

On Deconvolution as a First Stage Nonparametric Estimator

Yingyao Hu

Department of Economics,
University of Texas, Austin
E-mail: hu@eco.utexas.edu

Geert Ridder

Department of Economics
University of Southern California, Los Angeles
E-mail: ridder@usc.edu

August 10, 2005

Abstract

We reconsider Taupin's (2001) Integrated Nonlinear Regression (INLR) estimator for a nonlinear regression with a mismeasured covariate. We find that if we restrict the distribution of the measurement error to the class of range-restricted distributions, then weak smoothness assumptions suffice to ensure \sqrt{n} consistency of the estimator. The restriction to such distributions is innocuous, because it does not affect the fit to the data. Our results show that deconvolution can be used in a nonparametric first step without imposing restrictive smoothness assumptions on the parametric model.

1 Introduction

To estimate the parameters of a model that is nonlinear in a mismeasured covariate consistently it is necessary to identify and estimate (at an appropriate rate) the density of the latent true value given the reported value and the error-free covariates. There are a number of different identifying assumptions that can be used for this purpose. A common feature is that estimation of the conditional density involves deconvolution. Examples are Li and Vuong (1998), Li (2002), and Schennach(2004) who assume repeated measurements, Taupin (2001) who assumes that the distribution of the measurement error is known, and Hu and Ridder (2003) who consider the case that there is marginal information on the distribution of the latent true value. In all cases the first stage estimate of the conditional density is used in a second stage to integrate the latent value from the parametric model. The parameters of the model are estimated in the second stage by Maximum Likelihood (ML) or the Generalized Method of Moments (GMM).

In an influential paper Taupin (2001) has argued that if the distribution of the measurement error is normal and if the first-stage density is estimated by deconvolution, then a nonlinear regression model¹ has to be restricted to a polynomial or an exponential function, both sufficiently smooth to keep the variance of the second-stage nonlinear regression estimator finite. Her result suggests that deconvolution can only be used in a first-stage density estimator, if the parametric model satisfies restrictive smoothness assumptions.

In this note we reconsider Taupin's Integrated Nonlinear Regression (INLR) estimator. We argue that most economic variables have a restricted range. They are non-negative or they are bounded. This implies that measurement errors have a similar restricted range. If the distribution of the measurement error has a restricted range, then the convergence speed of the first-stage non-parametric estimator is sufficiently fast that minimal smoothness assumptions on the parametric model suffice to ensure \sqrt{n} consistent esti-

¹She only considers nonlinear regression, but analogous restrictions must be imposed on any parametric model.

mators in the second stage. Even if we believe that the measurement error distribution is close to normal, then, because the normal distribution can be approximated to any desired level of approximation by a two-sided truncated normal or a mixture of one-sided truncated normals, the integrated regression function² under the normal or truncated normal distributions are arbitrarily close. Hence, we can fit the data equally well with either model, so that we do not sacrifice fit, if we restrict the measurement error distribution to the class of range restricted distributions.

2 The integrated nonlinear regression estimator

We use the same setup as Taupin (2001). We have a random sample $y_i, x_i, i = 1, \dots, n$. The covariate x is measured with error

$$x = x^* + \varepsilon.$$

We assume that

$$\varepsilon \perp x^*$$

i.e. the measurement error is classical. To allow for nonclassical measurement error a validation sample is needed (see e.g. Chen, Hong, and Tamer (2005)).

The parametric model specifies the conditional mean function

$$E(y|x^*) = h(x^*, \theta_0)$$

that depends on the unobserved latent true value x^* . To concentrate on essentials we assume that there are no other covariates. By the Law of Iterated Expectations the conditional mean function given the observed x is

$$E(y|x) = \int_{x^*} h(x^*, \theta_0) g(x^*|x) dx^*$$

²In general this applies to the distributions of the observed variables under the two models.

with \mathcal{X}^* the support of x^* . The corresponding moment function is

$$m(y, x, \theta) = w(x) \left(y - \int_{\mathcal{X}^*} h(x^*, \theta_0) g(x^*|x) dx^* \right)$$

with w a (vector of) weighting function(s) with dimension at least as large as the number of parameters in θ . In this note we only consider the just identified case where the dimension of w and the number of parameters in θ are equal.

The final step is to estimate the conditional density of x^* given x . We have

$$g(x^*|x) = \frac{g(x|x^*)g_{x^*}(x^*)}{g_x(x)} = \frac{g_\varepsilon(x - x^*)}{g_x(x)} g_{x^*}(x^*)$$

As we assume that the density of the measurement error ε is known, we need to estimate the marginal densities of x and of x^* nonparametrically. Because x is observed, we can use a standard nonparametric estimator, e.g. a kernel estimator. In this note we assume that the density of x is known, i.e. we ignore sampling variation in the density estimate. Although the estimation of the density of x affects the sampling variance of the second stage estimator, it is not the reason that that variance can become infinite.

To estimate the density of x^* we use the fact that the measurement error model and the independence of x^* and ε imply that

$$\phi_{x^*}(t) = \frac{\phi_x(t)}{\phi_\varepsilon(t)}$$

with $\phi_x(t) = E(e^{itx})$ the characteristic function of x etc. If $\phi_{x^*}(t)$ is absolutely integrable then by the Fourier inversion theorem

$$g_{x^*}(x^*) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx^*} \phi_{x^*}(t) dt$$

The corresponding density estimator is

$$\hat{g}_{x^*}(x^*) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx^*} \frac{\hat{\phi}_x(t)}{\hat{\phi}_\varepsilon(t)} K_n^*(t) dt.$$

with

$$\hat{\phi}_x(t) = \frac{1}{n} \sum_{i=1}^n e^{itx_i} = \int_{\mathcal{X}} e^{itx} dF_n(x)$$

the empirical characteristic function (ecf) of x (F_n is the empirical cdf of the sample x_1, \dots, x_n). The Fourier inversion theorem does not hold for the ecf and for that reason the integrand is multiplied by the kernel $K_n^*(t) = K^*(\frac{t}{T_n})$ which ensures that the integral exists. The function K^* is the Fourier transform of the function K and $\frac{1}{T_n}$ is the bandwidth. By the convolution theorem multiplication of the ecf by $K^*(\frac{t}{T_n})$ smoothes the ecf. The kernel K satisfies

- (i) K is an even function and K^2 is integrable.
- (ii) Its Fourier transform is such that $K^*(t) = 1$ for $|t| \leq 1$.
- (iii) $|K^*(t)| \leq I_{[-2,2]}(t)$ all t .
- (iv) $\int K(z)dz = 1$, $\int |K(z)|dz < \infty$, $\int z^j K(z)dz = 0$ for $j = 1, 2, \dots, q-1$, and $|\int z^q K(z)dz| < \infty$, i.e. K is a higher order kernel of order q .

Substitution of this estimator in the moment function gives the INLR estimator as the solution to

$$m_n(\hat{\theta}) \equiv \frac{1}{n} \sum_{i=1}^n w(x_i) \left(y_i - \int_{\mathcal{X}^*} h(x^*, \hat{\theta}) \frac{g_\varepsilon(x_i - x^*)}{g_x(x_i)} \hat{g}_{x^*}(x^*) dx^* \right) = 0 \quad (1)$$

3 The characteristic function of range-restricted distributions

The asymptotic bias of the first-stage deconvolution estimator is determined by the behavior of the characteristic function of the measurement error if its argument is large. By the Riemann-Lebesgue theorem (Feller, 1971) the (absolute value of the) characteristic function $\phi_v(t)$ of a distribution is bounded

above by $C|t|^{-k}$ if the distribution has a density with a k -th derivative that is absolutely integrable³. The upper bound can converge faster for certain distributions. Fan (1991) introduced two classes of characteristic functions. The supersmooth characteristic functions are bounded from below and above by functions that decrease exponentially if t is large. The ordinary smooth characteristic functions are bounded from below and above by functions that decrease geometrically if t is large. The normal distribution has a supersmooth characteristic function.

In general, deconvolution estimators of densities converge at a logarithmic rate if the measurement error has a distribution with a supersmooth characteristic function. Hence, it is important to know how ‘prevalent’ distributions with such a characteristic function are. The next theorem establishes that all range restricted distributions have ordinary smooth characteristic functions.

Theorem 1 *If a distribution of a random variable v has a density (with respect to the Lebesgue measure) f_v that is positive on $[L, U]$ with either L or U finite and has $k + 2$ absolutely integrable derivatives $f_v^{(j)}$ with*

$$(i) \quad f_v^{(j)}(U) = f_v^{(j)}(L) = 0 \text{ for } j = 0, \dots, k - 1.$$

$$(ii) \quad |f_v^{(k)}(U)| \neq |f_v^{(k)}(L)|$$

then its characteristic function is ordinary smooth. In particular, there is a $t_0 > 0$ such that for all $|t| \geq t_0$

$$C_0 \left| |f_v^{(k)}(U)| - |f_v^{(k)}(L)| \right| t^{-(k+1)} \leq |\phi_v(t)| \leq C_1 (|f_v^{(k)}(U)| + |f_v^{(k)}(L)|) t^{-(k+1)} \quad (2)$$

Proof In appendix. \square

³Here and in the sequel C denotes a generic constant

Remark 1 For the normal distribution the theorem holds with $k = 0$ if the truncation points are not symmetric with respect to the mean. In that case the characteristic function also does not have any real zeros, which is convenient because in the deconvolution estimator of the density we divide by the characteristic function.

Remark 2 If the measurement error is truncated (standard) normal with density

$$\phi(\varepsilon; L, U) = I(L \leq \varepsilon \leq U) \frac{\phi(\varepsilon)}{\Phi(U) - \Phi(L)}$$

then

$$\lim_{L \rightarrow -\infty, U \rightarrow \infty} \sup_{-\infty < \varepsilon < \infty} |\phi(\varepsilon; L, U) - \phi(\varepsilon)| = 0$$

so that, if we define

$$E(y|x; L, U) = \int_{\mathcal{X}^*} h(x^*, \theta_0) \frac{\phi(x - x^*; L, U)}{g_x(x)} g_{x^*}(x^*) dx^*$$

for all $x = x^* + \varepsilon$ with $x^* \in \mathcal{X}^*$ and $L \leq \varepsilon \leq U$ and set it equal to 0 for all other values of x , we have

$$|E(y|x) - E(y|x; L, U)| \leq \int_{\mathcal{X}^*} |h(x^*, \theta_0)| \frac{|\phi(x - x^*) - \phi(x - x^*; L, U)|}{g_x(x)} g_{x^*}(x^*) dx^*$$

Hence if g_x is bounded from 0 on its support, then

$$\lim_{L \rightarrow -\infty, U \rightarrow \infty} \sup_{x \in \mathcal{X}} |E(y|x) - E(y|x; L, U)| = 0$$

Because the nonlinear regression function can be uniformly well approximated, the fit of a model based on truncated normal measurement error will be as good as one based on unrestricted normal measurement error.

Remark 3 If we consider a mixture of a random variable v_1 with a support that has an upper bound and a random variable v_2 with a support that has a lower bound, then

$$\phi_v(t) = p\phi_{v_1}(t) + (1 - p)\phi_{v_2}(t)$$

Hence v has unbounded support and is ordinary smooth. It is however hard to see which economic variables can be seen as such mixtures.

In the sequel we refer to a random variable with a distribution as in Theorem 1 as *range restricted of order k* . Note that in general both the lower and the upper bound converge to 0 at a lower rate for range-restricted distributions. That is also true for distributions that have an ordinary smooth characteristic function, if the support is unrestricted. For instance, if the density of a distribution with unbounded support has l absolutely integrable derivatives, then its characteristic function decreases as $|t|^{-l}$. If we truncate the support the rate can be as slow as $|t|^{-1}$.

4 The rate of uniform convergence of the deconvolution density estimator

The first application of theorem 1 is to the rate of convergence of the deconvolution density estimator of section 2. In particular, we show that its uniform rate of convergence can be at least $n^{-\frac{1}{4}}$, a rate that is required to ensure that the second stage estimator is \sqrt{n} consistent (Newey, 1994).

Theorem 2 *Let ϕ_ε be absolutely integrable and $|\phi_\varepsilon(t)| > 0$ for all $t \in \mathfrak{R}$. The distribution of ε is range restricted of order k . Choose $T_n = O\left(\left(\frac{n}{\log n}\right)^\gamma\right)$ with $0 < \gamma < \frac{1}{2}$ and let the kernel K be of order q . The distribution of x^* has a density that is q times differentiable and the q -th derivative is bounded on \mathcal{X}^* . Then a.s.*

$$\sup_{x^* \in \mathcal{X}^*} |\hat{g}_{x^*}(x^*) - g_{x^*}(x^*)| = O\left(\left(\frac{\log n}{n}\right)^{\frac{1}{2} - (k+3)\gamma - \eta}\right) + O\left(\left(\frac{\log n}{n}\right)^{q\gamma}\right). \quad (3)$$

for all $\eta > 0$.

Proof In appendix. \square

The best rate of convergence is $n^{-\frac{q}{2(k+3+q)} + \eta}$ and if $k = 0$ the rate is certainly faster than $n^{-\frac{1}{4}}$ if $q \geq 4$.

5 The asymptotic properties of the INLR estimator

5.1 Consistency

The next theorem provides a set of conditions that ensure that the INLR estimator is weakly consistent.

Theorem 3 *If*

- (i) $E_x \left[w(x) \int_{\mathcal{X}^*} (h(x^*, \theta_0) - h(x^*, \theta)) g(x^*|x) dx^* \right] = 0$ if and only if $\theta = \theta_0$.
- (ii) *The regression function h is bounded on $\mathcal{X}^* \times \Theta$.*
- (iii) *The density of x $g_x(x)$ is bounded from 0 on \mathcal{X} .*
- (iv) *$w(x)$ is bounded on \mathcal{X} .*
- (v) *The distribution of the measurement error is range-restricted of order k .*
- (vi) *ϕ_ε is absolutely integrable and $|\phi_\varepsilon(t)| > 0$ for all t .*
- (vii) $T_n = O \left(\left(\frac{n}{\log n} \right)^\gamma \right)$ with $0 < \gamma < \frac{1}{2(k+3)}$.

then the INLR estimator is weakly consistent.

Proof In appendix. \square

5.2 The asymptotic distribution

The next theorem gives the asymptotically linear representation of the INLR estimator and establishes that the estimator is asymptotically normally distributed. The proof is in the appendix.

Theorem 4 *If*

- (i) *Assumptions (ii)-(vii) of Theorem 2 hold.*
- (ii) $\frac{\partial h(x^*, \theta)}{\partial \theta'}$ *is bounded on* \mathcal{X}^* *and continuous in* θ *for (almost all)* $x^* \in \mathcal{X}^*$.
The derivative with respect to x^* , $h'(x^*, \theta_0)$, *is bounded on* \mathcal{X}^* .
- (iii) *If the distribution of the measurement error is range-restricted of order* k , *then* $h(x^*, \theta_0)$ *and* $g_\varepsilon(\varepsilon)$ *have at least* $k + 3$ *absolutely integrable derivatives.*
- (iv) $\text{rank } G = K$ *with* K *the dimension of* θ *and* $G = E \left[w(x) \int_{\mathcal{X}^*} \frac{\partial h(x^*, \theta_0)}{\partial \theta'} g(x^* | x) dx^* \right]$.
- (v) $\sqrt{n} T_n^{-q} \rightarrow 0$

then if we define

$$c^*(x, t, \theta) = \frac{1}{2\pi} \int_{\mathcal{X}^*} e^{-itx^*} w(x) h(x^*, \theta) \frac{g_\varepsilon(x - x^*)}{g_x(x)} dx^*$$

the INLR estimator is asymptotically linear with

$$\begin{aligned} \sqrt{n}(\hat{\theta} - \theta_0) = & G^{-1} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n w(x_i) \left(y_i - \int_{\mathcal{X}^*} h(x^*, \theta_0) g(x^* | x_i) dx^* \right) - \right. \\ & \left. - \frac{1}{\sqrt{n}} \sum_{i=1}^n \int \frac{E[c^*(x, t, \theta_0)]}{\phi_\varepsilon(t)} (e^{itx_i} - \phi_x(t)) K_n^*(t) dt \right) \end{aligned}$$

with

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \int \frac{E[c^*(x, t, \theta_0)]}{\phi_\varepsilon(t)} (e^{itx_i} - \phi_x(t)) K_n^*(t) dt$$

the correction term that accounts for the use of the deconvolution estimator in the first stage. The asymptotic variance of its limiting normal distribution is finite.

Proof In appendix. \square

As noted by Taupin (2001) the correction term in the asymptotic linear representation is asymptotically normally distributed, if its asymptotic variance is finite. If the distribution of the measurement error is range-restricted of order k this requires mild smoothness assumptions on the regression function. In the leading case that $k = 0$ the existence of three (absolutely integrable) derivatives suffices. This should be compared with the requirement that the regression function has to be polynomial or exponential as in Taupin. The trade-off is between a mild assumption on the measurement error distribution (that does not affect its ability to fit the data) and extreme smoothness assumptions on the parametric model.

Assumptions (v) of Theorem 4 and (vii) of Theorem 3 imply that $T_n = O\left(\left(\frac{n}{\log n}\right)^\gamma\right)$ with $\frac{1}{2q} < \gamma < \frac{1}{2(k+3)}$. For $k = 0$ this implies that $q \geq 4$ which is consistent with a convergence rate of at least $n^{-\frac{1}{4}}$ for the first-stage nonparametric deconvolution estimator.

6 Conclusion

We reconsider Taupin's (2001) Integrated Nonlinear Regression (INLR) estimator. We conclude that if we are prepared to restrict the distribution of the measurement error to the class of range restricted distributions, then weak smoothness assumptions suffice to ensure \sqrt{n} consistency of the estimator. The restriction to such distributions of the measurement error does not affect the fit to the data. For that reason we think that the bad reputation of deconvolution based density estimators as first-stage estimators in a semi-parametric procedure is undeserved.

The result of this note also applies to other semi-parametric estimators that have a first-stage nonparametric deconvolution estimator, e.g. Hu and Ridder's (2003) estimator for nonlinear parametric models with a mismeasured covariate.

7 Appendix

Proof of Theorem 1

We give the proof for $k = 0$. We use integration by parts twice to obtain

$$\begin{aligned}
 \phi_v(t) &= \int_L^U e^{itx} f_v(x) dx \\
 &= \frac{1}{it} \left[f_v(U) e^{itU} - f_v(L) e^{itL} - \int_L^U e^{itx} f'_v(x) dx \right] \\
 &= \frac{1}{it} \left[f_v(U) e^{itU} - f_v(L) e^{itL} - \frac{1}{it} \left(f'_v(U) e^{itU} - f'_v(L) e^{itL} - \int_L^U e^{itx} f''_v(x) dx \right) \right]
 \end{aligned}$$

Hence

$$|\phi_v(t)| \geq \frac{1}{|t|} \left| \left(f_v(U) - \frac{1}{it} f'_v(U) \right) e^{itU} - \left(f_v(L) - \frac{1}{it} f'_v(L) \right) e^{itL} - \frac{1}{it} \int_L^U e^{itx} f''_v(x) dx \right| \quad (4)$$

The first term in absolute value on the right-hand side is bounded from below

by

$$\begin{aligned}
 &\left| \left(f_v(U) - \frac{1}{it} f'_v(U) \right) e^{itU} - \left(f_v(L) - \frac{1}{it} f'_v(L) \right) e^{itL} \right| \geq \quad (5) \\
 &\left| \left(f_v(U) - \frac{1}{it} f'_v(U) \right) e^{itU} \right| - \left| \left(f_v(L) - \frac{1}{it} f'_v(L) \right) e^{itL} \right| = \\
 &= \left| \sqrt{f_v(U)^2 + \left(\frac{1}{t} f'_v(U) \right)^2} - \sqrt{f_v(L)^2 + \left(\frac{1}{t} f'_v(L) \right)^2} \right|
 \end{aligned}$$

The second term in absolute value on the right-hand side of (4) is bounded from above by

$$\left| \frac{1}{it} \int_L^U e^{itx} f''_v(x) dx \right| \leq \frac{1}{|t|} \int_L^U e^{itx} |f''_v(x)| dx$$

and this upper bound converges to 0 if $|t| \rightarrow \infty$.

If $|t| \rightarrow \infty$ the lower bound in (5) converges to $|f_v(U) - f_v(L)| > 0$, so that there is t_0 such that for $|t| \geq t_0$

$$\left| \left(f_v(U) - \frac{1}{it} f'_v(U) \right) e^{itU} - \left(f_v(L) - \frac{1}{it} f'_v(L) \right) e^{itL} - \frac{1}{it} \int_L^U e^{itx} f''_v(x) dx \right| \geq C |f_v(U) - f_v(L)|$$

and hence for $|t| \geq t_0$

$$|\phi_v(t)| \geq \frac{C_0|f_v(U) - f_v(L)|}{|t|}$$

For the upper bound we have

$$\begin{aligned} |\phi_v(t)| &\leq \frac{1}{|t|} \left(\left| \left(f_v(U) - \frac{1}{it} f'_v(U) \right) e^{itU} \right| + \left| \left(f_v(L) - \frac{1}{it} f'_v(L) \right) e^{itL} \right| + \left| \frac{1}{it} \int_L^U e^{itx} f''_v(x) dx \right| \right) \leq \\ &\leq \frac{1}{|t|} \left(f_v(U) + \frac{1}{|t|} |f'_v(U)| + f_v(L) + \frac{1}{|t|} |f'_v(L)| + \frac{1}{|t|} \int_L^U |f''_v(x)| dx \right) \leq \\ &\leq \frac{C_1(f_v(U) + f_v(L))}{|t|} \end{aligned}$$

if $|t| \geq t_0$ where if necessary we increase the t_0 we used earlier.

For the case that $k \geq 1$ the same method of proof applies after $k + 2$ integrations by parts. \square

Proof of Theorem 2

We first establish the rate of uniform convergence of the empirical characteristic function. This lemma corrects a result in Horowitz and Markatou (1996).

Lemma 1 *Let $\hat{\phi}_v(t) = \int_{-\infty}^{\infty} e^{itv} dF_n(v)$ be the empirical characteristic function of a random sample v_1, \dots, v_n from a distribution with cdf F and with $E(|v|) < \infty$. For $0 < \gamma < \frac{1}{2}$, let $T_n = o\left(\left(\frac{n}{\log n}\right)^\gamma\right)$. Then*

$$\sup_{|t| \leq T_n} \left| \hat{\phi}_v(t) - \phi_v(t) \right| = o(\alpha_n) \quad \text{a.s.} \quad (6)$$

with $\alpha_n = o(1)$ and $\frac{\left(\frac{\log n}{n}\right)^{\frac{1}{2}-\gamma}}{\alpha_n} = O(1)$, i.e the rate of convergence is at most $\left(\frac{\log n}{n}\right)^{\frac{1}{2}-\gamma}$.

Proof Consider the parametric class of functions $\mathcal{G}_n = \{e^{itx} \mid |t| \leq T_n\}$. The first step, is to find the L_1 covering number of \mathcal{G}_n . Because $e^{itx} = \cos(tx) +$

$i \sin(tx)$, we need covers of $\mathcal{G}_{1n} = \{\cos(tx) \mid |t| \leq T_n\}$ and $\{\mathcal{F}_{2n} = \sin(tx) \mid |t| \leq T_n\}$. Because $|\cos(t_2x) - \cos(t_1x)| \leq |x||t_2 - t_1|$ and $E(|x|) < \infty$, an $\frac{\varepsilon}{2}E(|x|)$ cover (with respect to the L_1 norm) of \mathcal{G}_{1n} is obtained from an $\frac{\varepsilon}{2}$ cover of $\{t \mid |t| \leq T_n\}$ by choosing $t_k, k = 1, \dots, K$ arbitrarily from the distinct covering sets, where K is the smallest integer larger than $\frac{2T_n}{\varepsilon}$. Because $|\sin(t_2x) - \sin(t_1x)| \leq |x||t_2 - t_1|$, the functions $\sin(t_kx), k = 1, \dots, K$ are an $\frac{\varepsilon}{2}E(|x|)$ cover of \mathcal{F}_{2n} . Hence $\cos(t_kx) + i \sin(t_kx), k = 1, \dots, K$ is an $\varepsilon E(|x|)$ cover of \mathcal{G}_n , and we conclude that

$$\mathcal{N}_1(\varepsilon, P, \mathcal{G}_n) \leq A \frac{T_n}{\varepsilon} \quad (7)$$

with P an arbitrary probability measure such that $E(|x|) < \infty$ and $A > 0$, a constant that does not depend on n . The next step is to apply the argument that leads to Theorem 2.37 in Pollard (1984). The theorem cannot be used directly, because the condition $\mathcal{N}_1(\varepsilon, P, \mathcal{G}_n) \leq A\varepsilon^{-W}$ is not met. In Pollard's proof we set $\delta_n = 1$ for all n , and $\varepsilon_n = \varepsilon\alpha_n$. Equations (30) and (31) in Pollard (1984), p. 31 are valid for $\mathcal{N}_1(\varepsilon, P, \mathcal{G}_n)$ defined above. Hence we have as in Pollard's proof using his (31)

$$\Pr \left(\sup_{|t| \leq T_n} |\hat{\phi}(t) - \phi(t)| > 2\varepsilon_n \right) \leq 2A \left(\frac{\varepsilon_n}{T_n} \right)^{-1} \exp \left(-\frac{1}{128} n \varepsilon_n^2 \right) + \Pr \left(\sup_{|t| \leq T_n} \hat{\phi}(2t) > 64 \right). \quad (8)$$

The second term on the right-hand side is obviously 0. The first term on the right-hand side is bounded by

$$2A\varepsilon^{-1} \exp \left(\log \left(\frac{T_n}{\alpha_n} \right) - \frac{1}{128} n \varepsilon^2 \alpha_n^2 \right). \quad (9)$$

The restrictions on α_n and T_n imply that $\frac{T_n}{\alpha_n} = o \left(\sqrt{\frac{n}{\log n}} \right)$, and hence $\log \left(\frac{T_n}{\alpha_n} \right) - \frac{1}{2} \log n \rightarrow -\infty$. The same restrictions imply that $\frac{n\alpha_n^2}{\log n} \rightarrow \infty$. The result now follows from the Borel-Cantelli lemma. \square

Proof of Theorem 2 We have

$$\begin{aligned} \sup_{x^* \in \mathcal{X}^*} |\hat{g}_{x^*}(x^*) - g_{x^*}(x^*)| &\leq \sup_{x^* \in \mathcal{X}^*} \left| \frac{1}{2\pi} \int e^{-itx^*} \left(\frac{\hat{\phi}_x(t) - \phi_x(t)}{\phi_\varepsilon(t)} \right) K_n^*(t) dt \right| + \\ &+ \sup_{x^* \in \mathcal{X}^*} \left| \frac{1}{2\pi} \int e^{-itx^*} \phi_{x^*}(t) [1 - K_n^*(t)] dt \right| \end{aligned} \quad (10)$$

We consider the first term on the right-hand side, the variance term, that is bounded by

$$\sup_{|t| \leq T_n} \left| \hat{\phi}_x(t) - \phi_x(t) \right| \frac{1}{2\pi} \int_{-T_n}^{T_n} \frac{1}{|\phi_\varepsilon(t)|} dt \quad (11)$$

Hence (11) is a.s. bounded by

$$O \left(\frac{T_n}{|\phi_\varepsilon(T_n)|} \left(\frac{\log n}{n} \right)^{\frac{1}{2} - \gamma - \eta} \right) = O \left(T_n^{k+2} \left(\frac{\log n}{n} \right)^{\frac{1}{2} - \gamma - \eta} \right) = O \left(\left(\frac{\log n}{n} \right)^{\frac{1}{2} - (k+3)\gamma - \eta} \right) \quad (12)$$

where $T_n = O \left(\left(\frac{n}{\log n} \right)^\gamma \right)$ and $0 < \gamma < \frac{1}{2}$.

Next we consider the second term on the right-hand side of (10) which is the bias term. Because K is a kernel of order q , we have by a Taylor series expansion of the density of x^*

$$\begin{aligned} \frac{1}{2\pi} \int e^{-itx^*} \phi_{x^*}(t) [1 - K_n^*(t)] dt &= g_{x^*}(x^*) - \int K(z) g_{x^*} \left(x^* - \frac{z}{T_n} \right) dz \\ &= T_n^{-q} \left(g_\varepsilon^{(q)}(\tilde{x}^*) \int z^q K(z) dz \right), \end{aligned} \quad (13)$$

where \tilde{x}^* is between x^* and $x^* - \frac{z}{T_n}$. Therefore, the bias term is $O(T_n^{-q})$. The results follow. \square

Proof of Theorem 3

Sufficient for weak consistency of the estimator is that

$$m_n(\theta) = \frac{1}{n} \sum_{i=1}^n w(x_i) \left(y_i - \int_{\mathcal{X}^*} h(x^*, \theta) \frac{g_\varepsilon(x_i - x^*)}{g_x(x_i)} \hat{g}_{x^*}(x^*) dx^* \right) \xrightarrow{p}$$

$$\xrightarrow{p} \mathbb{E}_x \left[w(x) \int_{\mathcal{X}^*} (h(x^*, \theta_0) - h(x^*, \theta)) g(x^*|x) dx^* \right] \equiv m(\theta, \theta_0)$$

uniformly for $\theta \in \Theta$. We have

$$\begin{aligned} m_n(\theta) - m(\theta, \theta_0) &= \frac{1}{n} \sum_{i=1}^n \left(w(x_i) y_i - \mathbb{E}_x \left[w(x) \int_{\mathcal{X}^*} h(x^*, \theta_0) g(x^*|x) dx^* \right] \right) - \\ &\quad - \frac{1}{n} \sum_{i=1}^n w(x_i) \int_{\mathcal{X}^*} h(x^*, \theta) \frac{g_\varepsilon(x_i - x^*)}{g_x(x_i)} (\hat{g}_{x^*}(x^*) - g_{x^*}(x^*)) dx^* - \\ &\quad - \left(\frac{1}{n} \sum_{i=1}^n w(x_i) \int_{\mathcal{X}^*} h(x^*, \theta) g(x^*|x_i) dx^* - \mathbb{E}_x \left[w(x) \int_{\mathcal{X}^*} h(x^*, \theta) g(x^*|x) dx^* \right] \right) \equiv A_1 + A_2 + A_3 \end{aligned}$$

The term A_2 involves the deconvolution estimator of the density of x^* . Obviously A_1 converges to 0 in probability. For A_3 we have by the uniform weak law of large numbers that it converges to 0 in probability uniformly for $\theta \in \Theta$, if

$$\mathbb{E}_x \left[\sup_{\theta \in \Theta} \left| w(x) \int_{\mathcal{X}^*} h(x^*, \theta) g(x^*|x) dx^* \right| \right] < \infty$$

which holds if w is bounded on \mathcal{X} and $h(x^*, \theta)$ is bounded on $\mathcal{X}^* \times \Theta$.

We now consider A_2

$$\begin{aligned} \sup_{\theta \in \Theta} |A_2| &= \sup_{\theta \in \Theta} \left| \frac{1}{n} \sum_{i=1}^n w(x_i) \int_{\mathcal{X}^*} h(x^*, \theta) \frac{g_\varepsilon(x_i - x^*)}{g_x(x_i)} (\hat{g}_{x^*}(x^*) - g_{x^*}(x^*)) dx^* \right| \\ &\leq \sup_{x^* \in \mathcal{X}^*} |\hat{g}_{x^*}(x^*) - g_{x^*}(x^*)| \sup_{\theta \in \Theta} \left| \frac{1}{n} \sum_{i=1}^n w(x_i) \int_{\mathcal{X}^*} h(x^*, \theta) \frac{g_\varepsilon(x_i - x^*)}{g_x(x_i)} dx^* \right| \end{aligned}$$

By Theorem 2 $\sup_{x^* \in \mathcal{X}^*} |\hat{g}_{x^*}(x^*) - g_{x^*}(x^*)|$ is $o_p(1)$. It is easy to show that the second term is finite. Therefore, we have

$$\sup_{\theta \in \Theta} |A_2| \xrightarrow{p} 0$$

□.

Proof of Theorem 4

Expanding (1) around θ_0 we have for some $\bar{\theta} = \lambda\hat{\theta} + (1-\lambda)\theta_0$, $0 \leq \lambda \leq 1$

$$\begin{aligned} & \frac{1}{\sqrt{n}} \sum_{i=1}^n w(x_i) \left(y_i - \int_{\mathcal{X}^*} h(x^*, \theta_0) \frac{g_\varepsilon(x_i - x^*)}{g_x(x_i)} \hat{g}_{x^*}(x^*) dx^* \right) - \\ & - \frac{1}{n} \sum_{i=1}^n w(x_i) \int_{\mathcal{X}^*} \frac{\partial h(x^*, \bar{\theta})}{\partial \theta'} \frac{g_\varepsilon(x_i - x^*)}{g_x(x_i)} \hat{g}_{x^*}(x^*) dx^* \sqrt{n}(\hat{\theta} - \theta_0) = 0 \end{aligned}$$

Define

$$B_1 = \frac{1}{\sqrt{n}} \sum_{i=1}^n w(x_i) \left(y_i - \int_{\mathcal{X}^*} h(x^*, \theta_0) \frac{g_\varepsilon(x_i - x^*)}{g_x(x_i)} \hat{g}_{x^*}(x^*) dx^* \right)$$

and

$$B_2 = \frac{1}{n} \sum_{i=1}^n w(x_i) \int_{\mathcal{X}^*} \frac{\partial h(x^*, \bar{\theta})}{\partial \theta'} \frac{g_\varepsilon(x_i - x^*)}{g_x(x_i)} \hat{g}_{x^*}(x^*) dx^*$$

We consider first B_2 and we show that

$$B_2 \xrightarrow{p} \mathbb{E} \left[w(x) \int_{\mathcal{X}^*} \frac{\partial h(x^*, \theta_0)}{\partial \theta'} g(x^*|x) dx^* \right]$$

We have

$$\begin{aligned} & B_2 - \mathbb{E} \left[w(x) \int_{\mathcal{X}^*} \frac{\partial h(x^*, \theta_0)}{\partial \theta'} g(x^*|x) dx^* \right] = \\ & = \frac{1}{n} \sum_{i=1}^n w(x_i) \int_{\mathcal{X}^*} \left(\frac{\partial h(x^*, \bar{\theta})}{\partial \theta'} - \frac{\partial h(x^*, \theta_0)}{\partial \theta'} \right) \frac{g_\varepsilon(x_i - x^*)}{g_x(x_i)} \hat{g}_{x^*}(x^*) dx^* + \\ & + \frac{1}{n} \sum_{i=1}^n w(x_i) \int_{\mathcal{X}^*} \frac{\partial h(x^*, \theta_0)}{\partial \theta'} \frac{g_\varepsilon(x_i - x^*)}{g_x(x_i)} (\hat{g}_{x^*}(x^*) - g_{x^*}(x^*)) dx^* + \\ & + \frac{1}{n} \sum_{i=1}^n w(x_i) \int_{\mathcal{X}^*} \frac{\partial h(x^*, \theta_0)}{\partial \theta'} g(x^*|x_i) dx^* - \mathbb{E} \left[w(x) \int_{\mathcal{X}^*} \frac{\partial h(x^*, \theta_0)}{\partial \theta'} g(x^*|x) dx^* \right] \equiv C_1 + C_2 + C_3 \end{aligned}$$

For C_1 we have, because assumption (ii) and dominated convergence imply that $\int_{\mathcal{X}^*} \frac{\partial h(x^*, \theta)}{\partial \theta'} \frac{g_\varepsilon(x_i - x^*)}{g_x(x_i)} \hat{g}_{x^*}(x^*) dx^*$ is continuous in θ , that for all $\eta > 0$ there is $\delta > 0$ such that

$$|\bar{\theta} - \theta_0| \leq \delta \Rightarrow \left| \frac{1}{n} \sum_{i=1}^n w(x_i) \int_{\mathcal{X}^*} \left(\frac{\partial h(x^*, \bar{\theta})}{\partial \theta'} - \frac{\partial h(x^*, \theta_0)}{\partial \theta'} \right) \frac{g_\varepsilon(x_i - x^*)}{g_x(x_i)} \hat{g}_{x^*}(x^*) dx^* \right| \leq \eta$$

Because $\bar{\theta} \xrightarrow{p} \theta_0$ we have that $C_1 \xrightarrow{p} 0$. Also $C_3 \xrightarrow{p} 0$. The term $C_2 \xrightarrow{p} 0$ due to the uniform convergence of \hat{g}_{x^*} .

Next, we consider B_1 . We write

$$B_1 = \frac{1}{\sqrt{n}} \sum_{i=1}^n w(x_i) \left(y_i - \int_{\mathcal{X}^*} h(x^*, \theta_0) g(x^* | x_i) dx^* \right) - \\ - \frac{1}{\sqrt{n}} \sum_{i=1}^n w(x_i) \int_{\mathcal{X}^*} h(x^*, \theta_0) \frac{g_\varepsilon(x_i - x^*)}{g_x(x_i)} (\hat{g}_{x^*}(x^*) - g_{x^*}(x^*)) dx^* \equiv D_1 - D_2$$

with D_1 the moment condition after substitution of the population density of x^* and D_2 the correction term that accounts for the fact that this density is estimated. Because D_1 is obviously $O_p(1)$ the rate of convergence of the INLR estimator is determined by the rate of convergence of D_2 . This is the main point of Taupin's (2001) result.

For D_2 we have

$$D_2 = \frac{1}{\sqrt{n}} \sum_{i=1}^n w(x_i) \int_{\mathcal{X}^*} h(x^*, \theta_0) \frac{g_\varepsilon(x_i - x^*)}{g_x(x_i)} (\hat{g}_{x^*}(x^*) - \tilde{g}_{x^*}(x^*)) dx^* + \\ + \frac{1}{\sqrt{n}} \sum_{i=1}^n w(x_i) \int_{\mathcal{X}^*} h(x^*, \theta_0) \frac{g_\varepsilon(x_i - x^*)}{g_x(x_i)} (\tilde{g}_{x^*}(x^*) - g_{x^*}(x^*)) dx^* = E_1 + E_2$$

where

$$\tilde{g}_{x^*}(x^*) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx^*} \frac{\phi_x(t)}{\phi_\varepsilon(t)} K_n^*(t) dt.$$

In the proof of Theorem 2 we showed that

$$\sup_{x^* \in \mathcal{X}^*} |\tilde{g}_{x^*}(x^*) - g_{x^*}(x^*)| \leq CT_n^{-q}$$

with q the order of the kernel. Now by assumptions (ii)-(iv) of Theorem 3

$$\left| \frac{1}{\sqrt{n}} \sum_{i=1}^n w(x_i) \int_{\mathcal{X}^*} h(x^*, \theta_0) \frac{g_\varepsilon(x_i - x^*)}{g_x(x_i)} dx^* \right| \leq C \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_L^U h(x_i - \varepsilon, \theta_0) g_\varepsilon(\varepsilon) d\varepsilon \right| \leq C\sqrt{n}$$

so that E_2 is bounded by

$$C\sqrt{n}T_n^{-q} \tag{14}$$

Finally, we consider E_1 . We can express it as a U-statistic

$$E_1 = \frac{1}{n\sqrt{n}} \sum_{i=1}^n \sum_{j=1}^n \int_{-\infty}^{\infty} \frac{c^*(x_i, t, \theta_0)}{\phi_\varepsilon(t)} (e^{itx_j} - \phi_x(t)) K_n^*(t) dt \quad (15)$$

with

$$c(x, x^*, \theta) = w(x)h(x^*, \theta) \frac{g_\varepsilon(x - x^*)}{g_x(x)} \quad c^*(x, t, \theta) = \frac{1}{2\pi} \int_{\mathcal{X}^*} e^{-itx^*} c(x, x^*, \theta) dx^*$$

i.e. c^* is a partial Fourier transform of c with respect to x^* . The projection is (\tilde{x} and x have the same distribution)

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \int_{-\infty}^{\infty} \frac{\mathbb{E}[c^*(\tilde{x}, t, \theta_0)]}{\phi_\varepsilon(t)} (e^{itx_i} - \phi_x(t)) K_n^*(t) dt$$

with variance

$$\mathbb{E} \left[\left(\int_{-\infty}^{\infty} \frac{\mathbb{E}[c^*(\tilde{x}, t, \theta_0)]}{\phi_\varepsilon(t)} (e^{itx} - \phi_x(t)) K_n^*(t) dt \right)^2 \right]$$

As noted by Taupin (2001) this variance may be infinite, if the distribution of the measurement error is supersmooth and $c(x, x^*, \theta)$ is not sufficiently smooth, i.e. the regression function is not a polynomial or exponential function. We will show that the variance is always finite if the distribution of the measurement error is range-restricted.

A sufficient condition for a finite variance is that

$$\begin{aligned} & \left| \int_{-\infty}^{\infty} \frac{\mathbb{E}_{\tilde{x}}[c^*(\tilde{x}, t, \theta_0)]}{\phi_\varepsilon(t)} (e^{itx} - \phi_x(t)) K_n^*(t) dt \right| = \\ & = \left| \mathbb{E}_{\tilde{x}} \left[\int_{-\infty}^{\infty} \frac{c^*(\tilde{x}, t, \theta_0)}{\phi_\varepsilon(t)} (e^{itx} - \phi_x(t)) K_n^*(t) dt \right] \right| \leq M < \infty \end{aligned}$$

for all $x \in \mathcal{X}$. Define

$$\kappa(\varepsilon, x) = h(x - \varepsilon, \theta_0) g_\varepsilon(\varepsilon) \quad \kappa^*(t, x) = \int_L^U e^{it\varepsilon} \kappa(\varepsilon, x) d\varepsilon \quad G(x) = \frac{w(x)}{g_x(x)}$$

Then

$$\left| \mathbb{E}_{\tilde{x}} \left[\int_{-\infty}^{\infty} \frac{c^*(\tilde{x}, t, \theta_0)}{\phi_\varepsilon(t)} (e^{itx} - \phi_x(t)) K_n^*(t) dt \right] \right| = \quad (16)$$

$$\begin{aligned}
&= \left| \mathbb{E}_{\tilde{x}} \left[G(\tilde{x}) \int_{-\infty}^{\infty} \frac{\kappa^*(t, \tilde{x})}{\phi_{\varepsilon}(t)} e^{-it\tilde{x}} (e^{itx} - \phi_x(t)) K_n^*(t) dt \right] \right| \leq \\
&\leq \left| \mathbb{E}_{\tilde{x}} \left[G(\tilde{x}) \int_{-\infty}^{\infty} \frac{\kappa^*(t, \tilde{x})}{\phi_{\varepsilon}(t)} e^{it(x-\tilde{x})} K_n^*(t) dt \right] \right| + \left| \mathbb{E}_{\tilde{x}} \left[G(\tilde{x}) \int_{-\infty}^{\infty} \frac{\kappa^*(t, \tilde{x})}{\phi_{\varepsilon}(t)} e^{-it\tilde{x}} \phi_x(t) K_n^*(t) dt \right] \right|
\end{aligned}$$

We consider the final term on the right-hand side first. The integrand involves the function (of t)

$$r(t, x) = \frac{\kappa^*(t, \tilde{x})}{\phi_{\varepsilon}(t)}$$

If the distribution of ε is range-restricted of order k , then a sufficient condition for this ratio being bounded in t is that $\kappa(\varepsilon, x)$ has at least $k + 1$ absolutely integrable derivatives which in turn requires that both $h(x^*, \theta_0)$ and $g_{\varepsilon}(\varepsilon)$ have $k + 1$ absolutely integrable derivatives with respect to x^* and ε respectively. Under this assumption (and by boundedness of w and boundedness from 0 of g_x) the second term is bounded by

$$C \int_{-\infty}^{\infty} |\phi_x(t) K_n^*(t)| dt$$

Because $K_n^*(t) \leq 1$ and ϕ_x is assumed to be absolutely integrable, this bound is finite.

The final step is to show that the first term on the right-hand side of (16) is bounded. We show this in two steps: (i) we expand the function $r(t, x)$ up to $\frac{1}{(it)^2}$, (ii) we show that the terms up to $\frac{1}{it}$ have a finite integral over $(-\infty, \infty)$. We

(i) *Expansion of $r(t, x)$*

If we partially integrate both numerator and denominator two times we obtain, using the notation $f(x)|_L^U = f(U) - f(L)$

$$r(t, x) = \frac{e^{it\varepsilon} \kappa(\varepsilon, x)|_L^U - \frac{1}{it} (e^{it\varepsilon} \kappa'(\varepsilon, x)|_L^U) + \frac{1}{(it)^2} \left(e^{it\varepsilon} \kappa''(\varepsilon, x)|_L^U - \int_L^U e^{it\varepsilon} \kappa'''(\varepsilon, x) d\varepsilon \right)}{e^{it\varepsilon} g_{\varepsilon}(\varepsilon)|_L^U - \frac{1}{it} (e^{it\varepsilon} g'_{\varepsilon}(\varepsilon)|_L^U) + \frac{1}{(it)^2} \left(e^{it\varepsilon} g''_{\varepsilon}(\varepsilon)|_L^U - \int_L^U e^{it\varepsilon} g'''_{\varepsilon}(\varepsilon) d\varepsilon \right)}$$

This suffices if the distribution of ε is range-restricted of order 0. Note that the denominator equals $it\phi_{\varepsilon}(t)$. If the distribution is range-restricted of order

k we need to apply partial integration $k + 3$ times. The proof is similar for this case with some obvious changes.

Using the identity

$$\frac{A}{B} = \frac{A' + (A - A')}{B' + (B - B')} = \frac{A'}{B'} + \frac{1}{B}(A - A') - \frac{A'}{B'B}(B - B')$$

with

$$\begin{aligned} A' &= e^{it\varepsilon} \kappa(\varepsilon, x)|_L^U - \frac{1}{it} (e^{it\varepsilon} \kappa'(\varepsilon, x)|_L^U) \\ B' &= e^{it\varepsilon} g_\varepsilon(\varepsilon)|_L^U - \frac{1}{it} (e^{it\varepsilon} g'_\varepsilon(\varepsilon)|_L^U) \end{aligned}$$

we have

$$\begin{aligned} r(t, x) &= \frac{A'}{B'} + \left[\frac{1}{it\phi_\varepsilon(t)} \left(e^{it\varepsilon} \kappa''(\varepsilon, x)|_L^U - \int_L^U e^{it\varepsilon} \kappa'''(\varepsilon, x) d\varepsilon \right) - \right. \\ &\quad \left. - \frac{1}{it\phi_\varepsilon(t)} \frac{A'}{B'} \left(e^{it\varepsilon} g''_\varepsilon(\varepsilon)|_L^U - \int_L^U e^{it\varepsilon} g'''_\varepsilon(\varepsilon) d\varepsilon \right) \right] \frac{1}{(it)^2} \end{aligned}$$

The next step is to use the identity

$$\begin{aligned} \frac{A'}{B'} &= \frac{A'' + (A' - A'')}{B'' + (B' - B'')} = \frac{A''}{B''} + \frac{1}{B''}(A' - A'') - \frac{A''}{(B'')^2}(B' - B'') + \frac{A''}{B'(B'')^2}(B' - B'')^2 - \\ &\quad - \frac{1}{B''B'}(A' - A'')(B' - B'') \end{aligned}$$

with

$$\begin{aligned} A'' &= e^{it\varepsilon} \kappa(\varepsilon, x)|_L^U \\ B'' &= e^{it\varepsilon} g_\varepsilon(\varepsilon)|_L^U \end{aligned}$$

to write

$$\begin{aligned} \frac{A'}{B'} &= \frac{e^{it\varepsilon} \kappa(\varepsilon, x)|_L^U}{e^{it\varepsilon} g_\varepsilon(\varepsilon)|_L^U} + \left(\frac{(e^{it\varepsilon} \kappa(\varepsilon, x)|_L^U) (e^{it\varepsilon} g'_\varepsilon(\varepsilon)|_L^U)}{(e^{it\varepsilon} g_\varepsilon(\varepsilon)|_L^U)^2} - \frac{e^{it\varepsilon} \kappa'(\varepsilon, x)|_L^U}{e^{it\varepsilon} g_\varepsilon(\varepsilon)|_L^U} \right) \frac{1}{it} + \\ &\quad + \left(\frac{e^{it\varepsilon} \kappa(\varepsilon, x)|_L^U}{B'} \frac{(e^{it\varepsilon} g'_\varepsilon(\varepsilon)|_L^U)^2}{(e^{it\varepsilon} g_\varepsilon(\varepsilon)|_L^U)^2} - \frac{(e^{it\varepsilon} g'_\varepsilon(\varepsilon)|_L^U) (e^{it\varepsilon} \kappa'(\varepsilon, x)|_L^U)}{(e^{it\varepsilon} g_\varepsilon(\varepsilon)|_L^U) B'} \right) \frac{1}{(it)^2} \end{aligned}$$

Substitution gives the following expansion

$$r(t, x) = \kappa_1(t, x) + \kappa_2(t, x) \frac{1}{it} + \kappa_3(t, x) \frac{1}{(it)^2}$$

with

$$\begin{aligned} \kappa_1(t, x) &= \frac{e^{itU} \kappa(U, x) - e^{itL} \kappa(L, x)}{e^{itU} g_\varepsilon(U) - e^{itL} g_\varepsilon(L)} \\ \kappa_2(t, x) &= \frac{(e^{itU} \kappa(U, x) - e^{itL} \kappa(L, x)) (e^{itU} g'_\varepsilon(U) - e^{itL} g'_\varepsilon(L))}{(e^{itU} g_\varepsilon(U) - e^{itL} g_\varepsilon(L))^2} - \frac{e^{itU} \kappa'(U, x) - e^{itL} \kappa'(L, x)}{e^{itU} g_\varepsilon(U) - e^{itL} g_\varepsilon(L)} \\ \kappa_3(t, x) &= \frac{1}{it\phi_\varepsilon(t)} \left(e^{itU} \kappa''(U, x) - e^{itL} \kappa''(L, x) - \int_L^U e^{it\varepsilon} \kappa'''(\varepsilon, x) d\varepsilon \right) - \\ &\quad - \frac{1}{it\phi_\varepsilon(t)} \frac{e^{itU} \kappa(U, x) - e^{itL} \kappa(L, x) - \frac{1}{it} (e^{itU} \kappa'(U, x) - e^{itL} \kappa'(L, x))}{e^{itU} g_\varepsilon(U) - e^{itL} g_\varepsilon(L) - \frac{1}{it} (e^{itU} g'_\varepsilon(U) - e^{itL} g'_\varepsilon(L))} \\ &\quad \cdot \left(e^{itU} g''_\varepsilon(U) - e^{itL} g''_\varepsilon(L) - \int_L^U e^{it\varepsilon} g'''_\varepsilon(\varepsilon) d\varepsilon \right) + \\ &\quad + \frac{e^{itU} \kappa(U, x) - e^{itL} \kappa(L, x)}{e^{itU} g_\varepsilon(U) - e^{itL} g_\varepsilon(L) - \frac{1}{it} (e^{itU} g'_\varepsilon(U) - e^{itL} g'_\varepsilon(L))} \cdot \frac{(e^{itU} g'_\varepsilon(U) - e^{itL} g'_\varepsilon(L))^2}{(e^{itU} g_\varepsilon(U) - e^{itL} g_\varepsilon(L))^2} \\ &\quad - \frac{(e^{itU} g'_\varepsilon(U) - e^{itL} g'_\varepsilon(L)) (e^{itU} \kappa'(U, x) - e^{itL} \kappa'(L, x))}{(e^{itU} g_\varepsilon(U) - e^{itL} g_\varepsilon(L)) (e^{itU} g_\varepsilon(U) - e^{itL} g_\varepsilon(L) - \frac{1}{it} (e^{itU} g'_\varepsilon(U) - e^{itL} g'_\varepsilon(L)))} \end{aligned}$$

Note that κ_1 and κ_2 are well-defined because

$$|e^{itU} g_\varepsilon(U) - e^{itL} g_\varepsilon(L)| \geq |g_\varepsilon(U) - g_\varepsilon(L)| > 0$$

If t is large, then by the same argument

$$\begin{aligned} &\left| e^{itU} g_\varepsilon(U, x) - e^{itL} g_\varepsilon(L, x) - \frac{1}{it} (e^{itU} g'_\varepsilon(U, x) - e^{itL} g'_\varepsilon(L, x)) \right| \geq \\ &\geq \left| e^{itU} g_\varepsilon(U, x) - e^{itL} g_\varepsilon(L, x) \right| - \frac{1}{t} \left| (e^{itU} g'_\varepsilon(U, x) - e^{itL} g'_\varepsilon(L, x)) \right| > 0 \end{aligned}$$

Also all numerators in κ_3 are bounded in t, x , if $h'''(x^*, \theta_0)$ and g_ε have three absolutely integrable derivatives, so that

$$\kappa_3(t, x) \leq M < \infty$$

(ii) *Finiteness of the integral*

We consider

$$\left| \mathbb{E}_{\tilde{x}} \left[G(\tilde{x}) \int_{-\infty}^{\infty} \frac{\kappa^*(t, \tilde{x})}{\phi_\varepsilon(t)} e^{it(x-\tilde{x})} K_n^*(t) dt \right] \right|$$

Let t_0 be such that the expansion in (i) is well-defined for $|t| \geq t_0$. Then

$$\begin{aligned} \left| \mathbb{E}_{\tilde{x}} \left[G(\tilde{x}) \int_{-\infty}^{\infty} \frac{\kappa^*(t, \tilde{x})}{\phi_\varepsilon(t)} e^{it(x-\tilde{x})} K_n^*(t) dt \right] \right| &\leq \mathbb{E}_{\tilde{x}} \left[|G(\tilde{x})| \int_{0 \leq |t| \leq t_0} \left| \frac{\kappa^*(t, \tilde{x})}{\phi_\varepsilon(t)} \right| |K_n^*(t)| dt \right] + \\ &+ \left| \mathbb{E}_{\tilde{x}} \left[G(\tilde{x}) \int_{t_0 \leq |t| < \infty} \frac{\kappa^*(t, \tilde{x})}{\phi_\varepsilon(t)} e^{it(x-\tilde{x})} K_n^*(t) dt \right] \right| \end{aligned}$$

The first term on the right-hand side is finite if $r(t, x)$ is bounded in t (see above) and x which holds if $h(x^*, \theta_0)$ is bounded in x^* .

We show that the second term is finite by substitution of the expansion in (i) which gives

$$\begin{aligned} \left| \mathbb{E}_{\tilde{x}} \left[G(\tilde{x}) \int_{t_0 \leq |t| < \infty} \frac{\kappa^*(t, \tilde{x})}{\phi_\varepsilon(t)} e^{it(x-\tilde{x})} K_n^*(t) dt \right] \right| &\leq \left| \mathbb{E}_{\tilde{x}} \left[G(\tilde{x}) \int_{t_0 \leq |t| < \infty} \kappa_1(t, \tilde{x}) e^{it(x-\tilde{x})} K_n^*(t) dt \right] \right| + \\ &+ \left| \mathbb{E}_{\tilde{x}} \left[G(\tilde{x}) \int_{t_0 \leq |t| < \infty} \frac{1}{it} \kappa_2(t, \tilde{x}) e^{it(x-\tilde{x})} K_n^*(t) dt \right] \right| + \\ &+ \left| \mathbb{E}_{\tilde{x}} \left[G(\tilde{x}) \int_{t_0 \leq |t| < \infty} \frac{1}{(it)^2} \kappa_3(t, \tilde{x}) e^{it(x-\tilde{x})} K_n^*(t) dt \right] \right| \end{aligned} \tag{17}$$

The final term is bounded by

$$C \mathbb{E}_{\tilde{x}} [|G(\tilde{x})|] \int_{t_0 \leq |t| < \infty} \frac{1}{|t|^2} |K_n^*(t)| dt < \infty$$

so that we only need to consider the first two terms on the right-hand side.

Substitution of $\kappa(\varepsilon, x)$ in κ_1 gives

$$\kappa_1(t, x) = \frac{h(x - U, \theta_0) - e^{it(L-U)} \frac{g_\varepsilon(L)}{g_\varepsilon(U)} h(x - L, \theta_0)}{1 - e^{it(L-U)} \frac{g_\varepsilon(L)}{g_\varepsilon(U)}}$$

is the characteristic function of a discrete random variable z with

$$\Pr(z = (L - U)j) = \left(1 - \frac{g_\varepsilon(L)}{g_\varepsilon(U)}\right) \left(\frac{g_\varepsilon(L)}{g_\varepsilon(U)}\right)^j$$

for $j = 0, 1, \dots$. Hence

$$\int_{-\infty}^{\infty} \frac{1}{1 - e^{it(L-U)} \frac{g_\varepsilon(L)}{g_\varepsilon(U)}} e^{it(x-\tilde{x})} K_n^*(t) dt = \frac{2\pi}{1 - \frac{g_\varepsilon(L)}{g_\varepsilon(U)}} \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{it(x-\tilde{x})} \phi_z(t) K_n^*(t) dt$$

Because the density corresponding to $K_n^*(t)$ is $T_n K(T_n v)$, this is equal to

$$\frac{2\pi}{1 - \frac{g_\varepsilon(L)}{g_\varepsilon(U)}} \sum_{j=0}^{\infty} T_n K(T_n(\tilde{x} - x - (L - U)j)) \left(\frac{g_\varepsilon(L)}{g_\varepsilon(U)}\right)^j$$

Hence

$$\begin{aligned} & \left| \mathbb{E}_{\tilde{x}} \left[G(\tilde{x}) h(\tilde{x} - U, \theta_0) \int_{-\infty}^{\infty} \frac{1}{1 - e^{it(L-U)} \frac{g_\varepsilon(L)}{g_\varepsilon(U)}} e^{it(x-\tilde{x})} K_n^*(t) dt \right] \right| \leq \\ & \leq C \sum_{j=0}^{\infty} \left(\frac{g_\varepsilon(L)}{g_\varepsilon(U)}\right)^j \left| \mathbb{E}_{\tilde{x}} [G(\tilde{x}) h(\tilde{x} - U, \theta_0) T_n K(T_n(\tilde{x} - x - (L - U)j))] \right| \end{aligned}$$

Finally, for $w = T_n(\tilde{x} - x - (L - U)j)$

$$\begin{aligned} & \left| \mathbb{E}_{\tilde{x}} [G(\tilde{x}) h(\tilde{x} - U, \theta_0) T_n K(T_n(\tilde{x} - x - (L - U)j))] \right| = \\ & = \left| \int G \left(\frac{w}{T_n} + x + (L - U)j \right) h \left(\frac{w}{T_n} + x + (L - U)j - U, \theta_0 \right) \right. \\ & \quad \left. \cdot K(w) g_x \left(\frac{w}{T_n} + x + (L - U)j \right) dw \right| \leq \\ & \leq C \int_{-\infty}^{\infty} |K(w)| dw < \infty \end{aligned}$$

Using

$$\frac{e^{it(L-U)} \frac{g_\varepsilon(L)}{g_\varepsilon(U)}}{1 - e^{it(L-U)} \frac{g_\varepsilon(L)}{g_\varepsilon(U)}} = \sum_{j=1}^{\infty} \left(\frac{g_\varepsilon(L)}{g_\varepsilon(U)}\right)^j e^{it(L-U)j}$$

we use the same proof to show that the second term on the right-hand side of (18) is finite.

Finally, we consider the second term on the right-hand side of (17). First, we have

$$\begin{aligned} \kappa_2(t, x) &= \frac{\left(h(x-U, \theta_0) - e^{it(L-U)} \frac{g_\varepsilon(L)}{g_\varepsilon(U)} h(x-L, \theta_0) \right) \left(\frac{g'_\varepsilon(U) - e^{it(L-U)} g'_\varepsilon(L)}{g_\varepsilon(U)} \right)}{\left(1 - e^{it(L-U)} \frac{g_\varepsilon(L)}{g_\varepsilon(U)} \right)^2} \\ &\quad - \frac{\frac{h(x-U, \theta_0) g'_\varepsilon(U) - e^{it(L-U)} h(x-L, \theta_0) g'_\varepsilon(L)}{g_\varepsilon(U)}}{1 - e^{it(L-U)} \frac{g_\varepsilon(L)}{g_\varepsilon(U)}} + \frac{h'(x-U, \theta_0) - e^{it(L-U)} \frac{g_\varepsilon(L)}{g_\varepsilon(U)} h'(x-L, \theta_0)}{1 - e^{it(L-U)} \frac{g_\varepsilon(L)}{g_\varepsilon(U)}} \end{aligned}$$

Substitution gives the bound

$$\begin{aligned} &\left| \mathbf{E}_{\tilde{x}} \left[G(\tilde{x}) \int_{t_0 \leq |t| < \infty} \frac{\left(h(\tilde{x}-U, \theta_0) - e^{it(L-U)} \frac{g_\varepsilon(L)}{g_\varepsilon(U)} h(\tilde{x}-L, \theta_0) \right) \left(\frac{g'_\varepsilon(U) - e^{it(L-U)} g'_\varepsilon(L)}{g_\varepsilon(U)} \right)}{\left(1 - e^{it(L-U)} \frac{g_\varepsilon(L)}{g_\varepsilon(U)} \right)^2} \right. \right. \\ &\quad \left. \left. + \frac{e^{it(x-\tilde{x})}}{it} K_n^*(t) dt \right] \right| + \\ &+ \left| \mathbf{E}_{\tilde{x}} \left[G(\tilde{x}) \int_{t_0 \leq |t| < \infty} \frac{\frac{h(\tilde{x}-U, \theta_0) g'_\varepsilon(U) - e^{it(L-U)} h(\tilde{x}-L, \theta_0) g'_\varepsilon(L)}{g_\varepsilon(U)}}{1 - e^{it(L-U)} \frac{g_\varepsilon(L)}{g_\varepsilon(U)}} \frac{e^{it(x-\tilde{x})}}{it} K_n^*(t) dt \right] \right| + \\ &+ \left| \mathbf{E}_{\tilde{x}} \left[G(\tilde{x}) \int_{t_0 \leq |t| < \infty} \frac{h'(\tilde{x}-U, \theta_0) - e^{it(L-U)} \frac{g_\varepsilon(L)}{g_\varepsilon(U)} h'(\tilde{x}-L, \theta_0)}{1 - e^{it(L-U)} \frac{g_\varepsilon(L)}{g_\varepsilon(U)}} \frac{e^{it(x-\tilde{x})}}{it} K_n^*(t) dt \right] \right| \end{aligned} \quad (20)$$

We show that the final term of (20) is bounded (in x). It is bounded by

$$\begin{aligned} &\left| \mathbf{E}_{\tilde{x}} \left[G(\tilde{x}) h'(\tilde{x}-U, \theta_0) \int_{t_0 \leq |t| < \infty} \frac{1}{1 - e^{it(L-U)} \frac{g_\varepsilon(L)}{g_\varepsilon(U)}} \frac{e^{it(x-\tilde{x})}}{it} K_n^*(t) dt \right] \right| + \quad (21) \\ &+ \left| \mathbf{E}_{\tilde{x}} \left[G(\tilde{x}) \frac{g_\varepsilon(L)}{g_\varepsilon(U)} h'(x-L, \theta_0) \int_{t_0 \leq |t| < \infty} \frac{1}{1 - e^{it(L-U)} \frac{g_\varepsilon(L)}{g_\varepsilon(U)}} \frac{e^{it(x-\tilde{x})+it(L-U)}}{it} K_n^*(t) dt \right] \right| \end{aligned}$$

The first term of (21) is bounded by

$$\begin{aligned} & \left| \mathbb{E}_{\tilde{x}} \left[G(\tilde{x}) h'(\tilde{x} - U, \theta_0) \int_{|t| \leq t_0} \frac{1}{1 - e^{it(L-U) \frac{g_\varepsilon(L)}{g_\varepsilon(U)}}} \frac{e^{it(x-\tilde{x})}}{it} K_n^*(t) dt \right] \right| + \quad (22) \\ & + \left| \mathbb{E}_{\tilde{x}} \left[G(\tilde{x}) h'(\tilde{x} - U, \theta_0) \int_{-\infty}^{\infty} \frac{1}{1 - e^{it(L-U) \frac{g_\varepsilon(L)}{g_\varepsilon(U)}}} \frac{e^{it(x-\tilde{x})}}{it} K_n^*(t) dt \right] \right| \end{aligned}$$

The second term of (22) contains the integral

$$s(x - \tilde{x}) \equiv \int_{-\infty}^{\infty} \frac{1}{1 - e^{it(L-U) \frac{g_\varepsilon(L)}{g_\varepsilon(U)}}} \frac{e^{it(x-\tilde{x})}}{it} K_n^*(t) dt$$

with derivative

$$\begin{aligned} s'(x - \tilde{x}) &= \int_{-\infty}^{\infty} \frac{1}{1 - e^{it(L-U) \frac{g_\varepsilon(L)}{g_\varepsilon(U)}}} e^{it(x-\tilde{x})} K_n^*(t) dt = \\ &= \frac{2\pi}{1 - \frac{g_\varepsilon(L)}{g_\varepsilon(U)}} \sum_{j=0}^{\infty} T_n K(T_n(\tilde{x} - x - (L - U)j)) \left(\frac{g_\varepsilon(L)}{g_\varepsilon(U)} \right)^j \end{aligned}$$

so that

$$s(x - \tilde{x}) = -\frac{2\pi}{1 - \frac{g_\varepsilon(L)}{g_\varepsilon(U)}} \sum_{j=0}^{\infty} H(T_n(\tilde{x} - x - (L - U)j)) \left(\frac{g_\varepsilon(L)}{g_\varepsilon(U)} \right)^j$$

with $H(v) = \int_{-\infty}^v K(s) ds$ the integral of K which is a bounded function.

Hence, the second term on the right-hand side of (22) is bounded by

$$C \sum_{j=0}^{\infty} \left(\frac{g_\varepsilon(L)}{g_\varepsilon(U)} \right)^j |\mathbb{E}_{\tilde{x}} [G(\tilde{x}) h'(\tilde{x} - U, \theta_0) H(T_n(\tilde{x} - x - (L - U)j))]| \leq M < \infty$$

because $G(x)$, $h'(x^*, \theta_0)$, $H(v)$ are all bounded.

To bound the first term on the right-hand side of (22) we note that $K^*(t) = \int_{-\infty}^{\infty} e^{itv} K(v) dv = \int_{-\infty}^{\infty} \cos(tv) K(v) dv$ because K is an even function, so that K^* is real and even. This implies that, because $\frac{\sin t}{t}$ is even, $\frac{\cos t}{t}$ is odd and $K^*\left(\frac{t}{T_n}\right) = 1$ if $|t| \leq T_n$

$$\int_{|t| \leq t_0} \frac{1}{1 - e^{it(L-U) \frac{g_\varepsilon(L)}{g_\varepsilon(U)}}} \frac{e^{it(x-\tilde{x})}}{it} K_n^*(t) dt = \sum_{j=0}^{\infty} \left(\frac{g_\varepsilon(L)}{g_\varepsilon(U)} \right)^j \int_{|t| \leq t_0} \frac{e^{it(x-\tilde{x}) + itj(L-U)}}{it} K^*\left(\frac{t}{T_n}\right) dt =$$

$$\sum_{j=0}^{\infty} \left(\frac{g_{\varepsilon}(L)}{g_{\varepsilon}(U)} \right)^j \int_{|t| \leq t_0} \frac{\sin t(x - \tilde{x} + j(L - U))}{t} dt$$

Now

$$\int_{|t| \leq t_0} \frac{\sin t(x - \tilde{x} + j(L - U))}{t} dt = \int_{|t| \leq t_0(x - \tilde{x} + j(L - U))} \frac{\sin t}{t} dt \leq M < \infty$$

so that the first term is also a bounded function of x .

The proof that the second term of (21) is bounded is completely analogous. The same method of proof also applies to the second term on the right-hand side of (20). For the first term of (20) we note that

$$\phi_z(t) \equiv \frac{\left(1 - \frac{g_{\varepsilon}(L)}{g_{\varepsilon}(U)}\right)^2}{\left(1 - \frac{g_{\varepsilon}(L)}{g_{\varepsilon}(U)} e^{it(L-U)}\right)^2}$$

is the characteristic function of $z = z_1 + z_2$ where z_1, z_2 are independent and have the same distribution

$$\Pr(z_k = (L - U)j) = \left(1 - \frac{g_{\varepsilon}(L)}{g_{\varepsilon}(U)}\right) \left(\frac{g_{\varepsilon}(L)}{g_{\varepsilon}(U)}\right)^j$$

for $j = 0, 1, \dots$ and $k = 1, 2$. Expressing

$$\frac{\left(1 - \frac{g_{\varepsilon}(L)}{g_{\varepsilon}(U)}\right)^2}{\left(1 - \frac{g_{\varepsilon}(L)}{g_{\varepsilon}(U)} e^{it(L-U)}\right)^2} = \left(1 - \frac{g_{\varepsilon}(L)}{g_{\varepsilon}(U)}\right)^2 \sum_{j=0}^{\infty} (j+1) \left(\frac{g_{\varepsilon}(L)}{g_{\varepsilon}(U)}\right)^j e^{it(L-U)j}$$

we see that the same method of proof can be applied to the first term of (20). We conclude that the variance is indeed finite. \square

References

- [1] Chen, X., H. Hong, and E. Tamer (2005), 'Measurement error models with auxiliary data', *Review of Economic Studies* 72, 343-366.
- [2] Fan, J. (1991), 'On the optimal rates of convergence for nonparametric deconvolution problems', *Annals of Statistics* 19, 1257-1272.
- [3] Feller, W. (1971), *An Introduction to Probability Theory and Its Applications*, vol II, Vol. 2., Wiley, New York.
- [4] Horowitz, J.L., and M. Markatou, (1996), 'Semiparametric estimation of regression models for panel data', *Review of Economic Studies* 63, 145-168.
- [5] Hu, Y. and G. Ridder (2003), 'Estimation of nonlinear models with measurement error using marginal information', Working paper.
- [6] Li, T. (2002), 'Robust and consistent estimation of nonlinear errors-in-variables models', *Journal of Econometrics* 110, 1-26.
- [7] Li, T. and Q. Vuong (1998), 'Nonparametric estimation of the measurement error model using multiple indicators', *Journal of Multivariate Analysis* 65, 139-165.
- [8] Newey, W. (1994), 'The asymptotic variance of semiparametric estimators', *Econometrica* 62, 1349-1382.
- [9] Pollard, D. (1984), *Convergence of Stochastic Processes*, Springer-Verlag, New York.
- [10] Schennach, S. (2004), 'Estimation of nonlinear models with measurement error', *Econometrica*, 72, 33-75.