Econometric Theory, **17**, 2001, 863–888. Printed in the United States of America.

CONDITIONAL MOMENT RESTRICTIONS IN CENSORED AND TRUNCATED REGRESSION MODELS

WHITNEY K. NEWEY
Massachusetts Institute of Technology

Censored and truncated regression models with unknown distribution are important in econometrics. This paper characterizes the class of all conditional moment restrictions that lead to $\sqrt{n}$-consistent estimators for these models. The semiparametric efficiency bound for each conditional moment restriction is derived. In the case of a nonzero bound it is shown how an estimator can be constructed and that an appropriately weighted version can attain the efficiency bound. These estimators also work when the disturbance is independent of the regressors. The paper discusses combining conditional moment restrictions for more efficient estimation in this case.

1. INTRODUCTION

Censored and truncated regression models are important for econometric data with a limited dependent variable. Unlike regression models without censoring or truncation, consistency of maximum likelihood estimators depends on the distributional specification. This property has motivated a search for estimators that are robust to distributional assumptions. This work includes Powell (1984, 1986a, 1986b), Newey (1987, 1989), Lee (1992, 1993), Honore and Powell (1994), and others.

In this paper we characterize the class of all conditional moment restrictions that lead to $\sqrt{n}$-consistent estimators for censored and truncated regression. We derive the semiparametric efficiency bound for each conditional moment restriction and show when it is nonzero. For the nonzero cases we describe how an estimator can be constructed and show that an appropriately weighted version can attain the semiparametric bound.

Because independence of disturbance and regressors will imply any conditional moment restriction, all the estimators will work in the independence case. Here we discuss how to combine several moment restrictions in both censored and truncated regression models.

In relation to previous work, the semiparametric efficiency bounds for the censored case generalize results of Newey and Powell (1990) and for the trun-

The NSF provided financial support. D. Andrews, J. Powell, T. Rothenberg, and two referees provided helpful comments. Address correspondence to: Whitney K. Newey, Department of Economics, Massachusetts Institute of Technology, E52-262D, Cambridge, MA 02139, USA.

© 2001 Cambridge University Press 0266-4666/01 $9.50

https://doi.org/10.1017/S0266466601175018 Published online by Cambridge University Press

cated case are new. Also, the censored regression estimators given here are based on a conditional moment restriction described by Newey (1989) that generalizes the moment restriction of Powell (1986a). Lee (1992) considers construction of $\sqrt{n}$-consistent estimators from a special case of these moment conditions. The truncated regression moment restriction is similar to that of Newey (1987) and generalizes Lee (1993). Here we dispense with the symmetry assumption imposed by Lee (1992, 1993). Also, we generalize previous results by characterizing the entire class of useful moment conditions. This leads to estimators that have improved properties over those previously proposed, including asymptotic efficiency and ease of asymptotic variance estimation.

2. THE MODEL

It is convenient to describe the class of models we consider in terms of a latent regression. Let $m(\varepsilon)$ be some scalar function. Then a latent regression equation with a conditional moment restriction can be described as

$$y^* = X'\beta_0 + \varepsilon, \qquad E^*[m(\varepsilon)|X] = 0, \tag{1}$$

where X and β are $K \times 1$ vectors of regressors and parameters and $E^*[\cdot]$ denotes the expectation in the latent regression. Each such condition corresponds to the location restriction $\mu(X) = 0$ for $\mu(X)$ solving $E^*[m(\varepsilon - \mu(X))|X] = 0$. For example, if $m(\varepsilon) = \varepsilon$ then ε has conditional mean zero, whereas if $m(\varepsilon) = 1(\varepsilon > 0) - 1(\varepsilon < 0)$ then ε has conditional median zero. Other specifications of $m(\varepsilon)$ correspond to other location restrictions, some of which are less familiar than the median and mean.

Censored and truncated regression models are ones where (y^*, X) is only partially observed. For censored regression,

$$y = \max\{0, y^*\} \quad \text{(Censored regression)}. \tag{2}$$

For truncated regression we have

$$(y^*, X) \text{ only observed if } y^* > 0 \quad \text{(Truncated regression)}. \tag{3}$$

These models are familiar in econometrics, and we focus on them to keep the exposition relatively simple. Our results could be extended to other models, including censored regression where the censoring point varies with X or censoring occurs above and also below.

In the latent model, where (y^*, X) is always observed, it is well understood how to use a conditional moment restriction to estimate β_0. Equation (1) implies that for any vector of functions $A(X)$ the unconditional moment restriction $E^*[A(X)m(y^* - X'\beta)] = 0$ will be satisfied at $\beta = \beta_0$ (assuming expectations exist). This moment condition could be used to form a generalized method of moments (GMM) estimator in the usual way. However, consistency

https://doi.org/10.1017/S0266466601175018 Published online by Cambridge University Press

of a GMM estimator requires that β_0 be the unique solution to the moment equation, which is difficult to show. This identification problem motivates obtaining $\hat{\beta}$ from minimizing a corresponding objective function that "integrates back from" $m(\varepsilon)$. It is often easier to show that such an objective function has a unique minimum. To this end, let $q(\varepsilon) = \int_0^{\varepsilon} m(u)\,du + C$, where C can be any constant. Also, let $w(X) \geq 0$ be a weight function that will be important for the efficiency discussion that follows. Then the moment restriction $E^*[w(X)Xm(\varepsilon)] = 0$ will be the first-order condition for $E^*[w(X)q(y^* - X'\beta)]$ to have an extremum (min or max) at β_0. Assume that the sign of $m(\varepsilon)$ is chosen so that $d(X) = \partial E^*[m(\varepsilon + \alpha)|X]/\partial\alpha|_{\alpha=0} \geq 0$. Then $E^*[w(X)d(X)XX']$ will be positive semidefinite, the necessary second-order condition for $E^*[w(X)q(y^* - X'\beta)]$ to have a minimum at β_0. Then $E^*[w(X)q(y^* - X'\beta)]$ becomes a function whose minimization corresponds to the moment restriction $E^*[w(X)Xm(y^* - X'\beta)] = 0$. The sample analog to the minimizer of $E^*[w(X)q(y^* - X'\beta)]$ is

$$\hat{\beta} = \operatorname*{argmin}_{\beta} \sum_{i=1}^{n} w(X_i)q(y_i^* - X_i'\beta). \tag{4}$$

The identification condition for consistency of this estimator is that $E^*[w(X)q(y^* - X'\beta)]$ has a unique minimum at β_0, which is easier to show than that $E^*[w(X)Xm(y^* - X'\beta)] = 0$ has a unique solution.

It turns out that in censored and truncated models an analogous approach works for some conditional moment restrictions and that for the rest no $\sqrt{n}$-consistent estimator exists. The nonexistence result will follow from the form of the semiparametric efficiency bound for this model. This bound is the infimum of the information bounds for β for regular parametric submodels (e.g., for the definition of *regular*, see Newey, 1990). It can often be computed by a projection. Define the tangent set $\mathcal{T}$ to be the mean-square closure of the set of all scores for parameters of the distribution of (ε, X) in parametric submodels passing through the truth and satisfying (1) and let S_β denote the score for β. Define the efficient score $\mathcal{S}$ to be the residual from the mean-square projection of S_β on $\mathcal{T}$, assuming $\mathcal{T}$ is linear. If $E[\mathcal{S}\mathcal{S}']$ is singular then no $\sqrt{n}$-consistent, regular estimator exists, whereas if $E[\mathcal{S}\mathcal{S}']$ is nonsingular then its inverse provides a bound on the asymptotic variance of regular $\sqrt{n}$-consistent estimators. Here we find that $\mathcal{S}$ is zero except for certain cases where the moment condition leads to a $\sqrt{n}$-consistent estimator analogous to that in (4). We also find that the asymptotic variance of this estimator is equal to the semiparametric bound when $w(X)$ is chosen to have a certain form.

3. CENSORED REGRESSION MODELS

For censored regression any moment condition where $m(\varepsilon)$ is constant for all ε small enough leads to $\sqrt{n}$-consistent estimation. The fact that $m(\varepsilon)$ is constant below a certain value means that when $X'\beta$ is large enough, the function

https://doi.org/10.1017/S0266466601175018 Published online by Cambridge University Press

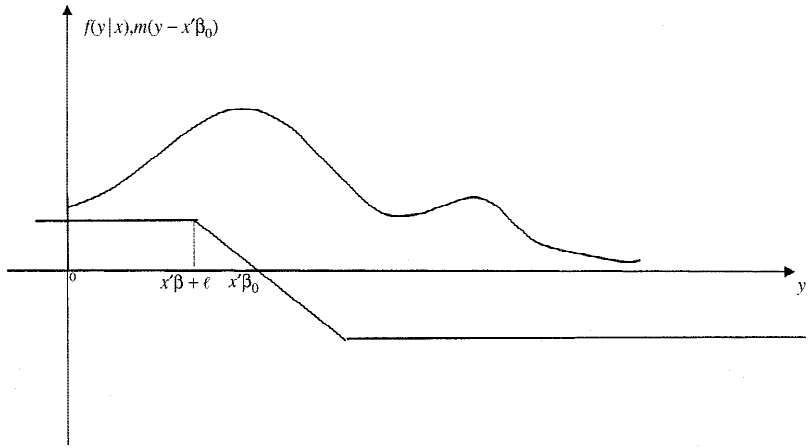


FIGURE 1. Censored regression moment condition.

$m(y - X'\beta)$ will have the same value at the censored y as the latent y, leading to the conditional moment restriction being satisfied in the censored data. Figure 1 illustrates how this occurs.

To be precise, let ℓ equal the supremum of all points where $m(\varepsilon)$ is constant below that point, i.e., $\ell = \sup\{\bar{\varepsilon} : m(\varepsilon) = m(\bar{\varepsilon}) \forall \varepsilon \leq \bar{\varepsilon}\}$, where we take $\ell = -\infty$ if the set is empty. Then

$$\begin{aligned} &1(X'\beta > -\ell)m(y - X'\beta) \\ &\quad = 1(-X'\beta < \ell)\{1(y = 0)m(-X'\beta) + 1(y > 0)m(y - X'\beta)\} \\ &\quad = 1(-X'\beta < \ell)\{1(y^* \leq 0)m(\ell) + 1(y^* > 0)m(y^* - X'\beta)\} \\ &\quad = 1(X'\beta > -\ell)m(y^* - X'\beta), \end{aligned} \tag{5}$$

where the second and third equalities hold by $m(-X'\beta) = m(\ell) = m(y - X'\beta)$ for $-X'\beta \leq \ell$ and $y \leq 0$ and by $y = y^*$ for $y^* > 0$. Letting $E[\cdot]$ denote the expectation in the censored data, and noting that $E[\cdot] = E^*[\cdot]$, leads to the conditional moment restriction

$$E[1(X'\beta_0 > -\ell)m(y - X'\beta_0)|X] = 1(v > -\ell)E^*[m(\varepsilon)|X] = 0, \tag{6}$$

where $v = X'\beta_0$.

As discussed in Section 2, integrating back to an objective function can lead to better identification conditions. To integrate back, note that for a scalar α,

$$1(\alpha > -\ell)m(y - \alpha) = -dq(y - \max\{\alpha, -\ell\})/d\alpha, \tag{7}$$

except where $\alpha = -\ell$. This means that $E[w(X)X \cdot 1(v > -\ell)m(y - v)] = 0$, as implied by (6), is the first-order condition corresponding to minimization at β_0

https://doi.org/10.1017/S0266466601175018 Published online by Cambridge University Press

of $E[w(X)q(y - \max\{X'\beta, -\ell\})]$. An estimator based on the sample analog of this minimization is

$$\hat{\beta} = \underset{\beta \in \mathcal{B}}{\operatorname{argmin}} \sum_{i=1}^{n} w(X_i)q(y_i - \max\{X_i'\beta, -\ell\}), \tag{8}$$

where $\mathcal{B}$ is the set of possible values for β. This estimator is the extension of that of (4) to censored regression.

Any function $m(\varepsilon)$ that is constant below some ℓ can be used in the construction of this type of estimator under the conditional moment restriction of equation (1). The censored regression quantile estimators of Powell (1986a) are a prime example, where $m(\varepsilon) = 1(\varepsilon < 0) - \alpha$ for some $0 < \alpha < 1$. Then equation (1) holds if the αth conditional quantile of ε is zero. Also, as long as X includes a constant, the assumption that the quantile is zero can be replaced by the assumption that the quantile is any constant without affecting the slopes (coefficients of nonconstant X's).

The Winsorized mean estimator of Lee (1992) is another example, where $m(\varepsilon) = -1(\varepsilon < -c)c + 1(-c \leq \varepsilon \leq c)\varepsilon + 1(\varepsilon > c)c$. Because this function is continuous, no density estimation is required for a consistent estimator of the asymptotic variance.

The moment condition of (6) is critically dependent on $m(\varepsilon)$ being constant for all ε small enough. Without this property, no $\sqrt{n}$-consistent estimator will exist. This result follows from the form of the semiparametric information bound. To derive that bound we impose the following condition. Let $\wp_X$ denote the probability distribution of X and U denote Lebesgue measure. Note that $d(X) = \partial E[m(\varepsilon + \alpha)|X]/\partial\alpha|_{\alpha=0}$.

Assumption 3.1. (ε_i, X_i') is independent and identically distributed (i.i.d.) with distribution that is absolutely continuous with respect to $U \times \wp_X$. For $\wp_X$-almost all X there is $f_\varepsilon(\varepsilon|X)$ such that $f(\varepsilon|X) = \int_{-\infty}^{\varepsilon} f_\varepsilon(u|X)\,du$ and $E^*[(1 + \|X\|^2)\{1 + \int[f_\varepsilon(u|X)^2/f(u|X)]\,du\}] < \infty$, as a function of α, $E^*[m(\varepsilon + \alpha)^2|X]$ is bounded in a neighborhood of every α and $E^*[m(\varepsilon)^2|X] > 0$ with probability one, $\text{Prob}(v = -\ell) = 0$, and $E^*[\|X\|^2 d(X)^2/E^*[m(\varepsilon)^2|X]] < \infty$.

Let $w^*(X) = (E^*[m(\varepsilon)^2|X])^{-1}d(X)$. We will also impose Assumption A.1 of the Appendix on the parametric submodels.

THEOREM 3.1. *If Assumptions 3.1 and A.1 are satisfied then the efficient score is*

$$\mathcal{S} = w^*(X)X \cdot 1(X'\beta_0 > -\ell)m(y - X'\beta_0).$$

If $E[\mathcal{S}\mathcal{S}']$ is nonsingular then $(E[\mathcal{S}\mathcal{S}'])^{-1}$ is the semiparametric variance bound.

Because the efficient score is identically zero unless ℓ is finite, $m(\varepsilon)$ being constant below some value is a necessary condition for existence of a (regular) $\sqrt{n}$-consistent estimator. That is, there will be no $\sqrt{n}$-consistent estimator un-

https://doi.org/10.1017/S0266466601175018 Published online by Cambridge University Press

less $\hat{\beta}$ from (8) is available. Furthermore, as shown subsequently, for $w(X) = w^*(X)$, the asymptotic variance of $\hat{\beta}$ will equal the bound. In this sense there is no additional information available to be used in estimation of β_0 other than that used by $\hat{\beta}$.

As examples consider the conditional median and mean zero cases. In the zero conditional median case, the function $m(\varepsilon) = 1(\varepsilon > 0) - 1(\varepsilon < 0)$ is constant below zero, leading to the efficient score $\mathcal{S} = 2f(0|X)X \cdot 1(X'\beta_0 > 0)m(y - X'\beta_0)$. In the conditional mean case, the function $m(\varepsilon) = \varepsilon$ is not constant below any ℓ, and hence the efficient score is zero. Consequently, no $\sqrt{n}$-consistent regular estimator will exist in this case.

This result sidesteps the identification question for β_0. Despite the lack of a $\sqrt{n}$-consistent estimator, it could be that β_0 is identified "at infinity," by approximate satisfaction of $E[m(y - X'\beta_0)|X] = 0$ for large values of $X'\beta_0$. Chamberlain (1986) shows that this is possible for the sample selection model, and similar reasoning may apply here.

To show $\sqrt{n}$-consistent of $\hat{\beta}$ it is useful to make additional assumptions. Let

$$Q = E^*[w(X)d(X)1(v > -\ell)XX'].$$

Assumption 3.2. $E^*[m(\varepsilon + \alpha)|X] \geq (\leq)\ 0$ for $\alpha \geq (\leq)\ 0$, and Q exists and is nonsingular. Also, $\beta_0 \in \text{interior}(\mathcal{B})$, $\mathcal{B}$ is compact, $m(\varepsilon)$ is bounded and continuous almost everywhere, and $w(X) \geq 0$ is bounded.

This condition imposes a "single crossing" property for $E^*[m(\varepsilon + \alpha)|X]$: that its sign does not change on either side of $\alpha = 0$. A simple sufficient condition for single crossing is that $m(\varepsilon)$ is monotonic increasing (i.e., $q(\varepsilon)$ is convex). Restricting attention to bounded $m(\varepsilon)$ does not seem too stringent, because its lower tail must be constant anyway. This restriction could be relaxed at the expense of complicating the conditions. With this condition in place we can obtain a consistency and asymptotic normality result for the estimator. Let $\Sigma = E[w(X)^2 1(v > -\ell)m(\varepsilon)^2 XX']$ and $V = Q^{-1}\Sigma Q^{-1}$.

THEOREM 3.2. *If Assumptions 3.1 and 3.2 are satisfied then* $\sqrt{n}(\hat{\beta} - \beta_0) \xrightarrow{d} N(0,V)$. *Furthermore, if* $w(X) = w^*(X) = d(X)/E[m(\varepsilon)^2|X]$ *then* $V = (E[\mathcal{S}\mathcal{S}'])^{-1}$.

This result also shows that the weighted *m*-estimator would attain the semiparametric efficiency bound if the weight was equal to $w^*(X)$. In this sense there is no information lost from using an estimator like that of (8).

An asymptotic variance estimator is needed for large sample inference procedures based on Theorem 3.2. Here it is important to use a "heteroskedasticity consistent" estimator, analogous to that of Powell (1984). This can be formed in the usual way as $\hat{Q}^{-1}\hat{\Sigma}\hat{Q}^{-1}$ for estimators $\hat{Q}$ and $\hat{\Sigma}$. An estimator of $\hat{\Sigma}$ is straightforward to construct, as

$$\hat{\Sigma} = \sum_{i=1}^{n} 1(\hat{v}_i > -\ell)w(X_i)^2 m(y_i - \hat{v}_i)^2 X_i X_i'/n,$$

https://doi.org/10.1017/S0266466601175018 Published online by Cambridge University Press

where $\hat{v}_i = X_i'\hat{\beta}$. An estimator of Q is more difficult, because it involves the derivative $d(X) = \partial E[m(\varepsilon + \alpha)|X]/\partial\alpha|_{\alpha=0}$. If $m(\varepsilon) = \int_0^\varepsilon m_\varepsilon(u)\,du + C$ for some $m_\varepsilon(\varepsilon)$ that is continuous almost everywhere and a constant C then

$$\hat{Q} = \sum_{i=1}^{n} 1(\hat{v}_i > -\ell)w(X_i)m_\varepsilon(y_i - \hat{v}_i)X_i X_i'/n$$

will do. Otherwise, $d(X)$ will need to be approximated. This may be done by a numerical derivative, as in

$$\hat{Q}_\delta = \sum_{i=1}^{n} w(X_i)X_i X_i'\hat{\Delta}_{i\delta}/n,$$

$$\hat{\Delta}_{i\delta} = q(y_i - \max\{\hat{v}_i + \delta, -\ell\}) + q(y_i - \max\{\hat{v}_i - \delta, -\ell\})$$
$$- 2q(y_i - \max\{\hat{v}_i, -\ell\})/\delta^2.$$

The following result shows consistency of the corresponding estimators.

THEOREM 3.3. *Suppose that Assumptions 3.1 and 3.2 are satisfied. If* $m(\varepsilon) = \int_0^\varepsilon m_\varepsilon(u)\,du + C$ *and* $m_\varepsilon(\varepsilon)$ *is continuous almost everywhere then* $\hat{Q}^{-1}\hat{\Sigma}\hat{Q}^{-1} \xrightarrow{p} V$. *Also, if* $X = (1, x')'$, $E[w(X)\|X\|^6] < \infty$, $\delta \to 0$, *and* $n^{1/4}\delta \to \infty$ *then* $\hat{Q}_\delta^{-1}\hat{\Sigma}\hat{Q}_\delta^{-1} \xrightarrow{p} V$.

Imposing the sixth moment condition simplifies the proof of this result, although it could probably be weakened. Also, it would be useful to have guidelines for the choice or δ in practice, but these are beyond the scope of this paper.

Although precise regularity conditions for feasible efficient estimation are beyond the scope of this paper, the construction of efficient estimators can be described. One approach involves estimating the optimal weight by nonparametric regression, analogously to Robinson (1987) and Newey and Powell (1990). Let $\hat{\nu}_j(X)$ be nonparametric weights satisfying $\sum_{i=1}^n \hat{\nu}_j(X) = 1$ (e.g., as for a kernel estimator where $\hat{\nu}_j(X) = K((X - X_j)/h)/\sum_{i=1}^n K((X - X_i)/h)$ for a kernel function $K(\cdot)$ and a bandwidth h). Then estimators of the optimal weights are, for $\hat{\varepsilon}_i = y_i - \hat{v}_i$,

$$\hat{w}(X) = \sum_{i=1}^{n} \hat{\nu}_i(X)1(\hat{v}_i > -\ell)m_\varepsilon(\hat{\varepsilon}_i)\Big/\sum_{i=1}^{n} \hat{\nu}_i(X)1(\hat{v}_i > -\ell)m(\hat{\varepsilon}_i)^2, \quad \text{or}$$

$$\hat{w}(X) = \sum_{i=1}^{n} \hat{\nu}_i(X)1(\hat{v}_i > -\ell)\hat{\Delta}_{i\delta}\Big/\sum_{i=1}^{n} \hat{\nu}_i(X)1(\hat{v}_i > -\ell)m(\hat{\varepsilon}_i)^2.$$

Then $\hat{w}(X)$ could be used in place of $w(X)$ in equation (8). Alternatively, a one-step version of the estimator could be formed using an initial estimator $\hat{\beta}$,

https://doi.org/10.1017/S0266466601175018 Published online by Cambridge University Press

by constructing $\tilde{Q}$ as either $\hat{Q}$ or $\hat{Q}_\delta$ as given previously with $\hat{w}(X)$ replacing $w(X)$ and then

$$\tilde{\beta} = \hat{\beta} - \tilde{Q}^{-1} \sum_{i=1}^{n} X_i \hat{w}(X_i) 1(\hat{v}_i > -\ell) m(\hat{\varepsilon}_i)/n.$$

As long as $\nu_i(X)$ do not became too "peaked" at X_i too fast (and δ does not shrink too fast), these estimators should be asymptotically efficient.

Another approach to efficient estimation is GMM based on a "rich" set of unconditional moment restrictions, analogously to Chamberlain (1987) and Newey (1993). Let $A(X)$ be an $r \times 1$ vector of functions of X. Note that the moment restrictions $E[A(X)1(v > -\ell)m(\varepsilon)] = 0$ will hold. Let $\hat{\beta}$ be an initial estimator and

$$\tilde{\Omega} = \sum_{i=1}^{n} A(X_i)A(X_i)'1(\hat{v}_i > -\ell)m(\hat{\varepsilon}_i)^2/n,$$

$$\tilde{G} = -\sum_{i=1}^{n} A(X_i)X_i'1(\hat{v}_i > -\ell)m_\varepsilon(\hat{\varepsilon}_i)/n \quad \text{or}$$

$$\tilde{G} = -\sum_{i=1}^{n} A(X_i)X_i'1(\hat{v}_i > -\ell)\hat{\Delta}_{i\delta}/n.$$

Then a one-step GMM estimator is given by

$$\tilde{\beta} = \hat{\beta} - (\tilde{G}'\tilde{\Omega}^{-1}\tilde{G})^{-1}\tilde{G}'\tilde{\Omega}^{-1} \sum_{i=1}^{n} A(X_i)1(\hat{v}_i > -\ell)m(\hat{\varepsilon}_i)/n.$$

As long as the dimension of $A(X)$ grows slowly enough with the sample size and the elements of $A(X)$ are chosen so that they can approximate in mean-square any function of X, this estimator should be asymptotically efficient.

4. TRUNCATED REGRESSION MODELS

For truncated regression the special characteristic that leads to a useful moment condition is that $m(\varepsilon)$ is zero for all ε small enough. That feature means that for $X'\beta$ large enough, $m(y^* - X'\beta) = 0$ as y^* goes from 0 to $-\infty$, so the conditional distribution of $m(y - X'\beta)$ will be the same in the truncated and latent data. Hence the conditional moment restriction $E[m(y - X'\beta_0)|X] = 0$ will be satisfied for $X'\beta_0$ large enough. Figure 2 illustrates this condition.

To be precise, let k equal the supremum of all points where $m(\varepsilon)$ is zero below that point; i.e., $k = \sup\{\bar{\varepsilon} : m(\varepsilon) = 0 \forall \varepsilon \leq \bar{\varepsilon}\}$, where we take $k = -\infty$ if the set is empty. For $\Pi(X) = E^*[1(y > 0)|X]$ note that $E[\cdot|X] = E^*[1(y > 0)(\cdot)|X]/\Pi(X)$. Also, because $m(\varepsilon)$ is zero below k, $m(y - v) = 0$

https://doi.org/10.1017/S0266466601175018 Published online by Cambridge University Press

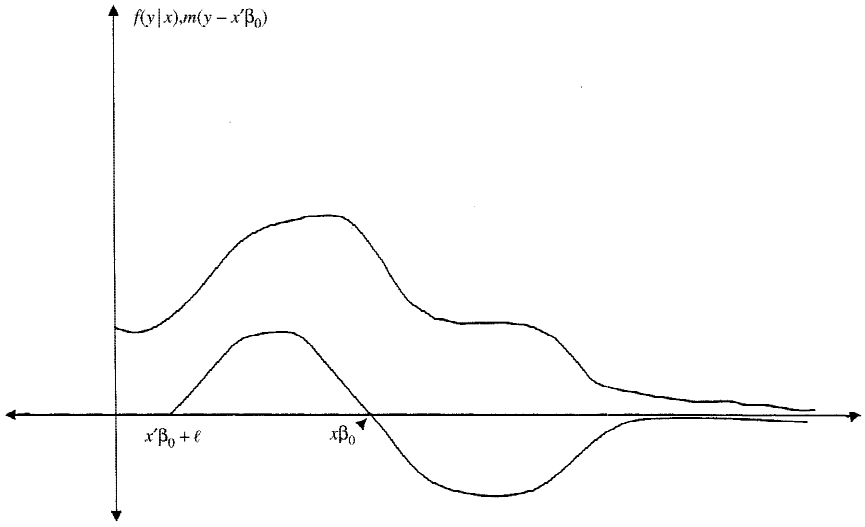


FIGURE 2. Truncated regression moment condition.

for $y \leq 0$ and $-v \leq k$, so that $1(y \leq 0)1(v > -k)m(y - v) = 0$. Then we have

$$\begin{aligned} &E[1(X'\beta_0 > -k)m(y - X'\beta_0)|X] \\ &\quad = \Pi(X)^{-1}E^*[1(v > -k)1(y > 0)m(y - v)|X] \\ &\quad = \Pi(X)^{-1}E^*[1(v > -k)m(\varepsilon)|X] = \Pi(X)^{-1}1(v > -k)E^*[m(\varepsilon)|X] \\ &\quad = 0. \end{aligned} \tag{9}$$

Integrating back to an objective function as was done in Section 3 leads to the estimator

$$\hat{\beta} = \underset{\beta \in \mathcal{B}}{\operatorname{argmin}} \sum_{i=1}^{n} w(X_i)q(y_i - \max\{X_i'\beta, -k\}). \tag{10}$$

Any function $m(\varepsilon)$ that is zero below some k can be used in the construction of this type of estimator, under the conditional moment restriction of equation (1). For example, a function similar to that of Newey (1987) would be $m(\varepsilon) = 1(\varepsilon > 0) - \gamma \cdot 1(\varepsilon > c)$ for $\gamma > 1$, $c > 0$. Here the conditional moment restriction is that $P^*(\varepsilon > 0|X) = \gamma P^*(\varepsilon > c|X)$, where $P^*(A)$ denotes the probability of an event A in the latent data. Another example is Lee's (1993) estimator, where $m(\varepsilon) = 1(-c \leq \varepsilon \leq c)\varepsilon$. Here the restriction is $E^*[\varepsilon| - c \leq \varepsilon \leq c] = 0$. This example is a special case of a class where $m(\varepsilon)$ is an odd function that is zero above some value c.

https://doi.org/10.1017/S0266466601175018 Published online by Cambridge University Press

A new estimator can be obtained by choosing $m(\varepsilon)$ to be an odd function where for $\varepsilon > 0$, $m(\varepsilon) = 1(0 < \varepsilon < c)[c\varepsilon - \varepsilon^2]$. In contrast to previous examples, $m(\varepsilon)$ is continuous here, so that density estimation is not needed to estimate the asymptotic variance of β. This seems to be the first moment condition estimator for the truncated regression model where nonparametric methods are not required for estimation of the asymptotic variance.

The analysis of the properties of this estimator is analogous to those of the censored regression case in Section 3. If $m(\varepsilon)$ is not zero below some value then no $\sqrt{n}$-consistent estimator will exist, and the semiparametric efficiency bound will correspond to the asymptotic variance of $\hat{\beta}$ for $w(X) = w^*(X) = d(X)/E^*[m(\varepsilon)^2|X]$. The semiparametric efficiency bound is given in the following result.

THEOREM 4.1. *If Assumptions 3.1 and A.1 are satisfied and with probability one the support of $f(\varepsilon|X)$ is $\mathbb{R}$, then the efficient score is*

$$\mathcal{S} = w^*(X)X \cdot 1(X'\beta_0 > -k)m(y - X'\beta_0).$$

If $E[\mathcal{S}\mathcal{S}']$ is nonsingular then $(E[\mathcal{S}\mathcal{S}'])^{-1}$ is the semiparametric variance bound.

Asymptotic normality will also hold under conditions similar to those in Section 3. Here let

$$Q = E[w(X)1(v > -k)d(X)XX'/\Pi(X)],$$

$$\Sigma = E[w(X)^2 1(v > -k)m(\varepsilon)^2 XX'], \qquad V = Q^{-1}\Sigma Q^{-1}.$$

The following result gives asymptotic normality.

THEOREM 4.2. *If Assumptions 3.1 and 3.2 are satisfied and $P^*(y > 0) > 0$ then $\sqrt{n}(\hat{\beta} - \beta_0) \xrightarrow{d} N(0, V)$. Furthermore, if $w(X) = w^*(X) = d(X)/E^*[m(\varepsilon)^2|X]$ then $V = (E[\mathcal{S}\mathcal{S}'])^{-1}$.*

Here it should be noted that $m(\varepsilon)$ being zero below some value rules out $m(\varepsilon)$ being monotonic, so that other primitive conditions for the "single crossing" identification condition of Assumption 3.2 must be found. One such condition is that the conditional distribution of ε given X is symmetric (in ε) around zero with unimodal conditional density and that $m(\varepsilon)$ is an odd function that is nonnegative for $\varepsilon \geq 0$. Then, for some choice of C, $-q(\varepsilon) = -\int_0^\varepsilon m(t)\,dt - C$ will be proportional to a unimodal symmetric density function, so that as in Bierens (1981, Lemma 3.2.1), $-E[q(\varepsilon + \alpha)|X]$ is proportional to a unimodal density function as a function of α. This unimodality property will imply that $E[m(\varepsilon + \alpha)|X] = -\partial\{-E[q(\varepsilon + \alpha)|X]\}/\partial\alpha$ is nonnegative for $\alpha \geq 0$ and will give the single crossing property of Assumption 3.2.

A consistent estimator of the asymptotic variance can be formed in a way analogous to that for censored regression. Let $\hat{\Sigma}$, $\hat{Q}$, and $\hat{Q}_\delta$ be exactly as spec-

https://doi.org/10.1017/S0266466601175018 Published online by Cambridge University Press

ified in Section 3 with k replacing ℓ in the formulae. The following result shows consistency of the corresponding estimators.

THEOREM 4.3. *Suppose that Assumptions 3.1 and 3.2 are satisfied and* $P^*(y^* > 0) > 0$. *If* $m(\varepsilon) = \int_0^\varepsilon m_\varepsilon(u)\,du + C$ *and* $m_\varepsilon(\varepsilon)$ *is continuous almost everywhere then* $\hat{Q}^{-1}\hat{\Sigma}\hat{Q}^{-1} \xrightarrow{p} V$. *Also, if* $X = (1, x')'$, $E[w(X)\|X\|^6] < \infty$, $\delta \to 0$, *and* $n^{1/4}\delta \to \infty$ *then* $\hat{Q}_\delta^{-1}\hat{\Sigma}\hat{Q}_\delta^{-1} \xrightarrow{p} V$.

Feasible efficient estimators could be constructed by estimating the optimal weight or by GMM in ways exactly as described in Section 3, with k replacing ℓ. For brevity the details are omitted.

5. INDEPENDENCE OF ε AND X

The sensitivity of the efficiency bound results to the form of conditional location restriction is troubling. For instance, for censored regression the parameters can be $\sqrt{n}$-consistently estimated when the disturbance has conditional median zero but not under conditional mean zero. Some economic models do imply specific location restrictions, such as conditional mean restrictions in rational expectations models. Often though, we do not have strong a priori reasons for choosing one location restriction over another.

Things are different when the disturbance is independent of the regressors. In that case, corresponding to each moment restriction should be a location shift of ε that satisfies it. Consequently, if a constant is included in the regressors, any conditional moment restriction will be satisfied at the true slopes and some constant, including the ones that lead to $\sqrt{n}$-consistent estimation of censored or truncated regression. Thus, any of the estimators will be $\sqrt{n}$-consistent for the slope coefficients in the independence case.

To be specific, consider any function $m(\varepsilon)$ such that there is some μ with $E^*[m(\varepsilon - \mu)] = 0$. Suppose that $X = (1, x')'$ includes a constant and partition $\beta = (\beta_1, \beta_2')'$ conformably with X. Then for $\tilde{\beta}_{10} = \beta_{10} + \mu$ and $\tilde{\beta}_0 = (\tilde{\beta}_{10}, \beta_{20}')'$,

$$E^*[m(y^* - X'\tilde{\beta}_0)|X] = E^*[m(\varepsilon - \mu)|X] = E^*[m(\varepsilon - \mu)] = 0. \quad \textbf{(11)}$$

So, a conditional moment restriction is satisfied in the latent data. Consequently, any of the estimators we have considered for the censored and truncated regression models that are $\sqrt{n}$-consistent under some conditional moment restriction should be $\sqrt{n}$-consistent under independence, for the slope parameters β_{20}. The estimator of the constant will also be $\sqrt{n}$-consistent for the original constant plus μ.

It should be emphasized that the conditional moment restriction in (11) depends on independence and not on any other restriction on the distribution of the disturbance. In particular, with independence, the symmetry assumption of Lee (1993) is not needed for the estimation of a truncated regression. Although symmetry is part of the primitive conditions for the single crossing property

https://doi.org/10.1017/S0266466601175018 Published online by Cambridge University Press

for truncated regression given in Section 4, it is not the fundamental identification condition that leads to (11) being satisfied.

When ε and X are independent, the asymptotic variance of the estimators simplifies and the optimal weight $w(X)$ is equal to one. By independence, neither $d(X)$ nor $E^*[m(\varepsilon)^2|X]$ depends on X. Let $d = \partial E^*[m(\varepsilon - \mu + \alpha)]/\partial\alpha|_{\alpha=0}$ and $\sigma^2 = E^*[m(\varepsilon - \mu)^2]$. Then for censored regression, Theorem 3.2 will hold with

$$V = d^{-2}\sigma^2 M^{-1}\Lambda M^{-1}, \qquad M = E[w(X)1(v > -\ell)XX'],$$
$$\Lambda = E[w(X)^2 1(v > -\ell)XX']. \tag{12}$$

Furthermore, by a standard result (see the proof in the Appendix of Theorem 5.1, which follows), V is minimized (in the positive semidefinite semiorder) at $w(X) = 1$, where

$$V = d^{-2}\sigma^2(E[1(v > -\ell)XX'])^{-1}.$$

The components of this asymptotic variance can be estimated in a similar way to that described in Section 3. Let

$$\hat{d} = \left[\sum_{i=1}^{n} 1(\hat{v}_i > -\ell)m_\varepsilon(y_i - \hat{v}_i)\right] \Big/ \sum_{i=1}^{n} 1(\hat{v}_i > -\ell),$$
$$\hat{d}_\delta = \sum_{i=1}^{n} 1(\hat{v}_i > -\ell)\hat{\Delta}_{i\delta} \Big/ \sum_{i=1}^{n} 1(\hat{v}_i > -\ell),$$
$$\hat{\sigma}^2 = \left[\sum_{i=1}^{n} 1(\hat{v}_i > -\ell)m(y_i - \hat{v}_i)^2\right] \Big/ \sum_{i=1}^{n} 1(\hat{v}_i > -\ell),$$
$$\hat{V}_\delta = \hat{d}_\delta^{-2}\hat{\sigma}^2\left(\sum_{i=1}^{n} 1(\hat{v}_i > -\ell)X_i X_i'/n\right)^{-1},$$
$$\hat{V} = \hat{d}^{-2}\hat{\sigma}^2\left(\sum_{i=1}^{n} 1(\hat{v}_i > -\ell)X_i X_i'/n\right)^{-1}, \tag{13}$$

where $m(\varepsilon)$ is assumed to be differentiable in the definition of $\hat{d}$. The estimators $\hat{V}_\delta$ and $\hat{V}$ are the analog for the independence case of those given in Section 3.

For truncated regression an analogous simplification of the asymptotic variance occurs. Here $E[m(\varepsilon)^2|X] = \sigma^2/\Pi(X)$, so the asymptotic variance will be

$$V = d^{-2}\sigma^2 M^{-1}\Lambda M^{-1}, \qquad M = E[w(X)1(v > -\ell)XX'/\Pi(X)],$$
$$\Lambda = E[w(X)^2 1(v > -\ell)XX'/\Pi(X)]. \tag{14}$$

When $w(X) = 1$ this expression simplifies to

$$V = d^{-2}\sigma^2(E[1(v > -\ell)XX'/\Pi(X)])^{-1}.$$

However, unlike the censored regression model there is not a straightforward way to estimate the components of V. Both d and σ^2 involve expectation

https://doi.org/10.1017/S0266466601175018 Published online by Cambridge University Press

from the latent data. Also, the presence of $\Pi(X)$ makes it difficult to estimate $E[1(v > -\ell)XX'/\Pi(X)]$. Intuitively, the truncation induces dependence between ε and X, so that one still has to use the "heteroskedasticity consistent" estimator previously given.

The following result makes this discussion precise.

THEOREM 5.1. *If Assumptions 3.1 and 3.2 are satisfied,* $P^*(y^* > 0) > 0$ *in the truncated case,* ε *and* X *are independent, and* $X = (1, x')'$ *then* $\sqrt{n}(\hat{\beta} - \tilde{\beta}_0) \xrightarrow{d} N(0,V)$ *for* V *in equation (12) for the censored case and equation (14) for the truncated case. Also, in each case* V *is minimized at* $w(X) = 1$. *In addition, in the censored case, if* $m(\varepsilon)$ *is continuously differentiable with bounded derivative then* $\hat{V} \xrightarrow{p} V$, *whereas if* $\delta \to 0$ *and* $n^{1/4}\delta \to \infty$ *then* $\hat{V}_\delta \xrightarrow{p} V$.

It should be noted that Assumptions 3.1 and 3.2 simplify under independence of ε, e.g., for $f(\varepsilon) = f(\varepsilon|X)$, requiring only $E^*[\|X\|^2]$ and $\int[f_\varepsilon(\varepsilon)^2/f(\varepsilon)]\,d\varepsilon$ finite.

Because many estimators will be consistent for the slope coefficients under independence, it is possible to choose from among a group of estimators the most efficient one or to combine several different estimators for improved efficiency. These approaches were suggested by Powell (1986a) for censored regression quantiles, and they apply equally well to other moment conditions for censored or truncated data.

It turns out that neither of these methods makes efficient use of information. Intuitively, combining several restrictions generally leads to higher efficiency than any of the individual ones and so dominates the highest efficiency from a single restriction. Furthermore, because of heteroskedasticity induced by the presence of the trimming indicator $1(X'\beta > -\ell)$, combining several restrictions turns out to be more efficient than combining the estimators obtained from the individual restrictions. Of course, these methods might have better small sample properties, but as yet we do not know.

To describe efficient estimation from multiple conditional moment restrictions, let $m_j(\varepsilon)$, $(j = 1,\ldots,J)$ be J functions of a scalar argument and ℓ_j (or k_j in the truncated case) be as given previously with m_j replacing m. Reparameterize so that β denotes the slope coefficients and γ_j a separate constant for each m_j. Suppose that there is γ_{j0} such that for $X = (1, x')'$ and $\varepsilon = y - x'\beta_0$, $E[m(\varepsilon - \gamma_{j0})] = 0$, $(j = 1,\ldots,J)$. Let $\theta = (\beta', \gamma_1,\ldots,\gamma_J)'$ and

$$\rho(z,\theta) = (1(\gamma_1 + x'\beta > -\ell_1)m_1(y - \gamma_1 - x'\beta),\ldots,1(\gamma_J + x'\beta > -\ell_J) \times m_J(y - \gamma_J - x'\beta))'$$

(with k_j replacing ℓ_j in the truncated case). Then

$$E[\rho(z,\theta_0)|x] = 0. \tag{15}$$

https://doi.org/10.1017/S0266466601175018 Published online by Cambridge University Press

This vector of conditional moment restrictions can be used in GMM estimation. For a conformable matrix function $A(x)$ and $\hat{g}(\theta) = \sum_{i=1}^{n} A(x_i)\rho(z_i,\theta)/n$, an estimator of θ_0 can be obtained by minimizing $\hat{g}(\theta)'\hat{W}\hat{g}(\theta)$, where $\hat{W}$ is some positive semidefinite matrix.

We focus here on the case with efficient $A(x)$, which minimizes the asymptotic variance. To obtain the efficient $A(x)$ we need to generalize the formula of Chamberlain (1987) to allow for $\Omega(x) = E[\rho(z,\theta_0)\rho(z,\theta_0)'|x]$ to be singular. Let $G(x) = \partial E[\rho(z,\theta)|x]/\partial\theta|_{\theta=\theta_0}$ and for any square matrix B let B^- denote any generalized inverse, satisfying $BB^-B = B$.

THEOREM 5.2. *If there is $F(x)$ such that $G(x) = \Omega(x)F(x)$ and $E[F(x)'\Omega(x)F(x)]$ exists and is nonsingular, then among all $A(x)$ and $W = \text{plim}(\hat{W})$ positive semidefinite where $WE[A(x)G(x)]$ has full column rank, the asymptotic variance of the* GMM *estimator is minimized at $A^*(x) = G(x)'\Omega(x)^-$.*

Dridi (1999) had previously derived a partial version of this result.

To apply this result to censored and truncated regression, let $I(x,\theta)$ be the selection matrix such that $I(x,\theta)\rho(z,\theta)$ consists of those elements of $\rho(z,\theta)$ with $\gamma_j + x'\beta > -\ell_j$ (with k_j instead of ℓ_j in the truncated case). Also, for $\varepsilon = y - x'\beta_0$ let Φ denote the $J \times J$ matrix with $\Phi_{jk} = E^*[m_j(\varepsilon - \gamma_{j0})m_k(\varepsilon - \gamma_{k0})]$, $d_j = -\partial E^*[m_j(\varepsilon - \gamma_{j0} + \alpha)]/\partial\alpha|_{\alpha=0}$, $D = (d_1,\ldots,d_J)'$, $\text{diag}(D)$ denote a diagonal matrix with the elements of D down the diagonal, and $I(x) = I(x,\theta_0)$. Also, let $M(x) = [Dx', \text{diag}(D)]'[I(x)\Phi I(x)']^{-1}[Dx', \text{diag}(D)]$.

THEOREM 5.3. *If Φ and $E^*[M(x)]$ are nonsingular then, for censored and truncated regression, the variance minimizing choice of $A(x)$ is*

$$A^*(x) = -[Dx', \text{diag}(D)]'I(x)[I(x)\Phi I(x)']^{-1}I(x). \tag{16}$$

An important feature of these optimal functions is that they consist of known functions of x and constant matrices d and Φ. This allows $A^*(x)$ to be estimated without resorting to nonparametric methods. In the censored case, direct estimation of the constants that appear in $A^*(x)$ is possible. Let $\hat{\theta}$ be some preliminary estimator of the parameters. For example, this estimator could be constructed by obtaining $\hat{\beta}$ from the estimator of equation (8) (for some $m(\varepsilon)$) and, for $dq_j(\varepsilon)/d\varepsilon = m_j(\varepsilon)$, $\hat{\gamma}_j = \text{argmin}_\gamma \sum_{i=1}^{n} q_j(y_i - \max\{x_i'\hat{\beta} + \gamma, -\ell_j\})$. Also, let $\hat{d}_j$ (or $\hat{d}_{j\delta}$) be as in equation (13), with m_j replacing m, and let $\hat{v}_i = x_i'\hat{\beta}$, $\hat{1}_{ji} = 1(\hat{\gamma}_j + \hat{v}_i > -\ell_j)$, $\hat{m}_{ji} = m_j(y_i - \hat{\gamma}_j - \hat{v}_i)$, and $\hat{\Phi}_{jk} = \sum_{i=1}^{n} \hat{1}_{ji}\hat{1}_{ki}\hat{m}_{ji}\hat{m}_{ki}/\sum_{i=1}^{n} \hat{1}_{ji}\hat{1}_{ki}$ and $\hat{I}(x) = I(x,\hat{\theta})$. Then an estimator of $A^*(x)$ is

$$\hat{A}(x) = \hat{G}(x)'\hat{I}(x)'[\hat{I}(x)\hat{\Phi}\hat{I}(x)']^{-1}\hat{I}(x), \qquad \hat{G}(x) = \hat{I}(x)'\hat{I}(x)[\hat{D}\cdot x', \text{diag}(\hat{D})].$$

https://doi.org/10.1017/S0266466601175018 Published online by Cambridge University Press

A one-step efficient GMM and asymptotic variance estimator could then be obtained as

$$\tilde{\theta} = \hat{\theta} - \left[\sum_{i=1}^{n} \hat{A}(x_i)\hat{G}(x_i)\right]^{-1} \sum_{i=1}^{n} \hat{A}(x_i)\rho(z_i,\hat{\theta}), \qquad \hat{V} = \left(\sum_{i=1}^{n} \hat{A}(x_i)\hat{G}(x_i)/n\right)^{-1}.$$

This estimator will be asymptotically equivalent to GMM with the efficient $A^*(x)$, under regularity conditions like those already given. For brevity this result is omitted.

In the truncated case direct estimation of D and Φ is problematic, as a result of the heteroskedasticity induced by the truncation. This problem can be circumvented by noting that the $\Omega(x)$ varies discretely, shifting as each $1(\gamma_{j0} + x'\beta_0 > -\ell_j)$ changes. Consequently, the conditional moment restrictions will be "spanned" by unconditional ones based on an exhaustive set of outcomes for these "trimming" indicators, leading to an efficient GMM estimator.

To describe this estimator, suppose that j can been ordered such that

$$\gamma_{j+1,0} + k_{j+1} < \gamma_{j,0} + k_j \qquad (j = 1,\ldots,J-1).$$

Let I_j denote a $j \times j$ identity matrix, $H_j = [I_j, 0]$, and

$$g^j(z,\theta) = 1(-\gamma_{j+1} - k_{j+1} \le x'\beta > -\gamma_j - k_j)(H_j \otimes X)\rho(z,\theta) \qquad (j = 1,\ldots,J-1),$$

$$g^J(z,\theta) = 1(x'\beta > -\gamma_j - k_J)\rho(z,\theta) \otimes X,$$

$$g(z,\theta) = (g^1(z,\theta)',\ldots,g^J(z,\theta)')'. \tag{17}$$

The unconditional moment restriction $E[g(z,\theta_0)] = 0$ is implied by equation (15). A one-step efficient GMM estimator can be based on this restriction. For an initial estimator $\hat{\theta}$ like that given previously (with k_j replacing ℓ_j) let $\hat{v}_i = x_i'\hat{\beta}$, $\hat{\varepsilon}_i = y_i - \hat{v}_i$,

$$\hat{\Omega}_j = \sum_{i=1}^{n} g_j(z_i,\hat{\theta})g_j(z_i,\hat{\theta})'/n, \qquad \hat{\Omega} = \operatorname{diag}(\hat{\Omega}_1,\ldots,\hat{\Omega}_J),$$

$$\tilde{1}_{ji} = 1(-\hat{\gamma}_{j+1} - k_{j+1} \le \hat{v}_i > -\hat{\gamma}_j - k_j),$$

$$\hat{G}_j = -\sum_{i=1}^{n} \tilde{1}_{ji}(m_{1\varepsilon}(\hat{\varepsilon}_i - \hat{\gamma}_1),\ldots,m_{j\varepsilon}(\hat{\varepsilon}_i - \hat{\gamma}_j))' \otimes X_i X_i'/n,$$

$$\hat{G} = [\hat{G}_1',\ldots,\hat{G}_J']'.$$

Alternatively, when $m_j(\varepsilon)$ is not differentiable, one could replace $m_{j\varepsilon}(\hat{\varepsilon}_i - \hat{\gamma}_j)$ by $\hat{\Delta}_{ij\delta}$ in the construction of $\hat{G}$, where $\hat{\Delta}_{ji\delta}$ is defined like $\Delta_{i\delta}$ with $q_j(\varepsilon) = \int_0^{\varepsilon} m_j(t)\,dt$ replacing $q(\varepsilon)$ and k_j replacing ℓ. Noting that $\hat{\Omega}^{-1} = \operatorname{diag}[\hat{\Omega}_1^{-1},\ldots,\hat{\Omega}_J^{-1}]$, a one-step efficient GMM estimator is

https://doi.org/10.1017/S0266466601175018 Published online by Cambridge University Press

$$\tilde{\theta} = \hat{\theta} - (\hat{G}'\hat{\Omega}^{-1}\hat{G})^{-1}\hat{G}'\hat{\Omega}^{-1}\sum_{i=1}^{n} g(z_i,\hat{\theta})/n, \qquad \hat{V} = (\hat{G}'\hat{\Omega}^{-1}\hat{G})^{-1}.$$

Under regularity conditions like those already given, this estimator will be asymptotically normal. For brevity, we omit this result, although we do use the asymptotic variance formula to give the following theorem.

THEOREM 5.4. *The asymptotic variance of $\tilde{\theta}$ is the same as a* GMM *estimator based on moment functions $A^*(x)\rho(z,\theta_0)$.*

One potential problem with this estimator is that the asymptotic distribution approximation may be poor, as a result of the large number of moment restrictions. The mutual orthogonality of the $g^j(z,\theta)$ vector may mitigate that problem, but there still seem to be many moment restrictions. An estimator that used more of the structure of $A^*(x)$ in its construction might have better properties. The development of such an estimator is left to future work.

An interesting question concerns the efficiency of these estimators relative to the variance bound derived by Cosslett (1987) for the censored case and Bickel, Klaasen, Ritov, and Wellner (1993) for the truncated case. In Newey and Chipty (2000) it is shown that when enough censored regression quantiles, or enough truncated moments in the truncated case, are combined the asymptotic variance of $\tilde{\beta}$ and $\tilde{\gamma}_j$ approaches the semiparametric bound. In this sense the type of estimator developed here uses all of the information available in censored or truncated regression with independent disturbance and regressors.

REFERENCES

Bickel, P.J., C.A.J. Klaasen, Y. Ritov, & J.A. Wellner (1993) *Efficient and Adaptive Estimation for Semiparametric Models.* Baltimore: Johns Hopkins Press.

Bierens, H.J. (1981) *Robust Methods and Asymptotic Theory in Nonlinear Econometrics.* New York: Springer-Verlag.

Chamberlain, G. (1986) Asymptotic efficiency in semiparametric models with censoring. *Journal of Econometrics* 32, 189–218.

Chamberlain, G. (1987) Asymptotic efficiency in estimation with conditional moment restrictions. *Journal of Econometrics* 34, 305–334.

Cosslett, S.R. (1987) Efficiency bounds for distribution-free estimators of the binary choice and censored regression models. *Econometrica* 55, 559–585.

Dridi, R. (1999) Simulated asymptotic least squares. Paper presented at Econometric Study Group, Bristol, 1999.

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

Ibragimov, I.A. & R.Z. Hasminskii (1981) *Statistical Estimation: Asymptotic Theory.* New York: Springer-Verlag.

Lee, M.J. (1992) Winsorized mean estimator for censored regression. *Econometric Theory* 8, 368–382.

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

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

https://doi.org/10.1017/S0266466601175018 Published online by Cambridge University Press

Newey, W.K. (1989) An introduction to semiparametric efficiency bounds. *Annals d'Economie et de Statistique* 13, 1–47.
Newey, W.K. (1990) Semiparametric efficiency bounds. *Journal of Applied Econometrics* 5, 99–135.
Newey, W.K. (1991) Efficient estimation of tobit models under conditional symmetry. In W. Barnett, J. Powell, & G. Tauchen (eds.), *Semiparametric and Nonparametric Methods in Statistics and Econometrics.* Cambridge: Cambridge University Press.
Newey, W.K. (1993) Efficient estimation of models with conditional moment restrictions. In G.S. Maddala, C.R. Rao, & H.D. Vinod (eds.), *Handbook of Statistics*, vol. 11. Amsterdam: North-Holland.
Newey, W.K. & T. Chipty (2000) Efficient Semiparametric Estimation via Moment Restrictions. Preprint, Massachusetts Institute of Technology.
Newey, W.K. & D. McFadden (1994) Large sample estimation and hypothesis testing. In R. Engle & D. McFadden (eds.), *Handbook of Econometrics*, vol. 4, pp. 2113–2245. Amsterdam: North-Holland.
Newey, W.K. & J.L. Powell (1990) Efficient estimation of linear and type I censored regression models under conditional quantile restrictions. *Econometric Theory* 6, 295–317.
Powell, J.L. (1984) Least absolute deviation estimation for the censored regression model. *Journal of Econometrics* 25, 303–325.
Powell, J.L. (1986a) Censored regression quantiles. *Journal of Econometrics* 32, 143–155.
Powell, J.L. (1986b) Symmetrically trimmed least squares estimation for tobit models. *Econometrica* 54, 1435–1460.
Robinson, P. (1987) Asymptotically efficient estimation in the presence of heteroskedasticity of unknown form. *Econometrica* 55, 875–891.
Van der Vaart, A. & J. Wellner (1996) *Weak Convergence and Empirical Processes.* New York: Springer-Verlag.

APPENDIX

Throughout the Appendix C will denote a generic constant that may be different in different uses. The following assumption is useful in the proofs of the semiparametric information bounds.

Assumption A.1. Parametric submodels correspond to latent densities $f(\varepsilon, X|\eta) = f(\varepsilon|X,\eta)f(X|\eta)$ such that $f(\varepsilon|X,\eta_1)f(X|\eta_2)$ is smooth in $(\eta_1', \eta_2')'$, $f(\varepsilon|X,\eta_0) = f(\varepsilon|X)$, and for almost all X, $f(\varepsilon|X,\eta)$ is a density for ε that is smooth in η and with $\int m(\varepsilon)^2 f(\varepsilon|X,\eta)\,d\varepsilon$ bounded in a neighborhood of η_0. Also, in the truncated case, for $\theta = (\beta', \eta')'$, $\text{Prob}_\theta(y > 0) > 0$ for θ.

The proof of Theorem 3.1 will make use of the following lemma. Let $s(\varepsilon, X) = f_\varepsilon(\varepsilon|X)/f(\varepsilon|X)$ be the conditional location score for the density $f(\varepsilon|X)$ and $S_\beta = -X\cdot\{1(y > 0)s(\varepsilon, X) + 1(y = 0)E[s(\varepsilon, X)|y = 0, X]\}$ be the score for β in the censored case.

LEMMA A.1. *If Assumptions 3.1 and A.1 are satisfied then* $1(v \leq -\ell)S_\beta \in \mathcal{T}$.

Proof. By Lemma A.5 of Newey and Powell (1990) it follows that the tangent set of the latent model is $\mathcal{T}^* = \{\delta = \delta(\varepsilon, X) : E[\delta] = 0,\ E[m(\varepsilon)\delta|X] = 0,\ E[\|\delta\|^2] < \infty\}$.

https://doi.org/10.1017/S0266466601175018 Published online by Cambridge University Press

Also, by Lemma A.2 of Newey and Powell (1990), for $\mathcal{C}(\delta) = E[\delta|y,x]$, it follows that the tangent set of the observed data is the mean-square closure of the set $\{\mathcal{C}(\delta): \delta \in \mathcal{T}\}$. Consider any $\bar{\ell} > \ell$. Then for $v \leq -\bar{\ell}$, $m(\varepsilon)$ is nonconstant on $(-\infty, -v]$, because $-v \geq \bar{\ell} > \ell$. Let $L = \{\varepsilon \leq -v\}$. Then for $v \geq -\ell$, $\mathrm{Var}(m(\varepsilon)|L,X) > 0$, and hence $A(X) = E[1(L)(1,m(\varepsilon))'(1,m(\varepsilon))|X]$ is nonsingular. For $M > 0$ let $D(X) = 0$ if $P(L|X) = 0$ and otherwise

$$D(X) = E[1(\varepsilon > -v)(1,m(\varepsilon))s(\varepsilon,X)|X]A(X)^{-1},$$

$$1_M = 1(\|D(X)\| \leq M)\cdot 1(v \leq -\bar{\ell}),$$

$$\tilde{t} = -X\cdot 1_M\cdot\{1(\varepsilon > -v)s(\varepsilon,X) - D(X)1(\varepsilon \leq -v)(1,m(\varepsilon))'\}.$$

Note that

$$\begin{aligned} E[\tilde{t}(1,m(\varepsilon))|X] &= -X\cdot 1_M\cdot E[\{1(\varepsilon > -v)s(\varepsilon,X) - D(X)1(\varepsilon \leq -v)(1,m(\varepsilon))'\}(1,m(\varepsilon))|X] \\ &= 0, \end{aligned}$$

where the last equality follows by the definition of $D(X)$. Also, $\tilde{t}$ has second moments by Assumption 3.1, so that $t = \mathcal{C}(\tilde{t}) \in \mathcal{T}$. Furthermore,

$$\begin{aligned} t &= -X\cdot 1_M\{1(y > 0)s(\varepsilon,X) - 1(y = 0)D(x)E[(1,m(\varepsilon))'|y,X]\} \\ &= -X\cdot 1_M\{1(y > 0)s(\varepsilon,X) - 1(y = 0)E[1(\varepsilon > -v)s(\varepsilon,X)|X]/P(L|X)\} \\ &= -X\cdot 1_M\{1(y > 0)s(\varepsilon,X) + 1(y = 0)E[s(\varepsilon,X)|L,X]\} = 1_M S_\beta, \end{aligned}$$

where the second equality follows by $1(y = 0)E[(1,m(\varepsilon))'|y,X]$ being equal to the first column of $A(X)$ divided by $P(L|X)$. It follows that $E[\|1(v \leq -\bar{\ell})S_\beta - t\|^2] = E[\|S_\beta\|^2 1(\|D(X)\| > M)1(v \leq -\bar{\ell})] \to 0$ as $M \to \infty$. Therefore, $1(v \leq -\bar{\ell})S_\beta \in \mathcal{T}$. Finally, because $\mathrm{Prob}(v = -\ell) = 0$, $1(\ell < v \leq \bar{\ell}) \to 0$ almost surely as $\bar{\ell} \to \ell$, so by the dominated convergence theorem, $E[\|1(v \leq -\ell)S_\beta - 1(v \leq -\bar{\ell})S_\beta\|^2] = E[1(\ell < v \leq -\bar{\ell})\|S_\beta\|^2] \to 0$ as $\bar{\ell} \to \ell$. ■

Proof of Theorem 3.1. Let

$$\delta(\varepsilon,X) \equiv -X\cdot 1(v > -\ell)\cdot\{s(\varepsilon,X) + w^*(X)m(\varepsilon)\}. \tag{A.1}$$

By Assumption 3.1 this random variable has finite mean-square. By Lemma 7.2 of Ibragimov and Hasminskii (1981) and $E[m(\varepsilon+\alpha)^2|X]$ bounded on a neighborhood of $\alpha = 0$, $E[m(\varepsilon+\alpha)|X] = \int m(\varepsilon)f(\varepsilon-\alpha|X)\,d\varepsilon$ is differentiable in α at $\alpha = 0$ and $d(X) = \partial E[m(\varepsilon+\alpha)|X]/\partial\alpha = -E[m(\varepsilon)s(\varepsilon,X)|X]$. Then by $E[s(\varepsilon,X)|X] = 0$,

$$\begin{aligned} E[\delta(\varepsilon,X)|X] &= -X1(v > -\ell)\{E[s(\varepsilon,X)|X] + w^*(X)E[m(\varepsilon)|X]\} = 0, \\ E[m(\varepsilon)\delta(\varepsilon,X)|X] &= -X1(v > -\bar{\ell})\{E[m(\varepsilon)s(\varepsilon,X)|X] + w^*(X)E(m(\varepsilon)^2|X]\} \\ &= 0. \end{aligned} \tag{A.2}$$

https://doi.org/10.1017/S0266466601175018 Published online by Cambridge University Press

Therefore, $\mathcal{C}(\delta) = 1(v > -\ell)S_\beta - \mathcal{S} \in \mathcal{T}$. Then, by linearity of $\mathcal{T}$ and Lemma A.1, $\mathcal{S}_\beta - \mathcal{S} = \mathcal{C}(\delta) + 1(v \leq -\ell)S_\beta \in \mathcal{T}$; i.e., $\mathcal{S} = S_\beta - \iota$ for $\iota \in \mathcal{T}$. Furthermore, by (1), for any $\delta(\varepsilon, X)$ satisfying $E[\delta(\varepsilon, X)|X] = 0$, $E[\delta(\varepsilon, X)m(\varepsilon)|X] = 0$,

$$\begin{aligned} E[\mathcal{S}'\mathcal{C}(\delta)] &= E[\mathcal{S}'E[\delta(\varepsilon, X)|y, X]] = E[E[\mathcal{S}'\delta(\varepsilon, X)|y, X]] \\ &= E[w^*(X)X'1(v > -\ell)m(\varepsilon)\delta(\varepsilon, X)] \\ &= E[w^*(X)X'1(v > -\ell)E[m(\varepsilon)\delta(\varepsilon, X)|X]] = 0, \end{aligned}$$

where the third equality follows by equation (5). Then, because $\mathcal{T}$ is the mean-square of objects of the form $\mathcal{C}(\delta)$, it follows that $\mathcal{S}$ is orthogonal to the tangent set. Therefore, $\mathcal{S}$ is the residual from the mean-square projection of S_β on the tangent set and hence is the efficient score. ■

The following lemma is useful for the proofs of Theorems 3.2 and 4.2.

LEMMA A.2. *Consider an estimator* $\hat{\beta} = \text{argmin}_{\beta \in \mathcal{B}} \sum_{i=1}^n q(z_i, \beta)$. *If (i)* $\hat{\beta} \xrightarrow{p} \beta_0 \in$ *interior*$(\mathcal{B})$; *(ii) there is* $b(z)$ *with* $|q(z, \tilde{\beta}) - q(z, \beta)| \leq b(z)\|\tilde{\beta} - \beta\|$ *and* $E[b(z)^2] < \infty$; *(iii) there is* $U(z)$ *with* $E[U(z)] = 0$ *and* $E[U(z)U(z)']$ *nonsingular such that with probability one,* $[q(z, \beta) - q(z, \beta_0) - U(z)'(\beta - \beta_0)]/\|\beta - \beta_0\| \to 0$ *as* $\beta \to \beta_0$; *(iv) for* $\bar{q}(\beta) = E[q(z_i, \beta)]$ *there is* Q *nonsingular with* $\bar{q}(\beta) = \bar{q}(\beta_0) + (\beta - \beta_0)'Q(\beta - \beta_0)/2 + o(\|\beta - \beta_0\|^2)$; *then* $\sqrt{n}(\hat{\beta} - \beta_0) \xrightarrow{d} N(0, Q^{-1}E[U(z)U(z)']Q^{-1})$.

Proof of Lemma A.2. By (ii) and the triangle inequality,

$$r(z, \beta) = |[q(z, \beta) - q(z, \beta_0) - U(z)'(\beta - \beta_0)]/\|\beta - \beta_0\|| \leq b(z) + \|U(z)\|.$$

Then by (iii) and the dominated convergence theorem, $E[r(z, \beta)^2] \to 0$ as $\beta \to \beta_0$. The conclusion then follows by Example 3.2.22 of Van der Vaart and Wellner (1996). ■

Proof of Theorem 3.2. Consider $\tilde{r} = \min\{-X'\beta, \ell\} \leq \ell$ and $r = \min\{-v, \ell\} \leq \ell$. Note that for $\varepsilon \leq \ell$, $q(\varepsilon)$ is linear in ε. Then for $y^* \leq 0$, $y^* - \max\{X'\beta, -\ell\} = y^* + \tilde{r} \leq \ell$, so that

$$\begin{aligned} &q(y - \max\{X'\beta, -\ell\}) - q(y - \max\{v, -\ell\}) \\ &\quad - [q(y^* - \max\{X'\beta, -\ell\}) - q(y^* - \max\{v, -\ell\})] \\ &= 1(y^* < 0)\{q(\tilde{r}) - q(r) - [q(y^* + \tilde{r}) - q(y^* + r)]\} \\ &= 1(y^* < 0)\{q(\tilde{r} - r - [y^* + \tilde{r} - (y^* + r)])\} = 1(y^* < 0)q(0). \end{aligned} \tag{A.3}$$

Then, because none of $q(y^* - \max\{v, -\ell\})$, $q(y - \max\{v, -\ell\})$, $1(y^* < 0)q(0)$, and $q(\varepsilon)$ depend on β, and adding terms to the objective function that do not depend on β does not change the minimum, it follows that for $z = (y^*, X)$,

$$\hat{\beta} = \text{argmin}_{\beta \in \mathcal{B}} \hat{Q}(\beta), \qquad \hat{Q}(\beta) = \sum_{i=1}^n q(z_i, \beta)/n,$$

$$q(z, \beta) = w(X)[q(y^* - \max\{X'\beta, -\ell\}) - q(\varepsilon)]. \tag{A.4}$$

https://doi.org/10.1017/S0266466601175018 Published online by Cambridge University Press

Note that $q(z,\beta)$ is continuous in β. Also, $q(\varepsilon)$ and $\max\{v,-\ell\}$ are Lipschitz in ε and v, respectively, so that

$$|q(z,\beta)| \le Cw(X)|v - \max\{X'\beta,-\ell\}| \le Cw(X)\|X\|(\|\beta\| + \|\beta_0\|)$$
$$\le Cw(X)\|X\|,$$
$$|q(z,\tilde\beta) - q(z,\beta)| \le Cw(X)\|X\|\|\tilde\beta - \beta\|. \qquad \textbf{(A.5)}$$

It follows by a standard uniform law of large numbers that $\bar Q(\beta) = E[q(z,\beta)]$ exists and is continuous in β and that $\sup_{\beta\in\mathcal{B}}|\hat Q(\beta) - \bar Q(\beta)| \xrightarrow{p} 0$. Because $\hat\beta = \operatorname{argmin}_{\beta\in\mathcal{B}}\hat Q(\beta)$, consistency will follow from the standard argument if $\bar Q(\beta)$ is uniquely minimized at β_0. By the definition of $q(\cdot)$, $q(y^* - \tilde\alpha) - q(y^* - \alpha) = \int_{y^*-\alpha}^{y^*-\tilde\alpha} m(u)\,du$. By $m(\varepsilon)$ bounded, $|[q(y^* - \tilde\alpha) - q(y^* - \alpha)]/(\tilde\alpha - \alpha)| \le C$, and by the fundamental theorem of calculus and $m(\varepsilon)$ continuous almost everywhere, as $\tilde\alpha \to \alpha$, $[q(y^* - \tilde\alpha) - q(y^* - \alpha)]/(\tilde\alpha - \alpha) \to m(y^* - \alpha)$ almost surely. Therefore, by the dominated convergence theorem, $E[q(y^* - \tilde\alpha) - q(y^* - \alpha)|X]/(\tilde\alpha - \alpha) = \int[\int_{y^*-\alpha}^{y^*-\tilde\alpha} m(u)\,du/(\tilde\alpha - \alpha)]f(y^*|X)\,dy^* \to E[m(y^* - \alpha)|X]$. It follows that $\bar q(\alpha,X) = E[q(y^* - \alpha) - q(\varepsilon)|X]$ is differentiable in α with derivative $E[m(y^* - \alpha)|X]$. Then, by Assumption 3.2, $\bar q(\alpha,X)$ is increasing in α for $\alpha \ge v$, decreasing in α for $\alpha \le v$, having a global minimum at $\alpha = v$. Furthermore, $\bar q(\max\{\alpha,-\ell\},X)$ also has a global minimum at $\alpha = v$, because $\bar q(\alpha,X)$ is increasing in α for $\alpha \ge v$. Therefore,

$$E[q(z,\beta)|X] = w(X)\bar q(\max\{X'\beta,-\ell\},X) \ge w(X)\bar q(\max\{v,-\ell\},X)$$
$$= E[q(z,\beta_0)|X]. \qquad \textbf{(A.6)}$$

Now, for any $\beta \ne \beta_0$ consider the event $\mathcal{A} = \{w(X) > 0,\ v > -\ell,\ d(X) > 0,$ and $X'\beta \ne v\}$. Assumption 3.2 implies that for $\delta = \beta - \beta_0 \ne 0$, $E[w(X)d(X)1(v > -\ell) \times (v - X'\beta)^2] = \delta'Q\delta > 0$, to that $\text{Prob}(\mathcal{A}) > 0$. Also, it follows from the proof of Theorem 3.1 that $E[m(y^* - \alpha)|X]$ is differentiable at $\alpha = v$ with derivative $-d(X)$, so that $d(X) = \partial^2\bar q(\alpha,X)/\partial\alpha^2|_{\alpha=v}$. When $d(X) > 0$, $\bar q(\alpha,X)$ will have a unique local minimum at $\alpha = v$, which is a unique global minimum by the previous argument. When $\mathcal{A}$ occurs, $\max\{X'\beta,-\ell\} \ne v = \max\{v,-\ell\}$, so that

$$E[q(z,\beta)|X] = w(X)\bar q(\max\{X'\beta,-\ell\},X) > w(X)\bar q(\max\{v,-\ell\},X) = E[q(z,\beta_0)|X].$$

Then by $\text{Prob}(\mathcal{A}) > 0$, $\bar Q(\beta) = E[q(z,\beta)] = P(\mathcal{A})E[q(z,\beta)|\mathcal{A}] + P(\mathcal{A}^c) \times E[q(z,\beta)|\mathcal{A}^c] = P(\mathcal{A})E[E[q(z,\beta)|X]|\mathcal{A}] + P(\mathcal{A}^c)E[E[q(z,\beta)|X]|\mathcal{A}^c] > P(\mathcal{A}) \times E[E[q(z,\beta_0)|X]|\mathcal{A}] + P(\mathcal{A}^c)E[E[q(z,\beta_0)|X]|\mathcal{A}^c] = \bar Q(\beta_0)$. Thus, the $\bar Q(\beta_0)$ has a unique minimum at β_0, giving $\hat\beta \xrightarrow{p} \beta$.

We have now shown that condition (i) of Lemma A.1 holds. Condition (ii) follows by (A.5). Now, for condition (iii), let $\mathcal{A} = \{(y^*,X)$: $v \ne -\ell$ and $q(\varepsilon)$ is continuously differentiable at $y^* - \max\{v,-\ell\}\}$ and note $\text{Prob}(\mathcal{A}) = 1$ by $m(\varepsilon)$ continuous almost everywhere and $\text{Prob}(v = -\ell) = 0$. On $\mathcal{A}$, $\max\{X'\beta,-\ell\}$ is linear in β in a neighborhood of β_0 with derivative $1(v > -\ell)X$ at β_0. Also, on $\mathcal{A}$, $q(y - \alpha)$ is continuously differentiable at $\alpha = \max\{v,-\ell\}$ with derivative $-m(y - \alpha)$. Then by the chain rule, on $\mathcal{A}$, $q(y - \max\{X'\beta,-\ell\})$ is differentiable with derivative $-m(y - \max\{v,-\ell\}) \times 1(v > -\ell)X = -1(v > -\ell)m(\varepsilon)X = U(z)$. Because $E[U(z)] = E[E[U(z)|X]] = -E[1(v > -\ell)E[m(\varepsilon)|X]X] = 0$ and $E[U(z)U(z)'] = \Sigma$, condition (iii) of Lemma A.1 is satisfied.

https://doi.org/10.1017/S0266466601175018 Published online by Cambridge University Press

For condition (iv), note that as shown previously $\bar{q}(\alpha, X)$ is differentiable with derivative $-E[m(y^* - \alpha)|X] = -E[m(\varepsilon + v - \alpha)|X]$. Also, it follows as in the proof of Theorem 3.1 that $E[m(\varepsilon + v - \alpha)|X]$ is differentiable in α with

$$\partial E[m(\varepsilon + v - \alpha)|X]/\partial\alpha = \int m(\varepsilon + v) f_\varepsilon(\varepsilon + \alpha|X)\, d\varepsilon = \int m(\varepsilon + v - \alpha) f_\varepsilon(\varepsilon|X)\, d\varepsilon$$

$$= E[m(\varepsilon + v - \alpha)s(\varepsilon|X)|X],$$

where $s(\varepsilon|X) = f_\varepsilon(\varepsilon|X)/f(\varepsilon|X)$. Because $m(\varepsilon)$ is continuous almost everywhere, it follows that $\partial E[m(\varepsilon + v - \alpha)|X]/\partial\alpha$ is continuous, so that $\bar{q}(\alpha, X)$ is twice continuously differentiable and

$$\bar{q}_{\alpha\alpha}(\alpha, X) \leq Ca(X), \qquad a(X) = E[|s(\varepsilon|X)||X].$$

For notational convenience, let $b = -\ell$, $\tilde{v} = X'\beta$, $\tilde{r} = \max\{\tilde{v}, b\}$, $r = \max\{v, b\}$, and suppress the * on y^*. Then a second order mean-value expansion gives

$$E[q(z,\beta) - q(z,\beta_0)|X] = w(X)[\bar{q}(\tilde{r}, X) - \bar{q}(r, X)]$$

$$= w(X)[\bar{q}_\alpha(r, X)(\tilde{r} - r) + \bar{q}_{\alpha\alpha}(\bar{r}, X)(\tilde{r} - r)^2/2], \tag{A.7}$$

where $\bar{r}$ lies between $\tilde{r}$ and r. Now, note that

$$\tilde{r} - r = 1(\tilde{v} > b)1(v > b)(\tilde{v} - v) + 1(\tilde{v} > b)1(v \leq b)(\tilde{v} - b)$$
$$+ 1(\tilde{v} \leq b)1(v > b)(b - v)$$
$$= 1(v > b)(\tilde{v} - v) + 1(\tilde{v} > b)1(v \leq b)(\tilde{v} - b) + 1(\tilde{v} \leq b)1(v > b)(b - \tilde{v}).$$

Also, $1(v > b)\bar{q}_\alpha(r, X) = 1(v > b)\bar{q}_\alpha(v, X) = -1(v > b)E[m(\varepsilon)|X] = 0$, so that for $\bar{v}$ between b and v,

$$|\bar{q}_\alpha(r, X)(\tilde{r} - r)| = |\bar{q}_\alpha(b, X)1(\tilde{v} > b)1(v \leq b)(\tilde{v} - b)|$$
$$\leq |\bar{q}_{\alpha\alpha}(\bar{v}, X)|1(\tilde{v} > b)1(v \leq b)|\tilde{v} - b||b - v|$$
$$\leq Ca(X)1(\tilde{v} > b)1(v \leq b)|\tilde{v} - v|^2$$
$$\leq Ca(X)1(\tilde{v} > b)1(v \leq b)\|X\|^2\|\beta - \beta_0\|^2.$$

By $\text{Prob}(v = b) = 0$, $1(\tilde{v} > b)1(v \leq b) \to 0$ w.p.1. as $\beta \to \beta_0$, so by the dominated convergence theorem and existence of $E[a(X)\|X\|^2]$ (as implied by Assumption 3.1), $E[a(X)1(\tilde{v} > b)1(v \leq b)\|X\|^2] \to 0$. Therefore, it follows that as $\beta \to \beta_0$,

$$E[w(X)|\bar{q}_\alpha(r, X)(\tilde{r} - r)|] \leq CE[a(X)1(\tilde{v} > b)1(v \leq b)\|X\|^2]\|\beta - \beta_0\|^2$$
$$= o(\|\beta - \beta_0\|^2). \tag{A.8}$$

By similar reasoning,

$$E[w(X)|\bar{q}_{\alpha\alpha}(\bar{r}, X)1(\tilde{v} > b)1(v \leq b)(\tilde{v} - b)^2|] = o(\|\beta - \beta_0\|^2), \tag{A.9}$$

$$E[w(X)|\bar{q}_{\alpha\alpha}(\bar{r}, X)1(\tilde{v} \leq b)1(v > b)(b - \tilde{v})^2|] = o(\|\beta - \beta_0\|^2). \tag{A.10}$$

https://doi.org/10.1017/S0266466601175018 Published online by Cambridge University Press

Also, as $\beta \to \beta_0$, $\bar{q}_{\alpha\alpha}(\bar{r}, X) \to d(X)$, so by the dominated convergence theorem,

$$E[w(X)\bar{q}_{\alpha\alpha}(\bar{r}, X)1(v > b)(\tilde{v} - v)^2]$$
$$= (\beta - \beta_0)'E[w(X)\bar{q}_{\alpha\alpha}(\bar{r}, X)1(v > b)XX'](\beta - \beta_0)$$
$$= (\beta - \beta_0)'Q(\beta - \beta_0) + o(\|\beta - \beta_0\|^2). \quad \textbf{(A.11)}$$

Taking expectations of (A.7) and applying the triangle inequality to (A.8)–(A.11) then gives condition (iv). The first conclusion then follows from the conclusion to Lemma A.1. The second follows upon noting that if $w(X) = w^*(X)$, $Q = E[\{d(X)/E[m(\varepsilon)^2|X]\}d(X)1(v > -\ell)XX'] = E[\mathcal{S}\mathcal{S}']$ and $\Sigma = E[\{d(X)/E[m(\varepsilon)^2|X]\}^2 \times E[m(\varepsilon)^2|X]1(v > -\ell)XX'] = E[\mathcal{S}\mathcal{S}']$. ■

Proof of Theorem 3.3. Note that if $m(\varepsilon)$ is differentiable then by Assumptions 3.1 and 3.2, $1(X'\beta > -\ell)w(X)^2m(y - X'\beta)^2XX'$ and $1(X'\beta > -\ell)w(X)m_\varepsilon(y - X'\beta)XX'$ are continuous at β_0 with probability one and dominated by functions with finite expectation. Consistency of $\hat{Q}$ and $\hat{\Sigma}$ then follows by Lemma 4.3 of Newey and McFadden (1994). In the other case, let $Q(z, \beta) = w(X)XX'[q(y - \max\{X'\beta, -\ell\}) - q(y - \max\{v, -\ell\})]$, so that for $e_1 = (1, 0, \ldots, 0)'$ the first unit vector,

$$\hat{Q}_\delta = \sum_{i=1}^{n} [Q(z_i, \hat{\beta} + e_1\delta) + Q(z_i, \hat{\beta} - e_1\delta) - 2Q(z_i, \hat{\beta})]/(\delta^2 n).$$

Note that by (A.5), $\|Q(z, \tilde{\beta}) - Q(z, \beta)\| \leq Cw(X)\|X\|^3\|\tilde{\beta} - \beta\|$ and $E[w(X)^2\|X\|^6] < \infty$. Then by 2.7.11 of Van der Vaart and Wellner (1996), for any $\Delta \to 0$ and $\tilde{\beta}, \beta$ in a small enough neighborhood of β_0, for $\tilde{Q}(\beta) = E[Q(z_i, \beta)]$,

$$\sup_{\|\tilde{\beta} - \beta\| \leq \Delta} \left\| n^{-1/2} \sum_{i=1}^{n} \{Q(z_i, \tilde{\beta}) - Q(z_i, \beta) - [\tilde{Q}(\tilde{\beta}) - \tilde{Q}(\beta)]\} \right\| \xrightarrow{p} 0.$$

It follows by $n^{1/2}\delta^2 \to \infty$ that

$$\tilde{Q}_\delta = [\tilde{Q}(\hat{\beta} + e_1\delta) + \tilde{Q}(\hat{\beta} - e_1\delta) - 2\hat{Q}(\hat{\beta})]/\delta^2 = \hat{Q} + o_p(1). \quad \textbf{(A.12)}$$

Furthermore, it follows by (A.3) and the associated discussion that we can take

$$\hat{Q}(\beta) = E[XX'q(z, \beta)] = E[w(X)XX'\bar{q}(\max\{X'\beta, -\ell\}, X)]$$

for $q(z, \beta)$ as defined in (A.2). Let $\tilde{Q}_{jk}(\beta) = E[X_jX_kq(z, \beta)]$. It follows by the expansion in (A.7) that for $\bar{r}$ and r as defined there,

$$\tilde{Q}_{jk}(\beta) - \tilde{Q}_{jk}(\beta_0) = E[w(X)X_jX_k\{\bar{q}_\alpha(r, X)(\bar{r} - r) + \bar{q}_{\alpha\alpha}(\bar{r}, X)(\bar{r} - r)^2/2\}].$$

Noting that $E[w(X)a(X)\|X\|^4] \leq CE[w(X)a(X)^2\|X\|^2] + CE[w(X)\|X\|^6] \leq CE[s(\varepsilon|X)^2\|X\|^2] + C$ is finite by Assumption 3.1, it follows similarly to the proof of (iv) for Lemma A.2 that for $M_{jk} = E[w(X)d(X)1(v > -\ell)X_jX_kXX']/2$,

$$\tilde{Q}_{jk}(\beta) - \tilde{Q}_{jk}(\beta_0) = (\beta - \beta_0)'M_{jk}(\beta - \beta_0) + o(\|\beta - \beta_0\|^2).$$

https://doi.org/10.1017/S0266466601175018 Published online by Cambridge University Press

Therefore, for $\hat{\gamma} = \hat{\beta} - \beta_0$, and noting that $\|\hat{\gamma} + e_1\delta\|^2 = O_p(\delta^2)$,

$$[\tilde{Q}_{jk}(\hat{\beta} + e_1\delta) + \tilde{Q}_{jk}(\hat{\beta} - e_1\delta) - 2\tilde{Q}_{jk}(\hat{\beta})]/\delta^2$$
$$= [(\hat{\gamma} + e_1\delta)'M_{jk}(\hat{\gamma} + e_1\delta) + (\hat{\gamma} - e_1\delta)'M_{jk}(\hat{\gamma} - e_1\delta) - 2\hat{\gamma}'M_{jk}\hat{\gamma}]/\delta^2 + o_p(1)$$
$$= 2e_1'M_{jk}e_1 + o_p(1) = E[w(X)d(X)1(v > -\ell)X_jX_k] + o_p(1) = Q_{jk} + o_p(1).$$

The conclusion then follows by (A.12). ■

For the truncated case and $T = \{y^* : y^* > 0\}$, note that $E[\cdot] = E^*[\cdot|T] = E^*[1(T)(\cdot)]/P^*(T)$ and $E[\cdot|X] = E^*[\cdot|T, X] = E^*[1(T)(\cdot)|X]/E^*[1(T)|X]$. For $\delta \in \mathcal{T}^*$, where $\mathcal{T}^*$ is the latent tangent set specified in the proof of Lemma A.1, let $\mathcal{C}(\delta)$ be the transformation from the latent to the observed data given by $\mathcal{C}(\delta) = \delta - E^*[\delta|T] = \delta - E[\delta]$. By Lemma B.2 of Newey (1991), the tangent set for the observed data is the mean-square closure of $\{\mathcal{C}(\delta) : \delta \in \mathcal{T}^*\}$. Also, the score for β is $-Xs(\varepsilon, X) + E[Xs(\varepsilon, X)]$. The following lemmas will be used in the proof of Theorem 4.1.

LEMMA A.3. *If* $E[\delta(X)] = 0$ *and* $E[\delta(X)^2] < \infty$ *then* $\delta(X) \in \mathcal{T}$.

Proof. By the distribution of X unrestricted in the latent data, any $\delta^*(X)$ with $E^*[\delta^*(X)] = 0$ and $E^*[\delta^*(X)^2]$ finite is an element of $\mathcal{T}^*$. Consider any $\delta(X)$ satisfying the conditions of the lemma and let $\delta_M(X) = 1(|\delta(X)| \leq M)\delta(X)$. Then $\delta(X)$ is bounded and so has finite second moment. Hence $\delta_M^*(X) = \delta_M(X) - E^*[\delta_M(X)] \in \mathcal{T}^*$. Also, note that $\mathcal{C}(\delta_M^*) = \delta_M(X) - E[\delta_M(X)]$ converges in mean-square to $\delta(X) - E[\delta(X)] = \delta(X)$ by $\delta(X)$ having finite mean-square. Therefore, $\delta(X)$ is in the mean-square closure of $\mathcal{C}(\mathcal{T}^*)$. ■

LEMMA A.4. *If Assumptions 3.1 and A.1 are satisfied then* $-1(v \leq -k)Xs(\varepsilon, X) + E[1(v \leq -k)Xs(\varepsilon, X)] \in \mathcal{T}$.

Proof. Let $T^c = \{y^* \leq 0\}$ and define $\tilde{s}(\varepsilon, X) = 1(T)\{s(\varepsilon, X) - E^*[s(\varepsilon, X)|T, X]\}$. Note that $E[\tilde{s}(\varepsilon, X)|X] = E^*[\tilde{s}(\varepsilon, X)|T, X] = 0$. Also, $E^*[\|X\|^2\tilde{s}(\varepsilon, X)^2] = P^*(T) \times E[\|X\|^2\tilde{s}(\varepsilon, X)^2] = P^*(T)E[\|X\|^2E[\tilde{s}(\varepsilon, X)^2|X]] = P^*(T)E[\|X\|^2\mathrm{Var}(s(\varepsilon, X)|X)] \leq P^*(T)E[\|X\|^2E[s(\varepsilon, X)^2|X]] = P^*(T)E[\|X\|^2s(\varepsilon, X)^2] \leq E^*[\|X\|^2s(\varepsilon, X)^2] < \infty$. Consider $\bar{k} > k$ and the case where $\bar{k} \leq -v$. By the definition of k, $m(\varepsilon)$ will be nonzero on $\varepsilon < -v$. Therefore, $E^*[1(T^c)m(\varepsilon)^2|X] > 0$. Let $A(X) = E^*[(1(T), 1(T^c) \times m(\varepsilon))'(1, m(\varepsilon))|X]$. Note that by $E^*[m(\varepsilon)|X] = 0$, we have $E^*[1(T^c)m(\varepsilon)|X] = -E^*[1(T)m(\varepsilon)|X]$, and hence

$$\mathrm{Det}(A(X)) = E^*[1(T^c)m(\varepsilon)^2|X]E^*[1(T)|X] - E^*[1(T^c)m(\varepsilon)|X]E^*[1(T)m(\varepsilon)|X]$$
$$= E^*[1(T^c)m(\varepsilon)^2|X]E^*[1(T)|X] + E^*[1(T^c)m(\varepsilon)|X]^2 > 0 \qquad (v \leq -\bar{k}).$$

Let $D(X) = 1(v \leq -\bar{k})E^*[\tilde{s}(\varepsilon, X)(1, m(\varepsilon))|X]A(X)^{-1}$, M be some positive constant, and

$$\tilde{\delta}(\varepsilon, X) = -X \cdot 1(v \leq -\bar{k}) \cdot 1(\|D(X)\| \leq M) \cdot [\tilde{s}(\varepsilon, X) - D(X)(1(T), 1(T^c)m(\varepsilon))'].$$

By the definition of $D(X)$, it follows that $E^*[\tilde{\delta}(1, m(\varepsilon))|X] = 0$. Also, $E^*[\tilde{\delta}'\tilde{\delta}] < \infty$ by $E^*[m(\varepsilon)^2\|X\|^2] < \infty$ and $E^*[\|X\|^2\tilde{s}(\varepsilon, X)^2] < \infty$. By Lemma A.5 of Newey and Powell (1990), $\tilde{\delta}$ is in the tangent set $\mathcal{T}^*$ for the latent model. Note also that $1(T^c)$ is zero in

https://doi.org/10.1017/S0266466601175018 Published online by Cambridge University Press

the observed data and that $1(T) - E^*[1(T)|T] = 0$ in the observed data, so that $\mathcal{C}((1(T),1(T^c)m(\varepsilon))) = 0$. Also, $\mathcal{C}(\tilde{s}(\varepsilon,x)) = \tilde{s}(\varepsilon,X)$. Therefore,

$$\mathcal{C}(\tilde{\delta}) = 1(v \le -\bar{k})\cdot 1(\|D(X)\| \le M)\cdot(-X\tilde{s}(\varepsilon,X)).$$

It then follows as in the proof of Lemma A.1 that $1(v \le -k)[-X\tilde{s}(\varepsilon,X)] \in \mathcal{T}$. Furthermore, by Lemma A.3,

$$\begin{aligned}1(v \le -k)XE^*[s(\varepsilon,X)|T,X] - E[1(v \le -k)XE^*[s(\varepsilon,X)|T,X]]\\ = 1(v \le -k)XE^*[s(\varepsilon,X)|T,X] - E[1(v \le -k)Xs(\varepsilon,X)] \in \mathcal{T}.\end{aligned}$$

Then linearity of $\mathcal{T}$ and subtraction give the conclusion. ■

Proof of Theorem 4.1. Let $\delta(\varepsilon,X)$ be as defined in (A.1) with k replacing ℓ. As shown in the proof of Theorem 3.1, $E^*[\delta(\varepsilon,X)(1,m(\varepsilon))|X] = 0$ and $E^*[\|\delta(\varepsilon,X)\|^2] < \infty$, so that $\mathcal{C}(\delta) \in \mathcal{T}$. Note that by $m(\varepsilon)$ zero for $\varepsilon \le k$, $1(v > -k)m(\varepsilon) = 1(T)1(v > -k)m(\varepsilon)$, so that $\mathcal{C}(X\cdot 1(v > -k)\cdot w^*(X)m(\varepsilon)) = \mathcal{S}$. Therefore,

$$\mathcal{C}(\delta) = -1(v > -k)Xs(\varepsilon,X) + E[1(v > -k)Xs(\varepsilon,X)] - \mathcal{S} \in \mathcal{T}.$$

By Lemma A.4, linearity of $\mathcal{T}$, and summing it then follows that $S_\beta - \mathcal{S} \in \mathcal{T}$, i.e., $\mathcal{S} = S_\beta - t$, $t \in \mathcal{T}$. Also, note that $E^*[1(T)\mathcal{S}|X] = X \times w^*(X)E^*[1(T)1(v > -k)m(\varepsilon)|X] = X \times w^*(X)E^*[1(v > -k)m(3)|X] = X1(v > -k)E^*[m(\varepsilon)|X] = 0$. Therefore, for any δ satisfying $E^*[\delta(\varepsilon,X)m(\varepsilon)|X] = 0$.

$$\begin{aligned}E[\mathcal{S}'\mathcal{C}(\delta)] &= E^*[1(T)\mathcal{S}'\{\delta(\varepsilon,x) - E^*[\delta(\varepsilon,X)|T]\}]/P^*(T)\\ &= E^*[1(T)\mathcal{S}'\delta(\varepsilon,x)]/P^*(T)\\ &= E^*[E^*[\mathcal{S}'\delta(\varepsilon,x)|X]]/P^*(T)\\ &= E^*[X1(v > -k)E^*[m(\varepsilon)\delta(\varepsilon,X)|X]]/P^*(T) = 0,\end{aligned}$$

where the third equality follows by $1(T)\mathcal{S} = \mathcal{S}$ in the latent model. Then, because $\mathcal{T}$ is the mean-square closure of objects of the form $\mathcal{C}(\delta)$, it follows that $\mathcal{S}$ is orthogonal to the tangent set. ■

Proof of Theorem 4.2. Let $q(z,\beta) = w(X)[q(y - \max\{X'\beta,-k\}) - q(\varepsilon)]$. It follows as in the proof of Theorem 3.1 that for all y (positive or negative),

$$|q(z,\beta)| \le Cw(X)\|X\|, \qquad |q(z,\tilde{\beta}) - q(z,\beta)| \le Cw(X)\|X\|\|\tilde{\beta} - \beta\|.$$

It follows that $E^*[1(T^c)|q(z,\beta)|] \le CE^*[1(T^c)w(X)\|X\|]$ is finite. Also, because $q(\varepsilon)$ is constant for $\varepsilon \le k$ and $-\max\{X'\beta,-k\}) \le k$, we have $1(T^c)q(y - \max\{X'\beta,-\ell\}) = 1(T^c)q(k)$. Therefore,

$$\begin{aligned}E[q(z,\beta)] &= E^*[1(T)q(z,\beta)]/P^*(T)\\ &= E^*[q(z,\beta)]/P^*(T) - E^*[1(T^c)q(z,\beta)]/P^*(T)\\ &= E^*[q(z,\beta)]/P^*(T) - E^*[1(T^c)w(X)\{q(k) - q(\varepsilon)\}]/P^*(T).\end{aligned}$$

That is, $E[q(z,\beta)] = C_1E^*[q(z,\beta)] + C_2$ for constants $C_1 > 0$ and C_2. It follows that the maximum of $E[q(z,\beta)]$ will coincide with the maximum of $E^*[q(z,\beta)]$, which has a unique maximum at β_0, as shown in the proof of Theorem 3.1. Furthermore, because $q(\varepsilon)$ is constant below k, it is also linear below k, and so $E^*[q(z,\beta)]$ inherits all the properties $E[q(z,\beta)]$ has in the censored case. In particular, $E[q(z,\beta)] = E[q(z,\beta_0)] +$

https://doi.org/10.1017/S0266466601175018 Published online by Cambridge University Press

$(\beta - \beta_0)'Q(\beta - \beta_0)/2 + o(\|\beta - \beta_0\|^2)$, as in the proof of Theorem 3.1. Thus, conditions (i), (ii), and (iv) of Lemma A.2 are satisfied. Condition (iii) of Lemma A.2 also follows as in the proof of Theorem 3.1, so that the conclusion follows from the conclusion of Lemma A.2. ■

Proof of Theorem 4.3. This follows by extending the results for the censored case to the truncated case analogously to the proof of Theorem 4.2. ■

Proof of Theorem 5.1. By (11) the conditional moment restriction is satisfied in the latent data at $\beta = \tilde{\beta}_0$, so that $\sqrt{n}(\hat{\beta} - \tilde{\beta}_0) \xrightarrow{d} N(0, Q^{-1}\Sigma Q^{-1})$ by the conclusion of Theorem 3.2 and 4.2. Also, by independence, $d = d(X)$ and $E^*[m(\varepsilon - \mu)^2 | X] = \sigma^2$, so that in the censored case $Q = d \cdot M$ and $\Sigma = \sigma^2 \Lambda$. In the truncated case

$$Q = E[w(X)1(v > -k)d(X)XX'/\Pi(x)] = d \cdot M,$$

$$\Sigma = E[w(X)^2 1(v > -k)E^*[m(\varepsilon)^2 | X]XX'/\Pi(x)] = \sigma^2 \Lambda,$$

so that the first conclusion also holds. For the second conclusion, note that for $Y = w(X)1(v > -\ell)X$ and $U = 1(v > -\ell)X$, in the censored case,

$$\begin{aligned} M^{-1}\Lambda M^{-1} - (E[1(v > -\ell)XX'])^{-1} \\ &= (E[YU'])^{-1}E[YY'](E[UY'])^{-1} - (E[UU'])^{-1} \\ &= (E[YU'])^{-1}\{E[YY'] - E[YU'](E[UU'])^{-1}E[UY']\}(E[UY'])^{-1}, \end{aligned}$$

that is positive semidefinite by the matrix in the angle brackets being positive semidefinite. In the censored case the same result holds with $E[\cdot]$ replaced by $E[(\cdot)/\Pi(X)]$.

For the third conclusion, note that similarly to the proof of Theorem 3.3, $\hat{\alpha} = \sum_{i=1}^n 1(\hat{v}_i > -\ell)/n \xrightarrow{p} E[1(v > -\ell)] = \alpha$, whereas $\hat{\alpha} \cdot \hat{d}$ ($\hat{\alpha} \cdot \hat{d}_\delta$) is the upper left element of $\hat{Q}$ ($\hat{Q}_\delta$) and $\alpha \cdot d$ the upper left element of Q. Then $\hat{\alpha} \cdot \hat{d} \xrightarrow{p} \alpha \cdot d$ and $\hat{\alpha} \cdot \hat{d}_\delta \to \alpha \cdot d$ follow as in the proof of Theorem 3.3, except that here the sixth moments of X are not needed to exist by virtue of only requiring convergence of the upper left element of $\hat{Q}_\delta$. ■

Proof of Theorem 5.2. We consider first the case where x is constant. Let A satisfy the conditions of the theorem, so that WAG has full column rank. By W positive semidefinite, $G'A'WAG$ is nonsingular. For $\rho = \rho(z, 0_0)$ let $Y = G'A'WA\rho$ and $U = G'\Omega^- \rho$. Note that $E[YU'] = G'A'WA\Omega\Omega^- G = G'A'WA\Omega\Omega^- \Omega F = G'A'WAG$. Also note that $E[UU'] = G'\Omega^- \Omega\Omega^- G = F'\Omega\Omega^- \Omega\Omega^- \Omega F = F'\Omega F$ is nonsingular. Furthermore, the asymptotic variance of a GMM estimator based on $A^* = G'\Omega^-$ will be invariant to W (which has the same dimension as θ) and will be equal to $(E[UU'])^{-1}$. Therefore, the difference of asymptotic variances for a GMM estimator with A and W and that with A^* is $(G'A'WAG)^{-1}G'A'WA\Omega AWAG(G'A'WAG) - (G'\Omega^- G)^{-1} = (E[UY'])^{-1}E[YY'](E[YU'])^{-1} - (E[UU'])^{-1}$, which is positive semidefinite by the proof of Theorem 5.1.

Now consider the general case and, for notational convenience, suppress the x arguments. For the moment let $Y = A\rho$ and $U = F'\rho$. Then there exist a random vector X with nonsingular variance and a matrix B with full column rank such that $Y = BX$. Note that

$$\begin{aligned} E[AG] = E[A\Omega F] = E[YU'] = BE[XU'] = BE[XX']B'B(B'B)^{-1}(E[XX'])^{-1}E[XU'] \\ = E[A\Omega A']B(B'B)^{-1}(E[XX'])^{-1}E[XU']. \end{aligned}$$

https://doi.org/10.1017/S0266466601175018 Published online by Cambridge University Press

Therefore, $E[AG]$ is in the range of $E[A\Omega A']$. It follows then from the preceding proof that the variance minimizing choice of W is $(E[A\Omega A'])^-$ and the corresponding asymptotic variance is $\{E[G'A'](E[A\Omega A'])^-E[AG]\}^{-1}$. Now let $U = F'\rho$ and $Y = E[G'A'](E[A\Omega A'])^-A\rho = E[F'\Omega A'](E[A\Omega A'])^-A\rho$. Then $E[YY'] = E[YU']$, so that the difference of the asymptotic variances is

$$\begin{aligned}
&\{E[G'AP](E[A\Omega A'])^-E[AG]\}^{-1} - (E[G'\Omega^-G])^{-1} \\
&\quad = \{E[F'\Omega A'](E[A\Omega A'])^-E[A\Omega F]\}^{-1} - (E[F'\Omega\Omega^-\Omega F])^{-1} \\
&\quad = (E[YY'])^{-1} - (E[UU'])^{-1} \\
&\quad = (E[YU'])^{-1}\{E[YY'] - E[YU'](E[UU'])^{-1}E[UY']\}(E[YU'])^{-1}.
\end{aligned}$$

The conclusion then follows as in the proof of Theorem 5.1. ■

Proof of Theorem 5.3. It follows similarly to equation (12) that $\Omega(x) = I(x)'I(x)\Phi I(x)'I(x)$ and $G(x) = -I(x)'I(x)[D\cdot x', \mathrm{diag}(D)]$. In the truncated case $G(x)$ and $\Omega(x)$ are the same divided by $\Pi(x)$. By Φ nonsingular, $I(x)\Phi I(x)'$ is also. It follows that

$$I(x)[D\cdot x', \mathrm{diag}(D)] = I(x)\Phi I(x)'I(x)F(x),$$
$$F(x) = I(x)'[I(x)\Phi I(x)']^{-1}[Dx', \mathrm{diag}(D)].$$

Thus, $G(x) = \Omega(x)F(x)$. Then the conditions of Theorem 5.2 are satisfied. Also, because $I(x)I(x)'$ is an identity matrix, a generalized inverse of $\Omega(x)$ is $\Omega(x)^- = I(x)'[I(x)\Phi I(x)']^{-1}I(x)$ in the censored case and this quantity multiplied by $\Pi(x)$ in the truncated case. Then $G(x)'\Omega(x)^-$ is equal to the expression in the statement of the result. ■

Proof of Theorem 5.4. From the definition of $g(z,\theta)$ it follows that there is a matrix $K(x,\theta)$ such that $g(z,\theta) = K(x,\theta)\rho(z,\theta)$. Therefore, any GMM estimator based on $g(z,\theta)$ is also a GMM estimator based on $\rho(z,\theta)$ and hence is no more efficient than the one with $A^*(x)$. Let $\mathcal{N}$ be a neighborhood of θ_0 such that $\gamma_{j+1} + \not{k}_{j+1} < \gamma_{j-1} + \not{k}_{j-1}$ for all $\theta \in \mathcal{N}$. Let I_j denote a $j \times j$ identity matrix. Note that in the event $-\gamma_{j+1} - \not{k}_{j+1} \geq x'\beta > -\gamma_j - \not{k}_j$, $I(x,\theta) = [I_j, 0]$ and $g^k(z,\theta) = 0$ for all $k \neq j$. Then for the constant matrix $H_j = [I_j, 0]$, in the same event,

$$A^*(x)\rho(z,\theta) = [Dx', \mathrm{diag}(D)]'H_j'(H_j\Phi H_j')^{-1}H_j\rho(z,\theta) = B_j g_j(z,\theta),$$
$$B_j = \begin{bmatrix} D'H_j'(H_j\Phi H_j')^{-1}H_j \otimes [0, I] \\ \mathrm{diag}(D)H_j'(H_j\Phi H_j')^{-1}H_j(1,0) \end{bmatrix}.$$

Because for each of these intervals at most one of $g^j(z,\theta)$ is zero, it follows that for $\theta \in \mathcal{N}$, $A^*(x)\rho(z,\theta) = \sum_{j=1}^J B_j g^j(z,\theta) = Bg(z,\theta)$, for $B = [B_1, \ldots, B_J]$. It follows that the estimator with $A(x) = A^*(x)$ is a GMM estimator and hence is no more efficient than the best GMM estimator. ■

https://doi.org/10.1017/S0266466601175018 Published online by Cambridge University Press