Economics Department

Working Papers in Economics

Boston College *Year* 2002

Weighted and Two Stage Least Squares Estimation of Semiparametric Truncated Regression Models

Shakeeb Khan
University of Rochester,

Arthur Lewbel
Boston College,

WEIGHTED AND TWO STAGE LEAST SQUARES ESTIMATION OF SEMIPARAMETRIC TRUNCATED REGRESSION MODELS

Shakeeb Khan
University of Rochester
Arthur Lewbel*
Boston College

February 2002

Abstract

This paper provides a root-n consistent, asymptotically normal weighted least squares estimator of the coefficients in the truncated regression model. The distribution of the errors is unknown and permits general forms of unknown heteroscedasticity. Also provided is an instrumental variables based two stage least squares estimator for this model, which can be used when the errors are correlated with some regressors. Estimation is based on a "special" regressor as in Lewbel(2000). Our limiting distributions include a new result regarding asymptotic trimming for root-n convegence of density weighed extemeum estimators.

JEL Classification: C14, C25, C13.

Key Words: Semiparametric, Truncated Regression, Heteroscedasticity, Latent Variable Models, Endogenous Regressors.

*Corresponding author. Department of Economics, Boston College, 140 Commonwealth Ave., Chestnut Hill, MA, 02467, USA. (781)-552-3678, lewbel@bc.edu. This research was supported in part by the National Science Foundation through grant SBR-9514977 to A. Lewbel. The authors wish to thank Dan McFadden, Jim Powell, Richard Blundell, Bo Honoré, Jim Heckman, Xiaohong Chen and Songnian Chen for helpful comments. Any errors are our own.

1 Introduction

The truncated regression model has received a great deal of attention in the econometrics literature, as many data sets in economics exhibit some form of truncation. One common model of truncation is

$$\begin{aligned} y_i &= w_i'\theta_0 + \widetilde{e}_i \\ \widetilde{e}_i &= e_i | e_i > -w_i'\theta_0 \end{aligned}$$

where y_i denotes the observed dependent variable, w_i denotes a $k+1$ dimensional vector of observed covariates, the random variable e_i denotes an an unobserved disturbance term, and $\widetilde{e}_i$ denotes the unobserved random variable that is equal to e_i, conditional on it exceeding $w_i'\theta_0$. Thus the $k+2$ dimensional random vector $(y_i, w_i')'$ is only observed when $y_i > 0$. The $k+1$ dimensional vector θ_0 is the parameter of interest to be identified and estimated under restrictions on e_i.

Parametric approaches restrict e_i to be distributed independently of x_i and lie in a parametic family, so that β_0 and nuisance parameters in the distribution of e_i can be estimated by MLE or (nonlinear) least squares. The drawback of this approach is that these estimators are generally inconsistent if the distribution of e_i is misspecified, correlated with w_i, or if conditional heteroskedasticity is present.

Semiparametric, or "distribution-free" estimators have also been proposed with various restrictions on e_i. These include papers by Powell(1986a,b), Lee(1989,1993), and Honoré and Powell(1994). With the exception of Lee(1989), which converges at the slow rate of the cube root of the sample size, these estimators converge as quickly as the parametric estimators, and have asymptotic normal distributions.

In this paper, a new estimator for the truncated regression model is proposed. The estimator is numerically simple, being equivalent to a linear weighted least squares, though the weights depend on a nonparametrically estimated (plug in) density. The error distribution is assumed to be unknown, and permits very general forms of heteroskedasticity, including forms not permitted by other semiparametric estimators. The estimator may also be applied to doubly truncated data, that is, to models where $(y_i, w_i')'$ is only observed when $0 < y_i < K$ for some known K.

A two stage least squares based estimator for the truncated model is also proposed. Given instruments z_i that are uncorrelated with the latent errors e_i, this estimator permits estimation of coefficients when these errors are correlated with the regressors (as would arise in models with endogenous or mismeasured regressors), analogous to a standard linear model two stage least squares regression. This is in contrast to the semiparametric approaches referred to above, which do not allow for any form of endogeneity.

The estimators depend on the density of a "special" regressor, as in Lewbel(2000). We provide the limiting distribution for a general class of special regressor estimators, namely, density weighted extremum estimators. Examples of this class are the proposed truncated weighted and two stage estimators. This limiting distribution theory includes a new result on the use of asymptotic trimming to deal with issues regarding boundary bias in first stage density estimation.

This paper's estimators treat one regressor as special. We decompose the regressor vector as $w_i = (v_i, x_i')'$ with v_i denoting the special regressor, and x_i denoting the k-dimensional vector of other regressors. We decompose the parameter vector as $\theta_0 = (\alpha_0, \beta_0')'$. With this notation the truncated regression model is

$$\begin{aligned} y_i &= v_i\alpha_0 + x_i'\beta_0 + \widetilde{e}_i \\ \widetilde{e}_i &= e_i | e_i > -v_i\alpha_0 - x_i'\beta_0 \end{aligned}$$

There may also be a vector of instruments z_i that are uncorrelated with e_i.The data are n observations of y_i, v_i, x_i, and possibly z_i. A useful equivalent representation is

$$\begin{aligned} \widetilde{y}_i &= v_i\alpha_0 + x_i'\beta_0 + e_i \\ y_i &= \widetilde{y}_i | \widetilde{y}_i > 0 \end{aligned} \tag{1.1}$$

where $\widetilde{y}_i$ is a latent variable that, were it observed without truncation, could be regressed on $(v_i, x_i')'$ using ordinary or two stage least squares to obtain $(\alpha_0, \beta_0')'$, but our data are only sampled from individuals having $\widetilde{y}_i$ positive.

The paper is organized as follows. The following section shows identification of the parameters of interest and motivates the weighted and two stage least squares estimation procedures. Section 3 provides general asymptotic results concerning functions that satisfy a density weighted moment condition.

Sections 4 and 5 apply the general results in section 3 and provide root-n normal limiting distribution of the resulting weighted least squares based parameter estimates. Sections 6 and 7 then similarly apply the results to give an associated two stage least squares based estimator and establish its root-n normal limiting distribution. Section 9 concludes. Details of the proofs of the main theorems are provided in the appendix.

2 Overview of the Weighted Least Squares Estimator

Let $f(v|x)$ denote the conditional probability density function of v given an observation x, which can be estimated from the data. Given a positive constant k, and any positive bounded function $w(x)$, define the function $\mu(\alpha, \beta)$ by

$$\mu(\alpha, \beta) = E\left(\frac{(y - v\alpha - x'\beta)^2\ \alpha^{-2}\ I(0 < y < k)\ w(x)}{f(v|x)}\right) \tag{2.1}$$

We show that with the truncated y data, under general conditions summarized below,

$$(\alpha_0, \beta_0) = \arg\min \mu(\alpha, \beta) \tag{2.2}$$

Moreover, there exists a unique, closed form expression for estimates of α_0 and β_0 based on 2.2. The estimator is numerically equivalent to a linear weighted least squares regression of v on y and x, with weights given by $I(0 < y < k)w(x)/\widehat{f}(v|x)$. Also, since the data in (2.2) is artificially truncated from above by the chosen constant k, the estimator can be applied without change to doubly truncated data.

Let $F_e(e|\cdot)$ denote the conditional distribution of an observation of e given data $\cdot$. The minimal uncorrelated error assumption for linear models,

$$E(ex) = 0 \tag{2.3}$$

is not generally sufficient to identify the coefficients in the truncated regression model. An additional assumption that is made for identification and estimation is that the distribution of e be conditionally independent of the one regressor v, or equivalently,

$$F_e(e|v, x) = F_e(e|x). \tag{2.4}$$

The distribution of v will also be assumed to have a large support. Equations 2.3 and 2.4 require that the error distribution F_e not depend on v, but permit virtually any form of heteroscedasticity with respect to the vector of other regressors x.

Lewbel (2000) used virtually the same identification conditions to estimate β_0 in a binary choice model. Lewbel (1998) employed related identification conditions, but that estimator used much stronger tail conditions and required a "tuning" function.

Both of these earlier papers provide examples of models that satisfy the identification conditions 2.3 and 2.4. In particular, Equation 2.4 holds when $e = \nu(e^*, x)$ for any function ν and any random vector e^*, where the distribution of e^* is independent of (v, x). The function ν, the vector e^* and its distribution do not need to be known, observed, or estimated. This permits virtually any form of conditional heteroscedasticity involving x, but not v. For example, a standard random coefficients model for the latent variable, $\widetilde{y} = v\alpha_0 + x'(\beta_0 + e^*) + e_0^*$ with $E(e^*) = 0$ and $E(e_0^*) = 0$ equals equation 1.1 with $e = x'e^* + e_0^*$, and satisfies equations 2.3 and 2.4. Every regressor except v could have a random coefficient.

For estimation based on 2.2, if

$$E(e \mid x) = 0 \tag{2.5}$$

then $w(x)$ in 2.1 can be any positive bounded function, and so can be selected to maximize efficiency. Alternatively, if only 2.3 and not 2.5 is assumed to hold, then the estimator based on 2.1 will still work, by taking $w(x)$ to equal one.

2.1 Overview of the Two Stage Least Squares Estimator

Suppose $E(ex) \neq 0$, as would happen when some elements of x are either endogenously determined or mismeasured. Given some instruments z, assume that $E(ez) = 0$ and that

$$F_{ex}(e, x|v, z) = F_{ex}(e, x|z) \tag{2.6}$$

where F_{ex} denotes the distribution of (e, x). Define y^* by

$$y^* = \frac{(y - v\alpha_0)\ I(0 < y < k)}{f(v|z)} / E\left(\frac{I(0 < y < k)}{f(v|z)}\right) \tag{2.7}$$

With truncated data we show that

$$E(zy^*) = E(zx')\beta_0 \tag{2.8}$$

so β_0 can be estimated by an ordinary linear two stage least squares regression of (an estimate of) y^* on x using instruments z. An estimator of α_0, which is required to estimate y^*, is also provided.

Note that equation 2.6 implies 2.4 when $z = x$. In general, 2.6 will hold if $x = \tau(z) + \varepsilon$ for any function τ and the joint distribution of (e, ε) does not depend on x or v (though it can depend on z). Lewbel (2000) uses the same identification condition to estimate a binary choice two stage least squares model, and discusses this identification condition in detail.

The next section provides a general result concerning functions that, like $E(zy^*)$ in (2.8), or like equation (2.1), are of the form

$$\psi(\theta) = E\left(\frac{h(v, x, z, e, \theta)}{f(v|z)}\right) \tag{2.9}$$

The appendix provides a root-n limiting distribution theory for estimators of $\psi(\theta)$, and for estimators of parameters θ based on minimizing an estimator of $\psi(\theta)$.

3 Special Regressor Density Weighting

Let $F_{ex}(e, x|\cdot)$ denote the joint distribution of (e, x), conditional on data $(\cdot)$, with support denoted $\Omega_{ex}(\cdot)$. Let $f(v|x)$ denote the conditional density of an observation of v given an observation of x. Let θ be a vector of parameters and let $h(v, x, z, e, \theta)$ be any function such that $\psi(\theta)$ defined by equation 2.9 exists.

Theorem 3.1 *If $F_{ex}(e, x|v, z) = F_{ex}(e, x|z)$, $\Omega_{ex}(v, z) = \Omega_{ex}(z)$, and the support of the random variable v is the interval $[L, K]$, then*

$$E\left(\frac{h(v, x, z, e, \theta)}{f(v|z)}|z\right) = E[\int_L^K h(v, x, z, e, \theta) dv|z] \tag{3.1}$$

Proof:

$$
\begin{aligned}
E\left(\frac{h}{f(v|z)}|z\right) &= E\left(\frac{E(h|v,z)}{f(v|z)}|z\right) \\
&= \int_L^K \left(\frac{E(h|v,z)}{f(v|z)} f(v|z)\right) dv \\
&= \int_L^K E(h|v,z) dv \\
&= \int_L^K \int_{\Omega_{ex}} h(v,x,z,e,\theta) dF_{ex}(e,x|z) dv \\
&= \int_{\Omega_{ex}} \int_L^K h(v,x,z,e,\theta) dv dF_{ex}(e,x|z)
\end{aligned}
$$

An immediate implication of Theorem 3.1 is

$$
\psi(\theta) = E\left(\frac{h(v,x,z,e,\theta)}{f(v|z)}\right) = E[\int_L^K h(v,x,z,e,\theta)dv] \tag{3.2}
$$

The usefulness of equations 3.1 or 3.2 is that h can be a function of a limited dependent variable, and appropriate choice of the function h can make $\int_L^K h(v,x,z,e,\theta)dv$ either linear or quadratic in e, which then permits direct estimation of θ from $\psi(\theta)$.

To illustrate Theorem 3.1 , consider the binary choice model $d = I(v + x'\beta_0 + e > 0)$ with data consisting of a sample of observations of d_i, v_i, x_i, z_i. Letting $h(v,x,z,e,\theta) = z[d - I(v > 0)]$ gives, by equation 3.2,

$$
E\left(z\frac{d - I(v>0)}{f(v|z)}\right) = E[z(x'\beta_0 + e)] \tag{3.3}
$$

which, if $E(ze) = 0$, shows that β in the binary choice model can be estimated by linearly regressing $[d_i - I(v_i > 0)]/f(v_i|z_i)$ on x_i using instruments z_i. This, using a plug in estimator for $f(v_i|z_i)$, is the binary choice model estimator proposed in Lewbel (2000).

Taking $z = x$ yields the following Corollary to Theorem 3.1, which will be useful for estimation of models in which the errors are uncorrelated the regressors.

Corollary 3.1 *If $F_e(e|v,x) = F_e(e|x)$, $\Omega_e(v,x) = \Omega_e(x)$, and the support of the random variable v is the interval $[L,K]$, then*

$$
E\left(\frac{h(v,x,e,\theta)}{f(v|x)}\right) = E[\int_L^K h(v,x,e,\theta)dv] \tag{3.4}
$$

Sections 4 and 7 of this paper provide applications of Theorem 3.1 and Corollary 3.1.

4 Identification

This section derives the truncated regression estimator based on equations 2.1 and 2.2.

ASSUMPTION A.1: Assume y is given by equation 1.1 with $\alpha_0 \neq 0$. The conditional distribution of v given x is absolutely continuous with respect to a Lebesgue measure with nondegenerate Radon-Nikodym conditional density $f(v|x)$.

ASSUMPTION A.2: Let Ω denote the support of the distribution of an observation of (v, x). Let $F_e(e|v, x)$ denote the conditional distribution of an observation of e given an observation of (v, x), with support denoted $\Omega_e(v, x)$. Assume $F_e(e|v, x) = F_e(e|x)$ and $\Omega_e(v, x) = \Omega_e(x)$ for all $(v, x) \in \Omega$.

ASSUMPTION A.3: The conditional distribution of v given x has support $[L, K]$ for some constants L and K, $-\infty \leq L < K \leq \infty$.

ASSUMPTION A.4: For all (x, e) on the support of (x, e), $[I(\alpha_0 > 0)L - I(\alpha_0 < 0)K]\alpha_0 + x'\beta_0 + e < 0$. Let $\widetilde{k}$ equal the largest number that satisfies the inequality $\widetilde{k} \leq [I(\alpha_0 > 0)K - I(\alpha_0 > 0)L]\alpha_0 + x'\beta_0 + e$ for all (x, e) on the support of (x, e). $\widetilde{k} > 0$.

These assumptions do not require independent observations, though the root-n estimator provided later will assume independence. The identification result in Theorem 3.1 below only requires that the expectation of a certain function of f be identified.

Assumption A.1 says that y is defined by the truncated regression model and that v has a continuous distribution. Assumption A.2 was discussed in the introduction and in Lewbel (1998),(2000). In terms of equations 2.3 and 2.4. Assumption A.2 does not require the distribution of e to be continuous, e.g., it can be discrete or contain mass points.

The vector of regressors x can include dummy variables. Squares and interaction terms, e.g., $x_{3i} = x_{2i}^2$, are also permitted. In addition, x can be related to (e.g., correlated with) v, though Assumption A.1 rules out having elements of x be deterministic functions of v.

Assumption A.3 and A.4 requires v to have a large support. Standard models like tobit have errors that can take on any value, which would by these assumptions require v to have support equal to the whole real line. These assumptions imply that the estimator is likely to perform best when the spread of observations of v is large relative to the spread of $x'\beta + e$ (since if the observed spread of v values were not large, then the observed data would resemble data drawn from a process that violated A.4).

The truncation takes the form $y = \widetilde{y}|\widetilde{y} > 0$. It follows that, for any function $h(y, x, e)$ we have

$$E[h(\widetilde{y}, x, e)I(0 < \widetilde{y} < k)] = \begin{bmatrix} E[h(\widetilde{y}, x, e)I(0 < \widetilde{y} < k)|\widetilde{y} > 0]prob(\widetilde{y} > 0) \\ +E[h(\widetilde{y}, x, e)I(0 < \widetilde{y} < k)|\widetilde{y} \leq 0]prob(\widetilde{y} \leq 0) \end{bmatrix}$$

and so

$$E[h(y, x, e)I(0 < y < k)] = \frac{E[h(\widetilde{y}, x, e)I(0 < \widetilde{y} < k)]}{prob(\widetilde{y} > 0)} \tag{4.1}$$

The following Corollary to Theorem 3.1, along with equation 4.1, provides most of the machinery required to derive the estimator, which is based on Theorem 2 below.

Corollary 4.1 *Let Assumptions A.1, A.2, A.3 and A.4 hold. Let $H(\widetilde{y}, x, e)$ be any function that is differentiable in $\widetilde{y}$. Let k be any constant that satisfies $0 < k < \widetilde{k}$. Then*

$$E\left(\frac{\partial H(\widetilde{y}, x, e)}{\partial \widetilde{y}} \frac{I(0 < \widetilde{y} < k)}{f(v|x)}\right) = E\left(\frac{H(k, x, e, \theta) - H(0, x, e, \theta)}{|\alpha_0|}\right) \tag{4.2}$$

provided that these expectations exist.

Proof: First apply Corollary 3.1, then do a change of variables in the integration from v to $\widetilde{y}$ to get

$$E\left(\frac{\partial H(\widetilde{y}, x, e)}{\partial \widetilde{y}} \frac{I(0 < \widetilde{y} < k)}{f(v|x)}\right) = E\left(\int_L^K \frac{\partial H[\widetilde{y}(v, x, e), x, e]}{\partial \widetilde{y}(v, x, e)} \; I[0 < \widetilde{y}(v, x, e) < k] \; dv\right)$$

$$= E\left(\int_{L\alpha_0 + x'\beta_0 + e}^{K\alpha_0 + x'\beta_0 + e} \frac{\partial H(\widetilde{y}, x, e)}{\partial \widetilde{y}} \; I(0 < \widetilde{y} < k) \; d\widetilde{y}/\alpha_0\right)$$

if $\alpha_0 > 0$, or

$$= -E\left(\int_{-K\alpha_0 + x'\beta_0 + e}^{-L\alpha_0 + x'\beta_0 + e} \frac{\partial H(\widetilde{y}, x, e)}{\partial \widetilde{y}} \; I(0 < \widetilde{y} < k) \; d\widetilde{y}/\alpha_0\right)$$

if $\alpha_0 < 0$. Either way, by Assumptions A.3, A.4, and $0 < k < \widetilde{k}$, we get

$$= E\left(\int_0^k \frac{\partial H(\widetilde{y}, x, e)}{\partial \widetilde{y}} d\widetilde{y}/|\alpha_0|\right)$$

which proves the result

Theorem 4.1 : *Let Assumptions A.1, A.2, A.3, and A.4 hold. Let k be any constant that satisfies $0 < k < \widetilde{k}$. Let $w(x)$ be any positive, bounded function. Assume $E[exw(x)] = 0$ and $E[w(x)xx']$ exists and is nonsingular. Define the function $\mu(\alpha, \beta)$ by equation (2.1). Then equation (2.2) holds, and (α_0, β_0) are the only finite solutions to the first order conditions $\partial\mu(\alpha, \beta)/\partial\alpha = 0$ and $\partial\mu(\alpha, \beta)/\partial\beta = 0$.*

Proof of Theorem 4.1: Equations 4.1, 2.1, and $v = (\widetilde{y} - x'\beta_0 - e_i)/\alpha_0$ yield

$$\mu(\alpha, \beta) = E\left(\frac{[\widetilde{y}(\frac{1}{\alpha} - \frac{1}{\alpha_0}) + x'(\frac{\beta_0}{\alpha_0} - \frac{\beta}{\alpha}) + \frac{e}{\alpha_0}]^2 \; I(0 < \widetilde{y} < k) \; w(x)}{f(v|x)}\right)/prob(\widetilde{y} > 0)$$

Next apply Corollary 4.1 and $E[exw(x)] = 0$ to get

$$\begin{aligned}\mu(\alpha, \beta) prob(\widetilde{y} &> 0) = \frac{1}{|\alpha_0|} E\left(\int_0^k [\widetilde{y}(\frac{1}{\alpha} - \frac{1}{\alpha_0}) + x'(\frac{\beta_0}{\alpha_0} - \frac{\beta}{\alpha}) + \frac{e}{\alpha_0}]^2 \; w(x) d\widetilde{y}\right) \\ &= \frac{k^3 E[w(x)]}{3|\alpha_0|}(\frac{1}{\alpha} - \frac{1}{\alpha_0})^2 + \frac{k^2}{|\alpha_0|}(\frac{1}{\alpha} - \frac{1}{\alpha_0}) E[w(x)x'](\frac{\beta_0}{\alpha_0} - \frac{\beta}{\alpha}) + \\ &\quad + \frac{k}{|\alpha_0|}(\frac{\beta_0}{\alpha_0} - \frac{\beta}{\alpha})' E[w(x)xx'](\frac{\beta_0}{\alpha_0} - \frac{\beta}{\alpha}) + \frac{kE[w(x)e^2]}{|\alpha_0|\alpha_0^2}\end{aligned}$$

Minimizing $\mu(\alpha, \beta)$ first over β gives the first order condition

$$(\frac{\beta}{\alpha} - \frac{\beta_0}{\alpha_0}) = \frac{k}{2}(\frac{1}{\alpha} - \frac{1}{\alpha_0}) E[w(x)xx']^{-1} E[w(x)x]$$

which is linear in β and so has a unique solution. Call this solution $\beta(\alpha)$. The second order condition

$$\frac{\partial^2 \mu(\alpha, \beta)}{\partial\beta\partial\beta'} = \frac{2k}{|\alpha_0|\alpha^2} E[w(x)xx']$$

is positive definite, so $\beta(\alpha)$ does indeed minimizes $\mu(\alpha, \beta)$ with respect to β. Substituting the above first order condition into $\mu(\alpha, \beta)$ gives,

$$\mu[\alpha, \beta(\alpha)] prob(\widetilde{y} > 0) = \frac{k^3}{|\alpha_0|}(\frac{1}{\alpha} - \frac{1}{\alpha_0})^2 \left(\frac{E[w(x)]}{3} + \frac{3}{4} E[w(x)x'] E[w(x)xx']^{-1} E[w(x)x]\right) + \frac{kE[w(}{|\alpha_0}$$

The first order condition for minimizing $\mu[\alpha, \beta(\alpha)]$ is

$$\frac{2k^3}{|\alpha_0|\alpha^2}(\frac{1}{\alpha_0} - \frac{1}{\alpha}) \left(\frac{E[w(x)]}{3} + \frac{3}{4} E[w(x)x'] E[w(x)xx']^{-1} E[w(x)x] \right) = 0$$

which has solutions $\alpha = \pm\infty$ and $\alpha = \alpha_0$. Now

$$\mu(\pm\infty, \beta) = \frac{1}{|\alpha_0|\alpha_0^2} \left(\frac{k^3 E[w(x)]}{3} + k^2 E[w(x)x']\beta_0 + k\beta' E[w(x)xx']\beta_0 + kE[w(x)e^2] \right)$$

while $\beta(\alpha_0) = \beta_0$ and

$$\mu(\alpha_0, \beta_0) = \frac{kE[w(x)e^2}{|\alpha_0|\alpha_0^2} < \mu(\pm\infty, \beta)$$

Also the second order condition

$$\frac{d^2\mu[\alpha_0, \beta(\alpha_0)]}{\partial\alpha^2} = \frac{2k^3}{|\alpha_0|\alpha_0^4} \left(\frac{E[w(x)]}{3} + \frac{3}{4} E[w(x)x'] E[w(x)xx']^{-1} E[w(x)x] \right)$$

is positive, and hence $\alpha = \alpha_0$ and $\beta = \beta_0$ is both the only finite solution to the first order conditions,and is the global minimizer of $\mu(\alpha, \beta)$.

5 Estimation

Let $u = u(x)$ be any vector of variables such that the conditional density of v given x equals the conditional density of v given u, that is, $f(v|u) = f(v|x)$, where no element of u equals a deterministic function of other elements of u. For example, if $x = (1, z, z^2)$, we could take $u = z$. This construction of u is employed because $f(v_i|x_i)$ will be estimated as $\widehat{f}(v_i|u_i)$ using a kernel density estimator. Also, if v were known to be independent of some elements of x, then u could exclude those elements.

To deal with boundary bias issues in density estimation, we incorporate a "trimming" function τ_{ni} into the estimator procedure. A novelty of this asymptotic trimming is that it is based directly on the distance of observation i to the boundary of the support (if known), or on the distance to the nearest (element by element) extreme observation in the data. This trimming permits root n convergence of a density weighted average over the entire support of the data.

The resulting estimator based on Theorem 4.1 is

$$(\widehat{\alpha}, \widehat{\beta}) = \arg\min n^{-1} \Sigma_{i=1}^n \tau_{ni} \frac{(y_i - v_i\alpha - x_i'\beta)^2 \ \alpha^{-2} \ I(0 < y_i < k) \ w(x_i)}{\widehat{f}(v_i|u_i)} \tag{5.1}$$

for some sensibly chosen scalar k and weighting function w.

Closed form expressions for $(\widehat{\alpha}, \widehat{\beta})$ can be obtained as follows. Let $a = 1/\alpha$ and $b = -\beta/\alpha$. Then $(y - v\alpha - x'\beta)^2\alpha^{-2}I(0 < y) = (v - ya - x'b)^2 I(0 < y)$, so from equation 4.2 $\widehat{\alpha} = 1/\widehat{a}$ and $\widehat{\beta} = -\widehat{b}/\widehat{a}$ where

$$(\widehat{a}, \widehat{b}) = \arg\min n^{-1}\Sigma_{i=1}^{n}\tau_{ni} \cdot (v_i - y_i a - x_i' b)^2 \frac{I(0 < y_i < k)\ w(x_i)}{\widehat{f}(v_i|u_i)} \tag{5.2}$$

and 5.2 is just a linear weighted least regression of v on y and x, using weights $I(0 < y_i < k)w(x_i)/\widehat{f}(v_i|u_i)$. Unlike an ordinary least squares regression, where weighting only affects efficiency, in equation 5.1 or 5.2 the weights are functions of the regressand and are required for consistency.

The following theorem characterizes the limiting distribution of this estimator. The conditions upon which the theorem is based, as well as its proof, can be found in the appendix.

Theorem 5.1 *Define the matrix*

$$M = \begin{bmatrix} M_{\alpha\alpha} & M_{\alpha\beta} \\ M_{\beta\alpha} & M_{\beta\beta} \end{bmatrix}$$

where

$$M_{\beta\beta} = 2k|\alpha_0|^{-3}E[w(x_i)x_i x_i']$$

$$M_{\alpha\alpha} = 2k^3|\alpha_0|^5 \left(E[w(x_i)]/3 + 3/4(E[w(x_i)x_i']E[w(x_i)x_i x_i']^{-1}E[w(x_i)x_i]\right)$$

$$M_{\alpha\beta} = \alpha_0^{-2}(k/2E[(w(x_i)x_i x_i']^{-1}E[w(x_i)x_i] - \beta_0)$$

and the vector $\hbar_i \equiv (\hbar_{1i}, \hbar_{2i}')'$ *where*

$$\hbar_{1i} = e_i I[0 < \alpha_0 v_i + x_i'\beta_0 + e_i < k]w(x_i)(v_i + e_i/\alpha_0)$$

$$\hbar_{2i} = e_i I[0 < \alpha_0 v_i + x_i'\beta_0 + e_i < k]w(x_i)\alpha_0^{-2}x_i$$

Finally, let

$$\Omega = E\left[f(v_i|u_i)^{-2}\hbar_i\hbar_i'\right]$$

then

$$\sqrt{n}(\hat{\theta} - \theta_0) \Rightarrow N(0, M^{-1}\Omega M^{-1})$$

6 Identification With Instrumental Variables

Theorem 6.1 below describes the instrumental variables identification of the truncated regression model, where equations 2.3 and 2.4 are replaced with 2.6 and $E(ez) = 0$. Note that Assumption A.5' below is the standard assumption regarding instruments in two stage least squares regressions.

ASSUMPTION A.1': Assume y is given by equation 1.1 with $\alpha_0 \neq 0$. The conditional distribution of v given z is absolutely continuous with respect to a Lebesgue measure with nondegenerate Radon-Nikodym conditional density $f(v|z)$.

ASSUMPTION A.2': Let Ω denote the support of the distribution of an observation of (v, z). Let $F_{ex}(e, x|v, z)$ denote the conditional distribution of an observation of (e, x) given an observation of (v, z), with support denoted $\Omega_{ex}(v, z)$. Assume $F_{ex}(e, x|v, z) = F_{ex}(e, x|z)$ and $\Omega_{ex}(v, z) = \Omega_{ex}(z)$ for all $(v, z) \in \Omega$.

ASSUMPTION A.3': The conditional distribution of v given z has support $[L, K]$ for some constants L and K, $-\infty \leq L < K \leq \infty$.

ASSUMPTION A.4': For all (x, e) on the support of (x, e), $[I(\alpha_0 > 0)L - I(\alpha_0 < 0)K]\alpha_0 + x'\beta_0 + e < 0$. Let $\widetilde{k}$ equal the largest number that satisfies the inequality $\widetilde{k} \leq [I(\alpha_0 > 0)K - I(\alpha_0 > 0)L]\alpha_0 + x'\beta_0 + e$ for all (x, e) on the support of (x, e). $\widetilde{k} > 0$.

ASSUMPTION A.5': $E(ez) = 0$, $E(zz')$ exists and is nonsingular, and the rank of $E(xz')$ is J (the dimension of x).

Define Σ_{xz}, Σ_{zz}, Δ, and y^* by $\Sigma_{xz} = E(xz')$, $\Sigma_{zz} = E(zz')$,

$$\Delta = (\Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_{xz}')^{-1}\Sigma_{xz}\Sigma_{zz}^{-1} \tag{6.1}$$

$$y^* = \left[E\left(\frac{I(0 < y < k)}{f(v|z)}\right)\right]^{-1} \frac{(y - v\alpha_0)\ I(0 < y < k)}{f(v|z)} \tag{6.2}$$

Theorem 6.1 *Let Assumptions A.1', A.2', A.3', A.4' and A.5' hold. Let k be any constant that satisfies $0 < k < \widetilde{k}$. Then $E(zy^*) = E(zx')\beta$ and*

$$\beta = \Delta E(zy^*) \tag{6.3}$$

Proof of Theorem 6.1: Let $\partial H(\widetilde{y},x,z,e,\theta)$ be any function that is differentiable in $\widetilde{y}$. If Assumptions A.1', A.2', A.3' and A.4' hold then

$$E\left(\frac{\partial H(\widetilde{y},x,z,e,\theta)}{\partial\widetilde{y}}\frac{I(0<\widetilde{y}<k)}{f(v|z)}\right)=E\left(\frac{H(k,x,z,e,\theta)-H(0,x,z,e,\theta)}{|\alpha_0|}\right) \tag{6.4}$$

provided these expectations exist. The proof follows the same steps as the proof of Corollary 4.1. Similarly, the analog to equation 4.1 is

$$E[h(y,x,z,e)I(0<y<k)]=\frac{E[h(\widetilde{y},z,x,e)I(0<\widetilde{y}<k)]}{prob(\widetilde{y}>0)} \tag{6.5}$$

The theorem follows by applying equations 6.4 and 6.5 to$E[I(0<y<k)/f(v|z)]$ and to $E[z(y-v\alpha_0)I(0<y<k)/f(v|z)]$. In the latter application use $E[z(\widetilde{y}-v\alpha_0)I(0<\widetilde{y}<k)/f(v|z)]=E[z(x'\beta_0+e)I(0<\widetilde{y}<k)/f(v|z)]=(k/|\alpha_0|)[E(zx')\beta+E(ze)]$, where the last equality is equation 6.4 with $H=z(x'\beta_0+e)\widetilde{y}$.

Define $\eta(k)$ by

$$\eta(k)=\left[E\left(\frac{I(0<y<k)}{f(v|z)}\right)\right]^{-1}\left(\frac{2v\ I(0<y<k)}{f(v|z)}\right) \tag{6.6}$$

Corollary 6.1 *Let Assumptions A.1', A.2', A.3' and A.4' hold. Let k and k^* be any constants that satisfy $0<k^*<k<\widetilde{k}$. Then*

$$\alpha_0=\frac{k-k^*}{\eta(k)-\eta(k^*)} \tag{6.7}$$

Proof of Corollary 6.1:

$$\begin{aligned}E[vI(0 &< \widetilde{y}<k)/f(v|z)]=E[\alpha_0^{-1}(\widetilde{y}-x'\beta-e)I(0<\widetilde{y}<k)/f(v|z)]\\ &= \frac{k^2}{2\alpha_0|\alpha_0|}-\frac{kE(x'\beta-e)}{\alpha_0|\alpha_0|}\end{aligned}$$

where the second equality above is from equation 6.4 with $H=\alpha_0^{-1}[(\widetilde{y}^2/2)-\widetilde{y}(x'\beta+e)]$. Similarly, $E[I(0<\widetilde{y}<k)/f(v|z)]=k/|\alpha_0|$. Using equation 6.5 to go from $\widetilde{y}$ to y in these expressions yields $\eta(k)=(k/\alpha_0)-2E(x'\beta-e)$, and equation 6.7 follows immediately.

Equation 6.3 in Theorem 6.1 shows that β is identified, and can be estimated by an ordinary linear two stage least squares regression of y^* on x, using instruments z. The variable y^* depends on $f(v|z)$, which can be estimated by a kernel density estimator, but it also depends on α_0. Estimation of equation 6.7 can be used to estimate α_0. A disadvantage of equation 6.7 is that estimates based on it require choosing a constant k^* in addition to k. If the assumptions of Theorem 6.1 hold for $z = x$, the either the weighted least squares or the two stage least squares estimator could be used, but in that case the weighted least squares is likely to be preferable, in part because it does not require this separate preliminary estimator for α_0.

7 Estimation with Instrumental Variables

Equations (6.6) and (6.3) suggest a natural estimator for α_0, β_0. Let τ_{ni} denote a trimming function as before. Let f_i and $\hat{f}_i$ denote $f(v_i|z_i)$ and $\hat{f}(v_i|z_i)$, respectively, the latter being a kernel estimator. Define $\mu_0(k) \equiv E\left[\frac{I[0<y_i<k]}{f_i}\right]$, and its estimator

$$\hat{\mu}(k) = \frac{1}{n}\sum_{i=1}^{n} \tau_{ni}\frac{I[0 < y_i < k]}{\hat{f}_i}$$

and define our estimator of $\eta(k)$ as

$$\hat{\eta}(k) = \hat{\mu}(k)^{-1}\frac{1}{n}\sum_{i=1}^{n} \tau_{ni}\frac{2v_i I[0 < y_i < k]}{\hat{f}_i}$$

Then our estimator of α_0 is

$$\hat{\alpha} = \frac{k - k^*}{\hat{\eta}(k) - \hat{\eta}(k^*)}$$

The following theorem characterizes the limiting distribution of this estimator. The conditions under which it holds, as well as its proof, are left to the appendix:

Theorem 7.1 *The estimator $\hat{\alpha}$ is root-n consistent and asymptotically normal. Specifically, we have*

$$\sqrt{n}(\hat{\alpha} - \alpha_0) \Rightarrow N(0, E[\psi_{\alpha i}^2])$$

where

$$\begin{aligned} \psi_{\alpha i} &= \frac{\alpha_0}{\eta(k) - \eta(k^*}(\eta(k)\mu_0(k)^{-2}\psi_{\mu i}(k) + \mu_0(k)^{-1}\psi_{\eta i}(k) - \\ &= \frac{\alpha_0}{\eta(k) - \eta(k^*}(\eta(k^*)\mu_0(k^*)^{-2}\psi_{\mu i}(k^*) + \mu_0(k^*)^{-1}\psi_{\eta i}(k^*) \end{aligned}$$

and

$$\psi_{\mu i}(k) = \frac{I[0<y_i<k]}{f_i} - \mu_0 - E\left[\frac{I[0<y_i<k]}{f_i}\middle| v_i, z_i\right] + E\left[\frac{I[0<y_i<k]}{f_i}\middle| z_i\right]$$

and

$$\psi_{\eta i}(k) = \frac{v_i I[0<y_i<k]}{f_i} - \eta(k) - E\left[\frac{v_i I[0<y_i<k]}{f_i}\middle| v_i, z_i\right] + E\left[\frac{v_i I[0<y_i<k]}{f_i}\middle| z_i\right]$$

To estimate β_0 we define the estimator of Δ by

$$\hat{\Delta} = \left(\left(\frac{1}{n}\sum_{i=1}^{n} x_i z_i'\right)\left(\frac{1}{n}\sum_{i=1}^{n} z_i z_i'\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} z_i x_i'\right)\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} x_i z_i'\right)\left(\frac{1}{n}\sum_{i=1}^{n} z_i z_i'\right)^{-1}$$

and the estimator of y_i^* by

$$\hat{y}_i^* = \hat{\mu}^{-1}\frac{(y_i - v_i\hat{\alpha})I[0<y_i<k]}{\hat{f}_i}$$

Then our proposed estimator of β_0 is

$$\hat{\beta} = \hat{\Delta}\frac{1}{n}\sum_{i=1}^{n}\tau_{ni} z_i \hat{y}_i^*$$

The following theorem characterizes the limiting distribution of our proposed instrumental variables estimator. The conditions on which it holds, as well as its proof, are left to the appendix:

Theorem 7.2 *Define the following mean zero vectors:*

$$\begin{aligned}
\psi_{\beta_1 i} &= -\left(\mu_0^{-2}\frac{k}{\alpha_0}E[z_i x_i']\beta_0\right)\cdot\psi_{\mu i} \\
\psi_{\beta_2 i} &= -\left(\frac{1}{2\alpha_0^2}(k^2 E[z_i] - kE[z_i x_i']\beta_0\right)\cdot\mu_0^{-1}\cdot\psi_{\alpha i} \\
\psi_{\beta_3 i} &= \frac{\mu_0^{-1} z_i(y_i - v_i\alpha_0)I[0<y_i<k]}{f_i} - z_i x_i'\beta_0 - E\left[\frac{\mu_0^{-1} z_i(y_i - v_i\alpha_0)I[0<y_i<k]}{f_i}\middle| z_i, v_i\right] \\
&+ E\left[\frac{\mu_0^{-1} z_i(y_i - v_i\alpha_0)I[0<y_i<k]}{f_i}\middle| z_i\right]
\end{aligned}$$

and let

$$\psi_{\beta i} = \psi_{\beta_1 i} + \psi_{\beta_2 i} + \psi_{\beta_3 i}$$

and

$$\Omega_\beta = E\left[\psi_{\beta i}\psi'_{\beta i}\right]$$

Then we have

$$\sqrt{n}(\hat{\beta} - \beta_0) \Rightarrow N(0, \Delta \cdot \Omega_\beta \cdot \Delta')$$

8 Monte Carlo Results

In this section, the finite sample properties of the estimators proposed in this paper are examined by a small simulation study. The performance of our estimators are compared to existing parametric and semiparametric estimators. The study was performed in GAUSS.

Simulation results for the weighted least squares (WLS) estimator are reported in Tables I-IV. We simulated data from the following model:

$$\begin{aligned} \widetilde{y}_i &= 1 + v_i + (0.5)x_i + \sigma(x_i)\epsilon_i \\ y_i &= \widetilde{y}_i | \widetilde{y}_i > 0 \end{aligned}$$

The random variable x_i was distributed uniform between -1 and 1. v_i was distributed as the sum of a uniform random variable between -6 and 6 and x_i. The error term ϵ_i was distributed independently of v_i, x_i, either truncated normal, with support[-2,2], or chi-squared with one degree of freedom, censored at 4, minus its mean. For homoskedastic designs the scale function $\sigma(x_i)$ was set to 1, and for heteroskedastic designs the scale function was set to $\exp(x_i/2)$.

To implement the WLS estimator, we used a bandwidth of order $n^{-1/5}$ to estimate the joint density of v_i, x_i, and a bandwidth of order $n^{-1/4}$ for estimating the marginal density of x_i. Silverburg's rule of thumb was used to calculate the constant, and a quartic kernel function was used. We set the right truncation point k to be the sample 90th percentile of y_i.

For comparison, results are also reported for the symmetrically trimmed least squares (STLS) estimator in Powell(1986b), the pairwise difference (PWD) estimator in Honoré and Powell(1994), and the maximum likelihood estimator (MLE) assuming a homoskedastic normal distribution. The PWD and STLS were computed using linear programming and iterative least squared methods, respectively. The MLE was computed using the BFGS algorithm. The summary statistics reported are mean bias, median bias, root-mean squared error (RMSE), and mean absolute deviation (MAD). Sample sizes of 100, 200, and 400 were simulated with 401 replications.

As the results in Table 1-4 indicate, the proposed WLS estimator performed very well at all sample sizes in the homoskedastic design, and moderately well in the heteroskedastic designs. In contrast, the MLE performs poorly everywhere, due to distributional misspecification and/or heteroskedasticity. The STLS performs poorly with the chi-squared errors, and PWD performs poorly in the heteroskedastic designs.

Tables 5 and 6 report results for the instrumental variables two stage least squares (2SLS) estimator. Here we simulated data from the following model:

$$
\begin{aligned}
\widetilde{y}_i &= 1 + v_i + x_i + \epsilon_i \\
y_i &= \widetilde{y}_i | \widetilde{y}_i > 0
\end{aligned}
$$

To incorporate endogeneity, we simulated a binary variable d_i which took the value 1 with probability 1/2 and 0 otherwise. When d_i was 1, the error term ϵ_i was equal to x_i, and when d_i was 0, the error term was drawn from one of the two distributions mentioned previously. The instrument z_i was independently distributed as uniform between -1 and 1 when d_i was one, and equal to x_i when d_i was 0.

Results using this endogenous model are reported for our 2SLS estimator, and for the STLS,PWD, and MLE estimators. As Tables 5,6 indicate, only the 2SLS performs at an acceptable level when the regressor is endogenous. The other estimators, which are inconsistent when the regressors are endogenous, perform very poorly, with biases as high as 50%, and not decreasing with the sample size.

Overall, the results of our simulation study indicate that the estimators introduced in this paper perform well in moderately sized samples. The results for the endogenous regressor design are especially encouraging when compared to other estimation procedures.

9 Conclusions

This paper proposes new estimators for truncated regression models. The estimators are “distribution free”, and are robust to general forms of conditional heteroskedasticity, as well as general forms of measurement error and endogeneity. The proposed estimators converge at the parametric rate and have a limiting normal distribution.

Our limiting distribution theory employs a new variant of asymptotic trimming to deal with boundary bias issues. This is demonstrated for estimation of density weighted averages, but should be usefully applicable in general contexts involving two step 'plug-in' estimators with a nonparametric first step.

We have focused on estimation of coefficients, but the proposed methodology may also be useful in recovering other information regarding the distribution of the latent $\widetilde{y}$. For example, a simple implication of intermediate results in the proof of corollary 6.1 is $(|\alpha_0|/k)E[I(0 < y < k)/f(v|z)] = prob(\widetilde{y} > 0)$,.which immediately suggests an estimator of this probability.

The results in this paper suggest areas for future research. For example, the semiparametric efficiency bound of the models considered needs to be derived under the exclusion restriction we imposed, so that the relative efficiency of our estimators can be computed. Furthermore, an extension of the exclusion restriction we impose to other limited dependent variable models is also worth exploring.

References

[1] Honoré, B.E. and J.L. Powell (1994) "Pairwise Difference Estimators of Censored and Truncated Regression Models", Journal of Econometrics, 64, 241-278.

[2] Lee, M.J. (1989), "Mode Regression", Journal of Econometrics, 42, 337-349.

[3] Lee, M.J. (1993), "Quadratic Mode Regression", Journal of Econometrics, 57, 1-19.

[4] Lewbel, A. (2000), "Semiparametric Qualitative Response Model Estimation With Unknown Heteroscedasticity or Instrumental Variables," Journal of Econometrics, 97, 145-177.

[5] Lewbel, A. (1998), "Semiparametric Latent Variable Model Estimation With Endogenous or Mismeasured Regressors," Econometrica, 66, 105–121.

[6] Newey, W. K. and D. McFadden (1994), "Large Sample Estimation and Hypothesis Testing," in Handbook of Econometrics, vol. iv, ed. by R. F. Engle and D. L. McFadden, pp. 2111-2245, Amsterdam: Elsevier.

[7] Pakes, A. and D. Pollard (1989), "Simulation and the Asymptotics of Optimization Estimators," Econometrica, 57, 1027-1058.

[8] Powell, J.L. (1986a) "Censored Regression Quantiles", Journal of Econometrics, 32, 143-155.

[9] Powell, J.L. (1986b) "Symmetrically Trimmed Least Squares Estimation of Tobit Models", Econometrica, 54, 1435-1460.

[10] Powell, J. L., J. H. Stock, and T. M. Stoker (1989), "Semiparametric Estimation of Index Coefficients," Econometrica 57, 1403–1430.

[11] Sherman, R.P. (1994) "U-Processes in the Analysis of a Generalized Semiparametric Regression Estimator", Econometric Theory, 10, 372-395.

A Appendix

A.1 General Theorem for Density Weighted Extremum Estimator

In this section, we establish the asymptotic properties of a general density weighted extremum estimator. The estimator is defined as the maximizer of an objective function that involves a preliminary root-n consistent estimator of a finite dimensional nuisance parameter, as well as a preliminary estimator of a conditional density function using kernel estimation. Furthermore, the objective function need not be differentiable in the parameters of interest. Therefore, each of the three estimators introduced in this paper will fall into this framework. Because of the generality of the class of estimators discussed in this section, most of the assumptions will be left as "high level". They will be verified for each of the estimators in the following sections.

We begin by defining the estimator. As mentioned, the objective function includes a finite dimensional nuisance parameter, which we denote here by $\kappa_0 \in \mathbf{R}^q$, which has been estimated in a preliminary stage. We denote the estimator by $\hat{\kappa}$. We will assume throughout this section that $\hat{\kappa}$ has an asymptotically linear representation. Letting the random variables y_i, v_i, and the random vectors z_i, x_i be as defined in the paper, we express the representation as:

$$\hat{\kappa} - \kappa_0 = \frac{1}{n}\sum_{i=1}^{n} \psi_i + o_p(n^{-1/2}) \tag{A.1}$$

where ψ_i denotes $\psi(y_i, x_i, v_i, z_i)$ and satisfies $E[\psi_i] = 0$ and $E[\|\psi_i\|^2] < \infty$. The objective function also involves an estimator of a conditional density function, which we denote here by $f(v_i|z_i)$. We assume that a kernel estimator is used to estimate this function, and denote the estimator by $\hat{f}(v_i|z_i)$. To define this estimator, we first assume that the vector $z_i \in \mathbf{R}^Z$ can be partitioned as $z_i = (z_i^{(c)}, z_i^{(d)})$, where $z_i^{(c)} \in \mathbf{R}^{Z_c}$ is continuously distributed, and $z_i^{(d)} \in \mathbf{R}^{Z_d}$ is discretely distributed. We let $\mathcal{Z} = \mathcal{Z}_c \times \mathcal{Z}_d$ denote the support of z_i. We assume the support of $\mathcal{Z}_c$ is an open, convex subset of $\mathbf{R}^{Z_c}$ and assume the support of v_i, denoted by $\mathcal{V}$ is an open interval in $\mathbf{R}$. We define the kernel estimator as:

$$\hat{f}(v_i|z_i) = \frac{\frac{1}{nh_n^{Z_c+1}}\sum_{j\neq i} I[z_i^{(d)} = z_j^{(d)}] K_1\left(\frac{z_j^{(c)} - z_i^{(c)}}{h_n}\right) K_2\left(\frac{v_j - v_i}{h_n}\right)}{\frac{1}{nh_n^{Z_c}}\sum_{j\neq i} I[z_i^{(d)} = z_j^{(d)}] K_1\left(\frac{z_j^{(c)} - z_i^{(c)}}{h_n}\right)} \tag{A.2}$$

Where K_1 and K_2 are "kernel" functions, and h_n is a bandwidth sequence. Properties of K_1, K_2 and h_n will be detailed in assumptions needed for the main theorems.

Here, we let $\theta_0 \in \mathbf{R}^k$ denote the parameter of interest, known to lie in some parameter space Θ. The density weighted extremum estimator, denoted by $\hat{\theta}$, is defined as

$$\hat{\theta} = \text{argsup}_{\theta} \frac{1}{n}\sum_{i=1}^{n} \tau_{ni} \frac{\hat{h}_i(\theta)}{\hat{f}_i} \tag{A.3}$$

where

- τ_{ni} is a trimming function depending on v_i,z_i and n only. It essentially serves to address technical problems that arise when the conditional density approaches 0.

- $\hat{h}_i(\theta)$ is shorthand notation for $h(y_i, v_i, x_i, z_i, \hat{\kappa}, \theta)$, where the properties of the function $h(\cdot)$ will be specified in the theorems.

- $\hat{f}_i$ is shorthand notation for $\hat{f}(v_i|z_i)$.

We first give sufficient conditions for the consistency of the estimator. The consistency theorem is based on the following assumptions:

AA1 Θ is a compact subset of $\mathbf{R}^k$.

AA2 The random vector $(y_i, v_i, x_i', z_i')'$ is identically and independently distributed.

AA3 Letting $h_i(\theta), f_i$ denote $h(y_i, v_i, x_i, z_i, \kappa_0, \theta)$ and $f(v_i|z_i)$ respectively, we assume that $E\left[\frac{h_i(\theta)}{f_i}\right]$ is continuous in θ and uniquely maximized at θ_0.

AA4 Let $h_i(\kappa, \theta)$ denote $h(y_i, v_i, x_i, z_i, \kappa, \theta)$, and let $\mathcal{A}$ be a neighborhood of κ_0. Then

AA4.1 $E\left[\sup_{\kappa\in\mathcal{A},\theta\in\Theta} \|h_i(\kappa,\theta)\|\right] < \infty$

AA4.2 $E\left[\sup_{\kappa\in\mathcal{A},\theta\in\Theta} \frac{\|h_i(\kappa,\theta)\|}{f_i}\right] < \infty$

AA4.3 There exists a constant $\gamma_0 \in (0, 1]$, and a function $g_0(\theta)$ that satisfies $\sup_{\theta\in\Theta} g(\theta) < \infty$, such that for all $\kappa \in \mathcal{A}$,

$$\left\|E\left[\frac{h_i(\kappa,\theta) - h_i(\theta)}{f_i}\right]\right\| \leq g(\theta)\|\kappa - \kappa_0\|^{\gamma_0}$$

AA5 Let $f(v_i, z_i^{(c)}|z_i^{(d)})$ denote the conditional (Lebesgue) density of $v_i, z_i^{(c)}$ given $z_i^{(d)}$, and let $f(z_i^{(d)})$ denote the probability mass function of $z_i^{(d)}$. Then we assume:

AA5.1 For all $v_i, z_i \in \mathcal{V} \times \mathcal{Z}$, $f(v_i, z_i^{(c)}|z_i^{(d)})$, considered as a function of $v_i, z_i^{(c)}$, is continuously differentiable.

AA5.2 There exists a constant $\mu_0 > 0$ such that for all $z_i \in \mathcal{Z}$, $f(z_i^{(d)}) \notin (0, \mu_0)$.

AA6 The trimming function τ_{ni} has the following properties:

AA6.1 τ_{ni} is a function of v_i, z_i and n only, and $0 \leq \tau_{ni} \leq 1$ for all $n \in \mathbf{N}$.

AA6.2 For each $v_i, z_i \in \mathcal{V} \times \mathcal{Z}$, $\tau_{ni} \to 1$ as $n \to \infty$.

AA6.3 For all $\delta > 0$, $\sup_{v_i,z_i\in\mathcal{V}\times\mathcal{Z}} \tau_{ni}/f_i = o(n^\delta)$, and $\sup_{v_i,z_i\in\mathcal{V}\times\mathcal{Z}} \tau_{ni}/f_{zvi} = o(n^\delta)$, where f_{vzi} denotes $f(v_i, z_i)$.

The following theorem establishes consistency for the density weighted extremum estimator:

Theorem A.1 *Assume that the kernel functions $K_1(\cdot)$ and $K_2(\cdot)$ are symmetric, have bounded support, are of bounded variation, and integrate to 1. Also assume that $h_n \to 0$, $nh_n^{Z_c}/\ln n \to \infty$, and (A.1) holds. Then under Assumptions* **AA1-AA6**

$$\hat{\theta} \xrightarrow{p} \theta_0 \tag{A.4}$$

Proof: The proof will be based on standard theorems for consistency with compact parameter spaces. (For example, see Newey and MacFadden(1994), Theorem 2.1.) We first show that uniformly in θ, the objective function is asymptotically equivalent to the infeasible objective function:

$$\frac{1}{n}\sum_{i=1}^{n} \tau_{ni} \frac{h_i(\theta)}{f_i}$$

We decompose the objective function as:

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\tau_{ni}\frac{h_i(\hat{\kappa},\theta)}{\hat{f}_i} &= \frac{1}{n}\sum_{i=1}^{n}\tau_{ni}\frac{h_i(\theta)}{f_i} \\ &+ \frac{1}{n}\sum_{i=1}^{n}\tau_{ni}\left(\frac{h_i(\hat{\kappa},\theta)-h_i(\theta)}{f_i}\right) \\ &+ \frac{1}{n}\sum_{i=1}^{n}\tau_{ni}h_i(\hat{\kappa},\theta)\left(\frac{1}{\hat{f}_i}-\frac{1}{f_i}\right)\end{aligned}$$

Thus we will show that

$$\sup_{\theta\in\Theta}\left\|\frac{1}{n}\sum_{i=1}^{n}\tau_{ni}\left(\frac{h_i(\hat{\kappa},\theta)-h_i(\theta)}{f_i}\right)\right\| = o_p(1) \tag{A.5}$$

and

$$\sup_{\theta\in\Theta}\left\|\frac{1}{n}\sum_{i=1}^{n}\tau_{ni}h_i(\hat{\kappa},\theta)\left(\frac{1}{\hat{f}_i}-\frac{1}{f_i}\right)\right\| = o_p(1) \tag{A.6}$$

To show (A.5), we first note that by (A.1), we have $\hat{\kappa}-\kappa_0 = O_p(n^{-1/2})$, so it will suffice to show that

$$\sup_{\theta\in\Theta,\|\kappa-\kappa_0\|\leq(\log n)n^{-1/2}}\left\|\frac{1}{n}\sum_{i=1}^{n}\tau_{ni}\left(\frac{h_i(\kappa,\theta)-h_i(\theta)}{f_i}\right)\right\| = o_p(1) \tag{A.7}$$

We proceed by adding and subtracting $E\left[\tau_{ni}\frac{h_i(\kappa,\theta)}{f_i}\right]$ inside the above summation. Under Assumption AA4.2, Lemma 2.4 in Newey and MacFadden(1994) implies that

$$\sup_{\theta\in\Theta,\|\kappa-\kappa_0\|\leq(\log n)n^{-1/2}}\left\|\frac{1}{n}\sum_{i=1}^{n}\tau_{ni}\frac{h_i(\kappa,\theta)}{f_i} - E\left[\tau_{ni}\frac{h_i(\kappa,\theta)}{f_i}\right]\right\| = o_p(1) \tag{A.8}$$

To show (A.7) it thus remains to establish

$$\sup_{\theta\in\Theta,\|\kappa-\kappa_0\|\leq(\log n)n^{-1/2}}\left\|\frac{1}{n}\sum_{i=1}^{n}E\left[\tau_{ni}\frac{h_i(\kappa,\theta)}{f_i}\right] - \tau_{ni}\frac{h_i(\theta)}{f_i}\right\| = o_p(1) \tag{A.9}$$

We add and subtract $E\left[\tau_{ni}\frac{h_i(\theta)}{f_i}\right]$ from the above summation. It follows from Assumption AA4.2 and Lemma 2.4 in Newey and MacFadden(1994) that

$$\sup_{\theta\in\Theta}\frac{1}{n}\sum_{i=1}^{n}E\left[\tau_{ni}\frac{h_i(\theta)}{f_i}\right] - \tau_{ni}\frac{h_i(\theta)}{f_i} = o_p(1) \tag{A.10}$$

We also have by Assumptions AA6.1,AA6.2 that

$$E\left[\tau_{ni}\frac{h_i(\theta)}{f_i}\right] - E\left[\frac{h_i(\theta)}{f_i}\right] = o(1)$$

By Assumption AA4.3, we have

$$\sup_{\theta\in\Theta,\|\kappa-\kappa_0\|\leq(\log n)n^{-1/2}}\left\|E\left[\tau_{ni}\frac{h_i(\kappa,\theta)-h_i(\theta)}{f_i}\right]\right\| \leq \left((\log n)n^{-1/2}\right)^{\gamma_0}\cdot\sup_{\theta\in\Theta}g_0(\theta) = o_p(1) \tag{A.11}$$

Thus (A.5) is established. We next establish (A.6). Let f_{vzi} denote $f(v_i,z_i)\equiv f(v_i,z_i^{(c)}|z_i^{(d)})\cdot f(z_i^{(d)})$ and let f_{zi} denote $f(z_i)\equiv f(z_i^{(c)}|z_i^{(d)})\cdot f(z_i^{(d)})$. Also, let $\hat{f}_{zvi}$ and $\hat{f}_{zi}$ denote the kernel estimators of these values. Let $\|\hat{f}_{zi}-f_{zi}\|_\infty$ denote

$\sup_{z_i \in \mathcal{Z}} |f_{zi} - \hat{f}_{zi}|$, and let $\|\hat{f}_{zvi} - f_{vzi}\|_\infty$ denote $\sup_{z_i, v_i \in \mathcal{Z} \times \mathcal{V}} |f_{vzi} - \hat{f}_{zvi}|$. We work with the identity

$$\frac{1}{\hat{f}_i} - \frac{1}{f_i} = \frac{\hat{f}_{zi} - f_{zi}}{f_{vzi}} \tag{A.12}$$

$$- \frac{f_{zi}(\hat{f}_{zvi} - f_{vzi})}{f_{vzi}^2} \tag{A.13}$$

$$- \frac{(\hat{f}_{zi} - f_{zi})(\hat{f}_{zvi} - f_{vzi})}{f_{vzi}\hat{f}_{zvi}} \tag{A.14}$$

$$+ \frac{f_{zi}(\hat{f}_{zvi} - f_{vzi})^2}{\hat{f}_{zvi} f_{vzi}^2} \tag{A.15}$$

By Assumption AA5 and the conditions on the bandwidth, by Lemma 8.10 in Newey and McFadden(1994), we have $\|\hat{f}_{zi} - f_{zi}\|_\infty$ and $\|\hat{f}_{zvi} - f_{vzi}\|_\infty$ are both $O_p(n^{-\delta})$ for some $\delta > 0$. Thus (A.6) follows from Assumption AA4.1 and AA6.3.

This establishes uniform convergence. Compactness follows from Assumption AA1, and identification follows from Assumption AA3. Consistency follows from Theorem 2.1 in Newey and MacFadden(1994). ■

Our next step is to derive the limiting distribution of the density weighted extremum estimator. We provide sufficient conditions for the estimator to converge at the root-n rate with an asymptotic normal distribution. We first impose stronger smoothness assumptions on the function $h_i(\kappa, \theta)$, requiring that it be right-differentiable in θ, and let $\hbar_i(\kappa, \theta)$ denote the subgradient with respect to θ. Our results will be based on the assumption that the density weighted extremum estimator satisfies the following asymptotic first order condition:

$$\frac{1}{n}\sum_{i=1}^{n} \tau_{ni} \frac{\hat{\hbar}_i(\hat{\theta})}{\hat{f}_i} = o_p(n^{-1/2}) \tag{A.16}$$

where $\hat{\hbar}_i(\theta) = \hbar_i(\hat{\kappa}, \theta)$. We now list the additional regularity conditions used for the proof of the limiting distribution of the density weighted extremum estimator. Throughout the arguments used in the proof, we let $\hbar_i(\theta), \hbar_i$ denote $\hbar_i(\kappa_0, \theta)$ and $\hbar_i(\kappa_0, \theta_0)$ respectively.

B1 For all $\kappa \in \mathcal{A},\ n \in \mathbf{N}$, $E\left[\tau_{ni}\frac{\hbar_i(\kappa,\theta)}{f_i}\right]$ is differentiable in θ in a neighborhood of θ_0.

B2 For all θ in a neighborhood of θ_0, $E\left[\tau_{ni}\frac{\hbar_i(\kappa,\theta)}{f_i}\right]$ is differentiable in κ for $\kappa \in \mathcal{A},\ n \in \mathbf{N}$.

B3 Let M_θ denote $\nabla_\theta E\left[\frac{\hbar_i(\kappa_0,\theta_0)}{f_i}\right]$, where ∇_θ denotes the gradient with respect to θ. Then we assume that $-M_\theta$ is a positive definite matrix.

B4 Let M_κ denote $\nabla_\kappa E\left[\frac{\hbar_i(\kappa_0,\theta_0)}{f_i}\right]$. Also, let δ_{ni} denote the vector:

$$M_\kappa \psi_i + \tau_{ni}\frac{\hbar_i}{f_i} - E\left[\tau_{ni}\frac{\hbar_i}{f_i}\middle| v_i, z_i\right] + E\left[\tau_{ni}\frac{\hbar_i}{f_i}\middle| z_i\right]$$

and let δ_i denote

$$M_\kappa \psi_i + \frac{\hbar_i}{f_i} - E\left[\frac{\hbar_i}{f_i}\middle| v_i, z_i\right] + E\left[\frac{\hbar_i}{f_i}\middle| z_i\right]$$

then we assume that

B4.1 $E\left[\|\delta_{ni}\|^2\right] < \infty$ for all $n \in \mathbf{N}$.

B4.2 $\frac{1}{n}\sum_{i=1}^{n} \delta_i - \delta_{ni} = o_p(n^{-1/2})$

B5 The condition in Assumption AA5.1 is strengthened to p times continuously differentiable, with bounded p^{th} derivatives.

B6 The functional space $\{\hbar_i(\kappa,\theta) : \kappa \in \mathcal{A}, \theta \in \Theta\}$ is Euclidean with respect to the envelope $\mathcal{F}_i$, where $E\left[\mathcal{F}_i^2\right] < \infty$.

B7 The functions $E\left[\left\|\tau_{ni}\frac{\hbar_i(\kappa,\theta)}{f_i}\right\|^2\right]$, $\nabla_\kappa E\left[\tau_{ni}\frac{\hbar_i(\kappa,\theta)}{f_i}\right]$, and $\nabla_\theta E\left[\tau_{ni}\frac{\hbar_i(\kappa,\theta)}{f_i}\right]$ are continuous at κ_0, θ_0 for all $n \in \mathbf{N}$.

B8 For two vectors of the same dimension, $u, \jmath$, we let $u^\jmath$ denote the product of each of the components of u raised to the corresponding component of $\jmath$. Also, for a vector l which has all integer components, we let $[l]$ denote the sum of its components. The kernel functions are assumed to have the following property:

$$\int K_j(u)u^\jmath du = 0 \quad j = 1, 2 \quad l \in \mathbf{N}, 1 \leq [l] < p$$

B9 The functions

$$E\left[\tau_{ni}\frac{\hbar_i}{f_i}\Big|z_i\right]$$

and

$$E\left[\tau_{ni}\frac{\hbar_i}{f_i}\Big|z_i, v_i\right]$$

are p times differentiable with bounded p^{th} derivatives, in $z_i^{(c)}$ and $z_i^{(c)}, v_i$ respectively, for all values of $z_i^{(d)}$ and all $n \in \mathbf{N}$.

B10 In addition to Assumption AA6, we assume the trimming function satisfies the following: letting $\mathcal{K}_1, \mathcal{K}_2$ denote the supports of the kernel functions $K_1(\cdot), K_2(\cdot)$, then $\tau_{ni} = 0$ if

$$\inf_{(u,v)\in\mathcal{K}_1\times\mathcal{K}_2} f(v_i + uh_n, z_i + vh_n) = 0$$

Remark A.1 *Before proceeding to the statement and proof of the theorem, which characterizes the limiting distribution of the density weighted extremum estimator, we remark on the regularity conditions imposed. Many of these assumptions (e.g. smoothness, moment conditions) are standard when compared to assumptions imposed for existing semiparametric estimators. However, some of our assumptions regarding the trimming function τ_{ni} have particular features which warrant comment.*

1. *Assumption AA6 implicitly makes assumptions regarding where and how quickly the densities f_i,f_{vzi} approach 0, as was assumed in Sherman(1994). Sherman(1994) provides concrete examples where WLS6.1 will be satisfied.*

2. *Assumption B4 ensures that the bias induced by the trimming function decreases to zero faster than the parametric rate. This assumption imposes conditions on the tail behavior of e_i, v_i, z_i, and can be satisfied in a variety of cases. For example, if the error term e_i has bounded support, the condition is satisfied if v_i (strictly) contains the support of e_i. The assumption can also be satisfied when e_i has an unbounded support, if the support of v_i has sufficiently "heavier" tails.*

3. *Assumption B10 is imposed specifically for situations when the supports of v_i, z_i are compact. In this case, the kernel density estimator has a bias of different order near the boundary of the support set. To ensure uniformity of the order of the bias, observations which are near the boundary, even those for which the value of the density function are bounded away from 0, are also trimmed away.*

Remark A.2 *As an alternative to the trimming conditions in AA6 and B10, which has the drawback of requiring that the researcher know where and how quickly regressor densities go to 0, we propose the following data dependent trimming procedure. This procedure only applies to situtations where the regressors which have a bounded, "rectangular" support; To explain this procedure we let x_i denote a $k-$dimensional vector, and h_n denote the bandwidth sequence used in density estimation of $f(x_i)$.*

Assume x_i has compact support. Let x_{mx} denote the $k-$dimensional vector of the maxima in the supports of each of the k components of x_i and x_{mn} denote the vector of minima. We also assume a "rectangular support" of x_i in the sense:

$$f(x) > 0 \quad \forall x \in [x_{mn}^{(1)}, x_{mx}^{(1)}] \times [x_{mn}^{(2)}, x_{mx}^{(2)}] \times ...[x_{mn}^{(k)}, x_{mx}^{(k)}]$$

where superscripts denote components of the vector x_i.

One form of the infeasible trimming function is the product of k indicator functions:

$$\tau_n(x_i) = I[x_i^{(1)} \in [[x_{mn}^{(1)} + h_n, x_{mx}^{(1)} - h_n]] \cdot I[x_i^{(2)} \in [[x_{mn}^{(2)} + h_n, x_{mx}^{(2)} - h_n]] \cdot ...I[x_i^{(k)} \in [[x_{mn}^{(k)} + h_n, x_{mx}^{(k)} - h_n]$$

To define the feasible, data-dependent trimming function, let $x_{\hat{mx}}$ denote the k vector obtained by taking the maximum of each of the components of x_i from a sample of n observations. Let $x_{\hat{mn}}$ denote the vector of sample minima. The feasible trimming function is

AA6' $\hat{\tau}_n(x_i) = I[x_i^{(1)} \in [[x_{\hat{mn}}^{(1)} + h_n, x_{\hat{mx}}^{(1)} - h_n]] \cdot I[x_i^{(2)} \in [[x_{\hat{mn}}^{(2)} + h_n, x_{\hat{mx}}^{(2)} - h_n]] \cdot ...I[x_i^{(k)} \in [[x_{\hat{mn}}^{(k)} + h_n, x_{\hat{mx}}^{(k)} - h_n]$

We show now that for our purposes, the feasible data dependent trimming function is asymptotically equivalent to the infeasible trimming function in density estimation.

Lemma A.1 *Let $\hat{f}(x_i)$ denote the kernel density estimator. Then*

$$\frac{1}{n}\sum_{i=1}^{n}(\hat{\tau}_n(x_i) - \tau_n(x_i))(\hat{f}(x_i) - f(x_i)) = o_p(n^{-1/2})$$

Proof: By the uniform consistency of kernel density estimators over compact sets, it will suffice to show that

$$\frac{1}{n}\sum_{i=1}^{n}(\hat{\tau}_n(x_i) - \tau_n(x_i)) = O_p(n^{-1/2})$$

Let A_n denote $\frac{1}{n}\sum_{i=1}^{n}\tau_n(x_i) - \hat{\tau}_n(x_i)$, and let B_n denote the event

$$|x_{\hat{mx}}^{(j)} - x_{mx}^{(j)}| < n^{-(1/2+\delta)}, |x_{\hat{mn}}^{(j)} - x_{mn}^{(j)}| < n^{-(1/2+\delta)} \quad j = 1, 2, ...k$$

We have for some arbitrarily small $\epsilon > 0$,

$$P(n^{1/2}|A_n| > \epsilon) \leq P(n^{1/2}|A_n| > \epsilon, B_n) + P(B_n^c)$$

where B_n^c denotes the compliment of the event B_n. We note that

$$P(B_n^c) \leq \sum_{j=1}^{k} P(|x_{\hat{mx}}^{(j)} - x_{mx}^{(j)}| \geq n^{-(1/2+\delta)}) + P(|x_{\hat{mn}}^{(j)} - x_{mn}^{(j)}| \geq n^{-(1/2+\delta)})$$

and the right hand side goes to 0 by the well known n-rate of convergence of the extreme estimators under the compact support conditions. Also, we note that

$$P(n^{1/2}|A_n| > \epsilon, B_n) \leq P(C_n > \epsilon)$$

where

$$C_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} I[\sum_{j=1}^{n} I[x_i^{(j)} \in [x_{mx}^{(j)} - h_n - n^{-(1/2+\delta)}, x_{mx}^{(j)} - h_n]] + I[x_i^{(j)} \in [x_{mn}^{(j)} + h_n - n^{-(1/2+\delta)}, x_{mn}^{(j)} + h_n]] > 0]$$

We note that by the assumption that x_i has positive density everywhere on the rectangle, $E[C_n] = o(1)$ and $\text{Var}(C_n) = o(1)$, so $P(C_n > \epsilon) \to 0$, establishing the desired result. ■

We now prove the main theorem, characterizing the limiting distribtution of the density weighted extremum estimator. The asymptotic properties hold either under the infeasible trimming function, or the feasible trimming with bounded, rectangular support conditions.

Theorem A.2 *Suppose either Assumptions AA1-AA6 and B1-B10 or alternatively, AA1-AA5,AA6' and B1-B9 hold, and the bandwidth h_n satisfies $\sqrt{n}h_n^p \to 0$,*

$n^{1/2+\delta}\log n/(nh_n^{Z_c}) \to 0$ for some arbitrarily small $\delta > 0$, then

$$\sqrt{n}(\hat{\theta}-\theta) \Rightarrow N(0, M_\theta^{-1}\Omega M_\theta^{-1}) \tag{A.17}$$

where $\Omega = E[\delta_i\delta_i']$.

The proof of the theorem will be carried out in a series of steps. The first step is to derive a linear representation for an estimator which solves the infeasible first order condition that one obtains when replacing $\hat{f}_i$ with f_i.

Lemma A.2 *Suppose $\hat{\theta} \xrightarrow{p} \theta_0$ and it can be established that $\hat{\theta}$ solves the asymptotic first order condition:*

$$\frac{1}{n}\sum_{i=1}^n \tau_{ni}\frac{\hat{\hbar}_i(\hat{\theta})}{f_i} = o_p(n^{-\ell}) \tag{A.18}$$

where $\ell \in [1/4, 1/2]$. Then under Assumptions B1-B4,B6,B7 and equation (A.1), $\hat{\theta}$ has the following linear representation:

$$\hat{\theta}-\theta_0 = -M_\theta^{-1}\frac{1}{n}\sum_{i=1}^n M_\kappa\psi_i + \tau_{ni}\frac{\hbar_i}{f_i} + o_p(n^{-\ell}) \tag{A.19}$$

Proof: We decompose the left hand side of (A.18) as:

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^n \tau_{ni}\frac{\hat{\hbar}_i(\hat{\theta})}{f_i} &= \frac{1}{n}\sum_{i=1}^n \tau_{ni}\frac{\hbar_i}{f_i} && \text{(A.20)}\\
&+ \frac{1}{n}\sum_{i=1}^n \tau_{ni}\frac{\hat{\hbar}_i(\hat{\theta})}{f_i} - E\left[\tau_{ni}\frac{\hbar_i(\hat{\kappa},\hat{\theta})}{f_i}\right] - \tau_{ni}\frac{\hbar_i}{f_i} && \text{(A.21)}\\
&+ E\left[\tau_{ni}\frac{\hbar_i(\hat{\kappa},\hat{\theta})}{f_i}\right] && \text{(A.22)}
\end{aligned}$$

We first show that the term in (A.21) is $o_p(n^{-1/2})$. We note that by the consistency of $\hat{\theta}$ and $\hat{\kappa}$, it will suffice to show that for a sequence of numbers ϵ_n converging to zero slowly enough,

$$\sup_{\|\theta-\theta_0\|<\epsilon_n, \|\kappa-\kappa_0\|<\epsilon_n}\left\|\frac{1}{n}\sum_{i=1}^n \tau_{ni}\frac{\hbar_i(\kappa,\theta)}{f_i} - E\left[\tau_{ni}\frac{\hbar_i(\kappa,\theta)}{f_i}\right] - \tau_{ni}\frac{\hbar_i}{f_i}\right\| = o_p(n^{-1/2}) \tag{A.23}$$

This follows from Assumptions B6 and B7 using Lemma 2.17 in Pakes and Pollard(1989). We next expand the term in (A.22) around κ_0, θ_0. Noting that $E\left[\tau_{ni}\frac{\hbar_i}{f_i}\right] = o(n^{-1/2})$ by Assumptions AA3 and B4.2, this yields:

$$\nabla_\kappa E\left[\tau_{ni}\frac{\hbar_i(\kappa^*,\theta^*)}{f_i}\right](\hat{\kappa}-\kappa_0) + \nabla_\theta E\left[\tau_{ni}\frac{\hbar_i(\kappa^*,\theta^*)}{f_i}\right](\hat{\theta}-\theta_0) + o(n^{-1/2})$$

where here κ^*, θ^* denote intermediate values. Thus we now have that (A.18) can be expressed as:

$$\begin{aligned}
\hat{\theta}-\theta_0 &= \left(-\nabla_\theta E\left[\tau_{ni}\frac{\hbar_i(\kappa^*,\theta^*)}{f_i}\right]\right)^{-1}\\
&\times \left(\frac{1}{n}\sum_{i=1}^n \tau_{ni}\frac{\hbar_i}{f_i} + \nabla_\kappa E\left[\tau_{ni}\frac{\hbar_i(\kappa^*,\theta^*)}{f_i}\right](\hat{\kappa}-\kappa_0)\right) + o_p(n^{-\ell}) && \text{(A.24)}
\end{aligned}$$

Note that $\left(-\nabla_\theta E\left[\tau_{ni}\frac{\hbar_i(\kappa^*,\theta^*)}{f_i}\right]\right)^{-1}$ can be expressed as $(M_\theta + o_p(1))^{-1}$ by Assumptions AA3, AA4.1, AA6.2, B3 and the consistency of $\hat{\theta}$. By the fact that $\frac{1}{n}\sum_{i=1}^n \tau_{ni}\frac{\hbar_i}{f_i}$ and $\hat{\kappa}-\kappa_0$ are both $O_p(n^{-1/2})$, it follows by Assumptions B3, B7 that we can express $\nabla_\kappa E\left[\tau_{ni}\frac{\hbar_i(\kappa^*,\theta^*)}{f_i}\right]$ as $(M_\kappa + o_p(1))$. This completes the proof. ■

An immediate corollary to this lemma is that the density weighted extremum estimator is fourth-root consistent.

Corollary A.1 *Under the Assumptions stated in Lemma A.2, along with Assumption B5,B8,B9 if h_n satisfies $\sqrt{n}h_n^p \to 0$, and $n^{1/2+\delta}\log n/(nh_n^{Z_c}) \to 0$ for some arbitrarily small $\delta > 0$, then*

$$\hat{\theta} - \theta_0 = o_p(n^{-1/4}) \tag{A.25}$$

Proof: By the restriction on h_n and Assumption B.5, Lemma 8.10 in Newey and MacFadden(1994) implies that $\|\hat{f}_i^{-1} - f_i^{-1}\|_\infty = o_p(n^{-1/4})$. By Assumption B6, we have

$$E\left[\sup_{\kappa\in\mathcal{A},\|\theta-\theta_0\|<\epsilon} \|\tau_{ni}\hbar_i(\kappa,\theta)\|\right] < \infty$$

, so we can replace $\hat{f}_i$ with f_i in (A.16) and the remainder term is $o_p(n^{-1/4})$. The corollary then follows from Lemma A.2.

A second corollary from Lemma A.2 is that by replacing $\hat{\kappa}, \hat{\theta}$ with κ_0, θ_0 in (A.18), the resulting remainder term is $o_p(n^{-1/4})$.

Corollary A.2 *Under the conditions in Corollary A.1,*

$$\frac{1}{n}\sum_{i=1}^n \tau_{ni}\frac{\hat{\hbar}_i(\hat{\theta}) - \hbar_i}{f_i} = o_p(n^{-1/4}) \tag{A.26}$$

Proof: We add and subtract $E\left[\tau_{ni}\frac{\hat{\hbar}_i(\hat{\theta})}{f_i}\right]$ from the above summation. By Corollary 3.1 and (A.1), it will suffice to show that, for a sequence of numbers $\delta_n = O((\log n)n^{-1/4})$, we have

$$\sup_{\|\kappa-\kappa_0\|<\delta_n,\|\theta-\theta_0\|<\delta_n} \left\|\frac{1}{n}\sum_{i=1}^n \tau_{ni}\frac{\hbar_i(\kappa,\theta) - \hbar_i}{f_i} - E\left[\tau_{ni}\frac{\hbar_i(\kappa,\theta)}{f_i}\right]\right\| = o_p(n^{-1/4}) \tag{A.27}$$

and

$$E\left[\tau_{ni}\frac{\hat{\hbar}_i(\hat{\theta})}{f_i}\right] = o_p(n^{-1/4}) \tag{A.28}$$

Under Assumptions B6 and B7, (A.27) follows from Lemma 2.17 in Pakes and Pollard(1989). By a mean value expansion, under Assumptions B1 and B2, (A.28) follows by Corollary A.1 and (A.1). ■

The next step involves deriving an expression for the variance of the density weighted extremum estimator which arises from estimating the conditional density function of the special regressor.

Lemma A.3 *Under Assumptions B5,B8,B9, if the bandwidth h_n satisfies $\sqrt{n}h_n^p \to 0$, and $n^{1/2+\delta}\log n/(nh_n^{Z_c}) \to 0$ for some arbitrarily small $\delta > 0$, then*

$$\frac{1}{n}\sum_{i=1}^n \tau_{ni}\hbar_i\left(\frac{1}{\hat{f}_i} - \frac{1}{f_i}\right) = \frac{1}{n}\sum_{i=1}^n E\left[\tau_{ni}\frac{\hbar_i}{f_i}\Big|z_i\right] - \frac{E\left[\tau_{ni}\hbar_i|v_i,z_i\right]}{f_i} + o_p(n^{-1/2}) \tag{A.29}$$

Proof: We again work with the identity

$$\begin{aligned}
\frac{1}{\hat{f}_i} - \frac{1}{f_i} &= \frac{\hat{f}_{zi} - f_{zi}}{f_{vzi}} && \text{(A.30)}\\
&- \frac{f_{zi}(\hat{f}_{zvi} - f_{vzi})}{f_{vzi}^2} && \text{(A.31)}\\
&- \frac{(\hat{f}_{zi} - f_{zi})(\hat{f}_{zvi} - f_{vzi})}{f_{vzi}\hat{f}_{zvi}} && \text{(A.32)}\\
&+ \frac{f_{zi}(\hat{f}_{zvi} - f_{vzi})^2}{\hat{f}_{zvi}f_{vzi}^2} && \text{(A.33)}
\end{aligned}$$

By Assumption B5 and the conditions on the bandwidth, by Lemma 8.10 in Newey and McFadden(1994), we have $\|\hat{f}_{zi} - f_{zi}\|_\infty$ and $\|\hat{f}_{zvi} - f_{vzi}\|_\infty$ are both $O_p(n^{-1/4-\delta})$ for some $\delta > 0$. It will thus suffice to derive representations for

$$\frac{1}{n}\sum_{i=1}^{n} \tau_{ni}\hbar_i \frac{\hat{f}_{zi} - f_{zi}}{f_{vzi}} \tag{A.34}$$

and

$$\frac{1}{n}\sum_{i=1}^{n} \tau_{ni}\hbar_i \frac{f_{zi}(\hat{f}_{zvi} - f_{vzi})}{f_{vzi}^2} \tag{A.35}$$

Turning attention to the first of the above terms, we let $\bar{f}_{zi}$ denote the expected value of the kernel estimator $\hat{f}_{zi}$ and work with the decomposition

$$\frac{1}{n}\sum_{i=1}^{n} \tau_{ni}\hbar_i \frac{\hat{f}_{zi} - \bar{f}_{zi}}{f_{vzi}} \tag{A.36}$$

$$+ \quad \frac{1}{n}\sum_{i=1}^{n} \tau_{ni}\hbar_i \frac{\bar{f}_{zi} - f_{zi}}{f_{vzi}} \tag{A.37}$$

We note that by the definition of $\hat{f}_{zi}$, (A.36) can be expressed as a U-statistic of the form:

$$\frac{1}{n(n-1)}\sum_{i\neq j} \tau_{ni}\hbar_i \frac{K_1\left(\frac{z_j^{(c)}-z_i^{(c)}}{h_n}\right) I[z_j^{(d)} = z_i^{(d)}] - \bar{f}_{zi}}{h_n^{Z_c} f_{vzi}} \tag{A.38}$$

We apply a projection theorem (see Powell et al.(1989), for example) to derive a linear representation for this U-statistic. Let ζ_i denote the vector (y_i, x_i, v_i, z_i) and let $\mathbf{F}_n(\zeta_i, \zeta_j)$ denote the term inside the double summation. We first note that $E\left[\|\mathbf{F}_n(\zeta_i, \zeta_j)\|^2\right] = O(h_n^{-Z_c})$ which is $o(n)$ by the conditions on h_n. We also note that $E\left[\mathbf{F}_n(\zeta_i, \zeta_j)\Big|\zeta_i\right] = 0$. It follows by Lemma 3.1 in Powell et al.(1989) that it will suffice to derive a representation for $E\left[\mathbf{F}_n(\zeta_i, \zeta_j)\Big|\zeta_j\right]$. We first show that

$$\frac{1}{n}\sum_{j=1}^{n} E\left[\tau_{ni}\hbar_i \frac{K_1\left(\frac{z_j^{(c)}-z_i^{(c)}}{h_n}\right) I[z_j^{(d)} = z_i^{(d)}]}{h_n^{Z_c} f_{vzi}}\Bigg| z_j\right] = \frac{1}{n}\sum_{i=1}^{n} E\left[\tau_{ni}\frac{\hbar_i}{f_i}\Big| z_i\right] + o_p(n^{-1/2}) \tag{A.39}$$

To show (A.39), it will be notationally convenient to let $\Xi(z_i)$ denote $E\left[\tau_{ni}\frac{\hbar_i}{f_i}|z_i\right]$. We note that

$$E\left[\tau_{ni}\hbar_i \frac{K_1\left(\frac{z_j^{(c)}-z_i^{(c)}}{h_n}\right) I[z_j^{(d)} = z_i^{(d)}]}{h_n^{Z_c} f_{vzi}}\Bigg| z_j\right]$$

can be written as

$$\frac{1}{h_n^{Z_c}}\int \Xi(z_i^{(c)}, z_j^{(d)}) K_1\left(\frac{z_j^{(c)} - z_i^{(c)}}{h_n}\right) dz_i^{(c)}$$

A change of variables $u = \frac{z_j^{(c)}-z_i^{(c)}}{h_n}$ yields the following integral:

$$\int \Xi(z_j^{(c)} - uh_n, z_j^{(d)}) K_1(u)\, du \tag{A.40}$$

By Assumptions B8,B9 a p^{th} order Taylor series expansion of $\Xi(z_j^{(c)} - uh_n, z_j^{(d)})$ around $\Xi(z_j)$ implies that the above integral can be expressed as the sum of $\Xi(z_j)$ and a remainder term which is of the form

$$h_n^p \int \Xi^{(p)}(z_j^{(c)} - uh_n^*, z_j^{(d)}) u^p K(u) du$$

where here $\Xi^{(p)}$ denotes the vector of p^{th} order derivatives of Ξ, the vector u raised to the integer p denotes the vector of the same dimension as $\Xi^{(p)}$ whose individual components are of the form $u^{\jmath}$ for all $[\jmath] = p$, and h_n^* denotes an intermediate value between 0 and h_n. It follows by the dominated convergence theorem and the conditions on h_n that:

$$E\left[E\left[\tau_{ni}\hbar_i \frac{K_1\left(\frac{z_j^{(c)}-z_i^{(c)}}{h_n}\right) I[z_j^{(d)} = z_i^{(d)}]}{h_n^{Z_c} f_{vzi}}\Big| z_j\right] - \Xi(z_j)\right] = o_p(n^{-1/2})$$

We also note by the continuity and boundedness of Ξ, an application of the dominated convergence theorem to (A.40) implies that:

$$\int \Xi(z_j^{(c)} - uh_n, z_j^{(d)}) K_1(u)\, du - \Xi(z_j) \to 0$$

as $h_n \to 0$. Another application of the dominated convergence theorem implies that

$$E\left[\left\| \int \Xi(z_j^{(c)} - uh_n, z_j^{(d)}) K_1(u)\, du - \Xi(z_j)\right\|^2\right] \to 0$$

as $h_n \to 0$. Thus (A.39) follows from Chebyshev's inequality. To complete the linear representation of $E\left[\mathbf{F}_n(\zeta_i, \zeta_j)\Big|\zeta_j\right]$ we show that

$$E\left[\tau_{ni}\frac{\hbar_i}{f_{vzi}}\bar{f}_{zi}\right] = o_p(n^{-1/2}) \tag{A.41}$$

Note that $E\left[\tau_{ni}\frac{\hbar_i}{f_{vzi}} f_{zi}\right] = o(n^{-1/2})$ by Assumption B4.2. Note also that $\left\|E\left[\tau_{ni}\frac{\hbar_i}{f_{vzi}}(\bar{f}_{zi} - f_{zi})\right]\right\|$ is bounded above by

$$\|\bar{f}_{zi} - f_{zi}\|_\infty \cdot E\left[\left\|\tau_{ni}\frac{\hbar_i}{f_{vzi}}\right\|\right]$$

$\|\bar{f}_{zi} - f_{zi}\|_\infty = O(h_n^p)$ by Lemma 8.9 in Newey and MacFadden(1994), and $E\left[\left\|\tau_{ni}\frac{\hbar_i}{f_{vzi}}\right\|\right]$ is bounded for all n by assumption. The desired result follows by the conditions on the bandwidth. To complete the linear representation in (A.34) we show that

$$\frac{1}{n}\sum_{i=1}^{n}\tau_{ni}\frac{\hbar_i}{f_{vzi}}(\bar{f}_{zi} - f_{zi}) = o_p(n^{-1/2}) \tag{A.42}$$

By (A.41) it will suffice to show that

$$\frac{1}{n}\sum_{i=1}^{n}\tau_{ni}\frac{\hbar_i}{f_{vzi}}(\bar{f}_{zi} - f_{zi}) - E\left[\tau_{ni}\frac{\hbar_i}{f_{vzi}}(\bar{f}_{zi} - f_{zi})\right] = o_p(n^{-1/2}) \tag{A.43}$$

By Chebyshev's inequality it will suffice to establish the above relation by showing that

$$E\left[\left\|\tau_{ni}\frac{\hbar_i}{f_{vzi}}(\bar{f}_{zi} - f_{zi})\right\|^2\right] \to 0$$

This follows by an application of the dominated convergence theorem and the condition that $h_n \to 0$.

Using virtually identical arguments, we can show that (A.35) has the following linear representation:

$$\frac{1}{n}\sum_{i=1}^{n} E\left[\tau_{ni}\frac{\hbar_i}{f_i}\Big| v_i, z_i\right] + o_p(n^{-1/2}) \tag{A.44}$$

completing the proof of the lemma. ■

We can now easily derive the limiting distribution of the density weighted extremum estimator. We decompose the left hand side of the asymptotic first order condition:

$$\frac{1}{n}\sum_{i=1}^{n}\tau_{ni}\frac{\hat{\hbar}_i(\hat{\theta})}{\hat{f}_i}=o_p(n^{-1/2})$$

as

$$\frac{1}{n}\sum_{i=1}^{n}\tau_{ni}\frac{\hat{\hbar}_i(\hat{\theta})}{f_i}+$$

$$\frac{1}{n}\sum_{i=1}^{n}\tau_{ni}\hbar_i\left(\frac{1}{\hat{f}_i}-\frac{1}{f_i}\right)+$$

$$\frac{1}{n}\sum_{i=1}^{n}\tau_{ni}\left(\hat{\hbar}_i(\hat{\theta})-\hbar_i\right)\left(\frac{1}{\hat{f}_i}-\frac{1}{f_i}\right)$$

By the established result that $\|\hat{f}_i-f_i\|_\infty=o_p(n^{-1/4-\delta})$ for some $\delta>0$ and Corollary A.2 the third term in the decomposition is $o_p(n^{-1/2})$. By Lemma A.3, the second term in the decomposition has the representation:

$$\frac{1}{n}\sum_{i=1}^{n}E\left[\tau_{ni}\frac{\hbar_i}{f_i}\Big|z_i\right]-E\left[\tau_{ni}\frac{\hbar_i}{f_i}\Big|v_i,z_i\right]+o_p(n^{-1/2})$$

The linear representation of the density weighted extremum estimator follows from Lemma A.2. ∎

B General Theorem for Density Weighted Closed Form Estimators

In this section we derive the asymptotic properties of a general density weighted closed form estimator. We define the parameter of interest to be:

$$\theta_0=E\left[\frac{h_i}{f_i}+g_i\right]$$

with $h_i, h_i(\kappa), g_i, f_i$ denoting $h(y_i,v_i,x_i,z_i,\kappa_0), h(y_i,v_i,x_i,z_i,\kappa), g(y_i,v_i,x_i,z_i)$ and $f(v_i|z_i)$ respectively. We define the estimator as

$$\hat{\theta}=\frac{1}{n}\sum_{i=1}^{n}\tau_{ni}\frac{\hat{h}_i}{\hat{f}_i}+g_i$$

with $\hat{h}_i, \hat{f}_i$ denoting $h_i(\hat{\kappa})$ and $\hat{f}(v_i|z_i)$ respectively, and we again assume (A.1) holds. Our theorem for the asymptotic properties of $\hat{\theta}$ are based on the following Assumptions:

C1 $h_i(\kappa)$ is continuously differentiable for $\kappa\in\mathcal{A}$, a neighborhood of κ_0.

C2 $E\left[\sup_{\kappa\in\mathcal{A}}\left\|\frac{h_i(\kappa)}{f_i}\right\|\right]<\infty$

C3 The function $E\left[\frac{h_i(\kappa)}{f_i}\right]$ is continuous at κ_0.

C4 Let M_κ denote $E\left[\nabla_\kappa \frac{h_i(\kappa_0)}{f_i}\right]$, and let δ_{ni} denote the vector

$$M_\kappa \psi_i + \tau_{ni}\frac{h_i}{f_i} + g_i - E\left[\tau_{ni}\frac{h_i}{f_i}\right] - E\left[\tau_{ni}\frac{h_i}{f_i}\middle| v_i, z_i\right] + E\left[\tau_{ni}\frac{h_i}{f_i}\middle| z_i\right]$$

and let δ_i denote the vector

$$M_\kappa \psi_i + \frac{h_i}{f_i} + g_i - \theta_0 - E\left[\frac{h_i}{f_i}\middle| v_i, z_i\right] + E\left[\frac{h_i}{f_i}\middle| z_i\right]$$

then we assume that

C4.1 $E\left[\|\delta_{ni}\|^2\right] < \infty$ for all $n \in \mathbf{N}$.

C4.2 $\frac{1}{n}\sum_{i=1}^n \delta_i - \delta_{ni} = o_p(n^{-1/2})$

C5 Assumption B5 holds.

C6 Assumptions B8 and B9 hold.

C7 Assumptions AA6,B10 hold, or alternatively, in the case of bounded rectangular supports, AA6' holds.

We now state the theorem for the density weighted closed form estimator:

Theorem B.1 *Suppose Assumptions C1-C7 hold and the bandwidth h_n satisfies $\sqrt{n}h_n^p \to 0$, and $n^{1/2+\delta}\log n/(nh_n^{2Z_c}) \to 0$ for some aribtrarly small $\delta > 0$, then*

$$\sqrt{n}(\hat{\theta} - \theta_0) \Rightarrow N(0, \Omega)$$

where $\Omega = E[\delta_i \delta_i']$.

Proof: We work with the relationship:

$$\begin{aligned} \hat{\theta} - \theta_0 &= \frac{1}{n}\sum_{i=1}^n \tau_{ni}\left(\frac{\hat{h}_i}{f_i} - \frac{h_i}{f_i}\right) && \text{(B.1)} \\ &+ \frac{1}{n}\sum_{i=1}^n \tau_{ni}\frac{h_i}{f_i} + g_i - E\left[\tau_{ni}\frac{h_i}{f_i} + g_i\right] && \text{(B.2)} \\ &+ E\left[\tau_{ni}\frac{h_i}{f_i} + g_i\right] - \theta_0 && \text{(B.3)} \end{aligned}$$

We note that the last term is $o(n^{-1/2})$ by Assumption C4.2. We thus focus attention on the first term. The difference in ratios can be linearized as before, yielding the terms:

$$\frac{1}{n}\sum_{i=1}^n \tau_{ni}\frac{\hat{h}_i - h_i}{f_i} - \quad \text{(B.4)}$$

$$\frac{1}{n}\sum_{i=1}^n \tau_{ni}\frac{h_i}{f_i^2}(\hat{f}_i - f_i) + R_n \quad \text{(B.5)}$$

The remainder term is of order $\frac{1}{n}\sum_{i=1}^n \tau_{ni}((\hat{h}_i - h_i)^2 + (\hat{f}_i - f_i)^2)$ and is $o_p(n^{-1/2})$ by Assumptions C1,C5,C6,C7 and the conditions on h_n, (which imply the fourth root consistency of the kernel density estimator) and (A.1). We derive a linear representation for (B.4). A mean value expansion of $\hat{h}_i$ around h_i implies we can express (B.4) as:

$$\left(\frac{1}{n}\sum_{i=1}^n \tau_{ni}\frac{h^*_{\kappa i}}{f_i}\right)(\hat{\kappa} - \kappa_0) \quad \text{(B.6)}$$

where $h_{\kappa i}$ denotes $\nabla_\kappa h(y_i, v_i, x_i, z_i, \kappa^*)$, with κ^* denoting an intermediate value. By Assumptions C1, C2, C3, and the root-n consistency of $\hat{\kappa}$, we can express $\left(\frac{1}{n}\sum_{i=1}^{n}\tau_{ni}\frac{h^*_{\kappa i}}{f_i}\right)$ as $M_\kappa + o_p(1)$. It thus follows by (A.1) that (B.4) has the following linear representation:

$$\frac{1}{n}\sum_{i=1}^{n} M_\kappa \psi_i + o_p(n^{-1/2}) \tag{B.7}$$

Turning attention to (B.5), we again linearize the difference of ratios in $\hat{f}_i - f_i$ yielding the term

$$\frac{1}{n}\sum_{i=1}^{n}\tau_{ni}\frac{h_i}{f_i}\frac{\hat{f}_{zvi} - f_{vzi}}{f_{vzi}} - \frac{1}{n}\sum_{i=1}^{n}\tau_{ni}h_i\frac{\hat{f}_{zi} - f_{zi}}{f_{vzi}} \tag{B.8}$$

plus a remainder term, which is $o_p(n^{-1/2})$ by Assumptions C5,C6, and C7 and the conditions on the bandwidth h_n. Using identical arguments as in the previous section, (B.8) has the following linear representation:

$$\frac{1}{n}\sum_{i=1}^{n} E\left[\tau_{ni}\frac{h_i}{f_i}\middle| v_i, z_i\right] - E\left[\tau_{ni}\frac{h_i}{f_i}\middle| z_i\right] + o_p(n^{-1/2}) \tag{B.9}$$

Combining all our results we have the following linear representation for the density weighted closed form estimator:

$$\hat{\theta} - \theta_0 = \frac{1}{n}\sum_{i=1}^{n}\tau_{ni}\frac{h_i}{f_i} - E\left[\tau_{ni}\frac{h_i}{f_i}\right] + M_\kappa\psi_i - E\left[\tau_{ni}\frac{h_i}{f_i}\middle| v_i, z_i\right] + E\left[\tau_{ni}\frac{h_i}{f_i}\middle| z_i\right] + o_p(n^{-1/2}) \tag{B.10}$$

The conclusion of the theorem follows from Assumption C4.2 and an application of the central limit theorem. ■

C Specific Examples

In this section, we apply the general theorems of the previous sections to derive the limiting distributions of the specific estimation procedures proposed in the paper, as well as the regularity conditions under which the limiting theory holds. The results are derived under the trimming conditions AA6,B10, but we note by Lemma A.1 the results would also hold under AA6' if the regressor support is bounded and rectangular.

C.1 Asymptotics for Weighted Least Squares Estimator

Here we derive the limiting distribution of the weighted least squares estimator for the truncated and censored regression model estimators. The limiting distribution is based on the following assumptions:

WLS1 The vector α_0, β_0 lies in the interior of the $\mathcal{A}\times\mathcal{B}$, a compact subset of $\Re^{k+1}$, where $|\alpha|$ is bounded away from 0 on $\mathcal{A}$.

WLS2 The random vector (y_i, v_i, x_i) is i.i.d.

WLS3 The k dimensional vector x_i can be partitioned as $(x_i^{(c)}, x_i^{(d)})$, where the k_c dimensional vector $x_i^{(c)}$ is continuously distributed, and the k_d dimensional vector $x_i^{(d)}$ is discretely distributed.

WLS4 The error term e_i satisfies either $E[e_i x_i] = 0$ or the stronger condition $E[e_i|x_i] = 0$.

WLS5 Letting $f(v_i, x_i^{(c)}|x_i^{(d)})$ denote the joint density function of $v_i, x_i^{(c)}$ conditional on $x_i^{(d)}$, we assume this function is p times continuously differentiable in its arguments $v_i, x_i^{(c)}$ for all values of $x_i^{(d)}$.

WLS6 The trimming function $0 \le \tau_{ni} \le 1$ depends on x_i, v_i, n only and satisfies $\tau_{ni} \to 1$ as $n \to \infty$ for all v_i, x_i. Also,

WLS6.1 Both $\sup_{v_i,x_i}\tau_{ni}/f_i$ and $\sup_{v_i,x_i}\tau_{ni}/f_{xvi}$ are $o(n^\delta)$ for all $\delta > 0$, where f_i, f_{xvi} denote $f(v_i|x_i)$ and $f(v_i, x_i)$ respectively.

WLS6.2 For the vector $\hbar_i \equiv (\hbar_{1i}, \hbar'_{2i})'$ where

$$h_{1i} = e_i I[0 < \alpha_0 v_i + x_i'\beta_0 + e_i < k] w(x_i)(v_i + e_i/\alpha_0) \tag{C.1}$$

$$h_{2i} = e_i I[0 < \alpha_0 v_i + x_i'\beta_0 + e_i < k] w(x_i)\alpha_0^{-2} x_i \tag{C.2}$$

The trimming function satisfies:

$$E\left[\tau_{ni}\frac{\hbar_i}{f_i}\right] = o(n^{-1/2}) \tag{C.3}$$

WLS7 The matrix $E[w(x_i)x_i x_i']$ is positive definite.

WLS8 $E[w(x_i)^2\|x_i\|^2 e_i^2] < \infty$.

We now state the theorem characterizing the limiting distribution of the weighted least squares estimator $\hat{\theta} \equiv (\hat{\alpha}, \hat{\beta}')'$ of $\theta_0 \equiv (\alpha_0, \theta_0')'$.

Theorem C.1 *Assume the bandwidth h_n satisfies $\sqrt{n}h_n^p \to 0$, and $n^{1/2+\delta}\log n/(nh_n^{Z_c}) \to 0$ for some arbitrarily small $\delta > 0$, the kernel function satisfies Assumption B8, and that Assumptions WLS1-WLS9 hold. Defining the matrix*

$$M = \begin{bmatrix} M_{\alpha\alpha} & M_{\alpha\beta} \\ M_{\beta\alpha} & M_{\beta\beta} \end{bmatrix} \tag{C.4}$$

where

$$M_{\beta\beta} = 2k|\alpha_0|^{-3} E[w(x_i)x_i x_i'] \tag{C.5}$$

$$M_{\alpha\alpha} = 2k^3|\alpha_0|^5 \left(E[w(x_i)]/3 + 3/4(E[w(x_i)x_i']E[w(x_i)x_i x_i']^{-1}E[w(x_i)x_i]\right) \tag{C.6}$$

$$M_{\alpha\beta} = \alpha_0^{-2}(k/2E[(w(x_i)x_i x_i']^{-1}E[w(x_i x_i] - \beta_0) \tag{C.7}$$

and the vector $\hbar_i \equiv (\hbar_{1i}, \hbar'_{2i})'$ where

$$h_{1i} = e_i I[0 < \alpha_0 v_i + x_i'\beta_0 + e_i < k] w(x_i)(v_i + e_i/\alpha_0) \tag{C.8}$$

$$h_{2i} = e_i I[0 < \alpha_0 v_i + x_i'\beta_0 + e_i < k] w(x_i)\alpha_0^{-2} x_i \tag{C.9}$$

Finally, let

$$\Omega = E\left[f_i^{-2}\hbar_i\hbar_i'\right]$$

then

$$\sqrt{n}(\hat{\theta} - \theta_0) \Rightarrow N(0, M^{-1}\Omega M^{-1})$$

Proof of Theorem: We simply verify Assumptions AA1-AA6 and B1-B10 from the previous section. note for this estimator, there is no preliminary estimator of a finite dimensional parameter, so the assumptions regarding κ_0 can be ignored. Assumptions AA1,AA2 follow from Assumptions WLS1 and WLS2. Assumption AA3 follows by the proof of Equation (1.5) in the text. AA4.1 and AA4.2 follow from the facts that $(y_i - v_i\alpha - x_i'\beta)^2$ is bounded for $0 < y_i < k$ and the compactness of the parameter space, as well as the behavior of the density f_i near 0 implicit in Assumption WLS6. AA5 follows from Assumptions WLS5, and AA6 follows from WLS6 This shows consistency.

For asymptotic normality, B1 follows by the fact that the objective function is twice differentiable in the parameters of interest. B3 follows from Assumption WLS7, and B4 follows from Assumption WLS6.2 and WLS8. B5 follows from WLS5. B6 follows since the objective function is twice differentiable in the parameters of interest, making the function in the first order condition sufficiently smooth. This smoothness condition also suffices for B7 to hold. B8 follows by the assumption stated in the theorem, and B9 follows by Assumption WLS5 and the smoothness of the objective function. ■

C.2 Asymptotics for Instrumental Variables Estimator

The asymptotic properties of the two stage least squares estimator are based on the following assumptions in addition to the identification assumptions A1'-A5' in the text.

IV1 The random vector $(y_i, v_i, x_i' z_i')'$ is identically and independently distributed.

IV2 The Z dimensional vector z_i can be partitioned as $(z_i^{(c)}, z_i^{(d)})$, where the Z_c dimensional vector $z_i^{(c)}$ is continuously distributed, and the Z_d dimensional vector $z_i^{(d)}$ is discretely distributed.

IV3 The trimming function $0 \leq \tau_{ni} \leq 1$ depends on z_i, v_i, n only and satisfies $\tau_{ni} \to 1$ as $n \to \infty$ for all v_i, z_i. Also,

IV3.1 Both $\sup_{v_i,z_i} \tau_{ni}/f_i$ and $\sup_{v_i,z_i} \tau_{ni}/f_{vzi}$ are $o(n^\delta)$ for all $\delta > 0$, where f_i, f_{vzi} denote $f(v_i|z_i)$ and $f(v_i, z_i)$ respectively.

IV3.2 For the vectors

$$\begin{aligned} h_{1i} &= \frac{I[0 < y_i < k]}{f_i} - \mu_0 \\ h_{2i} &= \mu_0^{-2} \frac{2 y_i I[0 < y_i < k]}{f_i} - \alpha_0 \\ h_{3i} &= z_i (y_i^* - x_i' \beta_0) \end{aligned}$$

we have

$$E[\tau_{ni} h_{ji}] = o(n^{-1/2}) \quad j = 1, 2, 3$$

IV4 The error term e_i has finite fourth moment.

IV5 Letting $f(v_i, z_i^{(c)} | z_i^{(d)})$ denote the joint density function of $v_i, z_i^{(c)}$ conditional on $z_i^{(d)}$, we assume this function is p times continuously differentiable in its arguments $v_i, z_i^{(c)}$ for all values of $z_i^{(d)}$. Furthermore, we assume that $f(v_i, z_i^{(c)} | z_i^{(d)})$ is bounded away from 0 for all $z_i, v_i \in \mathcal{Z} \times \mathcal{V}$.

We now derive the limiting distribution of the two stage estimator. Our arguments are based on applying Theorem B.1, so we will be verifying Assumptions C1-C7. We first derive a linear representations for $\hat{\mu}$ and $\hat{\alpha}$, assuming that $\alpha_0 > 0$. As mentioned in the text, this assumption is not problematic as the sign of α_0 can be estimated at an exponential rate, as shown in Lewbel(1998, 2000). The following lemma characterizes the limiting distribution of the estimator $\hat{\mu}$.

Lemma C.1 *If Assumptions A1'-A5' and IV1-IV5 hold, and the bandwidth sequence satisfies* $\sqrt{n} h_n^p \to 0$, $n^{1/2+\delta} \log n / (n h_n^{Z_c}) \to 0$ *for some arbitrarily small* $\delta > 0$, *and the kernel function satisfies Assumption B8, then*

$$\hat{\mu} - \mu_0 = \frac{1}{n} \sum_{i=1}^{n} \psi_{\mu i} + o_p(n^{-1/2})$$

where $\mu_0 \equiv E\left[\frac{I[0<y_i<k]}{f_i}\right]$ *and*

$$\psi_{\mu i} = \frac{I[0 < y_i < k]}{f_i} - \mu_0 - E\left[\frac{I[0 < y_i < k]}{f_i}\middle| v_i, z_i\right] + E\left[\frac{I[0 < y_i < k]}{f_i}\middle| z_i\right]$$

Proof: The result follows directly from Theorem B.1, with $h_i = I[0 < y_i < k], g_i = 0$. ■

We can now derive a limiting representation for $\hat{\alpha}$.

Theorem C.2 *If Assumptions A1'-A5' and IV1-IV5 hold, and the bandwidth sequence satisfies $\sqrt{n}h_n^p \to 0$, and $n^{1/2+\delta}\log n/(nh_n^{Z_c}) \to 0$ for some arbitrarily small $\delta > 0$, and the kernel function satisfies Assumption B8, then*

$$\hat{\alpha} - \alpha_0 = \frac{1}{n}\sum_{i=1}^{n}\psi_{\alpha i} + o_p(n^{-1/2})$$

with

$$\begin{aligned}\psi_{\alpha i} &= \frac{\alpha_0}{\eta(k) - \eta(k^*}(\eta(k)\mu_0(k)^{-2}\psi_{\mu i}(k) + \mu_0(k)^{-1}\psi_{\eta i}(k) - \\ &= \frac{\alpha_0}{\eta(k) - \eta(k^*}(\eta(k^*)\mu_0(k^*)^{-2}\psi_{\mu i}(k^*) + \mu_0(k^*)^{-1}\psi_{\eta i}(k^*)\end{aligned}$$

Proof: We again apply theorem B.1. In this case $h_i = \mu_0^{-2}2y_i I[0 < y_i < k]$, $g_i = 0$, and the plugged in estimator is $\hat{\mu}$. Note that $E\left[\nabla_\mu \frac{h_i}{f_i}\right] = -2\mu_0^{-1}\alpha_0$. ■

With the established linear representations, we can now derive the limiting distribution of $\hat{\beta}$.

Theorem C.3 *Suppose Assumptions A1'-A5' and IV1-IV5 hold, and the bandwidth sequence satisfies $\sqrt{n}h_n^p \to 0$, and $n^{1/2+\delta}\log n/(nh_n^{Z_c}) \to 0$ for some arbitrarily small $\delta > 0$, and the kernel function satisfies Assumption B8. Define the following mean 0 vectors:*

$$\begin{aligned}\psi_{\beta_1 i} &= -\left(\mu_0^{-2}\frac{k}{\alpha_0}E[z_i x_i']\beta_0\right)\cdot\psi_{\mu i} \\ \psi_{\beta_2 i} &= -\left(\frac{1}{2\alpha_0^2}(k^2 E[z_i] - kE[z_i x_i']\beta_0\right)\cdot\mu_0^{-1}\cdot\psi_{\alpha i} \\ \psi_{\beta_3 i} &= \frac{\mu_0^{-1}z_i(y_i - v_i\alpha_0)I[0 < y_i < k]}{f_i} - z_i x_i'\beta_0 - E\left[\frac{\mu_0^{-1}z_i(y_i - v_i\alpha_0)I[0 < y_i < k]}{f_i}\middle| z_i, v_i\right] \\ &+ E\left[\frac{\mu_0^{-1}z_i(y_i - v_i\alpha_0)I[0 < y_i < k]}{f_i}\middle| z_i\right]\end{aligned}$$

and let

$$\psi_{\beta i} = \psi_{\beta_1 i} + \psi_{\beta_2 i} + \psi_{\beta_3 i}$$

and

$$\Omega_\beta = E\left[\psi_{\beta i}\psi_{\beta i}'\right]$$

Then we have

$$\sqrt{n}(\hat{\beta} - \beta_0) \Rightarrow N(0, \Delta\cdot\Omega_\beta\cdot\Delta')$$

Proof : Define $\hat{\Delta}$ as:

$$\hat{\Delta} = \left(\left(\frac{1}{n}\sum_{i=1}^{n}x_i z_i'\right)\left(\frac{1}{n}\sum_{i=1}^{n}z_i z_i'\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}z_i x_i'\right)\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}x_i z_i'\right)\left(\frac{1}{n}\sum_{i=1}^{n}z_i z_i'\right)^{-1}$$

and $\hat{y}_i^*$ as

$$\hat{y}_i^* = \hat{\mu}^{-1}\frac{(y_i - v_i\hat{\alpha})I[0 < y_i < k]}{\hat{f}_i}$$

And note we can write

$$\hat{\beta} - \beta_0 = \hat{\Delta} \frac{1}{n} \sum_{i=1}^{n} z_i(\hat{y}_i^* - x_i'\beta_0) \tag{C.10}$$

We first note that an application of the law of large numbers and Slutsky's theorem immediately implies that

$$\hat{\Delta} \xrightarrow{p} \Delta$$

To complete the proof we apply theorem B.1 to derive a linear representation for

$$\frac{1}{n} \sum_{i=1}^{n} z_i(\hat{y}_i^* - x_i'\beta_0)$$

In this context, $h_i = \mu_0^{-1} z_i(y_i - v_i\alpha_0) I[0 < y_i < k]$ and $g_i = -z_i x_i'\beta_0$. The preliminary estimators are $\hat{\mu}$ and $\hat{\alpha}$. We note that:

$$E\left[\nabla_\mu \frac{h_i}{f_i}\right] = -\left(\mu_0^{-2} \frac{k}{\alpha_0} E[z_i x_i']\beta_0\right)$$

and

$$E\left[\nabla_\alpha \frac{h_i}{f_i}\right] = -\left(\frac{1}{2\alpha_0^2}(k^2 E[z_i] - kE[z_i x_i']\beta_0\right) \cdot \mu_0^{-1}$$

Hence the limiting distribution follows from this linear representation, the convergence of $\hat{\Delta}$ to Δ, and Slutsky's theorem. ■

TABLE 1
Simulation Results for Truncated Regression Estimators
Design 1: Homoskedastic Truncated Normal Errors

	Slope				Intercept			
	Mean Bias	Med. Bias	RMSE	MAD	Mean Bias	Med. Bias	RMSE	MAD
100 obs.								
WLS	-0.0539	-0.0719	0.2114	0.1736	-0.1458	-0.1374	0.2323	0.1873
STLS	0.0089	-0.0026	0.1559	0.1239	0.0313	0.0632	0.3779	0.2995
PWD	0.0186	0.0374	0.3042	0.2381	-	-	-	-
MLE	-0.0121	-0.0113	0.1395	0.1123	0.3003	0.3101	0.3244	0.3010
200 obs.								
WLS	-0.0566	-0.0591	0.1495	0.1198	-0.1034	-0.1008	0.1644	0.1305
STLS	-0.0048	0.0013	0.1085	0.0856	0.0532	0.0738	0.2775	0.2180
PWD	0.0126	0.0098	0.2142	0.1702	-	-	-	-
MLE	-0.0056	-0.0050	0.1056	0.0821	0.3125	0.3101	0.3242	0.3125
400 obs.								
WLS	-0.0567	-0.0534	0.1192	0.0963	-0.0892	-0.0878	0.1251	0.1031
STLS	-0.0016	-0.0018	0.0749	0.0596	0.0306	0.0372	0.1754	0.1405
PWD	0.0061	0.0061	0.1452	0.1148	-	-	-	-
MLE	-0.0100	-0.0107	0.0720	0.0577	0.3147	0.3165	0.3201	0.3147

TABLE 2
Simulation Results for Truncated Regression Estimators
Design 2: Homoskedastic Chi Squared Errors

	Slope				Intercept			
	Mean Bias	Med. Bias	RMSE	MAD	Mean Bias	Med. Bias	RMSE	MAD
100 obs.								
WLS	-0.0614	-0.0598	0.3009	0.2403	-0.2675	-0.2666	0.3722	0.3089
STLS	-0.0052	-0.0066	0.3983	0.3200	-0.3148	-0.3389	0.5527	0.4499
PWD	0.0195	0.0398	0.4667	0.3621	-	-	-	-
MLE	0.1711	0.1686	0.2846	0.2294	0.6261	0.6250	0.6560	0.6263
200 obs.								
WLS	-0.1035	-0.0972	0.2233	0.1806	-0.2057	-0.1983	0.2733	0.2250
STLS	0.0101	0.0060	0.2794	0.2184	-0.3454	-0.3542	0.4556	0.3908
PWD	0.0052	0.0160	0.3094	0.2431	-	-	-	-
MLE	0.1551	0.1531	0.2298	0.1886	0.6126	0.6155	0.6287	0.6126
400 obs.								
WLS	-0.0965	-0.0942	0.1656	0.1342	-0.1649	-0.1591	0.2027	0.1728
STLS	0.0165	0.0182	0.1785	0.1422	-0.3773	-0.3763	0.4230	0.3807
PWD	0.0008	-0.0029	0.2190	0.1734	-	-	-	-
MLE	0.1593	0.1625	0.1952	0.1644	0.6281	0.6320	0.6359	0.6281

TABLE 3
Simulation Results for Truncated Regression Estimators
Design 3: Heteroskedastic Truncated Normal Errors

	Slope				Intercept			
	Mean Bias	Med. Bias	RMSE	MAD	Mean Bias	Med. Bias	RMSE	MAD
100 obs.								
WLS	0.0311	0.0144	0.2316	0.1874	-0.1728	-0.1634	0.2623	0.2122
STLS	0.0448	0.0488	0.2643	0.2131	0.0701	0.1243	0.5891	0.4606
PWD	0.1594	0.1585	0.3780	0.3000	-	-	-	-
MLE	0.1171	0.1182	0.1996	0.1622	0.3618	0.3659	0.3846	0.3618
200 obs.								
WLS	0.0428	0.0484	0.1603	0.1277	-0.1286	-0.1226	0.1897	0.1526
STLS	0.0256	0.0307	0.1830	0.1447	0.0779	0.1110	0.4341	0.3403
PWD	0.1553	0.1513	0.2920	0.2363	-	-	-	-
MLE	0.1171	0.1123	0.1616	0.1311	0.3649	0.3718	0.3783	0.3649
400 obs.								
WLS	0.0512	0.0520	0.1273	0.1020	-0.1113	-0.1126	0.1471	0.1222
STLS	0.0110	0.0247	0.1294	0.1045	0.0304	0.0572	0.3111	0.2522
PWD	0.1523	0.1522	0.2257	0.1849	-	-	-	-
MLE	0.1093	0.1109	0.1390	0.1176	0.3682	0.3699	0.3739	0.3682

TABLE 4
Simulation Results for Truncated Regression Estimators
Design 4: Heteroskedastic Chi Squared Errors

	Slope				Intercept			
	Mean Bias	Med. Bias	RMSE	MAD	Mean Bias	Med. Bias	RMSE	MAD
100 obs.								
WLS	0.0441	0.0200	0.3361	0.2667	-0.3385	-0.3445	0.4521	0.3769
STLS	-0.1094	-0.1414	0.4807	0.3873	-0.3927	-0.3859	0.6719	0.5388
PWD	0.1910	0.1439	0.5878	0.4482	-	-	-	-
MLE	-0.0131	-0.0119	0.2035	0.1623	0.5332	0.5465	0.5620	0.5350
200 obs.								
WLS	0.0238	0.0262	0.2221	0.1805	-0.2672	-0.2424	0.3381	0.2791
STLS	-0.1532	-0.1788	0.3524	0.2878	-0.4231	-0.4277	0.5373	0.4616
PWD	0.1694	0.1496	0.4102	0.3188	-	-	-	-
MLE	-0.0020	-0.0013	0.1452	0.1182	0.5274	0.5266	0.5401	0.5274
400 obs.								
WLS	-0.1035	-0.0972	0.2233	0.1806	-0.2057	-0.1983	0.2733	0.2250
STLS	-0.1366	-0.1312	0.2446	0.1999	-0.4494	-0.4532	0.5023	0.4526
PWD	0.1511	0.1176	0.3174	0.2455	-	-	-	-
MLE	-0.0039	0.0013	0.1006	0.0799	0.5325	0.5326	0.5392	0.5325

TABLE 5
Simulation Results for Truncated Regression Estimators
Design 5: Endogenous Truncated Normal Errors

	Slope				**Intercept**			
	Mean Bias	Med. Bias	RMSE	MAD	Mean Bias	Med. Bias	RMSE	MAD
100 obs.								
2SLS	-0.2896	-0.2857	0.3687	0.3075	-0.3316	-0.2961	0.5292	0.4180
STLS	0.4629	0.4780	0.5356	0.4747	0.0312	0.0438	0.2586	0.2055
PWD	0.4227	0.4326	0.5017	0.4409	-	-	-	-
MLE	0.3688	0.3684	0.3910	0.3696	0.2312	0.2319	0.2540	0.2319
200 obs.								
2SLS	-0.2567	-0.2488	0.3107	0.2673	-0.2700	-0.2524	0.3984	0.3204
STLS	0.4177	0.4238	0.4567	0.4186	0.0236	0.0457	0.1977	0.1545
PWD	0.4375	0.4361	0.4726	0.4384	-	-	-	-
MLE	0.3802	0.3881	0.3929	0.3802	0.2250	0.2237	0.2381	0.2250
400 obs.								
2SLS	-0.2407	-0.2354	0.2678	0.2419	-0.2753	-0.2645	0.3342	0.2846
STLS	0.4395	0.4406	0.4576	0.4395	0.0160	0.0193	0.1262	0.1019
PWD	0.4233	0.4309	0.4425	0.4233	-	-	-	-
MLE	0.3807	0.3816	0.3869	0.3807	0.2207	0.2185	0.2272	0.2207

TABLE 6
Simulation Results for Truncated Regression Estimators
Design 6: Endogenous Chi Squared Errors

	Slope				**Intercept**			
	Mean Bias	Med. Bias	RMSE	MAD	Mean Bias	Med. Bias	RMSE	MAD
100 obs.								
2SLS	-0.2735	-0.2750	0.3847	0.3159	-0.3503	-0.3197	0.5463	0.4236
STLS	0.4287	0.4445	0.5410	0.4647	-0.1803	-0.1600	0.3682	0.2816
PWD	0.4414	0.4752	0.5792	0.4982	-	-	-	-
MLE	0.3657	0.3744	0.4056	0.3687	0.3597	0.3560	0.3920	0.3607
200 obs.								
2SLS	-0.2764	-0.2784	0.3390	0.2921	-0.2823	-0.2765	0.4078	0.3280
STLS	0.4605	0.4670	0.5079	0.4657	-0.1243	-0.1141	0.2283	0.1781
PWD	0.4482	0.4603	0.5112	0.4576	-	-	-	-
MLE	0.3676	0.3688	0.3856	0.3684	0.3652	0.3664	0.3825	0.3652
400 obs.								
2SLS	-0.2616	-0.2629	0.2946	0.2635	-0.2456	-0.2313	0.3091	0.2598
STLS	0.4747	0.4903	0.4992	0.4750	-0.1536	-0.1479	0.2156	0.1737
PWD	0.4290	0.4243	0.4629	0.4303	-	-	-	-
MLE	0.3669	0.3697	0.3775	0.3669	0.3681	0.3703	0.3766	0.3681