \pi_\alpha(\zeta) e^{-i\zeta z} d\zeta \right|$$

where $\pi_\alpha(\zeta)$ denotes the Fourier transform of $p_{z|\alpha}(z|\alpha)$ with respect to z . By a Taylor expansion,

$$\begin{aligned} B &= \sup_{\alpha \in \mathcal{A}} \sup_{z \in \mathbb{R}} \left| \int \left(\sum_{l=1}^{N_k-1} \frac{1}{l!} \frac{d^l \kappa(0)}{d\zeta^l} (h\zeta)^l + \frac{1}{N_k!} \frac{d^{N_k} \kappa(\bar{\zeta})}{d\zeta^{N_k}} (h\zeta)^{N_k} \right) \pi_\alpha(\zeta) e^{-i\zeta z} d\zeta \right| \text{ for some } \bar{\zeta} \in [0, \zeta] \\ &= \sup_{\alpha \in \mathcal{A}} \sup_{z \in \mathbb{R}} \left| \int \frac{1}{N_k!} \frac{d^{N_k} \kappa(\bar{\zeta})}{d\zeta^{N_k}} (h\zeta)^{N_k} \pi_\alpha(\zeta) e^{-i\zeta z} d\zeta \right|, \end{aligned}$$

since $\frac{d^l \kappa(0)}{d\zeta^l} = 0$ by the Moment Theorem and Assumption 16 (iii). Next,

$$\begin{aligned} B &\leq h^{N_k} \frac{1}{N_k!} \sup_{\alpha \in \mathcal{A}} \int \left| \frac{d^{N_k} \kappa(\bar{\zeta})}{d\zeta^{N_k}} \right| |\zeta|^{N_k} |\pi_\alpha(\zeta)| d\zeta \\ &\leq Ch^{N_k} \end{aligned}$$

for some $C < \infty$ since $\left| \frac{d^{N_k} \kappa(\bar{\zeta})}{d\zeta^{N_k}} \right| \leq C_2 < \infty$ by the Moment Theorem and Assumption 16 (iv) and $\int |\zeta|^{N_k} |\pi_\alpha(\zeta)| d\zeta < \infty$ by Assumption 17.

The second assertion is shown by noting that the difference between $\tilde{p}(z)$ and $\hat{p}(z)$ is at most $K(0)n^{-1}h^{-1}$ which is of an order less than $n^{-1/2}h^{-1}$ and can therefore be absorbed in the $O_p(n^{-1/2}h^{-1})$ remainder. ■

Proof of Theorem 3. Let $Q(\tilde{x}, \tilde{y}, \tilde{w}, \theta, \alpha)$, $Y(\tilde{x}, \tilde{y}, \tilde{w}, \theta, \alpha)$, $\hat{Q}(\theta, \alpha)$ and $Q(\theta, \alpha)$ be as defined in Section 5.1. We first show consistency of $\hat{\theta}$. This involves establishing the uniform convergence of $\hat{Q}(\theta, \hat{\alpha})$ to $Q(\theta, \alpha^*)$ for $\theta \in \Theta$. We first note that $\hat{\alpha} \xrightarrow{P} \alpha^*$, by Lemma 2.4 and Theorem 2.1 in Newey and McFadden (1994), under Assumptions 7 and 8. Hence $\hat{\alpha} \in \mathcal{A}$ with probability approaching 1 (hereafter w.p.a.1). We can then write, w.p.a.1,

$$\begin{aligned} \sup_{\theta \in \Theta} \left\| \hat{Q}(\theta, \hat{\alpha}) - Q(\theta, \alpha^*) \right\| &\leq \sup_{\theta \in \Theta} \left\| \hat{Q}(\theta, \hat{\alpha}) - Q(\theta, \hat{\alpha}) \right\| + \sup_{\theta \in \Theta} \left\| Q(\theta, \hat{\alpha}) - Q(\theta, \alpha^*) \right\| \\ &\leq \sup_{\alpha \in \mathcal{A}} \sup_{\theta \in \Theta} \left\| \hat{Q}(\theta, \alpha) - Q(\theta, \alpha) \right\| + \sup_{\theta \in \Theta} \left\| Q(\theta, \hat{\alpha}) - Q(\theta, \alpha^*) \right\| \end{aligned}$$

where $\sup_{\theta \in \Theta} \left\| Q(\theta, \hat{\alpha}) - Q(\theta, \alpha^*) \right\| \xrightarrow{P} 0$ by $\hat{\alpha} \xrightarrow{P} \alpha^*$ and Assumption 20. Next,

$$\sup_{\alpha \in \mathcal{A}} \sup_{\theta \in \Theta} \left\| \hat{Q}(\theta, \alpha) - Q(\theta, \alpha) \right\| \leq R_A + R_I + R_D$$

where

$$\begin{aligned} R_A &= \sup_{\alpha \in \mathcal{A}} \sup_{\theta \in \Theta} \left\| n^{-1} \sum_{j=1}^n \frac{Y(x_j, y_j, w_j, \theta, \alpha)}{p(X(w_j, \alpha) | \alpha)} - E \left[\frac{Y(x, y, w, \theta, \alpha)}{p(X(w, \alpha) | \alpha)} \right] \right\| \\ R_I &= \sup_{\alpha \in \mathcal{A}} \sup_{\theta \in \Theta} \left\| n^{-1} \sum_{j=1}^n \frac{Y(x_j, y_j, w_j, \theta, \alpha)}{p(X(w_j, \alpha) | \alpha)} (1(\hat{p}(X(w_j, \alpha) | \alpha) \geq \tau) - 1) \right\| \\ R_D &= \sup_{\alpha \in \mathcal{A}} \sup_{\theta \in \Theta} \left\| n^{-1} \sum_{j=1}^n Y(x_j, y_j, w_j, \theta, \alpha) \left(\frac{p(X(w_j, \alpha) | \alpha) - \hat{p}(X(w_j, \alpha) | \alpha)}{\hat{p}(X(w_j, \alpha) | \alpha) p(X(w_j, \alpha) | \alpha)} \right) 1(\hat{p}(X(w_j, \alpha) | \alpha) \geq \tau) \right\|. \end{aligned}$$

We then have $\sup_{\theta \in \Theta} \|R_A\| \xrightarrow{p} 0$ by Assumptions 7, 11 and 12 and Lemma 2.4 in Newey and McFadden (1994). Next, by Lemma 6, we have

$$\begin{aligned}
R_I &\leq \sup_{\alpha \in \mathcal{A}} \sup_{\theta \in \Theta} n^{-1} \sum_{j=1}^n \frac{\|Y(x_j, y_j, w_j, \theta, \alpha)\|}{p(X(w_j, \alpha) | \alpha)} |1(\hat{p}(X(w_j, \alpha) | \alpha) < \tau)| \\
&\leq \sup_{\alpha \in \mathcal{A}} \sup_{\theta \in \Theta} n^{-1} \sum_{j=1}^n \frac{\|Y(x_j, y_j, w_j, \theta, \alpha)\|}{p(X(w_j, \alpha) | \alpha)} \left|1\left(p(X(w_j, \alpha) | \alpha) - Cn^{\epsilon-1/2}h^{-1} < \tau\right)\right| \text{ w.p.a. 1 for } \epsilon \in]0, 1/4[\\
&= \sup_{\alpha \in \mathcal{A}} \sup_{\theta \in \Theta} n^{-1} \sum_{j=1}^n \frac{\|Y(x_j, y_j, w_j, \theta, \alpha)\|}{p(X(w_j, \alpha) | \alpha)} \left|1\left(p(X(w_j, \alpha) | \alpha) < \tau\left(1 + Cn^{\epsilon-1/2}h^{-1}/\tau\right)\right)\right| \\
&\leq \sup_{\alpha \in \mathcal{A}} \sup_{\theta \in \Theta} n^{-1} \sum_{j=1}^n \frac{\|Y(x_j, y_j, w_j, \theta, \alpha)\|}{p(X(w_j, \alpha) | \alpha)} |1(p(X(w_j, \alpha) | \alpha) < 2\tau)| \text{ by Assumption 18}
\end{aligned}$$

and $E[R_I] = E\left[\sup_{\alpha \in \mathcal{A}} \sup_{\theta \in \Theta} \frac{\|Y(x_j, y_j, w_j, \theta, \alpha)\|}{p(X(w_j, \alpha) | \alpha)} |1(p(z_j) < 2\tau)|\right] = o(n^{-1/2})$ by Assumption 19, thus implying that $R_I = o_p(n^{-1/2})$, by Markov's inequality. Next,

$$\begin{aligned}
R_D &\leq \sup_{\alpha \in \mathcal{A}} \sup_{\theta \in \Theta} n^{-1} \sum_{j=1}^n \|Y(x_j, y_j, w_j, \theta, \alpha)\| \left(\frac{|p(X(w_j, \alpha) | \alpha) - \hat{p}(X(w_j, \alpha) | \alpha)|}{\hat{p}(z_j) p(X(w_j, \alpha) | \alpha)}\right) \hat{I}_j \\
&\leq \sup_{\alpha \in \mathcal{A}} \sup_{\theta \in \Theta} \tau^{-1} n^{-1} \sum_{j=1}^n \|Y(x_j, y_j, w_j, \theta, \alpha)\| \left(\frac{|p(X(w_j, \alpha) | \alpha) - \hat{p}(X(w_j, \alpha) | \alpha)|}{p(X(w_j, \alpha) | \alpha)}\right) \hat{I}_j \\
&\leq \sup_{z \in \mathbb{R}} |p(z) - \hat{p}(z)| \tau^{-1} n^{-1} \sum_{j=1}^n \left(\frac{\|Y(x_j, y_j, w_j, \theta, \alpha)\|}{p(X(w_j, \alpha) | \alpha)}\right) \\
&= \left(O_p\left(n^{-1/2}h^{-1}\right) + O\left(h^{N\kappa}\right)\right) \tau^{-1} O_p(1)
\end{aligned}$$

by Lemma 6, and Lemma 2.4 in Newey and McFadden (1994) under Assumptions 7, 11 and 12. By Assumption 18, $n^{-1/2}h^{-1}\tau^{-1} \rightarrow 0$ and $h^{N\kappa} \rightarrow 0$ and it follows that $R_D \xrightarrow{p} 0$.

Having shown that $\sup_{\theta \in \Theta} \left\|\hat{Q}(\theta, \hat{\alpha}) - Q(\theta, \alpha^*)\right\| \xrightarrow{p} 0$, we now establish that this implies²⁵ that $\hat{\theta}$ converges to θ^* . Since $\hat{Q}(\hat{\theta}, \hat{\alpha}) = 0$ and $\sup_{\theta \in \Theta} \left\|\hat{Q}(\theta, \hat{\alpha}) - Q(\theta, \alpha^*)\right\| \xrightarrow{p} 0$ it follows that $\text{plim}_{n \rightarrow \infty} Q(\hat{\theta}, \alpha^*) = 0$. Since $\hat{Q}(\theta, \hat{\alpha})$ is continuous in θ (because $Y(x_j, y_j, w_j, \theta, \alpha)$ is), and its convergence to $Q(\theta, \alpha^*)$ is uniform in θ , $Q(\theta, \alpha^*)$ must be continuous in θ . Combining these two results yields $\text{plim}_{n \rightarrow \infty} Q(\hat{\theta}, \alpha^*) = Q(\text{plim}_{n \rightarrow \infty} \hat{\theta}, \alpha^*) = 0$. Since $\theta = \theta^*$ is the only solution to $Q(\theta, \alpha^*) = 0$ by Assumption 9, we conclude that $\text{plim}_{n \rightarrow \infty} \hat{\theta} = \theta^*$.

Having shown consistency, we turn to asymptotic normality and root n consistency. By a standard mean value expansion of the first-order conditions $\hat{Q}(\hat{\theta}, \hat{\alpha}) = 0$ around θ^* and the usual manipulations,

$$n^{1/2}(\hat{\theta} - \theta^*) = - \left(\frac{\partial \hat{Q}(\bar{\theta}, \hat{\alpha})}{\partial \theta'}\right)^{-1} n^{1/2} \hat{Q}(\theta^*, \hat{\alpha}), \tag{104}$$

for some mean value $\bar{\theta}$. Following the same steps as used above to show uniform convergence in probability of $\hat{Q}(\theta, \hat{\alpha})$, we can show that $\sup_{\theta \in \mathcal{N}} \left\|\frac{\partial \hat{Q}(\bar{\theta}, \hat{\alpha})}{\partial \theta'} - \frac{\partial Q(\theta, \alpha^*)}{\partial \theta'}\right\| \xrightarrow{p} 0$ and $\frac{\partial Q(\theta, \alpha^*)}{\partial \theta'}$ is continuous in θ , by simply

²⁵This would be obvious if $\hat{\theta}$ were defined as the maximizer of a random function. Here $\hat{\theta}$ is the solution to a set of equations and the usual consistency result (e.g. Theorem 2.1 in Newey and McFadden (1994)) does not directly apply.

replacing Assumption 12 by Assumption 13. Since $\hat{\theta} \xrightarrow{p} \theta^*$ it follows that $\bar{\theta} \xrightarrow{p} \theta^*$ and that $\frac{\partial Q(\bar{\theta}, \alpha^*)}{\partial \theta'} \xrightarrow{p} \frac{\partial Q(\theta^*, \alpha^*)}{\partial \theta'}$, thus implying that

$$\frac{\partial \hat{Q}(\bar{\theta}, \hat{\alpha})}{\partial \theta'} \xrightarrow{p} \frac{\partial Q(\theta^*, \alpha^*)}{\partial \theta'}. \quad (105)$$

Next, we let $Y_j = Y(x_j, y_j, w_j, \theta^*, \alpha^*)$, $z_j = X(w_j, \alpha^*)$, $\hat{p}(z_j) = \hat{p}(X(w_j, \alpha^*) | \alpha^*)$, $p(z_j) = p(X(w_j, \alpha^*) | \alpha^*)$, $\hat{I}_j = 1(\hat{p}(z_j) \geq \tau)$, $I_j = 1(p(z_j) \geq \tau)$ and decompose the term $n^{1/2} \hat{Q}(\theta^*, \hat{\alpha})$ in Equation 104 as

$$n^{1/2} \hat{Q}(\theta^*, \hat{\alpha}) = N + N_\alpha + R_{T1} + R_{T2} + R_{T3} + R_L + R_U + R_B + R_{\text{sec}}$$

where the asymptotically normal terms are given by

$$\begin{aligned} N &= n^{-1/2} \sum_{j=1}^n \frac{Y_j - E[Y_j | z_j]}{p(z_j)} \\ N_\alpha &= n^{1/2} (Q(\theta^*, \hat{\alpha}) - Q(\theta^*, \alpha^*)) \end{aligned}$$

while the remainder terms associated with trimming are

$$\begin{aligned} R_{T1} &= n^{-1/2} \sum_{j=1}^n \frac{Y_j}{\hat{p}(z_j)} (\hat{I}_j - I_j) \\ R_{T2} &= n^{1/2} E \left[\frac{Y_j}{p(z_j)} (1 - I_j) \right] \\ R_{T3} &= n^{-1/2} \sum_{j=1}^n \frac{(Y_j - E[Y_j | z_j])}{p(z_j)} (I_j - 1) \end{aligned}$$

the remainder from the linearization is given by

$$R_L = n^{-1/2} \sum_{j=1}^n \frac{Y_j}{\hat{p}(z_j) p^2(z_j)} (\hat{p}(z_j) - p(z_j))^2 I_j$$

the “ U -statistic” term is

$$R_U = -n^{-1/2} \sum_{j=1}^n \left(\frac{Y_j}{p^2(z_j)} (\hat{p}(z_j) - E[\hat{p}(z_j) | z_j]) I_j - \left(\frac{E[Y_j | z_j]}{p(z_j)} I_j - E \left[\frac{Y_j}{p(z_j)} I_j \right] \right) \right),$$

the “bias” term is

$$R_B = n^{-1/2} \sum_{j=1}^n \frac{Y_j}{p^2(z_j)} (p(z_j) - E[\hat{p}(z_j) | z_j]) I_j$$

and the “stochastic equicontinuity” remainder term is

$$R_{\text{sec}} = n^{1/2} \left(\left(\hat{Q}(\theta^*, \hat{\alpha}) - Q(\theta^*, \hat{\alpha}) \right) - \left(\hat{Q}(\theta^*, \alpha^*) - Q(\theta^*, \alpha^*) \right) \right).$$

We consider each remainder in turn.

$$\begin{aligned}
|R_{T1}| &\leq n^{-1/2} \sum_{j=1}^n \frac{|Y_j|}{\hat{p}(z_j)} |1(\hat{p}(z_j) \geq \tau) - 1(p(z_j) \geq \tau)| \\
&\leq n^{-1/2} \sum_{j=1}^n \frac{|Y_j|}{p(z_j) - Cn^{\epsilon-1/2}h^{-1}} |1(\hat{p}(z_j) \geq \tau) - 1(p(z_j) \geq \tau)| \text{ w.p.a. 1 for } \epsilon \in]0, 1/4[\\
&\leq n^{-1/2} \sum_{j=1}^n \frac{|Y_j|}{p(z_j) - \frac{p(z_j)}{\tau - Cn^{\epsilon-1/2}h^{-1}} Cn^{\epsilon-1/2}h^{-1}} |1(\hat{p}(z_j) \geq \tau) - 1(p(z_j) \geq \tau)| \text{ w.p.a. 1} \\
&= \frac{1}{1 - \frac{1}{C\tau n^{1/2-\epsilon}h^{-1}}} n^{-1/2} \sum_{j=1}^n \frac{|Y_j|}{p(z_j)} |1(\hat{p}(z_j) \geq \tau) - 1(p(z_j) \geq \tau)| \\
&= o(1) n^{-1/2} \sum_{j=1}^n \frac{|Y_j|}{p(z_j)} |1(\hat{p}(z_j) \geq \tau \text{ and } p(z_j) < \tau) - 1(p(z_j) \geq \tau \text{ and } \hat{p}(z_j) < \tau)| \\
&= o(1) n^{-1/2} \sum_{j=1}^n \frac{|Y_j|}{p(z_j)} 1(\hat{p}(z_j) \geq \tau \text{ and } p(z_j) < \tau) + o(1) n^{-1/2} \sum_{j=1}^n \frac{|Y_j|}{p(z_j)} 1(p(z_j) \geq \tau \text{ and } \hat{p}(z_j) < \tau) \\
&\leq o(1) n^{-1/2} \sum_{j=1}^n \frac{|Y_j|}{p(z_j)} 1(p(z_j) < \tau) + o(1) n^{-1/2} \sum_{j=1}^n \frac{|Y_j|}{p(z_j)} 1(\hat{p}(z_j) < \tau) \\
&\leq o(1) n^{-1/2} \sum_{j=1}^n \frac{|Y_j|}{p(z_j)} 1(p(z_j) < \tau) + o(1) n^{-1/2} \sum_{j=1}^n \frac{|Y_j|}{p(z_j)} 1(p(z_j) < \tau - Cn^{\epsilon-1/2}h^{-1}) \text{ w.p.a. 1}
\end{aligned}$$

where

$$\begin{aligned}
E \left[n^{-1/2} \sum_{j=1}^n \frac{|Y_j|}{p(z_j)} 1(p(z_j) < \tau) \right] &= n^{1/2} E \left[\frac{|Y_j|}{p(z_j)} 1(p(z_j) < \tau) \right] = o(1) \\
E \left[n^{-1/2} \sum_{j=1}^n \frac{|Y_j|}{p(z_j)} 1(p(z_j) < \tau - Cn^{\epsilon-1/2}h^{-1}) \right] &= n^{1/2} E \left[\frac{|Y_j|}{p(z_j)} 1(p(z_j) < \tau - Cn^{\epsilon-1/2}h^{-1}) \right] \\
&= n^{1/2} E \left[\frac{|Y_j|}{p(z_j)} 1(p(z_j) < \tau (1 - Cn^{\epsilon-1/2}h^{-1}/\tau)) \right] \\
&= n^{1/2} E \left[\frac{|Y_j|}{p(z_j)} 1(p(z_j) < \tau (1 - o(1))) \right] \\
&\rightarrow n^{1/2} E \left[\frac{|Y_j|}{p(z_j)} 1(p(z_j) < \tau) \right] = o(1)
\end{aligned}$$

and by Markov's inequality $R_{T1} = o_p(1)$. Next,

$$\begin{aligned}
|R_{T2}| &\leq n^{1/2} E \left[\frac{|Y_j|}{p(z_j)} |I_j - 1| \right] \\
&= n^{1/2} E \left[\frac{|Y_j|}{p(z_j)} 1(p(z_j) \leq \tau) \right] \\
&= n^{1/2} o(n^{-1/2}) = o(1)
\end{aligned}$$

and

$$\begin{aligned}
E[|R_{T3}|] &= E \left[\left| n^{-1/2} \sum_{j=1}^n \frac{Y_j - E[Y_j|z_j]}{p(z_j)} (I_j - 1) \right| \right] \\
&\leq n^{1/2} 2E \left[\frac{|Y_j|}{p(z_j)} |I_j - 1| \right] \\
&= n^{1/2} o(n^{-1/2}) = o(1)
\end{aligned}$$

implying that $|R_{T3}| = o_p(1)$ as well by the Markov inequality. The linearization remainder is then

$$\begin{aligned}
|R_L| &= \left| n^{-1/2} \sum_{j=1}^n \frac{Y_j}{\hat{p}(z_j) p^2(z_j)} (\hat{p}(z_j) - p(z_j))^2 I_j \right| \\
&\leq n^{-1/2} \sum_{j=1}^n \frac{|Y_j|}{\hat{p}(z_j) p^2(z_j)} |\hat{p}(z_j) - p(z_j)|^2 I_j \\
&\leq n^{-1/2} \sum_{j=1}^n \frac{|Y_j|}{(p(z_j) - Cn^{-1/2}h^{-1}) p^2(z_j)} |\hat{p}(z_j) - p(z_j)|^2 I_j \\
&\leq n^{-1/2} \sum_{j=1}^n \frac{|Y_j|}{(\tau - Cn^{-1/2}h^{-1}) \tau p(z_j)} |\hat{p}(z_j) - p(z_j)|^2 I_j \\
&= \frac{1}{\tau^2} (1 - Cn^{-1/2}h^{-1}/\tau)^{-1} n^{-1/2} \sum_{j=1}^n \frac{|Y_j|}{p(z_j)} |\hat{p}(z_j) - p(z_j)|^2 I_j \\
&\leq \frac{2}{\tau^2} n^{-1/2} \sum_{j=1}^n \frac{|Y_j|}{p(z_j)} |\hat{p}(z_j) - p(z_j)|^2 \\
&\leq \frac{2Cn^{-1}h^{-2}}{\tau^2} n^{1/2} n^{-1} \sum_{j=1}^n \frac{|Y_j|}{p(z_j)} \\
&\leq \frac{2Cn^{-1}h^{-2}}{\tau^2} n^{1/2} \left(n^{-1} \sum_{j=1}^n \frac{|Y_j|^2}{p^2(z_j)} \right)^{1/2} \\
&= \frac{2Cn^{-1}h^{-2}}{\tau^2} n^{1/2} O_p(1) \\
&= o(n^{-1/2}) n^{1/2} O_p(1) \\
&= o_p(1)
\end{aligned}$$

The “ U -statistic” term can be written as

$$\begin{aligned}
-R_U &= n^{-1/2} \sum_{j=1}^n (n-1)^{-1} \times \\
&\quad \times \sum_{i \neq j} \left(\frac{Y_j I_j}{p^2(z_j)} (K_h(z_i - z_j) - E[K_h(z_i - z_j) | z_j]) - \left(\frac{E[Y_j | z_j]}{p(z_j)} I_j - E \left[\frac{Y_j}{p(z_j)} I_j \right] \right) \right) \\
&= n^{-1/2} \sum_{j=1}^n (n-1)^{-1} \times \\
&\quad \times \sum_{i \neq j} \left(\left(\frac{Y_j I_j}{2p^2(z_j)} + \frac{Y_i I_i}{2p^2(z_i)} \right) (K_h(z_i - z_j) - E[K_h(z_i - z_j) | z_j]) - \left(\frac{E[Y_i | z_i]}{p(z_i)} I_i - E \left[\frac{Y_i}{p(z_i)} I_i \right] \right) \right) \\
&= n^{1/2} \binom{n}{2}^{-1} \sum_{j=1}^n \sum_{i=j+1}^n U((Y_j, z_j), (Y_i, z_i))
\end{aligned}$$

where $K_h(z) = h^{-1}K(z/h)$ and

$$U((Y_j, z_j), (Y_i, z_i)) = \left(\frac{Y_j I_j}{2p^2(z_j)} + \frac{Y_i I_i}{2p^2(z_i)} \right) (K_h(z_i - z_j) - E[K_h(z_i - z_j) | z_j]) - \left(\frac{E[Y_i | z_i]}{p(z_i)} I_i - E \left[\frac{Y_i}{p(z_i)} I_i \right] \right).$$

Using the “ U -statistic” projection Theorem (e.g. Lemma 3.1 in Powell, Stock, and Stoker (1989)), standard but tedious manipulations show that $R_U = o_p(1)$ under Assumptions 16 and 18. Finally, the bias term is

$$\begin{aligned}
|R_B| &\leq n^{-1/2} \sum_{j=1}^n \frac{|Y_j|}{p^2(z_j)} |p(z_j) - E[\hat{p}(z_j) | z_j]| I_j \\
&\leq \tau^{-1} n^{-1/2} \sum_{j=1}^n \frac{|Y_j|}{p(z_j)} |p(z_j) - E[\hat{p}(z_j) | z_j]| I_j \\
&\leq \tau^{-1} n^{-1/2} \sum_{j=1}^n \frac{|Y_j|}{p(z_j)} |p(z_j) - E[\hat{p}(z_j) | z_j]| \\
&\leq \tau^{-1} n^{-1/2} \sum_{j=1}^n \frac{|Y_j|}{p(z_j)} C h^{N_k} \text{ by Lemma 6} \\
&= \tau^{-1} n^{1/2} O_p(1) h^{N_k}
\end{aligned}$$

and $|R_B| = O_p(n^{1/2} h^{N_k} \tau^{-1}) = o_p(1)$ since $n^{1/2} h^{N_k} \tau^{-1} \rightarrow 0$ by Assumption 18.

To bound the R_{sec} term, let $S_\tau(t)$ be continuously differentiable in t for all $\tau \neq 0$ and such that (i) $1(t \geq \tau) = 0 \Leftrightarrow S_\tau(t) = 0$ (ii) $1(t \geq \tau) = 1 \Leftrightarrow S_{2\tau}(t) = 1$ (iii) $0 \leq S_\tau(t) \leq 1$. (iv) $\sup_{t \in \mathbb{R}} |dS_\tau(t)/dt| = O(\tau)$. We then decompose $\hat{Q}(\theta^*, \alpha)$ as

$$\hat{Q}(\theta^*, \alpha) = \hat{Q}_S(\theta^*, \alpha) + R_S(\alpha)$$

where $\hat{Q}_S(\theta^*, \alpha)$ is continuous in α while $R_S(\alpha)$ may not be and are given by

$$\begin{aligned}
\hat{Q}_S(\theta^*, \alpha) &= n^{-1} \sum_{j=1}^n \frac{Y(x_j, y_j, w_j, \theta^*, \alpha)}{\hat{p}(X(w_j, \alpha) | \alpha)} S_\tau(\hat{p}(X(w_j, \alpha) | \alpha)) \\
R_S(\alpha) &= n^{-1} \sum_{j=1}^n \frac{Y(x_j, y_j, w_j, \theta^*, \alpha)}{\hat{p}(X(w_j, \alpha) | \alpha)} (1(\hat{p}(X(w_j, \alpha) | \alpha) \geq \tau) - S_\tau(\hat{p}(X(w_j, \alpha) | \alpha))).
\end{aligned}$$

The remainder $R_S(\alpha)$ satisfies $\sup_{\alpha \in \mathcal{A}} \|R_S(\alpha)\| = o_p(n^{-1/2})$ since

$$\sup_{\alpha \in \mathcal{A}} \|R_S(\alpha)\| \leq \sup_{\alpha \in \mathcal{A}} n^{-1} \sum_{j=1}^n \frac{\|Y(x_j, y_j, w_j, \theta^*, \alpha)\|}{p(X(w_j, \alpha) | \alpha) - n^{-1/2} h^{-1}} 1(\hat{p}(X(w_j, \alpha) | \alpha) < 2\tau) = o_p(n^{-1/2})$$

By Assumption 19 and the same arguments as used for R_{T1} . We can then write R_{sec} as

$$\begin{aligned} R_{\text{sec}} &= n^{1/2} \left(\left(\hat{Q}(\theta^*, \hat{\alpha}) - Q(\theta^*, \hat{\alpha}) \right) - \left(\hat{Q}(\theta^*, \alpha^*) - Q(\theta^*, \alpha^*) \right) \right) \\ &= n^{1/2} \left(\left(\hat{Q}_S(\theta^*, \hat{\alpha}) - Q(\theta^*, \hat{\alpha}) \right) - \left(\hat{Q}_S(\theta^*, \alpha^*) - Q(\theta^*, \alpha^*) \right) \right) + o_p(1) \\ &= \left(\frac{\partial \hat{Q}(\theta^*, \bar{\alpha})}{\partial \alpha'} - \frac{\partial Q(\theta^*, \bar{\alpha})}{\partial \alpha'} \right) n^{1/2} (\hat{\alpha} - \alpha^*) + o_p(1) \end{aligned} \quad (106)$$

for some mean value $\bar{\alpha}$. We then decompose $\frac{\partial}{\partial \alpha'} \hat{Q}_S(\theta^*, \alpha)$ as

$$\frac{\partial}{\partial \alpha'} \hat{Q}_S(\theta^*, \alpha) = D_1 + D_2 + R_{DS}$$

where

$$\begin{aligned} D_1 &= n^{-1/2} \sum_{j=1}^n \left(\frac{\frac{\partial}{\partial \alpha'} Y(x_j, y_j, w_j, \theta, \alpha)}{\hat{p}(X(w_j, \alpha) | \alpha)} \right) S_\tau(\hat{p}(X(w_j, \alpha) | \alpha)) \\ D_2 &= -n^{-1/2} \sum_{j=1}^n \left(\frac{Y(x_j, y_j, w_j, \theta, \alpha)}{\hat{p}^2(X(w_j, \alpha) | \alpha)} \frac{\partial}{\partial \alpha'} \hat{p}(X(w_j, \alpha) | \alpha) \right) S_\tau(\hat{p}(X(w_j, \alpha) | \alpha)) \\ R_{DS} &= n^{-1/2} \sum_{j=1}^n \frac{Y(x_j, y_j, w_j, \theta, \alpha)}{\hat{p}(X(w_j, \alpha) | \alpha)} \frac{\partial S_\tau(\hat{p}(X(w_j, \alpha) | \alpha))}{\partial \alpha'}. \end{aligned}$$

The R_{DS} term is negligible, since

$$\begin{aligned} & \left\| n^{-1/2} \sum_{j=1}^n \frac{Y(x_j, y_j, w_j, \theta, \alpha)}{\hat{p}(X(w_j, \alpha) | \alpha)} \frac{\partial S_\tau(\hat{p}(X(w_j, \alpha) | \alpha))}{\partial \alpha'} \right\| \\ & \leq n^{-1/2} \sum_{j=1}^n \frac{\|Y(x_j, y_j, w_j, \theta, \alpha)\|}{\tau} \left| \frac{\partial S_\tau(\hat{p}(X(w_j, \alpha) | \alpha))}{\partial \alpha'} \right| \\ & \leq n^{-1/2} \sum_{j=1}^n \frac{\|Y(x_j, y_j, w_j, \theta, \alpha)\|}{\tau} C\tau 1(p(X(w_j, \alpha) | \alpha) > \tau) \\ & = Cn^{-1/2} \sum_{j=1}^n \|Y(x_j, y_j, w_j, \theta, \alpha)\| 1(p(X(w_j, \alpha) | \alpha) > \tau) \\ & = Cn^{-1/2} \sum_{j=1}^n \frac{\|Y(x_j, y_j, w_j, \theta, \alpha)\|}{p(X(w_j, \alpha) | \alpha)} p(X(w_j, \alpha) | \alpha) 1(p(X(w_j, \alpha) | \alpha) > \tau) \\ & \leq Cn^{1/2} n^{-1} \sum_{j=1}^n \frac{\|Y(x_j, y_j, w_j, \theta, \alpha)\|}{p(X(w_j, \alpha) | \alpha)} 1(p(X(w_j, \alpha) | \alpha) > \tau) \text{ since } p(z | \alpha) \text{ is bounded by Assumption 17} \\ & = n^{1/2} o_p(n^{-1/2}) = o_p(1) \text{ by Markov's inequality and Assumption 19.} \end{aligned}$$

Now, the terms D_1 and D_2 can be handled through the same techniques as the ones used to show uniform convergence of $\hat{Q}(\theta, \hat{\alpha})$ after noting that trimming by $S_\tau(\hat{p}(X(w_j, \alpha) | \alpha))$ is asymptotically equivalent to

trimming by $1(\hat{p}(X(w_j, \alpha) | \alpha) \geq \tau)$. Under Assumption 21, and by using an expansion of the form

$$\begin{aligned}
D_1 &= n^{-1/2} \sum_{j=1}^n \frac{\frac{\partial}{\partial \alpha'} Y(x_j, y_j, w_j, \theta, \alpha)}{p(X(w_j, \alpha) | \alpha)} S_\tau(\hat{p}(X(w_j, \alpha) | \alpha)) + \\
&\quad - n^{-1/2} \sum_{j=1}^n \frac{\frac{\partial}{\partial \alpha'} Y(x_j, y_j, w_j, \theta, \alpha)}{p(X(w_j, \alpha) | \alpha)} \frac{(\hat{p}(X(w_j, \alpha) | \alpha) - p(X(w_j, \alpha) | \alpha))}{\hat{p}(X(w_j, \alpha) | \alpha)} S_\tau(\hat{p}(X(w_j, \alpha) | \alpha)) \\
D_2 &= n^{-1/2} \sum_{j=1}^n \frac{Y(x_j, y_j, w_j, \theta, \alpha)}{p(X(w_j, \alpha) | \alpha)} \left(1 - \frac{(\hat{p}(X(w_j, \alpha) | \alpha) - p(X(w_j, \alpha) | \alpha))}{\hat{p}(X(w_j, \alpha) | \alpha)} \right) \times \\
&\quad \times \frac{\frac{\partial}{\partial \alpha'} p(X(w_j, \alpha) | \alpha) + \left(\frac{\partial}{\partial \alpha'} \hat{p}(X(w_j, \alpha) | \alpha) - \frac{\partial}{\partial \alpha'} p(X(w_j, \alpha) | \alpha) \right)}{\hat{p}(X(w_j, \alpha) | \alpha)}
\end{aligned}$$

it can be shown that $D_1 \xrightarrow{p} E \left[\frac{\frac{\partial}{\partial \alpha'} Y(x_j, y_j, w_j, \theta, \alpha)}{p(X(w_j, \alpha) | \alpha)} \right]$ and $D_2 \xrightarrow{p} E \left[\frac{Y(x_j, y_j, w_j, \theta, \alpha)}{p^2(X(w_j, \alpha) | \alpha)} \frac{\partial p(X(w_j, \alpha) | \alpha)}{\partial \alpha'} \right]$ uniformly in α for $\alpha \in \mathcal{A}$. (The convergence rate of $\frac{\partial}{\partial \alpha'} \hat{p}(X(w_j, \alpha) | \alpha) - \frac{\partial}{\partial \alpha'} p(X(w_j, \alpha) | \alpha)$ is obtained as in the proof of Lemma 6, with N_k replaced by $N_k - 1$.) This implies, by Assumption 20, that

$$\sup_{\alpha \in \mathcal{A}} \left(\frac{\partial \hat{Q}(\theta^*, \alpha)}{\partial \alpha'} - \frac{\partial Q(\theta^*, \alpha)}{\partial \alpha'} \right) \xrightarrow{p} 0$$

and by Equation (106) and the fact that $\hat{\alpha} - \alpha^* = O_p(n^{-1/2})$, we have that $R_{\text{sec}} = o_p(1)$.

Having bounded all remainder terms, we note that the N term clearly satisfies

$$N = n^{-1/2} \sum_{j=1}^n \psi_\theta(x_j, y_j, w)$$

where $E[\psi_\theta(x_j, y_j, w) \psi'_\theta(x_j, y_j, w)]$ is finite under Assumption 14.

By a mean-value expansion, the N_α term is equal to

$$N_\alpha = \frac{\partial Q(\theta^*, \bar{\alpha})}{\partial \alpha'} n^{1/2} (\hat{\alpha} - \alpha^*)$$

for some mean value $\bar{\alpha}$. Since $\hat{\alpha} \xrightarrow{p} \alpha^*$ and therefore $\bar{\alpha} \xrightarrow{p} \alpha^*$, Assumption 20 implies that $\frac{\partial Q(\theta^*, \bar{\alpha})}{\partial \alpha'} \xrightarrow{p} \frac{\partial Q(\theta^*, \alpha^*)}{\partial \alpha'}$.

By standard results (such as Theorem 3.1 in Newey and McFadden (1994)), Assumptions 7 and 8 imply that the first-step estimate $\hat{\alpha}$ is a root n consistent estimator of α^* with influence function equal to

$$\psi_\alpha(\tilde{x}, \tilde{w}) = - \left(E \left[\frac{\partial X(w, \alpha^*)}{\partial \alpha} \frac{\partial X(w, \alpha^*)}{\partial \alpha'} \right] \right)^{-1} \frac{\partial X(\tilde{w}, \alpha^*)}{\partial \alpha} (\tilde{x} - X(\tilde{w}, \alpha^*))$$

and such that $E[\psi_\alpha(\tilde{x}, \tilde{w}) \psi'_\alpha(\tilde{x}, \tilde{w})]$ exists. Hence, we can write

$$N_\alpha = n^{-1/2} \sum_{j=1}^n \frac{\partial Q(\theta^*, \alpha)}{\partial \alpha'} \psi_\alpha(x_j, w_j).$$

We have just established that

$$n^{1/2} \hat{Q}(\theta^*, \hat{\alpha}) = n^{-1/2} \sum_{j=1}^n \left(\psi_\theta(x_j, y_j, w_j) + \frac{\partial Q(\theta^*, \alpha)}{\partial \alpha'} \psi_\alpha(x_j, w_j) \right) + o_p(1)$$

and by the finiteness of $E[\psi_\theta(x_j, y_j, w_j) \psi'_\theta(x_j, y_j, w_j)]$ and $E[\psi_\alpha(\tilde{x}, \tilde{w}) \psi'_\alpha(\tilde{x}, \tilde{w})]$, the Cauchy-Schwartz inequality, Assumptions 7 and the Lindeberg-Levy Central Limit Theorem, this sum is asymptotically normal. By Equations (104), (105) and the Slutsky Theorem, the conclusion of the Theorem follows. \blacksquare

Lemma 7 Let $\sigma(\zeta)$ be the Fourier transform of $s(z)$. For $\alpha \in \mathbb{R}^+$ and $\gamma \in \mathbb{N}$, if

$$\sum_{t=0}^{\infty} \frac{\alpha^t}{t!} \int \left| \frac{d^{\gamma t} \sigma(\zeta)}{d\zeta^{\gamma t}} \right| d\zeta < \infty$$

then, for some $C > 0$,

$$|s(z)| \leq C \exp(-\alpha |z|^\gamma).$$

Proof. Let $T(z) = \exp(\alpha z^\gamma)$. Since the radius of convergence of the Taylor series of the exponential function is infinite, we can also write $T(z) = \sum_{t=0}^{\infty} \frac{\alpha^t}{t!} z^{\gamma t}$ for all $z \in \mathbb{R}$. Let Θ denote the linear operator defined by

$$\Theta \sigma(\zeta) = \sum_{t=0}^{\infty} \frac{\alpha^t}{t!} \frac{(-\mathbf{i})^{\gamma t} d^{\gamma t} \sigma(\zeta)}{d\zeta^{\gamma t}}.$$

Since the Fourier transform of $z^t s(z)$ is $(-\mathbf{i})^t \frac{d^t \sigma(\zeta)}{d\zeta^t}$, the Fourier transform of $T(z) s(z)$ is $\Theta \sigma(\zeta)$. We can then write, for $z \geq 0$,

$$\begin{aligned} |s(z)| &= \frac{1}{|T(z)|} |T(z) s(z)| \\ &= \frac{1}{|T(z)|} \left| \int \Theta \sigma(\zeta) e^{-\mathbf{i}\zeta z} d\zeta \right| \\ &\leq \frac{1}{|T(z)|} \int |\Theta \sigma(\zeta)| d\zeta \\ &= \frac{1}{|T(z)|} \int \left| \sum_{t=0}^{\infty} \frac{\alpha^t}{t!} \frac{(-\mathbf{i})^{\gamma t} d^{\gamma t} \sigma(\zeta)}{d\zeta^{\gamma t}} \right| d\zeta \\ &\leq \frac{1}{|T(z)|} \sum_{t=0}^{\infty} \frac{\alpha^t}{t!} \int \left| \frac{d^{\gamma t} \sigma(\zeta)}{d\zeta^{\gamma t}} \right| d\zeta \\ &= \frac{C}{|T(z)|} \\ &= C \exp(-\alpha |z|^\gamma). \end{aligned}$$

with $C = \sum_{t=0}^{\infty} \frac{\alpha^t}{t!} \int \left| \frac{d^{\gamma t} \sigma(\zeta)}{d\zeta^{\gamma t}} \right| d\zeta < \infty$. For $z < 0$, we can similarly write

$$\begin{aligned} |s(z)| &= \frac{1}{|T(-z)|} |T(-z) s(z)| \\ &= \frac{1}{|T(-z)|} \left| \int \Theta \sigma(\zeta) e^{\mathbf{i}\zeta z} d\zeta \right| \\ &\leq \frac{1}{|T(-z)|} \int |\Theta \sigma(\zeta)| d\zeta \\ &\leq \frac{C}{|T(|z|)|} = C \exp(-\alpha |z|^\gamma). \end{aligned}$$

■

Theorem 6 The inverse Fourier transform $s(\zeta)$ of the function $\sigma(\zeta) = \exp(-\cos^{-2}(\zeta)) \mathbf{1}(|\zeta| \leq \pi/2)$ is such that $|s(z)| \leq C \exp(-\alpha |z|)$ for $\alpha \in [0, 1/3[$ and some positive $C < \infty$.

Proof. The proof consists of verifying that $\sigma(\zeta)$ satisfies the hypothesis of Lemma 7. The t -th derivative of $\exp(-\cos^{-2}(\zeta))$ consists of a sum of at most 3^t terms of the form

$$C \exp(-\cos^{-2}(\zeta)) \cos^{-p}(\zeta) \sin^q(\zeta), \quad (107)$$

where $q \geq 0$, $0 \leq p \leq 2t$, and $|C| \leq 1 + t$. Since $p \leq 2t$, $|\sin(\zeta)| \leq 1$ and $X^t \exp(-X) \leq t^t \exp(-t)$ for all $X \in \mathbb{R}^+$ and all $t \in \mathbb{N}$, we have

$$\begin{aligned} & \left| \exp(-\cos^{-2}(\zeta)) \cos^{-p}(\zeta) \sin^q(\zeta) \right| \\ & \leq \exp(-\cos^{-2}(\zeta)) \cos^{-2t}(\zeta) \\ & \leq t^t \exp(-t). \end{aligned}$$

Consequently, for some $C > 0$,

$$\begin{aligned} \sum_{t=0}^{\infty} \frac{\alpha^t}{t!} \int \left| \frac{d^t \sigma(\zeta)}{d\zeta^t} \right| d\zeta & \leq C \sum_{t=0}^{\infty} \frac{\alpha^t}{t!} 3^t (1+2t) t^t \exp(-t) \\ & \leq C \sum_{t=0}^{\infty} \alpha^t (3 + \varepsilon_1)^t \frac{t^t \exp(-t)}{t!} \text{ for any } \varepsilon_1 > 0 \\ & \leq C \sum_{t=0}^{\infty} ((3 + \varepsilon_2) \alpha)^t, \text{ for any } \varepsilon_2 > 0 \end{aligned}$$

which converges if $\alpha < 1/3$, choosing $\varepsilon_2 < 1/\alpha - 3$. ■