Quantile regression with mismeasured covariates

Susanne M. Schennach*
Department of Economics
University of Chicago
smschenn@uchicago.edu

First version: February 6, 2005
This version: October 22, 2006

Abstract

This paper establishes that the availability of instrumental variables enables the identification and the consistent estimation of nonparametric quantile regression models in the presence of measurement error in the regressors. The proposed estimator takes the form of a nonlinear functional of derivatives of conditional expectations and is shown to provide estimated quantile functions that are uniformly consistent over a compact set.

1 Introduction

There is currently substantial interest in the development of estimation methods based on quantile restrictions (among others, Chesher (2003), Chernozhukov, Imbens, and Newey (2006), Chernozhukov and Hansen (2005), Angrist, Chernozhukov, and Fernandez-Val (2006), Abadie, Angrist, and Imbens (2002), Matzkin (2003)). These methods offer the advantage of describing not only averages of possible outcomes but their entire distribution as well. Quantile methods also exhibit robustness to outliers since quantiles of the dependent variable are always defined even when some of its moments may not exist. Finally, in multi-equations settings, quantile-based models enable a natural treatment of nonseparable (i.e. nonadditive) disturbances (for instance, Chesher (2003), Chernozhukov, Imbens, and Newey (2006), Matzkin (2003)).

This paper extends the applicability of quantile methods to the common situation where the covariate is measured with error. We thus consider the following quantile restriction model

$$P\left[y \leq Q_{\tau}\left(x^{*}\right)|x^{*}\right] = \tau \tag{1}$$

$$x = x^{*} + \Delta x \tag{2}$$

where the function of interest, $Q_{\tau}(x^{*})$, is the τ quantile of y conditional on x^{*}, the true unobserved value of the explanatory variable, while x is its observed counterpart, contaminated by a measurement error Δx. Our aim is to describe how the availability of an instrument vector w enables the identification and the consistent

*This work is made possible through financial support from the National Science Foundation via grant SES-0452089.

estimation of such a model. In the spirit of quantile analysis, which imposes no specific functional form on the distribution of the dependent variable, the proposed approach makes no functional form assumption regarding the quantile function $Q_\tau(x^*)$, thus enabling a full characterization of how the distribution of the dependent variable varies as a function of the covariates.

Although a number of powerful approaches have been developed to handle nonseparable disturbances (e.g. Chesher (2003), Chernozhukov, Imbens, and Newey (2006), Chernozhukov and Hansen (2005), Matzkin (2003)), these methods are unable to handle endogeneity of the specific form associated with the presence of mismeasured regressors. The basic problem is that when the above model is cast into the form $y = h(x, \varepsilon)$, there is no way to define an appropriate disturbance ε so that the assumptions traditionally made in the nonseparable error literature hold. First, since ε must depend on Δx and since Δx is correlated with x, methods relying on independence between x and ε are not applicable (e.g. Matzkin (2003)). In addition, the disturbance ε is, in general, not independent from the true value x^* and therefore cannot reasonably be assumed to be independent from an instrument for x^*, thus violating the most crucial assumption employed in existing endogenous quantile regression methods (e.g. Chernozhukov, Imbens, and Newey (2006), Chernozhukov and Hansen (2005)).

To see this second point, consider a case where y is generated according to $y = g(x^*, \eta)$ where, for simplicity, the nonlinear function $g(\cdot, \cdot)$ is monotone both in the scalar x^* and in the scalar disturbance η. If x^* is measured with error, there exist many possible equivalent models of the general form $y = h(x, \varepsilon)$, where $h(x, \varepsilon)$ is a function of a scalar x (the observed but mismeasured value of x^*) and some scalar disturbance ε. We will now illustrate that instrumental quantile methods cannot handle this model for any function $h(x, \varepsilon)$ (so, in particular, for the desired function $h(x, \varepsilon)$ that would happen to exhibit the same dependence on x as $g(x^*, \eta)$ has on x^*). Let $h_\varepsilon^{-1}(x, \cdot)$ denote the inverse of the function $h(x, \cdot)$ for a fixed value of x. It can be verified by direct substitution that, for any $h(x, \varepsilon)$ that is strictly increasing in ε, the model $y = h(x, \varepsilon)$ generates the same distribution of y if the scalar disturbance ε is set to $h_\varepsilon^{-1}(x, g(x - \Delta x, \eta))$, or equivalently, to $h_\varepsilon^{-1}(x^* + \Delta x, g(x^*, \eta))$. Since the disturbance ε explicitly depends on x^*, it is, in general, impossible to find an instrument that would be correlated with x^* while being statistically independent from ε. As a result, the model cannot generally be cast into a form that satisfies the assumptions traditionally made in the quantile instrumental variable literature, except in special cases, such as when $g(x^*, \eta)$ is linear in both of its arguments.

Despite their wide applicability, approaches based on triangular systems of equations (Chesher (2003), Imbens and Newey (2003)) are also unable to handle mismeasured covariates. These methods require that the effective number of distinct unobservable variables[1] is equal to the number of equations, which is not the case in errors-in-variables models, since there are three unobservables (x^*, Δx, η) but only two equations ($y = g(x^*, \eta)$ and $x = x^* + \Delta x$). Including an instrument equation does not help, since it also increases the

[1] Strictly speaking, Imbens and Newey (2003) allow for multiple disturbances per equation. However, as they point out on p.4, they do not have the "ability to identify effects corresponding to particular elements" of the multivariate disturbances, which is crucial for errors-in-variables models.

number of disturbances by one.

To our knowledge, the study of the effect of general forms of measurement error on quantile methods has so far been limited to the development of approximate expressions for the bias, resulting from classical measurement error, that become exact in the "small errors" limit (Chesher (2001)). While this approach provides very helpful insight into the nature of measurement error-induced distortions, there remains considerable interest in devising exactly consistent quantile estimation methods based on instrumental variables.[2]

This paper adapts the instrumental variable identification result for conditional means found in Schennach (2006) to quantile restrictions. In particular, since quantile restrictions take the form of expectations of bounded step functions, their special form enables considerable simplifications in terms of identification, thereby facilitating the derivation of a corresponding nonparametric estimator and allowing for the use of more primitive regularity conditions. We also provide a uniformly consistent nonparametric estimator based on this new identification result, thus paralleling the treatment found in Chernozhukov, Imbens, and Newey (2006) for a complementary class of endogenous models.

2 Identification

It is clear that $Q_\tau(x^*)$ in Equation (1) will be nonparametrically identified over some compact interval $\mathcal{X}$ if it is possible to identify conditional expectations of the form

$$E\left[S\left(y-q\right)|x^*\right] \tag{3}$$

for $x^* \in \mathcal{X} \subset \mathbb{R}$ and $q \in \mathbb{R}$, where

$$S\left(v\right) = 1\left(v \leq 0\right), \tag{4}$$

and where $1(A)$ denotes the indicator function of some event A. Indeed, in this case, $Q_\tau(x^*)$ for given values of x^* and τ would simply be the solution q to the equation $E\left[S\left(y-q\right)|x^*\right] = \tau$, which we assume to be unique.

Assumption 1 *The equation* $E\left[S\left(y-q\right)|x^*\right] = \tau$ *has a unique solution* $q = Q_\tau(x^*)$ *for each* $x^* \in \mathcal{X}$ *and* $\tau \in [0,1]$.

The problem is, of course, that x^* is not directly observed, only its error-contaminated counterpart x is. In this section, we show that the availability of an instrument vector w enables the identification of expectations of the form (3).The instrument is assumed to satisfy the following restrictions.

[2] When repeated measurements of the mismeasured variables are available, the identification of quantile-based models with measurement error follows directly from Schennach (2004a) for parametric models and from Schennach (2004b) in nonparametric settings, although a proper analysis of the asymptotic properties of the corresponding estimators would require a separate treatment due to the nonsmooth nature of quantile restrictions.

Assumption 2 *Let* $\Delta x^* = x^* - E[x^*|w]$*. For any* y*,* x^**,* w*,* Δx^* *and* Δx *in the support of their respective densities (denoted by* f *with the appropriate subscripts),*

$$f_{y|x^*,w}(y|x^*,w) = f_{y|x^*}(y|x^*) \tag{5}$$

$$f_{\Delta x^*|w}(\Delta x^*|w) = f_{\Delta x^*}(\Delta x^*) \tag{6}$$

$$E[\Delta x|w,y] = 0. \tag{7}$$

The first part of the assumption basically means that the instrument does not give any information about the dependent variable that the true regressor does not already provide, which is a fairly standard exclusion restriction. The second part states that the "prediction error of the instruments" $\Delta x^* = x^* - E[x^*|w]$ and the instruments w are independent. This assumption has been previously used in Newey (2001), Hausman, Newey, Ichimura, and Powell (1991), Hausman, Newey, and Powell (1995), Wang and Hsiao (1995) in the directly related context of instrumental variable estimation of nonlinear conditional mean regressions in the presence of mismeasured regressors. Outside of the measurement error context, similar relationships between a regressor and an instrument are extensively considered in Chalak and White (2006) and White and Chalak (2006) in their systematic classification of types of instruments. Intuitively, this independence assumption presumes that variations in x^* are "caused" by variations in w and unrelated additive random factors through an instrumental variable equation of the general form $x^* = Z(w) + \Delta x^*$ where $Z(w)$ is some nonlinear function. For instance, if x^* is cognitive skills during adulthood (imperfectly measured through an IQ score x), the instrument vector w could contain parent's education level and income, number of siblings and availability of reading material during childhood, etc.

In practice, the plausibility of this independence assumption can be tested by regressing x on w and verifying whether the residuals are independent from w. This test is not perfect, because it tests the independence of $\Delta x + \Delta x^*$ from w, instead of merely Δx^* from w. However, if the hypothesis of independence is not rejected, this can be regarded as strong support for the validity of this assumption. In fact, an example of an application relying on the same assumption in the IQ score example mentioned above can be found in Schennach (2006) (Section H), where the above test was clearly found not to reject the hypothesis of independence.

The measurement error Δx only needs to satisfy a weak conditional mean assumption (thus allowing for heteroskedasticity). Although one might be critical of our somewhat classical assumption regarding the measurement error Δx, it is worth mentioning that most approaches devised to handle measurement error require the stronger assumption of independence between the measurement error and the true regressor, including the only existing study of the effect of measurement error in quantile regressions (Chesher (2001)).

The fact that the independence assumption we need involves instrumental variables rather than the main variables of the model can be viewed as advantageous. In a given application, researchers rarely have the freedom to choose their regressor and their dependent variables. However, they have the ability to select the instruments without changing the main relationship of interest. Hence, researchers can simply explore various instruments in search for the ones such that the independence assumption is plausible, according to the test described above.

It is convenient to define a scalar-valued random variable z that summarizes the predictive power of the instruments:

$$z \equiv E[x^*|w] = E[x|w]. \tag{8}$$

The second equality, which shows that z can be expressed solely in terms of observable variables, follows from $E[x|w] = E[x^*|w] + E[\Delta x|w] = E[x^*|w]$ since $E[\Delta x|w] = E[E[\Delta x|w,y]|w] = 0$ by Assumption 2.

The implementation of this approach and the requisite regularity conditions are best expressed in terms of the following quantities and associated Fourier transforms.

Definition 1 *Let* $\mathbf{i} = \sqrt{-1}$,

$$\phi(\zeta) = E\left[e^{-\mathbf{i}\zeta\Delta x^*}\right] \tag{9}$$

$$\begin{aligned} c(q) &= \frac{1}{2}\left(c_+(q) + c_-(q)\right) && (10)\\ c_+(q) &= \lim_{z\to+\infty} E[S(y-q)|z] && (11)\\ c_-(q) &= \lim_{z\to-\infty} E[S(y-q)|z] && (12) \end{aligned}$$

and, for $k = 0, 1$,

$$\begin{aligned} s_{kx^*}(x^*,q) &= E\left[(z-x)^k\left(S(y-q)-c(q)\right)|x^*\right] && (13)\\ s_{k\partial x^*}(x^*,q) &= \frac{\partial E\left[(z-x)^k S(y-q)|x^*\right]}{\partial x^*} && (14) \end{aligned}$$

$$\begin{aligned} s_{kz}(z,q) &= E\left[(z-x)^k\left(S(y-q)-c(q)\right)|z\right] && (15)\\ s_{k\partial z}(z,q) &= \frac{\partial E\left[(z-x)^k S(y-q)|z\right]}{\partial z} && (16) \end{aligned}$$

$$\begin{aligned} \sigma_{kx^*}(\zeta,q) &= \int s_{kx^*}(x^*,q)\,e^{\mathbf{i}\zeta x^*}dx^* && (17)\\ \sigma_{k\partial x^*}(\zeta,q) &= \int s_{k\partial x^*}(x^*,q)\,e^{\mathbf{i}\zeta x^*}dx^*. && (18) \end{aligned}$$

$$\begin{aligned} \sigma_{kz}(\zeta,q) &= \int s_{kz}(z,q)\,e^{\mathbf{i}\zeta z}dz && (19)\\ \sigma_{k\partial z}(\zeta,q) &= \int s_{k\partial z}(z,q)\,e^{\mathbf{i}\zeta z}dz. && (20) \end{aligned}$$

(By convention, integrals without explicit bounds are taken over the whole real line.) Let $\mathcal{Q}$ *be a set containing the set* $\{Q_\tau(x^*) : x^* \in \mathcal{X},\ \tau \in [0,1]\}$.

To ensure that the above Fourier transforms are well-defined we require the following assumption.

Assumption 3 *The distributions of x^* and z are supported on $\mathbb{R}$.*

Fourier transforms are integrals over the whole real line that therefore demand the existence, over the whole real line, of the quantities to be transformed. Fourier transforms of probability distributions pose no problem because, even if a distribution has compact support, it is still defined outside of the support — it just happens to be zero. Conditional expectations are a different story, however, since they are not defined outside of the support of the conditioning variable. Hence, 3 is used ensures that Fourier transforms of expectations conditional on x^* or z are well-defined. The assumption on x^* is included for clarity, but is actually redundant, since having $f_z(z)$ supported on $\mathbb{R}$ actually implies that $f_{x^*}(x^*)$ is also supported on $\mathbb{R}$ (under Assumption 2). Assumption 2 implies that $f_x(x)$ is supported on $\mathbb{R}$ as well, although we do not make use of that fact. It is important to note that none of the remaining variables, i.e. Δx, $\Delta x^*, y$ or w, need to have distributions supported on $\mathbb{R}$.

Clearly, all quantities that involve x^* or Δx^* are not directly observable. However, quantities involving z, such as $\sigma_{kz}(\zeta, q)$, $\sigma_{k\partial z}(\zeta, q)$ and $c(q)$ are observable, and our goal will be to express $E[S(y-q)|x^*]$ in terms of these observable quantities.

We first need to make few primitive assumptions.

Assumption 4 *(i) $\phi(\zeta) \neq 0$ for all $\zeta \in \mathbb{R}$ and (ii) for all $q \in \mathcal{Q}$, $\sigma_{0\partial z}(\zeta, q) \neq 0$ for almost any ζ in $\mathbb{R}$.*

The assumption that various Fourier transforms be nonvanishing is widely used in the deconvolution literature (e.g. Fan (1991), Fan and Truong (1993), Liu and Taylor (1989), Li and Vuong (1998)). The only common distributions that are excluded by the requirement that $\phi(\zeta) \neq 0$ are the uniform and the triangular distributions. The normal, t, χ^2, gamma, and double exponential distributions all satisfy this assumption. Note that the restrictions on $\sigma_{0\partial z}(\zeta, q)$ are even weaker, thanks to the "almost everywhere" refinement, allowing $\sigma_{0\partial z}(\zeta, q)$ to cross zero at isolated points, perhaps infinitely many times. Assumption 4(ii) is only violated in cases where $\sigma_{0\partial z}(\zeta, q) = 0$ vanishes over an interval, which is highly unusual. The only simple example of such a function would be a polynomial, but $E[S(y-q)|z]$ cannot be a polynomial since it is bounded (except for the uninteresting case of $E[S(y-q)|z]$ constant).

We also need to impose weak constraints on the behavior of the tails of various functions.

Assumption 5 *$E[|\Delta x^*|] < \infty$.*

Assumption 6 *(i) There exist $B < \infty$ and $\beta > 0$ such that $\sup_{q\in\mathcal{Q}} |\partial E[S(y-q)|x^*]/\partial x^*| \leq B(1+|x^*|)^{-2-\beta}$ for all $x^* \in \mathbb{R}$ and (ii) the density of Δx^* is bounded.*

Intuitively, Assumption 6(i) imposes a minimum amount of smoothness and prevents the tails of $E[S(y-q)|x^*]$ from oscillating too much as $|x^*| \to \infty$. Since most economic models yield conditional quantiles that either diverge towards $\pm\infty$ or reach a constant limiting value as $|x^*| \to \infty$, this assumption is very likely to be satisfied in most applications. This assumption can perhaps best be understood when phrased in terms of a restriction imposed on the behavior of y as a function of x^*. For simplicity, consider the case when y has a homoskedastic distribution centered on $g(x^*)$ conditional on x^*. (We further assume that the density of y given x^* is bounded and that $g(x^*)$ is differentiable.) In this case, there are two alternative situations:

1. If $\lim_{x^* \to \infty} g(x^*)$ is finite, then Assumption 6(i) requires that $\partial g(x^*)/\partial x^* = O\left((x^*)^{-2-\beta}\right)$ as $|x^*| \to \infty$.

2. If $g(x^*)$ diverges as $|x^*| \to \infty$, then Assumption 6(i) requires that the density of y given x^* has sufficiently thin tails and/or that $g(x^*)$ diverges sufficiently rapidly as $|x^*| \to \infty$. Specifically, we need that $f_\varepsilon(g(x^*))\, \partial g(x^*)/\partial x^* = O\left((x^*)^{-2-\beta}\right)$ as $|x^*| \to \infty$, where $f_\varepsilon(\cdot)$ denotes the density of $y - g(x^*)$.

As shown in Lemma 1 in the Appendix, Assumption 6 implies three important facts. First, it implies that the limits of $E[S(y-q)|x^*]$ as $x^* \to +\infty$ or $x^* \to -\infty$ are well-defined. It implies that the limits of $E[S(y-q)|z]$ as $z \to +\infty$ or $z \to -\infty$ exist as well, so that the $c(q)$ in Definition 1 is well-defined. This assumption also guarantees that all Fourier transforms considered exist as *ordinary functions*. The term "ordinary function" is used here to describe a Fourier transform that is entirely defined by the value it takes at each frequency, in contrast with *generalized functions* (e.g., Schwartz (1966), Lighthill (1962)), which contain singular entities such as Dirac delta functions that need to be defined through the limit of a sequence of functions. The fact that all Fourier transforms involved remain ordinary functions in the present setup represents a considerable simplification over the treatment found in Schennach (2006) and will enable the construction of a natural nonparametric quantile estimator in the next section.[3]

Under these assumptions we can obtain a closed-form expression for $E[S(y-q)|x^*]$ (or, equivalently, $s_{0x^*}(x^*, q)$) and an expression for the quantile function $Q_\tau(x^*)$ in terms of observable quantities. Note that our identification result holds for general stationary processes, since it is expressed in term of the marginal distributions of individual observations.

Theorem 1 *Under Assumptions 2-6, for any $q, \tilde{q} \in \mathcal{Q}$,*

$$s_{0x^*}(x^*, q) = \frac{1}{2\pi} \int \frac{\sigma_{0z}(\zeta, q)}{\phi(\zeta)} e^{-\mathbf{i}\zeta x^*} d\zeta \tag{21}$$

where

$$\phi(\zeta) = \exp\left(\int_0^\zeta \frac{\mathbf{i}\sigma_{1\partial z}(\xi, \tilde{q})}{\sigma_{0\partial z}(\xi, \tilde{q})} d\xi\right). \tag{22}$$

Under the additional Assumption 1, the solution $Q_\tau(x^*)$ to $s_{0x^*}(x^*, Q_\tau(x^*)) = \tau$ gives the conditional τ quantile function for any $\tau \in [0, 1]$.

[3] The use of delta functions cannot be avoided in Schennach (2006), because Fourier transforms of unbounded functions are necessary in that more general setting. However, merely applying the results of Schennach (2006) in the present context does not automatically yield a method that is free of delta functions — additional techniques are needed, as will be explained below.

While a formal proof of this result can be found in the Appendix, an outline of the proof can be given as follows. First, Assumption 2 enables us to relate the observable quantity $E\left[S\left(y-q\right)|z\right]$ to the unobservable quantity of interest $E\left[S\left(y-q\right)|x^{*}\right]$ through a simple convolution operation

$$E\left[S\left(y-q\right)|z\right]=\int E\left[S\left(y-q\right)|x^{*}=z-u\right]f_{\Delta x^{*}}\left(-u\right)du. \tag{23}$$

To facilitate further manipulations (such as Fourier transforms), we consider the derivative of this expression with respect to z because is exhibits the property that it decays to zero as $|z|\rightarrow\infty$ under our regularity conditions:

$$\frac{\partial}{\partial z}E\left[S\left(y-q\right)|z\right]=\int\left[\frac{\partial E\left[S\left(y-q\right)|x^{*}\right]}{\partial x^{*}}\right]_{x^{*}=z-u}f_{\Delta x^{*}}\left(-u\right)du. \tag{24}$$

This expression cannot merely be inverted to establish identification since we have two unknown functions $\partial E\left[S\left(y-q\right)|x^{*}\right]/\partial x^{*}$ and $f_{\Delta x^{*}}\left(u\right)$ on the right-hand side. We need another equation. Fortunately, Assumption 2 also implies, in a similar fashion, that

$$\frac{\partial}{\partial z}E\left[\left(z-x\right)S\left(y-q\right)|z\right]=\int\left[\frac{\partial E\left[S\left(y-q\right)|x^{*}\right]}{\partial x^{*}}\right]_{x^{*}=z-u}uf_{\Delta x^{*}}\left(-u\right)du. \tag{25}$$

We now have two functional equations in two unknown functions $\partial E\left[S\left(y-q\right)|x^{*}\right]/\partial x^{*}$ and $f_{\Delta x^{*}}\left(u\right)$ and we can solve for the unknowns using Fourier transform techniques.

Since both expressions (24) and (25) involve convolutions, their respective Fourier transform take the form of a product of Fourier transforms (by the Convolution Theorem)

$$\sigma_{0\partial z}\left(\zeta,q\right) = \sigma_{0\partial x^{*}}\left(\zeta,q\right)\phi\left(\zeta\right) \tag{26}$$

$$\mathbf{i}\sigma_{1\partial z}\left(\zeta,q\right) = \sigma_{0\partial x^{*}}\left(\zeta,q\right)\dot{\phi}\left(\zeta\right), \tag{27}$$

where we have used Definition 1 in conjunction with the fact that the Fourier transform of $\mathbf{i}uf_{\Delta x^{*}}\left(-u\right)$ is $d\phi\left(\zeta\right)/d\zeta\equiv\dot{\phi}\left(\zeta\right)$. Taking the ratio of Equations (27) and (26), we obtain a simple differential equation in $\phi\left(\zeta\right)$,

$$\frac{\mathbf{i}\sigma_{1\partial z}\left(\zeta,q\right)}{\sigma_{0\partial z}\left(\zeta,q\right)}=\frac{\dot{\phi}\left(\zeta\right)}{\phi\left(\zeta\right)}, \tag{28}$$

which can be solved for $\phi\left(\zeta\right)$ to yield Equation (22), after exploiting the known initial condition $\phi\left(0\right)=\int f_{\Delta x^{*}}\left(-u\right)du=1$.

Next, we use a version of Equation (23) that is centered by the constant $c\left(q\right)$

$$E\left[S\left(y-q\right)-c\left(q\right)|z\right]=\int E\left[S\left(y-q\right)-c\left(q\right)|x^{*}=z-u\right]f_{\Delta x^{*}}\left(-u\right)du.$$

Thanks to the centering by $c\left(q\right)$, the corresponding Fourier transforms can be shown to be ordinary functions and we can write the Fourier transform of this espression as

$$\sigma_{0z}\left(\zeta,q\right)=\sigma_{0x^{*}}\left(\zeta,q\right)\phi\left(\zeta\right),$$

in which $\sigma_{0x^{*}}\left(\zeta,q\right)$ can be isolated in term of the observable $\sigma_{0z}\left(\zeta,q\right)$ and the quantity $\phi\left(\zeta\right)$ which we have just identified. Equation (21) then follows, after an inverse Fourier transform and the Theorem is proven.

Solving for the quantile of interest $Q_\tau(x^*)$ at some given x^* numerically will typically require calculating $s_{0x^*}(x^*, q)$ through Equation (21) for various trial values of q. In contrast, the quantity $\tilde{q}$ in Equation (22) can be held fixed throughout the estimation process. Naturally, it will be advantageous to select a value of $\tilde{q}$ that minimizes the statistical noise, e.g. that is such that $\sigma_{0\partial z}(\xi, \tilde{q})$ goes to zero slowly as $|\xi| \to \infty$.

It is interesting to point out two key steps that circumvent the potential problem that Fourier transforms often do not exist within the space of conventional functions. First, we rely on the fact that the Fourier transform of a centered step function (e.g. $1(z \geq 0) - 1/2$) is an ordinary function, even though the original step function is not absolutely integrable. A consequence of this fact is that subtracting the constant term $c(q)$ in the expression of $s_{kz}(z, q)$ in Definition 1 implies that the corresponding Fourier transform $\sigma_{kz}(\zeta, q)$ exists as an ordinary function. In a sense, our identification result relies on a form of "identification at infinity" (Heckman (1990)) to determine the constant term. However, unlike in the conventional "identification at infinity" scheme, it is relatively common that the constants $c_-(q)$ and $c_+(q)$ are known to be exactly equal to 0 or 1 based on known limiting behavior of the economic model considered. This occurs whenever $|Q_\tau(x^*)| \to \infty$ as $|x^*| \to \infty$ for any given $\tau \in]0, 1[$.

A second way we can ensure the existence of all Fourier transforms as ordinary functions is through our use of derivatives of conditional expectations (which can reasonably be assumed absolutely integrable) for some of the terms, so that the corresponding Fourier transforms ($\sigma_{k\partial z}(\xi, \tilde{q})$ for $k = 0, 1$) are guaranteed to be bounded and continuous.

Another useful consequence of Theorem 1 is that Equation (22) provides the characteristic function of $-\Delta x^*$ in terms of observable variables. Since $x^* = z + \Delta x^*$, where Δx^* is independent from z by Assumption 2, the characteristic function of x^* is given by

$$E\left[e^{\mathbf{i}\zeta x^*}\right] = E\left[e^{\mathbf{i}\zeta z}\right] E\left[e^{\mathbf{i}\zeta \Delta x^*}\right] = E\left[e^{\mathbf{i}\zeta z}\right] \phi(-\zeta), \tag{29}$$

where the right-hand side is expressed solely in terms of known quantities. Hence, the distribution of x^* is fully identified. This, in conjunction with the possibility to identify $E[S(y-q)|x^*]$ for any q, implies that the joint distribution of y and x^* is identified.

A multivariate extension of Theorem 1 that, in addition, allows for correctly measured covariates r is straightforward. The multivariate analogues of Equations (21) and (22) are

$$E[S(y-q) - \bar{c}(q,r)|x^*, r] = \frac{1}{(2\pi)^J} \int \frac{\int E[S(y-q) - \bar{c}(q,r)|z, r] e^{\mathbf{i}\zeta \cdot z} dz}{\phi(\zeta, r)} e^{-\mathbf{i}\zeta \cdot x^*} d\zeta \tag{30}$$

where the constant $\bar{c}(q,r)$ is selected so that $\int E[S(y-q) - \bar{c}(q,r)|z,r] e^{\mathbf{i}\zeta \cdot z} dz$ is an ordinary function[4] and where x^*, $z \equiv E[x^*|w]$ and ζ now take values in $\mathbb{R}^J$ and r is a vector of correctly measured regressors (if any) that can be discrete and where $\phi(\zeta, r) = \prod_{j=1}^J \phi_j(\zeta_j, r)$ with

$$\phi_j(\zeta_j, r) = \exp\left(\int_0^{\zeta_j} \frac{\mathbf{i} \int (\partial E[(z_j - x_j) S(y-\tilde{q})|z_j, r]/\partial z_j) e^{\mathbf{i}\zeta_j z_j} dz_j}{\int (\partial E[S(y-\tilde{q})|z_j, r]/\partial z_j) e^{\mathbf{i}\zeta_j z_j} dz_j} d\xi_j\right) \tag{31}$$

[4] For instance, $\bar{c}(q,r) = \lim_{R_1 \to \infty}\left[\lim_{R_2 \to \infty}\left(\int_{R_1 \leq \|z\| \leq R_2} E[S(y-q)|z,r]\, dz\right) / \left(\int_{R_1 \leq \|z\| \leq R_2} dz\right)\right]$.

where x and ξ also take values in $\mathbb{R}^J$. The above result holds under the additional assumptions that the "prediction errors" $x_j^* - E\left[x_j^*|w\right]$ are mutually independent[5] (for different values of j) conditional on r. Note that the integrals over vectors in Equation (30) stand for J nested integrals, while all integrals in Equation (31) remain univariate.

3 Estimation

We will now formally establish that a consistent estimator of $Q_\tau(x^*)$ can be obtained by substituting estimates of $E[S(y-q)|z]$ and $E[(z-x)S(y-q)|z]$ into Theorem 1 and inserting various trimming parameters. For estimation purposes, we consider iid settings, although an extension to serially dependent data is clearly possible.

Assumption 7 *(y_i, x_i, w_i) for $i = 1, \ldots, n$ is an iid sample of observations of the random variables (y, x, w).*

The first step in the estimation is to construct the scalar-valued variable z from the instrument vector w, which is accomplished by a standard least-squares projection. We consider the case where x^* is related to w through

$$x^* = Z(w, \alpha) + \Delta x^*, \tag{32}$$

where $Z(w, \alpha)$ is a known function of an unknown parameter α and where Δx^* is independent from w, and may be assumed to have zero mean, without loss of generality, by including an intercept in the function $Z(w, \alpha)$. The parameter α can be estimated from a nonlinear regression of x on w, since

$$x = Z(w, \alpha) + \Delta x^* + \Delta x, \tag{33}$$

where $E[\Delta x^* + \Delta x|w] = E[\Delta x^*] + E[E[\Delta x|w, y]|w] = 0$. The following Assumption collects all the standard regularity conditions traditionally used to show consistency of the first-step estimator $\hat{\alpha}$ in iid settings.

Assumption 8 *Let $\mathcal{A}$ be a compact set under some norm $\|\cdot\|_\alpha$ such that*
(i) $\alpha^* = \arg\min_{\alpha \in \mathcal{A}} E\left[(x - Z(w, \alpha))^2\right]$ is unique,
(ii) $E\left[\sup_{\alpha \in \mathcal{A}} |Z(w, \alpha)|^2\right] < \infty$ and $E\left[x^2\right] < \infty$,
(iii) $Z(w, \alpha)$ is continuous in α for $\alpha \in \mathcal{A}$ under the norm $\|\cdot\|_\alpha$.

[5] The mutual independence assumption can be avoided at the expense of employing higher-dimensional nonparametric quantities. Instead of writing the joint characteristic function of the vector $x^* - E(x^*|w)$ as a product of marginals, $\phi(\zeta, r) = \prod_{j=1}^J \phi_j(\zeta_j, r)$, the joint characteristic function can be directly obtained as a path integral from 0 to ζ:

$$\phi(\zeta, r) = \exp\left(\int_0^\zeta \rho(\xi, r) \cdot d\xi\right),$$

where the vector $\rho(\xi, r)$ is a ratio of multivariate Fourier transforms:

$$\rho(\xi, r) = \frac{\mathbf{i} \int \left(\nabla_z^2 E[(z-x)S(y-\tilde{q})|z, r]\right) e^{\mathbf{i}\zeta \cdot z} dz}{\int \left(\nabla_z^2 E[S(y-\tilde{q})|z, r]\right) e^{\mathbf{i}\zeta \cdot z} dz}$$

and where it should be noted that the numerator is vector-valued but not the denominator. The operator ∇_z^2 denotes $\partial^2/\partial z_1^2 + \ldots + \partial^2/\partial z_J^2$. Note that first derivatives in Equation (31) are replaced here by second derivatives, because the latter admit a natural spherically symmetric extension to multivariate settings, unlike first-derivatives. The proof of our identification result can be straightforwardly adapted to allow for second derivatives, provided Assumption 6(i) is replaced by $\sup_{q \in \mathcal{Q}} \left|\nabla_{x^*}^2 E[S(y-q)|x^*]\right| \le B(1 + \|x^*\|)^{-2-\beta}$.

Note that Assumption 8 covers not only the case where $Z(w,\alpha)$ is parametrically specified, but also allows for a nonparametric first step, provided that it is plausible to assume that α^* belongs to a compact set under some norm $\|\cdot\|_\alpha$. Newey and Powell (2003), for instance, propose that functions can be assumed to belong to a compact set if one is willing to assume that the "tails" of these functions follow a parametric trend while their "middle" portion is fully nonparametric but satisfies Lipschitz constraints. With an estimate $\hat{\alpha}$ in hand, we define

$$z_i = Z(w_i, \hat{\alpha})$$

for each observation i in the sample. Since the parameter α is estimated, it is necessary to consider α-dependent counterparts of the quantities defined in Definition 1 for the purpose of obtaining asymptotic properties.

Definition 2 *For any $q \in \mathcal{Q}$ and $\alpha \in \mathcal{A}$, let*

$$\begin{aligned} s_{kz}(z,q,\alpha) &= E\left[(Z(w,\alpha)-x)^k S(y-q)\,|Z(w,\alpha)=z\right] \\ s_{k\partial z}(z,q,\alpha) &= \partial E\left[(Z(w,\alpha)-x)^k S(y-q)\,|Z(w,\alpha)=z\right]/\partial z \end{aligned}$$

and let $s_{0x^*}(x^*,q,\alpha)$ and $\phi(\zeta,\alpha)$ respectively denote the value of $s_{0x^*}(x^*,q)$ and $\phi(\zeta)$ obtained from Theorem 1 when setting $z = Z(w,\alpha)$.

A consistent estimator of $E[S(y-q)\,|x^*]$ can be obtained by substituting kernel estimates of $E[S(y-q)\,|z]$ and $E[(z-x)S(y-q)\,|z]$ into Theorem 1 and inserting various trimming parameters.

Let h_n, $\bar{z}_n$, and $\bar{\zeta}_n$ be sequences such that $h_n \to 0$, $\bar{z}_n \to \infty$ and $\bar{\zeta}_n \to \infty$ as $n \to \infty$ and define the following quantities. First, let $K(z)$ be a standard bias-reducing kernel of order N_K:

Assumption 9 *The kernel $K(z)$satisfies (i) $\int K(z)\,dz = 1$, (ii) $K(z) = K(-z)$ (iii) $\int K(z)\,z^j dz = 0$ for $j = 1,\ldots,N_K-1$ (iv) $\int |K(z)|\,|z|^{N_k} dz < \infty$ and (v) $K(z)$ is differentiable.*

Definition 3 *Let*[6]

$$\hat{s}_{0x^*}(x^*,q,\alpha) = \frac{1}{2\pi}\int_{-\bar{\zeta}_n}^{\bar{\zeta}_n} \hat{\sigma}_{0x^*}(\zeta,q,\alpha)\,e^{-\mathbf{i}\zeta x^*} d\zeta \quad (34)$$

$$\hat{\sigma}_{0x^*}(\zeta,q,\alpha) = \hat{\sigma}_{0z}(\zeta,q,\alpha)\,\hat{\phi}^{-1}(\zeta,\tilde{q},\alpha) \quad (35)$$

$$\hat{\phi}^{-1}(\zeta,\tilde{q},\alpha) = \exp\left(-\int_0^{\zeta} \frac{\mathbf{i}\hat{\sigma}_{1\partial z}(\xi,\tilde{q},\alpha)}{\hat{\sigma}_{0\partial z}(\xi,\tilde{q},\alpha)} d\xi\right) \quad (36)$$

where $\tilde{q} \in \mathcal{Q}$ is some constant and

$$\hat{\sigma}_{0z}(\xi,q,\alpha) = \int \hat{s}_{0z}(z,q,\alpha)\,e^{\mathbf{i}\xi z} dz \quad (37)$$

$$\hat{\sigma}_{0\partial z}(\xi,q,\alpha) = \int \hat{s}_{0\partial z}(z,q,\alpha)\,e^{\mathbf{i}\xi z} dz \quad (38)$$

$$\hat{\sigma}_{1\partial z}(\zeta,q,\alpha) = \int \hat{s}_{1\partial z}(z,q,\alpha)\,e^{\mathbf{i}\zeta z} dz \quad (39)$$

[6] Operations involving complex numbers should not pose implementation problems, as most modern computer languages can handle them directly.

and where the integrands are trimmed kernel estimates, i.e., for $k = 0, 1$,

$$\hat{s}_{k\partial z}(z,q,\alpha) = 1(|z| \le \bar{z}_n) \frac{\partial}{\partial z} \frac{(nh_n)^{-1}\sum_{i=1}^n (Z(w_i,\alpha) - x_i)^k S(y_i - q) K\left(h_n^{-1}(Z(w_i,\alpha) - z)\right)}{(nh_n)^{-1}\sum_{i=0}^n K\left(h_n^{-1}(Z(w_i,\alpha) - z)\right)} \quad (40)$$

$$\hat{s}_{0z}(z,q,\alpha) = 1(|z| \le \bar{z}_n) \frac{(nh_n)^{-1}\sum_{i=1}^n S(y_i - q) K\left(h_n^{-1}(Z(w_i,\alpha) - z)\right)}{(nh_n)^{-1}\sum_{i=1}^n K\left(h_n^{-1}(Z(w_i,\alpha) - z)\right)} + \quad (41)$$
$$+ c_+(q) 1(z > \bar{z}_n) + c_-(q) 1(z < -\bar{z}_n) - c(q),$$

where $c_+(q)$, $c_-(q)$ *and* $c(q)$ *are given in Definition 1.*

Since our estimator involves Fourier transforms of conditional expectations (and derivatives thereof), it is necessary to establish uniform convergence results over the whole real line. While this is not possible in general, it becomes possible when the "tails" of these quantities have a known behavior, so that the kernel estimate in the region $|z| \ge \bar{z}_n$ where few sample points lie, can be trimmed and replaced by a constant. In the case of $\hat{s}_{k\partial z}(z,q,\alpha)$, Assumption 6 implies that the appropriate constant is zero. In the case of $\hat{s}_{0z}(z,q,\alpha)$, the appropriate constants are $c_-(q) \equiv \lim_{z\to-\infty} E[S(y-q)|z]$ and for $z < -\bar{z}_n$ and $c_+(q) \equiv \lim_{z\to\infty} E[S(y-q)|z]$ for $z > \bar{z}_n$, which necessarily exist by Assumption 6 and Lemma 1 in the Appendix. These limiting values are often known exactly. For instance, a very common case is when $c_-(q) = 0$ and $c_+(q) = 1$, occurring whenever $Q_\tau(x^*) \to -\infty$ as $x^* \to -\infty$ and $Q_\tau(x^*) \to \infty$ as $x^* \to \infty$ for any $\tau \in]0,1[$. If these constants are not known, they can be replaced by consistent estimates, such as

$$\hat{c}_+(q) = \frac{\sum_{i=1}^n S(y_i - q) 1(Z(w_i,\hat{\alpha}) \ge \bar{z}_n)}{\sum_{i=1}^n 1(Z(w_i,\hat{\alpha}) \ge \bar{z}_n)} \quad (42)$$

$$\hat{c}_-(q) = \frac{\sum_{i=1}^n S(y_i - q) 1(Z(w_i,\hat{\alpha}) \le -\bar{z}_n)}{\sum_{i=1}^n 1(Z(w_i,\hat{\alpha}) \le -\bar{z}_n)}. \quad (43)$$

while $c(q)$ is replaced by $\hat{c}(q) = (\hat{c}_+(q) + \hat{c}_-(q))/2$ in Equation (41). Our asymptotic analysis given below trivially allows for these estimated quantities since fluctuations in these quantities simply translate into fluctuations in $\hat{s}_{0x^*}(x^*,q,\hat{\alpha})$ of the same magnitude. Hence, we only need to establish that $\hat{c}_+(q) \xrightarrow{p} c_+(q)$ and that $\hat{c}_-(q) \xrightarrow{p} c_-(q)$, uniformly in q. A simple sufficient condition is that[7]

$$n \min\{E[1(Z(w_i,\alpha) \le -\bar{z}_n)], E[1(Z(w_i,\alpha) \ge \bar{z}_n)]\} \to \infty, \quad (44)$$

since it ensures that the averages in Equations (42) and (43) are taken over a number of observations going to infinity. The convergence is uniform in q and α, since the summand is bounded and continuous with probability one.

Next, we also need fairly weak boundedness assumptions.

Assumption 10 *There exist* $M < \infty$ *and* $\mu > 0$ *such that* $\sup_{\alpha\in\mathcal{A}} \sup_{q\in\mathcal{Q}} |s_{k\partial z}(z,q,\alpha)| \le M(1+|z|)^{-2-\mu}$ *for* $k = 0, 1$ *and for all* $z \in \mathbb{R}$.

[7] If the tails of the distribution of $Z(w_i,\alpha)$ are thin, then Equation (44) can be satisfied by having $\bar{z}_n$ grow sufficiently slowly with sample size. The trade-off is that a slowly growing $\bar{z}_n$ would cause a larger truncation bias, making the last line of Assumption 13 below more difficult to satisfy. This would only be a problem if we simultaneously have three issues:
(i) the constants $c_+(q)$ and $c_-(q)$ need to be estimated;
(ii) the tails of the distribution of $Z(w_i,\alpha)$ are thin;
(iii) $E[S(y-q)|z]$ approaches its limiting value slowly as $|z| \to \infty$ (implying that the truncation bias is slowly decaying).

Assumption 11 *For $Y = 1, S(y-q), (Z(w,\alpha) - x) S(y-q)$,*

$$\sup_{q\in\mathcal{Q}} \sup_{\alpha\in\mathcal{A}} \sup_{z\in\mathbb{R}} \left|\partial^j \left(E\left[Y|Z(w,\alpha) = z\right] p(z|\alpha)\right)/\partial z^j\right| < \infty,$$

where $p(z|\alpha)$ denotes the density of $z = Z(w,\alpha)$ for $j = 1$ or $j = N_K$, where N_k is as in Assumption 9.

Assumption 12 $\sup_{\alpha\in\mathcal{A}} \sup_{q\in\mathcal{Q}} \int \left|\partial^2 s_{0x^*}(x^*, q, \alpha)/\partial (x^*)^2\right| dx^* < \infty.$

In order to determine sequences for the trimming parameters $\bar{\zeta}_n$ and $\bar{z}_n$ and smoothing parameter h_n suitable for consistent estimation, we first define various bounds on the quantities defining the model.

Definition 4 *Let*

$$\begin{aligned}
\bar{\phi}_n &= \sup_{\alpha\in\mathcal{A}} \sup_{|\zeta|\leq\zeta_n} \left|\frac{d\ln\phi(\zeta,\alpha)}{d\zeta}\right| \\
\underline{\sigma}_n &= \inf_{\alpha\in\mathcal{A}} \inf_{|\zeta|\leq\zeta_n} \left|\sigma_{0\partial z}(\zeta,\tilde{q},\alpha)\right| \\
\underline{\phi}_n &= \inf_{\alpha\in\mathcal{A}} \inf_{|\zeta|\leq\zeta_n} \left|\phi(\zeta,\alpha)\right| \\
\underline{f}_n &= \inf_{\alpha\in\mathcal{A}} \inf_{|z|\leq\bar{z}_n} p(z|\alpha) \\
T_n &= \max\left\{\sup_{\alpha\in\mathcal{A}} \sup_{q\in\mathcal{Q}} \int_{|z|\geq\bar{z}_n} \left|s_{0z}(z,q,\alpha) + c(q) - c_+(q) 1(z\geq 0) - c_-(q) 1(z<0)\right| dz, \right. \\
&\qquad \left. \max_{k=0,1} \int_{|z|\geq\bar{z}_n} \left|s_{k\partial z}(z,q,\alpha)\right| dz\right\}
\end{aligned}$$

The existence of sample size-dependent lower or upper bounds (such as $\underline{\sigma}_n$, $\underline{\phi}_n$ or $\bar{\phi}_n$ above) on various Fourier transforms over an expanding interval is commonly assumed in the deconvolution literature (e.g., Fan (1991), Fan and Truong (1993), Li and Vuong (1998), Schennach (2004b)). Typically, these bounds take the form of some power or some exponential of the interval length $\bar{\zeta}_n$ and are determined by the asymptotic behavior of the various Fourier transforms as $\zeta \to \infty$. Smooth functions that admit a finite number of derivatives lead to bounds $\underline{\sigma}_n$ and $\underline{\phi}_n$ of the form $\bar{\zeta}_n^{-k}$ for some $k \in \mathbb{N}$, while so-called supersmooth functions imply bounds of the form $\exp\left(-\bar{\zeta}_n^k\right)$ for some $k \in \mathbb{N}$. The corresponding bound $\bar{\phi}_n$ would typically be of the form $\ln\bar{\zeta}_n$ or $\bar{\zeta}_n^k$, in the smooth and supersmooth cases, respectively. Lower bounds on densities (such as $\underline{f}_n$) over an expanding interval are common in the nonparametric and semiparametric estimation literature (e.g. Andrews (1995)). Bounds on tail-trimming terms (such as T_n) are also frequently used in semiparametric estimation involving nonparametric estimates defined over the whole real line (e.g. Hardle and Stoker (1989), Lewbel (1998)).

The constraints on the sample-size dependent parameters can then by expressed in terms of the quantities found in Definition 4 as follows.

Assumption 13 *As $n \to \infty$, we have $h_n \to 0$, $\bar{z}_n \to \infty$, $\bar{\zeta}_n \to \infty$,*

$$
\begin{aligned}
n^{-1/2} h_n^{-2} \frac{\bar{z}_n}{\underline{f}_n} \left(1 + h_n \underline{f}_n^{-1}\right) \frac{\bar{\zeta}_n^2 \bar{\phi}_n}{\underline{\phi}_n \underline{\sigma}_n} &\to 0 \\
h_n^{N_K - 1} \frac{\bar{z}_n}{\underline{f}_n} \left(1 + h_n \underline{f}_n^{-1}\right) \frac{\bar{\zeta}_n^2 \bar{\phi}_n}{\underline{\phi}_n \underline{\sigma}_n} &\to 0 \\
T_n \frac{\bar{\zeta}_n^2 \bar{\phi}_n}{\underline{\phi}_n \underline{\sigma}_n} &\to 0.
\end{aligned}
$$

One can readily recognize some familiar terms in the above expressions. The terms $n^{-1/2} h_n^{-2}$ and $h_n^{N_K - 1}$ are the uniform rates of convergence over a fixed interval for the standard deviation and bias, respectively, of kernel estimates of derivatives of conditional expectations (see Andrews (1995), Theorem 1, specialized to the i.i.d. case[8]). The penalty term $\bar{z}_n / \underline{f}_n$ accounts for the use of an expanding interval of length $2\bar{z}_n$ over which the density of z is bounded below by $\underline{f}_n$. The $1 + h_n \underline{f}_n^{-1}$ term is also a correction to the convergence rates that is needed when an expanding interval is used.[9] The bias due to trimming everything outside that expanding interval is T_n. Finally, the term $\left(\bar{\zeta}_n^2 \bar{\phi}_n\right) / \left(\underline{\phi}_n \underline{\sigma}_n\right)$ is the worst-case scenario noise amplification resulting from plugging the kernel estimates into the nonlinear functional defined in Theorem 1.

Although it is beyond the scope of this paper to devise data-dependent selection rules regarding the optimal trimming parameters $\bar{\zeta}_n$ and $\bar{z}_n$ and smoothing parameter h_n, a few helpful guidelines can be given. First, a reasonably good value for the bandwidth parameter h_n can be obtained through standard cross-validation rules applied to the estimates of the intermediate nonparametric quantities $s_{kz}(z, q, \alpha)$ or $s_{k\partial z}(z, q, \alpha)$. Second, the trimming parameter $\bar{z}_n$ can be optimized by looking for a region of values of z where $\hat{s}_{kz}(z, q, \alpha)$ seems to level-off. Then, $\bar{z}_n$ should be set so as to include as much as possible of this flatter region, while excluding any noisy "spikes" lying beyond. Naturally, different trimming parameters can be used for positive and negative z, although our notation does not allow for that, for simplicity. It may be necessary to alternate between the optimization of $\bar{z}_n$ and h_n a few times for best results. Once satisfactory estimates of $s_{kz}(z, q, \alpha)$ or $s_{k\partial z}(z, q, \alpha)$ have been found, the frequency cutoff $\bar{\zeta}_n$ can be optimized by identifying the largest frequency such that the ratio $\hat{\sigma}_{1\partial z}(\xi, \tilde{q}, \alpha) / \hat{\sigma}_{0\partial z}(\xi, \tilde{q}, \alpha)$ is roughly decaying in magnitude for $\xi \in \left[0, \bar{\zeta}_n\right]$. In other words, $\bar{\zeta}_n$ should be such that large spikes in $\hat{\sigma}_{1\partial z}(\xi, \tilde{q}, \alpha) / \hat{\sigma}_{0\partial z}(\xi, \tilde{q}, \alpha)$ are avoided. It should be noted that the Fourier space trimming parameter $\bar{\zeta}_n$ and the reciprocal of the bandwidth h_n^{-1} play a very similar role in that they both determine an effective frequency cutoff. Hence, for a given type of kernel, the product $\bar{\zeta}_n h_n$ will not differ much between different samples.

Our estimator also relies on a user-specified value of $\tilde{q}$. Typically, this value would be chosen so as to minimizes the statistical noise, e.g. chosen such that $\sigma_{0\partial z}(\xi, \tilde{q}, \alpha)$ (or, in practice, its estimate $\hat{\sigma}_{0\partial z}(\xi, \tilde{q}, \hat{\alpha})$) goes to zero slowly as $|\xi| \to \infty$. It may sometimes be beneficial to average the estimates obtained with different values of $\tilde{q}$ in order to improve efficiency. Our proof of consistency will directly imply that such a

[8]Different rates have also been obtained in the literature, under slightly different assumptions (e.g. Lemma 8.10 in Newey and McFadden (1994)).

[9]For fixed interval, $\underline{f}_n$ is constant so that $h_n \underline{f}_n^{-1} = O(h_n) = o(1)$ and this correction term would be asymptotically negligible.

procedure is consistent as well, since an average of consistent estimators is necessarily consistent.

To ensure consistency of our estimator in the presence of a plugged-in first step estimate $\hat{\alpha}$, we rely on a standard continuity condition (under the norm $\|\cdot\|_\alpha$ used in Assumption 8).

Assumption 14 *$s_{0x^*}(x^*, q, \alpha)$ is continuous in α for $\alpha \in \mathcal{A}$ and uniformly for $(x^*, q) \in \mathcal{X} \times \mathcal{Q}$.*

Consistency also demands that the behavior of the quantile function be suitably constrained. Following Chernozhukov, Imbens, and Newey (2006), we impose a Lipschitz constraint that implies compactness of the space of candidate functions.

Assumption 15 *$Q_\tau \in \mathcal{L}$, where, for some given $C > 0$, $\mathcal{L} = \{Q : \mathbb{R} \mapsto \mathbb{R}$ such that $|Q(x_1) - Q(x_2)| \leq C|x_1 - x_2|$ for all $x_1, x_2 \in \mathcal{X}$ and such that $Q(x_1) \in \mathcal{Q}$ for all $x_1 \in \mathcal{X}.\}$.*

A nonparametric estimator $\hat{Q}_\tau(x^*)$ can be obtained from

$$\hat{Q}_\tau \in \arg\min_{Q_\tau \in \mathcal{L}} \int_{x^* \in \mathcal{X}} \left(\hat{s}_{0x^*}(x^*, Q_\tau(x^*), \hat{\alpha}) - \tau\right)^2 dx^*, \tag{45}$$

where the use of the symbol "$\in$" instead of "$=$" allows for the fact that the minimand may not be unique, as is commonly the case in quantile approaches. Simple tie-breaking rules can be implemented to pick a single $\hat{Q}_\tau$ among the set of possible minimands, such as taking $\hat{Q}_\tau(x^*)$ so that it lies in the "middle" of the set of possible values of $\hat{Q}_\tau(x^*)$ at x^*. Our proof of consistency will hold no matter which rule is used to select a specific function $\hat{Q}_\tau$.

In practice, the minimization over a set of functions defined in Equation (45) can be carried out in a number of computationally convenient ways. One would typically consider a "sieve" method (Grenander (1981)) where the set of allowed functions $\mathcal{L}_n$ increases with sample size n:

$$\hat{Q}_\tau \in \arg\min_{Q_\tau \in \mathcal{L}_n} \int_{x^* \in \mathcal{X}} \left(\hat{s}_{0x^*}(x^*, Q_\tau(x^*), \hat{\alpha}) - \tau\right)^2 dx^*, \tag{46}$$

where $\mathcal{L}_n$ satisfies standard denseness condition.

Assumption 16 *$\mathcal{L}_n$ is a sequence of compact sets such that, for each n, there exists $Q_{\tau,n} \in \mathcal{L}_n$ satisfying $\lim_{n\to\infty} \sup_{x^* \in \mathcal{X}} |Q_{\tau,n}(x^*) - Q_\tau(x^*)| = 0$.*

For instance, one could solve for q in $\hat{s}_{0x^*}(x^*, q, \hat{\alpha}) = \tau$ on a grid $x^* = x_1^*, \ldots, x_m^*$ that becomes denser with increasing sample size and employ a linear interpolation between grid points. Alternatively, one could represent $Q_\tau(x^*)$ by a series with a number of terms that increases with sample size. Note that, thanks to the fact that smoothing parameters have already been introduced to deal with all nonparametric quantities, this sieve approach in our context is purely a numerical convenience aimed at representing an infinite-dimensional quantity in a necessarily finite computer memory. Our method is consistent for any sup norm-convergent approximation scheme $\mathcal{L}_n$ and even without the use of a sieve ($\mathcal{L}_n = \mathcal{L}$). One advantage of using a sieve is that, for reasonable choices of the grid $x_1^*, \ldots, x_m^*$ (that is, not too fine) or a reasonable number of terms in the series, the Lipschitz constraint on $\mathcal{L}_n$ will rarely be binding in practice and can therefore typically be ignored in the implementation of the estimator.

Our main consistency result, shown in the Appendix, can now be stated as follows.

Theorem 2 *Under Assumptions 1-15, $\hat{Q}_\tau$ given by Equation (45) satisfies* $\sup_{x^* \in \mathcal{X}} \left| \hat{Q}_\tau(x^*) - Q_\tau(x^*) \right| \xrightarrow{p} 0$ *for any* $\tau \in [0,1]$. *Under the additional Assumption 16, the same conclusion holds for $\hat{Q}_\tau$ given by Equation (46).*

The ability to obtain uniform consistency should prove useful in applications because the quantile function is often "plugged in" into some functional F aimed at obtaining some summary aggregate value. Continuity of such a functional F in the sup norm, which implies that $F\left(\hat{Q}_\tau\right)$ is consistent, is typically easy to show. Note that, while this consistency result only holds uniformly over x^* in a compact interval $\mathcal{X}$, this does not imply that the random variable x^* must be compactly supported.

It should be possible to extend Theorem 2 to obtain uniform convergence of $\hat{Q}_\tau$ over an expanding interval, although it is beyond the scope of the present paper to do so. A consistency result for multivariate x^* based on Equations (30) and (31) is also conceptually straightforward to obtain, although it is beyond the scope of the present paper to give the detailed regularity conditions needed.

As a by-product of obtaining our consistency result, we obtain uniform convergence results for families of nonparametric kernel estimates of conditional expectation, and derivatives thereof, over expanding intervals (see Lemma 4, in the Appendix), cases apparently not covered by existing results (e.g. Andrews (1995)). Another noteworthy aspect of the proof of Theorem 2 lies in Lemma 6 in the Appendix. The quantity $\sigma_{0z}(\zeta, q, \alpha)$ and its estimated counterpart $\hat{\sigma}_{0z}(\zeta, q, \alpha)$ diverge at the rate ζ^{-1} when $\zeta \to 0$, because the Fourier transform of a centered step function is $-(\mathbf{i}\zeta)^{-1}$. This appears to cause a problem in Equation (35), since the noise in the estimation of $\hat{\phi}^{-1}(\zeta, q, \alpha)$ is magnified by the explosion in $\hat{\sigma}_{0z}(\zeta, q, \alpha)$ near the origin. However, the noise in the estimated $\hat{\phi}^{-1}(\zeta, q, \alpha)$ happens to be proportional to ζ as $\zeta \to 0$ and the factors ζ^{-1} and ζ nicely cancel each other to yield a finite noise that decreases with sample size.

Finally, an interesting alternative application of our result arises if one exchanges the role of the dependent variable and the regressor. In this different setup, the dependent variable would satisfy a standard conditional mean assumption and the measurement error would satisfy a conditional quantile restriction.[10] This is a very plausible form of *nonclassical* measurement error that cannot be handled with any existing method. This form of nonclassical measurement error with nonzero conditional mean but with zero conditional median has been observed in a validation data study by Bollinger (1998). Of course, this alternative setup does require monotonicity assumptions regarding the regression function, to ensure that it is one-to-one, unlike our main model.

[10] The model could have the form:

$$\begin{array}{ll} x = g(y^*) + \Delta x & E[\Delta x | y^*] = 0 \\ y = y^* + \Delta y & \text{Median}(\Delta y | y^*) = 0 \end{array}$$

and the assumptions made throughout the paper would have to hold with $x^* \equiv g(y^*)$.

4 Conclusion

Given the prevalence of measurement error in economic data and its impact on inference based on conditional quantiles (Chesher (2001)), the ability to obtain consistent estimates of a conditional quantile function in the presence of measurement error in the covariates considerably expands the range of datasets that can be analyzed via models based on quantile restrictions. The proposed method nicely complements powerful existing instrumental quantile methods (Chesher (2003), Chernozhukov, Imbens, and Newey (2006), Chernozhukov and Hansen (2005), Angrist, Chernozhukov, and Fernandez-Val (2006), Abadie, Angrist, and Imbens (1998)) which are not applicable to the specific form of endogeneity resulting from measurement error.

References

Abadie, A., J. D. Angrist, and G. W. Imbens (1998): "Instrumental Variables Estimation of Quantile Treatment Effects," Working Paper 229, National Bureau of Economic Research (NBER).

——— (2002): "Instrumental Variables Estimates of the E ect of Subsidized Training on the Quantiles of Trainee Earnings," *Econometrica*, 70, 91–117.

Andrews, D. W. K. (1995): "Nonparametric Kernel Estimation for Semiparametric Models," *Econometric Theory*, 11, 560–596.

Angrist, J. D., V. Chernozhukov, and I. Fernandez-Val (2006): "Quantile Regression under Misspecification, with an Application to the U.S. Wage Structure," *Econometrica*, 74, 539–563.

Bollinger, C. R. (1998): "Measurement Error in the Current Population Survey: A Nonparametric Look," *Journal of Labor Economics*, 16, 576–594.

Chalak, K., and H. White (2006): "An Extended Class of Instrumental Variables for the Estimation of Causal Effects," Working Paper, UCSD.

Chernozhukov, V., and C. Hansen (2005): "An IV Model of Quantile Treatment Effects," *Econometrica*, 73, 245–261.

Chernozhukov, V., G. W. Imbens, and W. K. Newey (2006): "Instrumental Variable Identification and Estimation of Nonseparable Models via Quantile Conditions," *Journal of Econometrics*, forthcoming.

Chesher, A. (2001): "Parameter Approximations for Quantile Regressions with Measurement Error," Working Paper CWP02/01, Department of Economics, University College London.

——— (2003): "Identification in Nonseparable Models," *Econometrica*, 71, 1405–1441.

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.

Grenander, U. (1981): *Abstract Inference*. Wiley Series, 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., W. Newey, H. Ichimura, and J. Powell (1991): "Measurement Errors in Polynomial Regression Models," *Journal of Econometrics*, 50, 273–295.

Hausman, J., W. Newey, and J. Powell (1995): "Nonlinear Errors in Variables. Estimation of Some Engel Curves," *Journal of Econometrics*, 65, 205–233.

Heckman, J. (1990): "Varieties of Selection Bias," *American Economic Review*, 80, 313–Ű318.

Imbens, G. W., and W. K. Newey (2003): "Identification and Estimation of Triangular Simultaneous Equations Models without Additivity," Working Paper, MIT, Department of Economics.

Lewbel, A. (1998): "Semiparametric Latent Variable Model Estimation with Endogenous or Mismeasured Regressors," *Econometrica*, 66, 105–121.

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.

Liu, M., and R. Taylor (1989): "A consistent nonparametric density estimator for the deconvolution problem," *The Canadian Journal of Statistics*, 17(4), 427–438.

Matzkin, R. L. (2003): "Nonparametric Estimation of Nonparametric Nonadditive Random Functions," *Econometrica*, 71, 1339–1375.

Newey, W. (2001): "Flexible Simulated Moment Estimation of Nonlinear Errors-in-Variables Models," *Review of Economics and Statistics*, 83, 616–627.

Newey, 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.

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

Schennach, S. M. (2004a): "Estimation of Nonlinear Models with Measurement Error," *Econometrica*, 72, 33–75.

——— (2004b): "Nonparametric Estimation in the Presence of Measurement Error," *Econometric Theory*, 20, 1046–1093.

——— (2006): "Instrumental Variable Estimation of Nonlinear Errors-in-Variables Models," *Econometrica*, forthcoming.

Schwartz, L. (1966): *Théorie des distributions*. Paris, Hermann.

Wang, L., and C. Hsiao (1995): "Simulation-Based Semiparametric Estimation of Nonlinear Errors-in-Variables Models," Working Paper, University of Southern California.

White, H., and K. Chalak (2006): "A Unified Framework for Defining and Identifying Causal Effects," Working Paper, UCSD.

A Proofs

All lemmas and associated theorems given below make use of the definitions provided in the main text.

A.1 Identification

Lemma 1 *Assumptions 2 and 6 imply that (i)* $\lim_{x^*\to-\infty} E[S(y-q)|x^*]$ *and* $\lim_{x^*\to+\infty} E[S(y-q)|x^*]$ *exist, (ii) that* $\lim_{x^*\to-\infty} E[S(y-q)|x^*] = \lim_{z\to-\infty} E[S(y-q)|z]$ *and* $\lim_{x^*\to+\infty} E[S(y-q)|x^*] = \lim_{z\to+\infty} E[S(y-q)|z]$ *and (iii) that* $\sigma_{0x^*}(\zeta,q)$ *is an ordinary function.*

Proof. First, $\lim_{x^*\to+\infty} E[S(y-q)|x^*]$ is well defined because $E[S(y-q)|x^*=i]$ for $i=1,2,\ldots$ is Cauchy sequence, as it can be verified that for $i,j\in\mathbb{N}$ with $j>i$ and some $B',B''<\infty$ we have $|E[S(y-q)|x^*=j]-E[S(y-q)|x^*=i]| = \left|\int_i^j (\partial E[S(y-q)|x^*]/\partial x^*)\,dx^*\right| \le \int_i^j |\partial E[S(y-q)|x^*]/\partial x^*|\,dx^*$ which, by Assumption 6(i), is bounded by $\int_i^j B(1+|x^*|)^{-2-\beta}dx^* = \left[B'(1+|x^*|)^{-1-\beta}\right]_i^j \le B'' i^{-1-\beta}\to 0$ as $i\to\infty$. A similar reasoning shows that $\lim_{x^*\to-\infty} E[S(y-q)|x^*]$ is well defined.

Second, we will show that $\lim_{x^*\to+\infty} E[S(y-q)|x^*] = \lim_{z\to+\infty} E[S(y-q)|z] \equiv c_+(q)$. Assumption 2 implies that

$$E[S(y-q)|z] = \int E[S(y-q)|x^*=z+\Delta x^*]\, f_{\Delta x^*}(\Delta x^*)\, d(\Delta x^*).$$

By the Dominated Convergence Theorem (since $E[S(y-q)|x^*]$ is bounded by construction and Assumption 6(ii) states that $f_{\Delta x^*}(\Delta x^*)$ is bounded),

$$\begin{aligned}
\lim_{z\to+\infty} E[S(y-q)|z] &= \int \lim_{z\to+\infty} E[S(y-q)|x^*=z+\Delta x^*]\, f_{\Delta x^*}(\Delta x^*)\, d(\Delta x^*) \\
&= \int \left(\lim_{x^*\to+\infty} E[S(y-q)|x^*]\right) f_{\Delta x^*}(\Delta x^*)\, d(\Delta x^*) \\
&= \left(\lim_{x^*\to+\infty} E[S(y-q)|x^*]\right) \int f_{\Delta x^*}(\Delta x^*)\, d(\Delta x^*) \\
&= \lim_{x^*\to+\infty} E[S(y-q)|x^*],
\end{aligned}$$

a limit just shown to exist. Hence, the statement $c_+(q)\equiv \lim_{z\to+\infty} E[S(y-q)|z]$ is well-defined and we also have $c_+(q) = \lim_{x^*\to+\infty} E[S(y-q)|x^*]$. We can similarly state that $\lim_{x^*\to-\infty} E[S(y-q)|x^*] = \lim_{z\to-\infty} E[S(y-q)|z] \equiv c_-(q)$.

Third,

$$\int |E[S(y-q)|x^*] - H(x^*)|\, dx^* < \infty \tag{47}$$

where $H(x^*) = c_+(q)1(x^*>0) + c_-(q)1(x^*\le 0)$, since

$$\begin{aligned}
\int_{-\infty}^{0} |E[S(y-q)|x^*] - H(x^*)|\, dx^* &= \int_{-\infty}^{0} \left| c_-(q) + \int_{-\infty}^{x^*} \left[\frac{\partial E[S(y-q)|x^*]}{\partial x^*}\right]_{x^*=u} du - c_-(q)\right| dx^* \\
&\le \int_{-\infty}^{0}\int_{-\infty}^{x^*} \left|\left[\frac{\partial E[S(y-q)|x^*]}{\partial x^*}\right]_{x^*=u}\right| du dx^* \\
&\le \int_{-\infty}^{0}\int_{-\infty}^{x^*} B(1+|u|)^{-2-\beta}\, du dx^* = \int_{-\infty}^{0} B'(1+|x^*|)^{-1-\beta}\, dx^* \le B''
\end{aligned}$$

and similarly for $\int_0^\infty |E[S(y-q)|x^*] - H(x^*)|\,dx^*$. Hence, Equation (47) implies that $E[S(y-q)|x^*] - c(q)$ can be written as the sum of the function $H(x^*) - c(q)$ whose Fourier transform is $(c_+(q) - c_-(q))/(-\mathbf{i}\xi)$, and a function that is absolutely integrable, whose Fourier transform is necessarily bounded. In follows that $\sigma_{0x^*}(\zeta, q)$ is an ordinary function. ■

Proof of Theorem 1. We first note that, Assumption 2 implies that $f_{y|x^*,z}(y|x^*,z) = f_{y|x^*}(y|x^*)$ and that $f_{\Delta x^*|z}(\Delta x^*|z) = f_{\Delta x^*}(\Delta x^*)$. Hence, the quantity $E[S(y-q) - c(q)|z]$ can be related to $E[S(y-q) - c(q)|x^*]$ as follows:

$$E[S(y-q) - c(q)|z] = \int (S(y-q) - c(q)) f_{y|z}(y|z)\,dy \tag{48}$$

$$= \int (S(y-q) - c(q)) \int f_{y|x^*,z}(y|z-u,z) f_{\Delta x^*|z}(-u|z)\,dudy \tag{49}$$

$$= \int (S(y-q) - c(q)) \int f_{y|x^*,z}(y|z-u,z) f_{\Delta x^*}(-u)\,dudy \tag{50}$$

where we have used the change of variable $u = -\Delta x^*$, to make the innermost integral look like a convolution. Next, we can write

$$E[S(y-q) - c(q)|z] = \int \left(\int (S(y-q) - c(q)) f_{y|x^*,z}(y|z-u,z)\,dy \right) f_{\Delta x^*}(-u)\,du \tag{51}$$

$$= \int \left(\int (S(y-q) - c(q)) f_{y|x^*}(y|z-u)\,dy \right) f_{\Delta x^*}(-u)\,du \tag{52}$$

$$= \int E[(S(y-q) - c(q))|x^* = z-u] f_{\Delta x^*}(-u)\,du, \tag{53}$$

where we have used Fubini's Theorem to interchange the integrals after noting that, since $|S(y-q) - c(q)|$ is bounded by 1 (as $S(\cdot)$ is an indicator function and $c(q) \in [0,1]$) and since densities are positive, $\int\int |S(y-q) - c(q)| \left|f_{y|x^*,z}(y|z-u,z)\right| |f_{\Delta x^*}(-u)|\,dudy \leq \int\int f_{y|x^*,z}(y|z-u,z) f_{\Delta x^*}(-u)\,dudy = \int f_{y|z}(y|z)\,dy = 1$. Differentiating with respect to z on each side yields:

$$\frac{\partial}{\partial z} E[S(y-q)|z] = \int \left[\frac{\partial E[S(y-q)|x^*]}{\partial x^*} \right]_{x^*=z-u} f_{\Delta x^*}(-u)\,du,$$

where the interchange between integration and differentiation is allowed since $\partial E[S(y-q)|x^*]/\partial x^*$ is bounded by Assumption 6(i) and since $f_{\Delta x^*}(u)$ is absolutely integrable. Similarly, using the fact that Assumption 2 implies that $E[\Delta x|y,z]=0$, and $f_{y|x^*,z}(y|x^*,z)=f_{y|x^*}(y|x^*)$, we have

$$
\begin{aligned}
& E\left[(z-x)\,S(y-q)\,|z\right] && (54)\\
= \;& E\left[zS(y-q)\,|z\right]-E\left[x^*S(y-q)\,|z\right]-E\left[\Delta xS(y-q)\,|z\right] && (55)\\
= \;& E\left[zS(y-q)\,|z\right]-E\left[x^*S(y-q)\,|z\right]-E\left[E\left[\Delta x|y,z\right]S(y-q)\,|z\right] && (56)\\
= \;& E\left[zS(y-q)\,|z\right]-E\left[x^*S(y-q)\,|z\right] && (57)\\
= \;& E\left[(-\Delta x^*)\,S(y-q)\,|z\right] && (58)\\
= \;& \int S(y-q)\int f_{y|x^*,z}(y|z-u,z)\,uf_{\Delta x^*|z}(-u|z)\,dudy && (59)\\
= \;& \int S(y-q)\int f_{y|x^*}(y|z-u)\,uf_{\Delta x^*}(-u)\,dudy && (60)\\
= \;& \int\left(\int S(y-q)\,f_{y|x^*}(y|z-u)\,dy\right)uf_{\Delta x^*}(-u)\,du && (61)\\
= \;& \int E\left[S(y-q)\,|x^*=z-u\right]uf_{\Delta x^*}(-u)\,du, && (62)
\end{aligned}
$$

where we have used Fubini's Theorem again, since $\int\int|S(y-q)|\,f_{y|x^*}(y|z-u)\,|u|\,f_{\Delta x^*}(-u)\,dudy\le\int\int f_{y|x^*}(y|z-u)\,|u|\,f_{\Delta x^*}(-u)\,dudy=\int|u|\,f_{\Delta x^*}(-u)\,du<\infty$ by Assumption 5. Differentiating with respect to z on each side yields:

$$\frac{\partial}{\partial z}E\left[(z-x)\,S(y-q)\,|z\right]=\int\left[\frac{\partial E\left[S(y-q)\,|x^*\right]}{\partial x^*}\right]_{x^*=z-u}uf_{\Delta x^*}(-u)\,du,$$

where interchange of the integral and differential are allowed since $\partial E\left[S(y-q)\,|x^*\right]/\partial x^*$ is bounded by Assumption 6(i) and $|u|\,f_{\Delta x^*}(u)$ is absolutely integrable by Assumption 5. Applying a Fourier transform to Equations (53) and (62) enables us to use the Convolution Theorem (i.e. the Fourier transform of a convolution of two functions is equal to the product of their respective Fourier transforms) and a generalization of the Moment Theorem (which implies that the Fourier transform of $\mathbf{i}uf_{\Delta x^*}(-u)$ is $d\phi(\zeta)/d\zeta$). Equations (53) and (62) then become

$$
\begin{aligned}
\sigma_{0\partial z}(\zeta,q) &= \sigma_{0\partial x^*}(\zeta,q)\,\phi(\zeta) && (63)\\
\mathbf{i}\sigma_{1\partial z}(\zeta,q) &= \sigma_{0\partial x^*}(\zeta,q)\,\dot{\phi}(\zeta), && (64)
\end{aligned}
$$

where these quantities are given in Definition 1 and where dots denote derivatives with respect to ζ. Note that all Fourier transforms in Equations (63) and (64) exist as ordinary functions because they are the Fourier transforms of absolutely integrable functions by Assumptions 6(i) and 5. After taking the ratio of Equations (64) and (63) we obtain

$$\frac{\mathbf{i}\sigma_{1\partial z}(\zeta,q)}{\sigma_{0\partial z}(\zeta,q)}=\frac{\dot{\phi}(\zeta)}{\phi(\zeta)}. \qquad (65)$$

The ratio $\mathbf{i}\sigma_{1\partial z}(\zeta,q)/\sigma_{0\partial z}(\zeta,q)$ is almost everywhere defined because $\sigma_{0\partial z}(\zeta,q)\neq0$ almost everywhere by Assumption 4(ii). Under Assumption 5 and 4(i), the right-hand side is bounded and continuous in ζ. Hence, at any ζ where the left-hand side is undefined due to a division by zero, the numerator must be simultaneously zero and the limiting value of $\mathbf{i}\sigma_{1z}(\zeta, q)/\sigma_{0z}(\zeta, q)$ at the undefined point is well-defined and bounded. Hence, if we take the convention that $\mathbf{i}\sigma_{1z}(\zeta, q)/\sigma_{0z}(\zeta, q) \equiv \lim_{\xi\to\zeta} \mathbf{i}\sigma_{1z}(\xi, q)/\sigma_{0z}(\xi, q)$, Equation (65) holds for all $\zeta \in \mathbb{R}$. After integration of Equation (65) with respect to ζ and observing that $\phi(0) = 1$, we obtain:

$$\phi(\zeta) = \exp\left(\int_0^\zeta \frac{\mathbf{i}\sigma_{1\partial z}(\xi, q)}{\sigma_{0\partial z}(\xi, q)} d\xi\right). \tag{66}$$

Since this holds for any q, it holds in particular for $q = \tilde{q}$, thus establishing Equation (22) of the Theorem.

Now, starting from Equation (53) and applying a Fourier transform, we obtain

$$\sigma_{0z}(\zeta, q) = \sigma_{0x^*}(\zeta, q)\,\phi(\zeta). \tag{67}$$

Lemma 1 establishes that $\sigma_{0x^*}(\zeta, q)$ is a conventional function (it contains no "delta functions"), even though $E[S(y-q)|x^*] - c(q)$ is not absolutely integrable. Since $\phi(\zeta)$ is also an ordinary function, so is $\sigma_{0z}(\zeta, q)$. We can multiply each side of Equation (67) by $\phi^{-1}(\zeta)$ and obtain

$$\sigma_{0x^*}(\zeta, q) = \sigma_{0z}(\zeta, q)/\phi(\zeta) \tag{68}$$

Finally, Equation (21) merely states that $s_{0x^*}(x^*, q)$ is the inverse Fourier transform of $\sigma_{0x^*}(\zeta, q)$ given by Equation (68). ■

A.2 Consistency

Lemmas 2-6 will be used in the proof of Theorem 2.

Lemma 2 *Given functions $Y(W, \beta)$ and $X(W, \beta)$ of some random vector W and some parameter vector β belonging to some set $\mathcal{B}$, let $p(z|\beta)$ denote the density of $z = X(W, \beta)$ and let*[11] *$G_Y^{(d)}(z, \beta) = \partial^d(E[Y(W, \beta)|X(W, \beta) = z]\,p(z|\beta))/\partial z^d$ for $d \in \mathbb{N}$. Let W_j for $j = 1, \ldots, n$ be an iid sample of realizations of the random vector W. Under Assumption 9(i)-(iv), if $K(z)$ is d times differentiable (with $d < N_K$) and if $\sup_{\beta\in\mathcal{B}}\left(E\left[|Y(W, \beta)|^2\right]\right) < \infty$ and $\sup_{\beta\in\mathcal{B}}\sup_{z\in\mathbb{R}}\left|G_Y^{(N_k)}(z, \beta)\right| < \infty$ (for N_K as in Assumption 9), then*

$$\sup_{\beta\in\mathcal{B}}\sup_{z\in\mathbb{R}}\left|\hat{G}_Y^{(d)}(z, \beta) - G_Y^{(d)}(z, \beta)\right| = O_p\left(n^{-1/2}h_n^{-1-d}\right) + O\left(h_n^{N_K - d}\right)$$

where

$$\hat{G}_Y^{(d)}(z, \beta) = (nh_n)^{-1} h_n^{-d}\sum_{j=1}^n Y(W_j, \beta)\,K^{(d)}((z_j - z)/h_n)$$

and where $z_j = X(W_j, \beta)$ and $K^{(d)}(z) = \partial^d K(z)/\partial z^d$.

Proof. Note that $\sup_{\beta\in\mathcal{B}}\sup_{z\in\mathbb{R}}\left|\hat{G}_Y^{(d)}(z, \beta) - G_Y^{(d)}(z, \beta)\right| \leq R + B$, where

$$\begin{aligned} R &= \sup_{\beta\in\mathcal{B}}\sup_{z\in\mathbb{R}}\left|\hat{G}_Y^{(d)}(z, \beta) - E\left[\hat{G}_Y^{(d)}(z, \beta)\right]\right|, \\ B &= \sup_{\beta\in\mathcal{B}}\sup_{z\in\mathbb{R}}\left|E\left[\hat{G}_Y^{(d)}(z, \beta)\right] - G_Y^{(d)}(z, \beta)\right|. \end{aligned}$$

[11] By convention $G_Y^{(0)}(z, \beta) \equiv G_Y(z, \beta)$.

Using the Convolution Theorem, a kernel estimate of $G_Y^{(0)}(z,\beta)$ can be written as

$$\hat{G}_Y^{(0)}(z,\beta) = \int \kappa(h_n\zeta)\, n^{-1}\sum_{j=1}^{n} Y(W_j,\beta)\, e^{\mathbf{i}\zeta z_j} e^{-\mathbf{i}\zeta z} d\zeta$$

where $\kappa(\zeta)$ denotes the Fourier transform of $K(z)$. Subtracting the corresponding expected value, differentiating this expression with respect to z and using the moment Theorem yields the following expression for R

$$\begin{aligned} R &= \sup_{\beta\in\mathcal{B}}\sup_{z\in\mathbb{R}}\left|\int(-\mathbf{i}\zeta)^d \kappa(h_n\zeta)\, n^{-1}\sum_{j=1}^{n}\left(Y(W_j,\beta)\, e^{\mathbf{i}\zeta z_j} - E\left[Y(W_j,\beta)\, e^{\mathbf{i}\zeta z_j}\right]\right) e^{-\mathbf{i}\zeta z} d\zeta\right| \\ &\le h_n^{-d}\sup_{\beta\in\mathcal{B}}\int |h_n\zeta|^d\, |\kappa(h_n\zeta)| \left| n^{-1}\sum_{j=1}^{n}\left(Y(W_j,\beta)\, e^{\mathbf{i}\zeta z_j} - E\left[Y(W_j,\beta)\, e^{\mathbf{i}\zeta z_j}\right]\right)\right| d\zeta. \end{aligned}$$

We then have

$$\begin{aligned} E[R] &\le h_n^{-d}\sup_{\beta\in\mathcal{B}}\int |h_n\zeta|^d\, |\kappa(h_n\zeta)|\, E\left[\left| n^{-1}\sum_{j=1}^{n}\left(Y(W_j,\beta)\, e^{\mathbf{i}\zeta z_j} - E\left[Y(W_j,\beta)\, e^{\mathbf{i}\zeta z_j}\right]\right)\right|\right] d\zeta \\ &\le h_n^{-d}\sup_{\beta\in\mathcal{B}}\int |h_n\zeta|^d\, |\kappa(h_n\zeta)| \left(E\left[\left| n^{-1}\sum_{j=1}^{n}\left(Y(W_j,\beta)\, e^{\mathbf{i}\zeta z_j} - E\left[Y(W_j,\beta)\, e^{\mathbf{i}\zeta z_j}\right]\right)\right|^2\right]\right)^{1/2} d\zeta \\ &= h_n^{-d}\sup_{\beta\in\mathcal{B}}\int |h_n\zeta|^d\, |\kappa(h_n\zeta)| (n^{-1}E[\left(Y(W_j,\beta)\, e^{\mathbf{i}\zeta z_j} - E\left[Y(W_j,\beta)\, e^{\mathbf{i}\zeta z_j}\right]\right)\times \\ &\quad \times\left(Y(W_j,\beta)\, e^{-\mathbf{i}\zeta z_j} - E\left[Y(W_j,\beta)\, e^{-\mathbf{i}\zeta z_j}\right]\right)])^{1/2} d\zeta \\ &= h_n^{-d} n^{-1/2}\sup_{\beta\in\mathcal{B}}\int |h_n\zeta|^d\, |\kappa(h_n\zeta)| (E[\left(Y(W_j,\beta)\, e^{\mathbf{i}\zeta z_j} - E\left[Y(W_j,\beta)\, e^{\mathbf{i}\zeta z_j}\right]\right)\times \\ &\quad \times\left(Y(W_j,\beta)\, e^{-\mathbf{i}\zeta z_j} - E\left[Y(W_j,\beta)\, e^{-\mathbf{i}\zeta z_j}\right]\right)])^{1/2} d\zeta \\ &\le h_n^{-d} n^{-1/2} 2^{1/2}\sup_{\beta\in\mathcal{B}}\left(E\left[Y^2(W_j,\beta)\right]\right)^{1/2}\int |h_n\zeta|^d\, |\kappa(h_n\zeta)|\, d\zeta \\ &= h_n^{-d} n^{-1/2} h_n^{-1} 2^{1/2}\sup_{\beta\in\mathcal{B}}\left(E\left[Y^2(W_j,\beta)\right]\right)^{1/2}\int |\zeta|^d\, |\kappa(\zeta)|\, d\zeta \\ &= O\left(n^{-1/2} h_n^{-1-d}\right), \end{aligned}$$

where, on the third line, we have used the fact that, for any bounded variance complex-valued random i.i.d. sequence a_j,

$$\begin{aligned} E\left[\left| n^{-1}\sum_{j=1}^{n}(a_j - E[a_j])\right|^2\right] &= E\left[\left(n^{-1}\sum_{j=1}^{n}(a_j - E[a_j])\right)\left(n^{-1}\sum_{j=1}^{n}(a_j - E[a_j])\right)^{\dagger}\right] \\ &= n^{-2}\sum_{j=1}^{n}\sum_{j'=1}^{n} E\left[(a_j - E[a_j])(a_{j'} - E[a_{j'}])^{\dagger}\right] \\ &= n^{-2}\sum_{j=1}^{n} E\left[(a_j - E[a_j])(a_{j'} - E[a_{j'}])^{\dagger}\right] \\ &= n^{-1} E\left[(a_j - E[a_j])(a_{j'} - E[a_{j'}])^{\dagger}\right], \end{aligned}$$

where † denotes complex conjugates. On the fourth line, we have used the fact that

$$\begin{aligned}
& E\left[\left(Y(W_j,\beta)e^{\mathbf{i}\zeta z_j}-E\left[Y(W_j,\beta)e^{\mathbf{i}\zeta z_j}\right]\right)\left(Y(W_j,\beta)e^{-\mathbf{i}\zeta z_j}-E\left[Y(W_j,\beta)e^{-\mathbf{i}\zeta z_j}\right]\right)\right] \\
= & E\left[Y(W_j,\beta)e^{\mathbf{i}\zeta z_j}Y(W_j,\beta)e^{-\mathbf{i}\zeta z_j}\right]-E\left[Y(W_j,\beta)e^{\mathbf{i}\zeta z_j}E\left[Y(W_j,\beta)e^{-\mathbf{i}\zeta z_j}\right]\right]+ \\
& -E\left[E\left[Y(W_j,\beta)e^{\mathbf{i}\zeta z_j}\right]Y(W_j,\beta)e^{-\mathbf{i}\zeta z_j}\right]+E\left[E\left[Y(W_j,\beta)e^{\mathbf{i}\zeta z_j}\right]E\left[Y(W_j,\beta)e^{-\mathbf{i}\zeta z_j}\right]\right] \\
= & E\left[Y^2(W_j,\beta)\right]-E\left[Y(W_j,\beta)e^{\mathbf{i}\zeta z_j}\right]E\left[Y(W_j,\beta)e^{-\mathbf{i}\zeta z_j}\right] \\
\leq & E\left[Y^2(W_j,\beta)\right]+E\left[Y^2(W_j,\beta)\right] \\
= & 2E\left[Y^2(W_j,\beta)\right].
\end{aligned}$$

Since $E[R]=O\left(n^{-1/2}h_n^{-1-d}\right)$, it follows that $R=O_p\left(n^{-1/2}h_n^{-1-d}\right)$ by Markov's inequality.

Next, by integration by parts,

$$\begin{aligned}
B &= \sup_{\beta\in\mathcal{B}}\sup_{z\in\mathbb{R}}\left|E\left[\hat{G}_Y^{(d)}(z,\beta)\right]-G_Y^{(d)}(z,\beta)\right| \\
&= \sup_{\beta\in\mathcal{B}}\sup_{z\in\mathbb{R}}\left|\int h_n^{-1}K\left(h_n^{-1}v\right)\left(G_Y^{(d)}(z+v,\beta)-G_Y^{(d)}(z,\beta)\right)dv\right|.
\end{aligned}$$

By a Taylor expansion,

$$\begin{aligned}
B &= \sup_{\beta\in\mathcal{B}}\sup_{z\in\mathbb{R}}\left|\int h_n^{-1}K\left(h_n^{-1}v\right)\left(\sum_{j=1}^{N_k-d-1}G_Y^{(d+j)}(z,\beta)\frac{v^j}{j!}+G_Y^{(N_k)}(\tilde{z},\beta)\frac{v^{N_k-d}}{(N_k-d)!}\right)dv\right| \text{ for } \tilde{z}\in[z,z+v] \\
&= \sup_{\beta\in\mathcal{B}}\sup_{z\in\mathbb{R}}\left|\int h_n^{-1}K\left(h_n^{-1}v\right)G_Y^{(N_k)}(\tilde{z},\beta)\frac{v^{N_k-d}}{(N_k-d)!}dv\right|
\end{aligned}$$

by Assumption 9(iii). Then, by a change of variable,

$$\begin{aligned}
B &= \sup_{\beta\in\mathcal{B}}\sup_{z\in\mathbb{R}}\left|\int K(u)\,G_Y^{(N_k)}(\tilde{z},\beta)\frac{u^{N_k-d}h_n^{N_k-d}}{(N_k-d)!}du\right| \\
&\leq h_n^{N_k-d}\left(\sup_{\beta\in\mathcal{B}}\sup_{\tilde{z}\in\mathbb{R}}\left|G_Y^{(N_k)}(\tilde{z},\beta)\right|\right)\frac{1}{N_k!}\left|\int|K(u)|\,|u|^{N_k-d}du\right|=O\left(h_n^{N_k-d}\right)
\end{aligned}$$

by Assumptions 9(iv) and the assumed boundedness of $G_Y^{(N_k)}(\tilde{z},\beta)$. ■

Lemma 3 *Given a sequence of sets $\mathbb{T}_n$, two sequences of functions a_n, b_n and two sequences of random functions $\hat{a}_n$, $\hat{b}_n$ each mapping $\mathbb{T}_n$ to $\mathbb{R}$, let ε_n^a and ε_n^b be deterministic sequences such that* $\sup_{t\in\mathbb{T}_n}|\hat{a}_n(t)-a_n(t)|=O_p(\varepsilon_n^a)$ *and* $\sup_{t\in\mathbb{T}_n}\left|\hat{b}_n(t)-b_n(t)\right|=O_p\left(\varepsilon_n^b\right)$, *and let* $d_n=\inf_{t\in\mathbb{T}_n}|b_n(t)|$ *and* $R_n\equiv\sup_{t\in\mathbb{T}_n}|a_n(t)/b_n(t)|$. *If* $\varepsilon_n^b/d_n\to 0$, *then*

$$\sup_{t\in\mathbb{T}_n}\left|\frac{\hat{a}_n(t)}{\hat{b}_n(t)}-\frac{a_n(t)}{b_n(t)}\right|=O_p\left(\frac{\varepsilon_n^a}{d_n}\right)+O_p\left(R_n\frac{\varepsilon_n^b}{d_n}\right).$$

Proof. After simple manipulations, we have

$$\frac{\hat{a}_n(t)}{\hat{b}_n(t)}-\frac{a_n(t)}{b_n(t)}=\left(\frac{\hat{a}_n(t)-a_n(t)}{b_n(t)}+\frac{a_n(t)}{b_n(t)}\frac{b_n(t)-\hat{b}_n(t)}{b_n(t)}\right)\left(1+\frac{\hat{b}_n(t)-b_n(t)}{b_n(t)}\right)^{-1}$$

which implies that

$$\sup_{t\in\mathbb{T}_n}\left|\frac{\hat{a}(t)}{\hat{b}(t)}-\frac{a(t)}{b(t)}\right| = \left(O_p\left(\frac{\varepsilon_n^a}{d_n}\right)+O_p\left(R_n\frac{\varepsilon_n^b}{d_n}\right)\right)\left(1+O_p\left(\frac{\varepsilon_n^b}{d_n}\right)\right)^{-1}.$$

The conclusion follows by noting that $\varepsilon_n^b/d_n \to 0$ implies that $\left(1+O_p\left(\varepsilon_n^b/d_n\right)\right)^{-1} = (1+o_p(1))^{-1} = 1+o_p(1)$. ■

Lemma 4 *Let Assumptions 7, 9, 10 and 11 hold. If* $\underline{f}_n$ *as in Definition 4 satisfies* $\underline{f}_n^{-1} = o\left(n^{1/2}h_n\right) + o\left(h_n^{-N_k}\right)$ *then, for* $k=0,1$,

$$\begin{aligned}\sup_{q\in\mathcal{Q}}\sup_{\alpha\in\mathcal{A}}\sup_{|z|\le\bar{z}_n}\left|\hat{s}_{k\partial z}(z,q,\alpha)-s_{k\partial z}(z,q,\alpha)\right| &= O_p\left(\left(1+h_n\underline{f}_n^{-1}\right)\underline{f}_n^{-1}n^{-1/2}h_n^{-2}\right)+\\ &\quad +O\left(\left(1+h_n\underline{f}_n^{-1}\right)\underline{f}_n^{-1}h_n^{N_K-1}\right)\\ \sup_{q\in\mathcal{Q}}\sup_{\alpha\in\mathcal{A}}\sup_{|z|\le\bar{z}_n}\left|\hat{s}_{0z}(z,q,\alpha)-s_{0z}(z,q,\alpha)\right| &= O_p\left(\underline{f}_n^{-1}n^{-1/2}h_n^{-1}\right)+O\left(\underline{f}_n^{-1}h_n^{N_K}\right).\end{aligned}$$

Proof. First observe that under the assumptions made, all the hypotheses of Lemma 2 hold for $\beta=(q',\alpha')'$, $\mathcal{B}=\mathcal{Q}\times\mathcal{A}$, $W=(y,x,w)$, $X(W,\beta)=Z(w,\alpha)$, for $d=0$ or $d=1$ and for $Y(W,\beta)$ set to either $S(y-q)$, $(Z(w,\alpha)-x)S(y-q)$ or 1. Let $\hat{G}_Y^{(d)}(z,\beta)$ and $G_Y^{(d)}(z,\beta)$ be defined as in Lemma 2. Let $\varepsilon_n^{(d)}$ be such that $\max_Y\sup_{\beta\in\mathcal{B}}\sup_{|z|\le z_n}\left|\hat{G}_Y^{(d)}(z,\beta)-G_Y^{(d)}(z,\beta)\right| = O_p\left(\varepsilon_n^{(d)}\right)$ (where the maximum over Y is taken over $Y=S(y-q)$, $(Z(w,\alpha)-x)S(y-q)$ or 1). We bound the estimation error in $\hat{s}_{k\partial z}(z,q,\alpha)$ as follows.

$$\begin{aligned}&\hat{s}_{k\partial z}(z,q,\alpha)-s_{k\partial z}(z,q,\alpha)=\frac{\partial}{\partial z}\left(\frac{\hat{G}_Y(z,\beta)}{\hat{G}_1(z,\beta)}-\frac{G_Y(z,\beta)}{G_1(z,\beta)}\right)\\ =&\left(\frac{\hat{G}_Y^{(1)}(z,\beta)}{\hat{G}_1(z,\beta)}-\frac{G_Y^{(1)}(z,\beta)}{G_1(z,\beta)}\right)-\left(\frac{\hat{G}_Y(z,\beta)}{\hat{G}_1(z,\beta)}\frac{\hat{G}_1^{(1)}(z,\beta)}{\hat{G}_1(z,\beta)}-\frac{G_Y(z,\beta)}{G_1(z,\beta)}\frac{G_1^{(1)}(z,\beta)}{G_1(z,\beta)}\right)\\ =&\left(\frac{\hat{G}_Y^{(1)}(z,\beta)}{\hat{G}_1(z,\beta)}-\frac{G_Y^{(1)}(z,\beta)}{G_1(z,\beta)}\right)-\frac{G_Y(z,\beta)}{G_1(z,\beta)}\left(\frac{\hat{G}_1^{(1)}(z,\beta)}{\hat{G}_1(z,\beta)}-\frac{G_1^{(1)}(z,\beta)}{G_1(z,\beta)}\right)+\\ &-\frac{G_1^{(1)}(z,\beta)}{G_1(z,\beta)}\left(\frac{\hat{G}_Y(z,\beta)}{\hat{G}_1(z,\beta)}-\frac{G_Y(z,\beta)}{G_1(z,\beta)}\right)-\left(\frac{\hat{G}_Y(z,\beta)}{\hat{G}_1(z,\beta)}-\frac{G_Y(z,\beta)}{G_1(z,\beta)}\right)\left(\frac{\hat{G}_1^{(1)}(z,\beta)}{\hat{G}_1(z,\beta)}-\frac{G_1^{(1)}(z,\beta)}{G_1(z,\beta)}\right),\end{aligned}$$

where $Y=(Z(w,\alpha)-x)^k S(y-q)$ for $k=0,1$. By Lemma 2, $\varepsilon_n^{(0)}=n^{-1/2}h_n^{-1}+h_n^{N_K}$ and it follows that $\underline{f}_n^{-1}\varepsilon_n^{(0)}\to 0$ since $\underline{f}_n^{-1}\equiv\sup_{|z|\le\bar{z}_n}(G_1(z,\beta))^{-1}$ is assumed to be $o\left(n^{1/2}h_n\right)+o\left(h_n^{-N_k}\right)$. We can therefore use Lemma 3 to bound each difference of ratios after noting that $|G_Y(z,\beta)/G_1(z,\beta)|\equiv|s_{k\partial z}(z,q,\alpha)|$ is bounded by Assumption 10 and $\left|G_1^{(1)}(z,\beta)\right|$ is bounded by Assumption 11, for $j=1$. We obtain

$$\sup_{q\in\mathcal{Q}}\sup_{\alpha\in\mathcal{A}}\sup_{|z|\le\bar{z}_n}\left|\hat{s}_{k\partial z}(z,q,\alpha)-s_{k\partial z}(z,q,\alpha)\right|=O_p\left(\underline{f}_n^{-1}\varepsilon_n^{(1)}\right)+O_p\left(\underline{f}_n^{-1}\varepsilon_n^{(1)}\right)+O_p\left(\underline{f}_n^{-2}\varepsilon_n^{(0)}\right)+O_p\left(\underline{f}_n^{-2}\varepsilon_n^{(0)}\varepsilon_n^{(1)}\right).$$

By Lemma 2, $\varepsilon_n^{(d)}=n^{-1/2}h_n^{-1-d}+h_n^{N_K-d}$ and it follows that

$$\begin{aligned}&\sup_{q\in\mathcal{Q}}\sup_{\alpha\in\mathcal{A}}\sup_{|z|\le\bar{z}_n}\left|\hat{s}_{k\partial z}(z,q,\alpha)-s_{k\partial z}(z,q,\alpha)\right|\\ =&\ O_p\left(\underline{f}_n^{-1}n^{-1/2}h_n^{-2}\right)+O\left(\underline{f}_n^{-1}h_n^{N_K-1}\right)+O_p\left(\underline{f}_n^{-2}n^{-1/2}h_n^{-1}\right)+O\left(\underline{f}_n^{-2}h_n^{N_K}\right)\\ =&\ O_p\left(\left(1+h_n\underline{f}_n^{-1}\right)\underline{f}_n^{-1}n^{-1/2}h_n^{-2}\right)+O\left(\left(1+h_n\underline{f}_n^{-1}\right)\underline{f}_n^{-1}h_n^{N_K-1}\right).\end{aligned}$$

We proceed similarly for $Y(W,\beta) = S(y-q)$ and for $d=0$, noting that

$$\hat{s}_{0z}(z,q,\alpha) - s_{0z}(z,q,\alpha) = \frac{\hat{G}_Y(z,\beta)}{\hat{G}_1(z,\beta)} - \frac{G_Y(z,\beta)}{G_1(z,\beta)}.$$

By Lemma 2 we then have

$$\sup_{\tilde{Y}=1,S(y-q)} \sup_{\beta\in\mathcal{B}} \sup_{|z|\le\bar{z}_n} \left|\hat{G}_{\tilde{Y}}(z,\beta) - G_{\tilde{Y}}(z,\beta)\right| = O_p\left(n^{-1/2}h_n^{-1}\right) + O\left(h_n^{N_K}\right),$$

which implies that $\sup_{q\in\mathcal{Q}}\sup_{\alpha\in\mathcal{A}}\sup_{|z|\le\bar{z}_n}|\hat{s}_{0z}(z,q,\alpha) - s_{0z}(z,q,\alpha)| = O_p\left(\underline{f}_n^{-1}n^{-1/2}h_n^{-1}\right) + O\left(\underline{f}_n^{-1}h_n^{N_K}\right)$ by Lemma 3. ■

Lemma 5 *Under Assumptions 7, 9, 10, 11, 13, we have, for* $k=0,1$,

$$\begin{aligned}
\sup_{q\in\mathcal{Q}}\sup_{\zeta\in\mathbb{R}}\sup_{\alpha\in\mathcal{A}}|\hat{\sigma}_{k\partial z}(\zeta,q,\alpha) - \sigma_{k\partial z}(\zeta,q,\alpha)| &= O_p\left(\left(1+h_n\underline{f}_n^{-1}\right)\bar{z}_n n^{-1/2}h_n^{-2}\underline{f}_n^{-1}\right) + \\
&\quad + O\left(\left(1+h_n\underline{f}_n^{-1}\right)\bar{z}_n h_n^{N_K-1}\underline{f}_n^{-1}\right) + O(T_n) \\
\sup_{q\in\mathcal{Q}}\sup_{\zeta\in\mathbb{R}}\sup_{\alpha\in\mathcal{A}}|\hat{\sigma}_{0z}(z,q,\alpha) - \sigma_{0z}(z,q,\alpha)| &= O_p\left(\bar{z}_n n^{-1/2}h_n^{-1}\underline{f}_n^{-1}\right) + O\left(\bar{z}_n h_n^{N_K}\underline{f}_n^{-1}\right) + O(T_n).
\end{aligned}$$

Proof. Observe that

$$\begin{aligned}
&\sup_{q\in\mathcal{Q}}\sup_{\zeta\in\mathbb{R}}\sup_{\alpha\in\mathcal{A}}|\hat{\sigma}_{k\partial z}(\zeta,q,\alpha) - \sigma_{k\partial z}(\zeta,q,\alpha)| \\
=\ &\sup_{q\in Q}\sup_{\zeta\in\mathbb{R}}\sup_{\alpha\in\mathcal{A}}\left|\int \hat{s}_{k\partial z}(z,q,\alpha) - s_{k\partial z}(z,q,\alpha)e^{\mathbf{i}\zeta z}dz\right| \\
\le\ &\sup_{q\in Q}\sup_{\zeta\in\mathbb{R}}\sup_{\alpha\in\mathcal{A}}\left|\int_{|z|\le\bar{z}_n} \hat{s}_{k\partial z}(z,q,\alpha) - s_{k\partial z}(z,q,\alpha)e^{\mathbf{i}\zeta z}dz\right| + \\
&+\sup_{q\in Q}\sup_{\zeta\in\mathbb{R}}\sup_{\alpha\in\mathcal{A}}\left|\int_{|z|\ge\bar{z}_n} \hat{s}_{k\partial z}(z,q,\alpha) - s_{k\partial z}(z,q,\alpha)e^{\mathbf{i}\zeta z}dz\right| \\
\le\ &\sup_{q\in Q}\sup_{\alpha\in\mathcal{A}}\int_{|z|\le\bar{z}_n}|\hat{s}_{k\partial z}(z,q,\alpha) - s_{k\partial z}(z,q,\alpha)|dz + \sup_{q\in Q}\sup_{\alpha\in\mathcal{A}}\int_{|z|\ge\bar{z}_n}|0 - s_{k\partial z}(z,q,\alpha)|dz \\
\le\ &2\bar{z}_n\sup_{q\in Q}\sup_{\alpha\in\mathcal{A}}|\hat{s}_{k\partial z}(z,q,\alpha) - s_{k\partial z}(z,q,\alpha)| + O(T_n) \\
=\ &O_p\left(\left(1+h_n\underline{f}_n^{-1}\right)\bar{z}_n n^{-1/2}h_n^{-2}\underline{f}_n^{-1}\right) + O\left(\left(1+h_n\underline{f}_n^{-1}\right)\bar{z}_n h_n^{N_K-1}\underline{f}_n^{-1}\right) + O(T_n).
\end{aligned}$$

by Lemma 4 (since Assumption 13 and the fact that $\bar{z}_n\bar{\zeta}_n^2\bar{\phi}_n/\left(\underline{\phi}_n\underline{\sigma}_n\right)\to\infty$ imply that $\underline{f}_n^{-1}$ has the appropriate order of magnitude for Lemma 4 to apply[12]) and we have used the definition of T_n given in Definition 4. The second conclusion of the Lemma follows similarly. ■

[12] This can be seen as follows. The first two conditions in Assumption 13 are equivalent to $\underline{f}_n^{-1}\left(n^{-1/2}h_n^{-2} + h_n^{N_K-1}\right)\left(1+h_n\underline{f}_n^{-1}\right)\frac{\bar{z}_n\bar{\zeta}_n^2\bar{\phi}_n}{\underline{\phi}_n\underline{\sigma}_n}\to 0$. Moreover, $\bar{z}_n\bar{\zeta}_n^2\bar{\phi}_n/\left(\underline{\phi}_n\underline{\sigma}_n\right)\to\infty$ by construction and $\left(1+h_n\underline{f}_n^{-1}\right)$ cannot go to zero. Hence, this condition can only be satisfied if $\underline{f}_n^{-1}\left(n^{-1/2}h_n^{-2} + h_n^{N_K-1}\right) = o(1)$, i.e., if $\underline{f}_n^{-1} = o\left(n^{1/2}h_n^2 + h_n^{-N_K+1}\right) = o\left(n^{1/2}h_n + h_n^{-N_K}\right)h_n$. Since $h_n = o(1)$, having $\underline{f}_n^{-1} = o\left(n^{1/2}h_n + h_n^{-N_K}\right)h_n$ implies that $\underline{f}_n^{-1} = o\left(n^{1/2}h_n + h_n^{-N_K}\right)$, which is the hypothesis of Lemma 4.

Lemma 6 *Under Assumptions 7, 9-13,*

$$\sup_{x^*\in\mathbb{R}}\sup_{q\in\mathcal{Q}}\sup_{\alpha\in\mathcal{A}}\left|\hat{s}_{0x^*}\left(x^*,q,\alpha\right)-s_{0x^*}\left(x^*,q,\alpha\right)\right|=o_p\left(1\right)$$

Proof. We first establish that $\sigma_{0z}(\zeta,q,\alpha)$ is bounded by $\bar{\sigma}\max\left(|\zeta|^{-1},1\right)$ for some $\bar{\sigma}>0$. As in the proof of Lemma 1, part (iii), Assumption 10 implies that $s_{0z}(z,q,\alpha)$ for given values of q and α can be written as the sum of the function $\bar{H}(z)=c_+(q)\,1(z\geq 0)+c_-(q)\,1(z<0)-c(q)$ and an absolutely integrable function. Hence, the Fourier transform $\sigma_{0z}(\zeta,q,\alpha)$ can be written as the sum of the Fourier transform of $\bar{H}(z)$ (which is equal to $\left(c_+(q)-c_-(q)\right)/\left(-\mathbf{i}\zeta\right)$) and the Fourier transform of an absolutely integrable function, which is necessarily bounded. It follows that $|\sigma_{0z}(\zeta,q,\alpha)|\leq\bar{\sigma}\max\left(|\zeta|^{-1},1\right)$ for some $\bar{\sigma}>0$ that does not depend on q or α since the bound in Assumption 10 is uniform in q and α.

Lemma 5 provides the rates of convergence of the following quantities:

$$\begin{aligned}
\hat{\varepsilon}_n^{(1)} &= \sup_{\zeta\in\mathbb{R}}\sup_{q\in\mathcal{Q}}\sup_{\alpha\in\mathcal{A}}\max\left\{\left|\hat{\sigma}_{0\partial z}(\zeta,q,\alpha)-\sigma_{0\partial z}(\zeta,q,\alpha)\right|,\left|\hat{\sigma}_{1\partial z}(\zeta,q,\alpha)-\sigma_{1\partial z}(\zeta,q,\alpha)\right|\right\}\\
\hat{\varepsilon}_n^{(0)} &= \sup_{\zeta\in\mathbb{R}}\sup_{q\in\mathcal{Q}}\sup_{\alpha\in\mathcal{A}}\left|\hat{\sigma}_{0z}(\zeta,q,\alpha)-\sigma_{0z}(\zeta,q,\alpha)\right|.
\end{aligned}$$

We can then bound the estimation error as follows:

$$\begin{aligned}
& 2\pi\left|\hat{s}_{0x^*}(x^*,q)-s_{0x^*}(x^*,q)\right|\\
=\ & \Bigg|\int_{-\bar{\zeta}_n}^{\bar{\zeta}_n}\hat{\sigma}_{0z}(\zeta,q,\alpha)\exp\left(-\int_0^\zeta\frac{\mathbf{i}\hat{\sigma}_{1\partial z}(\xi,\tilde{q},\alpha)}{\hat{\sigma}_{0\partial z}(\xi,\tilde{q},\alpha)}d\xi\right)e^{-\mathbf{i}\zeta x^*}d\zeta+\\
& -\int_{-\bar{\zeta}_n}^{\bar{\zeta}_n}\sigma_{0z}(\zeta,q,\alpha)\exp\left(-\int_0^\zeta\frac{\mathbf{i}\sigma_{1\partial z}(\xi,\tilde{q},\alpha)}{\sigma_{0\partial z}(\xi,\tilde{q},\alpha)}d\xi\right)e^{-\mathbf{i}\zeta x^*}d\zeta-\int_{|\zeta|\geq\bar{\zeta}_n}\sigma_{0x^*}(\zeta,q,\alpha)\,e^{-\mathbf{i}\zeta x^*}d\zeta\Bigg|\\
\leq\ & R_1+R_2+R_3
\end{aligned}$$

where

$$\begin{aligned}
R_1 &= \left|\int_{-\bar{\zeta}_n}^{\bar{\zeta}_n}\hat{\sigma}_{0z}(\zeta,q,\alpha)\left(\exp\left(-\int_0^\zeta\frac{\mathbf{i}\hat{\sigma}_{1\partial z}(\xi,\tilde{q},\alpha)}{\hat{\sigma}_{0\partial z}(\xi,\tilde{q},\alpha)}d\xi\right)-\exp\left(-\int_0^\zeta\frac{\mathbf{i}\sigma_{1\partial z}(\xi,\tilde{q},\alpha)}{\sigma_{0\partial z}(\xi,\tilde{q},\alpha)}d\xi\right)\right)e^{-\mathbf{i}\zeta x^*}d\zeta\right|\\
&\leq \int_{-\bar{\zeta}_n}^{\bar{\zeta}_n}\left|\hat{\sigma}_{0z}(\zeta,q,\alpha)\right|\left|\exp\left(-\int_0^\zeta\frac{\mathbf{i}\hat{\sigma}_{1\partial z}(\xi,\tilde{q},\alpha)}{\hat{\sigma}_{0\partial z}(\xi,\tilde{q},\alpha)}d\xi\right)-\exp\left(-\int_0^\zeta\frac{\mathbf{i}\sigma_{1\partial z}(\xi,\tilde{q},\alpha)}{\sigma_{0\partial z}(\xi,\tilde{q},\alpha)}d\xi\right)\right|d\zeta\\
&\leq \int_{-\bar{\zeta}_n}^{\bar{\zeta}_n}\left(\left|\sigma_{0z}(\zeta,q,\alpha)\right|+\hat{\varepsilon}_n^{(0)}\right)\left|\exp\left(-\int_0^\zeta\frac{\mathbf{i}\hat{\sigma}_{1\partial z}(\xi,\tilde{q},\alpha)}{\hat{\sigma}_{0\partial z}(\xi,\tilde{q},\alpha)}d\xi\right)-\exp\left(-\int_0^\zeta\frac{\mathbf{i}\sigma_{1\partial z}(\xi,\tilde{q},\alpha)}{\sigma_{0\partial z}(\xi,\tilde{q},\alpha)}d\xi\right)\right|d\zeta\\
&\leq \int_{-\bar{\zeta}_n}^{\bar{\zeta}_n}\left(\bar{\sigma}\max\left(|\zeta|^{-1},1\right)+\hat{\varepsilon}_n^{(0)}\right)\left|\exp\left(-\int_0^\zeta\frac{\mathbf{i}\hat{\sigma}_{1\partial z}(\xi,\tilde{q},\alpha)}{\hat{\sigma}_{0\partial z}(\xi,\tilde{q},\alpha)}d\xi\right)-\exp\left(-\int_0^\zeta\frac{\mathbf{i}\sigma_{1\partial z}(\xi,\tilde{q},\alpha)}{\sigma_{0\partial z}(\xi,\tilde{q},\alpha)}d\xi\right)\right|d\zeta\\
&= \int_{-\bar{\zeta}_n}^{\bar{\zeta}_n}\left(\bar{\sigma}\max\left(|\zeta|^{-1},1\right)+\hat{\varepsilon}_n^{(0)}\right)\left|\exp\left(-\int_0^\zeta\frac{\mathbf{i}\sigma_{1\partial z}(\xi,\tilde{q},\alpha)}{\sigma_{0\partial z}(\xi,\tilde{q},\alpha)}d\xi\right)\right|\times\\
&\quad\times\left|\exp\left(-\int_0^\zeta\frac{\mathbf{i}\hat{\sigma}_{1\partial z}(\xi,\tilde{q},\alpha)}{\hat{\sigma}_{0\partial z}(\xi,\tilde{q},\alpha)}d\xi+\int_0^\zeta\frac{\mathbf{i}\sigma_{1\partial z}(\xi,\tilde{q},\alpha)}{\sigma_{0\partial z}(\xi,\tilde{q},\alpha)}d\xi\right)-1\right|d\zeta
\end{aligned}$$

$$
\begin{aligned}
&\leq \underline{\phi}_n^{-1}\int_{-\bar{\zeta}_n}^{\bar{\zeta}_n}\left(\bar{\sigma}\max\left(|\zeta|^{-1},1\right)+\hat{\varepsilon}_n^{(0)}\right)\left|\exp\left(-\int_0^{\zeta}\left(\frac{\mathbf{i}\hat{\sigma}_{1\partial z}(\xi,\tilde{q},\alpha)}{\hat{\sigma}_{0\partial z}(\xi,\tilde{q},\alpha)}-\frac{\mathbf{i}\sigma_{1\partial z}(\xi,\tilde{q},\alpha)}{\sigma_{0\partial z}(\xi,\tilde{q},\alpha)}\right)d\xi\right)-1\right|d\zeta \\
&\leq \underline{\phi}_n^{-1}\int_{-\bar{\zeta}_n}^{\bar{\zeta}_n}\left(\bar{\sigma}\max\left(|\zeta|^{-1},1\right)+\hat{\varepsilon}_n^{(0)}\right)\left|\exp\left(|\zeta|\,\bar{\phi}_n\hat{\varepsilon}_n^{(1)}\underline{\sigma}_n^{-1}\right)-1\right|d\zeta
\end{aligned}
$$

by Lemma 3. Next, since Assumption 13 implies that the argument of the exponential goes to zero for $|\zeta|\leq\zeta_n$ and since $\exp(\eta)-1=\eta+o(\eta)$, we have, for some $C,C'>0$ and for n sufficiently large,

$$
\begin{aligned}
R_1 &\leq C\underline{\phi}_n^{-1}\int_{-\bar{\zeta}_n}^{\bar{\zeta}_n}\left(\bar{\sigma}\max\left(|\zeta|^{-1},1\right)+\hat{\varepsilon}_n^{(0)}\right)|\zeta|\,\bar{\phi}_n\hat{\varepsilon}_n^{(1)}\underline{\sigma}_n^{-1}d\zeta \\
&= C\underline{\phi}_n^{-1}\int_{-\bar{\zeta}_n}^{\bar{\zeta}_n}\left(\bar{\sigma}\max(1,|\zeta|)+\hat{\varepsilon}_n^{(0)}|\zeta|\right)\bar{\phi}_n\hat{\varepsilon}_n^{(1)}\underline{\sigma}_n^{-1}d\zeta \\
&\leq 2C\underline{\phi}_n^{-1}\left(\bar{\sigma}\max\left(1,\bar{\zeta}_n\right)+\hat{\varepsilon}_n^{(0)}\bar{\zeta}_n\right)\bar{\zeta}_n\bar{\phi}_n\hat{\varepsilon}_n^{(1)}\underline{\sigma}_n^{-1} \\
&\leq C'\hat{\varepsilon}_n^{(1)}\bar{\zeta}_n^{2}\bar{\phi}_n\underline{\phi}_n^{-1}\underline{\sigma}_n^{-1}
\end{aligned}
$$

Similarly for R_2, we have

$$
\begin{aligned}
R_2 &= \left|\int_{-\bar{\zeta}_n}^{\bar{\zeta}_n}\left(\hat{\sigma}_{0z}(\zeta,q,\alpha)-\sigma_{0z}(\zeta,q,\alpha)\right)\exp\left(-\int_0^{\zeta}\frac{\mathbf{i}\sigma_{1z}(\xi,\tilde{q},\alpha)}{\sigma_{0z}(\xi,\tilde{q},\alpha)}d\xi\right)e^{-\mathbf{i}\zeta x^*}d\zeta\right| \\
&= \left|\int_{-\bar{\zeta}_n}^{\bar{\zeta}_n}\left(\hat{\sigma}_{0z}(\zeta,q,\alpha)-\sigma_{0z}(\zeta,q,\alpha)\right)\phi^{-1}(\zeta,\alpha)\,e^{-\mathbf{i}\zeta x^*}d\zeta\right| \\
&\leq \int_{-\bar{\zeta}_n}^{\bar{\zeta}_n}\left|\hat{\sigma}_{0z}(\zeta,q,\alpha)-\sigma_{0z}(\zeta,q,\alpha)\right|\left|\phi(\zeta,\alpha)\right|^{-1}d\zeta \\
&\leq \int_{-\bar{\zeta}_n}^{\bar{\zeta}_n}\hat{\varepsilon}_n^{(0)}\underline{\phi}_n^{-1}d\zeta \\
&= 2\bar{\zeta}_n\hat{\varepsilon}_n^{(0)}\underline{\phi}_n^{-1}.
\end{aligned}
$$

Finally, by Assumption 12, $\sigma_{0x^*}(\zeta,q,\alpha)=O\left(\zeta^{-2}\right)$ uniformly in q and α as $|\zeta|\to\infty$, thus implying that

$$
R_3=\left|\int_{|\zeta|\geq\bar{\zeta}_n}\sigma_{0x^*}(\zeta,q,\alpha)\,e^{-\mathbf{i}\zeta x^*}d\zeta\right|\leq\int_{|\zeta|\geq\bar{\zeta}_n}\left|\sigma_{0x^*}(\zeta,q,\alpha)\right|d\zeta=o(1).
$$

It follows that

$$
\sup_{x^*\in\mathbb{R}}\sup_{q\in\mathcal{Q}}\sup_{\alpha\in\mathcal{A}}\left|\hat{s}_{0x^*}(x^*,q,\alpha)-s_{0x^*}(x^*,q,\alpha)\right|\leq C'\hat{\varepsilon}_n^{(1)}\bar{\zeta}_n^{2}\bar{\phi}_n\underline{\phi}_n^{-1}\underline{\sigma}_n^{-1}+2\varepsilon_n^{0}\bar{\zeta}_n\underline{\phi}_n^{-1}+o(1).
$$

Substituting the orders of magnitude of $\hat{\varepsilon}_n^{(0)}$ and $\hat{\varepsilon}_n^{(1)}$ from Lemma 5 and using Assumption 13 establishes that the right-hand side is $o_p(1)$. ∎

Proof of Theorem 2. The proof consists in verifying the assumptions of Theorem 2.1 in Newey and McFadden (1994). Write

$$
\varepsilon_n=\sup_{x^*\in\mathcal{X}}\sup_{q\in\mathcal{Q}}\left|\hat{s}_{0x^*}(x^*,q,\hat{\alpha})-s_{0x^*}(x^*,q,\alpha^*)\right|\leq R_1+R_2 \tag{69}
$$

where R_1 and R_2 are bounded as follows:

$$\begin{aligned} R_1 &= \sup_{x^*\in\mathcal{X}}\sup_{q\in\mathcal{Q}}\left|\hat{s}_{0x^*}\left(x^*,q,\hat{\alpha}\right)-s_{0x^*}\left(x^*,q,\hat{\alpha}\right)\right| \\ &\le \sup_{x^*\in\mathcal{X}}\sup_{q\in\mathcal{Q}}\sup_{\alpha\in\mathcal{A}}\left|\hat{s}_{0x^*}\left(x^*,q,\alpha\right)-s_{0x^*}\left(x^*,q,\alpha\right)\right| = o_p\left(1\right) \end{aligned}$$

by Lemma 6 and

$$R_2 = \sup_{x^*\in\mathcal{X}}\sup_{q\in\mathcal{Q}}\left|s_{0x^*}\left(x^*,q,\hat{\alpha}\right)-s_{0x^*}\left(x^*,q,\alpha^*\right)\right| = o_p\left(1\right)$$

since (i) Assumptions 7 and 8 imply that $\hat{\alpha}\xrightarrow{p}\alpha^*$ and (ii) $s_{0x^*}\left(x^*,q,\alpha\right)$ is continuous in α uniformly for $x^*\in\mathcal{X}$ and $q\in\mathcal{Q}$, by Assumption 14. We now show that the functional

$$\hat{F}\left(Q_\tau\right)\equiv\int_{x^*\in\mathcal{X}}\left(\hat{s}_{0x^*}\left(x^*,Q_\tau\left(x^*\right),\hat{\alpha}\right)-\tau\right)^2dx^*$$

converges to the functional

$$F\left(Q_\tau\right)\equiv\int_{x^*\in\mathcal{X}}\left(s_{0x^*}\left(x^*,Q_\tau\left(x^*\right),\alpha^*\right)-\tau\right)^2dx^*$$

in the sense that $\sup_{Q_\tau\in\mathcal{L}}\left|\hat{F}\left(Q_\tau\right)-F\left(Q_\tau\right)\right| = o_p\left(1\right)$. Indeed,

$$\begin{aligned} &\sup_{Q_\tau\in\mathcal{L}}\left|\hat{F}\left(Q_\tau\right)-F\left(Q_\tau\right)\right| \\ =& \sup_{Q_\tau\in\mathcal{L}}\left|\int_{x^*\in\mathcal{X}}\left(\left(\hat{s}_{0x^*}\left(x^*,Q_\tau\left(x^*\right),\hat{\alpha}\right)-\tau\right)^2-\left(s_{0x^*}\left(x^*,Q_\tau\left(x^*\right),\alpha^*\right)-\tau\right)^2\right)dx^*\right| \\ \le& \sup_{Q_\tau\in\mathcal{L}}\left|\int_{x^*\in\mathcal{X}}\left(2\left(s_{0x^*}\left(x^*,Q_\tau\left(x^*\right),\alpha^*\right)-\tau\right)\varepsilon_n+\varepsilon_n^2\right)dx^*\right| \\ \le& \left|\int_{x^*\in\mathcal{X}}\left(2\varepsilon_n+\varepsilon_n^2\right)dx^*\right| \\ \le& B\varepsilon_n = o_p\left(1\right) \end{aligned}$$

for ε_n as in Equation (69) and for some finite B since $\left|s_{0x^*}\left(x^*,Q_\tau\left(x^*\right),\alpha^*\right)\right|\le 1$, $\tau\in[0,1]$ and the set $\mathcal{X}$ is a compact interval.

By Assumption 1, $s_{0x^*}\left(x^*,q,\alpha^*\right)$ is a continuous function of q for any $q\in\mathcal{Q}$ and $x^*\in\mathcal{X}$ and since $\mathcal{Q}$ and $\mathcal{X}$ are compact, the continuity is uniform in q and x^*. Hence for any $\eta>0$ there exists a $\delta>0$ such that $\sup_{x^*\in\mathcal{X}}\left|Q_\tau\left(x^*\right)-\tilde{Q}_\tau\left(x^*\right)\right|\le\delta$ implies $\sup_{x^*\in\mathcal{X}}\left|s_{0x^*}\left(x^*,Q_\tau\left(x^*\right),\alpha^*\right)-s_{0x^*}\left(x^*,\tilde{Q}_\tau\left(x^*\right),\alpha^*\right)\right|\le\eta$. We can then write for Q_τ and $\tilde{Q}_\tau$ such that $\sup_{x^*\in\mathcal{X}}\left|Q_\tau\left(x^*\right)-\tilde{Q}_\tau\left(x^*\right)\right|<\delta$,

$$\begin{aligned} \left|F\left(Q_\tau\right)-F\left(\tilde{Q}_\tau\right)\right| &= \int_{x^*\in\mathcal{X}}\left(\left(s_{0x^*}\left(x^*,Q_\tau\left(x^*\right)\right)-\tau\right)^2-\left(s_{0x^*}\left(x^*,\tilde{Q}_\tau\left(x^*\right)\right)-\tau\right)^2\right)dx^* \\ &\le \int_{x^*\in\mathcal{X}}\left(2\left(s_{0x^*}\left(x^*,Q_\tau\left(x^*\right)\right)-\tau\right)\eta+\eta^2\right)dx^* \\ &\le B\eta. \end{aligned}$$

Hence, the functional $F\left(Q_\tau\right)$ is a continuous mapping from $\mathcal{L}$ to $\mathbb{R}$, when $\mathcal{L}$ is endowed with the sup norm. This norm also makes the closed set $\mathcal{L}$ compact under Assumption 15, by the Arzela-Ascoli Theorem. Finally, $F\left(Q_\tau\right)$ is uniquely minimized at the true Q_τ by Theorem 1. By Theorem 2.1 in Newey and

McFadden (1994), it follows that any $\hat{Q}_\tau$ satisfying $\hat{Q}_\tau \in \arg\min_{Q_\tau \in \mathcal{L}} \hat{F}(Q_\tau)$ is consistent in the sense that $\sup_{x^* \in \mathcal{X}} \left|\hat{Q}_\tau(x^*) - Q_\tau(x^*)\right| \xrightarrow{p} 0$. Under the additional Assumption 16, the same conclusion holds for $\hat{Q}_\tau \in \arg\min_{Q_\tau \in \mathcal{L}_n} \hat{F}(Q_\tau)$, by Lemma A1 in Newey and Powell (2003). ■