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

SHAKEEB KHAN
Duke University

ARTHUR LEWBEL
Boston College

This paper provides a root-n consistent, asymptotically normal weighted least squares estimator of the coefficients in a truncated regression model. The distribution of the errors is unknown and permits general forms of unknown heteroskedasticity. Also provided is an instrumental variables based two-stage least squares estimator for this model, which can be used when some regressors are endogenous, mismeasured, or otherwise correlated with the errors. A simulation study indicates that the new estimators perform well in finite samples. Our limiting distribution theory includes a new asymptotic trimming result addressing the boundary bias in first-stage density estimation without knowledge of the support boundary.

1. INTRODUCTION

Many statistical and econometric data sets exhibit some form of truncation. In this paper we consider estimation of a truncated regression model, which corresponds to a regression model where the sample is selected on the basis of the dependent variable.[1] We define the truncated regression model within the latent variable framework. Let $\tilde{y}_i$, w_i, and e_i denote a latent dependent variable, a $J + 1$–dimensional vector of observed covariates (which may include a constant), and a random unobserved disturbance term. We have the following latent variable relationship:

$$\tilde{y}_i = w_i' \theta_0 + e_i.$$

309

For our truncated model the econometrician does not observe $\tilde{y}_i$, but the non-negative variable y_i, where

$$y_i = \tilde{y}_i \,|\, \tilde{y}_i \geq 0.$$

If $\tilde{y}_i$ were observed without truncation, then it could be linearly regressed on w_i using ordinary or two-stage least squares to estimate θ_0, which is the parameter of interest. However, this simple estimator cannot be used because our data are only sampled from individuals having $\tilde{y}_i$ positive.

Our method of obtaining identification makes relatively weak assumptions regarding e_i, but it assumes that we can estimate the population (as opposed to the truncated) distribution of the regressors w_i. One way this may be accomplished is by having two distinct data sets, one being a sample of y_i, w_i observations generated by the truncation model $y_i = \tilde{y}_i \,|\, \tilde{y}_i \geq 0$ and the other a possibly independent sample of just w_i observations that are drawn from the population distribution of w_i. The latter, augmented data would be used to construct an estimate of the population density of w_i, which is the feature of the untruncated population that our estimator requires. For example, $\tilde{y}_i$ could be an attribute of consumers or workers that we sample with truncation, and w_i could be a vector of demographic characteristics with a population distribution that can be estimated from census data. See, e.g., Devereux and Tripathi (2004) and Chen, Hong, and Tamer (2005)[2] for recent work on optimally combining primary and augmented data in semiparametric models.

Another sampling design that fits our framework is a design using data derived from a classical censored regression model, where the econometrician observes regressor w_i values for both censored and uncensored observations and can thus infer the population regressor distribution from the data. However, we emphasize that the truncated regression model considered here is more general than the censored regression model, because censored regression data also provide information on the probability of the variable $\tilde{y}_i$ being negative and hence the probability of censoring, whereas our estimators do not require that information. Our estimators do require either regressors or instruments that are uncorrelated with e_i, and they require one regressor to satisfy some additional conditional independence and support requirements.

Most parametric truncated regression models restrict e_i to be distributed independently of w_i and lie in a parametric family, so that θ_0 and nuisance parameters in the distribution of e_i could be estimated by maximum likelihood estimation or (nonlinear) least squares. These estimators are generally inconsistent if the distribution of e_i is misspecified, if e_i is correlated with w_i, or if conditional heteroskedasticity is present.

Semiparametric, or "distribution-free," estimators for truncated models have been proposed with various restrictions on e_i by researchers including Bhattacharya, Chernoff, and Yang (1983), Powell (1986), Newey (1987, 2001), Lee (1989, 1993), and Honoré and Powell (1994). With the exception of the estimator in Lee (1989), which converges at a rate of the cube root of the sample size,

these estimators converge at parametric rates and have asymptotic normal distributions. Attaining this rate is more difficult in the truncated model than it is for standard models. For example, Newey (2001) shows that attaining the parametric rate is not possible with only a conditional mean restriction on the disturbance term.

In this paper, two new estimators for the truncated regression model are proposed. The estimators are numerically simple, being equivalent to linear weighted least squares or weighted two-stage least squares, though the weights depend on an 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. Unlike the previously listed estimators, our estimator does not require conditional independence, conditional symmetry, or conditional mode restrictions on the errors. The estimators may also be applied to doubly truncated data.

Given instruments z_i that are uncorrelated with the latent errors e_i, the two-stage least squares estimator we propose 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 previously, which do not allow for any form of endogeneity.

The new estimators involve weighting the data by an estimate of the population probability density of one of the regressors. We provide the limiting distribution for a general class of density weighted estimators. This limiting distribution theory includes a new result on the use of asymptotic trimming to deal with issues regarding first-stage density estimation, specifically addressing the boundary bias without knowledge of the support boundary. It also encompasses the case where the density of w_i might be estimated using an auxiliary data set.

Turning attention to the notation we will be adopting in the rest of the paper, we decompose the regressor vector as $w_i = (v_i, x_i')'$ with v_i denoting a regressor that satisfies restrictions discussed subsequently and x_i denoting the J-dimensional vector of other regressors. Correspondingly we decompose the parameter vector as $\theta_0 = (\alpha_0, \beta_0')'$. With this notation the truncated regression model is

$$\tilde{y}_i = v_i \alpha_0 + x_i' \beta_0 + e_i, \tag{1.1}$$

$$y_i = \tilde{y}_i \mid \tilde{y}_i \geq 0. \tag{1.2}$$

There may also be a vector of instruments z_i that are uncorrelated with e_i. The primary data set consists of n observations of y_i, v_i, x_i, and possibly z_i.

We assume a fixed underlying or untruncated distribution for the random variables V, X, Z, e or equivalently for $V, X, Z, \tilde{Y}$. We will refer to this as the underlying, or untruncated, population and use E^* to denote expectations over this

population. Our estimator depends on knowing or estimating a conditional distribution of the regressors and instruments, V, X, Z, and so our augmented data generating process consists of draws from the underlying or untruncated distribution of V, X, Z. In what follows, we will let (v_i^*, x_i^*, z_i^*), $i = 1, 2, \ldots, n^*$ denote the independent and identically distributed (i.i.d.) draws from this distribution, with n^* denoting the sample size.

Our primary data generating process is draws of truncated data. These are draws, denoted by (v_i, x_i, z_i, e_i), from the joint distribution of V, X, Z, e, conditional on $v_i \alpha_0 + x_i' \beta_0 + e_i \geq 0$ (i.e., discarding draws where this inequality does not hold) and y_i defined by equation (1.2). We refer to these draws as coming from the truncated population, use E to denote expectations over this truncated distribution, and let n denote the sample size.

The rest of 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 discusses the asymptotic properties of the proposed estimators, first by establishing general asymptotic results concerning functions that satisfy a density weighted moment condition and then by applying the general results to the estimators at hand. Section 4 explores the finite-sample properties of the estimators by means of a Monte Carlo study, and Section 5 concludes. Details regarding the asymptotics of our estimators are provided in the Appendix.

2. IDENTIFICATION

2.1. Preliminary Results

Our identification results are based on conditions imposed on the relationships between e_i, x_i, v_i for the heteroskedastic truncated regression model and on the relationships between e_i, x_i, v_i, z_i for the endogenous truncated regression model, where z_i is a vector of instrumental variables.

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

$$E^*[ex] = 0, \tag{2.1}$$

is not generally sufficient to identify the coefficients in the truncated regression model. We make two additional assumptions for identification and estimation. These assumptions are analogous to those imposed in Lewbel (1998, 2000), though the identification and estimation results in those papers do not apply to truncated regression models. One such assumption is that the underlying distribution of e is conditionally independent of the one regressor v, or equivalently,

$$F_e^*(e | v, x) = F_e^*(e | x). \tag{2.2}$$

The other is that the underlying distribution of v is assumed to have a large support.

The conditional independence restriction in (2.2) is an example of what Powell (1994, Sect. 2.5) calls a strong exclusion restriction. He notes that it permits general forms of conditional heteroskedasticity. This exclusion assumption arises naturally in some economic models. For example, in demand models where e represents unobserved preference variation, prices satisfy the conditional independence condition if they are determined by supply, such as under constant returns to scale production. An analogous exclusion restriction in the endogenous setting can be interpreted as a form of exogeneity; e.g., Blundell and Powell (2003) show that $e, x | v, z \sim e, x | z$ is very closely related to their control function assumption.

The other assumption for identification is that v have large support. Assuming a regressor to have large or infinite support is common in the literature on semiparametric limited dependent variable models. Examples include Manski (1975, 1985) and Horowitz (1992) for heteroskedastic binary choice models and Han (1987) and Cavanagh and Sherman (1998) for homoskedastic transformation models.

Let $F_{ex}^*(e, x | \cdot)$ denote the underlying, untruncated joint distribution of (e, x) conditional on data $(\cdot)$, with support denoted $\Omega_{ex}(\cdot)$. Let $f^*(v | \cdot)$ denote the underlying, untruncated conditional density of an observation of v, conditional on an observation of $(\cdot)$. In the exogenous setting we condition on the regressors x, whereas in the endogenous model we condition on a vector of instruments z.

THEOREM 2.1. *Let θ be a vector of parameters and let $h(v, x, z, e, \theta)$ be any function*

$$\psi(\theta) = E^* \left[\frac{h(v, x, z, e, \theta)}{f^*(v | x)} \right]. \tag{2.3}$$

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

$$E^* \left[\frac{h(v, x, z, e, \theta)}{f^*(v | z)} \middle| z \right] = E^* \left[\int_L^K h(v, x, z, e, \theta) \, dv \middle| z \right]. \tag{2.4}$$

Proof.

$$E^* \left[g \frac{h}{f^*(v | z)} \middle| z \right] = E^* \left[\frac{E[h | v, z]}{f^*(v | z)} \middle| z \right] = \int_L^K \left(\frac{E[h | v, z]}{f^*(v | z)} f^*(v | z) \right) dv$$

$$= \int_L^K E[h | v, z] \, dv = \int_L^K \int_{\Omega_{ex}} h(v, x, z, e, \theta) \, dF_{ex}^*(e, x | z) \, dv$$

$$= \int_{\Omega_{ex}} \int_L^K h(v, x, z, e, \theta) \, dv \, dF_{ex}^*(e, x | z). \tag{2.5}$$

∎

An immediate implication of Theorem 2.1 is

$$\psi(\theta) = E^* \left[\frac{h(v, x, z, e, \theta)}{f^*(v|z)} \right] = E^* \left[\int_L^K h(v, x, z, e, \theta) \, dv \right]. \tag{2.6}$$

The usefulness of equation (2.4) or (2.6) 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 $\psi(\theta)$.

Taking $z = x$ yields the following corollary to Theorem 2.1, which will be useful for estimation of models in which the errors are uncorrelated with the regressors.

COROLLARY 2.1. *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*

$$E^* \left[\frac{h(v, x, e, \theta)}{f^*(v|x)} \right] = E^* \left[\int_L^K h(v, x, e, \theta) \, dv \right]. \tag{2.7}$$

To illustrate Theorem 2.1, consider as a special case the binary choice model $d = I(v + x'\beta_0 + e \geq 0)$ with data consisting of a sample of observations of d_i, v_i, x_i, z_i. Letting $h(v, x, z, e, \theta) = z[d - I(v \geq 0)]$ gives, by equation (2.6),

$$E^* \left[z \frac{d - I(v \geq 0)}{f^*(v|z)} \right] = E^* [z(x'\beta_0 + e)], \tag{2.8}$$

which, if $E^*[ze] = 0$, shows that β_0 in the binary choice model can be estimated by linearly regressing $[d_i - I(v_i \geq 0)]/f^*(v_i|z_i)$ on x_i using instruments z_i. This is the binary choice model identification result proposed in Lewbel (2000).

We will now apply Theorem 2.1 and its corollary to obtain identification results for truncated regression models.

2.2. Exogenous Truncated Regression Model Identification

Our identification for the truncated regression model with exogenous regressors and possibly heteroskedastic errors is based on the following assumptions.

Assumption A.1. Assume that the truncated data are draws v, x, e, y conditional on $\tilde{y}_i \geq 0$ as described by equations (1.1) and (1.2) with $\alpha_0 \neq 0$. The underlying, untruncated conditional distribution of v given x is absolutely continuous with respect to a Lebesgue measure with conditional density $f^*(v|x)$.

Assumption A.2. Let Ω denote the underlying, untruncated support of the distribution of an observation of (v, x). Let $F_e^*(e|v, x)$ denote the underlying, untruncated conditional distribution of an observation of e given an observa-

tion 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 underlying, untruncated 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 underlying, untruncated support of (x, e), $[I(\alpha_0 > 0)L - I(\alpha_0 < 0)K]\alpha_0 + x'\beta_0 + e < 0$. Let $\tilde{k}$ equal the largest number that satisfies the inequality $\tilde{k} \leq [I(\alpha_0 > 0)K - I(\alpha_0 < 0)L]\alpha_0 + x'\beta_0 + e$ for all (x, e) on the support of (x, e). Note that $\tilde{k} > 0$.

Assumption A.5. For some positive, bounded function $w(x)$ chosen by the econometrician, $E^*[exw(x)] = 0$ and $E^*[w(x)xx']$ exists and is nonsingular.

Assumption A.1 defines the truncated regression model and says that v has a continuous distribution. The assumptions do not require the distributions of e or x to be continuous; e.g., they 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.2 is the conditional (or partial) independence exclusion assumption discussed earlier. Assumptions A.3 and A.4 together imply that whatever value x, e take on, there exists some value of v that results in $\tilde{y} \geq 0$ and in this sense requires v to have a large support. Standard models like tobit have errors that can take on any value, which would 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 $x'\beta + e$, because if the observed spread of v values were not large, then the observed data might resemble data drawn from a process that violated Assumption A.4.

For Assumption A.5, the function $w(x)$ will be chosen for efficiency. If $E^*[e|x] = 0$, $w(x)$ can be almost any positive, bounded function. Under the weaker assumption $E^*[ex] = 0$, we can just let $w(x) \equiv 1$.

The truncation takes the form $y = \tilde{y}|\tilde{y} \geq 0$. It follows that for any function $h(y, x, e)$ and constant $k > 0$ the relationship of the truncated to untruncated expectation is

$$E[h(y, x, v, e)I(0 \leq y \leq k)] = \frac{E^*[h(\tilde{y}, x, v, e)I(0 \leq \tilde{y} \leq k)]}{prob(\tilde{y} \geq 0)}. \tag{2.9}$$

The following corollary to Theorem 2.1, along with equation (2.9), provides the main identification result that is the basis for our estimator of the heteroskedastic model.

COROLLARY 2.2. *Let Assumptions A.1–A.4 hold. Let $H(\tilde{y}, x, e)$ be any function that is differentiable in $\tilde{y}$. Let k be any constant that satisfies $0 \leq k \leq \tilde{k}$. Then*

$$E^* \left[\frac{\partial H(\tilde{y}, x, e)}{\partial \tilde{y}} \frac{I(0 \leq \tilde{y} \leq k)}{f^*(v|x)} \right] = E^* \left[\frac{H(k, x, e) - H(0, x, e)}{|\alpha_0|} \right] \tag{2.10}$$

provided that these expectations exist.

Proof. First apply Corollary 2.1 and then do a change of variables in the integration from v to $\tilde{y}$ to get

$$E^* \left[\frac{\partial H(\tilde{y}, x, e)}{\partial \tilde{y}} \frac{I(0 \leq \tilde{y} \leq k)}{f^*(v|x)} \right]$$

$$= E^* \left[\int_L^K \frac{\partial H[\tilde{y}(v, x, e), x, e]}{\partial \tilde{y}(v, x, e)} I[0 \leq \tilde{y}(v, x, e) \leq k] \, dv \right] \tag{2.11}$$

$$= E^* \left[\int_{L\alpha_0 + x'\beta_0 + e}^{K\alpha_0 + x'\beta_0 + e} \frac{\partial H(\tilde{y}, x, e)}{\partial \tilde{y}} I(0 \leq \tilde{y} \leq k) \, d\tilde{y}/\alpha_0 \right] \tag{2.12}$$

if $\alpha_0 > 0$ or

$$= -E^* \left[\int_{-K\alpha_0 + x'\beta_0 + e}^{-L\alpha_0 + x'\beta_0 + e} \frac{\partial H(\tilde{y}, x, e)}{\partial \tilde{y}} I(0 \leq \tilde{y} \leq k) \, d\tilde{y}/\alpha_0 \right] \tag{2.13}$$

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

$$= E^* \left[\int_0^k \frac{\partial H(\tilde{y}, x, e)}{\partial \tilde{y}} \, d\tilde{y}/|\alpha_0| \right], \tag{2.14}$$

which proves the result.   ∎

THEOREM 2.2. *Let Assumptions A.1–A.5 hold. Let k be any constant that satisfies $0 < k \leq \tilde{k}$. Define the function $\mu(\alpha, \beta)$ by*

$$\mu(\alpha, \beta) = E \left[\frac{(y - v\alpha - x'\beta)^2 \alpha^{-2} I[0 \leq y \leq k] w(x)}{f^*(v|x)} \right]. \tag{2.15}$$

Then

$$(\alpha_0, \beta_0) = \arg \min \mu(\alpha, \beta), \tag{2.16}$$

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 2.2. Define $h(y, x, v, e)$ in equation (2.9) by $h(y, x, v, e) = (y - v\alpha - x'\beta)^2 \alpha^{-2} w(x)/f^*(v|x)$. Equations (2.9) and (2.10) will be combined by defining H such that $[\partial H(\tilde{y}, x, e)/\partial \tilde{y}]/f^*(v|x) = h(\tilde{y}, x, v, e)$. Specifically by equations (2.9) and (2.15) and $v = (\tilde{y} - x'\beta_0 - e_i)/\alpha_0$ we obtain

$$
\mu(\alpha, \beta) = E^* \frac{\left[\dfrac{\left[\tilde{y}\left(\dfrac{1}{\alpha} - \dfrac{1}{\alpha_0} \right) + x'\left(\dfrac{\beta_0}{\alpha_0} - \dfrac{\beta}{\alpha} \right) + \dfrac{e}{\alpha_0} \right]^2 I(0 \le \tilde{y} \le k)\, w(x)}{f^*(v|x)} \right]}{prob(\tilde{y} \ge 0)}.
$$

(2.17)

Next apply Corollary 2.2, obtaining $H(\tilde{y}, x, e)$ by integrating $h(\tilde{y}, x, e)f^*(v|x)$ and making use of $E^*[exw(x)] = 0$ to get

$$
\mu(\alpha, \beta) prob(\tilde{y} \ge 0)
$$

$$
= \frac{1}{|\alpha_0|} E^* \left[\int_0^k \left[\tilde{y}\left(\frac{1}{\alpha} - \frac{1}{\alpha_0} \right) + x'\left(\frac{\beta_0}{\alpha_0} - \frac{\beta}{\alpha} \right) + \frac{e}{\alpha_0} \right]^2 w(x)\, d\tilde{y} \right] \quad (2.18)
$$

$$
= \frac{k^3 E^*[w(x)]}{3|\alpha_0|} \left(\frac{1}{\alpha} - \frac{1}{\alpha_0} \right)^2 + \frac{k^2}{|\alpha_0|}\left(\frac{1}{\alpha} - \frac{1}{\alpha_0} \right) E^*[w(x)x']\left(\frac{\beta_0}{\alpha_0} - \frac{\beta}{\alpha} \right)
$$

$$
+ \frac{k}{|\alpha_0|}\left(\frac{\beta_0}{\alpha_0} - \frac{\beta}{\alpha} \right)' E^*[w(x)xx']\left(\frac{\beta_0}{\alpha_0} - \frac{\beta}{\alpha} \right) + \frac{k E^*[w(x)e^2]}{|\alpha_0|\alpha_0^2}. \quad (2.19)
$$

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

$$
\left(\frac{\beta}{\alpha} - \frac{\beta_0}{\alpha_0} \right) = \frac{k}{2}\left(\frac{1}{\alpha} - \frac{1}{\alpha_0} \right) E^*[w(x)xx']^{-1} E^*[w(x)x], \quad (2.20)
$$

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

$$
\frac{\partial^2 \mu(\alpha, \beta)}{\partial\beta\partial\beta'} = \frac{2k}{|\alpha_0|\alpha^2} E^*[w(x)xx'] \quad (2.21)
$$

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

$$
\mu[\alpha, \beta(\alpha)] prob(\tilde{y} \ge 0)
$$

$$
= \frac{k^3}{|\alpha_0|}\left(\frac{1}{\alpha} - \frac{1}{\alpha_0} \right)^2
$$

$$
\times \left(\frac{E^*[w(x)]}{3} + \frac{3}{4} E^*[w(x)x'] E^*[w(x)xx']^{-1} E^*[w(x)x] \right)
$$

$$
+ \frac{k E^*[w(x)e^2]}{|\alpha_0|\alpha_0^2}. \quad (2.22)
$$

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

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

(2.23)

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

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

$$\left. + k\beta' E^*[w(x)xx']\beta_0 + kE^*[w(x)e^2]\right),$$

(2.24)

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

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

(2.25)

Also the second-order condition

$$\frac{d^2\mu[\alpha_0, \beta(\alpha_0)]}{\partial\alpha^2}$$

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

(2.26)

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

Theorem 2.2 shows that α_0 and β_0 are identified and can be estimated by a linear weighted least squares regression of y^* on x. The variable y^* depends on the untruncated population density $f^*(v|z)$, which we will estimate using a kernel density estimator.

2.3. Endogenous Truncated Regression Model Identification

Now consider identification of the truncated regression model with endogenous or mismeasured regressors in addition to heteroskedastic errors. Theorem 2.3, which follows, describes instrumental variables based identification of this model, where we assume $E^*[ez] = 0$ (i.e., the standard assumption regarding instruments in two-stage least squares regressions) and the underlying, untruncated conditional independence

$$F_{ex}^*(e, x|v, z) = F_{ex}^*(e, x|z).$$

(2.27)

Assumption A.1$'$. Assume that the truncated data are draws v, x, z, e, y conditional on $\tilde{y}_i \geq 0$ as described by equations (1.1) and (1.2) with $\alpha_0 \neq 0$. The underlying, untruncated conditional distribution of v given x is absolutely continuous with respect to a Lebesgue measure with conditional density $f^*(v|z)$.

Assumption A.2$'$. Let Ω denote the underlying, untruncated support of the distribution of an observation of (v, z). Let $F_e^*(e, x|v, z)$ denote the underlying, untruncated 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 underlying, untruncated 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$'$. This is the same as Assumption A.4.

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 $\Sigma_{xz}, \Sigma_{zz}, \Delta$, and y^* by $\Sigma_{xz} = E^*[xz']$, $\Sigma_{zz} = E^*[zz']$,

$$\Delta = (\Sigma_{xz} \Sigma_{zz}^{-1} \Sigma_{xz}')^{-1} \Sigma_{xz} \Sigma_{zz}^{-1}, \tag{2.28}$$

$$y^* = \frac{(y - v\alpha_0) I(0 \leq y \leq k)/f^*(v|z)}{E[I(0 \leq y \leq k)/f^*(v|z)]}. \tag{2.29}$$

THEOREM 2.3. *Let Assumptions A.1$'$–A.5$'$ hold. Let k be any constant that satisfies $0 < k \leq \tilde{k}$. Then $E[zy^*] = E^*[zx']\beta_0$, and so*

$$\beta_0 = \Delta E[zy^*]. \tag{2.30}$$

Proof of Theorem 2.3. Let $H(\tilde{y}, x, z, e)$ be any function that is differentiable in $\tilde{y}$. If Assumptions A.1$'$–A.4$'$ hold then

$$E^*\left[\frac{\partial H(\tilde{y}, x, z, e)}{\partial \tilde{y}} \frac{I(0 \leq \tilde{y} \leq k)}{f^*(v|z)} \right] = E^*\left[\frac{H(k, x, z, e) - H(0, x, z, e)}{|\alpha_0|} \right] \tag{2.31}$$

provided these expectations exist. The proof follows the same steps as the proof of Corollary 2.2. Similarly, the analogue to equation (2.9) is

$$E[h(y, x, v, z, e)I(0 \leq y \leq k)] = \frac{E^*[h(\tilde{y}, x, v, z, e)I(0 \leq \tilde{y} \leq k)]}{prob(\tilde{y} \geq 0)}, \tag{2.32}$$

and these two equations are combined by defining H such that $[\partial H(\tilde{y}, x, z, e)/ \partial \tilde{y}]/f^*(v|z) = h(\tilde{y}, x, v, z, e)$.

Applying (2.31) and (2.32) with $h(y,x,z,e) = 1/f^*(v|z)$, which makes $H(\tilde{y},x,z,e,\theta) = \tilde{y}$, gives

$$E[I(0 \leq y \leq k)/f^*(v|z)] = \frac{k}{|\alpha_0| prob(\tilde{y} \geq 0)}, \tag{2.33}$$

and applying (2.31) and (2.32) with $h(y,x,z,e) = z(y - v\alpha_0)/f^*(v|z)$ gives

$$E[z(y - v\alpha_0)I(0 \leq y \leq k)/f^*(v|z)] = \frac{E^*[z(\tilde{y} - v\alpha_0)I(0 \leq \tilde{y} \leq k)/f^*(v|z)]}{prob(\tilde{y} \geq 0)}$$

$$= \frac{E^*[z(x'\beta_0 + e)I(0 \leq \tilde{y} \leq k)/f^*(v|z)]}{prob(\tilde{y} \geq 0)}$$

$$= \frac{k[E^*[zx']\beta + E^*[ze]]}{|\alpha_0| prob(\tilde{y} \geq 0)},$$

where the last equality applies (2.31) with $H(\tilde{y},x,z,e,\theta) = z(x'\beta_0 + e)\tilde{y}$. With $E^*[ze] = 0$, it follows that $E^*[zy^*] = E^*[zx']\beta_0$. ∎

We next provide an identification result for α_0. Define $\eta(k)$ by

$$\eta(k) = \left(\frac{2v\, I(0 \leq y \leq k)/f^*(v|z)}{E[I(0 \leq y \leq k)/f^*(v|z)]} \right). \tag{2.34}$$

COROLLARY 2.3. *Let Assumptions A.1′–A.5′ hold. Let k and k^* be any constants that satisfy $0 < k^* < k \leq \tilde{k}$. Then*

$$\alpha_0 = \frac{k - k^*}{\eta(k) - \eta(k^*)}. \tag{2.35}$$

Proof of Corollary 2.3.

$$E[vI(0 \leq y \leq k)/f^*(v|z)]$$

$$= E[\alpha_0^{-1}(y - x'\beta - e)I(0 \leq \tilde{y} \leq k)/f^*(v|z)] \tag{2.36}$$

$$= \left(\frac{k^2}{2\alpha_0|\alpha_0|} - \frac{kE^*[x'\beta - e]}{\alpha_0|\alpha_0|} \right) \Big/ prob(\tilde{y} \geq 0), \tag{2.37}$$

where the second equality applies (2.31) and (2.32) with $h(y,x,v,z,e) = \alpha_0^{-1}(y - x'\beta - e)/f^*(v|z)$, which implies $H(\tilde{y},x,z,e) = \alpha_0^{-1}[(\tilde{y}^2/2) - \tilde{y}(x'\beta + e)]$. Similarly, $E[I(0 \leq \tilde{y} \leq k)/f^*(v|z)]$ is given by equation (2.33), and so $\eta(k) = (k/\alpha_0) - 2E^*[x'\beta - e]$ and equation (2.35) follows immediately. ∎

Equation (2.30) in Theorem 2.3 shows that β_0 is identified and can be esti-
mated 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 we will estimate using
a kernel density estimator, and it also depends on α_0. Equation (2.35) can be
used to construct an estimator for α_0. A disadvantage of equation (2.35) is that
it requires choosing a constant k^* in addition to k. If the assumptions of Theo-
rem 2.3 hold for $z = x$, then either the weighted least squares estimator of Theo-
rem 2.2 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. Identification was achieved
in one step in the exogenous setting because the coefficient of the regressor v_i
(which equals the intercept after dividing by v_i) is obtained along with the other
parameters by minimizing a least squares based criterion, which is valid under
exogeneity. In contrast, in the endogenous case, identification of coefficients
other than the coefficient of v are obtained by instrumenting. In this case, the
coefficient of v_i cannot be obtained in the same way as the others, because we
must integrate over v_i to obtain properly weighted instruments.

3. ESTIMATION

In this section we provide descriptions and limiting distributions of the weighted
and two-stage least squares estimators based on the identification results in the
previous section.

3.1. Weighted Least Squares Estimation of the
Heteroskedastic Truncated Regression Model

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, i.e., $f^*(v|u) = f^*(v|x)$,
where no element of u equals a deterministic function of other elements of u.
This construction of u is employed because $f^*(v|x)$ will be estimated as $\hat{f}^*(v|u)$
using a kernel density estimator. Also, if v is known to be conditionally inde-
pendent of some elements of x, then this construction allows u to exclude those
elements of x, thereby reducing the dimension of this conditional density esti-
mation. As mentioned previously, $f^*(v|x)$ and hence $f^*(v|u)$ refers to the under-
lying population density before truncation, and consequently its kernel density
estimator requires availability of an augmented data set on regressor observa-
tions (either regressor observations for the truncated data, as in a censored regres-
sion data set, or data from another source such as a census). As a result $\hat{f}^*(v|u)$
is estimated from the augmented data set but is evaluated at observations that
are drawn from truncated data.

To deal with boundary bias issues or vanishing marginal densities that
arise in kernel density estimation, we incorporate a "trimming" function into
the estimator procedure. A novelty of the asymptotic trimming we apply to

address boundary bias 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 2.2 is

$$(\hat{\alpha}, \hat{\beta}) = \arg\min \frac{1}{n} \sum_{i=1}^{n} \tau_{ni} \frac{(y_i - v_i\alpha - x_i'\beta)^2 \, \alpha^{-2} \, I(0 \le y_i \le k) \, w(x_i)}{\hat{f}^*(v_i|u_i)} \tag{3.1}$$

for some chosen scalar k, weighting function $w(x)$, and trimming function $\tau_{ni} \equiv \tau(x_i, n)$, with properties that are detailed in the Appendix. The n observations in equation (3.1) are of truncated data, whereas the function $\hat{f}^*(v|u)$ is constructed from an augmented data set, with sample size n^*. The resulting limiting distribution depends on the asymptotic relative sample sizes of these data sets.

The function $w(x)$ is chosen by the researcher and so may be selected to ensure that the assumptions of Theorem 2.2 are likely to hold in a particular application, e.g., taking $w(x) = 1$ if economic theory suggests only that $E^*[ex] = 0$. Alternatively, if $E^*[e|x] = 0$, then $w(x)$ may be chosen to maximize an estimated measure of efficiency. Similarly, the truncation point k may be chosen either by a sensible rule of thumb based on not discarding much data or more formally to maximize an efficiency measure. In our simulations we simply take $w(x) = 1$ and set the right truncation point k to be the sample 75th percentile of y_i.

Closed form expressions for $(\hat{\alpha}, \hat{\beta})$ can be obtained as follows. Let $a = 1/\alpha$ and $b = -\beta/\alpha$. Then $(y - v\alpha - x'\beta)^2\alpha^{-2}I(0 \le y) = (v - ya - x'b)^2I(0 \le y)$, and so from equation (2.10) $\hat{\alpha} = 1/\hat{a}$ and $\hat{\beta} = -\hat{b}/\hat{a}$ where

$$(\hat{a}, \hat{b}) = \arg\min \frac{1}{n} \sum_{i=1}^{n} \tau_{ni} \cdot (v_i - y_i a - x_i'b)^2 \frac{I(0 \le y_i \le k) \, w(x_i)}{\hat{f}^*(v_i|u_i)} \tag{3.2}$$

and (3.2) is just a linear weighted least squares regression of v on y and x, using weights $I(0 \le y_i \le k)w(x_i)/\hat{f}^*(v_i|u_i)$. Unlike an ordinary least squares regression, where weighting only affects efficiency, in equations (3.1) and (3.2) the weights are functions of the regress and are required for consistency.

The following theorem characterizes the limiting distribution of this estimator. Asymptotic theory corresponds to the primary (i.e., truncated) data set, with sample size n going to infinity. To allow for an augmented data set to be used for density estimation, with n^* denoting its sample size, we let $c_p = \lim_{n\to\infty}(n/n^*)$, with $0 \le c_p < \infty$.

The conditions upon which the theorem is based, in addition to its proof, can be found in the Appendix.

THEOREM 3.1. *Define the* $(J + 1) \times (J + 1)$ *matrix* **Q** *as*

$$\mathbf{Q} = \begin{bmatrix} \alpha_0^2 & \mathbf{0}_{1 \times J} \\ \alpha_0 \beta_0 & \alpha_0 \mathbf{I}_{J \times J} \end{bmatrix}, \tag{3.3}$$

where $\mathbf{0}_{1 \times J}$ *refers to a* $(1 \times J)$ *vector of 0's and* $\mathbf{I}_{J \times J}$ *refers to a* $(J \times J)$ *identity matrix, and define the* $(J + 1) \times (J + 1)$ *matrix* H_0 *as*

$$H_0 = \frac{k}{|\alpha_0|} \begin{bmatrix} E[w(x_i)] \dfrac{k^2}{3} & E[w(x_i)x_i'] \dfrac{k}{2} \\ E[w(x_i)x_i] \dfrac{k}{2} & E[w(x_i)x_i x_i'] \end{bmatrix}. \tag{3.4}$$

Also, define the $(J + 1) \times 1$ *vector* $\hbar_i \equiv (\hbar_{1i}, \hbar_{2i}')'$ *as*

$$\hbar_{1i} = I[0 \le y_i \le k] w(x_i) \alpha_0^{-1} y_i (v_i - y_i - x_i' \beta_0),$$

$$\hbar_{2i} = I[0 \le y_i \le k] w(x_i) \alpha_0^{-1} x_i (v_i - y_i - x_i' \beta_0).$$

Let $f(x_i)$ *denote the density function of X from the truncated sample evaluated at* x_i. *Similarly, let* $f(v_i, x_i)$ *denote the joint density function of V, X from the truncated sample evaluated at* v_i, x_i. *Let* $r_{xi} = f(x_i)/f^*(x_i)$ *denote the ratio of the density functions of X from the truncated sample over the density function of X from the underlying (untruncated) sample, evaluated at draws from the truncated sample,* x_i. *Furthermore, let* $r_{vxi} = f(v_i, x_i)/f^*(v_i, x_i)$ *denote the analogous ratio (truncated over underlying) of joint density functions of X, V. Furthermore, we let* $\varphi_a(\cdot) = E[(r_{xi} \hbar_i / f_i) | x_i = \cdot]$ *and* $\varphi_b(\cdot, \cdot) = E[(r_{vxi} \hbar_i / f_i) | v_i = \cdot, x_i = \cdot]$, *where* f_i *is shorthand notation for* $f^*(v_i | u_i)$. *Define the "score" vector*

$$\delta_i = \frac{\hbar_i}{f_i} - c_p(\varphi_b(v_i^*, x_i^*) - \varphi_a(x_i^*)). \tag{3.5}$$

Note that δ_i *depends on values drawn from both the truncated and untruncated distributions. Finally, set* $\Omega = \tilde{E}[\delta_i \delta_i']$, *where the operator* $\tilde{E}[\cdot]$ *denotes that the expectation is taken over both the truncated and underlying distributions. Then*

$$\sqrt{n}(\hat{\theta} - \theta_0) \Rightarrow N(0, \mathbf{Q} H_0^{-1} \Omega H_0^{-1} \mathbf{Q}'). \tag{3.6}$$

The last two terms in the definition of δ_i correspond to the correction term for replacing the true conditional density function with its estimator. If an augmented data set is used to estimate this density function and is sufficiently large with respect to the primary data set, then $c_p = 0$ and this correction term in δ_i disappears asymptotically, and so in that case one can treat the density function as known.

3.2. Two-Stage Least Squares Estimation of the Endogenous Truncated Regression Model

Equations (2.30) and (2.35) suggest natural estimators 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 of the underlying conditional density function, using the augmented sample, evaluated at observations in the truncated sample. Define $\mu_0(k) \equiv E[I[0 \le y_i \le k]/f_i]$ and its estimator

$$\hat{\mu}(k) = \frac{1}{n} \sum_{i=1}^{n} \tau_{ni} \frac{I[0 \le y_i \le k]}{\hat{f}_i^*} \tag{3.7}$$

and define our estimator of $\eta(k)$ from equation (2.34) as

$$\hat{\eta}(k) = \hat{\mu}(k)^{-1} \frac{1}{n} \sum_{i=1}^{n} \tau_{ni} \frac{2v_i I[0 \le y_i \le k]}{\hat{f}_i^*}. \tag{3.8}$$

Then our estimator of α_0 is

$$\hat{\alpha} = \frac{k - k^*}{\hat{\eta}(k) - \hat{\eta}(k^*)}. \tag{3.9}$$

The following theorem characterizes the limiting distribution of this estimator. The conditions under which it holds, in addition to its proof, are left to the Appendix.

THEOREM 3.2. *The estimator $\hat{\alpha}$ is root-n consistent and asymptotically normal. Specifically, we have*

$$\sqrt{n}(\hat{\alpha} - \alpha_0) \Rightarrow N(0, \tilde{E}[\psi_{\alpha i}^2]), \tag{3.10}$$

where

$$\psi_{\alpha i} = \frac{\alpha_0}{\eta(k) - \eta(k^*)} \left(\eta(k)\mu_0(k)^{-2}\psi_{\mu i}(k) + \mu_0(k)^{-1}\psi_{\eta i}(k) \right)$$

$$- \frac{\alpha_0}{\eta(k) - \eta(k^*)} \left(\eta(k^*)\mu_0(k^*)^{-2}\psi_{\mu i}(k^*) + \mu_0(k^*)^{-1}\psi_{\eta i}(k^*) \right)$$

$$\tag{3.11}$$

and

$$\psi_{\mu i}(k) = \frac{I[0 \le y_i \le k]}{f_i} - \mu_0 - c_p(\varphi_{\mu b}(v_i^*, z_i^*) - \varphi_{\mu a}(z_i^*)) \tag{3.12}$$

and

$$\psi_{\eta i}(k) = \frac{v_i I[0 \leq y_i \leq k]}{f_i} - \eta(k) - c_p(\varphi_{\eta b}(v_i^*, z_i^*) - \varphi_{\eta a}(z_i^*)), \tag{3.13}$$

where

$$\varphi_{\mu b}(\cdot, \cdot) = E\left[r_{vzi} \frac{I[0 \leq y_i \leq k]}{f_i} \middle| v_i = \cdot, z_i = \cdot \right],$$

$$\varphi_{\mu a}(\cdot) = E\left[r_{zi} \frac{I[0 \leq y_i \leq k]}{f_i} \middle| z_i = \cdot \right],$$

and

$$\varphi_{\eta b}(\cdot, \cdot) = E\left[r_{vzi} \frac{v_i I[0 \leq y_i \leq k]}{f_i} \middle| v_i = \cdot, z_i = \cdot \right],$$

$$\varphi_{\eta a}(\cdot) = E\left[r_{zi} \frac{v_i I[0 \leq y_i \leq k]}{f_i} \middle| z_i = \cdot \right]$$

with, analogous to the notation we adopted before, r_{zi}, r_{vzi} denoting ratios of density functions for truncated and underlying distributions.

To estimate β_0 we define the estimator of Δ by

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

$$\times \left(\frac{1}{n^*} \sum_{i=1}^{n^*} x_i^* z_i^{*\prime} \right) \left(\frac{1}{n^*} \sum_{i=1}^{n^*} z_i^* z_i^{*\prime} \right)^{-1}. \tag{3.14}$$

Note that $\hat{\Delta}$ is, like $\hat{f}^*$, estimated from the augmented data set, i.e., from a sample drawn from the untruncated population of the regressors and instruments. Define the estimator of y_i^* by

$$\hat{y}_i^* = \hat{\mu}^{-1} \frac{(y_i - v_i \hat{\alpha}) I[0 \leq y_i \leq k]}{\hat{f}_i^*}. \tag{3.15}$$

Then our proposed estimator of β_0 is

$$\hat{\beta} = \hat{\Delta} \frac{1}{n} \sum_{i=1}^{n} \tau_{ni} z_i \hat{y}_i^*. \tag{3.16}$$

The following theorem characterizes the limiting distribution of our proposed instrumental variables estimator. The conditions on which it holds, in addition to its proof, are left to the Appendix.

THEOREM 3.3. *Define the following mean zero vectors:*

$$\psi_{\beta_1 i} = -\left(\mu_0^{-2} \frac{k}{\alpha_0} E^*[zx']\beta_0\right) \cdot \psi_{\mu i},$$

$$\psi_{\beta_2 i} = -\left(\frac{1}{2\alpha_0^2}(k^2 E^*[z] - kE^*[zx']\beta_0)\right) \cdot \mu_0^{-1} \cdot \psi_{\alpha i},$$

$$\psi_{\beta_3 i} = \frac{\mu_0^{-1} z_i(y_i - v_i\alpha_0)I[0 \le y_i \le k]}{f_i} - z_i x_i'\beta_0 - c_p(\varphi_{\beta b}(v_i^*, z_i^*) - \varphi_{\beta a}(z_i^*)),$$

where

$$\varphi_{\beta b}(\cdot, \cdot) = E\left[r_{vzi} \frac{\mu_0^{-1} z_i(y_i - v_i\alpha_0)I[0 \le y_i \le k]}{f_i} \,\bigg|\, v_i = \cdot, z_i = \cdot\right],$$

$$\varphi_{\beta a}(\cdot) = E\left[r_{zi} \frac{\mu_0^{-1} z_i(y_i - v_i\alpha_0)I[0 \le y_i \le k]}{f_i} \,\bigg|\, z_i = \cdot\right],$$

and let $\psi_{\beta i} = \psi_{\beta_1 i} + \psi_{\beta_2 i} + \psi_{\beta_3 i}$ *and* $\Omega_\beta = \tilde{E}[\psi_{\beta i}\psi_{\beta i}']$. *Then we have*

$$\sqrt{n}(\hat{\beta} - \beta_0) \Rightarrow N(0, \Delta \cdot \Omega_\beta \cdot \Delta'). \tag{3.17}$$

COROLLARY 3.1. *It immediately follows from the proofs of Theorems 3.2 and 3.3 in the Appendix that the limiting distribution of* $\hat{\theta} = (\hat{\alpha}, \hat{\beta}')'$ *is*

$$\sqrt{n}(\hat{\theta} - \theta_0) \Rightarrow N(0, \tilde{E}[\psi_i \psi_i']), \tag{3.18}$$

where

$$\psi_i = (\psi_{\alpha i}, (\Delta\psi_{\beta i})')'. \tag{3.19}$$

4. 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 performances 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, both with the regressor density known, and with the density estimated from an augmented data set, are reported in Tables 1–3. We simulated data from the following model:

$$\tilde{y}_i = 1 + v_i + 0.5 \cdot x_i + \sigma(x_i)e_i, \tag{4.1}$$

$$y_i = \tilde{y}_i | \tilde{y}_i \ge 0. \tag{4.2}$$

TABLE 1. Homoskedastic truncated normal errors

	Slope				Intercept			
	Mean bias	Median bias	RMSE	MAD	Mean bias	Median bias	RMSE	MAD
200 observations								
WLS	−0.2660	−0.2565	0.4695	0.3740	−0.0055	−0.0033	0.0873	0.0699
AWLS	−0.0720	−0.0576	0.4346	0.3435	−0.0090	−0.0073	0.0998	0.0797
STLS	0.0042	0.0003	0.3431	0.2708	0.0853	0.0980	0.1666	0.1388
PWD	0.0026	0.0132	0.3024	0.2408	—	—	—	—
MLE	−0.1002	−0.0959	0.2475	0.1970	0.3416	0.3405	0.3461	0.3416
400 observations								
WLS	−0.2234	−0.2300	0.3688	0.3026	−0.0018	0.0004	0.0639	0.0507
AWLS	−0.0586	−0.0869	0.3377	0.2731	0.0012	0.0021	0.0690	0.0551
STLS	−0.0405	−0.0399	0.2328	0.1828	0.0917	0.0998	0.1422	0.1199
PWD	−0.0197	−0.0243	0.2142	0.1703	—	—	—	—
MLE	−0.1149	−0.1219	0.2013	0.1643	0.3412	0.3441	0.3437	0.3412
800 observations								
WLS	−0.1681	−0.1782	0.2571	0.2108	0.0013	0.0030	0.0473	0.0380
AWLS	−0.0171	−0.0276	0.2275	0.1778	−0.0013	−0.0022	0.0526	0.0414
STLS	−0.0321	−0.0323	0.1646	0.1294	0.0897	0.1003	0.1181	0.1006
PWD	−0.0138	−0.0219	0.1452	0.1163	—	—	—	—
MLE	−0.1111	−0.1151	0.1580	0.1303	0.3390	0.3394	0.3400	0.3390

327

TABLE 2. Homoskedastic chi-squared errors

	Slope				Intercept			
	Mean bias	Median bias	RMSE	MAD	Mean bias	Median bias	RMSE	MAD
200 observations								
WLS	−0.2424	−0.2234	0.6936	0.5272	−0.1246	−0.1211	0.2941	0.1641
AWLS	−0.0403	−0.1456	0.9074	0.6695	0.0052	−0.0132	0.2305	0.1749
STLS	0.1170	0.1351	0.4528	0.3487	−0.4990	−0.4989	0.5135	0.4990
PWD	0.0358	0.0392	0.6091	0.4690	—	—	—	—
MLE	−0.1414	−0.1391	0.3626	0.2867	0.4969	0.4935	0.5059	0.4969
400 observations								
WLS	−0.2245	−0.2717	0.4694	0.3809	−0.1127	−0.1204	0.1516	0.1297
AWLS	−0.0732	−0.1457	0.6271	0.4622	0.0180	0.0088	0.1531	0.1224
STLS	0.0846	0.0811	0.2954	0.2386	−0.5183	−0.5179	0.5260	0.5183
PWD	−0.0018	−0.0337	0.4710	0.3706	—	—	—	—
MLE	−0.1492	−0.1595	0.2995	0.2453	0.4960	0.4921	0.5005	0.4960
800 observations								
WLS	−0.1761	−0.1950	0.3305	0.2711	−0.0950	−0.0986	0.1182	0.1024
AWLS	−0.0187	−0.0826	0.5945	0.3752	0.0503	0.0425	0.1390	0.1060
STLS	0.1048	0.1041	0.2189	0.1764	−0.5204	−0.5202	0.5238	0.5204
PWD	0.0083	−0.0160	0.2974	0.2403	—	—	—	—
MLE	−0.1442	−0.1524	0.2230	0.1822	0.4962	0.4970	0.4986	0.4962

328

TABLE 3. Heteroskedastic truncated normal errors

	Slope				Intercept			
	Mean bias	Median bias	RMSE	MAD	Mean bias	Median bias	RMSE	MAD
200 observations								
WLS	−0.2775	−0.2860	0.4948	0.4028	−0.0221	−0.0221	0.1008	0.0799
AWLS	−0.0879	−0.1069	0.4762	0.3774	−0.0128	−0.0149	0.1144	0.0903
STLS	0.0567	0.0718	0.3643	0.2894	0.0856	0.0992	0.1699	0.1414
PWD	0.3755	0.3565	0.4989	0.4188	—	—	—	—
MLE	0.1001	0.0970	0.2629	0.2103	0.3379	0.3396	0.3419	0.3379
400 observations								
WLS	−0.2517	−0.2490	0.3999	0.3286	−0.0151	−0.0141	0.0674	0.0538
AWLS	−0.0847	−0.1322	0.3819	0.3100	−0.0053	−0.0049	0.0746	0.0598
STLS	0.0099	0.0327	0.2512	0.2008	0.0947	0.1004	0.1400	0.1187
PWD	0.3544	0.3603	0.4259	0.3703	—	—	—	—
MLE	0.0836	0.0876	0.1908	0.1549	0.3389	0.3412	0.3414	0.3389
800 observations								
WLS	−0.1803	−0.1711	0.2810	0.2287	−0.0094	−0.0089	0.0511	0.0402
AWLS	−0.0180	−0.0460	0.2970	0.2164	−0.0019	0.0022	0.0580	0.0457
STLS	0.0074	0.0101	0.1790	0.1392	0.0895	0.0940	0.1207	0.1013
PWD	0.3581	0.3586	0.3934	0.3607	—	—	—	—
MLE	0.0852	0.0909	0.1491	0.1223	0.3354	0.3365	0.3365	0.3354

329

The random variables x_i, v_i were distributed bivariate normal, with correlation of 0.25 and marginals with mean zero and standard deviation of 0.25. The error term e_i was distributed independently of v_i, x_i, either standard normal truncated at -2 and 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 $\exp(0.5 \cdot x_i)$.

To simulate the model we first generated a censored data set sequentially until the desired sample size n for the truncated data set was achieved. For the proposed estimator with estimated density function (AWLS in the tables), we estimate the density function from exogenous variables in the censored data set using kernel methods with 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, n^* denoting the number of observations in the censored data set. The Silverman (1986) rule of thumb was used to calculate the constant, and an Epanechnikov kernel function was used. We set the right truncation point k to be the sample 75th percentile of y_i.

For comparison, results are also reported for the symmetrically trimmed least squares (STLS) estimator in Powell (1986), the pairwise difference (PWD) estimator in Honoré and Powell (1994) (this estimator can only identify the slope coefficient, and so no results are reported for its intercept term), and the maximum likelihood estimator (MLE) assuming a homoskedastic normal distribution. The PWD and STLS were computed using linear programming and iterative least squares 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 200, 400, and 800 were simulated with 801 replications.

As the results in Tables 1–3 indicate, the proposed WLS estimator performs moderately well at all sample sizes in both the homoskedastic and heteroskedastic designs. The AWLS performs worse for the intercept term, reflecting perhaps the method of bandwidth selection. The MLE performs poorly for all designs as a result of distributional misspecification and/or heteroskedasticity. The STLS poorly estimates the intercept with the chi-squared errors, and PWD performs poorly in the heteroskedastic designs.

Table 4 reports results for the instrumental variables two-stage least squares (2SLS in the table) estimator. Here we simulated data from the following model:

$$\tilde{y}_i = 1 + v_i + 0.5 \cdot x_i + e_i, \tag{4.3}$$

$$y_i = \tilde{y}_i \mid \tilde{y}_i \geq 0. \tag{4.4}$$

To incorporate endogeneity, we simulated a binary variable d_i that took the value 1 with probability $\frac{1}{2}$ and 0 otherwise. When d_i was 1, the error term e_i was equal to x_i, and when d_i was 0, the error term was drawn from the truncated normal distribution mentioned previously. The instrument z_i was indepen-

TABLE 4. Endogenous truncated normal errors

	Slope				Intercept			
	Mean bias	Median bias	RMSE	MAD	Mean bias	Median bias	RMSE	MAD
200 observations								
2SLS	0.1439	0.1605	1.1743	0.9431	−0.0440	−0.0360	0.1157	0.0873
A2SLS	−0.0069	−0.0161	1.0063	0.7719	−0.0048	−0.0055	0.1055	0.0808
STLS	0.3353	0.3413	0.4400	0.3687	0.0302	0.0303	0.0984	0.0795
PWD	0.2841	0.2884	0.3927	0.3254	—	—	—	—
MLE	0.1348	0.1434	0.2558	0.2069	0.2641	0.2633	0.2691	0.2641
400 observations								
2SLS	0.0584	0.0738	0.8483	0.6704	−0.0257	−0.0297	0.0748	0.0604
A2SLS	−0.0111	−0.0368	0.7379	0.5867	−0.0043	−0.0042	0.0697	0.0547
STLS	0.3532	0.3526	0.4023	0.3598	0.0360	0.0387	0.0706	0.0567
PWD	0.2974	0.3077	0.3454	0.3053	—	—	—	—
MLE	0.1466	0.1498	0.2025	0.1687	0.2634	0.2639	0.2658	0.2634
800 observations								
2SLS	−0.0569	−0.0569	0.5444	0.4390	−0.0136	−0.0115	0.0482	0.0385
A2SLS	−0.0366	−0.0508	0.4701	0.3823	−0.0033	−0.0025	0.0448	0.0359
STLS	0.3532	0.3597	0.3793	0.3533	0.0373	0.0358	0.0582	0.0471
PWD	0.3035	0.3107	0.3287	0.3042	—	—	—	—
MLE	0.1516	0.1539	0.1826	0.1586	0.2653	0.2657	0.2667	0.2653

331

dently distributed as uniform between -1 and 1 when d_i was 1 and equal to x_i when d_i was 0.

Here again we first simulated a censored data set for which the density functions could be estimated and then truncated the data. For implementing the proposed two-stage least squares procedure, we used the same kernel function and bandwidth selection procedure as in the heteroskedastic designs. The constants k and k^* needed for this procedure were chosen to be the 25th and 75th percentiles of the dependent variable.

Results using this endogenous model are reported in Table 4 for our two-stage least squares estimator(2SLS) and our estimator using the augmented data set to estimate the density function (A2SLS), and for the STLS, PWD, and MLE. As Table 4 indicates, only the two-stage least squares estimator 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 35% 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.

5. CONCLUSIONS

This paper proposes new estimators for truncated regression models. The estimators are "distribution free" and are robust to general forms of conditional heteroskedasticity, in addition to 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.

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. Magnac and Maurin (2003) compute the bound for a binary choice model under similar identifying assumptions, and Jacho-Chavez (2005) finds the bound for other similar density weighted estimators. Both find that such estimators are generally semiparametrically efficient.

NOTES

1. See, e.g., Hausman and Wise (1977) and Wooldridge (2002) for empirical examples of this type of sampling.

2. Chen et al. (2005) develop general asymptotic theory results for two sample sizes, one corresponding to a primary data set and the other to an auxiliary data set. Their asymptotic theory is

for an estimation procedure based on the method of sieves, which cannot be applied to our procedures, which require estimation of probability density functions.

REFERENCES

Bhattacharya, P.K., H. Chernoff, & S.S. Yang (1983) Nonparametric estimation of the slope of a truncated regression. *Annals of Statistics* 11, 505–514.

Blundell, R. & J.L. Powell (2003) Endogeneity in Semiparametric Binary Response Models. Manuscript, UC Berkeley.

Cavanagh, C. & R.P. Sherman (1998) Rank estimators for monotonic index models. *Journal of Econometrics* 84, 351–381.

Chen, X., H. Hong, & E. Tamer (2005) Nonlinear models with measurement error and auxiliary data. *Review of Economic Studies* 72, 343–366.

Devereux, P.J. & G. Tripathi (2004) Combining data sets to overcome selection caused by censoring and truncation in moment based models. Manuscript, University of Connecticut.

Han, A. (1987) Non parametric analysis of a generalized regression model. *Journal of Econometrics* 35, 303–316.

Hausman, J.A. & D.A. Wise (1977) Social experimentation, truncated distributions, and efficient estimation. *Econometrica* 45, 319–339.

Honoré, B.E. & J.L. Powell (1994) Pairwise difference estimators of censored and truncated regression models. *Journal of Econometrics* 64, 241–278.

Horowitz, J.L. (1992) A smoothed maximum score estimator for the binary response model. *Econometrica* 60, 505–531.

Jacho-Chavez, D.T. (2005) Efficiency Bounds in Semiparametric Models Defined by Moment Restrictions Using an Estimated Conditional Probability Density. Manuscript, London School of Economics.

Lee, M.J. (1989) Mode regression. *Journal of Econometrics* 42, 337–349.

Lee, M.J. (1993) Quadratic mode regression. *Journal of Econometrics* 57, 1–19.

Lewbel, A. (1998) Semiparametric latent variable model estimation with endogenous or mismeasured regressors. *Econometrica* 66, 105–121.

Lewbel, A. (2000) Semiparametric qualitative response model estimation with unknown heteroscedasticity or instrumental variables. *Journal of Econometrics* 97, 145–177.

Magnac, T. & E. Maurin (2003) Identification and Information in Monotone Binary Models. Manuscript, University of Toulouse.

Manski, C.F. (1975) Maximum score estimation of the stochastic utility model of choice. *Journal of Econometrics* 3, 205–228

Manski, C.F. (1985) Semiparametric analysis of discrete response: Asymptotic properties of maximum score estimation. *Journal of Econometrics* 27, 313–334.

Newey, W.K. (1987) Interval Moment Estimation of the Truncated Regression Model. Mimeo. Paper presented at the summer meeting of the Econometric Society, Berkeley.

Newey, W.K. (2001) Conditional moment restrictions in censored and truncated regression models. *Econometric Theory* 17, 863–888.

Newey, W.K. & D. McFadden (1994) Large sample estimation and hypothesis testing. In R.F. Engle & D.L. McFadden (eds.), *Handbook of Econometrics*, vol. 4, pp. 2111–2245. Elsevier.

Powell, J.L. (1986) Symmetrically trimmed least squares estimation of tobit models. *Econometrica* 54, 1435–1460.

Powell, J.L. (1994) Estimation of semiparametric models. In R.F. Engle & D.L. McFadden (eds.), *Handbook of Econometrics*, vol. 4, pp. 2443–2521. Elsevier.

Powell, J.L., J.H. Stock, & T.M. Stoker (1989) Semiparametric estimation of index coefficients. *Econometrica* 57, 1403–1430.

Sherman, R.P. (1994) U-processes in the analysis of a generalized semiparametric regression esti-
    mator. *Econometric Theory* 10, 372–395.
Silverman, B.W. (1986) *Density Estimation for Statistics and Data Analysis*. Chapman and Hall.
Wooldridge, J.M. (2002) *Econometrica Analysis of Cross Section and Panel Data*. MIT Press.

APPENDIX

The Appendix first develops regularity conditions for a general density weighted closed
form estimator and then applies the results to the weighted and two-stage least squares
estimators introduced in this paper. Throughout this section $\|\cdot\|$ will denote the euclid-
ean norm, i.e., for a matrix A with components $\{a_{ij}\}, \|A\| = (\sum_{i,j} a_{ij}^2)^{1/2}$; $\|\cdot\|_\infty$ will de-
note the sup norm over the regressor support.

A.1. Theorem for Density Weighted Estimators. In this section, we establish the
asymptotic properties of a general density weighted estimator. The estimator is defined
as a function of the data, a preliminary root-n consistent estimator of a finite-dimensional
nuisance parameter (denoted by κ_0), and a preliminary estimator of the underlying
conditional density function using kernel estimation. Here, we let $\Xi_0 \in \mathbf{R}^k$ denote the
parameter of interest in the general setting, defined here as

$$\Xi_0 = E\left[\frac{\hbar_i}{f_i^*}\right] \tag{A.1}$$

with $\hbar_i, f_i^*$ denoting $\hbar(y_i, v_i, x_i, z_i, \kappa_0)$ and $f^*(v_i|z_i)$, respectively. For any other possi-
ble value of the nuisance parameter, κ, we will let $\hbar_i(\kappa)$ denote $\hbar(y_i, v_i, x_i, z_i, \kappa)$. We
define the estimator as a sample analogue to the preceding equation:

$$\hat{\Xi} = \frac{1}{n}\sum_{i=1}^{n} \tau_{ni}\frac{\hat{\hbar}_i}{\hat{f}_i^*} \tag{A.2}$$

with $\hat{\hbar}_i, \hat{f}_i^*$ denoting $\hbar_i(\hat{\kappa})$ and $\hat{f}^*(v_i|z_i)$, respectively; τ_{ni} denotes the trimming function
as before, and $\hat{\kappa}$ denotes an estimator of κ_0.

We will assume throughout this section that $\hat{\kappa}$ has an asymptotically linear represen-
tation. Letting the random variables y_i, v_i and the random vectors z_i, x_i be as defined
previously, we express the representation as

$$\hat{\kappa} - \kappa_0 = \frac{1}{n}\sum_{i=1}^{n} \psi_i + o_p(n^{-1/2}), \tag{A.3}$$

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 esti-
mator of the parameter of interest also involves an estimator of the underlying condi-
tional density function $f^*(v_i|z_i)$. We assume that a kernel estimator is used to estimate
this function, and we denote the estimator by $\hat{f}^*(v_i|z_i)$. To define this estimator, we first
assume that the vector z_i 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. As alluded to in the
paper, for identification in the truncated regression model we assume that the popula-
tion density of the regressors (without truncation) is either known or can be estimated

from an alternative or augmented data set. In the latter case, to avoid notational confusion we will distinguish observations from this data set by the superscript $*$ and let n^* denote the number of observations for this data set. Regarding relative sample sizes, we will assume $\lim_{n\to\infty} n/n^* = c_p \in [0,\infty)$. Furthermore, as the two regressor density functions may be different, we will let r_{zi}, r_{vzi} denote $[f(z_i)/f^*(z_i)], [f(v_i,z_i)/f^*(v_i,z_i)]$, respectively.

We define the kernel estimator as

$$
\hat{f}^*(v_i|z_i) = \frac{\dfrac{1}{n^* h_n^{Z_c+1}} \sum_{j=1}^{n^*} I[z_i^{(d)} = z_j^{*(d)}] K_1\left(\dfrac{z_j^{*(c)} - z_i^{(c)}}{h_n}\right) K_2\left(\dfrac{v_j^* - v_i}{h_n}\right)}{\dfrac{1}{n^* h_n^{Z_c}} \sum_{j=1}^{n^*} I[z_i^{(d)} = z_j^{*(d)}] K_1\left(\dfrac{z_j^{*(c)} - z_i^{(c)}}{h_n}\right)}, \tag{A.4}
$$

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.

Our theorem for the asymptotic properties of $\hat{\hat{\Xi}}$ is based on an i.i.d. assumption on the sequences of vectors $(y_i, v_i, x_i', z_i')'$ and $(v_i^*, x_i^{*'}, z_i^{*'})'$ and the following assumptions.

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

C2. $E[\sup_{\kappa \in \mathcal{A}} \|\hbar_i(\kappa)/f_i^*\|] < \infty$

C3. The function $E[\hbar_i(\kappa)/f_i^*]$ is continuous at κ_0.

C4. With ∇_κ denoting the partial derivative with respect to the nuisance parameter, let M_κ denote $E[\nabla_\kappa(\hbar_i(\kappa_0)/f_i^*)]$ and let δ_{ni} denote the vector

$$
M_\kappa \psi_i + \tau_{ni} \frac{\hbar_i}{f_i^*} - E\left[\tau_{ni} \frac{\hbar_i}{f_i^*}\right] - c_p \varphi_{nb}(v_i^*, z_i^*) + c_p \varphi_{na}(z_i^*), \tag{A.5}
$$

where $\varphi_{nb}(\cdot, \cdot) = E[\tau_{ni} r_{vzi}(\hbar_i/f_i^*)|v_i = \cdot, z_i = \cdot]$ and $\varphi_{na}(\cdot) = E[\tau_{ni} r_{zi}(\hbar_i/f_i^*)|z_i = \cdot]$, and let δ_i denote the mean 0 vector

$$
M_\kappa \psi_i + \frac{\hbar_i}{f_i^*} - \Xi_0 - c_p \varphi_b(v_i^*, z_i^*) + c_p \varphi_a(z_i^*), \tag{A.6}
$$

where $\varphi_b(\cdot, \cdot) = E[r_{vzi}(\hbar_i/f_i^*)|v_i = \cdot, z_i = \cdot]$ and $\varphi_a(\cdot) = E[r_{zi}(\hbar_i/f_i^*)|z_i = \cdot]$. Then we make the following assumptions.

C4.1. $\tilde{E}[\|\delta_{ni}\|^2] < \infty$ uniformly in $n \in \mathbf{N}$.

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

C5. We let $\mathcal{Z} = \mathcal{Z}_c \times \mathcal{Z}_d$ denote the support of z_i, which we assume to be the same for the truncated and untruncated populations. We assume that the support set $\mathcal{Z}_c$ is an open, convex subset of $\mathbf{R}^{Z_c}$ and assume that the support of v_i, denoted by $\mathcal{V}$, is an open interval in $\mathbf{R}$. Let $f^*(v, z^{(c)}|z^{(d)})$ denote the (population untruncated) conditional (Lebesgue) density of $v_i, z_i^{(c)}$ given $z_i^{(d)}$ and let $f^*(z^{(d)})$ denote the probability mass function of $z_i^{(d)}$. Furthermore, let $f^*(v, z)$ denote $f^*(v, z^{(c)}|z^{(d)}) \cdot f^*(z^{(d)})$. Then we make the following assumptions.

C5.1. $f^*(v, z^{(c)}|z^{(d)})$, considered as a function of $v, z^{(c)}$, is p times continuously differentiable, with bounded pth derivatives on $\mathcal{V} \times \mathcal{Z}$.

C5.2. There exists a constant $\mu_0 > 0$ such that for all $z_i \in \mathcal{Z}, f^*(z_i^{(d)}) \notin (0, \mu_0)$.

C6. The kernel functions $K_1(\cdot), K_2(\cdot)$ satisfy the following properties.

C6.1. They are each the product of a common univariate function that integrates to 1, has support $[-1,1]$, and is assumed to be p times continuously differentiable.

C6.2. For two vectors of the same dimension, u, l, we let u^l denote the product of each of the components of u raised to the corresponding component of l. Also, for a vector l that has all integer components, we let $s(l)$ denote the sum of its components. The kernel functions are assumed to have the following property:

$$\int K_j(u) u^l \, du = 0 \qquad j = 1, 2 \qquad l \in \mathbf{N}, 1 \leq s(l) < p. \tag{A.7}$$

C7. The functions $\varphi_{na}(z)$ and $\varphi_{nb}(v, z)$ are p times differentiable with bounded pth derivatives, in $z^{(c)}$ and $z^{(c)}, v$, respectively, for all values of $z^{(d)}$ and all $n \in \mathbf{N}$.

C8. The trimming function satisfies the following properties.

C8.1. τ_{ni} is a function of v_i, z_i, and n only, and $0 \leq \tau_{ni} \leq 1$ for all $n \in \mathbf{N}$.

C8.2. For each $v_i, z_i \in \mathcal{V} \times \mathcal{Z}, \tau_{ni} \to 1$ as $n \to \infty$.

C8.3. For all $\delta > 0$, $\sup_{v_i, z_i \in \mathcal{V} \times \mathcal{Z}} \tau_{ni}/f_i^* = o(n^\delta)$, and $\sup_{v_i, z_i \in \mathcal{V} \times \mathcal{Z}} \tau_{ni}/f_{vzi}^* = o(n^\delta)$, where f_{vzi}^* denotes $f^*(v_i, z_i)$.

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

THEOREM A.1. *Suppose Assumptions C1–C8 hold and the bandwidth h_n satisfies $\sqrt{n^*} h_n^p \to 0$ and $n^{*1/2+\delta}/(n^* h_n^{2Z_c}) \to 0$ for some arbitrarily small $\delta > 0$; then*

$$\sqrt{n}(\hat{\Xi} - \Xi_0) \Rightarrow N(0, \Omega), \tag{A.8}$$

where $\Omega = \tilde{E}[\delta_i \delta_i']$.

Remark A.1. Our assumptions regarding the trimming function τ_{ni} have particular features that warrant comment. Assumption C8 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 C8 will be satisfied. Assumption C4 ensures that the bias induced by the trimming function decreases to 0 faster than the parametric rate. For the estimators proposed in this paper, this assumption imposes conditions on the tail behavior of e_i, v_i, z_i and can be satisfied, e.g., if the error term e_i has bounded support and v_i (strictly) contains the support of e_i.

As an alternative to the trimming conditions in C8, 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 situations where the regressors have a bounded, "rectangular" support, as opposed to the support assumptions stated at the beginning of Assumption C5. Here we assume that $z_i^{(c)}, v_i$ have compact support that is independent of $z_i^{(d)}$.

Specifically, we let z_{mx} denote the Z_c-dimensional vector of the maxima in the supports of each of the Z_c components of $z_i^{(c)}$ and z_{mn} denote the vector of minima. Let v_{mx}, v_{mn} denote the maximum and minimum of v_i.

We assume a "rectangular" support of $z_i^{(c)}, v_i$, providing an alternative condition to C5.

C5'. $f^*(v, z^{(c)}|z^{(d)}) > 0 \; \forall z^{(d)} \in \mathcal{Z}_d, \; (z^{(c)}, v) \in [z_{mn}^{[1]}, z_{mx}^{[1]}] \times [z_{mn}^{[2]}, z_{mx}^{[2]}] \times \ldots$ $[z_{mn}^{[Z_c]}, z_{mx}^{[Z_c]}] \times [v_{mn}, v_{mx}]$ where superscripts $[\cdot]$ denote components of a vector. Furthermore, the smoothness condition in C5.1 is satisfied on the interior of the rectangular support of $z_i^{(c)}, v_i$.

Also, before imposing the trimming conditions for these support conditions we slightly modify the smoothness conditions in C7 to account for the rectangular support assumption.

C7'. The functions $\varphi_{na}(z)$ and $\varphi_{nb}(v, z)$ are p times differentiable with bounded pth derivatives on the interior of the (rectangular) support of $z_i^{(c)}, v_i$.

Turning to the trimming procedure, one form of the infeasible trimming function is the product of $Z_c + 1$ indicator functions:

$$\tau_n(v_i, z_i) = I[v_i \in [v_{mn} + h_n, v_{mx} - h_n]] \cdot I[z_i^{[1]} \in [z_{mn}^{[1]} + h_n, z_{mx}^{[1]} - h_n]]$$

$$\cdot I[z_i^{[2]} \in [z_{mn}^{[2]} + h_n, z_{mx}^{[2]} - h_n]] \cdot \ldots I[z_i^{[Z_c]} \in [z_{mn}^{[Z_c]} + h_n, z_{mx}^{[Z_c]} - h_n]].$$

$$(A.9)$$

To define the feasible, data dependent trimming function, let $z_{m\hat{x}}$ denote the Z_c vector obtained by taking the maximum of each of the components of $z_i^{(c)}$ from a sample of n observations. Let $z_{m\hat{n}}$ denote the vector of sample minima and analogously denote sample minima and maxima for v_i as $v_{m\hat{n}}, v_{m\hat{x}}$, respectively. The feasible trimming function is

C8'.

$$\hat{\tau}_n(v_i, z_i) = I[v_i \in [v_{m\hat{n}} + h_n, v_i \in v_{m\hat{x}} - h_n]] \cdot I[z_i^{[1]} \in [z_{m\hat{n}}^{[1]} + h_n, z_{m\hat{x}}^{[1]} - h_n]]$$

$$\cdot I[z_i^{[2]} \in [z_{m\hat{n}}^{[2]} + h_n, z_{m\hat{x}}^{[2]} - h_n]] \cdot \ldots I[z_i^{[Z_c]} \in [z_{m\hat{n}}^{[Z_c]} + h_n, z_{m\hat{x}}^{[Z_c]} - h_n]].$$

$$(A.10)$$

We now show that for our purposes, the feasible data dependent trimming function is asymptotically equivalent to the infeasible trimming function in density estimation and so we can work with the latter in later proofs.

LEMMA A.1.

$$\frac{1}{n} \sum_{i=1}^{n} (\hat{\tau}_n(v_i, z_i) - \tau_n(v_i, z_i)) = o_p(n^{-1/2}).$$

$$(A.11)$$

Proof. Let A_n denote $(1/n)\sum_{i=1}^{n} \tau_n(v_i, z_i) - \hat{\tau}_n(v_i, z_i)$ and for an arbitrarily small $\delta > 0$, let B_n denote the event

$$|v_{\hat{m}x} - v_{mx}| < n^{-(1/2+\delta)}, |v_{\hat{m}n} - v_{mn}| < n^{-(1/2+\delta)},$$

$$|z_{\hat{m}x}^{[j]} - z_{mx}^{[j]}| < n^{-(1/2+\delta)}, |z_{\hat{m}n}^{[j]} - z_{mn}^{[j]}| < n^{-(1/2+\delta)} \qquad j = 1, 2, \dots Z_c. \tag{A.12}$$

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

$$P(n^{1/2}|A_n| > \epsilon) \leq P(n^{1/2}|A_n| > \epsilon, B_n) + P(B_n^c), \tag{A.13}$$

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

$$P(B_n^c) \leq P(|v_{\hat{m}x} - v_{mx}| \geq n^{-(1/2+\delta)}) + P(|v_{\hat{m}n} - v_{mn}| \geq n^{-(1/2+\delta)})$$

$$+ \sum_{j=1}^{Z_c} P(|z_{\hat{m}x}^{[j]} - z_{mx}^{[j]}| \geq n^{-(1/2+\delta)}) + P(|z_{\hat{m}n}^{[j]} - z_{mn}^{[j]}| \geq n^{-(1/2+\delta)}) \tag{A.14}$$

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

$$P(n^{1/2}|A_n| > \epsilon, B_n) \leq P(C_n > \epsilon), \tag{A.15}$$

where

$$C_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Bigg(I[v_i \in [v_{mx} - h_n - n^{-(1/2+\delta)}, v_{mx} - h_n]]$$

$$+ I[v_i \in [v_{mn} + h_n - n^{-(1/2+\delta)}, v_{mn} + h_n]]$$

$$+ \sum_{j=1}^{Z_c} I[z_i^{[j]} \in [z_{mx}^{[j]} - h_n - n^{-(1/2+\delta)}, z_{mx}^{[j]} - h_n]]$$

$$+ I[z_i^{[j]} \in [z_{mn}^{[j]} + h_n - n^{-(1/2+\delta)}, z_{mn}^{[j]} + h_n] > 0] \Bigg). \tag{A.16}$$

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

We now prove the theorem for the density weighted closed form estimator (Theorem A.1). The proof applies to either of the two trimming assumptions and their corresponding support assumptions. For clarity of exposition, we focus on the first set of assumptions and simply note that Assumptions C5′, C7′, and C8′ could be used whenever C5, C7, and C8 are referred to in the proof.

Proof of Theorem A.1. We work with the relationship

$$\hat{\Xi} - \Xi_0 = \frac{1}{n} \sum_{i=1}^{n} \tau_{ni} \left(\frac{\hat{\hbar}_i}{\hat{f}_i^*} - \frac{\hbar_i}{f_i^*} \right) \tag{A.17}$$

$$+ \frac{1}{n} \sum_{i=1}^{n} \left(\tau_{ni} \frac{\hbar_i}{f_i^*} - E\left[\tau_{ni} \frac{\hbar_i}{f_i^*} \right] \right) \tag{A.18}$$

$$+ E\left[\tau_{ni} \frac{\hbar_i}{f_i^*} \right] - \Xi_0. \tag{A.19}$$

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

$$\frac{1}{n} \sum_{i=1}^{n} \tau_{ni} \frac{\hat{\hbar}_i - \hbar_i}{f_i^*} \tag{A.20}$$

$$+ \frac{1}{n} \sum_{i=1}^{n} \tau_{ni} \hat{\hbar}_i \left(\frac{1}{\hat{f}_i^*} - \frac{1}{f_i^*} \right) + R_n. \tag{A.21}$$

The remainder term is of order $(1/n) \sum_{i=1}^{n} \tau_{ni}(\hat{\hbar}_i - \hbar_i)(\hat{f}_i^* - f_i^*)$ and is $o_p(n^{-1/2})$ by Assumptions C1 and C5–C7 and the conditions on h_n (which imply the fourth root consistency with respect to $\|\cdot\|_\infty$ of the kernel density estimator; for the support and trimming conditions C5′ and C8′, see, e.g., Newey and McFadden, 1994, Lem. 8.10; for the conditions C5 and C8, see Sherman, 1994, Cor. 5A). We derive a linear representation for (A.20). A mean value expansion of $\hat{\hbar}_i$ around $\hbar_i$ implies that we can express (A.20) as

$$\left(\frac{1}{n} \sum_{i=1}^{n} \tau_{ni} \frac{\hbar_{\kappa i}^*}{f_i^*} \right) (\hat{\kappa} - \kappa_0), \tag{A.22}$$

where $\hbar_{\kappa i}$ denotes $\nabla_\kappa \hbar(y_i, v_i, x_i, z_i, \kappa^*)$, with κ^* denoting an intermediate value. By Assumptions C1–C3 and the root-n consistency of $\hat{\kappa}$, we can express $((1/n) \sum_{i=1}^{n} \tau_{ni}(\hbar_{\kappa i}^*/f_i^*))$ as $M_\kappa + o_p(1)$. It thus follows by (A.3) that (A.20) has the following linear representation:

$$\frac{1}{n} \sum_{i=1}^{n} M_\kappa \psi_i + o_p(n^{-1/2}). \tag{A.23}$$

Turning attention to (A.21), we again first replace $\hat{\hbar}_i$ with $\hbar_i$. By the root-n consistency of $\hat{\kappa}$ and the uniform consistency of the kernel density estimator (for the bounded support case, see Newey and McFadden, 1994; for the support and trimming conditions C5 and C8, see Sherman, 1994), the resulting remainder term is $o_p(n^{-1/2})$. We next establish a linear representation for (A.21) with $\hbar_i$ replacing $\hat{\hbar}_i$. As the steps are more involved, we state this as a separate lemma.

LEMMA A.2. *Under Assumptions C5–C8, if the bandwidth h_n satisfies $\sqrt{n^*}h_n^p \to 0$ and $n^{*1/2+\delta}/(n^*h_n^{Z_c}) \to 0$ for some arbitrarily small $\delta > 0$, then*

$$\frac{1}{n}\sum_{i=1}^{n} \tau_{ni}\hbar_i\left(\frac{1}{\hat{f}_i^*} - \frac{1}{f_i^*}\right) = \frac{c_p}{n}\sum_{i=1}^{n}\varphi_{na}(z_i^*) - \varphi_{nb}(v_i^*, z_i^*) + o_p(n^{-1/2}).$$

Proof. We again work with the identity

$$\frac{1}{\hat{f}_i^*} - \frac{1}{f_i^*} = \frac{\hat{f}_{zi}^* - f_{zi}^*}{f_{vzi}^*} - \frac{f_{zi}^*(\hat{f}_{vzi}^* - f_{vzi}^*)}{(f_{vzi}^*)^2} \tag{A.24}$$

$$- \frac{(\hat{f}_{zi}^* - f_{zi}^*)(\hat{f}_{vzi}^* - f_{vzi}^*)}{f_{vzi}^*\hat{f}_{vzi}^*} + \frac{f_{zi}^*(\hat{f}_{vzi}^* - f_{vzi}^*)^2}{\hat{f}_{vzi}^*(f_{vzi}^*)^2}, \tag{A.25}$$

where f_{zi}^* denotes the true (population) conditional density function of the continuous components of the instrument vector times the probability function of the discrete components and $\hat{f}_{zi}^*$ denotes estimated values. By Assumptions C5 and C6 and the conditions on the bandwidth, by Lemma 8.10 in Newey and McFadden (1994) for the bounded support case, or Sherman (1994) for the unbounded support case, we have that $\|\hat{f}_{zi}^* - f_{zi}^*\|_\infty$ and $\|\hat{f}_{vzi}^* - f_{vzi}^*\|_\infty$ are both $o_p(n^{*-1/4})$. It will thus suffice to derive representations for

$$\frac{1}{n}\sum_{i=1}^{n}\tau_{ni}\hbar_i\frac{\hat{f}_{zi}^* - f_{zi}^*}{f_{vzi}^*} \tag{A.26}$$

and

$$\frac{1}{n}\sum_{i=1}^{n}\tau_{ni}\hbar_i\frac{f_{zi}^*(\hat{f}_{vzi}^* - f_{vzi}^*)}{(f_{vzi}^*)^2}. \tag{A.27}$$

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

$$\frac{1}{n}\sum_{i=1}^{n}\tau_{ni}\hbar_i\frac{\hat{f}_{zi}^* - \bar{f}_{zi}^*}{f_{vzi}^*} \tag{A.28}$$

$$+ \frac{1}{n}\sum_{i=1}^{n}\tau_{ni}\hbar_i\frac{\bar{f}_{zi}^* - f_{zi}^*}{f_{vzi}^*}. \tag{A.29}$$

We note that by the definition of $\hat{f}_{zi}^*$, (A.29) can be expressed as

$$\frac{1}{n^*\cdot n}\sum_{i=1}^{n}\sum_{j=1}^{n^*}\tau_{ni}\hbar_i\frac{h_n^{-Z_c}K_1\left(\frac{z_j^{*(c)} - z_i^{(c)}}{h_n}\right)I[z_j^{*(d)} = z_i^{(d)}] - \bar{f}_{zi}^*}{f_{vzi}^*}. \tag{A.30}$$

To write this in U-statistic form, we will multiply the terms in the double summation by the sequence of numbers $v_{ni} = I[i \le n]$. Therefore, the preceding equation can be expressed as

$$\frac{1}{n^* \cdot n^*} \sum_{i=1}^{n^*} \sum_{j=1}^{n^*} \tau_{ni} v_{ni} \hbar_i \frac{h_n^{-Z_c} K_1\left(\dfrac{z_j^{*(c)} - z_i^{(c)}}{h_n}\right) I[z_j^{*(d)} = z_i^{(d)}] - \bar{f}_{zi}^*}{f_{vzi}^*}. \tag{A.31}$$

We apply a projection theorem (see, e.g., Powell et al., 1989) to derive a linear representation. Let ζ denote the vector corresponding to observed data, either from the primary or augmented data set, and let $\chi_{n^*}(\zeta_i, \zeta_j^*)$ denote the term inside the double summation. We first note that $\tilde{E}[\|\chi_{n^*}(\zeta_i, \zeta_j^*)\|^2] = O(h_n^{-Z_c})$, which is $o(n^*)$ by the conditions on h_n. We also note that $E^*[\chi_{n^*}(\zeta_i, \zeta_j^*)|\zeta_i] = 0$. It follows by Lemma 3.1 in Powell, Stock, and Stoker (1989) that it will suffice to derive a representation for $E[\chi_{n^*}(\zeta_i, \zeta_j^*)|\zeta_j^*]$. We first show that

$$\frac{1}{n^*} \sum_{j=1}^{n^*} E\left[\tau_{ni} v_{ni} \hbar_i \frac{K_1\left(\dfrac{z_j^{*(c)} - z_i^{(c)}}{h_n}\right) I[z_j^{*(d)} = z_i^{(d)}]}{h_n^{Z_c} f_{vzi}^*} \middle| z_j^*\right]$$

$$= \frac{1}{n^*} \sum_{j=1}^{n^*} E\left[\tau_{ni} v_{ni} r_{zi} \frac{\hbar_i}{f_i^*} \middle| z_j^*\right] + o_p(n^{*-1/2}) \tag{A.32}$$

$$= \frac{n}{n^*} \frac{1}{n} \sum_{j=1}^{n} E\left[\tau_{ni} r_{zi} \frac{\hbar_i}{f_i^*} \middle| z_j^*\right] + o_p(n^{-1/2}). \tag{A.33}$$

To show (A.33), we note that

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

can be written as

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

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

$$\int \varphi_{na}(z_j^{*(c)} - u h_n, z_j^{*(d)}) K_1(u) \, du. \tag{A.34}$$

By Assumptions C5–C7 a pth-order Taylor series expansion of $\varphi_{na}(z_j^{*(c)} - u h_n, z_j^{*(d)})$ around $\varphi_{na}(z_j^*)$ implies that the preceding integral can be expressed as the sum of $\varphi_{na}(z_j^*)$ and a remainder term that is of the form

$$\frac{h_n^p}{p!} \int \sum_{j: s(p_j)=p} \nabla_{p_j} \varphi_{na}(z_j^{*(c)} - u h_n^*, z_j^{*(d)}) u^{p_j} K_1(u) \, du,$$

where here p_j denotes a vector of nonnegative integers, $\nabla_{p_j}\varphi_{na}(\cdot)$ denotes partial derivatives of $\varphi_{na}(\cdot)$ with respect to its components, and the order of each partial corresponds to components of p_j; the vector u raised to the vector of integers p_j denotes the product of the components of u raised to the corresponding component of p_j. Therefore, each term in the summation is a scalar, and we sum over all vectors p_j where the sum of its components, $s(p_j)$, is p. Finally, 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

$$E\left[E\left[\tau_{ni}\hbar_i \frac{K_1\left(\dfrac{z_j^{*(c)}-z_i^{(c)}}{h_n}\right)I[z_j^{*(d)}=z_i^{*(d)}]}{h_n^{Z_c}f_{vzi}^*}\middle| z_j^*\right]-\varphi_{na}(z_j^*)\right]=o_p(n^{-1/2}). \qquad \text{(A.35)}$$

We also note that by the continuity and boundedness of $\varphi_{na}(\cdot)$, an application of the dominated convergence theorem to (A.34) implies that

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

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

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

as $h_n\to 0$. Thus (A.33) follows from Chebyshev's inequality.

To complete the linear representation of $E[\chi_n(\zeta_i,\zeta_j^*)|\zeta_j^*]$ we show that

$$E\left[\tau_{ni}\frac{\hbar_i}{f_{vzi}^*}\bar{f}_{zi}^*\right]=E\left[\frac{\hbar_i}{f_i^*}\right]+o_p(n^{-1/2}). \qquad \text{(A.36)}$$

Note that

$$E\left[\tau_{ni}\frac{\hbar_i}{f_{vzi}^*}f_{zi}^*\right]=E\left[\frac{\hbar_i}{f_i^*}\right]+o(n^{-1/2})$$

by Assumption C4.2. Note also that $\|E[\tau_{ni}(\hbar_i/f_{vzi}^*)(\bar{f}_{zi}^*-f_{zi}^*)]\|$ 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 Sherman (1994) (or, in the bounded support case, by Newey and McFadden, 1994, Lem. 8.9), and $E[\|\tau_{ni}(\hbar_i/f_{vzi}^*)\|]$ is bounded for all n by assumption. The desired result follows by the conditions on the bandwidth. To complete the linear representation in (A.26) we show that

$$\frac{1}{n} \sum_{i=1}^{n} \tau_{ni} \frac{\hbar_i}{f_{vzi}^*} (\bar{f}_{zi}^* - f_{zi}^*) = o_p(n^{-1/2}).$$ (A.37)

First note that $\|E[\tau_{ni}(\hbar_i/f_{vzi}^*)(\bar{f}_{zi}^* - f_{zi}^*)]\|$ is bounded above by

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

$\|\bar{f}_{zi}^* - f_{zi}^*\|_\infty = O(h_n^p)$ (in the bounded support assumption, by Newey and McFadden, 1994, Lem. 8.9; otherwise, Sherman, 1994), and $E[\|\tau_{ni}(\hbar_i/f_{vzi}^*)\|]$ is bounded for all n by assumption. The desired result follows by the conditions on the bandwidth. Therefore, it will suffice to show that

$$\frac{1}{n} \sum_{i=1}^{n} \tau_{ni} \frac{\hbar_i}{f_{vzi}^*} (\bar{f}_{zi}^* - f_{zi}^*) - E\left[\tau_{ni} \frac{\hbar_i}{f_{vzi}^*} (\bar{f}_{zi}^* - f_{zi}^*) \right] = o_p(n^{-1/2}).$$ (A.38)

By Chebyshev's inequality it will suffice to establish the preceding 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.27) has the following linear representation:

$$\frac{1}{n^*} \sum_{i=1}^{n^*} \varphi_{nb}(v_i^*, z_i^*) - E\left[\frac{\hbar_i}{f_i^*} \right] + o_p(n^{-1/2}).$$ (A.39)

This completes the proof of the lemma.   ∎

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

$$\hat{\Xi} - \Xi_0 = \frac{1}{n} \sum_{i=1}^{n} \left(\tau_{ni} \frac{\hbar_i}{f_i^*} - E\left[\tau_{ni} \frac{\hbar_i}{f_i^*} \right] + M_\kappa \psi_i - c_p \varphi_{nb}(v_i^*, z_i^*) + c_p \varphi_{na}(z_i^*) \right) + o_p(n^{-1/2}).$$

(A.40)

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

A.2. Truncated Model Estimators. In this section, we apply the general theorems of the previous sections to derive the limiting distributions of the estimation procedures proposed in the paper. The results are derived under the support, smoothness, and trimming conditions in C5, C7, and C8, but we note that the result also hold under the conditions C5′, C7′, and C8′.

A.2.1. Asymptotics for the Weighted Least Squares Estimator (Proof of Theorem 3.1). Here we derive the limiting distribution of the weighted least squares estimator for the truncated regression model estimator. For notational convenience, here we set $u_i = x_i$. We will derive the limiting distribution for the closed form estimators obtained from (3.2). Specifically, here we will let π_0 denote $(a_0, b_0')'$ with $a_0 = 1/\alpha_0$ and $b_0 = -\beta_0/\alpha_0$. From (3.2) we can define

$$\hat{\pi} = \left(\frac{1}{n} \sum_{i=1}^{n} \tau_{ni} \hat{w}_{ki} \mathbf{x}_i \mathbf{x}_i' \right)^{-1} \frac{1}{n} \sum_{i=1}^{n} \tau_{ni} \hat{w}_{ki} \mathbf{x}_i v_i, \tag{A.41}$$

where here $\mathbf{x}_i = (y_i, x_i')'$ and $\hat{w}_{ki} = (I[0 \le y_i \le k] w(x_i)) / \hat{f}^*(v_i | x_i)$. Note that we have

$$\hat{\pi} - \pi_0 = \left(\frac{1}{n} \sum_{i=1}^{n} \tau_{ni} \hat{w}_{ki} \mathbf{x}_i \mathbf{x}_i' \right)^{-1} \frac{1}{n} \sum_{i=1}^{n} \tau_{ni} \hat{w}_{ki} \mathbf{x}_i (v_i - \mathbf{x}_i' \pi_0) = \hat{H}_n^{-1} \hat{S}_n, \tag{A.42}$$

where $\hat{H}_n = (1/n) \sum_{i=1}^{n} \tau_{ni} \hat{w}_{ki} \mathbf{x}_i \mathbf{x}_i'$ and $\hat{S}_n = (1/n) \sum_{i=1}^{n} \tau_{ni} \hat{w}_{ki} \mathbf{x}_i (v_i - \mathbf{x}_i' \pi_0)$.

The proof of limiting distribution theory is based on the following assumptions.

WLS1. $0 < \alpha_0 < \infty$.

WLS2. The sequences of random vectors (y_i, v_i, x_i) and (v_i^*, x_i^*) are i.i.d.

WLS3. The J-dimensional vector x_i can be partitioned as $(x_i^{(c)}, x_i^{(d)})$, where the J_c-dimensional vector $x_i^{(c)}$ is continuously distributed and the J_d-dimensional vector $x_i^{(d)}$ is discretely distributed. Furthermore, setting $z_i = x_i$, assume either the support condition at the beginning of Assumption C5 or the rectangular support assumption in C5′.

WLS4. The error term e_i satisfies either $E^*[ex] = 0$ or the stronger condition $E^*[e|x] = 0$.

WLS5. Letting $f^*(v, x^{(c)} | x^{(d)})$ denote the (population untruncated) joint density function of $v_i, x_i^{(c)}$ conditional on $x_i^{(d)}$, we assume that this function is p times continuously differentiable in its arguments $v, x^{(c)}$ with bounded derivatives, for all values of $x^{(d)}$. This smoothness assumption holds over the entire support defined at the beginning of Assumption C5 or on the interior of the rectangular support in C5′.

WLS6.1. The trimming function, here a function of v_i, x_i, n, satisfies either C8 (if the support condition at the beginning of C5 is satisfied) or C8′ (if condition C5′ is satisfied).

WLS6.2. For the vector $\hbar_i = I[0 \le y_i \le k] w(x_i) \mathbf{x}_i (v_i - \mathbf{x}_i' \pi_0)$, the trimming function satisfies

$$E \left[\tau_{ni} \frac{\hbar_i}{f_i^*} \right] = o(n^{-1/2}). \tag{A.43}$$

WLS7. Because we have defined $\hbar_i, \tau_{ni}$ for the estimator at hand, condition C7 is satisfied. Alternatively, C7′ holds if the support and trimming conditions in C5′ and C8′ are satisfied.

WLS8. The matrix $H_0 \equiv E[w_{ki} \mathbf{x}_i \mathbf{x}_i']$ is finite and positive definite, where $w_{ki} = (I[0 \le y_i \le k] w(x_i)) / f^*(v_i | x_i)$.

WLS9. With $\hbar_i$ defined in WLS6.2, δ_i is defined as in Assumption C4, with $z_i = x_i$, $z_i^* = x_i^*$. Defining $\Omega = \tilde{E}[\delta_i \delta_i']$, assume $\|\Omega\| < \infty$.

We now state the theorem characterizing the limiting distribution of the weighted least squares estimator $\hat{\pi} \equiv (\hat{a}, \hat{b}')'$ of $\pi_0 \equiv (a_0, b_0')'$.

THEOREM A.2. *Assume that the bandwidth h_n satisfies $\sqrt{n^*} h_n^p \to 0$, and $n^{*1/2+\delta}/(n^* h_n^{Z_c}) \to 0$ for some arbitrarily small $\delta > 0$, the kernel function satisfies Assumption C6, and Assumptions WLS1–WLS9 hold. We then have*

$$\sqrt{n}(\hat{\pi} - \pi_0) \Rightarrow N(0, H_0^{-1} \Omega H_0^{-1}). \tag{A.44}$$

Proof. Our proof strategy will be to show that the probability limit of the term $\hat{H}_n$ is the matrix H_0 and use the results from the previous section to derive a linear representation for $H_0^{-1} \hat{S}_n$. Finally, with the asymptotic distribution theory for $\pi_0 \equiv (a_0, b_0')'$ we can apply the delta method to attain the asymptotic distribution of our estimators of $\theta_0 = (\alpha_0, \beta_0')'$.

As a first step note that under the conditions on the bandwidth sequence and the kernel functions (i.e., Assumption WLS3) we can replace kernel density estimates with true values and the remainder term is uniformly (in the support of x_i, v_i) $o_p(1)$. (For the bounded support and trimming in C5′ and C8′, see, e.g., Newey and McFadden, 1994, Lem. 8.10; for the support and trimming in C5 and C8, see Sherman, 1994, Cor. 5A.) Thus we are left with the term

$$\frac{1}{n} \sum_{i=1}^{n} \tau_{ni} w_{ki} \mathbf{x}_i \mathbf{x}_i'. \tag{A.45}$$

Thus by the law of large numbers and Assumptions WLS8 and WLS9 the preceding term converges to $H_0 \equiv E[w_{ki} \mathbf{x}_i \mathbf{x}_i']$. Note that using (2.10) we can alternatively express H_0 as

$$H_0 = \frac{k}{|\alpha_0|} \begin{bmatrix} E[w(x_i)] \dfrac{k^2}{3} & E[w(x_i) x_i'] \dfrac{k}{2} \\[2ex] E[w(x_i) x_i] \dfrac{k}{2} & E[w(x_i) x_i x_i'] \end{bmatrix}. \tag{A.46}$$

By assumption WLS8 it follows that $\hat{H}_n^{-1} \xrightarrow{P} H_0^{-1}$.

Next we apply results from the previous section to derive a linear representation for

$$H_0^{-1} \frac{1}{n} \sum_{i=1}^{n} \tau_{ni} \hat{w}_{ki} \mathbf{x}_i (v_i - \mathbf{x}_i' \pi_0). \tag{A.47}$$

First we note that by (2.10) and the conditions on the trimming function behavior (specifically WLS6.2), $E[\tau_{ni} w_{ki} \mathbf{x}_i (v_i - \mathbf{x}_i' \pi_0)] = o(n^{-1/2})$. Conditions C1–C8 follow immediately from Assumptions WLS1–WLS9. To state the form of the limiting variance matrix of the estimator of π_0 we note that $\hbar_i / f_i^*$ is mean 0 by (2.10), and so δ_i in WLS9 is mean 0.

Finally, to get the asymptotic variance of the parameters of interest $\theta_0 = (\alpha_0, \beta_0)$ we can simply apply the delta method and pre- and postmultiply the matrix $H_0^{-1} \Omega H_0^{-1}$ by

the Jacobian of transformation. This $(J+1) \times (J+1)$ matrix, referred to here as $\mathbf{Q}$, is of the form

$$\mathbf{Q} = \begin{bmatrix} \alpha_0^2 & \mathbf{0}_{1 \times J} \\ \alpha_0 \beta_0 & \alpha_0 \mathbf{I}_{J \times J} \end{bmatrix}, \tag{A.48}$$

where $\mathbf{0}_{1 \times J}$ refers to a $(1 \times J)$ vector of 0's and $\mathbf{I}_{J \times J}$ refers to a $(J \times J)$ identity matrix. ∎

A.2.2. Asymptotics for the Instrumental Variables Estimator (Proofs of Theorems 3.2 and 3.3). The asymptotic properties of the two-stage least squares estimator are based on the following assumptions in addition to the identification assumptions A.1′–A.5′ in the text.

IV1. The random vectors $(y_i, v_i, x_i', z_i')', (v_i^*, x_i^{*'}, z_i^{*'})'$ are i.i.d.

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. Assume that the support and (population untruncated) density smoothness conditions in Assumption C5 are satisfied or, alternatively, the conditions in C5′.

IV3. $E^*[\|z\|^2] < \infty$, $E^*[\|x\|^2] < \infty$.

IV4. For the vectors $\hbar_{1i} = I[0 < y_i < k]$, $\hbar_{2i} = \mu_0^{-2} 2 y_i I[0 < y_i < k]$, $\hbar_{3i} = z_i(y_i^* - x_i'\beta_0)$, define $\delta_i^j, \delta_{ni}^j, j = 1,2,3$ as in Assumption C4, with $\hbar_i = \hbar_{ji}, j = 1,2,3$, and assume that C4 is satisfied.

IV5. For $j = 1,2,3$, the functions $\tau_{ni}(r_{vzi} \hbar_{ji}/f_i^*)$ satisfy the smoothness of conditional expectation conditions in Assumption C7 or C7′, depending on support conditions.

IV6. The trimming function depends on v_i, z_i, n and satisfies either Assumption C8 or C8′, depending on support conditions.

We now derive the limiting distribution of the two-stage estimator. Our arguments are based on applying Theorem A.1, and so we will be verifying Assumptions C1–C8. We first derive linear representations for $\hat{\mu}$ and $\hat{\alpha}$, assuming that $\alpha_0 > 0$. The following lemma characterizes the limiting distribution of the estimator $\hat{\mu}$.

LEMMA A.3. *If Assumptions A1′–A5′ and IV1–IV6 hold, and the bandwidth sequence satisfies* $\sqrt{n^*} h_n^p \to 0$, $n^{*1/2+\delta}/(n^* h_n^{Z_c}) \to 0$ *for some arbitrarily small* $\delta > 0$, *and the kernel function satisfies Assumption C6, then*

$$\hat{\mu} - \mu_0 = \frac{1}{n} \sum_{i=1}^n \psi_{\mu i} + o_p(n^{-1/2}), \tag{A.49}$$

where $\mu_0 \equiv E[I[0 < y_i < k]/f_i^*]$ *and* $\psi_{\mu i}$ *is defined as in equation (3.12).*

Proof. The result follows directly from Theorem A.1, with $\hbar_i = I[0 < y_i < k]$. ∎

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

THEOREM A.3. *If Assumptions A1′–A5′ and IV1–IV6 hold, and the bandwidth sequence satisfies* $\sqrt{n^*} h_n^p \to 0$, *and* $n^{*1/2+\delta}/(n^* h_n^{Z_c}) \to 0$ *for some arbitrarily small* $\delta > 0$, *and the kernel function satisfies Assumption C6, then*

$$\hat{\alpha} - \alpha_0 = \frac{1}{n} \sum_{i=1}^{n} \psi_{\alpha i} + o_p(n^{-1/2}) \tag{A.50}$$

with $\psi_{\alpha i}$ defined as in equation (3.11).

Proof. We again apply Theorem A.1. In this case $\hbar_i = \mu_0^{-2} 2 y_i I[0 < y_i < k]$, and the plugged-in estimator is $\hat{\mu}$. Note that $E[\nabla_\mu(\hbar_i/f_i^*)] = -2\mu_0^{-1}\alpha_0$. ■

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

THEOREM A.4. *Suppose that Assumptions A1′–A5′ and IV1–IV6 hold and the bandwidth sequence satisfies $\sqrt{n^*}h_n^p \to 0$, and $n^{*1/2+\delta}/(n^*h_n^{Z_c}) \to 0$ for some arbitrarily small $\delta > 0$, and the kernel function satisfies Assumption C6. With Δ defined in equation (2.28), and Ω_β as defined right above equation (3.17),*

$$\sqrt{n}(\hat{\beta} - \beta_0) \Rightarrow N(0, \Delta \cdot \Omega_\beta \cdot \Delta'). \tag{A.51}$$

Proof. Define $\hat{\Delta}$, $\hat{y}_i^*$ as was done in equations (3.14) and (3.15). Note that we can write

$$\hat{\beta} - \beta_0 = \hat{\Delta} \frac{1}{n} \sum_{i=1}^{n} z_i(\hat{y}_i^* - x_i'\beta_0). \tag{A.52}$$

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

$$\hat{\Delta} \xrightarrow{p} \Delta. \tag{A.53}$$

To complete the proof we apply Theorem A.1 to derive a linear representation for

$$\frac{1}{n} \sum_{i=1}^{n} z_i(\hat{y}_i^* - x_i'\beta_0). \tag{A.54}$$

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

$$E\left[\nabla_\mu \frac{\hbar_i}{f_i^*}\right] = -\left(\mu_0^{-2} \frac{k}{\alpha_0} E[z_i x_i']\beta_0\right) \tag{A.55}$$

and

$$E\left[\nabla_\alpha \frac{\hbar_i}{f_i^*}\right] = -\left(\frac{1}{2\alpha_0^2}(k^2 E[z_i] - kE[z_i x_i']\beta_0)\right) \cdot \mu_0^{-1}. \tag{A.56}$$

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