On Deconvolution as a First Stage Nonparametric Estimator

Yingyao Hu
Department of Economics,
Johns Hopkins University
E-mail: yhu@jhu.edu

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

December 16, 2008

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 a class of distributions with restricted support, then much weaker smoothness assumptions than hers suffice to ensure $\sqrt{n}$ consistency of the estimator. In addition we show that the INLR estimator remains consistent under these weaker smoothness assumptions if the support of the measurement error distribution expands with the sample size. In that case the estimator remains also asymptotically normal with a rate of convergence that is arbitrarily close to $\sqrt{n}$. 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 model[1] 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

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

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 estimators in the second stage.

This result can be generalized to the case that the support of the measurement error distribution expands with the number of observations.[2] The rate of uniform convergence of the first stage nonparametric estimator is with an appropriate choice of the rate at which the support expands arbitrarily close to that obtained for measurement errors with restricted range, even if measurement error distribution that is truncated at a decreasing rate is supersmooth, e.g. has a normal distribution. For instance, if the measurement error distribution is truncated normal with truncation points that diverge at the $\sqrt{\log n}$ rate, then the INLR estimator is asymptotically normal with a rate of convergence that is arbitrarily close to $\sqrt{n}$.

2 The integrated nonlinear regression estimator

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

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

We assume that

$$\varepsilon \perp x^*$$

i.e. the measurement error is classical. Note that restricted supports of x and x^* are compatible with independence of these variables, if the support of x is larger than that of x^*[3]. Independence also implies that the variance

[2] This extension was suggested by one of the referees.

[3] If x and x^* are discrete independence only holds in special cases.

of x is larger than that of x^*. Under the assumptions made the parameters are not identified if the measurement error is nonclassical and the correlation between x and x^* is unknown. With a validation sample no assumptions on the correlation are needed (see e.g. Chen, Hong, and Tamer (2005)).

The parametric model specifies the conditional mean function

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

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

$$\mathrm{E}(y|x) = \int_{\mathcal{X}^*} h(x^*, \theta_0) g(x^*|x) \mathrm{d}x^*$$

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

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

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

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

The corresponding density estimator is

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

with

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

the empirical characteristic function (ecf) of x (F_n is the empirical cdf of the sample $x_1, \ldots, x_n$). 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 $j = 1, 2, ..., q-1$, and $\left|\int z^q K(z)dz\right| < \infty$, i.e. K is a higher order kernel of order q, for $q \geq 2$.

A detailed discussion of the kernel functions can be found in Taupin (2001, remark 2.2). Substitution of this estimator in the moment function gives the INLR estimator as the solution to

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

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 integrable[4]. 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. We say that the distribution of a random variable v is *range restricted of order* k if v has a density (with respect to the Lebesgue measure) f_v that is positive on $[L,U]$ with either L or U finite, equals zero outside $[L,U]$, and has $k+2$ for $k \geq 0$ absolutely integrable derivatives $f_v^{(j)}$ with

(i) $|f_v^{(k)}(U)| \neq |f_v^{(k)}(L)|$,

[4]Here and in the sequel C denotes a generic constant.

(ii) if $k > 1$, $f_v^{(j)}(U) = f_v^{(j)}(L) = 0$ for $j = 0, \ldots, k-1$.

The next theorem establishes that all range restricted distributions have ordinary smooth characteristic functions.

Theorem 1 *If a random variable v is range restricted of order k, then its characteristic function is ordinary smooth. In particular, there is a $t_0 > 0$ such that for all $|t| \geq t_0$ and some C_0, $C_1 > 0$*

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

Proof In appendix. □

Remark 1 *For the truncated 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 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 represented by such mixtures.

Note that in general both the lower and the upper bounds converge to 0 at a slower rate for range-restricted distributions if it is a truncation of some distribution that has a characteristic function that decreases at a faster rate. That is also true for distributions that have an ordinary smooth characteristic function when 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 at least as fast as $|t|^{-l}$. If we truncate the support, the rate can be as slow as $|t|^{-1}$.

A random variable with a distribution that is obtained by truncating a distribution with unbounded support from below at L_n and from above by U_n is denoted by v_n. The underlying untruncated random variable is v. The density function of v_n is

$$f_{v_n}(v) = \frac{f_v(v)}{\int_{L_n}^{U_n} f_v(z)\,\mathrm{d}z}$$

for $L_n \leq x \leq U_n$. If the range restricted distribution is obtained by truncating a distribution with unbounded support, the following corollary gives a lower bound on the rate at which $|\phi_{v_n}(T_n)|$ converges to 0 if $T_n \to \infty$ and the bounds on the support $L_n \to -\infty$ and $U_n \to \infty$.

Corollary 1 *Suppose that f_v has absolutely integrable derivatives $f_v^{(j)}$ on $\Re$ for $j = 1, 2$. Let $T_n = O\left(\left(\frac{n}{\log n}\right)^{\gamma}\right)$ and $U_n \to \infty$ such that*

$$\frac{1}{f_{v_n}(U_n)} = O\left(\left(\frac{n}{\log n}\right)^{\delta}\right)$$

with $\delta < \gamma$. Moreover, let $L_n \to -\infty$ such that

$$\frac{f_{v_n}(L_n)}{f_{v_n}(U_n)} \to 0.$$

Then there is an n_0 such that for $n \geq n_0$

$$|\phi_{v_n}(T_n)| \geq C\left(\frac{\log n}{n}\right)^{\delta+\gamma}.$$

The conditions on L_n are satisfied if $L_n = -\infty$. For the standard normal distribution the condition on U_n is satisfied if

$$U_n = O\left(\sqrt{\log\left(\frac{n}{\log n}\right)}\right).$$

Note that the lower and upper bound are treated asymmetrically. We can interchange their roles and reach the same conclusion.

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 *Suppose that $|\phi_\varepsilon(t)| > 0$ for all $t \in \Re$, that the distribution of ε is range restricted of order k, and that the distribution of x^* has a density that is q times differentiable and the q-th derivative is continuous and bounded on $\mathcal{X}^*$. 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. Then for an arbitrary $\eta > 0$ a.s.*

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

Proof In appendix. □

As shown in the proof, we introduce the constant η in order to allow the smoothing parameter T_n to increase at the specified rate (and not slower than that rate). This theorem shows that 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 $q \geq 4$.

If the distribution of the measurement error is obtained by truncating a distribution with unbounded support at L_n and U_n we can replace the assumption that the distribution of the measurement error is range restricted by the assumptions on the rate at which the truncation points expand as in Corollary 1. Note that these assumptions imply that the support of the distribution of the measurement error grows at (a slightly) slower rate than T_n, the truncating parameter in the nonparametric density estimator.

Corollary 2 *Let the distribution of ε_n be obtained by truncating a distribution with unbounded support at L_n from below and U_n from above and let the* *first two derivatives of the untruncated density be absolutely integrable. For sequences U_n, L_n, T_n as specified in Corollary 1 and conditions on the kernel K and the distribution of x^* as in Theorem 2, we have*

$$\sup_{x^*\in\mathcal{X}^*} |\hat{g}_{x^*}(x^*) - g_{x^*}(x^*)| = O\left(\left(\frac{\log n}{n}\right)^{\frac{1}{2}-3\gamma-\delta-\eta}\right) + O\left(\left(\frac{\log n}{n}\right)^{q\gamma}\right). \quad (4)$$

for $0 < \delta < \gamma < 1/2$ and all $\eta > 0$.

Proof In appendix. □

The best rate of convergence is now $n^{-\frac{q}{2(q+4)}+\eta}$ which is slightly slower than that if L, U are fixed (and $k = 0$).

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) $|\phi_\varepsilon(t)| > 0$ *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. □

The assumption on the boundedness of $h(\cdot)$ is sufficient but by no means necessary. It can be replaced by boundedness assumptions on the moment function and the Frechet differential of the moment function. However, we prefer to give sufficient conditions that can be verified more easily in most applications.

We can also establish consistency if the distribution of ε_n is obtained by truncating a distribution with unbounded support at L_n and U_n.

Corollary 3 *If we replace* (v) *by the assumption that the first two derivatives of the untruncated measurement error density are absolutely integrable,* L_n, U_n *are as in Corollary 1, and in (vii)* $0 < \gamma < 1/8$ *and* $0 < \delta < \gamma$*, then the INLR estimator is weakly consistent.*

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 (i)-(vii) of Theorem 3 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) $g_\varepsilon(\varepsilon)$ *have at least* $k+3$ *absolutely integrable derivatives.*

(iv) rank $G = d_\theta$ *with* d_θ *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) $\sqrt{n}T_n^{-q} \to 0$

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

$$\sqrt{n}(\hat{\theta}-\theta_0) = G^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} w(x_i)\left(y_i - \int_{\mathcal{X}^*} h(x^*,\theta_0)g(x^*|x_i)\,dx^*\right) - \right.$$

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

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. □

If the distribution of the measurement error is obtained by truncating an unbounded distribution and we let the lower and upper bound on the support diverge, then the INLR estimator has a slower rate of convergence, but at that slower rate the estimator is still asymptotically normal.

Corollary 4 *Suppose that the assumptions of Corollary 3 and Theorem 4 hold with the exception of (iii) in Theorem 4, then*

$$\sqrt{n}g_{\varepsilon_n}(U_n)(\hat{\theta}-\theta_0) = G^{-1}\left(\frac{g_{\varepsilon_n}(U_n)}{\sqrt{n}}\sum_{i=1}^{n} w(x_i)\left(y_i - \int_{\mathcal{X}^*} h(x^*,\theta_0)g(x^*|x_i)\,dx^*\right) - \right.$$

$$-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int \frac{E[c^*(x,t,\theta_0)]}{\phi_{\varepsilon_n}(t)/g_{\varepsilon_n}(U_n)}(e^{itx_i}-\phi_x(t))K_n^*(t)\,dt\Bigg)+o_p(1)$$

The variance of the final term is finite (and in general strictly positive).

We can replace $g_{\varepsilon_n}(U_n)$ by $g_\varepsilon(U_n)$ with g_ε the underlying untruncated density. The rate of convergence is essentially $n^{\frac{1}{2}}g_\varepsilon(U_n)=n^{\frac{1}{2}-\delta}$. Hence the estimator converges slower if we let the support of the measurement error expand faster. The upper limit is $\delta<1/8$. Because of the slower rate the main term in the asymptotically linear representation is asymptotically negligible and the asymptotic distribution depends only on the correction term.

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.

The assumption that the support of the measurement error distribution is bounded may not be appealing to everyone. Corollary 4 shows that we can let the bounded support expand with the sample size. The resulting INLR estimator is still consistent and asymptotically normal, be it at a marginally slower rate that depends on the rate of expansion. Therefore we can adjust the bounds on the support to an increasing sample size, as one likely do in practice.

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.

Finally, we show the performance of a deconvolution estimator when the error distribution is range-restricted. Consider a probit model with $f(y|x^*, w) = \Phi(\beta_0 + \beta_1 x^* + \beta_2 w)^y [1 - \Phi(\beta_0 + \beta_1 x^* + \beta_2 w)]^{1-y}$, where observables include (y, x, w) with $x = x^* + \varepsilon$ and $w \sim N(0, .25)$. In the simulation, the distributions of x^* and ε are truncated normal generated by truncating $N(0, .25)$ from above at point C_{trunc}. Instead of using the true characteristic function of x^*, we use a secondary sample only containing x^* to estimate the characteristic function of x^*, and then use the deconvolution estimator in Hu and Ridder (2003) to estimate $(\beta_0, \beta_1, \beta_2)$. The smoothing parameters of the density estimators (equations 17 and 18 in Hu and Ridder (2003)) are 0.6 and (0.7,0.7). Tables 1 and 2 provide the simulation results, together with the sample sizes and the repetition times. The results imply that the deconvolution estimator reduces the bias caused by the measurement error with a larger variance due to the nonparametric estimation. Figures 1 and 2 show the normal probability plots of the deconvolution estimates $\widehat{\beta}_1$ and $\widehat{\beta}_2$ with $C_{trunc} = 4$. The figures imply that the empirical distributions of the estimates are very close to a normal except in the tail area. The results show that the deconvolution estimator performs well when the error distribution is range-restricted.

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. Moreover, we can expand the support with the sample size and still have a consistent and asymptotically normal estimator, be it with a somewhat slower rate of convergence, that may still be acceptable. Therefore, semi-parametric estimators with a nonparametric deconvolution estimator as a first stage can be applied if the model is not in the class considered by Taupin, if we are prepared to make relatively weak assumptions on the distribution of the measurement error.

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

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

Hence

$$|\phi_v(t)| \geq \frac{1}{|t|}\left|\left|\left(f_v(U) - \frac{1}{it}f_v'(U)\right)e^{itU} - \left(f_v(L) - \frac{1}{it}f_v'(L)\right)e^{itL}\right| - \left|\frac{1}{it}\int_L^U e^{itx} f_v''(x)\mathrm{d}x\right|\right| \tag{5}$$

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

$$\left|\left(f_v(U) - \frac{1}{it}f_v'(U)\right)e^{itU} - \left(f_v(L) - \frac{1}{it}f_v'(L)\right)e^{itL}\right| \geq \tag{6}$$

$$\left|\left|\left(f_v(U) - \frac{1}{it}f_v'(U)\right)e^{itU}\right| - \left|\left(f_v(L) - \frac{1}{it}f_v'(L)\right)e^{itL}\right|\right| =$$

$$= \left|\sqrt{f_v(U)^2 + \left(\frac{1}{t}f_v'(U)\right)^2} - \sqrt{f_v(L)^2 + \left(\frac{1}{t}f_v'(L)\right)^2}\right|$$

If $|t| \to \infty$ the lower bound in (6) converges to $|f_v(U) - f_v(L)| > 0$, so that there are a t_1 (that depends on L, U) and $C_1 < 1$ such that (6) is greater than $C_1|f_v(U) - f_v(L)|$ for $t \geq t_1$.

The second term in absolute value on the right-hand side of (5) is bounded from above by

$$\left|\frac{1}{it}\int_L^U e^{itx} f_v''(x)\mathrm{d}x\right| \leq \frac{1}{|t|}\int_L^U |f_v''(x)|\mathrm{d}x$$

and this upper bound converges to 0 if $|t| \to \infty$. Therefore there is a t_2 (that depends on L, U) such that for $t \geq t_2$

$$\left|\frac{1}{it}\int_L^U e^{itx} f_v''(x)\mathrm{d}x\right| \leq \frac{C_1}{2}|f_v(U) - f_v(L)|$$

so that for $t \geq t_0 = \max\{t_1, t_2\}$

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

with $C_0 = C_1/2$.

For the upper bound we have

$$|\phi_v(t)| \leq \frac{1}{|t|}\left(\left|\left(f_v(U) - \frac{1}{it}f_v'(U)\right)e^{itU}\right| + \left|\left(f_v(L) - \frac{1}{it}f_v'(L)\right)e^{itL}\right| + \left|\frac{1}{it}\int_L^U e^{itx} f_v''(x)\mathrm{d}x\right|\right) \leq$$

$$\leq \frac{1}{|t|}\left(f_v(U) + \frac{1}{|t|}|f_v'(U)| + f_v(L) + \frac{1}{|t|}|f_v'(L)| + \frac{1}{|t|}\int_L^U |f_v''(x)|\mathrm{d}x\right) \leq$$

$$\leq \frac{C_1(f_v(U) + f_v(L))}{|t|}$$

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. □

Proof of Corollary 1

Consider the lower bound in (6) with U_n, L_n substituted for U, L and T_n for t. Under the assumptions

$$f_{v_n}(U_n)^2 + \left(\frac{1}{T_n} f'_{v_n}(U_n)\right)^2 > f_{v_n}(L_n)^2 + \left(\frac{1}{T_n} f'_{v_n}(L_n)\right)^2$$

for n sufficiently large. Hence the expression is bounded from below by

$$f_{v_n}(U_n)\left[1 - \sqrt{\left(\frac{f_{v_n}(L_n)}{f_{v_n}(U_n)}\right)^2 + \left(\frac{f'_{v_n}(L_n)}{T_n f_{v_n}(U_n)}\right)^2}\right].$$

Note that $f'_{v_n}(U_n) \to 0$, $f'_{v_n}(L_n) \to 0$, and $T_n f_{v_n}(U_n) \to \infty$ as $n \to \infty$. Under the assumptions this lower bound is larger than

$$\frac{1}{T_n} \int_{-\infty}^{\infty} f''_v(x) \mathrm{d}x$$

for n sufficiently large, so that

$$\left|\left|\left(f_{v_n}(U_n) - \frac{1}{iT_n} f'_{v_n}(U_n)\right) e^{iT_n U_n} - \left(f_{v_n}(L_n) - \frac{1}{iT_n} f'_{v_n}(L_n)\right) e^{iT_n L_n}\right| - \left|\frac{1}{iT_n} \int_{L_n}^{U_n} e^{iT_n x} f''_{v_n}(x) \mathrm{d}x\right|\right| \geq$$

$$f_{v_n}(U_n)\left[1 - \sqrt{\left(\frac{f_{v_n}(L_n)}{f_{v_n}(U_n)}\right)^2 + \left(\frac{f'_{v_n}(L_n)}{T_n f_{v_n}(U_n)}\right)^2} - \frac{1}{f_{v_n}(U_n) T_n} \int_{-\infty}^{\infty} f''_v(x) \mathrm{d}x\right] \geq C f_{v_n}(U_n)$$

and the result follows. $\square$.

Proof of Theorem 2

We first establish the rate of uniform convergence of the empirical characteristic function. This lemma corrects a result in Lemma 1 in Horowitz and Markatou (1996, p. 163).

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

$$\sup_{|t| \leq T_n} \left|\hat{\phi}_v(t) - \phi_v(t)\right| = o(\alpha_n) \qquad \text{a.s.} \tag{7}$$

with $\alpha_n = O\left(\left(\frac{\log n}{n}\right)^{\frac{1}{2}-\gamma}\right)$, *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 $\mathcal{G}_n = \{e^{itx} | |t| \le T_n\}$. 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 $\mathcal{G}_{1n} = \{\cos(tx) | |t| \le T_n\}$ and $\{\mathcal{F}_{2n} = \sin(tx) | |t| \le T_n\}$. Because $|\cos(t_2 x) - \cos(t_1 x)| \le |x||t_2 - t_1|$ and $\mathrm{E}(|x|) < \infty$, an $\frac{\varepsilon}{2}\mathrm{E}(|x|)$ cover (with respect to the L_1 norm) of $\mathcal{G}_{1n}$ is obtained from an $\frac{\varepsilon}{2}$ cover of $\{t | |t| \le T_n\}$ by choosing $t_k, k = 1, \ldots, K$ arbitrarily from the distinct covering sets, where K is the smallest integer larger than $\frac{2T_n}{\varepsilon}$. Because $|\sin(t_2 x) - \sin(t_1 x)| \le |x||t_2 - t_1|$, the functions $\sin(t_k x), k = 1, \ldots, K$ are an $\frac{\varepsilon}{2}\mathrm{E}(|x|)$ cover of $\mathcal{F}_{2n}$. Hence $\cos(t_k x) + i\sin(t_k x), k = 1, \ldots, K$ is an $\varepsilon\mathrm{E}(|x|)$ cover of $\mathcal{G}_n$, and we conclude that

$$\mathcal{N}_1(\varepsilon, P, \mathcal{G}_n) \le A\frac{T_n}{\varepsilon} \tag{8}$$

with P an arbitrary probability measure such that $\mathrm{E}(|x|) < \infty$ and $A > 0$, a constant that does not depend on n. The next step is to apply an argument in Pollard (1984). With $\delta_n = 1$ 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)

$$\Pr\left(\sup_{|t|\le T_n} |\hat{\phi}(t) - \phi(t)| > 2\varepsilon_n\right) \le 2A\left(\frac{\varepsilon_n}{T_n}\right)^{-1} \exp\left(-\frac{1}{128}n\varepsilon_n^2\right) + \Pr\left(\sup_{|t|\le T_n} \hat{\phi}(2t) > 64\right). \tag{9}$$

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

$$2A\varepsilon^{-1}\exp\left(\log\left(\frac{T_n}{\alpha_n}\right) - \frac{1}{128}n\varepsilon^2\alpha_n^2\right). \tag{10}$$

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 $\log\left(\frac{T_n}{\alpha_n}\right) - \frac{1}{2}\log n \to -\infty$. The same restrictions imply that $\frac{n\alpha_n^2}{\log n} \to \infty$. The result now follows from the Borel-Cantelli lemma. □

In Lemma 1 we can choose $T_n = O\left(\left(\frac{n}{\log n}\right)^{\gamma}\right)$, but the rate of convergence is then at most $\left(\frac{\log n}{n}\right)^{\frac{1}{2}-\gamma-\eta}$ for an arbitrary $\eta > 0$, i.e. strictly less than the rate of convergence in Lemma 1. In the sequel we prefer this choice of T_n because it suggests how the sequence of T_n should be chosen proportionally to $\left(\frac{n}{\log n}\right)^{\gamma}$.

Proof of Theorem 2 Given that $|\phi_{\varepsilon}(t)| > 0$, we have

$$\sup_{x^*\in\mathcal{X}^*} |\hat{g}_{x^*}(x^*) - g_{x^*}(x^*)| \leq \sup_{x^*\in\mathcal{X}^*} \left| \frac{1}{2\pi} \int e^{-itx^*} \left(\frac{\hat{\phi}_x(t) - \phi_x(t)}{\phi_{\varepsilon}(t)} \right) K_n^*(t)\mathrm{d}t \right| + \tag{11}$$
$$+ \sup_{x^*\in\mathcal{X}^*} \left| \frac{1}{2\pi} \int e^{-itx^*} \phi_{x^*}(t) \left[1 - K_n^*(t)\right] \mathrm{d}t \right|$$

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

$$\sup_{|t|\leq T_n} \left|\hat{\phi}_x(t) - \phi_x(t)\right| \frac{1}{2\pi} \int_{-T_n}^{T_n} \frac{1}{|\phi_{\varepsilon}(t)|} \mathrm{d}t \tag{12}$$

Hence (12) is a.s. bounded by

$$O\left(\frac{T_n}{\inf_{|t|\leq T_n} |\phi_{\varepsilon}(t)|} \left(\frac{\log n}{n}\right)^{\frac{1}{2}-\gamma-\eta} \right) = O\left(T_n^{k+2} \left(\frac{\log n}{n}\right)^{\frac{1}{2}-\gamma-\eta} \right) \tag{13}$$
$$= O\left(\left(\frac{\log n}{n}\right)^{\frac{1}{2}-(k+3)\gamma-\eta} \right) \tag{14}$$

where $T_n = O\left(\left(\frac{n}{\log n}\right)^{\gamma}\right)$ and $0 < \gamma < \frac{1}{2}$. The first equality is due to Theorem 1 and the assumption that the distribution of ε is range restricted of order k.

Next we consider the second term on the right-hand side of (11) 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^*

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

where $\tilde{x}^*$ is between x^* and $x^* - \frac{z}{T_n}$. The last equality is due to the assumption that the density of x^* is q times differentiable and the q-th derivative is continuous and bounded on $\mathcal{X}^*$. Therefore, the bias term is $O\left(T_n^{-q}\right)$. The results follow. □

Proof of Corollary 2

In (13) we substitute the lower bound of Corollary 1 (and $k = 0$) to obtain the result. □

Proof of Theorem 3

Sufficient for weak consistency of the estimator is that

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

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

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

$$m_n(\theta) - m(\theta,\theta_0) = \frac{1}{n}\sum_{i=1}^{n}\left(w(x_i)y_i - \mathrm{E}_x\left[w(x)\int_{\mathcal{X}^*} h(x^*,\theta_0)g(x^*|x)\mathrm{d}x^*\right]\right) -$$

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

$$-\left(\frac{1}{n}\sum_{i=1}^{n} w(x_i)\int_{\mathcal{X}^*} h(x^*,\theta)g(x^*|x_i)\mathrm{d}x^* - \mathrm{E}_x\left[w(x)\int_{\mathcal{X}^*} h(x^*,\theta)g(x^*|x)\mathrm{d}x^*\right]\right) \equiv A_1 + A_2 + A_3$$

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

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

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

We now consider A_2

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

By Theorem 2 $\sup_{x^*\in\mathcal{X}^*}|\hat{g}_{x^*}(x^*)-g_{x^*}(x^*)|$ is $o_p(1)$. Since g_x is bounded away from zero on $\mathcal{X}$ and the functions w, h are bounded on $\mathcal{X}$ and $\mathcal{X}^*\times\Theta$ respectively, the second term is bounded by $C\int_{-\infty}^{\infty} g_\varepsilon(\varepsilon)d\varepsilon$ for a constant $0\le C<\infty$. Therefore, we have

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

$\square$.

Proof of Theorem 4

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

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

$$-\frac{1}{n}\sum_{i=1}^{n} w(x_i)\int_{\mathcal{X}^*}\frac{\partial h(x^*,\overline{\theta})}{\partial\theta'}\frac{g_\varepsilon(x_i-x^*)}{g_x(x_i)}\hat{g}_{x^*}(x^*)\mathrm{d}x^*\sqrt{n}(\hat{\theta}-\theta_0)=0$$

Define

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

and

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

We consider first B_2 and we show that

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

We have

$$B_2 - \mathrm{E}\left[w(x)\int_{\mathcal{X}^*}\frac{\partial h(x^*,\theta_0)}{\partial\theta'}g(x^*|x)\mathrm{d}x^*\right] =$$

$$= \frac{1}{n}\sum_{i=1}^n w(x_i)\int_{\mathcal{X}^*}\left(\frac{\partial h(x^*,\overline{\theta})}{\partial\theta'} - \frac{\partial h(x^*,\theta_0)}{\partial\theta'}\right)\frac{g_\varepsilon(x_i - x^*)}{g_x(x_i)}\hat{g}_{x^*}(x^*)\mathrm{d}x^* +$$

$$+\frac{1}{n}\sum_{i=1}^n w(x_i)\int_{\mathcal{X}^*}\frac{\partial h(x^*,\theta_0)}{\partial\theta'}\frac{g_\varepsilon(x_i - x^*)}{g_x(x_i)}\left(\hat{g}_{x^*}(x^*) - g_{x^*}(x^*)\right)\mathrm{d}x^* +$$

$$+\frac{1}{n}\sum_{i=1}^n w(x_i)\int_{\mathcal{X}^*}\frac{\partial h(x^*,\theta_0)}{\partial\theta'}g(x^*|x_i)\mathrm{d}x^* - \mathrm{E}\left[w(x)\int_{\mathcal{X}^*}\frac{\partial h(x^*,\theta_0)}{\partial\theta'}g(x^*|x)\mathrm{d}x^*\right] \equiv C_1 + C_2 + C_3$$

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

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

Because $\overline{\theta} \xrightarrow{p} \theta_0$ we have that $C_1 \xrightarrow{p} 0$ by a uniform law of large numbers. 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

$$B_1 = \frac{1}{\sqrt{n}}\sum_{i=1}^n w(x_i)\left(y_i - \int_{\mathcal{X}^*} h(x^*,\theta_0)g(x^*|x_i)\mathrm{d}x^*\right) -$$

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

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

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

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

where

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

In the proof of Theorem 2 we showed that

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

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

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

so that E_2 is bounded by

$$C\sqrt{n}T_n^{-q} \tag{15}$$

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

$$E_1 = \frac{1}{n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n}\int_{-\infty}^{\infty}\frac{c^*(x_i,t,\theta_0)}{\phi_\varepsilon(t)}(e^{itx_j}-\phi_x(t))K_n^*(t)\mathrm{d}t \tag{16}$$

with

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

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)

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

with variance

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

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

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

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

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

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

Then

$$\left|\mathrm{E}_{\tilde{x}}\left[\int_{-\infty}^{\infty} \frac{c^*(\tilde{x}, t, \theta_0)}{\phi_\varepsilon(t)}(e^{itx} - \phi_x(t))K_n^*(t)\mathrm{d}t\right]\right| = \tag{17}$$

$$= \left|\mathrm{E}_{\tilde{x}}\left[G(\tilde{x})\int_{-\infty}^{\infty} \frac{\kappa^*(t, \tilde{x})}{\phi_\varepsilon(t)}e^{-it\tilde{x}}(e^{itx} - \phi_x(t))K_n^*(t)\mathrm{d}t\right]\right| \leq$$

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

We consider the final term on the right-hand side first. Consider

$$\left|\int_{-\infty}^{\infty} \frac{\kappa^*(t, \tilde{x})}{\phi_\varepsilon(t)}e^{-it\tilde{x}}\phi_x(t)K_n^*(t)\mathrm{d}t\right| \leq \int_{-T_n}^{T_n} |\kappa^*(t, \tilde{x})|\,|\phi_{x^*}(t)|\mathrm{d}t$$

Now

$$|\kappa^*(t, x)| \leq \int_L^U |h(x - \varepsilon, \theta_0)|g_\varepsilon(\varepsilon)\mathrm{d}\varepsilon \tag{18}$$

which is bounded because $h(x^*, \theta_0)$ is bounded on $\mathcal{X}^*$.

The final step is to show that the first term on the right-hand side of (17) is bounded on $\mathcal{X}$. Define the function (of t)

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

We take 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.

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

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

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

This suffices if the distribution of ε is range-restricted of order 0. 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

$$A' = e^{it\varepsilon}\kappa(\varepsilon,x)|_L^U - \frac{1}{it}\left(e^{it\varepsilon}\kappa'(\varepsilon,x)|_L^U\right)$$

$$B' = e^{it\varepsilon}g_\varepsilon(\varepsilon)|_L^U - \frac{1}{it}\left(e^{it\varepsilon}g_\varepsilon'(\varepsilon)|_L^U\right)$$

we have

$$r(t,x) = \frac{A'}{B'} + \left[\frac{1}{it\phi_\varepsilon(t)}\left(e^{it\varepsilon}\kappa''(\varepsilon,x)|_L^U - \int_L^U e^{it\varepsilon}\kappa'''(\varepsilon,x)\mathrm{d}\varepsilon\right) - \right.$$
$$\left. - \frac{1}{it\phi_\varepsilon(t)}\frac{A'}{B'}\left(e^{it\varepsilon}g_\varepsilon''(\varepsilon)|_L^U - \int_L^U e^{it\varepsilon}g_\varepsilon'''(\varepsilon)\mathrm{d}\varepsilon\right)\right]\frac{1}{(it)^2}$$

The next step is to use the identity

$$\frac{A'}{B'} = \frac{A'' + (A' - A'')}{B'' + (B' - B'')} = \frac{A''}{B''} + \frac{1}{B''}(A' - A'') - \frac{A''}{(B'')^2}(B' - B'') + \frac{A''}{B'\,(B'')^2}(B' - B'')^2 -$$
$$- \frac{1}{B''B'}(A' - A'')(B' - B'')$$

with

$$A'' = e^{it\varepsilon}\kappa(\varepsilon, x)|_L^U$$

$$B'' = e^{it\varepsilon}g_\varepsilon(\varepsilon)|_L^U$$

to write

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

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

Substitution gives the following expansion

$$r(t,x) = \kappa_1(t,x) + \kappa_2(t,x)\frac{1}{it} + \kappa_3(t,x)\frac{1}{(it)^2} \tag{19}$$

with

$$\kappa_1(t,x) = \frac{e^{itU}\kappa(U,x) - e^{itL}\kappa(L,x)}{e^{itU}g_\varepsilon(U) - e^{itL}g_\varepsilon(L)} \tag{20}$$

$$\kappa_2(t,x) = \frac{\left(e^{itU}\kappa(U,x) - e^{itL}\kappa(L,x)\right)\left(e^{itU}g'_\varepsilon(U) - e^{itL}g'_\varepsilon(L)\right)}{\left(e^{itU}g_\varepsilon(U) - e^{itL}g_\varepsilon(L)\right)^2} - \frac{e^{itU}\kappa'(U,x) - e^{itL}\kappa'(L,x)}{e^{itU}g_\varepsilon(U) - e^{itL}g_\varepsilon(L)} \tag{21}$$

$$\kappa_3(t,x) = \frac{1}{it\phi_\varepsilon(t)}\left(e^{itU}\kappa''(U,x) - e^{itL}\kappa''(L,x) - \int_L^U e^{it\varepsilon}\kappa'''(\varepsilon,x)\mathrm{d}\varepsilon\right) - \tag{22}$$

$$-\frac{1}{it\phi_\varepsilon(t)}\frac{e^{itU}\kappa(U,x) - e^{itL}\kappa(L,x) - \frac{1}{it}\left(e^{itU}\kappa'(U,x) - e^{itL}\kappa'(L,x)\right)}{e^{itU}g_\varepsilon(U) - e^{itL}g_\varepsilon(L) - \frac{1}{it}\left(e^{itU}g'_\varepsilon(U) - e^{itL}g'_\varepsilon(L)\right)}.$$

$$.\left(e^{itU}g''_\varepsilon(U) - e^{itL}g''_\varepsilon(L) - \int_L^U e^{it\varepsilon}g'''_\varepsilon(\varepsilon)\mathrm{d}\varepsilon\right)+$$

$$+\frac{e^{itU}\kappa(U,x) - e^{itL}\kappa(L,x)}{e^{itU}g_\varepsilon(U) - e^{itL}g_\varepsilon(L) - \frac{1}{it}\left(e^{itU}g'_\varepsilon(U) - e^{itL}g'_\varepsilon(L)\right)}.\frac{\left(e^{itU}g'_\varepsilon(U) - e^{itL}g'_\varepsilon(L)\right)^2}{\left(e^{itU}g_\varepsilon(U) - e^{itL}g_\varepsilon(L)\right)^2}-$$

$$-\frac{\left(e^{itU}g'_\varepsilon(U) - e^{itL}g'_\varepsilon(L)\right)\left(e^{itU}\kappa'(U,x) - e^{itL}\kappa'(L,x)\right)}{\left(e^{itU}g_\varepsilon(U) - e^{itL}g_\varepsilon(L)\right)\left(e^{itU}g_\varepsilon(U) - e^{itL}g_\varepsilon(L) - \frac{1}{it}\left(e^{itU}g'_\varepsilon(U) - e^{itL}g'_\varepsilon(L)\right)\right)}$$

Note that κ_1 and κ_2 are well-defined because

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

If t is sufficiently large, say $|t| \geq t_0$, then by the same argument

$$\left| e^{itU} g_\varepsilon(U) - e^{itL} g_\varepsilon(L) - \frac{1}{it} \left(e^{itU} g'_\varepsilon(U) - e^{itL} g'_\varepsilon(L) \right) \right| \geq$$

$$\geq \left| \left| e^{itU} g_\varepsilon(U) - e^{itL} g_\varepsilon(L) \right| - \frac{1}{|t|} \left| \left(e^{itU} g'_\varepsilon(U) - e^{itL} g'_\varepsilon(L) \right) \right| \right| > 0$$

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

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

on $\mathcal{X}$ and for $|t| \geq t_0$.

(ii) Finiteness of the integral

We consider

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

We have

$$\left| \mathrm{E}_{\tilde{x}} \left[G(\tilde{x}) \int_{-\infty}^{\infty} \frac{\kappa^*(t, \tilde{x})}{\phi_\varepsilon(t)} e^{it(x-\tilde{x})} K_n^*(t) \mathrm{d}t \right] \right| \leq \mathrm{E}_{\tilde{x}} \left[|G(\tilde{x})| \int_{0 \leq |t| \leq t_0} \left| \frac{\kappa^*(t, \tilde{x})}{\phi_\varepsilon(t)} \right| |K_n^*(t)| \mathrm{d}t \right] +$$

$$+ \left| \mathrm{E}_{\tilde{x}} \left[G(\tilde{x}) \int_{t_0 \leq |t| < \infty} \frac{\kappa^*(t, \tilde{x})}{\phi_\varepsilon(t)} e^{it(x-\tilde{x})} K_n^*(t) \mathrm{d}t \right] \right|$$

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

$$\begin{aligned} \left| \mathrm{E}_{\tilde{x}} \left[G(\tilde{x}) \int_{t_0 \leq |t| < \infty} \frac{\kappa^*(t, \tilde{x})}{\phi_\varepsilon(t)} e^{it(x-\tilde{x})} K_n^*(t) \mathrm{d}t \right] \right| &\leq \left| \mathrm{E}_{\tilde{x}} \left[G(\tilde{x}) \int_{t_0 \leq |t| < \infty} \kappa_1(t, \tilde{x}) e^{it(x-\tilde{x})} K_n^*(t) \mathrm{d}t \right] \right| + \\ &+ \left| \mathrm{E}_{\tilde{x}} \left[G(\tilde{x}) \int_{t_0 \leq |t| < \infty} \frac{1}{it} \kappa_2(t, \tilde{x}) e^{it(x-\tilde{x})} K_n^*(t) \mathrm{d}t \right] \right| + \\ &+ \left| \mathrm{E}_{\tilde{x}} \left[G(\tilde{x}) \int_{t_0 \leq |t| < \infty} \frac{1}{(it)^2} \kappa_3(t, \tilde{x}) e^{it(x-\tilde{x})} K_n^*(t) \mathrm{d}t \right] \right| \end{aligned} \tag{23}$$

The final term is bounded by

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

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

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

and substitution of this expression in the relevant term in (23) gives

$$\left|\mathrm{E}_{\tilde{x}}\left[G(\tilde{x})\int_{t_0\leq|t|<\infty}\frac{h(\tilde{x}-U,\theta_0)-e^{it(L-U)}\frac{g_\varepsilon(L)}{g_\varepsilon(U)}h(\tilde{x}-L,\theta_0)}{1-e^{it(L-U)}\frac{g_\varepsilon(L)}{g_\varepsilon(U)}}e^{it(x-\tilde{x})}K_n^*(t)\mathrm{d}t\right]\right|\leq \tag{24}$$

$$\leq\left|\mathrm{E}_{\tilde{x}}\left[G(\tilde{x})h(\tilde{x}-U,\theta_0)\int_{t_0\leq|t|<\infty}\frac{1}{1-e^{it(L-U)}\frac{g_\varepsilon(L)}{g_\varepsilon(U)}}e^{it(x-\tilde{x})}K_n^*(t)\mathrm{d}t\right]\right|+$$

$$+\left|\mathrm{E}_{\tilde{x}}\left[G(\tilde{x})h(\tilde{x}-L,\theta_0)\int_{t_0\leq|t|<\infty}\frac{e^{it(L-U)}\frac{g_\varepsilon(L)}{g_\varepsilon(U)}}{1-e^{it(L-U)}\frac{g_\varepsilon(L)}{g_\varepsilon(U)}}e^{it(x-\tilde{x})}K_n^*(t)\mathrm{d}t\right]\right|$$

without loss of generality we assume that

$$\frac{g_\varepsilon(L)}{g_\varepsilon(U)}<1$$

Now consider the first term on the right-hand side of (24) that is bounded by

$$\left|\mathrm{E}_{\tilde{x}}\left[G(\tilde{x})h(\tilde{x}-U,\theta_0)\int_{0\leq|t|\leq t_0}\frac{1}{1-e^{it(L-U)}\frac{g_\varepsilon(L)}{g_\varepsilon(U)}}e^{it(x-\tilde{x})}K_n^*(t)\mathrm{d}t\right]\right|+ \tag{25}$$

$$+\left|\mathrm{E}_{\tilde{x}}\left[G(\tilde{x})h(\tilde{x}-U,\theta_0)\int_{-\infty}^{\infty}\frac{1}{1-e^{it(L-U)}\frac{g_\varepsilon(L)}{g_\varepsilon(U)}}e^{it(x-\tilde{x})}K_n^*(t)\mathrm{d}t\right]\right|$$

Because

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

the first term of (25) is bounded by (if $T_n > t_0$)

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

because $K^*(t) = 1$ if $|t| \le 1$.

For the second term of (25) we note that

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

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 $j = 0, 1, \ldots$. Hence

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

Because the density corresponding to $K_n^*(t)$ is $T_nK(T_nv)$, this is equal to

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

Hence

$$\left|\mathrm{E}_{\tilde{x}}\left[G(\tilde{x})h(\tilde{x}-U,\theta_0)\int_{-\infty}^{\infty}\frac{1}{1-e^{it(L-U)}\frac{g_\varepsilon(L)}{g_\varepsilon(U)}}e^{it(x-\tilde{x})}K_n^*(t)\mathrm{d}t\right]\right| \le$$

$$\le C\sum_{j=0}^{\infty}\left(\frac{g_\varepsilon(L)}{g_\varepsilon(U)}\right)^j \left|\mathrm{E}_{\tilde{x}}\left[G(\tilde{x})h(\tilde{x}-U,\theta_0)T_nK(T_n(\tilde{x}-x-(L-U)j))\right]\right|$$

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

$$\left|\mathrm{E}_{\tilde{x}}\left[G(\tilde{x})h(\tilde{x}-U,\theta_0)T_nK(T_n(\tilde{x}-x-(L-U)j))\right]\right| =$$

$$= \left|\int G\left(\frac{w}{T_n}+x+(L-U)j\right)h\left(\frac{w}{T_n}+x+(L-U)j-U,\theta_0\right).\right.$$

$$.K(w)g_x\left(\frac{w}{T_n}+x+(L-U)j\right)\mathrm{d}w\Bigg| \leq$$

$$\leq C\int_{-\infty}^{\infty}|K(w)|\mathrm{d}w<\infty$$

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 (24) is finite.

Finally, we consider the second term on the right-hand side of (23). First, we have

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

$$-\frac{\frac{h(x-U,\theta_0)g'_\varepsilon(U)-e^{it(L-U)}h(x-L,\theta_0)g'_\varepsilon(L)}{g_\varepsilon(U)}}{1-e^{it(L-U)}\frac{g_\varepsilon(L)}{g_\varepsilon(U)}}+\frac{h'(x-U,\theta_0)-e^{it(L-U)}\frac{g_\varepsilon(L)}{g_\varepsilon(U)}h'(x-L,\theta_0)}{1-e^{it(L-U)}\frac{g_\varepsilon(L)}{g_\varepsilon(U)}}$$

Substitution gives the bound

$$\Bigg|\mathrm{E}_{\tilde{x}}\Bigg[G(\tilde{x})\int_{t_0\leq|t|<\infty}\frac{\left(h(\tilde{x}-U,\theta_0)-e^{it(L-U)}\frac{g_\varepsilon(L)}{g_\varepsilon(U)}h(\tilde{x}-L,\theta_0)\right)\left(\frac{g'_\varepsilon(U)-e^{it(L-U)}g'_\varepsilon(L)}{g_\varepsilon(U)}\right)}{\left(1-e^{it(L-U)}\frac{g_\varepsilon(L)}{g_\varepsilon(U)}\right)^2}$$

$$\frac{e^{it(x-\tilde{x})}}{it}K_n^*(t)\mathrm{d}t\Bigg]\Bigg|+$$

$$+\Bigg|\mathrm{E}_{\tilde{x}}\Bigg[G(\tilde{x})\int_{t_0\leq|t|<\infty}\frac{\frac{h(\tilde{x}-U,\theta_0)g'_\varepsilon(U)-e^{it(L-U)}h(\tilde{x}-L,\theta_0)g'_\varepsilon(L)}{g_\varepsilon(U)}}{1-e^{it(L-U)}\frac{g_\varepsilon(L)}{g_\varepsilon(U)}}\frac{e^{it(x-\tilde{x})}}{it}K_n^*(t)\mathrm{d}t\Bigg]\Bigg|+$$

$$+\Bigg|\mathrm{E}_{\tilde{x}}\Bigg[G(\tilde{x})\int_{t_0\leq|t|<\infty}\frac{h'(\tilde{x}-U,\theta_0)-e^{it(L-U)}\frac{g_\varepsilon(L)}{g_\varepsilon(U)}h'(\tilde{x}-L,\theta_0)}{1-e^{it(L-U)}\frac{g_\varepsilon(L)}{g_\varepsilon(U)}}\frac{e^{it(x-\tilde{x})}}{it}K_n^*(t)\mathrm{d}t\Bigg]\Bigg| \tag{26}$$

We show that the final term of (26) is bounded (in x). It is bounded by

$$\left| \mathrm{E}_{\tilde{x}} \left[G(\tilde{x}) h'(\tilde{x}-U,\theta_0) \int_{t_0 \le |t| < \infty} \frac{1}{1-e^{it(L-U)}\frac{g_\varepsilon(L)}{g_\varepsilon(U)}} \frac{e^{it(x-\tilde{x})}}{it} K_n^*(t) \mathrm{d}t \right] \right| + \quad (27)$$

$$+ \left| \mathrm{E}_{\tilde{x}} \left[G(\tilde{x}) \frac{g_\varepsilon(L)}{g_\varepsilon(U)} h'(x-L,\theta_0) \int_{t_0 \le |t| < \infty} \frac{1}{1-e^{it(L-U)}\frac{g_\varepsilon(L)}{g_\varepsilon(U)}} \frac{e^{it(x-\tilde{x})+it(L-U)}}{it} K_n^*(t) \mathrm{d}t \right] \right|$$

The first term of (27) is bounded by

$$\left| \mathrm{E}_{\tilde{x}} \left[G(\tilde{x}) h'(\tilde{x}-U,\theta_0) \int_{|t| \le t_0} \frac{1}{1-e^{it(L-U)}\frac{g_\varepsilon(L)}{g_\varepsilon(U)}} \frac{e^{it(x-\tilde{x})}}{it} K_n^*(t) \mathrm{d}t \right] \right| + \quad (28)$$

$$+ \left| \mathrm{E}_{\tilde{x}} \left[G(\tilde{x}) h'(\tilde{x}-U,\theta_0) \int_{-\infty}^{\infty} \frac{1}{1-e^{it(L-U)}\frac{g_\varepsilon(L)}{g_\varepsilon(U)}} \frac{e^{it(x-\tilde{x})}}{it} K_n^*(t) \mathrm{d}t \right] \right|$$

The second term of (28) contains the integral

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

with derivative

$$s'(x-\tilde{x}) = \int_{-\infty}^{\infty} \frac{1}{1-e^{it(L-U)}\frac{g_\varepsilon(L)}{g_\varepsilon(U)}} e^{it(x-\tilde{x})} K_n^*(t) \mathrm{d}t =$$

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

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 $H(v) = \int_{-\infty}^{v} K(s)\mathrm{d}s$ the integral of K which is a bounded function. Hence, the second term on the right-hand side of (28) is bounded by

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

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

To bound the first term on the right-hand side of (28) we note that $K^*(t) = \int_{-\infty}^{\infty} e^{itv} K(v)\mathrm{d}v = \int_{-\infty}^{\infty} \cos(tv)K(v)\mathrm{d}v$ 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 $|t| \leq T_n$

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

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

Now

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

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

The proof that the second term of (27) is bounded is completely analogous. The same method of proof also applies to the second term on the right-hand side of (26). For the first term of (26) 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 $j = 0, 1, \ldots$ 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 (26). We conclude that the variance is indeed finite. □

Proof of Corollary 4

In the proof of Corollary 1 we established that for the sequences in that corollary the c.f. of ε_n obtained by truncating a distribution with unbounded support at L_n and U_n satisfies

$$|\phi_{\varepsilon_n}(T_n)| \geq C\frac{g_{\varepsilon_n}(U_n)}{T_n}.$$

We consider

$$\sqrt{n}g_{\varepsilon_n}(U_n)(\hat{\theta}-\theta)$$

so that in the proof of Theorem 4 we multiply by $g_{\varepsilon_n}(U_n)$ throughout.

The key problem in establishing that the variance of INLR is finite is with the first term on the right hand side of (17). This term is bounded by (we multiply by $g_{\varepsilon_n}(U_n)$)

$$\mathbb{E}_{\tilde{x}}\left[|G(\tilde{x})|\int_{-T_n}^{T_n}\frac{|\kappa^*(t,\tilde{x})|}{|\phi_{\varepsilon_n}(t)|}\mathrm{d}t\right]$$

If ε_n has a truncated normal distribution, then if the truncation points diverge, the denominator behaves as $e^{-\frac{1}{2}t^2}$ and even if the numerator has many absolutely integrable derivatives the decrease in the numerator is at most proportional to t^{-k}. Now if the first three derivatives of $\kappa(\varepsilon,x)$ are absolutely integrable, i.e.

$$\int_{-\infty}^{\infty}|\kappa^{(k)}(\varepsilon,x)|\mathrm{d}\varepsilon<\infty \qquad k=0,\ldots,3$$

then[5]

$$\left|\int_{-\infty}^{\infty}e^{it\varepsilon}\kappa(\varepsilon,x)\mathrm{d}\varepsilon\right|\leq\frac{C}{t^3}$$

so that for sufficiently large n

$$|\kappa^*(t,x)|=\left|\int_{L_n}^{U_n}e^{it\varepsilon}\kappa(\varepsilon,x)\mathrm{d}\varepsilon\right|\leq\frac{2C}{t^3}$$

[5]Note that this upper bound holds under infinite support. The more complicated proof of Theorem 4 is needed, because there the support is bounded.

This combined with the lower bound on ϕ_{ε_n} given above, implies that the integrand (after multiplication by $g_{\varepsilon_n}(U_n)$ behaves as t^{-2} so that the integral is bounded. Note that because we multiply by $g_{\varepsilon_n}(U_n)$, the second term on the right hand side of (17) has an asymptotically negligible contribution to the variance. $\square$

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.

[11] Taupin, M. L., 2001, "Semi-parametric estimation in the nonlinear structural errors-in-variables model," Annals of Statistics, 29, pp. 66-93.

Table 1: Simulation results with Probit model: n=500, n1=600, reps=100.

	$\beta_1 = 1$			$\beta_2 = -1$			$\beta_0 = 0.5$		
$C_{trunc} = 1$	Root MSE	Mean	Std. dev.	Root MSE	Mean	Std. dev.	Root MSE	Mean	Std. dev.
Ignoring meas. error	0.5321	0.4770	0.0976	0.1274	-0.9523	0.1182	0.0669	0.4923	0.0664
True x^*	0.1300	1.0098	0.1296	0.1154	-1.0073	0.1151	0.0669	0.5175	0.0646
Deconvolution	0.3044	1.0597	0.2985	0.1504	-1.0206	0.1490	0.0850	0.5320	0.0787
$C_{trunc} = 1.2$	Root MSE	Mean	Std. dev.	Root MSE	Mean	Std. dev.	Root MSE	Mean	Std. dev.
Ignoring meas. error	0.5227	0.4871	0.1007	0.1399	-0.9576	0.1334	0.0725	0.4900	0.0718
True x^*	0.1393	1.0279	0.1365	0.1338	-1.0190	0.1324	0.0750	0.5190	0.0725
Deconvolution	0.2862	1.0539	0.2811	0.1776	-1.0340	0.1743	0.0929	0.5283	0.0885
$C_{trunc} = 4$	Root MSE	Mean	Std. dev.	Root MSE	Mean	Std. dev.	Root MSE	Mean	Std. dev.
Ignoring meas. error	0.5412	0.4655	0.0849	0.1474	-0.9580	0.1413	0.0702	0.4819	0.0678
True x^*	0.1414	0.9810	0.1401	0.1425	-1.0126	0.1419	0.0690	0.5103	0.0682
Deconvolution	0.2359	1.0067	0.2358	0.1800	-1.0343	0.1767	0.0859	0.5137	0.0848

Table 2: Simulation results with Probit model: n=500, n1=600, reps=100.

90% confi. interval	$\beta_1 = 1$			$\beta_2 = -1$			$\beta_0 = 0.5$		
$C_{trunc} = 1$	Mean	5th quant	95th quant	Mean	5th quant	95th quant	Mean	5th quant	95th quant
Ignoring meas. error	0.4770	0.3415	0.6347	-0.9523	-1.1503	-0.8012	0.4923	0.4029	0.6102
True x^*	1.0098	0.7921	1.2146	-1.0073	-1.2015	-0.8354	0.5175	0.4245	0.6288
Deconvolution	1.0597	0.6574	1.6053	-1.0206	-1.3003	-0.8277	0.5320	0.4290	0.6645
$C_{trunc} = 1.2$	Mean	5th quant	95th quant	Mean	5th quant	95th quant	Mean	5th quant	95th quant
Ignoring meas. error	0.4871	0.3239	0.6403	-0.9576	-1.2062	-0.7361	0.4900	0.3753	0.6079
True x^*	1.0279	0.8137	1.2628	-1.0190	-1.2642	-0.8236	0.5190	0.4054	0.6310
Deconvolution	1.0539	0.6095	1.5066	-1.0340	-1.3144	-0.7681	0.5283	0.4026	0.6963
$C_{trunc} = 4$	Mean	5th quant	95th quant	Mean	5th quant	95th quant	Mean	5th quant	95th quant
Ignoring meas. error	0.4655	0.3109	0.6168	-0.9580	-1.1946	-0.7629	0.4819	0.3801	0.5975
True x^*	0.9810	0.7855	1.2255	-1.0126	-1.2571	-0.8002	0.5103	0.4154	0.6189
Deconvolution	1.0067	0.6411	1.3323	-1.0343	-1.3389	-0.7949	0.5137	0.4033	0.6720

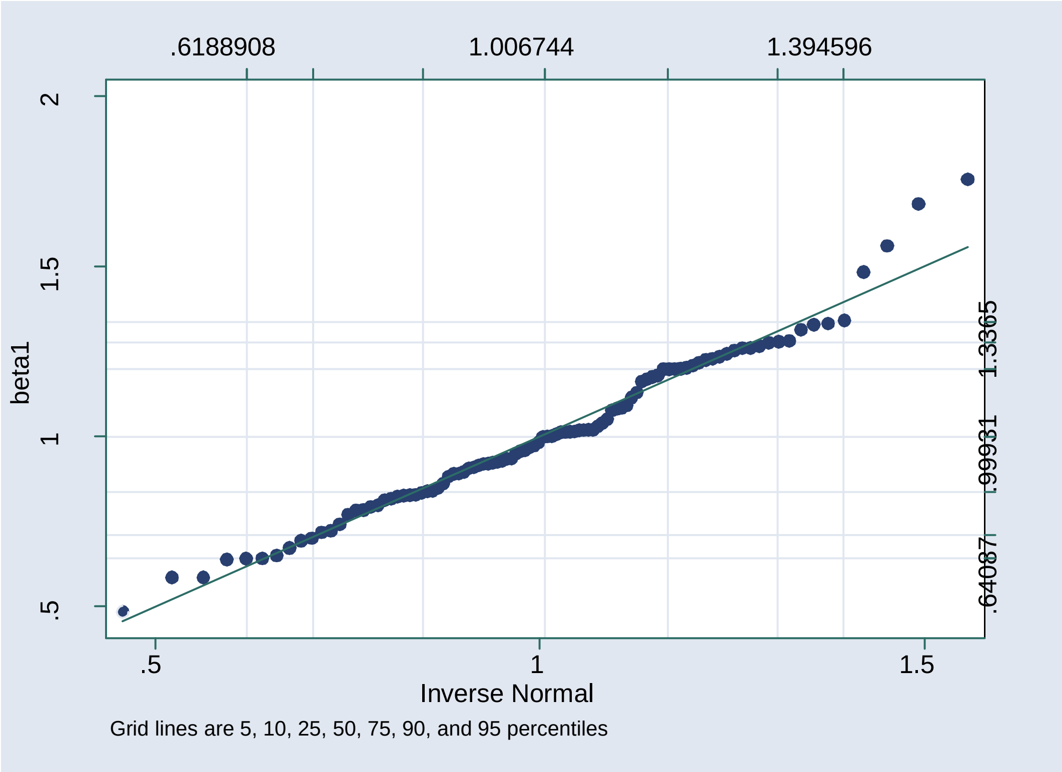

Figure 1: normal probability plot of 100 deconvolution estimates $\widehat{\beta}_1$ with $C_{trunc} = 4$

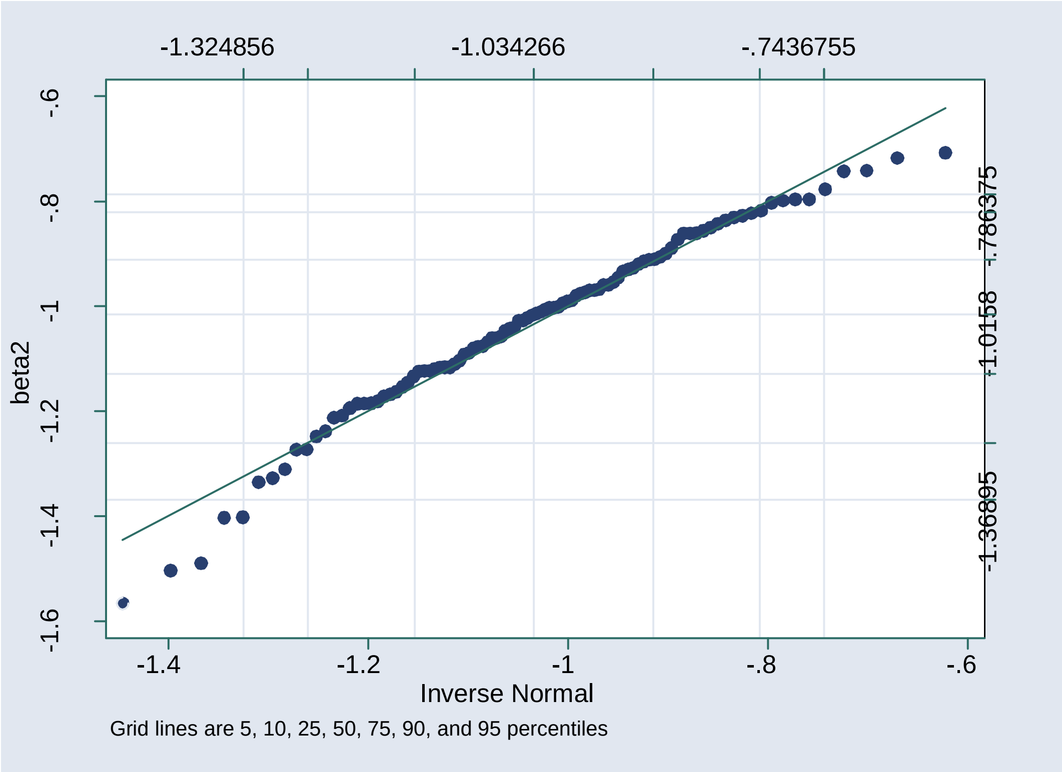

Figure 2: normal probability plot of 100 deconvolution estimates $\widehat{\beta}_2$ with $C_{trunc} = 4$