Trimmed and winsorized semiparametric estimator for left-truncated and right-censored regression models

Myoung-jae Lee · Maria Karlsson

Abstract For a linear regression model subject to left-truncation and right-censoring where the truncation and censoring points are known constants (or always observed if random), Karlsson and Laitila (Stat Probab Lett 78:2567–2571, 2008) proposed a semiparametric estimator which deals with left-truncation by trimming and right-censoring by 'winsorizing'. The estimator was motivated by a zero moment condition where a transformed error term appears with trimmed and winsorized tails. This paper takes the semiparametric estimator further by deriving the asymptotic distribution that was not shown in Karlsson and Laitila (Stat Probab Lett 78:2567–2571, 2008) and discusses its implementation aspects in practice, albeit brief.

Keywords Left truncation · Right censoring · LTRC · Semiparametrics · Trimmed mean · Winsorized mean

1 Introduction

Left-truncated and right-censored (LTRC) data occur in many situations due to, e.g., the sampling procedure, the study design and/or the measuring instrument. Left-truncation of a response variable in a regression model means that neither the response or the explanatory variables are observed if the value of the response is smaller than a truncation point t. Right-censoring of the response variable means that the observed value,

M. Lee
Department of Economics, Korea University, Seoul, South Korea

M. Karlsson (✉)
Department of Statistics, Umeå School of Business and Economics, Umeå University,
Umeå, Sweden
e-mail: maria.karlsson@stat.umu.se

Y, is the minimum of the (latent) value of the response, Y^*, and the censoring point c, i.e., whenever the value of the response is larger than c its true value is not observed but c is recorded instead.

One example of LTRC can be seen in an income survey where only individuals with income above a certain level get sampled (truncation) and then their income is top-coded (censoring). Another example is studies of damages sizes measured as claims paid by insurance companies, where only those damages larger than their deductibles are reported to the insurance companies (truncation) and where the is an upper limit of indemnification (censoring). LTRC also occurs when an individual is observed for an event A_1, but gets sampled only when another event A_0 holds. For instance, studying the duration of un-employment, suppose the sampling is done only from individuals that has been un-employed more than a specific amount of days, e.g., long enough to receive un-employment insurance contributions. Here the truncation point is a known constant, i.e., fixed. Depending on how the study ends, the censoring point can be fixed or random, and known or unknown. If the study ends on a fixed-calendar date, the censoring point is random, but known for everybody. If the study follows everybody until a fixed duration of un-employment (e.g., until the compensation days from un-employment insurance has run out) the censoring point is known and fixed. Suppose instead that, the sampling is done among the individuals (still) being un-employed at a fixed-calender date, then the truncation point is random and unknown.

In this paper we consider linear regression models where the truncation and censoring points are both known constants or always observed (i.e., known) if they are random.

Ignoring the truncation and/or censoring, e.g., by using the least squares (LS) approach to estimate a regression model, when the response variable is in fact LTRC, can yield very misleading estimates. Moreover, maximum likelihood based estimators are rather sensitive to distributional misspecification, i.e., the specification of the distribution of the error term, when data is LTRC. The goal of this paper is to estimate the regression coefficients without specifying the distribution of the error term in the regression model under LTRC. This is accomplished by using the semiparametric estimator proposed by Karlsson and Laitila (2008). This paper takes the estimator further by deriving the asymptotic distribution that was not shown in their paper.

There are many estimators for LTRC data (see e.g., Shen 2009, 2012, and the references therein) but they are mostly for two-sample location differences. For regression parameters, the estimators are rather complicated (Lai and Ying 1991, 1994; Gross and Lai 1996) involving estimation of the Y^* distribution. Moreover, most of them require random LTRC data. The estimator of Karlsson and Laitila (2008) is a simple semiparametric estimator for LTRC when the truncation and censoring points are fixed (or always observed if random). While the simplicity is its advantage, the disadvantage is inapplicability if the censoring point is unknown when greater than Y^*.

Section 2 explains the estimator, and Sect. 3 examines its asymptotic distribution and discusses its implementation. The paper ends with a section with conclusions.

2 The estimator and its main idea

For each individual i in the population,

$$Y_i^* = X_i'\beta + U_i, \text{ independent and identically distributed } (iid), \tag{1}$$

where $X_i = (1, \tilde{X}_i')'$, $\tilde{X}_i \equiv (X_{i2}, ..., X_{ik})'$ is a $(k - 1) \times 1$ regressor vector, $\beta \equiv (\beta_1, \tilde{\beta}')'$ is a $k \times 1$ coefficient vector with $\tilde{\beta}$ being the $(k - 1) \times 1$ slope coefficients, and U_i is an error term independent of X_i ("$U \amalg X$") with its density function (pdf) and distribution function (cdf) denoted by f_U and F_U.

Now, consider a LTRC linear regression model with, for simplicity, left-truncation at $t = 0$ and right-censoring at a known constant $c > 0$ for all individuals. Let $Y_i \equiv \min(Y_i^*, c)$ and sample N units (X_i, Y_i) from the sub-population (or stratum) where $Y_i^* > t = 0$. Henceforth, if possible, the subscripts i of Y_i^*, Y_i, X_i and U_i are dropped for convenience.

It is easy to allow for a known non-zero truncation constant by subtracting it from the response, the intercept and from the censoring point. Similarly, we can, as long as they are observed (i.e., known), allow left-truncation at t_i and right-censoring at c_i, i.e., truncation and censoring points varying across individuals. The main finding of this paper still holds conditionally on (t_i, c_i) so long as $Y_i^* \amalg (t_i, c_i)$ given X_i.

Without truncation/censoring, an estimate of β in (1) can be calculated by applying the LS estimator under the moment condition of $E(XU) = 0$. With LTRC data, Karlsson and Laitila (2008) proposed an estimator motivated by a zero moment condition where a transformed error trimmed at the lower end and winsorized at the upper tail appears. Henceforth, the estimator of Karlsson and Laitila (2008) is called the "trimmed and winsorized estimator" (TWE). The moment condition of TWE is a combination of the two moment conditions under which, the two estimators suggested in Karlsson (2006) for left truncation and left censoring respectively, are derived. They, in turn, are generalizations of the "quadratic mode regression estimator" (QME) by Lee (1993) and the "winsorized mean estimator" (WME) by Lee (1992). In order to explain the TWE, we first take a closer look at QME and WME and their properties.

2.1 The QME and the WME

If data is left-truncated at 0 (but not right-censored), i.e., only observed if $0 < Y^*$ or equivalently only observed if $-X'\beta < U$, the QME can be used. The QME is derived from the moment condition

$$E\{1[X'\beta > w]X \cdot 1[-w < U < w]U\} = 0 \tag{2}$$

where $1[A]$ denotes an indicator function taking value 1 if condition A is valid and 0 otherwise and $w > 0$ is a "trimming" constant. This means that the "trimmed error" $1[-w < U < w]U$ is orthogonal to X. The factor $1[X'\beta > w]$ in (2) is for $-w < U$ to hold when $-X'\beta < U$. QME is consistent under semiparametric assumptions—e.g., *partial symmetry* of $U|X$ up to $\pm w$.

In contrast to QME, when there is only left censoring at 0, the WME used another modified moment condition

$$E\{1[X'\beta > w]X \cdot (-w1[U \leq -w] + 1[-w < U < w]U + w1[w \leq U])\} = 0; \quad (3)$$

instead of trimming U when $|U| \geq w$, U gets replaced by $\pm w$. The "winsorized error" $-w1[U \leq -w] + 1[-w < U < w]U + w1[w \leq U]$ is orthogonal to X under the partial symmetry of U up to $\pm w$ and $P(U \leq -w) = P(U \geq w)$.

Whereas the LS estimator has $E(XX')$ as its second order matrix, QME has $(-1$ times)

$$H_{QME} \equiv E\left[XX'\left\{\left(\text{``area under density } \frac{f_U}{1 - F_U(-X'\beta)} \text{ between } \pm w"\right)\right.\right.$$
$$\left.\left. - 2w\frac{f_U(w)}{1 - F_U(-X'\beta)}\right\}\right], \quad (4)$$

where $1 - F_U(-X'\beta)$ is the truncation normalizing factor (see the left panel of Fig. 1). For H_{QME} to be positive definite, f_U should be strictly unimodal at 0, explaining the word "mode" in QME.

The asymptotic variance of WME has as its second order matrix

$$H_{WME} \equiv E\left[XX'\left\{(\text{``area under } f_U \text{ between } \pm w")\right\}\right], \quad (5)$$

(see right panel of Fig. 1) which differs somewhat from (4).

Although we used f_U for QME and WME, QME and WME allow heteroskedasticity of unknown form, in which case $f_{U|X}$ replaces f_U. It is the local symmetry assumption up to $\pm w$ that assures $E(1[|U| < w]U) = 0$ while allowing for heteroskedasticity of unknown form.
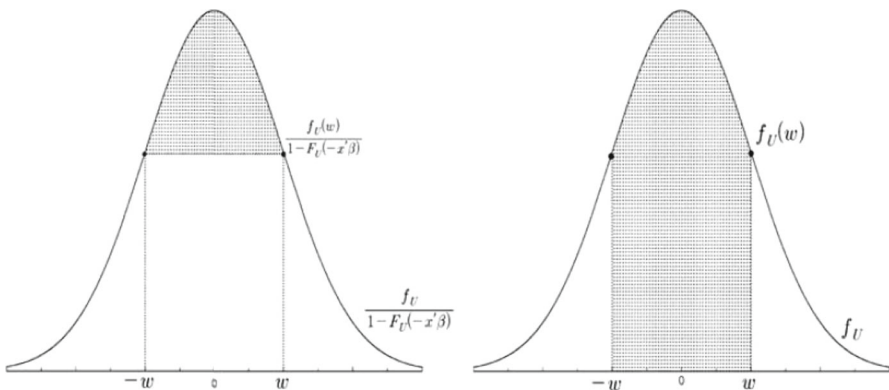

Fig. 1 Density areas for QME (*left*) and WME (*right*) hessians

2.2 The trimmed and winsorized estimator

In QME and WME where only one tail of U is subject to truncation/censoring, symmetric trimming/winsorizing is done artificially at the other tail of U. But symmetric trimming/winsorizing is not feasible under LTRC as both tails of U are subject to truncation/censoring. The left-truncation and right-censoring points give the interval $(-X'\beta, c - X'\beta]$ for U. In this case, a transformed error trimmed at the lower end and winsorized at the upper end can be used; for some chosen trimming constants $w_l, w_u > 0$. The trimming and winsorizing interval $(-w_l, w_u)$ should be placed within the range of U, such that,

$$-X'\beta < -w_l < U < w_u \le c - X'\beta.$$

For this, it is necessary to have

$$w_l < X'\beta \le c - w_u \tag{6}$$

for which it is in turn necessary to have $w_l + w_u < c$. Added to this condition is '$0 < w_u < w_l$' to help the expected value of the trimmed and winsorized error

$$1[-w_l < U < w_u]U + w_u 1[w_u \le U]. \tag{7}$$

to be equal to zero conditional on X. Because of the restrictions $w_l + w_u < c$ and $0 < w_u < w_l$, we can choose w_u from $(0, c/2)$ first, and then w_l over $(w_u, c - w_u)$.

In view of (6), one may try to use

$$1[-X'\beta < U < c - X'\beta]U + (c - X'\beta)1[c - X'\beta \le U]$$

instead of (7), but this fails because the intercept that makes its expected value, conditional on X, equal to zero depends on X. Using the *constant* interval $(-w_l, w_u)$ falling within the X-varying interval $(-X'\beta, c - X'\beta)$ is the key idea.

The restriction on $X'\beta$ in (6) combined with the trimmed and winsorized error in (7), yields the key unconditional moment condition under LTRC; $E\{m(X, Y^*; \beta)\} = 0$ (or just $E\{m(\beta)\} = 0$ if (X, Y^*) is omitted), where

$$m(\beta) \equiv 1\left[w_l < X'\beta < c - w_u\right] X \cdot \left(1\left[-w_l < Y^* - X'\beta < w_u\right]\left(Y^* - X'\beta\right)\right.$$
$$\left. + w_u 1\left[w_u \le Y^* - X'\beta\right]\right).$$

This is thus a combination of the corresponding moment conditions of QME and WME, (2) and (3) respectively.

Recall from (1) that $Y^* = \beta_1 + \tilde{X}'\tilde{\beta} + U$ and consider

$$E\left(1\left[-w_l < Y^* - \beta_1^* - \tilde{X}'\tilde{\beta} < w_u\right]\left(Y^* - \beta_1^* - \tilde{X}'\tilde{\beta}\right) + w_u 1\left[w_u \le Y^* - \beta_1^* - \tilde{X}'\tilde{\beta}\right] | X\right)$$
$$= E\left(1\left[-w_l < U - \Delta < w_u\right](U - \Delta) + w_u 1\left[w_u \le U - \Delta\right]\right)$$

with $\Delta \equiv \beta_1^* - \beta_1$, where β_1^* is for the intercept shift (i.e., β_1^* that does not necessarily equal β_1), which is assumed to exist (Assumption A1, Karlsson and Laitila 2008) for $E\{m(\beta_1^*, \tilde{\beta})\} = 0$ to hold. If U is not independent of X, then β_1^* may have to change depending on X because w_l and w_u are fixed. This would then hamper the identification of $\tilde{\beta}$, which is why $U \amalg X$ is invoked in this paper although it was not assumed in Karlsson and Laitila (2008). '$U \amalg X$' can be relaxed if the form of heteroskedasticity of $U|X$ is known. For instance, if $U = \exp(X_k)V$ where X_k is the k th element of X and V is an error term with $V \amalg X$, we can replace w_l and w_u with $w_l \exp(X_k)$ and $w_u \exp(X_k)$, respectively. Then the above moment condition becomes a moment condition involving V, w_l and w_u. But since a known form of heteroskedasticity is rare, we will proceed with $U \amalg X$.

The TWE is defined using a minimand in Karlsson and Laitila (2008),

$$\frac{1}{N} \sum_i \left\{ 1\left[X_i'b \leq w_l \right] g\left(Y_i - w_l\right) + 1\left[w_l < X_i'b < c - w_u \right] g\left(Y_i - X_i'b\right) \right.$$
$$\left. + 1\left[c - w_u \leq X_i'b \right] g\left(Y_i - c + w_u\right) \right\}$$

where

$$g(r) \equiv 1[r \leq -w_l]\frac{w_l^2}{2} + 1[-w_l < r < w_u]\frac{r^2}{2} + 1[w_u \leq r]\left(w_u r - \frac{w_u}{2}\right).$$

This minimand has three components depending on where $X'b$ locates relative to w_l and $c - w_u$, and each component can take three different forms as can be seen in $g(\cdot)$. The first term $w_l^2/2$ of $g(\cdot)$ appears for the QME maximand to make use of the truncated part, the second term $r^2/2$ is the usual squared residual in the LS estimator, and the last term $w_u r - w_u/2$ appears for the WME minimand to make use of the censored part.

Karlsson and Laitila (2008) obtained this minimand by integrating the population moment condition back. Their proof uses the proofs in Newey (2001), who showed that integrating a population moment condition back, a minimand is obtained with a unique minimum, and that the estimator minimizing the sample version is $\sqrt{N}$-consistent satisfying the sample moment condition $N^{-1/2} \sum_i m(X_i, Y_i; b_N) = o_p(1)$. Karlsson and Laitila (2008) defined $\beta^* \equiv (\beta_1^*, \tilde{\beta}')'$ and showed that b_N converged in probability to β^*, i.e. $b_N \to^p \beta^*$.

3 Asymptotic distribution

In addition to the assumptions in Karlsson and Laitila (2008) for the consistency of TWE, with $|X|$ denoting the Euclidean norm of X, suppose the following assumptions hold:

Assumption U $U \amalg X$, $E(U) < \infty$, and f_U is continuous and satisfies, for $\Delta \equiv \beta_1^* - \beta_1$, $P(\Delta - w_l < U < \Delta + w_u) > w_l f_U(\Delta - w_l)/\{1 - F_U(-X'\beta)\}$.

Assumption X $E\{XX'1[w_l < X'\beta^* < c - w_u]\}$ is positive definite, $E|X|^2 < \infty$ and for some constants $v_0, v_1, v_2 > 0$,

(i): $E\{1[|w_l - X'\beta^*| < |X||b - \beta^*|] \cdot |X|\} \leq v_1|b - \beta^*|$ and
(ii): $E\{1[|c - w_u - X'\beta^*| < |X||b - \beta^*|] \cdot |X|\} \leq v_2|b - \beta^*|$

for all $|b - \beta^*| < v_0$.

Assumption X is for smoothness of indicator functions of b such as $E(1[X'b \geq w_l])$. Almost the same assumptions appeared in Powell (1984, p. 310), Newey and Powell (1990, p. 304), and, Lee (1993, p. 4) in the proofs of asymptotic normality of the estimators of censored or truncated regression models suggested in those papers. Assumption X can be replaced with another assumption that makes $X'\beta$ smooth, e.g., $F_{k|-k}(x_k|x_{-k})$ having a continuously differentiable derivative $f_{k|-k}(x_k|x_{-k})$ for x_k that is uniformly bounded over $x = (x'_{-k}, x_k)'$, where $F_{k|-k}$ is the distribution function of X_k given the other elements X_{-k} of X. The motivation for making Assumption U will be explained shortly.

Theorem AD *Under Assumptions U and X, with '$\rightsquigarrow$' denoting convergence in law,*

$$\sqrt{N}(b_N - \beta^*) \rightsquigarrow N\left[0, H_{TWE}^{-1}E\{m(\beta^*)m(\beta^*)'\}H_{TWE}^{-1}\right]$$

where

$$H_{TWE} \equiv E\left[1\left[w_l < X'\beta^* < c - w_u\right]XX'\right.$$
$$\left. \times \left\{1[\Delta - w_l < U < \Delta + w_u] - w_l \frac{f_U(\Delta - w_l)}{1 - F_U(-X'\beta)}\right\}\right].$$

Theorem AD is proven in the Appendix. Roughly speaking, H_{TWE} falls between H_{QME} and H_{WME}, see (4) and (5), because the X-conditional mean of $\{\cdot\}$ in H_{TWE} is "area under the density $\frac{f_U}{1-F_U(-X'\beta)}$ between $\Delta - w_l$ and $\Delta + w_u$" minus $w_l \frac{f_U(\Delta - w_l)}{1-F_U(-X'\beta)}$. If $\beta_1^* = \beta_1 \Longleftrightarrow \Delta = 0$, then the strict unimodality of U at 0 is enough to make this display positive. But β_1^* might differ much from β_1, depending on f_U and $(-w_l, w_u)$ so that the interval $(-w_l, w_u)$ may fall under a downward sloping part of f_U to make the last display non-positive. This is why Assumption U was imposed.

3.1 Estimation and implementation aspects of TWE

Let $\hat{U}_i \equiv Y_i - X'_i b_N \rightarrow^p Y_i - X'_i\beta + \beta_1 - \beta_1^* = U_i - \Delta$. Then $\hat{H}_{TWE}^{-1}\hat{G}\hat{H}_{TWE}^{-1}$ can be used to estimate the asymptotic variance, where

$$\hat{G} \equiv \frac{1}{N}\sum_i 1\left[w_l < X'_i b_N < c - w_u\right]X_iX'_i$$
$$\times \left\{1\left[-w_l < \hat{U}_i < w_u\right]\hat{U}_i + w_u 1\left[w_u \leq \hat{U}_i\right]\right\}^2$$

and

$$\hat{H}_{TWE} \equiv \frac{1}{N} \sum_i 1 \left[w_l < X_i' b_N < c - w_u \right] X_i X_i'$$

$$\times \left\{ 1 \left[-w_l < \hat{U}_i < w_u \right] - w_l \frac{1}{h} 1 \left[-w_l \le \hat{U}_i \le -w_l + h \right] \right\}$$

for a bandwidth $h \to^+ 0$ as $N \to \infty$. The expression'$h^{-1}1[-w_l \le \hat{U}_i \le -w_l + h]$' in $\hat{H}_{TWE}$ is to estimate $f_U(\Delta - w_l)/\{1 - F_U(-X'\beta)\}$ nonparametrically. As $\hat{H}_{TWE}$ is bandwidth-dependent, practitioners would prefer using nonparametric bootstrap instead. This is one reason why we omit proving $\hat{H}_{TWE} \to^p H_{TWE}$ and $\hat{G} \to^p E\{m(\beta^*)m(\beta^*)'\}$ that need more assumptions.

The above derivation suggests an iteration algorithm: with $U_i(b_0) \equiv Y_i - X_i' b_0$,

$$b_1 = b_0 + \hat{H}_{TWE}^{-1} \frac{1}{N} \sum_i 1 \left[w_l < X_i' b_0 < c - w_u \right] X_i$$

$$\times \{ 1 \left[-w_l < U_i(b_0) < w_u \right] U_i(b_0) + w_u 1 \left[w_u \le U_i(b_0) \right] \}$$

which is to be iterated until convergence; b_N in $\hat{H}_{TWE}$ should be replaced by b_0 as well. In the iteration, $\hat{H}_{TWE}$ may pose a trouble as it may not be positive definite, in which case $w_l h^{-1} 1[-w_l \le U_i(b_0) \le -w_l + h]$ in $\hat{H}_{TWE}$ may be dropped. If the problem persists, $1[w_l < X_i' b_0 < c - w_u]$ may be dropped as well.

The finite sample bias could be large for the intercept estimator since it plays the role of satisfying $E\{m(\beta_1^*, \tilde{\beta})\} = 0$, and the intercept bias magnitude would depend on $(-w_l, w_u)$. Instead of modifying the intercept, TWE might try to adjust the scale of U, but this cannot be done easily because the scale of Y is fixed and there is no scale parameter to estimate. What could still happen though is scaling all slope estimates up/down to alter the scale of the residual $Y - X'b_N$—this would be how the intercept bias affects the slope estimates. In this case, all slope estimates will be biased up/down by the same positive factor. This, however, would not bias signs of slope estimates, nor their ratios.

4 Conclusions

LTRC data is a common problem that face researchers and it is important to use an estimator that takes the truncation and censoring into account in order not to draw incorrect conclusions due to misleading results. In this paper we derive the asymptotic distribution of the semiparametric estimator TWE, suggested by Karlsson and Laitila (2008). $\sqrt{N}$-consistency for the slope estimates were already shown in Karlsson and Laitila (2008). The results in this paper together with the results in Karlsson and Laitila (2008) suggest that the TWE is a suitable estimator when data is LTRC. However, there are also situations in practice, where it has drawbacks as discussed in the previous section.

Appendix

We will first show that $E\{m(b)\}$ is continuously differentiable at $b = \beta^*$ with derivative matrix H_{TWE}. Rewrite $E\{m(b)\} - E\{m(\beta^*)\}$ as

$$
\begin{aligned}
& E\left\{1\left[w_l < X'b < c - w_u\right] X \left\{1\left[-w_l < Y - X'b < w_u\right](Y - X'b)\right.\right. \\
& \quad \left.\left. + w_u 1\left[w_u \leq Y - X'b\right]\right\}\right\} \\
& - E\left\{1\left[w_l < X'b < c - w_u\right] X \left\{1\left[-w_l < Y - X'\beta^* < w_u\right](Y - X'\beta^*)\right.\right. \\
& \quad \left.\left. + w_u 1\left[w_u \leq Y - X'\beta^*\right]\right\}\right\} \\
& + E\left\{1\left[w_l < X'b < c - w_u\right] X \left\{1\left[-w_l < Y - X'\beta^* < w_u\right](Y - X'\beta^*)\right.\right. \\
& \quad \left.\left. + w_u 1\left[w_u \leq Y - X'\beta^*\right]\right\}\right\} \\
& - E\left\{1\left[w_l < X'\beta^* < c - w_u\right] X \left\{1\left[-w_l < Y - X'\beta^* < w_u\right](Y - X'\beta^*)\right.\right. \\
& \quad \left.\left. + w_u 1\left[w_u \leq Y - X'\beta^*\right]\right\}\right\}.
\end{aligned}
$$

The last two terms are 0 because $U \amalg X$ and the transformed error with $Y - X'\beta^*$ has mean zero. The first two terms share $1[w_l < X'b < c - w_u]X$, and thus we rewrite them as

$$
\begin{aligned}
& E\left(1\left[w_l < X'\beta^* < c - w_u\right] X \left[\left\{1\left[-w_l < Y - X'b < w_u\right](Y - X'b)\right.\right.\right. \\
& \quad \left.\left. + w_u 1\left[w_u \leq Y - X'b\right]\right\} - \left\{1\left[-w_l < Y - X'\beta^* < w_u\right](Y - X'\beta^*)\right.\right. \\
& \quad \left.\left.\left. + w_u 1\left[w_u \leq Y - X'\beta^*\right]\right\}\right]\right) + E\left(\left\{1[w_l < X'b < c - w_u]\right.\right. \\
& \quad \left. - 1[w_l < X'\beta^* < c - w_u]\right\} X \left[\left\{1[-w_l < Y - X'b < w_u](Y - X'b)\right.\right. \\
& \quad \left.\left. + w_u 1\left[w_u \leq Y - X'b\right]\right\} - \left\{1\left[-w_l < Y - X'\beta^* < w_u\right](Y - X'\beta^*)\right.\right. \\
& \quad \left.\left.\left. + w_u 1\left[w_u \leq Y - X'\beta^*\right]\right\}\right]\right).
\end{aligned} \tag{8}
$$

The following shows that the first two terms give H_{TWE} and that the remainder is $o(|b - \beta^*|)$.

With $f_{U,X'\beta}(\cdot) \equiv f_U(\cdot)/\{1 - F_U(-X'\beta)\}$, observe, for the first two terms of (8),

$$
\begin{aligned}
& E\{1[-w_l < Y - X'b < w_u](Y - X'b) + w_u 1[w_u \leq Y - X'b] \,|X\} \\
& = \int_{X'b-X'\beta-w_l}^{X'b-X'\beta+w_u} (u + X'\beta - X'b) f_{U,X'\beta}(u) du + w_u \int_{X'b-X'\beta+w_u}^{\infty} f_{U,X'\beta}(u) du.
\end{aligned}
$$

Differentiate this for b to get a continuous derivative vector

$$
\begin{aligned}
& X'\{w_l f_{U,X'\beta}(X'b - X'\beta - w_l) + w_u f_{U,X'\beta}(X'b - X'\beta + w_u) \\
& \quad - \int_{X'b-X'\beta-w_l}^{X'b-X'\beta+w_u} f_{U,X'\beta}(u) du - w_u f_{U,X'\beta}(X'b - X'\beta + w_u)\} \\
& = -X'\{\int_{X'b-X'\beta-w_l}^{X'b-X'\beta+w_u} f_{U,X'\beta}(u) du - w_l f_{U,X'\beta}(X'b - X'\beta - w_l)\}.
\end{aligned}
$$

Set $b = \beta^*$, attach $1[w_l < X'\beta^* < c - w_u]X$ and then take $E(\cdot)$ to get $-H_{TWE}$.

Since the transformed error is bounded by $-w_l$ and w_l, the last two terms of (8) is less than a constant times

$$
\begin{aligned}
E&\{|1[w_l < X'b < c - w_u] - 1[w_l < X'\beta^* < c - w_u]| \cdot |X|\} \\
&\leq E[\{1[X'b < w_l < X'\beta^*] + 1[1[X'\beta^* < w_l < X'b] \\
&\quad + 1[X'b < c - w_u < X'\beta^*] + 1[X'\beta^* < c - w_u < X'b]\} \cdot |X|] \\
&= E[\{1[X'b - X'\beta^* < w_l - X'\beta^* < 0] \\
&\quad + 1[1[0 < w_l - X'\beta^* < X'b - X'\beta^*] \\
&\quad + 1[X'b - X'\beta^* < c - w_u - X'\beta^* < 0] \\
&\quad + 1[0 < c - w_u - X'\beta^* < X'b - X'\beta^*]\} \cdot |X|] \\
&\leq E[\{1[0 < |w_l - X'\beta^*| < |X||b - \beta^*|] \\
&\quad + 1[0 < |c - w_u - X'\beta^*| < |X||b - \beta^*|]\} \cdot |X|] \\
&\leq E[\{1[|w_l - X'\beta^*| < |X||b - \beta^*|] \\
&\quad + 1[|c - w_u - X'\beta^*| < |X||b - \beta^*|]\}|X|] \\
&\leq (\nu_1 + \nu_2)|b - \beta^*|.
\end{aligned}
$$

Since the transformed error difference is $O(|b - \beta^*|)$, this shows that the last two terms of (8) are $O(|b - \beta^*|^2)$, and thus $o(|b - \beta^*|)$.

Consider the stochastic equicontinuity of the empirical process

$$
G_N(b) \equiv \frac{1}{\sqrt{N}} \sum_i [m(X_i, Y_i, b) - E\{m(X_i, Y_i, b)\}].
$$

Since $m(b)$ consists of indicator and polynomial functions, $\{m(b), b \in B\}$ with B being a compact parameter space is a Donsker class with a square-integrable envelope under $E|X|^2$; see, e.g., Van der Vaart (1998). From the stochastic equicontinuity, $o_p(1) = G_N(b_N) - G_N(\beta^*)$ holds. Substitute $N^{-1/2} \sum_i m(X_i, Y_i, b_N) = o_p(1)$ into this to obtain

$$
\begin{aligned}
o_p(1) &= -\frac{1}{\sqrt{N}} \sum_i m(X_i, Y_i, \beta^*) - \sqrt{N}[E\{m(b_N)\} - E\{m(\beta^*)\}] \\
&= -\frac{1}{\sqrt{N}} \sum_i m(X_i, Y_i, \beta^*) - \frac{\partial E\{m(b_N^*)\}}{\partial b} \sqrt{N}(b_N - \beta^*) \\
&\quad \text{(for some } b_N^* \to^p \beta^*) \\
&\implies \sqrt{N}(b_N - \beta^*) = -[\frac{\partial E\{m(\beta^*)\}}{\partial b}]^{-1} \frac{1}{\sqrt{N}} \sum_i m(X_i, Y_i, \beta^*) + o_p(1).
\end{aligned}
$$

From this, Theorem AD follows.

References

Gross ST, Lai TL (1996) Nonparametric estimation and regression analysis with left-truncated and right-censored data. J Am Stat Assoc 91:1166–1180

Karlsson M (2006) Estimators of regression parameters for truncated and censored data. Metrika 63:329–341

Karlsson M, Laitila T (2008) A semiparametric regression estimator under left truncation and right censoring. Stat Probab Lett 78:2567–2571

Lai T, Ying Z (1991) Rank regression methods for left-truncated and right-censored data. Ann Stat 19:531–556

Lai TL, Ying Z (1994) A missing information principle and M-estimators in regression analysis with censored and truncated data. Ann Stat 22:1222–1255

Lee MJ (1992) Winsorized mean estimator for censored regression. Econ Theory 8:368–382

Lee MJ (1993) Quadratic mode regression. J Econ 57:1–19

Newey WK (2001) Conditional moment restrictions in censored and truncated regression models. Econ Theory 17:863–888

Newey WK, Powell JL (1990) Efficient estimation of linear and type I censored regression models under conditional quantile restrictions. Econ Theory 6:295–317

Powell J (1984) Least absolute deviations estimation for the censored regression model. J Econ 25:303–325

Shen P (2009) A class of rank-based test for left-truncated and right-censored data. Ann Inst Stat Math 61:461–476

Shen P (2012) Median regression model with left truncated and right censored data. J Stat Plan Inference 142:1757–1766

Van der Vaart AW (1998) Asymptotic statistics. Cambridge University Press, Cambridge