Econometric Theory, **34**, 2018, 543–573.
doi:10.1017/S0266466617000111

NONPARAMETRIC IDENTIFICATION AND ESTIMATION OF TRUNCATED REGRESSION MODELS WITH HETEROSKEDASTICITY

SONGNIAN CHEN AND XUN LU
Hong Kong University of Science and Technology

XIANBO ZHOU
Sun Yat-sen University

YAHONG ZHOU
Shanghai University of Finance and Economics

We consider nonparametric identification and estimation of truncated regression models with unknown conditional heteroskedasticity. The existing methods (e.g., Chen (2010, *Review of Economic Studies* 77, 127–153)) that ignore heteroskedasticity often result in inconsistent estimators of regression functions. In this paper, we show that both the regression and heteroskedasticity functions are identified in a location-scale setting. Based on our constructive identification results, we propose kernel-based estimators of regression and heteroskedasticity functions and show that the estimators are asymptotically normally distributed. Our simulations demonstrate that our new method performs well in finite samples. In particular, we confirm that in the presence of heteroskedasticity, our new estimator of the regression function has a much smaller bias than Chen's (2010, *Review of Economic Studies* 77, 127–153) estimator.

1. INTRODUCTION

Consider the nonparametric location-scale model

$$Y^* = m_0(X) + \sigma_0(X) \cdot \varepsilon, \tag{1.1}$$

where X is an observable d-dimensional vector of regressors, $m_0(\cdot)$ and $\sigma_0(\cdot)$ are unknown regression (location) and positive heteroskedasticity (scale) functions, respectively, and ε is an unobservable error term that is independent of X. This model has been extensively studied in statistics and econometrics (see, e.g., Fan

We gratefully thank the Co-editor Liangjun Su and three anonymous referees for their many constructive comments. We also thank Arthur Lewbel and the participants of the workshop on "Advances in Microeconometrics" (Hong Kong), the 2014 Asian meeting of the Econometric Society (Taipei), and the seminar at University of Hong Kong for their helpful comments. X. Zhou's research was supported by the National Natural Science Foundation of China Grant 71371199. Address correspondence to Songnian Chen, Department of Economics, Hong Kong University of Science and Technology, Hong Kong, China; e-mail: snchen@ust.hk.

© Cambridge University Press 2017

https://doi.org/10.1017/S0266466617000111 Published online by Cambridge University Press

and Gijbels (1996) for kernel estimators and Newey (1997) for series estimators). However, in empirical research, data are often truncated. That is, Y^* is latent, and the observed data (Y, X) are from the conditional distribution of (Y^*, X) given the event $\{Y^* > 0\}$. The goal of this paper is to provide identification and estimation results for the two unknown functions $m_0(\cdot)$ and $\sigma_0(\cdot)$ in the presence of data truncation. The multiplicative heteroskedasticity in equation (1.1) is often used in empirical research. For example, it is commonly assumed in the nonparametric censored models (see, e.g., Fan and Gijbels (1994), Van Keilegom and Akritas (1999), and Chen, Dahl and Khan (2005)).

In two closely related papers, Lewbel and Linton (2002, LL hereafter) and Chen (2010, Chen hereafter) consider nonparametric truncated regression models with homoskedasticity, i.e., $\sigma_0(X)$ is assumed to be a constant. When data are truncated, ignoring the presence of heteroskedasticity is likely to lead to inconsistent estimators of the regression function. This is in sharp contrast with the case without truncation where ignoring heteroskedasticity only results in possible efficiency loss. To the best of our knowledge, none of the existing methods can be applied to truncated regression models with the form of heteroskedasticity in equation (1.1).[1,2] As discussed in Chen, the identification condition in Chen is weaker than that in LL, and the finite sample performance of Chen's method is demonstrated to be better than that of LL's. Hence, we generalize Chen's approach to deal with the case of heteroskedasticity.[3] Although this paper mainly builds on Chen, the identification conditions and asymptotic properties are substantially different from those in Chen. The complication arises from the fact that here we have to identify and estimate two unknown functions jointly.

For identification, we show that for two different points of the regressor, their conditional survival functions of Y, after being appropriately rescaled, should be matched. Specifically, we choose a reference point x_0 and impose the location and scale normalization condition that $m_0(x_0) = 0$ and $\sigma_0(x_0) = 1$. Let x be an evaluation point, and thus the parameter of interest is $(m_0(x), \sigma_0(x))$. Let $S^*(\cdot, x)$ be the conditional survival function of Y^* given $(X = x, Y^* > 0)$:

$$S^*(t, x) \equiv P\left(Y^* > t \mid X = x, Y^* > 0\right) = P(Y > t \mid X = x),$$

which can be estimated from the truncated data. We can show that when $m_0(x) \geq 0$,

$$S^*(t, x_0) = \frac{S^*(\sigma_0(x)t + m_0(x), x)}{S^*(m_0(x), x)} \quad \text{for all } t \geq 0. \tag{1.2}$$

The left-hand side of equation (1.2) is the conditional survival function of Y given the reference point $X = x_0$, whereas the right-hand side is the conditional survival function of Y given $X = x$, after being appropriately rescaled by the true parameters $m_0(x)$ and $\sigma_0(x)$. Equation (1.2) holds because both sides are equal to a rescaled survival function of the error term. Similarly, we can show that when

https://doi.org/10.1017/S0266466617000111 Published online by Cambridge University Press

$m_0(x) < 0$,

$$S^*(t,x) = \frac{S^*\left(\frac{t-m_0(x)}{\sigma_0(x)}, x_0\right)}{S^*\left(\frac{-m_0(x)}{\sigma_0(x)}, x_0\right)} \text{ for all } t \geq 0. \tag{1.3}$$

Thus, the true parameter $(m_0(x), \sigma_0(x))$ must satisfy equations (1.2) and (1.3). Further, we show that $(m_0(x), \sigma_0(x))$ is the unique solution to equations (1.2) and (1.3) if the error term ε does not follow certain special distributions.

Building on equations (1.2) and (1.3), we further propose a kernel-based estimator of $(m_0(x), \sigma_0(x))$ by minimizing a criterion function. Our estimator is consistent and asymptotically normally distributed with a typical nonparametric convergence rate. We conduct Monte Carlo simulations to examine the finite sample performance of our new estimator and find that our estimator substantially outperforms Chen's estimator in most cases, especially when the variation of the heteroskedasticity function is large.

The rest of the paper is organized as follows. In Section 2, we discuss the conditions for the identification of $(m_0(x), \sigma_0(x))$. In Section 3, we propose an estimation procedure based on our identification results and study its asymptotic properties. Section 4 provides simulation evidence of the finite sample performance of our new estimator. Section 5 concludes. All the mathematical proofs are relegated to the Appendix.

2. IDENTIFICATION

Suppose that the observations consist of data $\{Y_i, X_i,\ i = 1, \ldots n\}$ from the conditional distribution of (Y^*, X) given the event $\{Y^* > 0\}$, where the latent dependent variable Y^* is generated as in equation (1.1).[4] The unknown parameters here are $\{m_0(\cdot), \sigma_0(\cdot), F(\cdot)\}$, where $F(\cdot)$ is the distribution of ε. Apparently, $m_0(\cdot)$ and $\sigma_0(\cdot)$ are not identified without suitable normalizations. Equation (1.1) can be written as

$$Y^* = [m_0(X) + c_1 \cdot \sigma_0(X)] + \left[\frac{\sigma_0(X)}{c_2}\right] \cdot \varepsilon^*,$$

where c_1 and c_2 are two arbitrary constants, and $\varepsilon^* \equiv [c_2 \cdot (\varepsilon - c_1)]$. Thus, without any normalizations, $\{m_0(\cdot), \sigma_0(\cdot), F(\cdot)\}$ is observationally equivalent to $\left\{m_0(\cdot) + c_1 \cdot \sigma_0(\cdot), \frac{\sigma_0(\cdot)}{c_2}, F^*(\cdot)\right\}$, where $F^*(\cdot)$ is the distribution of ε^*. Essentially, we need location and scale normalizations. For this, we choose a reference point $x_0 \in \mathbb{X}$ such that $P(Y^* > 0 | X = x_0) > 0$, where $\mathbb{X}$ is the support of X. Without loss of generality, we impose the following normalization conditions:

$$m_0(x_0) = 0 \text{ and } \sigma_0(x_0) = 1. \tag{2.1}$$

(See Remark 2.1 below for the general case in which $m_0(x_0) = c_1$ and $\sigma_0(x_0) = c_2$ for two known constants c_1 and c_2.)

https://doi.org/10.1017/S0266466617000111 Published online by Cambridge University Press

We consider an evaluation point $x \in \mathbb{X}$ such that $P(Y^* > 0|X = x) > 0$.[5] Note that throughout the paper, the identification and estimation results are all pointwise, i.e., for a given evaluation point x. With the normalization above, we now show how to identify $m_0(x)$ and $\sigma_0(x)$. Let $S(\cdot)$ and $\lambda(\cdot)$ be the survival function and hazard function of ε, respectively, i.e.,

$$S(t) \equiv 1 - F(t) \text{ and } \lambda(t) \equiv -\frac{d(\ln S(t))}{dt} = -\frac{1}{S(t)}\frac{d(S(t))}{dt}.$$

Let $S^*(\cdot, x)$ denote the truncated conditional survival function of Y given $X = x$. That is,

$$S^*(t,x) \equiv P\left(Y^* > t | Y^* > 0, X = x\right) = \frac{P(m_0(X) + \sigma_0(X)\cdot\varepsilon > t | X = x)}{P(m_0(X) + \sigma_0(X)\cdot\varepsilon > 0 | X = x)} \text{ for } t \geq 0.$$

Further, it follows that

$$S^*(t,x) = \frac{P\left(\varepsilon > \frac{t - m_0(X)}{\sigma_0(X)} \middle| X = x\right)}{P\left(\varepsilon > \frac{-m_0(X)}{\sigma_0(X)} \middle| X = x\right)} = \frac{S\left(\frac{t-m_0(x)}{\sigma_0(x)}\right)}{S\left(\frac{-m_0(x)}{\sigma_0(x)}\right)} \text{ for all } t \geq 0, \tag{2.2}$$

where the second equality follows from the assumption that ε and X are independent. Equation (2.2) provides the key link between the conditional survival function of Y and the survival function of the error term ε.

Consider two cases for $m_0(x)$: (i) $m_0(x) \geq 0$ and (ii) $m_0(x) < 0$. For case (i), by equation (2.2), we have

$$\frac{S^*(\sigma_0(x)t + m_0(x), x)}{S^*(m_0(x), x)} = \frac{S(t)}{S(0)} \text{ and } S^*(t, x_0) = \frac{S(t)}{S(0)} \text{ for all } t \geq 0.$$

Therefore,

$$S^*(t,x_0) = \frac{S^*(\sigma_0(x)t + m_0(x), x)}{S^*(m_0(x), x)} \text{ for all } t \geq 0. \tag{2.3}$$

This suggests that when $m_0(x) \geq 0$, the true parameter $(m_0(x), \sigma_0(x))$ must satisfy equation (2.3). Note that $S^*(t,x)$ is always identified for $t \geq 0$; thus, equation (2.3) forms the basis for identification of $(m_0(x), \sigma_0(x))$ when $m_0(x) \geq 0$. For case (ii), again by equation (2.2), we have

$$\frac{S^*\left(\frac{t-m_0(x)}{\sigma_0(x)}, x_0\right)}{S^*\left(\frac{-m_0(x)}{\sigma_0(x)}, x_0\right)} = \frac{S\left(\frac{t-m_0(x)}{\sigma_0(x)}\right)}{S\left(\frac{-m_0(x)}{\sigma_0(x)}\right)} \text{ and } S^*(t,x) = \frac{S\left(\frac{t-m_0(x)}{\sigma_0(x)}\right)}{S\left(\frac{-m_0(x)}{\sigma_0(x)}\right)} \text{ for all } t \geq 0.$$

Hence,

$$S^*(t,x) = \frac{S^*\left(\frac{t-m_0(x)}{\sigma_0(x)}, x_0\right)}{S^*\left(\frac{-m_0(x)}{\sigma_0(x)}, x_0\right)} \text{ for all } t \geq 0. \tag{2.4}$$

https://doi.org/10.1017/S0266466617000111 Published online by Cambridge University Press

Equation (2.4) forms the basis for identification $(m_0(x), \sigma_0(x))$ when $m_0(x) < 0$. Combining equations (2.3) and (2.4) suggests that $(m_0(x), \sigma_0(x))$ must satisfy $T(m_0(x), \sigma_0(x)) = 0$, where

$$T(m,\sigma) \equiv T_1(m,\sigma) \cdot \mathbf{1}\{m \geq 0\} + T_2(m,\sigma) \cdot \mathbf{1}\{m < 0\}, \tag{2.5}$$

$$\begin{aligned} T_1(m,\sigma) &\equiv \int_0^\infty \left[S^*(t,x_0) - \frac{S^*(t\sigma + m, x)}{S^*(m,x)} \right]^2 dt, \\ T_2(m,\sigma) &\equiv \int_0^\infty \left[S^*(t,x) - \frac{S^*(\frac{t-m}{\sigma}, x_0)}{S^*(\frac{-m}{\sigma}, x_0)} \right]^2 dt, \end{aligned} \tag{2.6}$$

and $\mathbf{1}\{.\}$ is the usual indicator function. To identify $m_0(\cdot)$ and $\sigma_0(\cdot)$, we actually only need equations (2.3) and (2.4) to hold for a few values of t. Therefore, essentially our model is over-identified. By incorporating all $t \geq 0$ as in equation (2.5), we can improve estimation efficiency.

To formalize the above discussions on identification, we make the following assumptions.

Assumption A1. (i) The data $\{Y, X\}$ are generated based on equation (1.1) conditional on $Y^* > 0$. (ii) ε is continuously distributed with a finite first moment and independent of X. (iii) $m_0(x_0) = 0$ and $\sigma_0(x_0) = 1$, where $x_0 \in \mathbb{X}$ is a reference point with $P(Y^* > 0 | X = x_0) > 0$. (iv) The evaluation point $x \in \mathbb{X}$ satisfies $P(Y^* > 0 | X = x) > 0$ and $\sigma_0(x) > 0$. (v) $\lim_{t\to\infty} t S(t) = 0$. (vi) The conditional survival function $S^*(\cdot, x)$ is continuous for all $x \in \mathbb{X}$.

Define $\delta_x = \max\left\{0, -\frac{m_0(x)}{\sigma_0(x)}\right\}$.

Assumption A2. The survival function of ε, $S(\cdot)$, satisfies the property that there does not exist a pair (α, β) with $\beta > 0$ and $(\alpha, \beta) \neq (0, 1)$ such that

$$\frac{S(t)}{S(\delta_x)} = \frac{S(\beta t + \alpha)}{S(\beta \delta_x + \alpha)} \text{ for all } t \in [\delta_x, \infty). \tag{2.7}$$

LEMMA 2.1 (Identification). *Suppose that A1 and A2 hold. Then* $(m_0(x), \sigma_0(x))$ *is identified as the unique solution to* $T(m, \sigma) = 0$.

In general, identification fails when A2 does not hold. The lack of identification is not particular to any specific method, as shown below.

LEMMA 2.2 (Identification failure). *Suppose that A1 holds and that there exists* $(\alpha, \beta) \in \mathbb{R}^2$ *such that* $\beta > 0$, $(\alpha, \beta) \neq (0, 1)$ *and*

$$\frac{S(t)}{S(\delta_x)} = \frac{S(\beta t + \alpha)}{S(\beta \delta_x + \alpha)} \textit{for all } t \in \mathbb{R}. \tag{2.8}$$

Then, $(m_0(x), \sigma_0(x))$ *and* $(m_0(x) + \alpha\sigma_0(x), \beta\sigma_0(x))$ *are observationally equivalent, thus* $(m_0(x), \sigma_0(x))$ *is not identified.*

https://doi.org/10.1017/S0266466617000111 Published online by Cambridge University Press

Chen's identification result for the homoskedasticity case is a special case of our Lemma 2.1, where he imposes $\sigma_0(x) = 1$ and lets $\beta = 1$. In particular, Chen imposes the condition that there does not exist any α such that

$$\frac{S(t)}{S(\delta_x)} = \frac{S(t+\alpha)}{S(\delta_x+\alpha)} \text{ for all } t \in [\max\{0, -m_0(x)\}, \infty), \tag{2.9}$$

which corresponds to our equation (2.7) with $\beta = 1$. Chen shows that the condition in equation (2.9) can be satisfied under more primitive conditions. Because we have two functions, $m_0(\cdot)$ and $\sigma_0(\cdot)$, to be identified, our condition in equation (2.7) is stronger than Chen's in equation (2.9). Note that when ε has a positive density on the real line, equation (2.7) in A2 is equivalent to

$$\lambda(t) = \beta\lambda(\beta t + \alpha) \text{ for all } t \in [\delta_x, \infty),$$

and equation (2.9) for Chen's homoskedasticity case is equivalent to

$$\lambda(t) = \lambda(t+\alpha) \text{ for all } t \in [\max\{0, -m_0(x)\}, \infty).$$

We now discuss the identification assumptions A1 and A2 in detail. A1(i) describes the data generating process for the truncated regression model with multiplicative heteroskedasticity. A1(ii) is the key independence assumption that is indispensable for identification.[6] A1(iii) and (iv) impose conditions on the reference and evaluation points, respectively. A1(v) is a condition that guarantees that the integral in $T(\cdot,\cdot)$ is finite. A2 is an important assumption that is not innocuous; thus, it deserves a more detailed discussion. In fact, A2 can be violated if the error term ε follows certain special distributions. Below, we provide two concrete examples in which A2 is violated. ε has unbounded and bounded support from the right in Examples 1 and 2 below, respectively.

Example 1
Suppose ε has the support $\left[\underline{K}, \infty\right]$, where $\underline{K}$ is finite or $-\infty$. Its survival function $S(\cdot)$ can be written as

$$S(t) = \frac{c_3}{(t+c_2)^{c_1}} \text{ for all } t \geq \delta_x, \tag{2.10}$$

where c_1, c_2, and c_3 are some constants such that $S(\cdot)$ is an appropriate survival function. Equivalently, its hazard function $\lambda(\cdot)$ can be written as

$$\lambda(t) = \frac{c_1}{t+c_2} \text{ for all } t \geq \delta_x.$$

We can find (α, β) such that $\alpha = c_2(\beta - 1)$ and $\beta > 1$. Suppose that $\delta_x \geq \frac{\alpha}{1-\beta}$. Then

$$\frac{S(\beta t+\alpha)}{S(\beta\delta_x+\alpha)} = \frac{(\beta\delta_x+\alpha+c_2)^{c_1}}{(\beta t+\alpha+c_2)^{c_1}} = \frac{(\beta\delta_x+c_2(\beta-1)+c_2)^{c_1}}{(\beta t+c_2(\beta-1)+c_2)^{c_1}} = \frac{(\delta_x+c_2)^{c_1}}{(t+c_2)^{c_1}} = \frac{S(t)}{S(\delta_x)}, \text{ and}$$

$$\beta\lambda(\beta t+\alpha) = \beta\frac{c_1}{(\beta t+\alpha)+c_2} = \beta\frac{c_1}{(\beta t+c_2(\beta-1))+c_2} = \frac{c_1}{t+c_2} = \lambda(t).$$

https://doi.org/10.1017/S0266466617000111 Published online by Cambridge University Press

Therefore, this example violates A2. In this case, both $(m_0(x), \sigma_0(x))$ and $(m_0(x)+\alpha\sigma_0(x), \beta\sigma_0(x))$ solve the equation $T(m_0(x), \sigma_0(x)) = 0$. Thus, $(m_0(x), \sigma_0(x))$ is not identified using our approach. Note that this example includes a location-scale family of Pareto distributions as a special case.[7]

Example 2

Suppose that the support of ε is $[\underline{K}, \bar{K}]$, where $\underline{K} < \delta_x < \bar{K}$, and $\bar{K}$ is finite. Further, its survival function $S(\cdot)$ can be written as

$$S(t) = \begin{cases} c_2\left(\bar{K}-t\right)^{c_1}, & \text{if } \delta_x \le t < \bar{K}, \\ 0, & \text{if } t \ge \bar{K}, \end{cases} \tag{2.11}$$

where c_1 and c_2 are some constants such that $S(\cdot)$ is an appropriate survival function. Its corresponding hazard function $\lambda(\cdot)$ can be written as

$$\lambda(t) = \frac{c_1}{\bar{K}-t} \text{ if } \delta_x \le t < \bar{K}.$$

Find (α, β) such that $\alpha = (1-\beta)\bar{K}$ and $0 < \beta < 1$. Then,

$$\frac{S(\beta t+\alpha)}{S(\beta\delta_x+\alpha)} = \frac{c_2\left(\bar{K}-\left(\beta t+(1-\beta)\bar{K}\right)\right)^{c_1}}{c_2\left(\bar{K}-\left(\beta\delta_x+(1-\beta)\bar{K}\right)\right)^{c_1}} = \frac{\left(\bar{K}-t\right)^{c_1}}{\left(\bar{K}-\delta_x\right)^{c_1}} = \frac{S(t)}{S(\delta_x)} \text{ if } \delta_x < t < \bar{K}, \text{ and}$$

$$\frac{S(\beta t+\alpha)}{S(\beta\delta_x+\alpha)} = \frac{S\left(\bar{K}+\beta\left(t-\bar{K}\right)\right)}{S(\beta\delta_x+\alpha)} = 0 = \frac{S(t)}{S(\delta_x)} \text{ if } t \ge \bar{K}.$$

Thus, Example 2 violates A2. It can also be seen that

$$\beta\lambda(\beta t+\alpha) = \frac{\beta c_1}{\bar{K}-(\beta t+\alpha)} = \frac{\beta c_1}{\bar{K}-\left(\beta t+(1-\beta)\bar{K}\right)} = \frac{c_1}{\bar{K}-t} = \lambda(t) \text{ if } \delta_x < t < \bar{K},$$

and $\lambda(t)$ is not defined when $t \ge \bar{K}$. The uniform distribution is a special case of this example ($c_1 = 1$).

The next two lemmas essentially show that if we rule out these two kinds of special distributions in Examples 1 and 2, A2 holds.

LEMMA 2.3 (The case of ε having unbounded support from the right).

(*i*) *Suppose that* ε *has the support* $[\underline{K}, \infty]$, *where* $\underline{K}$ *is finite or* $-\infty$ *and* $\lim_{t\to\infty} t^r \cdot S(t)$ *exists (finite or infinite) for all* $r > 0$. *Further, the survival function* $S(\cdot)$ *cannot be written as equation (2.10). Then, equation (2.7) holds only if* $\beta = 1$.

(*ii*) *Suppose that in addition,* $\lambda(\cdot)$ *is not a periodic function such that there does not exist any* α *with* $\lambda(t) = \lambda(t+\alpha)$ *for all* $t \in [\delta_x, \infty)$. *Then, equation (2.7) holds only if* $(\alpha, \beta) = (0, 1)$. *That is, A2 holds.*

Lemma 2.3(*i*) states that if the distribution of the error term does not follow the particular form of equation (2.10), then $\sigma_0(x)$ is identified. Once $\sigma_0(x)$ is identified, the identification of $m_0(x)$ is identical to that in Chen. Lemma 2.3(*ii*) thus

https://doi.org/10.1017/S0266466617000111 Published online by Cambridge University Press

directly follows from Chen. The additional condition in Lemma 2.3(*ii*) can be satisfied by more primitive conditions. For example, if X is continuously distributed, $m_0(\cdot)$ is continuous over the support of X, and $\lambda(\cdot)$ is not a constant on $[\delta_x, \infty)$, then $m_0(x)$ is identified (Chen, Lemma 1(*iii*)).

LEMMA 2.4 (The case of ε having bounded support from the right). *Suppose that ε has bounded support $\left[\underline{K}, \bar{K}\right]$, where $\underline{K} < \delta_x < \bar{K}$, $\underline{K}$ is finite or $-\infty$ and $\bar{K}$ is finite, and $\lim_{t \to +0} \frac{S(\bar{K}-t)}{t^r}$ exists (finite or infinite) for all $r > 0$. Further, $S(\cdot)$ cannot be written as equation (2.11). Then, equation (2.7) holds only if $(\alpha, \beta) = (0, 1)$. That is, A2 holds.*

Lemma 2.4 shows that if the error term has bounded support from the right and its distribution does not satisfy equation (2.11), then both $m_0(x)$ and $\sigma_0(x)$ are identified.

Other than $(m_0(\cdot), \sigma_0(\cdot))$, another object of interest is the distribution of the error term ε, $F(\cdot)$. We now briefly discuss how to identify $F(\cdot)$. Let

$$H(u) \equiv \mathbb{E}\left(\frac{Y^*}{\sigma_0(X)} \middle| \frac{m_0(X)}{\sigma_0(X)} = u, Y^* > 0 \right) = \mathbb{E}\left(\frac{Y}{\sigma_0(X)} \middle| \frac{m_0(X)}{\sigma_0(X)} = u \right),$$

and $H'(u)$ be its derivative. Note that given the identification of $m_0(\cdot)$ and $\sigma_0(\cdot)$, both $H(\cdot)$ and $H'(\cdot)$ are identified. The next lemma shows how to identify $F(\cdot)$. Let $\left[\underline{K}, \bar{K}\right]$ be the support of ε, where $\underline{K}$ can be $-\infty$ and $\bar{K}$ can be ∞. Let $\mathcal{U}$ be the support of $\frac{m_0(X)}{\sigma_0(X)}$.

LEMMA 2.5. *Suppose that A1 and A2 hold. Further suppose that $\left[-\bar{K}, -\underline{K}\right] \subseteq \mathcal{U}$. Then, for any $e \in \left[\underline{K}, \bar{K}\right]$,*

$$F(e) = 1 - \exp\left[\int_{-e}^{-\underline{K}} \frac{H'(u) - 1}{H(u)} du \right]. \tag{2.12}$$

One limitation of the identification result for $F(\cdot)$ above is that the support of $\frac{m_0(X)}{\sigma_0(X)}$ has to be sufficiently large, which may not be satisfied in practice. Also, although after obtaining the estimators of $m_0(\cdot)$ and $\sigma_0(\cdot)$, we can propose a sample analog to estimate $F(\cdot)$, deriving the asymptotic properties is quite involved due to the problem of nonparametrically generated regressors.

Remark 2.1. Consider the general normalization

$$m_0(x_0) = c_1 \text{ and } \sigma_0(x_0) = c_2, \tag{2.13}$$

where c_1 and c_2 are two known constants. Then, equation (1.1) can be written as

$$Y^* = \tilde{m}_0(X) + \tilde{\sigma}_0(X) \cdot \tilde{\varepsilon},$$

where $\tilde{m}_0(x) \equiv \left[m_0(x) - \frac{c_1}{c_2} \cdot \sigma_0(x) \right]$, $\tilde{\sigma}_0(x) \equiv \left[\frac{\sigma_0(x)}{c_2} \right]$, and $\tilde{\varepsilon} \equiv c_2 \varepsilon + c_1$. By construction, $\tilde{m}_0(x_0) = 0$ and $\tilde{\sigma}_0(x_0) = 1$. Thus, $\tilde{m}_0(\cdot)$ and $\tilde{\sigma}_0(\cdot)$ are identified

https://doi.org/10.1017/S0266466617000111 Published online by Cambridge University Press

as discussed above. Therefore, $m_0(\cdot)$ and $\sigma_0(\cdot)$ are identified as

$$\begin{bmatrix} m_0(x) \\ \sigma_0(x) \end{bmatrix} = \begin{bmatrix} 1 & c_1 \\ 0 & c_2 \end{bmatrix} \begin{bmatrix} \tilde{m}_0(x) \\ \tilde{\sigma}_0(x) \end{bmatrix}.$$

Remark 2.2. Chen assumes homoskedasticity and proposes an estimator of $m_0(x)$ based on the following two equations:

$$S^*(t, x_0) = \frac{S^*(t + m_0(x), x)}{S^*(m_0(x), x)} \text{ for all } t \geq 0 \text{ if } m_0(x) \geq 0 \text{ and} \tag{2.14}$$

$$S^*(t, x) = \frac{S^*(t - m_0(x), x_0)}{S^*(-m_0(x), x_0)} \text{ for all } t \geq 0 \text{ if } m_0(x) < 0. \tag{2.15}$$

That is, $\sigma_0(x)$ is forced to be 1 in our equations (2.3) and (2.4). In the presence of heteroskedasticity, equations (2.14) and (2.15) are not necessarily true. Thus, Chen's estimator of $m_0(x)$ is in general inconsistent in the presence of heteroskedasticity.

Remark 2.3. In the location-scale model considered here, basically we assume that the conditional distribution of Y given X depends on X only through $m_0(x)$ and $\sigma_0(x)$. Without normalizations, $m_0(x)$ can represent a broad class of location functions (see, e.g., Akritas and Van Keilegom (2001) and Heuchenne and Van Keilegom (2010)). Here, we have considered the normalization in equation (2.13). In the literature, the following two normalizations are often used.

1. Alternative normalization 1:

   $Y^* = \mu_0(X) + s_0(X) \cdot e$

   with the normalization

   $\mathbb{E}(e) = 0$ and $var(e) = 1$.

   In this case

   $\mu_0(x) = \mathbb{E}\left(Y^* | X = x\right)$ and $s_0(x) = \sqrt{var\left(Y^* | X = x\right)}$.

   That is, μ_0 is the conditional expectation of Y^* given $X = x$ and $s_0^2(x)$ is the conditional variance of Y^* given $X = x$.

2. Alternative normalization 2:

   $Y^* = q_{\tau 0}(X) + s_0(X) \cdot \eta_\tau$

   with the normalization

   $\mathbb{P}(\eta_\tau \leq 0) = \tau$ and $var(\eta_\tau) = 1$,

   for some $\tau \in (0, 1)$. In this case

   $\mathbb{P}\left(Y^* \leq q_{\tau 0}(x) | X = x\right) = \tau$ and $s_0(x) = \sqrt{var\left(Y^* | X = x\right)}$.

   That is, $q_{\tau 0}(x)$ is the τth conditional quantile function of Y^* given $X = x$ and $s_0^2(x)$ is the conditional variance of Y^* given $X = x$.

https://doi.org/10.1017/S0266466617000111 Published online by Cambridge University Press

We can show that the link between our normalization and alternative normalization 1 is that

$$m_0(x) = \mu_0(x) + \left[\frac{c_1 - \mu_0(x_0)}{s_0(x_0)}\right] \cdot s_0(x) \text{ and } \sigma_0(x) = \left[\frac{c_2}{s_0(x_0)}\right] \cdot s_0(x). \quad \textbf{(2.16)}$$

We can also show that the link between our normalization and alternative normalization 2 is that

$$m_0(x) = q_{\tau 0}(x) + \left[\frac{c_1 - q_{\tau 0}(x_0)}{s_0(x_0)}\right] \cdot s_0(x) \text{ and } \sigma_0(x) = \left[\frac{c_2}{s_0(x_0)}\right] \cdot s_0(x). \quad \textbf{(2.17)}$$

Remark 2.4. One potential issue with our approach is that we need to choose x_0 and $m_0(\cdot)$ does not necessarily correspond to the conditional mean or conditional quantile functions as shown in Remark 2.3 above. Nevertheless, our approach is useful in the following aspects.

1. If we know the true value of $\mu_0(x_0)$ at a particular point x_0 and set $c_1 = \mu_0(x_0)$, then based on equation (2.16), $m_0(x) = \mu_0(x)$. Similarly, if we know the true value of $q_{\tau 0}(x_0)$ at a particular point x_0 and set $c_1 = q_{\tau 0}(x_0)$, then based on equation (2.17), $m_0(x) = q_{\tau 0}(x)$. Here x_0 can be thought of as a "representative" point, such as the mean or median of X. We may have knowledge of the population conditional mean or quantile at this point. For example, some census data may report the food expenditure for the consumer with the median income. In the literature, Matzkin (2007, 2012), for example, impose a similar normalization.
2. With our normalization, we can identify and estimate the difference of conditional expectations and conditional quantiles between any two points x_1 and x_2 such that $\sigma_0(x_1) \approx \sigma_0(x_2)$. If $\sigma_0(x_1) \approx \sigma_0(x_2)$, then equations (2.16) and (2.17) imply that

   $$\mu_0(x_1) - \mu_0(x_2) \approx m_0(x_1) - m_0(x_2) \text{ and } q_{\tau 0}(x_1) - q_{\tau 0}(x_2) \approx m_0(x_1) - m_0(x_2).$$

   That is, we can estimate the effects of changing the regressor from x_1 to x_2 on the conditional mean or conditional quantile of Y. Note that for any two given points, x_1 and x_2, we know whether $\sigma_0(x_1) \approx \sigma_0(x_2)$ is true, as $\sigma_0(\cdot)$ can be directly estimated using our method.
3. With our normalization, we can show that

   $$\frac{\sigma_0(x)}{c_2} = \frac{q_{\tau' 0}(x) - q_{\tau 0}(x)}{q_{\tau' 0}(x_0) - q_{\tau 0}(x_0)}.$$

   The left hand side can be estimated directly using our method. The numerator (denominator) of the right hand side is the difference of the τ'th and τth conditional quantiles evaluated at x (x_0). Therefore, $\frac{\sigma_0(x)}{c_2}$ can be interpreted as the ratio of the difference of conditional quantiles at x and x_0.
4. The estimation of $m_0(\cdot)$ and $\sigma_0(\cdot)$ provides the basis for the specification test for heteroskedasticity. We can test heteroskedasticity by testing

   $$\mathbb{E}\left[\left(\sigma^2(X) - c_2\right)^2 \cdot a(X)\right] = 0 \text{ or } \mathbb{E}\left[\left(m_0(X) - m_0^*(X)\right)^2 \cdot a(X)\right] = 0,$$

https://doi.org/10.1017/S0266466617000111 Published online by Cambridge University Press

where $a(\cdot)$ is a positive weight function and $m_0^*(X)$ is the probability limit of Chen's estimator under the assumption of homoskedasticity.

5. In the case where we cannot find an appropriate x_0, the identification and estimation of the conditional mean function $\mu_0(x)$ and the conditional quantile function $q_{\tau 0}(x)$ become much more complicated. In fact, the identification and estimation of $m_0(x)$ and $\sigma_0(x)$ are actually the first step towards identifying and estimating $\mu_0(x)$ and $q_{\tau 0}(x)$ for the model considered in our paper. As we have shown in our Lemma 2.5, after $m_0(x)$ and $\sigma_0(x)$ are identified, the whole distribution of ε, $F(\cdot)$, is identified. Based on the identification of m_0, σ_0 and F, $\mu_0(x)$ and $q_{\tau 0}(x)$ are identified. We are not aware of alternative ways of identifying $\mu_0(x)$ and $q_{\tau 0}(x)$ for nonparametric truncated regressions with heteroskedasticity.

3. ESTIMATION

We consider nonparametric estimation of the location function $m_0(x)$ and the heteroskedasticity scale function $\sigma_0(x)$ in this section. Our estimator of $(m_0(x), \sigma_0(x))$ is based on the identification result that $(m_0(x), \sigma_0(x))$ is the unique minimizer of $T(m, \sigma)$ defined in equation (2.5). The objective function for our estimation is constructed by replacing various elements in the expression of $T(m, \sigma)$ by their consistent estimators. Specifically, our estimator, $(\hat{m}(x), \hat{\sigma}(x))$, is defined as the minimizer of the following function:

$$T_n(m, \sigma) = T_{n1}(m, \sigma) w(m) + T_{n2}(m, \sigma) w^c(m),$$

over (m, σ) on a compact parameter space $M \subset \mathbb{R}^2$, where $w(m)$ is a smoothed version of the indicator function $\mathbf{1}\{m \geq 0\}$, $w^c(m) = 1 - w(m)$, and $T_{n1}(m, \sigma)$ and $T_{n2}(m, \sigma)$ are the sample analogues of $T_1(m, \sigma)$ and $T_2(m, \sigma)$ (equation (2.6)), respectively, with some minor adjustment for technical reasons.[8] Specifically,

$$T_{n1}(m, \sigma) = \int_0^\infty \left[\frac{S_n^*(t, x_0)}{S_n^*(\varepsilon_0, x_0)} - \frac{S_n^*(t\sigma + m, x)}{S_n^*(\varepsilon_0 \sigma + m, x)} \right]^2 w_1(t) dt \tag{3.1}$$

and

$$T_{n2}(m, \sigma) = \int_0^\infty \left[\frac{S_n^*(t, x)}{S_n^*(\varepsilon_0, x)} - \frac{S_n^*(\frac{t-m}{\sigma}, x_0)}{S_n^*(\frac{\varepsilon_0 - m}{\sigma}, x_0)} \right]^2 w_1(t) dt, \tag{3.2}$$

where ε_0 is a small and positive constant,[9] $w_1(\cdot)$ is a non-negative weighting function,

$$S_n^*(t, x) = \sum_{i=1}^n K_y\left(\frac{Y_i - t}{h_y}\right) K\left(\frac{X_i - x}{h}\right) \Big/ \sum_{i=1}^n K\left(\frac{X_i - x}{h}\right), \tag{3.3}$$

which is a smoothed version of the nonparametric estimator of $S^*(t, x)$ based on the random sample $\{Y_i, X_i : i = 1, 2, \ldots, n\}$, $K(\cdot)$ is a d-dimensioned product of

https://doi.org/10.1017/S0266466617000111 Published online by Cambridge University Press

univariate kernels $k(\cdot)$ with h as its bandwidth, and $K_y(t) = \int_{-\infty}^{t} k_y(v)dv$ is a smoothed step function with h_y as the smoothing parameter.

To study the large sample properties of our estimator, we make the following assumptions.

Assumption B1. $\{Y_i, X_i : i = 1, 2, \ldots, n\}$ is a random sample from the distribution of (Y^*, X) conditional on $Y^* > 0$ with $P_0 \equiv P(Y^* > 0) > 0$. The error term ε has positive density on the real line.

Assumption B2. (i) Both x_0 and x are interior points in the support of X. The density function of X, $p(\cdot)$, the regression function, $m_0(\cdot)$, and the positive heteroskedasticity scale function, $\sigma_0(\cdot)$, are continuously differentiable up to order q in some neighborhoods of x_0 and x with both $p(x)$ and $p(x_0)$ being positive, and the cumulative distribution function of ε, $F(\cdot)$, is continuously differentiable up to order $q^* = \max\{q, q_y\}$ for q_y to be specified below, and these derivatives are continuously bounded. (ii) The parameter $(m_0(x), \sigma_0(x))$ is an interior point in the compact parameter space $M \subset \mathbb{R}^2$. (iii) The weight function $w(\cdot)$ is twice continuously differentiable such that $w(m) \geq 0$, $w(m) = 1$ for $m \geq \varepsilon_0/2$, and $w(m) = 0$ for $m \leq -\varepsilon_0/2$. (iv) The weight function $w_1(\cdot)$ is continuous, non-negative, and integrable such that $w_1(t) = 0$ for $t \in [0, \varepsilon_0]$ and $w_1(t) > 0$ for $t > \varepsilon_0$.

Assumption B3. Let $k_y(t) = dK_y(t)/dt$. The kernel function $k(\cdot)$ is continuously differentiable, and $k_y(\cdot)$ is twice continuously differentiable, both with bounded support. Moreover, they are, respectively, q and q_y-order kernel functions, i.e.,

$$\int v^j k(v)dv = \begin{cases} 1, & \text{if } j = 0 \\ 0, & \text{if } 1 \leq |j| < q \\ B_k, & \text{if } |j| = q \end{cases} \quad \text{and} \quad \int u^r k_y(u)du = \begin{cases} 1, & \text{if } r = 0, \\ 0, & \text{if } 1 \leq r < q_y \\ B_{k_y}, & \text{if } r = q_y, \end{cases},$$

where $v = (v_1, \ldots, v_d)'$, $j = (j_1, \ldots, j_d)'$, $|j| = \sum_{i=1}^{d} j_i$, and $v^j = v_1^{j_1} \ldots v_d^{j_d}$.

Assumption B4. The sequence of bandwidths (h, h_y) satisfies $nh^d h_y^{3/2} / \ln n \to \infty$, $h_y^{q_y} = o(h^q)$, $h^{2q}/h_y = o(1)$, and $nh^{d+2q} = O(1)$ as $n \to \infty$.

B1–B4 are standard in the nonparametric estimation literature and almost identical to those in Chen. B1 assumes that the cross-sectional data constitute a random sample. B2 mainly imposes the smoothness and boundedness conditions. B3 and B4 specify conditions on the kernel functions and the rates of the bandwidths, respectively. For example, when $d = 1$ and the second-order kernels ($q = q_y = 2$) are used, the bandwidths can be chosen as $h \propto n^{-1/5}$ and $h_y \propto n^{-1/3}$.

THEOREM 3.1. *Under Assumptions A1–A2 and B1–B4, $\hat{m}(x)$ and $\hat{\sigma}(x)$ are consistent estimators of $m_0(x)$ and $\sigma_0(x)$, respectively, and asymptotically*

https://doi.org/10.1017/S0266466617000111 Published online by Cambridge University Press

normally distributed:

$$\sqrt{nh^d}\left(\begin{pmatrix}\hat{m}(x)-m_0(x)\\ \hat{\sigma}(x)-\sigma_0(x)\end{pmatrix}-h^q b(x)\right)\xrightarrow{d} N(0,\Gamma(x)),$$

with

$$\Gamma(x)=V(x)^{-1}(\Gamma_1(x)+\Gamma_2(x))V(x)^{-1}\int k^2(v)dv,$$

where $b(x)$, $V(x)$, $\Gamma_1(x)$, *and* $\Gamma_2(x)$ *are defined in the proof.*

For the purpose of conducting statistical inference, we can make the bias terms negligible asymptotically by choosing a bandwidth h with a sub-optimal rate such that $n^{\frac{1}{2}}h^{\frac{d}{2}+q}=o(1)$. For the asymptotic variance terms, in principle, we can estimate them by replacing the various elements with their sample analogues, although it would be tedious. Alternatively, we can consider subsampling as in Politis, Romano, and Wolf (1999, PRW).[10] Specifically, consider estimating the variance of $\hat{m}(x)$. Let the bandwidth $h=h_n=n^{-\alpha}$ and $b=b_n$ be a sequence of positive integers such that $b\to\infty$ and $b/n\to 0$ as $n\to\infty$. Let B be a large integer. Define $\tau_n=\sqrt{nh_n^d}=n^{\frac{1-\alpha d}{2}}$ and $\tau_r=r^{\frac{1-\alpha d}{2}}$, where $r=bn/(n-b)$. The subsampling procedure is as follows:

1. Randomly draw B subsamples $\left\{\left(X_i^{*(k)},Y_i^{*(k)}\right),i=1,\ldots,b\right\}_{k=1}^{B}$ of size b from the original sample $\{(X_i,Y_i)\}_{i=1}^{n}$.
2. For $k=1,\ldots,B$, compute $\hat{m}(x)$ using the subsample $\left\{\left(X_i^{*(k)},Y_i^{*(k)}\right)\right\}_{i=1}^{b}$ and denote this as $\hat{m}^{*(k)}(x)$.
3. Estimate the variance of $\hat{m}(x)$ as

$$\frac{\tau_r^2}{\tau_n^2}\left[\frac{1}{B}\sum_{k=1}^{B}\left[\hat{m}^{*(k)}(x)\right]^2-\left[\frac{1}{B}\sum_{k=1}^{B}\left[\hat{m}^{*(k)}(x)\right]\right]^2\right].$$

Given the general theory of subsampling (see, e.g., PRW), the validity of the above procedure can be readily established.

Remark 3.1. With the general normalization $m_0(x_0)=c_1$ and $\sigma_0(x_0)=c_2$, $(m_0(x),\sigma_0(x))'$ can be estimated by

$$\begin{bmatrix}\bar{m}(x)\\ \bar{\sigma}(x)\end{bmatrix}=\mathbb{C}\begin{bmatrix}\hat{m}(x)\\ \hat{\sigma}(x)\end{bmatrix},\ \text{where}\ \mathbb{C}=\begin{bmatrix}1 & c_1\\ 0 & c_2\end{bmatrix}.$$

It is easy to show that

$$\sqrt{nh^d}\left(\begin{pmatrix}\bar{m}(x)-m_0(x)\\ \bar{\sigma}(x)-\sigma_0(x)\end{pmatrix}-\mathbb{C}h^q b(x)\right)\xrightarrow{d} N(0,\mathbb{C}\Gamma(x)\mathbb{C}').$$

https://doi.org/10.1017/S0266466617000111 Published online by Cambridge University Press

Remark 3.2. Here, we consider the case in which the regressor X is continuous. Discrete regressors can be easily accommodated. Specifically, suppose that $X = (X_c, X_d)$, where X_c is continuous and X_d is discrete. To estimate $S^*(t,x) = S^*(t,(x_c,x_d))$, we can simply stratify the sample by each distinct discrete outcome. That is, it can be estimated using the data with $X_d = x_d$. More sophisticated methods, such as smoothing across the discrete outcomes, can also be applied (see, e.g., Li and Racine (2003)).

Remark 3.3. We propose our estimator based on our constructive identification results. We do not consider the efficiency issue. For example, different choices of the weight functions $w_1(\cdot)$ used in equations (3.1) and (3.2) are likely to lead to different efficiencies of the estimators. Also as pointed out by a referee, given that the model is identified, alternative estimators, such as nonparametric maximum likelihood estimators (MLEs), may be more efficient. We leave this important question for future research.

Remark 3.4. Let $\hat{m}(x_1)$ and $\hat{m}(x_2)$ be the estimators of $m_0(x_1)$ and $m_0(x_2)$ at two different points, x_1 and x_2, respectively. Unlike in the typical nonparametric regression case, $\hat{m}_0(x_1)$ and $\hat{m}_0(x_2)$ are not asymptotically independent due to the fact that the same reference point x_0 is used. Specifically, as shown in the proof of Theorem 3.1, the influence function (equation (A.9)) of $\sqrt{nh^d}\left(\hat{m}(x) - m_0(x)\right)$ contains the term $K\left(\frac{X_i - x_0}{h}\right)$. Therefore, the limiting covariance between $\sqrt{nh^d}\left(\hat{m}(x_1) - m_0(x_1)\right)$ and $\sqrt{nh^d}\left(\hat{m}(x_2) - m_0(x_2)\right)$ is not $o(1)$ due to the presence of the common term $K\left(\frac{X_i - x_0}{h}\right)$ in both expressions.

4. MONTE CARLO SIMULATIONS

In this section, we examine the finite sample performance of our estimator.

4.1. Data Generating Processes and Implementation

We consider four data generating processes (DGPs). In all four DGPs, $Y^* = m_0(X) + \sigma_0(X) \cdot \varepsilon$, where $m_0(x) = 0.8 + x$. The heteroskedasticity functions are chosen as follows:

DGP 1: $\sigma_0(x) = 0.6(1 + x^2)$,
DGP 2: $\sigma_0(x) = 0.6\exp(0.4x^2)$,
DGP 3: $\sigma_0(x) = 0.6/(0.1 + x^2)$,
DGP 4: $\sigma_0(x) = 0.6/(0.1 + |x|)$.

In all four DGPs, X is uniformly distributed on $[-1, 2]$, and ε is independent of X and follows the standard normal distribution.[11] Figures 1(a)–(d) show the heteroskedasticity functions for the four DGPs, respectively.

In all the simulations, we choose the reference point $x_0 = 0$. We impose the normalization $m_0(x_0) = 0.8$ for all four DGPs, and $\sigma_0(x_0) = 0.6$ for DGPs 1 and 2 and $\sigma_0(x_0) = 6$ for DGPs 3 and 4. See Remarks 2.1 and 3.1 for how to handle

https://doi.org/10.1017/S0266466617000111 Published online by Cambridge University Press

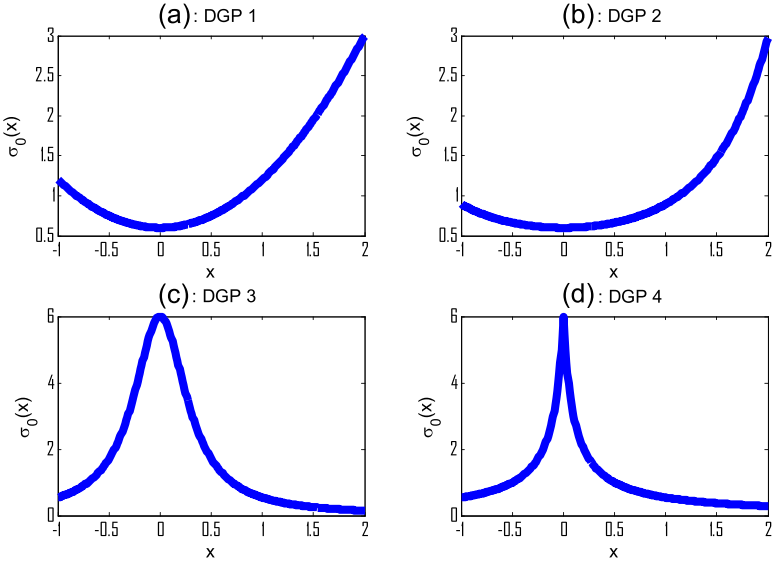


FIGURE 1. Heteroskedasticity functions.

the general normalization. We choose seven evaluation points (x) : -0.9, -0.6, -0.3, 0.3, 0.6, 1.3, and 1.9 in the support of X, $[-1, 2]$. Following Chen, we use the nonsmoothed estimator of the conditional survival function, thus we avoid choosing k_y and h_y. The weight function is chosen as $w_1(t) = \mathbf{1}\{0 \leq t < y_{0.975}\}$, where $y_{0.975}$ is the 0.975th sample quantile of $\{Y_i\}$. We let $w(m) = \mathbf{1}\{m \geq 0\}$ and $\varepsilon_0 = 0$. Different choices of ε_0 do not produce noticeable changes in our simulations. We adopt the standard normal density function as the kernel function, $k(\cdot)$. To choose the bandwidth h, we propose two practical ways. The first way is a rule of thumb. Specifically, we set $h = c \cdot s_x \cdot n^{-1/5}$, where s_x is the sample standard deviation of $\{X_i\}$ and $c = 1.06$ according to Silverman's (1986) rule of thumb. To examine the sensitivity to bandwidth h, different values of the constant c, 0.8 and 1.2, are also tried. The second way is based on the idea of cross-validation (CV). We may consider the leave-one-out CV for the estimator of the function $m_0(\cdot)$, i.e., we minimize

$$\frac{1}{n}\sum_{i=1}^{n}\left[Y_i - \hat{m}_{-i}(X_i)\right]^2 \cdot w_2(X_i),$$

where $\hat{m}_{-i}(X_i)$ is the estimator of $m_0(X_i)$ with the ith observation being deleted and $w_2(\cdot)$ is a weight function.[12] However, there are two issues with minimizing this criterion function. First, as we point out above, $m_0(\cdot)$ is not necessarily the conditional expectation function, therefore it may not be reasonable to minimize the distance between Y_i and the estimator of $m_0(X_i)$. Second, this minimization procedure can be computationally intensive, as the estimator of $m_0(\cdot)$

https://doi.org/10.1017/S0266466617000111 Published online by Cambridge University Press

itself involves a minimization problem. Considering these two issues, we propose to cross-validate the conditional survival function $S^*(\cdot,\cdot)$, which is the key element in the construction of our estimators of $m_0(\cdot)$ and $\sigma_0(\cdot)$. Specifically, let $S^*_{n,-i}(t, X_i)$ be the leave-one-out estimator of $S^*(t, X_i)$ with the ith observation being deleted. The CV bandwidth is constructed as

$$h_{cv} = \arg\min \int \left(\frac{1}{n} \sum_{i=1}^{n} \left[\mathbf{1}\{Y_i > t\} - S^*_{n,-i}(t, X_i) \right]^2 \cdot w_2(X_i) \right) \cdot w_1(t) dt,$$

where $w_2(\cdot)$ is a weight function. In our simulations, we simply let $w_2(\cdot) = 1$. In practice, CV might be preferred, as it is more adaptive to the underlying DGPs.

For each DGP, we report the bias, standard deviation (SD), and root mean square error (RMSE) for the estimators of $m_0(x)$ and $\sigma_0(x)$. For $m_0(x)$, we compare our estimator with Chen's estimator that ignores heteroskedasticity. We consider the sample size $n = 400$.[13] The number of replications is 500.

4.2. Simulation Results

The simulation results for DGP 1 are reported in Table 1. As shown in Figure 1(a), the range of the heteroskedasticity function $\sigma_0(\cdot)$ is $[0.6, 3]$. In general, our estimator performs well and substantially outperforms Chen's estimator. In terms of RMSE, our estimator has a much smaller value than Chen's for most of the evaluation points, except those close to 0. As expected, Chen's estimator has a much larger bias than ours in the presence of heteroskedasticity. For example, for $c = 1.06$, at the evaluation point $x = 1.3$, the bias of Chen's estimator (1.741) is about eight times as large as that of ours (-0.211). When x is close to 0, say in the area of $[-0.3, 0.3]$, the bias of Chen's estimator is relatively small, which is not surprising. Essentially, both our estimator of $(m_0(x), \sigma_0(x))$ and Chen's estimator of $m_0(x)$ rely only on the information in two local areas: the area around the evaluation point x and that around the reference point x_0. Intuitively, Chen's estimator, which ignores heteroskedasticity, performs well if the following three conditions are met: (i) $\sigma_0(\cdot)$ is flat around the evaluation point x, (ii) $\sigma_0(\cdot)$ is flat around the reference point x_0, and (iii) $\sigma_0(x)$ is close to $\sigma_0(x_0)$. It turns out that when $x \in [-0.3, 0.3]$ and $x_0 = 0$, these three conditions are approximately true. As shown in Figure 1(a), the heteroskedasticity function is almost flat in the area $x \in [-0.3, 0.3]$, with a range of $[0.6, 0.654]$. Hence, ignoring the heteroskedasticity in this area does not result in a large bias.

Table 2 presents the results for DGP 2. The range of the heteroskedasticity in this case is $[0.6, 2.97]$, which is similar to that in DGP 1. When $x \in [-0.6, 0.6]$, the range of $\sigma_0(x)$, [0.6, 0.693], is quite flat. Again, the performance of our estimator is much better than that of Chen's in the region of large heteroskedasticity.

Table 3 reports the simulation results for DGP 3. As shown in Figure 1(c), the range of the heteroskedasticity function, $[0.15, 6]$, is quite large. In this DGP, our estimator is much better than Chen's for all evaluation points. For example,

https://doi.org/10.1017/S0266466617000111 Published online by Cambridge University Press

TABLE 1. DGP 1, $n = 400$

		Our estimator				Our estimator			Chen's estimator		
x	$\sigma_0(x)$	Bias	SD	RMSE	$m_0(x)$	Bias	SD	RMSE	Bias	SD	RMSE
Rule of thumb with $c = 0.80$											
−0.9	1.086	−0.239	0.277	0.365	−0.1	0.226	0.592	0.633	0.675	0.290	0.734
−0.6	0.816	−0.071	0.227	0.238	0.2	0.031	0.498	0.499	0.287	0.159	0.328
−0.3	0.654	−0.024	0.182	0.184	0.5	0.007	0.380	0.380	0.073	0.107	0.129
0.3	0.654	0.037	0.148	0.152	1.1	−0.053	0.255	0.260	0.031	0.102	0.107
0.6	0.816	0.032	0.219	0.221	1.4	−0.079	0.338	0.347	0.216	0.243	0.325
1.3	1.614	−0.002	0.387	0.387	2.1	−0.158	0.638	0.657	1.711	0.449	1.769
1.9	2.766	−0.466	0.578	0.743	2.7	−0.019	0.882	0.881	1.299	0.463	1.379
Rule of thumb with $c = 1.06$											
−0.9	1.086	−0.274	0.257	0.376	−0.1	0.162	0.591	0.612	0.601	0.251	0.651
−0.6	0.816	−0.099	0.215	0.236	0.2	0.041	0.500	0.501	0.270	0.137	0.303
−0.3	0.654	−0.047	0.136	0.143	0.5	0.042	0.300	0.302	0.077	0.085	0.115
0.3	0.654	0.059	0.133	0.145	1.1	−0.105	0.248	0.269	0.019	0.091	0.093
0.6	0.816	0.052	0.197	0.203	1.4	−0.131	0.357	0.380	0.238	0.304	0.386
1.3	1.614	−0.036	0.338	0.339	2.1	−0.211	0.579	0.615	1.741	0.454	1.799
1.9	2.766	−0.590	0.519	0.786	2.7	−0.138	0.808	0.819	1.339	0.421	1.403
Rule of thumb with $c = 1.20$											
−0.9	1.086	−0.325	0.242	0.405	−0.1	0.182	0.597	0.623	0.547	0.196	0.581
−0.6	0.816	−0.149	0.194	0.244	0.2	0.100	0.472	0.482	0.250	0.130	0.282
−0.3	0.654	−0.070	0.129	0.147	0.5	0.075	0.314	0.322	0.075	0.084	0.113
0.3	0.654	0.051	0.117	0.127	1.1	−0.099	0.257	0.275	0.012	0.085	0.085
0.6	0.816	0.046	0.184	0.189	1.4	−0.149	0.327	0.359	0.227	0.290	0.368
1.3	1.614	−0.034	0.322	0.323	2.1	−0.280	0.605	0.666	1.787	0.445	1.842
1.9	2.766	−0.608	0.495	0.784	2.7	−0.326	0.837	0.897	1.348	0.412	1.409
CV (cross-validation)											
−0.9	1.086	−0.254	0.271	0.371	−0.1	0.178	0.621	0.645	0.634	0.265	0.687
−0.6	0.816	−0.089	0.229	0.246	0.2	0.043	0.514	0.516	0.279	0.160	0.322
−0.3	0.654	−0.031	0.172	0.174	0.5	0.015	0.370	0.370	0.073	0.104	0.127
0.3	0.654	0.052	0.148	0.157	1.1	−0.092	0.273	0.289	0.017	0.104	0.105
0.6	0.816	0.049	0.214	0.220	1.4	−0.117	0.360	0.378	0.219	0.247	0.330
1.3	1.614	−0.024	0.370	0.371	2.1	−0.210	0.632	0.665	1.681	0.482	1.748
1.9	2.766	−0.524	0.547	0.757	2.7	−0.074	0.897	0.899	1.330	0.419	1.394

for $c = 1.06$, at $x = -0.9$, the bias, SD, and RMSE of Chen's estimator are about 911, 2.8, and 4.4 times as large as those of ours, respectively. In contrast to the two previous DGPs, Chen's estimator performs much worse than ours even in the area where the heteroskedasticity function is relatively flat, say around $x = 1.9$. As explained above, Chen's estimator works well when the three conditions are met. But in this case, conditions (ii) and (iii) are clearly not satisfied. In the neighborhood of the evaluation point $x_0 = 0$, the heteroskedasticity function

https://doi.org/10.1017/S0266466617000111 Published online by Cambridge University Press

TABLE 2. DGP 2, $n = 400$

		Our estimator				Our estimator			Chen's estimator		
x	$\sigma_0(x)$	Bias	SD	RMSE	$m_0(x)$	Bias	SD	RMSE	Bias	SD	RMSE
Rule of thumb with $c = 0.80$											
−0.9	0.830	−0.207	0.397	0.447	−0.1	0.144	0.525	0.544	0.368	0.233	0.435
−0.6	0.693	−0.054	0.357	0.361	0.2	0.006	0.442	0.441	0.122	0.177	0.215
−0.3	0.622	0.011	0.318	0.317	0.5	−0.049	0.383	0.386	0.015	0.133	0.134
0.3	0.622	0.037	0.259	0.262	1.1	−0.018	0.258	0.258	0.019	0.109	0.110
0.6	0.693	0.012	0.296	0.296	1.4	−0.006	0.283	0.283	0.069	0.154	0.168
1.3	1.180	0.055	0.482	0.485	2.1	−0.092	0.413	0.423	1.099	0.524	1.217
1.9	2.543	−0.670	0.833	1.068	2.7	−0.146	0.712	0.726	1.324	0.398	1.383
Rule of thumb with $c = 1.06$											
−0.9	0.830	−0.144	0.218	0.261	−0.1	0.173	0.496	0.525	0.355	0.200	0.408
−0.6	0.693	−0.033	0.205	0.207	0.2	0.010	0.464	0.464	0.135	0.158	0.208
−0.3	0.622	−0.005	0.168	0.168	0.5	−0.023	0.382	0.382	0.038	0.092	0.099
0.3	0.622	0.032	0.127	0.131	1.1	−0.040	0.242	0.245	0.002	0.083	0.083
0.6	0.693	0.023	0.152	0.154	1.4	−0.025	0.252	0.253	0.067	0.144	0.159
1.3	1.180	0.056	0.247	0.253	2.1	−0.126	0.353	0.375	1.245	0.577	1.372
1.9	2.543	−0.570	0.466	0.736	2.7	−0.195	0.649	0.677	1.324	0.393	1.381
Rule of thumb with $c = 1.20$											
−0.9	0.830	−0.188	0.201	0.275	−0.1	0.216	0.489	0.534	0.332	0.184	0.379
−0.6	0.693	−0.052	0.195	0.202	0.2	0.023	0.459	0.459	0.128	0.158	0.203
−0.3	0.622	−0.019	0.152	0.153	0.5	0.000	0.361	0.361	0.039	0.085	0.093
0.3	0.622	0.032	0.102	0.107	1.1	−0.038	0.196	0.199	−0.002	0.068	0.068
0.6	0.693	0.026	0.143	0.145	1.4	−0.045	0.234	0.238	0.066	0.135	0.150
1.3	1.180	0.027	0.228	0.229	2.1	−0.164	0.358	0.394	1.237	0.600	1.375
1.9	2.543	−0.654	0.407	0.770	2.7	−0.315	0.598	0.676	1.286	0.419	1.353
CV (cross-validation)											
−0.9	0.830	−0.139	0.234	0.271	−0.1	0.172	0.513	0.541	0.351	0.221	0.415
−0.6	0.693	−0.032	0.213	0.215	0.2	0.017	0.454	0.454	0.133	0.165	0.211
−0.3	0.622	−0.006	0.172	0.172	0.5	−0.018	0.361	0.361	0.027	0.115	0.118
0.3	0.622	0.024	0.151	0.153	1.1	−0.023	0.261	0.262	0.005	0.096	0.096
0.6	0.693	0.020	0.178	0.179	1.4	−0.024	0.275	0.276	0.067	0.148	0.162
1.3	1.180	0.039	0.271	0.274	2.1	−0.104	0.396	0.409	1.166	0.563	1.295
1.9	2.543	−0.453	0.502	0.676	2.7	−0.175	0.702	0.723	1.291	0.437	1.363

varies substantially. In addition, $\sigma_0(x)$ is quite different from $\sigma_0(x_0)$ when x is close to 1.9. This suggests that when the heteroskedasticity function varies much globally over the support of X, the performance of Chen's estimator is in general poor, even in the region where the heteroskedasticity function is flat locally.

Table 4 reports the simulation results for DGP 4. As shown in Figure 1(d), the range of the heteroskedasticity function $\sigma_0(\cdot)$, $[0.3, 6]$, is also quite wide. Strictly speaking, Assumption B2(i) is not satisfied here, as the

https://doi.org/10.1017/S0266466617000111 Published online by Cambridge University Press

TABLE 3. DGP 3, $n = 400$

		Our estimator				Our estimator			Chen's estimator		
x	$\sigma_0(x)$	Bias	SD	RMSE	$m_0(x)$	Bias	SD	RMSE	Bias	SD	RMSE
Rule of thumb with $c = 0.80$											
−0.9	0.659	0.552	0.378	0.672	−0.1	−0.009	0.647	0.647	2.412	2.373	3.382
−0.6	1.304	0.888	0.720	1.146	0.2	−0.272	1.026	1.061	−0.234	2.961	2.967
−0.3	3.158	1.080	0.978	1.458	0.5	0.256	1.583	1.602	−1.596	1.945	2.515
0.3	3.158	0.906	0.876	1.260	1.1	−0.126	1.320	1.325	−2.477	1.414	2.852
0.6	1.304	0.582	0.702	0.912	1.4	−0.119	1.010	1.016	−3.312	1.367	3.583
1.3	0.335	0.114	0.156	0.192	2.1	0.096	0.180	0.203	−3.463	1.056	3.621
1.9	0.162	0.204	0.114	0.234	2.7	−0.169	0.117	0.206	−3.118	1.267	3.365
Rule of thumb with $c = 1.06$											
−0.9	0.659	0.975	0.487	1.090	−0.1	−0.003	0.796	0.795	2.733	2.246	3.536
−0.6	1.304	1.491	0.723	1.656	0.2	0.049	1.332	1.331	0.561	3.016	3.065
−0.3	3.158	1.581	0.719	1.737	0.5	0.397	1.459	1.511	−1.046	1.979	2.236
0.3	3.158	1.165	0.714	1.366	1.1	−0.115	1.13	1.134	−1.972	1.739	2.628
0.6	1.304	0.720	0.649	0.969	1.4	0.168	0.917	0.932	−2.857	1.953	3.460
1.3	0.335	0.152	0.306	0.341	2.1	0.217	0.223	0.311	−3.165	1.277	3.413
1.9	0.162	0.238	0.282	0.369	2.7	−0.108	0.257	0.279	−2.513	1.261	2.811
Rule of thumb with $c = 1.20$											
−0.9	0.659	1.210	0.513	1.314	−0.1	0.031	0.903	0.902	2.901	2.087	3.573
−0.6	1.304	1.884	0.836	2.061	0.2	0.192	1.528	1.538	1.276	2.882	3.149
−0.3	3.158	1.895	0.659	2.006	0.5	0.323	1.376	1.412	−0.421	1.933	1.976
0.3	3.158	1.262	0.623	1.407	1.1	−0.051	1.068	1.068	−1.521	1.932	2.458
0.6	1.304	0.866	0.612	1.061	1.4	0.347	0.833	0.902	−2.259	2.399	3.293
1.3	0.335	0.146	0.290	0.324	2.1	0.329	0.279	0.431	−3.054	1.389	3.354
1.9	0.162	0.264	0.393	0.473	2.7	−0.084	0.277	0.289	−2.160	1.215	2.477
CV (cross-validation)											
−0.9	0.659	0.252	0.391	0.465	−0.1	0.007	0.639	0.639	2.027	2.456	3.183
−0.6	1.304	0.433	0.660	0.789	0.2	−0.226	0.962	0.987	−1.006	2.559	2.747
−0.3	3.158	0.389	1.148	1.211	0.5	0.223	1.585	1.599	−2.156	1.534	2.645
0.3	3.158	0.466	1.254	1.336	1.1	−0.087	1.605	1.605	−2.719	1.279	3.004
0.6	1.304	0.459	0.649	0.794	1.4	−0.376	0.993	1.061	−3.415	1.016	3.562
1.3	0.335	0.123	0.155	0.198	2.1	−0.045	0.195	0.200	−3.623	1.069	3.777
1.9	0.162	0.194	0.056	0.202	2.7	−0.187	0.114	0.218	−3.750	1.191	3.934

heteroskedasticity function $\sigma_0(\cdot)$ is not differentiable at $x = 0$. Even so, our estimator performs well. Similar to DGP 3, our estimator substantially outperforms Chen's for all evaluation points. This DGP confirms that Chen's estimator performs poorly when the heteroskedasticity function has substantial variation globally.

For all four DGPs, the two bandwidth choice rules (the rule of thumb and CV) perform similarly. Also, it appears that the performance of our estimator is not very sensitive to the choice of bandwidth at least for the DGPs considered here. The same phenomenon is also reported for Chen's estimator for the case of homoskedasticity.

https://doi.org/10.1017/S0266466617000111 Published online by Cambridge University Press

TABLE 4. DGP 4, $n = 400$

		Our estimator				Our estimator			Chen's estimator		
x	$\sigma_0(x)$	Bias	SD	RMSE	$m_0(x)$	Bias	SD	RMSE	Bias	SD	RMSE
Rule of thumb with $c = 0.80$											
−0.9	0.600	0.810	0.432	0.918	−0.1	0.243	0.365	0.438	1.840	3.179	3.670
−0.6	0.858	1.284	0.570	1.410	0.2	0.036	0.534	0.534	0.734	3.248	3.327
−0.3	1.500	2.382	0.930	2.562	0.5	0.080	0.892	0.895	−0.590	2.400	2.470
0.3	1.500	2.028	0.870	2.202	1.1	0.247	0.721	0.761	−1.521	1.958	2.478
0.6	0.858	1.032	0.666	1.224	1.4	0.217	0.483	0.529	−2.398	1.924	3.073
1.3	0.426	0.582	0.336	0.672	2.1	0.098	0.216	0.237	−2.046	1.762	2.699
1.9	0.300	0.504	0.282	0.576	2.7	−0.090	0.162	0.186	−1.450	1.319	1.959
Rule of thumb with $c = 1.06$											
−0.9	0.600	1.233	0.438	1.308	−0.1	0.259	0.424	0.497	2.716	2.693	3.823
−0.6	0.858	1.901	0.643	2.006	0.2	0.028	0.578	0.578	1.503	2.991	3.345
−0.3	1.500	3.148	0.790	3.245	0.5	−0.011	0.899	0.898	−0.451	1.936	1.986
0.3	1.500	2.455	0.617	2.531	1.1	0.219	0.451	0.500	−0.875	1.604	1.826
0.6	0.858	1.194	0.509	1.298	1.4	0.344	0.325	0.473	−1.213	2.509	2.784
1.3	0.426	0.634	0.339	0.719	2.1	0.189	0.211	0.283	−0.896	1.772	1.985
1.9	0.300	0.567	0.280	0.632	2.7	−0.084	0.151	0.173	−0.783	1.189	1.423
Rule of thumb with $c = 1.20$											
−0.9	0.600	1.456	0.480	1.533	−0.1	0.284	0.415	0.502	2.690	2.675	3.792
−0.6	0.858	2.204	0.668	2.303	0.2	0.076	0.615	0.619	1.423	2.840	3.174
−0.3	1.500	3.545	0.814	3.637	0.5	−0.060	0.931	0.932	−0.328	1.550	1.583
0.3	1.500	2.608	0.574	2.671	1.1	0.252	0.405	0.477	−0.561	1.470	1.572
0.6	0.858	1.336	0.541	1.442	1.4	0.415	0.392	0.570	−0.566	2.561	2.620
1.3	0.426	0.682	0.387	0.784	2.1	0.243	0.242	0.343	−0.447	2.026	2.073
1.9	0.300	0.632	0.401	0.748	2.7	−0.071	0.186	0.199	−0.540	1.349	1.452
CV (cross-validation)											
−0.9	0.600	0.478	0.486	0.687	−0.1	0.153	0.435	0.461	0.695	3.238	3.309
−0.6	0.857	0.741	0.627	0.971	0.2	0.042	0.528	0.529	−0.858	2.937	3.058
−0.3	1.500	1.331	1.124	1.733	0.5	0.107	0.894	0.900	−1.649	2.315	2.841
0.3	1.500	1.377	1.121	1.778	1.1	−0.022	0.877	0.877	−2.547	1.610	3.013
0.6	0.857	0.912	0.727	1.160	1.4	−0.107	0.586	0.595	−3.184	1.353	3.459
1.3	0.429	0.496	0.367	0.608	2.1	−0.060	0.251	0.257	−3.076	1.555	3.447
1.9	0.300	0.418	0.319	0.518	2.7	−0.132	0.269	0.299	−2.667	1.668	3.145

5. CONCLUSION

In this paper, we establish conditions for the nonparametric identification of regression and heteroskedasticity functions in truncated regression models with unknown heteroskedasticity. Building on the constructive identification results, a kernel-based estimation procedure is proposed and is shown to have desirable asymptotic properties. Our Monte Carlo simulations demonstrate that in the presence of heteroskedasticity, our new estimator performs much better than the existing method that ignores the heteroskedasticity.

https://doi.org/10.1017/S0266466617000111 Published online by Cambridge University Press

There are many interesting topics for further research. First, asymptotic theories should be established for the estimators of the conditional mean and conditional quantile functions and the error distribution. Second, it is desirable to develop specification tests for homoskedasticity for nonparametric truncated models. Third, applying our methods to some truncated data empirically might yield interesting insights.

NOTES

1. Both LL and Chen briefly discuss a form of heteroskedasticity in their concluding remarks. They consider

$$Y^* = m_0(X) + \eta,$$

where $X = (X_1, X_2)$. They assume that X_1 and η are conditionally independent given X_2. Thus, η is allowed to depend on X_2 in an arbitrary way but cannot depend on X_1. Here, we allow η to depend on both X_1 and X_2.

2. There is also a relatively large body of literature on the parametric or semiparametric estimation of truncated regression models (see, e.g., Powell (1986), Lee (1992), Honoré and Powell (1994), Newey (2001, 2004), Cosslett (2004), Khan and Lewbel (2007), and Chen and Zhou (2012)).

3. It is not clear how to extend the method in LL to the heteroskedasticity in equation (1.1).

4. In this section of identification, we suppress the observation subscript i for notational convenience.

5. If $P\left(Y^* > 0 | X = x\right) = 0$, neither $m_0(x)$ nor $\sigma_0(x)$ can be identified.

6. When X and ε are not independent, the endogenous problem arises. Typically, we need to rely on certain instrumental variables, in which case the identification and estimation would become much more complicated. We leave this topic for future research.

7. Suppose $\mathcal{E}$ follows a Pareto distribution on $[\underline{E}, \infty)$ with $\underline{E} > 0$:

$$P(\mathcal{E} > t) = \left(\frac{\underline{E}}{t}\right)^{c_1} \text{ for all } t \geq \underline{E}, \text{ where } c_1 > 0.$$

Consider $\varepsilon = a\mathcal{E} - c_2$, where $a > 0$. Then, the survival function of ε is

$$S(t) = \frac{c_3}{(t + c_2)^{c_1}} \text{ for all } t \geq a\underline{E} - c_2, \text{ where } c_3 = \left(a\underline{E}\right)^{c_1}.$$

8. To avoid technical difficulty, we adopt smoothing in the conditional survival function and the weight function $w(\cdot)$. For the smoothed $w(\cdot)$, the cut-off point ε_0 is needed, as stated in Assumption B2. Correspondingly, we need to adjust our criterion function. Thus, we base our estimator on the following population equations. For all $t \geq \varepsilon_0$,

$$\frac{S^*(t, x_0)}{S^*(\varepsilon_0, x_0)} = \frac{S^*(t\sigma_0(x) + m_0(x), x)}{S^*(\varepsilon_0\sigma_0(x) + m_0(x), x)} \text{ when } m_0(x) \geq 0, \text{ and}$$

$$\frac{S^*(t, x)}{S^*(\varepsilon_0, x)} = \frac{S^*\left(\frac{t - m_0(x)}{\sigma_0(x)}, x_0\right)}{S^*\left(\frac{\varepsilon_0 - m_0(x)}{\sigma_0(x)}, x_0\right)} \text{ when } m_0(x) < 0.$$

9. Here we assume that the small number ε_0 is fixed. Therefore we do not make use of the information for $t \in [0, \varepsilon_0)$. As suggested by an anonymous referee, alternatively, we could let ε_0 go to zero at an appropriate rate to improve the estimation efficiency. Through simulations, we find that the estimation results are not sensitive to the choice of ε_0.

10. We may also consider nonparametric bootstrap. However, it is challenging to establish its validity theoretically.

https://doi.org/10.1017/S0266466617000111 Published online by Cambridge University Press

11. We also try the logistic distribution and a mixture distribution. The results are available upon request.
12. The weight function is often used to handle the boundary bias or random denominator issue.
13. The results for $n = 800$ are also available upon request.

REFERENCES

Akritas, M. & I. Van Keilegom (2001) Nonparametric estimation of the residual distribution. *Scandinavian Journal of Statistics* 28, 549–567.

Amemiya, T. (1985) *Advanced Econometrics*. Harvard University Press.

Cosslett, S. (2004) Efficient semiparametric estimation of censored and truncated regressions via smoothed self-consistency equation. *Econometrica* 72, 1277–1293.

Chen, S. (2010) Nonparametric identification and estimation of truncated regression models. *Review of Economic Studies* 77, 127–153.

Chen, S., G. Dahl, & S. Khan (2005) Nonparametric identification and estimation of a censored location-scale regression model. *Journal of the American Statistical Association* 100, 212–221.

Chen, S. & X. Zhou (2012) Semiparametric estimation of a truncated regression model. *Journal of Econometrics* 167, 297–304.

Fan, J. & I. Gijbels (1994) Censored regression: Local linear approximations and their applications. *Journal of the American Statistical Association* 89, 560–570.

Fan, J. & I. Gijbels (1996) *Local Polynomial Modelling and its Applications.* Chapman and Hall.

Heuchenne, C. & I. Van Keilegom (2010) Estimation in nonparametric location-scale regression models with censored data. *Annals of the Institute of Statistical Mathematics* 62, 439–463.

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

Khan, S. & A. Lewbel (2007) Weighted and two stage least squares estimation of semiparametric truncated regression models. *Econometric Theory* 23, 309–347.

Lee, L.-F. (1992) Semiparametric nonlinear least-square estimation of truncated regression models. *Econometric Theory* 8, 52–94.

Lewbel, A. & O. Linton (2002) Nonparametric censored and truncated regression. *Econometrica* 70, 765–779.

Li, Q. & J. Racine (2003) Nonparametric estimation of distributions with categorical and continuous data. *Journal of Multivariate Analysis* 86, 266–292.

Matzkin, R. (2007) Nonparametric identification. In J.J. Heckman & E.E. Leamer (eds.), *The Handbook of Econometrics*, vol. 6B, pp. 5307–5368. North-Holland.

Matzkin, R. (2012) Identification in nonparametric limited dependent variable models with simultaneity and unobserved heterogeneity. *Journal of Econometrics* 166, 106–115.

Newey, W.K. (1997) Convergence rates and asymptotic normality for series estimators. *Journal of Econometrics* 79, 147–168.

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

Newey, W.K. (2004) Efficient semiparametric estimation via moment restrictions. *Econometrica* 72, 1877–1897.

Politis, D., J. Romano, & M. Wolf (1999) *Subsampling*. Springer-Verlag.

Pollard, D. (1995) Uniform ratio limit theorems for empirical processes. *Scandinavian Journal of Statistics* 22, 271–278.

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

Silverman, B.W. (1986) *Density Estimation for Statistics and Data Analysis*. Chapman and Hall.

Van Keilegom, I. & G. Akritas (1999) Transfer of tail information in censored regression models. *Annals of Statistics* 27, 1745–1784.

https://doi.org/10.1017/S0266466617000111 Published online by Cambridge University Press

APPENDIX

Proof of Lemma 2.1. We first show that $T(m,\sigma)$ in equation (2.5) is well defined for any m and $\sigma \geq \sigma^*$, where σ^* is a positive constant. First, consider $T_1(m,\sigma)$ for $m \geq 0$. It is clear that

$$\begin{aligned}
T_1(m,\sigma) &= \int_0^\infty \left[S^*(t,x_0) - \frac{S^*(t\sigma+m,x)}{S^*(m,x)} \right]^2 dt \\
&\leq \int_0^\infty \left[S^*(t,x_0) + \frac{S^*(t\sigma+m,x)}{S^*(m,x)} \right] dt = \int_0^\infty \left[\frac{S(t)}{S(0)} + \frac{S\left(\frac{t\sigma+m-m_0(x)}{\sigma_0(x)}\right)}{S\left(\frac{m-m_0(x)}{\sigma_0(x)}\right)} \right] dt \\
&= \frac{1}{S(0)} \int_0^\infty t f(t) dt + \frac{1}{S\left(\frac{m-m_0(x)}{\sigma_0(x)}\right)} \frac{\sigma_0(x)}{\sigma} \int_{\frac{m-m_0(x)}{\sigma_0(x)}}^\infty S(u) du \\
&= \mathbb{E}[\varepsilon | \varepsilon > 0] + \frac{\sigma_0(x)}{\sigma} \mathbb{E}\left[\left(\varepsilon - \frac{(m-m_0(x))}{\sigma_0(x)} \right) \middle| \left(\varepsilon - \frac{m-m_0(x)}{\sigma_0(x)} \right) > 0 \right] \\
&< \infty,
\end{aligned}$$

where the last two equalities use $\lim_{t\to\infty} tS(t) = 0$, and the last inequality follows from the existence of the first moment for ε. Similarly, we can show that $T_2(m,\sigma)$ is also well defined for $m < 0$.

Clearly, $T(m_0(x),\sigma_0(x)) = 0$. We will show that $(m_0(x),\sigma_0(x))$ is the unique solution of $T(m,\sigma) = 0$. We first consider the case where $m_0(x) \geq 0$. Suppose that there is a pair $(\tilde{m}_x,\tilde{\sigma}_x) \neq (m_0(x),\sigma_0(x))$ such that $T(\tilde{m}_x,\tilde{\sigma}_x) = 0$, where $\tilde{\sigma}_x \geq \sigma^* > 0$. Without loss of generality, further assume that $\tilde{m}_x \geq m_0(x) \geq 0$. It follows from $T(\tilde{m}_x,\tilde{\sigma}_x) = 0$ that for all $t \geq 0$,

$$S^*(t,x_0) = \frac{S^*(t\tilde{\sigma}_x+\tilde{m}_x,x)}{S^*(\tilde{m}_x,x)} \text{ or } \frac{S(t)}{S(0)} = \frac{S\left(\frac{t\tilde{\sigma}_x+\tilde{m}_x-m_0(x)}{\sigma_0(x)}\right)}{S\left(\frac{\tilde{m}_x-m_0(x)}{\sigma_0(x)}\right)}.$$

Define $\alpha_x \equiv \frac{\tilde{m}_x-m_0(x)}{\sigma_0(x)}$ and $\beta_x \equiv \frac{\tilde{\sigma}_x}{\sigma_0(x)}$, then $\beta_x > 0$ and $(\alpha_x,\beta_x) \neq (0,1)$. Then, we have

$$\frac{S(t)}{S(0)} = \frac{S(\beta_x t+\alpha_x)}{S(\alpha_x)} \text{ for all } t \geq 0.$$

Therefore, $S(\cdot)$ can be expressed in the form of equation (2.7) for all $t \geq 0$, which leads to a contradiction.

Next, we consider the case where $m_0(x) \leq 0$. Similarly, suppose that there is a pair $(\tilde{m}_x,\tilde{\sigma}_x) \neq (m_0(x),\sigma_0(x))$ such that $T(\tilde{m}_x,\tilde{\sigma}_x) = 0$, where $\tilde{\sigma}_x \geq \sigma^* > 0$. Without loss of generality, further assume that $0 \geq \tilde{m}_x \geq m_0(x)$. It follows from $T(\tilde{m}_x,\tilde{\sigma}_x) = 0$ that

$$S^*(t,x) = \frac{S^*\left(\frac{t-\tilde{m}_x}{\tilde{\sigma}_x},x_0\right)}{S^*\left(\frac{-\tilde{m}_x}{\tilde{\sigma}_x},x_0\right)} \text{ for all } t \geq 0.$$

Further, note that for all $t \geq 0$,

$$S^*(t,x) = \frac{S\left(\frac{t-m_0(x)}{\sigma_0(x)}\right)}{S\left(\frac{-m_0(x)}{\sigma_0(x)}\right)} \text{ and } \frac{S^*\left(\frac{t-\tilde{m}_x}{\tilde{\sigma}_x},x_0\right)}{S^*\left(\frac{-\tilde{m}_x}{\tilde{\sigma}_x},x_0\right)} = \frac{S\left(\frac{t-\tilde{m}_x}{\tilde{\sigma}_x}\right)}{S\left(\frac{-\tilde{m}_x}{\tilde{\sigma}_x}\right)}.$$

https://doi.org/10.1017/S0266466617000111 Published online by Cambridge University Press

Therefore,

$$\frac{S\left(\frac{t-m_0(x)}{\sigma_0(x)}\right)}{S\left(\frac{-m_0(x)}{\sigma_0(x)}\right)} = \frac{S\left(\frac{t-\tilde{m}_x}{\tilde{\sigma}_x}\right)}{S\left(\frac{-\tilde{m}_x}{\tilde{\sigma}_x}\right)} \text{ for all } t \geq 0. \tag{A.1}$$

Define $T \equiv \frac{t-m_0(x)}{\sigma_0(x)}$, $\alpha_x \equiv \frac{m_0(x)-\tilde{m}_x}{\tilde{\sigma}_x}$ and $\beta_x \equiv \frac{\sigma_0(x)}{\tilde{\sigma}_x}$. Then $\beta_x > 0$ and $(\alpha_x, \beta_x) \neq (0,1)$. Equation (A.1) implies that

$$\frac{S(T)}{S(\delta_x)} = \frac{S(\beta_x T + \alpha_x)}{S(\beta_x \delta_x + \alpha_x)} \text{ for all } T \geq \delta_x = -\frac{m_0(x)}{\sigma_0(x)}.$$

Note that when $t \geq 0$, $T \geq -\frac{m_0(x)}{\sigma_0(x)}$. Therefore, $S(\cdot)$ can be expressed in the form of equation (2.7) for all $t \geq \delta_x$, which leads to a contradiction. ■

Proof of Lemma 2.2. We first show that the two sets of parameters, $(m_0(x), \sigma_0(x))$ and $(m_0(x)+\alpha\sigma_0(x), \beta\sigma_0(x))$, give the same conditional survival function of Y given X. Let S^* and $\tilde{S}^*$ be the corresponding conditional survival functions of Y given X with these two sets of parameters. Specifically, when the parameters are $(m_0(x), \sigma_0(x))$,

$$S^*(t,x) = P\left(Y^* > t | Y^* > 0, X = x\right) = \frac{P\left(\varepsilon > \frac{t-m_0(X)}{\sigma_0(X)} \middle| X = x\right)}{P\left(\varepsilon > \frac{-m_0(X)}{\sigma_0(X)} \middle| X = x\right)} = \frac{S\left(\frac{t-m_0(x)}{\sigma_0(x)}\right)}{S\left(\frac{-m_0(x)}{\sigma_0(x)}\right)} \text{ for all } t \geq 0.$$

Consider the alternative set of parameters: $(m_0(x)+\alpha\sigma_0(x), \beta\sigma_0(x))$, then

$$\begin{aligned}\tilde{S}^*(t,x) &= P\left(Y^* > t | Y^* > 0, X = x\right) = \frac{P\left(\varepsilon > \frac{t-m_0(X)-\alpha\sigma_0(X)}{\beta\sigma_0(X)} \middle| X = x\right)}{P\left(\varepsilon > \frac{-m_0(X)-\alpha\sigma_0(X)}{\beta\sigma_0(X)} \middle| X = x\right)} \\ &= \frac{S\left(\frac{t-m_0(x)-\alpha\sigma_0(x)}{\beta\sigma_0(x)}\right)}{S\left(\frac{-m_0(x)-\alpha\sigma_0(x)}{\beta\sigma_0(x)}\right)} \text{ for all } t \geq 0. \end{aligned} \tag{A.2}$$

Then by applying equation (2.8) to both the numerator and denominator in equation (A.2), we have for all $t \geq 0$,

$$S\left(\frac{t-m_0(x)-\alpha\sigma_0(x)}{\beta\sigma_0(x)}\right) = \frac{S(\delta_x)}{S(\beta\delta_x+\alpha)} S\left(\beta\left(\frac{t-m_0(x)-\alpha\sigma_0(x)}{\beta\sigma_0(x)}\right)+\alpha\right) = \frac{S(\delta_x)}{S(\beta\delta_x+\alpha)} S\left(\frac{t-m_0(x)}{\sigma_0(x)}\right) \text{ and}$$

$$S\left(-\frac{m_0(x)+\alpha\sigma_0(x)}{\beta\sigma_0(x)}\right) = \frac{S(\delta_x)}{S(\beta\delta_x+\alpha)} S\left(\beta\left(-\frac{m_0(x)+\alpha\sigma_0(x)}{\beta\sigma_0(x)}\right)+\alpha\right) = \frac{S(\delta_x)}{S(\beta\delta_x+\alpha)} S\left(-\frac{m_0(x)}{\sigma_0(x)}\right)$$

Therefore,

$$\tilde{S}^*(t,x) = \frac{S\left(\frac{t-m_0(x)}{\sigma_0(x)}\right)}{S\left(-\frac{m_0(x)}{\sigma_0(x)}\right)} = S^*(t,x).$$

Next, we show that the distributions of (Y^*, X) given $Y^* > 0$ under the two sets of parameters are identical. Let P^* and $\tilde{P}^*$ be the corresponding distributions of (Y^*, X) given Y^* under these two sets of parameters. Let p^* be the density function of X given $Y^* > 0$.

https://doi.org/10.1017/S0266466617000111 Published online by Cambridge University Press

Note that p^* is the same under the two sets of parameters, as it is directly identified from the data. Therefore, we have

$$P^*(t,x) = \Pr\left(Y^* \le t, X \le x \mid Y^* > 0\right) = \int_{-\infty}^{x} \left[1 - S^*(t,s)\right] p^*(s)\, ds \text{ and}$$

$$\tilde{P}^*(t,x) = \Pr\left(Y^* \le t, X \le x \mid Y^* > 0\right) = \int_{-\infty}^{x} \left[1 - \tilde{S}^*(t,s)\right] p^*(s)\, ds.$$

Given that $S^*(t,s) = \tilde{S}^*(t,s)$, we have $P^*(t,x) = \tilde{P}^*(t,x)$. We have shown that the observed data have the same distribution under the two sets of parameters, therefore these two sets of parameters are observationally equivalent. ■

Proof of Lemma 2.3. For part (i), we show that the following three cases are impossible: (a) $\frac{\alpha}{1-\beta} = \delta_x$ and $\beta \neq 1$, (b) $\frac{\alpha}{1-\beta} > \delta_x$ and $\beta \neq 1$, and $(c)\, x$ $\frac{\alpha}{1-\beta} < \delta_x$ and $\beta \neq 1$. For ease of reference, we restate the key equation for identification (equation (2.7)) here:

$$\frac{S(t)}{S(\delta_x)} = \frac{S(\beta t + \alpha)}{S(\beta\delta_x + \alpha)} \text{ for all } t \ge \delta_x. \tag{A.3}$$

Consider case (a). Then equation (A.3) implies that

$$S(t) = S(\beta t + \alpha) \text{ for all } t \ge \delta_x, \tag{A.4}$$

as $\delta_x = \beta\delta_x + \alpha$. Recursively applying equation (A.4) for k times yields

$$S(t) = S\left(\beta^k\left(t - \frac{\alpha}{1-\beta}\right) + \frac{\alpha}{1-\beta}\right).$$

Letting $k \to \infty$, for all $t > \delta_x$, we have

$$S(t) = S(\infty) = 0 \text{ when } \beta > 1 \text{ and } S(t) = S\left(\frac{\alpha}{1-\beta}\right) \text{ when } \beta < 1.$$

Both contradict the assumption that ε has unbounded support from the right. Therefore case (a) is impossible.

Consider case (b). Letting $t = \frac{\alpha}{1-\beta} > \delta_x$, equation (A.3) implies that

$$\frac{S\left(\frac{\alpha}{1-\beta}\right)}{S(\delta_x)} = \frac{S\left(\frac{\alpha}{1-\beta}\right)}{S(\beta\delta_x + \alpha)}.$$

This equation holds either if $S(\delta_x) = S(\beta\delta_x + \alpha)$ or if $S\left(\frac{\alpha}{1-\beta}\right) = 0$. $S\left(\frac{\alpha}{1-\beta}\right) = 0$ contradicts the assumption that ε has unbounded support from the right. $S(\delta_x) = S(\beta\delta_x + \alpha)$ implies that

$$S(t) = S(\beta t + \alpha) \text{ for all } t \ge \delta_x. \tag{A.5}$$

Following the proof for case (a), we can show that equation (A.5) cannot hold if $\beta \neq 1$. Therefore case (b) is impossible.

Consider case (c). Recursively applying equation (2.7) for k times yields

$$1 - F(t) = \left(\frac{1 - F(\delta_x)}{1 - F(\beta\delta_x + \alpha)}\right)^k \cdot \left[1 - F\left(\beta^k\left(t - \frac{\alpha}{1-\beta}\right) + \frac{\alpha}{1-\beta}\right)\right].$$

https://doi.org/10.1017/S0266466617000111 Published online by Cambridge University Press

Define $s = \beta^k \left(t - \frac{\alpha}{1-\beta}\right)$. Then, for all $t \geq \delta_x > \frac{\alpha}{1-\beta}$, the equation above becomes

$$1 - F(t) = \left(\frac{1 - F(\delta_x)}{1 - F(\beta\delta_x + \alpha)}\right)^{\frac{\ln(s) - \ln\left(t - \frac{\alpha}{1-\beta}\right)}{\ln(\beta)}} \cdot \left[1 - F\left(s + \frac{\alpha}{1-\beta}\right)\right] = \frac{s^r \cdot \left[1 - F\left(s + \frac{\alpha}{1-\beta}\right)\right]}{\left(t - \frac{\alpha}{1-\beta}\right)^r},$$

where $r = \frac{\ln\left(\frac{1-F(\delta_x)}{1-F(\beta\delta_x+\alpha)}\right)}{\ln(\beta)}$. Consider two subcases: $(c.i)$: $\beta > 1$ and $(c.ii)$: $\beta < 1$. For subcase $(c.i)$, letting $k \to \infty$, i.e., $s \to \infty$, then

$$1 - F(t) = \frac{\lim_{s\to\infty} s^r \cdot \left[1 - F\left(s + \frac{\alpha}{1-\beta}\right)\right]}{\left(t - \frac{\alpha}{1-\beta}\right)^r} \text{ for all } t \geq \delta_x,$$

where $r > 0$. If $\lim_{s\to\infty} s^r \cdot (1 - F(s)) = \infty$, then $F(t) = -\infty$ for all $t \geq \delta_x$, which is impossible. If $\lim_{s\to\infty} s^r \cdot (1 - F(s)) = 0$, then $F(t) = 1$ for all $t \geq \delta_x$, which is also impossible. If $\lim_{s\to\infty} s^r \cdot (1 - F(s))$ is some positive constant, then equation (2.10) is satisfied. This leads to a contradiction. For subcase $(c.ii)$, letting $k \to \infty$, i.e., $s \to 0$, then

$$1 - F(t) = \frac{\lim_{s\to 0} s^r \cdot \left[1 - F\left(s + \frac{\alpha}{1-\beta}\right)\right]}{\left(t - \frac{\alpha}{1-\beta}\right)^r} \text{ for all } t \geq \delta_x,$$

where $r > 0$. This means that $F(t) = 1$ for all $t \geq \delta_x$, which is impossible. Therefore, case (c) is impossible. Thus, the proof of part (i) is complete. The proof of part (ii) follows directly from Lemma 1 in Chen. ■

Proof of Lemma 2.4. We prove this lemma by showing the following four cases are impossible: (a) $\beta = 1$, $\alpha \neq 0$, (b) $\frac{\alpha}{1-\beta} < \bar{K}$, $\beta \neq 1$, (c) $\frac{\alpha}{1-\beta} > \bar{K}$, $\beta \neq 1$, (d) $\frac{\alpha}{1-\beta} = \bar{K}$, $\beta \neq 1$. For ease of reference, we restate the key equation for identification (equation (2.7)) here:

$$\frac{S(t)}{S(\delta_x)} = \frac{S(\beta t + \alpha)}{S(\beta\delta_x + \alpha)} \text{ for all } t \geq \delta_x. \tag{A.6}$$

Consider case (a). Equation (A.6) implies that

$$\frac{S(t)}{S(\delta_x)} = \frac{S(t + \alpha)}{S(\delta_x + \alpha)} \text{ for all } t \geq \delta_x. \tag{A.7}$$

Suppose that $\alpha > 0$. Let $t = \bar{K} - \epsilon$ where $0 < \epsilon < a$. Then $t < \bar{K}$ and $t + \alpha > \bar{K}$. Therefore the RHS of equation (A.7) is zero, while the LHS is positive. This is impossible. Similarly, we can show that it is impossible when $\alpha < 0$. Therefore, case (a) is impossible.

Consider case (b). We further consider two subcases: $(b.i)$: $\beta < 1$ and $(b.ii)$: $\beta > 1$. For subcase $(b.i)$, let $t = \bar{K}$ and $\beta t + \alpha = \beta\bar{K} + \alpha < \bar{K}$.[14] Then the LHS of equation (A.6) is zero, while the RHS is positive, which leads to a contradiction. For subcase $(b.ii)$, let $t = \frac{\bar{K}-\alpha}{\beta} < \bar{K}$,[15] and $\beta t + \alpha = \bar{K}$. Then the LHS of equation (A.6) is positive, while the RHS is zero, which leads to a contradiction. Therefore, case (b) is impossible.

Consider case (c). Similarly, we further consider two subcases: $(c.i)$: $\beta < 1$ and $(c.ii)$: $\beta > 1$. For subcase $(c.i)$, find an ϵ such that $0 < \epsilon < \frac{(\beta-1)\bar{K}+\alpha}{\beta}$.[16] Let $t = \bar{K} - \epsilon <$

https://doi.org/10.1017/S0266466617000111 Published online by Cambridge University Press

$\bar{K}$ and $\beta t+\alpha=\beta\left(\bar{K}-\epsilon\right)+\alpha>\bar{K}$. Then the LHS of equation (A.6) is positive, while the RHS is zero, which leads to a contradiction. For subcase $(c.ii)$, find an ϵ such that $0<\epsilon<(1-\beta)\,\bar{K}-\alpha$.[17] Let $t=\frac{\bar{K}-\alpha-\epsilon}{\beta}>\bar{K}$, and $\beta t+\alpha=\bar{K}-\epsilon<\bar{K}$. Then the LHS of equation (A.6) is zero, while the RHS is positive, which leads to a contradiction. Therefore, case (c) is impossible.

Finally, consider case (d). Recursively substituting equation (A.6) for k times yields

$$1-F(t)=\left(\frac{1-F(\delta_x)}{1-F(\beta\delta_x+\alpha)}\right)^k\cdot\left[1-F\left(\beta^k\left(t-\frac{\alpha}{1-\beta}\right)+\frac{\alpha}{1-\beta}\right)\right].$$

Substituting $\frac{\alpha}{1-\beta}=\bar{K}$ yields

$$1-F(t)=\left(\frac{1-F(\delta_x)}{1-F(\beta\delta_x+\alpha)}\right)^k\cdot\left[1-F\left(\bar{K}-\beta^k\left(\bar{K}-t\right)\right)\right].$$

Define $s=\beta^k\left(\bar{K}-t\right)$. Then, the equation above becomes

$$1-F(t)=\left(\frac{1-F(\delta_x)}{1-F(\beta\delta_x+\alpha)}\right)^{\frac{\ln(s)-\ln\left(\bar{K}-t\right)}{\ln(\beta)}}\cdot\left[1-F\left(\bar{K}-s\right)\right]=\left(\bar{K}-t\right)^r\left[\frac{1-F\left(\bar{K}-s\right)}{s^r}\right],\quad\textbf{(A.8)}$$

where $r=-\frac{1}{\ln(\beta)}\cdot\ln\left(\frac{1-F(\delta_x)}{1-F(\beta\delta_x+\alpha)}\right)$. We further consider two subcases: $(d.i)$: $\beta<1$ and $(d.ii)$: $\beta>1$. For subcase $(d.i)$, let $k\to\infty$ or $s\to+0$. Then, equation (A.8) implies that for $\delta_x\le t<\bar{K}$,

$$1-F(t)=\left(\bar{K}-t\right)^r\cdot\lim_{s\to+0}\left[\frac{1-F\left(\bar{K}-s\right)}{s^r}\right]$$

where $r>0$. By assumption, $\lim_{s\to+0}\frac{1-F\left(\bar{K}-s\right)}{s^r}$ exists. $\lim_{s\to+0}\frac{1-F\left(\bar{K}-s\right)}{s^r}=0$ and $\lim_{s\to+0}\frac{1-F\left(\bar{K}-s\right)}{s^r}=\infty$ imply that $F(t)=1$ and $F(t)=0$ for all $\delta_x\le t<\bar{K}$, respectively, both of which are impossible. If $\lim_{s\to+0}\frac{\left(1-F\left(\bar{K}-s\right)\right)}{s^r}$ equals some positive constant, then equation (2.11) holds. This leads to a contradiction. For subcase $(d.ii)$, let $k\to\infty$ or $s\to\infty$. Then, equation (A.8) implies that for $\delta_x\le t<\bar{K}$,

$$1-F(t)=\left(\bar{K}-t\right)^r\cdot\lim_{s\to\infty}\left[\frac{1-F\left(\bar{K}-s\right)}{s^r}\right],$$

where $r>0$. This means that $F(t)=1$ for all $\delta_x\le t<\bar{K}$, which is impossible. Therefore, case (d) is impossible. ■

Proof of Lemma 2.5. Observe that

$$H(u)=\mathbb{E}\left(\frac{Y^*}{\sigma_0(X)}\middle|\frac{m_0(X)}{\sigma_0(X)}=u,Y^*>0\right)=u+\mathbb{E}(\varepsilon|\varepsilon>-u)=u+\frac{\int_{-u}^{\bar{K}}sf(s)\,ds}{\int_{-u}^{\bar{K}}f(s)\,ds}$$

$$=u+\frac{\int_{-u}^{\bar{K}}sf(s)\,ds}{1-F(-u)},$$

https://doi.org/10.1017/S0266466617000111 Published online by Cambridge University Press

where $f(\cdot)$ is the density function of ε. This is equivalent to

$$[H(u)-u]\cdot(1-F(-u))=\int_{-u}^{\bar{K}} sf(s)\,ds.$$

Taking the derivative with respect to u on both sides yields

$$\frac{f(-u)}{1-F(-u)}=\frac{1-H'(u)}{H(u)}.$$

Integrating over $u\in[e,-\underline{K}]$ yields

$$\int_e^{-\underline{K}}\frac{f(-u)}{1-F(-u)}du=\int_e^{-\underline{K}}\frac{1-H'(u)}{H(u)}du.$$

This implies that

$$\ln(1-F(-u))|_e^{-\underline{K}}=\int_e^{-\underline{K}}\left(\frac{1-H'(u)}{H(u)}\right)du.$$

Hence,

$$F(-e)=1-\exp\left(-\int_e^{-\underline{K}}\left(\frac{1-H'(u)}{H(u)}\right)du\right),$$

which is equivalent to

$$F(e)=1-\exp\left(\int_{-e}^{-\underline{K}}\left(\frac{H'(u)-1}{H(u)}\right)du\right).$$ ■

Proof of Theorem 3.1. Similar to Chen, under our Assumptions B1–B4, we can follow Pollard (1995) to establish that $\sup_{t\in[\varepsilon_0,\infty)}|S_n^*(t,\tilde{x})-S^*(t,\tilde{x})|=O_p(\delta_n)$, $\sup_{t\in[\varepsilon_0,\infty)}|\partial S_n^*(t,\tilde{x})/\partial t-\partial S^*(t,\tilde{x})/\partial t|=O_p(\delta_{n1})$, and $\sup_{t\in[\varepsilon_0,\infty)}|\partial^2 S_n^*(t,\tilde{x})/\partial t^2-\partial^2 S^*(t,\tilde{x})/\partial t^2|=O_p(\delta_{n2})$ for $\tilde{x}$ uniformly in some neighborhoods of x and x_0, where $\delta_n=(\ln n/nh^d)^{1/2}+h^q+h_y^{q_y}$, $\delta_{n1}=(\ln n/nh^d h_y)^{1/2}+h^q+h_y^{q_y}$, $\delta_{n2}=(\ln n/nh^d h_y^3)^{1/2}+h^q+h_y^{q_y}$. Then, $T_n(m,\sigma)$ converges to

$$T^0(m,\sigma)=\int_0^\infty\left(\frac{S^*(t,x_0)}{S^*(\varepsilon_0,x_0)}-\frac{S^*(t\sigma+m,x)}{S^*(\varepsilon_0\sigma+m,x)}\right)^2 w(m)w_1(t)dt$$
$$+\int_0^\infty\left(\frac{S^*(t,x)}{S^*(\varepsilon_0,x)}-\frac{S^*(\frac{t-m}{\sigma},x_0)}{S^*(\frac{\varepsilon_0-m}{\sigma},x_0)}\right)^2 w^c(m)w_1(t)dt$$

uniformly in $(m,\sigma)\in M$. By Assumptions A1–A2 and Lemma 2.1, $T^0(m,\sigma)$ achieves the unique minimum at $(m_0(x),\sigma_0(x))$. Consistency follows from the standard arguments in Amemiya (1985).

For the asymptotic normality, we first define

$$V(x)=\int\Lambda_1(t,x)w_1(t)dt\cdot w(m_0(x))+\int\Lambda_2(t,x)w_1(t)dt\cdot w^c(m_0(x)),$$

https://doi.org/10.1017/S0266466617000111 Published online by Cambridge University Press

where

$$\Lambda_1(t,x)=\begin{pmatrix}\lambda_1^2(t,x) & \lambda_1(t,x)\lambda_2(t,x)\\ \lambda_1(t,x)\lambda_2(t,x) & \lambda_2^2(t,x)\end{pmatrix},\quad \Lambda_2(t,x)=\begin{pmatrix}\gamma_1^2(t,x) & \gamma_1(t,x)\gamma_2(t,x)\\ \gamma_1(t,x)\gamma_2(t,x) & \gamma_2^2(t,x)\end{pmatrix}$$

$$\lambda_1(t,x)=\frac{S(t)}{\sigma_0(x)S(\varepsilon_0)}(\lambda(\varepsilon_0)-\lambda(t)),\qquad \lambda_2(t,x)=\frac{S(t)}{\sigma_0(x)S(\varepsilon_0)}(\varepsilon_0\lambda(\varepsilon_0)-t\lambda(t))$$

$$\gamma_1(t,x)=\frac{S\left(\frac{t-m_0(x)}{\sigma_0(x)}\right)}{\sigma_0^2(x)S\left(\frac{\varepsilon_0-m_0(x)}{\sigma_0(x)}\right)}\left(\lambda\left(\frac{\varepsilon_0-m_0(x)}{\sigma_0(x)}\right)-\lambda\left(\frac{t-m_0(x)}{\sigma_0(x)}\right)\right)\text{ and}$$

$$\gamma_2(t,x)=\frac{S\left(\frac{t-m_0(x)}{\sigma_0(x)}\right)}{\sigma_0^2(x)S\left(\frac{\varepsilon_0-m_0(x)}{\sigma_0(x)}\right)}\left(\lambda\left(\frac{\varepsilon_0-m_0(x)}{\sigma_0(x)}\right)(\varepsilon_0-m_0(x))-\lambda\left(\frac{t-m_0(x)}{\sigma_0(x)}\right)(t-m_0(x))\right).$$

It follows from a Taylor expansion of the first-order condition $\nabla T_n(\hat{m}(x),\,\hat{\sigma}(x))=0$ that

$$\nabla T_n(m_0(x),\sigma_0(x))+\nabla^2 T_n(m^*(x),\sigma^*(x))\begin{pmatrix}\hat{m}(x)-m_0(x)\\ \hat{\sigma}(x)-\sigma_0(x)\end{pmatrix}=0,$$

where $(m^*(x),\sigma^*(x))$ is between $(m_0(x),\sigma_0(x))$ and $(\hat{m}(x),\hat{\sigma}(x))$. Then,

$$\begin{pmatrix}\hat{m}(x)-m_0(x)\\ \hat{\sigma}(x)-\sigma_0(x)\end{pmatrix}=V(x)^{-1}U_n(x)+o_p\left((nh^d)^{-1/2}\right)\begin{pmatrix}1\\1\end{pmatrix},$$

where

$$U_n(x)=U_{n1}(x)w(m_0(x))+U_{n2}(x)w^c(m_0(x)),$$

with

$$U_{n1}(x)=\int_0^\infty\left[\frac{S_n^*(t,x_0)}{S_n^*(\varepsilon_0,x_0)}-\frac{S_n^*(t\sigma_0(x)+m_0(x),x)}{S_n^*(\varepsilon_0\sigma_0(x)+m_0(x),x)}\right]\begin{pmatrix}\lambda_1(t,x)\\ \lambda_2(t,x)\end{pmatrix}w_1(t)dt\text{ and}$$

$$U_{n2}(x)=\int_0^\infty\left[\frac{S_n^*(t,x)}{S_n^*(\varepsilon_0,x)}-\frac{S^*\left(\frac{t-m_0(x)}{\sigma_0(x)},x_0\right)}{S^*\left(\frac{\varepsilon_0-m_0(x)}{\sigma_0(x)},x_0\right)}\right]\begin{pmatrix}\gamma_1(t,x)\\ \gamma_2(t,x)\end{pmatrix}w_1(t)dt.$$

Through linearization,

$$U_{n1}(x)=\frac{1}{n}\sum_{i=1}^n[\phi_{11}(Y_i,X_i)+B_{11}(X_i,x_0,x)]\frac{1}{h^d}K\left(\frac{X_i-x_0}{h}\right)$$
$$-\frac{1}{n}\sum_{i=1}^n[\phi_{12}(Y_i,X_i)+B_{12}(X_i,x_0,x)]\frac{1}{h^d}K\left(\frac{X_i-x}{h}\right)+o_p\left((nh^d)^{-1/2}\right)\begin{pmatrix}1\\1\end{pmatrix}$$

if $w(m_0(x))\neq 0$, and

$$U_{n2}(x)=\frac{1}{n}\sum_{i=1}^n[\phi_{21}(Y_i,X_i)+B_{21}(X_i,x_0,x)]\frac{1}{h^d}K\left(\frac{X_i-x}{h}\right)$$
$$-\frac{1}{n}\sum_{i=1}^n[\phi_{22}(Y_i,X_i)+B_{22}(X_i,x_0,x)]\frac{1}{h^d}K\left(\frac{X_i-x_0}{h}\right)+o_p\left((nh^d)^{-1/2}\right)\begin{pmatrix}1\\1\end{pmatrix}$$

https://doi.org/10.1017/S0266466617000111 Published online by Cambridge University Press

if $w^c(m_0(x)) \neq 0$, where

$$\phi_{11}(Y_i, X_i) = \frac{P_0}{S(\varepsilon_0)p(x_0)} \int \left[1\{Y_i > t\} - 1\{Y_i > \varepsilon_0\}\frac{S^*(t, X_i)}{S^*(\varepsilon_0, X_i)}\right]\begin{pmatrix}\lambda_1(t,x)\\ \lambda_2(t,x)\end{pmatrix} w_1(t)dt$$

$$\phi_{12}(Y_i, X_i) = \frac{P_0}{S(\varepsilon_0)p(x)} \times \int \left[1\{Y_i > t\sigma_0(x) + m_0(x)\} - 1\{Y_i > \varepsilon_0\sigma_0(x) + m_0(x)\}\frac{S^*(t\sigma_0(x) + m_0(x), X_i)}{S^*(\varepsilon_0\sigma_0(x) + m_0(x), X_i)}\right] \times \begin{pmatrix}\lambda_1(t,x)\\ \lambda_2(t,x)\end{pmatrix} w_1(t)dt$$

$$\phi_{21}(Y_i, X_i) = \frac{P_0}{S\left(\frac{\varepsilon_0 - m_0(x)}{\sigma_0(x)}\right)p(x)} \int \left[1\{Y_i > t\} - 1\{Y_i > \varepsilon_0\}\frac{S^*(t, X_i)}{S^*(\varepsilon_0, X_i)}\right]\begin{pmatrix}\gamma_1(t,x)\\ \gamma_2(t,x)\end{pmatrix} w_1(t)dt$$

$$\phi_{22}(Y_i, X_i) = \frac{P_0}{S\left(\frac{\varepsilon_0 - m_0(x)}{\sigma_0(x)}\right)p(x_0)} \times \int \left[1\left\{Y_i > \frac{t - m_0(x)}{\sigma_0(x)}\right\} - 1\left\{Y_i > \frac{\varepsilon_0 - m_0(x)}{\sigma_0(x)}\right\}\frac{S^*\left(\frac{t-m_0(x)}{\sigma_0(x)}, X_i\right)}{S^*\left(\frac{\varepsilon_0 - m_0(x)}{\sigma_0(x)}, X_i\right)}\right] \times \begin{pmatrix}\gamma_1(t,x)\\ \gamma_2(t,x)\end{pmatrix} w_1(t)dt$$

$$B_{11}(X_i, x_0, x) = \frac{P_0 S^*(\varepsilon_0, X_i)}{S(\varepsilon_0)p(x_0)} \int \left[\frac{S^*(t, X_i)}{S^*(\varepsilon_0, X_i)} - \frac{S(t)}{S(\varepsilon_0)}\right]\begin{pmatrix}\lambda_1(t,x)\\ \lambda_2(t,x)\end{pmatrix} w_1(t)dt$$

$$B_{12}(X_i, x_0, x) = \frac{P_0 S^*(\varepsilon_0\sigma_0(x) + m_0(x), X_i)}{S(\varepsilon_0)p(x)} \int \left[\frac{S^*(t\sigma_0(x) + m_0(x), X_i)}{S^*(\varepsilon_0\sigma_0(x) + m_0(x), X_i)} - \frac{S(t)}{S(\varepsilon_0)}\right] \times \begin{pmatrix}\lambda_1(t,x)\\ \lambda_2(t,x)\end{pmatrix} w_1(t)dt$$

$$B_{21}(X_i, x_0, x) = \frac{P_0 S^*(\varepsilon_0, X_i)}{S\left(\frac{\varepsilon_0 - m_0(x)}{\sigma_0(x)}\right)p(x)} \int \left[\frac{S^*(t, X_i)}{S^*(\varepsilon_0, X_i)} - \frac{S(t)}{S(\varepsilon_0)}\right]\begin{pmatrix}\gamma_1(t,x)\\ \gamma_2(t,x)\end{pmatrix} w_1(t)dt$$

$$B_{22}(X_i, x_0, x) = \frac{P_0 S^*\left(\frac{\varepsilon_0 - m_0(x)}{\sigma_0(x)}, X_i\right)}{S\left(\frac{\varepsilon_0 - m_0(x)}{\sigma_0(x)}\right)p(x_0)} \int \left[\frac{S^*\left(\frac{t-m_0(x)}{\sigma_0(x)}, X_i\right)}{S^*\left(\frac{\varepsilon_0 - m_0(x)}{\sigma_0(x)}, X_i\right)} - \frac{S^*\left(\frac{t-m_0(x)}{\sigma_0(x)}, x_0\right)}{S^*\left(\frac{\varepsilon_0 - m_0(x)}{\sigma_0(x)}, x_0\right)}\right] \times \begin{pmatrix}\gamma_1(t,x)\\ \gamma_2(t,x)\end{pmatrix} w_1(t)dt.$$

Then, we have

$$\sqrt{nh^d}\left(\begin{pmatrix}\hat{m}(x) - m_0(x)\\ \hat{\sigma}(x) - \sigma_0(x)\end{pmatrix} - h^q b(x)\right) = \frac{1}{\sqrt{nh^d}}\sum_{i=1}^{n} V(x)^{-1}\left[\phi_1(X_i, Y_i)K\left(\frac{X_i - x_0}{h}\right) + \phi_2(X_i, Y_i)K\left(\frac{X_i - x}{h}\right)\right] + \begin{pmatrix}o_p(1)\\ o_p(1)\end{pmatrix}, \quad \textbf{(A.9)}$$

https://doi.org/10.1017/S0266466617000111 Published online by Cambridge University Press

where

$$\phi_1(X_i,Y_i)=\phi_{11}(X_i,Y_i)w(m_0(x))+\phi_{22}(X_i,Y_i)w^c(m_0(x)),$$
$$\phi_2(X_i,Y_i)=\phi_{12}(X_i,Y_i)w(m_0(x))+\phi_{21}(X_i,Y_i)w^c(m_0(x)) \text{ and}$$
$$b(x)=\int b_1(x,t)\begin{pmatrix}\lambda_1(t,x)\\ \lambda_2(t,x)\end{pmatrix}w_1(t)dt\,w(m_0(x))$$
$$+\int b_2(x,t)\begin{pmatrix}\gamma_1(t,x)\\ \gamma_2(t,x)\end{pmatrix}w_1(t)dt\,w^c(m_0(x))$$

with

$$b_1(x,t)=\frac{P_0}{S(\varepsilon_0)p(x_0)}\sum_{|j|=q}\left[\frac{\partial S^*(t,u)}{\partial u^j}|_{u=x_0}-\frac{S(t)}{S(\varepsilon_0)}\frac{\partial S^*(\varepsilon_0,u)}{\partial u^j}|_{u=x_0}\right]\int\frac{1}{j!}v^jk(v)dv$$
$$-\frac{P_0}{S(\varepsilon_0)p(x)}\sum_{|j|=q}\left[\frac{\partial S^*(t\sigma_0(x)+m_0(x),u)}{\partial u^j}|_{u=x}-\frac{S(t)}{S(\varepsilon_0)}\frac{\partial S^*(\varepsilon_0\sigma_0(x)+m_0(x),u)}{\partial u^j}|_{u=x}\right]$$
$$\times\int\frac{1}{j!}v^jk(v)dv$$
$$b_2(x,t)=\frac{P_0}{S\left(\frac{\varepsilon_0-m_0(x)}{\sigma_0(x)}\right)p(x)}\sum_{|j|=q}\left[\frac{\partial S^*(t,u)}{\partial u^j}|_{u=x}-\frac{S^*(t,x)}{S^*(\varepsilon_0,x)}\frac{\partial S^*(\varepsilon_0,u)}{\partial u^j}|_{u=x}\right]\int\frac{1}{j!}v^jk(v)dv$$
$$-\frac{P_0}{S\left(\frac{\varepsilon_0-m_0(x)}{\sigma_0(x)}\right)p(x_0)}\sum_{|j|=q}\left[\frac{\partial S^*\left(\frac{t-m_0(x)}{\sigma_0(x)},u\right)}{\partial u^j}|_{u=x_0}-\frac{S^*(t,x)}{S^*(\varepsilon_0,x)}\frac{\partial S^*\left(\frac{\varepsilon_0-m_0(x)}{\sigma_0(x)},u\right)}{\partial u^j}|_{u=x_0}\right]$$
$$\times\int\frac{1}{j!}v^jk(v)dv.$$

Therefore, the triangular-array central limit theorem gives

$$\sqrt{nh^d}\left(\begin{pmatrix}\hat m(x)-m_0(x)\\ \hat\sigma(x)-\sigma_0(x)\end{pmatrix}-h^qb(x)\right)\xrightarrow{d}N(0,\Gamma(x)),$$

where

$$\Gamma(x)=V(x)^{-1}(\Gamma_1(x)+\Gamma_2(x))V(x)^{-1}\int K^2(v)dv,$$
$$\Gamma_1(x)=\frac{S(0)p(x_0)}{P_0}\mathbb{E}[\phi_1(X_i,Y_i)\phi_1'(X_i,Y_i)|Y_i>0,X_i=x_0]$$
$$\Gamma_2(x)=\frac{S\left(-\frac{m_0(x)}{\sigma_0(x)}\right)p(x)}{P_0}\mathbb{E}[\phi_2(X_i,Y_i)\phi_2'(X_i,Y_i)|Y_i>0,X_i=x_0].$$ ■

https://doi.org/10.1017/S0266466617000111 Published online by Cambridge University Press