GMM AND EMPIRICAL LIKELIHOOD WITH INCOMPLETE DATA

GAUTAM TRIPATHI

Preliminary. Comments welcome.

Abstract. In applied work economists often encounter data generating mechanisms that produce censored or truncated observations. These dgp’s induce a probability distribution on the realized observations that differs from the underlying distribution for which inference is to be made. If this dichotomy between the target and realized populations is not taken into account, statistical inference can be severely biased. In this paper, we show how to do efficient semiparametric inference in moment condition models by supplementing the incomplete observations with some additional data that is not subject to censoring or truncation.

1. Introduction

Many econometric models are expressed in terms of moment conditions. The typical setup is as follows. Let \( Z^* = (Z^{*(1)}, \ldots, Z^{*(d)}) \) be a random vector (following usual mathematical convention, “vector” means a column vector) in \( \mathbb{R}^d \) and \( \Theta \) a subset of \( \mathbb{R}^p \) such that

\[
H_0: \quad \mathbb{E}_{f^*}\{g(Z^*, \theta^*)\} = 0 \quad \text{for some } \theta^* \in \Theta, \tag{1.1}
\]

where \( g \) is a \( q \times 1 \) vector of known functions with \( q \geq p \) and \( f^* \) denotes the unknown density of \( Z^* \) with respect to a dominating measure \( \mu^* = \otimes_{i=1}^d \mu_i^* \). Since \( Z^* \) can have discrete components, the \( \mu_i^* \)’s need not be Lebesgue measures. The symbol \( \mathbb{E}_{f^*} \) indicates that expectation is with respect to the pdf \( f^* \).

Examples of (1.1) include linear and nonlinear regression models of the form \( Y^* = \psi(X^*, \theta^*) + \varepsilon \), where \( \psi \) is known up to \( \theta^* \) and the identifying assumption is that the error term is uncorrelated with the regressors; i.e., \( \mathbb{E}_{f^*}\{X^*\varepsilon\} = 0 \). Here \( g(Z^*, \theta^*) = X^*\{Y^* - \psi(X^*, \theta^*)\} \) and \( Z^* = (Y^*, X^*) \), where \( Y^* \) denotes the endogenous variable and \( X^* \) the vector of exogenous variables. Multivariate generalizations include linear and nonlinear simultaneous equations models where the error terms are uncorrelated with the exogenous variables. This class also contains instrumental variables (IV) models. Suppose we have a conditional mean restriction of the form \( \mathbb{E}_{Y^*|X^*}\{\tilde{g}(Y^*, X^*, \theta^*)|X^*\} = 0 \) w.p.1, where \( \tilde{g} \) is a \( k \times 1 \) vector of known functions, \( Y^* \) the vector of endogenous variables, and “w.p.1” abbreviates “with (target) probability one”. Then, letting \( A(X^*) \) denote a \( q \times k \) matrix of instruments, it yields unconditional moment restrictions of the form \( \mathbb{E}_{f^*}\{g(Z^*, \theta^*)\} = 0 \), where \( g(Z^*, \theta^*) = A(X^*)\tilde{g}(Y^*, X^*, \theta^*) \).

Date: This version: December 12, 2003.

Key words and phrases. Censored and truncated data, Empirical likelihood, GMM.

We thank Yuichi Kitamura and Arthur Lewbel for helpful suggestions.
If the target variable $Z^*$ is fully observed, these models are easily handled. But in many cases economists cannot fully observe $Z^*$. For instance, variables often get censored due to administrative reasons; e.g., government agencies routinely “top-code” income data before releasing it for public use. Similarly, studies investigating the length of unemployment spells can terminate prematurely due to financial constraints before all subjects have found employment. Censoring can also have behavioral origins; classic economic examples include female labor supply and household demand for durable goods. Biostatistics is another rich source of partially observed or incomplete data; see van der Laan and Robins (2003) for a comprehensive discussion and several examples.

Suppose that all coordinates of $Z^*$ are right-censored; i.e., instead of observing $Z^*$ we observe the random variable $Z = (Z^{(1)}, \ldots, Z^{(d)})_{d \times 1}$, where

$$Z^{(i)} = \begin{cases} Z^{*(i)} & \text{if } Z^{*(i)} < c^{(i)} \\ c^{(i)} & \text{otherwise} \end{cases}$$

for $i = 1, \ldots, d$

and $c = (c^{(1)}, \ldots, c^{(d)})$ is a $d \times 1$ vector of known constants. This allows for the possibility that some components of $Z^*$ may not be censored: If, say, the $i$th coordinate of $Z^*$ is not subject to the censoring mechanism, simply set $c^{(i)} = \infty$; if the $i$th and $j$th coordinates of $Z^*$, denoted by $Z^{*(i,j)}$, are not subject to censoring, then set $c^{(i,j)} = (\infty, \infty)$; etc.. Hence, in applications where the target variable can be decomposed into endogenous and exogenous parts as $Z^* = (Y^*, X^*)$, we can handle situations where only $Y^*$ is censored (pure endogenous censoring), or only $X^*$ is censored (pure exogenous censoring$^1$), or only some coordinates of both variables are censored. Left censoring of, say, the $i$th, $j$th, and $k$th coordinates can also be accommodated by replacing $Z^{*(i,j,k)}$ with $-Z^{*(i,j,k)}$ and $c^{(i,j,k)}$ with $-c^{(i,j,k)}$.

Let $S^*$ denote the survival function induced by $f^*$, and $\delta_{\{c\}}$ the Dirac measure at $\{c\}$; i.e.,

$$\delta_{\{c\}}(A) = \mathbb{I}\{c \in A\},$$

where $\mathbb{I}$ is the indicator function. It is implicitly understood that $\mu^*$ does not have an atom at $\{c\}$ because otherwise the censored and uncensored observations are indistinguishable. If $d = 1$, the density of $Z$ with respect to the dominating measure $\mu = \mu^* + \delta_{\{c\}}$ is given by

$$f(z) = f^*(z)\mathbb{I}\{z < c\} + S^*(c)\mathbb{I}\{z = c\}. \quad (1.2)$$

The density of $Z$ when it is vector valued is also straightforward to derive but requires some additional notation. So let $Z^{*-\{i,j,k\}}$ denote those coordinates of $Z^*$ that remain after the $i$th, $j$th, and $k$th ones have been deleted, $f^*_{-\{i,j,k\}}$ the joint density of $Z^{*-\{i,j,k\}}$, and $S^*_{i,j,k|-\{i,j,k\}}$ the conditional survival function induced by $f^*_{i,j,k|-\{i,j,k\}}$, where $f^*_{i,j,k|-\{i,j,k\}}$ denotes the conditional density of $Z^{*(i,j,k)}$ given $Z^{*-\{i,j,k\}}$. It is then easy to show that for $d > 1$ the density of $Z$ with respect to $\mu = \otimes_{i=1}^d \mu_i$, where

---

$^1$The term “exogenous” is, strictly speaking, an abuse of terminology since (1.1) does not involve any conditioning. We use it merely for the lack of a better alternative. The careful reader may want to substitute “censoring (resp. truncation) based on explanatory variables” for “exogenous censoring (resp. truncation)” whenever the latter is encountered.
\[ \mu_i = \mu_i^* + \delta_{(c(i))}, \]

is given by

\[
f(z) = f^*(z)\mathbb{I}\{z < c \} + \sum_{r=1}^{d-1} \sum_{i_1=1}^{d-r+1} \sum_{i_2=i_1+1}^{d-r+2} \cdots \sum_{i_r=i_{r-1}+1}^{d} S_{i_1, \ldots, i_r}^*(z^{(i_1, \ldots, i_r)} - (z^{(i_1, \ldots, i_r)})f^*(z^{(i_1, \ldots, i_r)}(z^{-(i_1, \ldots, i_r)})) \]
\[
\times \mathbb{I}\{z^{(i_1, \ldots, i_r)} = c^{(i_1, \ldots, i_r)}, z^{-(i_1, \ldots, i_r)} < c^{-(i_1, \ldots, i_r)}\} + S^*(c)\mathbb{I}\{z = c\}. \quad (1.3)
\]

Note that \( f \) has support \((-\infty, c^{(1)}] \times \ldots \times (-\infty, c^{(d)}]\) with an atom at \( \{c\} \). Also, equalities and inequalities involving vectors are assumed to hold coordinatewise.

**Example 1.1** (Censored mean). Suppose we want to estimate \( \theta^* = \mathbb{E}_{f^*}\{Z^*\} \), the mean of the target population. Since \( Z^* \) is censored from above, instead of a random sample from the target density \( f^* \) we have a random sample \( Z_1, \ldots, Z_n \) from the realized density \( f \) defined in (1.2) or (1.3). Therefore, the naive estimator \( \sum_{j=1}^n Z_j/n \) will not consistently estimate \( \theta^* \) because \( \sum_{j=1}^n Z_j/n \overset{P}{\neq} \mathbb{E}_f\{Z\} \) by the weak law of large numbers (WLLN)\(^2\) but

\[
\mathbb{E}_f\{Z\} = \begin{cases} 
\mathbb{E}_{f^*}\{Z^*\mathbb{I}\{Z^* < c\}\} + cS^*(c) & \text{if } d = 1 \\
\mathbb{E}_{f^*}\{Z^*\mathbb{I}\{Z^* < c\}\} + \sum_{r=1}^{d-1} \mathbb{E}_{f^*}\{Z_r^*\} + cS^*(c) & \text{if } d > 1,
\end{cases}
\]

where, for a function \( h(\cdot) \),

\[
h_r(Z^*) = \sum_{i_1=1}^{d-r+1} \sum_{i_2=i_1+1}^{d-r+2} \cdots \sum_{i_r=i_{r-1}+1}^{d} h(Z^*[i_1, \ldots, i_r])\mathbb{I}\{Z^{(i_1, \ldots, i_r)} > c^{(i_1, \ldots, i_r)}, Z^{-(i_1, \ldots, i_r)} < c^{-(i_1, \ldots, i_r)}\}
\]

denotes \( h \) evaluated at exactly \( r \) censored coordinates and the symbol \( Z^[i_1, \ldots, i_r] \) stands for \( Z^* \) with its \( i_1, \ldots, i_r \)th coordinates replaced by \( c^{(i_1)}, \ldots, c^{(i_r)} \), respectively, and the remaining coordinates unchanged; i.e., \( Z^[i_1, \ldots, i_r] = Z^*|_{Z^{(i_1, \ldots, i_r)} = c^{(i_1, \ldots, i_r)}} \). Since \( \mathbb{E}_f\{Z\} \neq \mathbb{E}_{f^*}\{Z^*\} \), the naive approach clearly fails to work. Intuitively, this happens because even though the parameter of interest is defined in terms of target density \( f^* \), censoring ensures that data comes from the realized density \( f \) instead of \( f^* \). Hence, any statistical inference using the realized observations is actually about \( f \) and not \( f^* \). Therefore, censored data should be handled carefully.

**Example 1.2** (Censored linear regression). Let \( Y^* = X^*\theta^* + \varepsilon \), where \( \mathbb{E}_{f^*}\{X^*\varepsilon\} = 0 \). Suppose that both \( Y^* \) and \( X^* \) are censored. Hence, instead of observing \( Z^* = (Y^*, X^*)_{(p+1) \times 1} \) from the target density \( f^* \), we observe \( Z = (Y, X) \) from the realized density \( f \) defined in (1.3). If we ignore censoring and simply regress \( Y \) on \( X \), then \( \theta^* \) cannot be consistently estimated by the least squares estimator \( \hat{\theta} = (\sum_{j=1}^n X_j X'_j)^{-1} \sum_{j=1}^n X_j Y_j \). To see this, observe that the probability limit of \( \hat{\theta} \) is given by

\[
(\mathbb{E}_f XX')^{-1}(\mathbb{E}_f XY) = (\mathbb{E}_{f^*}\{X^* X^{*'}\mathbb{I}\{Y^* < c^{(1)}, X^* < c^{-(1)}\}\} + \sum_{r=1}^{d-1} (X^* X^{*'})_r + c^{-1}(1)S^*(c))^{-1}
\]
\[
\times \mathbb{E}_{f^*}[X^* Y^*\mathbb{I}\{Y^* < c^{(1)}, X^* < c^{-(1)}\} + \sum_{r=1}^{d-1} (X^* Y^*)_r + c^{-1}(1)S^*(c)], \quad (1.4)
\]

\(^2\)Throughout the paper, all limits are taken as the sample size \( n \uparrow \infty \).
where \( d = p + 1 \) and, in the notation introduced in Example 1.1,

\[
(X^*Y^*)_r = \sum_{i_1=1}^{d-r+1} \sum_{i_2=i_1+1}^{d-r+2} \cdots \sum_{i_r=i_{r-1}+1}^{d} (X^*Y^*)_{(Y^*,X^*)(i_1,\ldots,i_r)=c(i_1,\ldots,i_r)}
\]

\[
\{Z_{(i_1,\ldots,i_r)} > c(i_1,\ldots,i_r), Z_{(i_1,\ldots,i_r)} \leq c(i_1,\ldots,i_r)\}
\] (1.5)

and \((X^*X^*)_r\) is obtained by replacing \(Y^*\) in (1.5) with \(X^*\). Hence, plim(\(\bar{\theta}\)) \(\neq (E_{f^*}X^*X^*)^{-1}(E_{f^*}X^*Y^*)\).

The special case of pure endogenous censoring, frequently called the tobit or limited dependent variable model, is obtained by letting \(c(1) = (\infty, \ldots, \infty)\) and using the convention that \(0 \cdot \infty = 0\). On doing so, (1.4) implies that

\[
\text{plim}(\bar{\theta}) = \theta^* - (E_{f^*}X^*X^*)^{-1}E_{f^*}[X^*(Y^* - c(1))]\{Y^* > c(1)\} \neq \theta^*,
\]

as is well known from tobit theory. However, a fact that does not seem to be as widely known is that the least squares estimator remains inconsistent even if censoring is purely exogenous. In particular, by letting \(c(1) = \infty\) in (1.4) it can be seen that

\[
\text{plim}(\bar{\theta}) = (E_{f^*}[X^*X^*])^{-1}E_{f^*}[X^*Y^*\{X^* < c(1)\}] + \sum_{r=1}^{d-1} (X^*Y^*)_r \neq \theta^*,
\]

where \((X^*Y^*)_r\) is now equal to

\[
\sum_{i_1=2}^{d-r+1} \sum_{i_2=i_1+1}^{d-r+2} \cdots \sum_{i_r=i_{r-1}+1}^{d} (X^*Y^*)_{(Y^*,X^*)(i_1,\ldots,i_r)=c(i_1,\ldots,i_r)}
\]

\[
\times \{X^*(i_1,\ldots,i_r-1) > c(i_1,\ldots,i_r), X^*-(i_1-1,\ldots,i_r-1) < c(1,i_1,\ldots,i_r)\}
\] (1.6)

and \((X^*X^*)_r\) follows by replacing \(Y^*\) in (1.6) with \(X^*\). Hence, pure exogenous censoring cannot be ignored here. In fact, pure exogenous censoring is not ignorable even if \(E_{f^*}\{X^*\varepsilon\} = 0\) is replaced by the stronger condition \(E_{Y^*X^*}\{\varepsilon|X^*\} = 0\) w.p.1\(^3\). However, as shown in Example 1.4, the situation changes if \(X^*\) is truncated instead of censored.

Sometimes, censoring is so severe that the target variable is completely unobserved outside a certain region. This phenomenon is called truncation. For e.g., in many job training programs subjects are allowed entry only if their household income falls below a certain level. In the physical sciences, truncation occurs when measuring instruments are not sensitive enough to detect events.

\[3\]To get some intuition behind this result, consider the case when \(d = 2\); i.e., when \(Y^* = X^*\theta^* + \varepsilon, X^*\) is scalar, and \(c = (\infty,c(2))_{2\times1}\). Then

\[
\text{plim}(\bar{\theta}) = \frac{E_{f^*}[XY]}{E_{f^*}[X^2]} = \frac{E_{f^*}[X^2Y^*\{X^* < c(2)\}] + [c(2)^2][I(X^* > c(2))]}{E_{f^*}[X^2\{X^* < c(2)\}] + [c(2)^2][I(X^* > c(2))]} = \frac{E_{f^*}[X^2\{X^* < c(2)\}] + [c(2)^2][I(X^* > c(2))]}{E_{f^*}[X^2\{X^* < c(2)\}] + [c(2)^2][I(X^* > c(2))]}\theta^*,
\]

where the last equality follows because \(E_{Y^*X^*}\{Y^*|X^*\} = X^*\theta^*\) w.p.1. Therefore, \text{plim}(\bar{\theta}) \neq \theta^* and \(\bar{\theta}\) remains inconsistent under pure exogenous censoring even when \(E_{Y^*X^*}\{\varepsilon|X^*\} = 0\) w.p.1.
below a certain threshold. If \( Z^* \) is a truncated random variable, then instead of observing \( Z^* \) we observe

\[
Z = \begin{cases} 
Z^* & \text{if } Z^* \in T \\
\text{unobserved} & \text{otherwise},
\end{cases}
\]

where \( T \) denotes a known region in \( \mathbb{R}^d \) such that \( Z^* \) lies in \( T \) with positive probability. In this case, the density of \( Z \) with respect to \( \mu^* \) is given by

\[
f(z) = \frac{f^*(z)\mathbb{1}\{z \in T\}}{\int_T f^*(z) d\mu^*}, \tag{1.7}
\]

Note that \( f \) has support \( T \). As before, we allow for the possibility that some coordinates of \( Z^* \) may not be truncated: In typical applications, \( T \) will be a rectangle of the form \( I_1 \times \ldots \times I_d \), where the \( I_j \)'s are known fixed intervals. If, say, \( Z^*_{(i,j,k)} \) are not truncated, then simply let \( I_i = I_j = I_k = \mathbb{R} \).

**Example 1.3** (Truncated mean). Again, suppose that we want to estimate the mean of the target population but now \( Z^* \) is truncated outside the region \( T \). Since \( \mathbb{E}_{f^*}\{Z\} = \mathbb{E}_{f^*}\{Z^*\mathbb{1}\{Z^* \in T\}\}/\int_T f^*(z) d\mu^* \), as in Example 1.1 the naive estimator is not consistent for \( \mathbb{E}_{f^*}\{Z^*\} \).

**Example 1.4** (Truncated linear regression). Consider the linear model of Example 1.2, but now suppose that instead of being censored, \( Z^* \) is truncated outside \( T = T_1 \times T_2 \). Since the probability limit of the least squares estimator is now given by

\[
\text{plim}(\hat{\theta}) = (\mathbb{E}_{f^*}\{X^*X^*\mathbb{1}\{Y^* \in T_1, X^* \in T_2\}\})^{-1}\mathbb{E}_{f^*}\{X^*Y^*\mathbb{1}\{Y^* \in T_1, X^* \in T_2\}\},
\]

it is immediate that \( \hat{\theta} \) is not consistent for \( \theta^* \). For pure endogenous truncation, \( T_2 = \mathbb{R}^p \). In this case,

\[
\text{plim}(\hat{\theta}) = (\mathbb{E}_{f^*}\{X^*X^*\mathbb{1}\{Y^* \in T_1\}\})^{-1}\mathbb{E}_{f^*}\{X^*Y^*\mathbb{1}\{Y^* \in T_1\}\} \neq \theta^*.
\]

Similarly, for pure exogenous truncation, \( T_1 = \mathbb{R} \). Hence,

\[
\text{plim}(\hat{\theta}) = (\mathbb{E}_{f^*}\{X^*X^*\mathbb{1}\{X^* \in T_2\}\})^{-1}\mathbb{E}_{f^*}\{X^*Y^*\mathbb{1}\{X^* \in T_2\}\} \neq \theta^*.
\]

Therefore, even pure exogenous truncation is not ignorable. But unlike Example 1.2, if the identifying assumption \( \mathbb{E}_{f^*}\{X^*\varepsilon\} = 0 \) is replaced by \( \mathbb{E}_{Y^*|X^*}\{\varepsilon|X^*\} = 0 \) w.p.1, then it is easy to see that ignoring pure exogenous truncation does not make the least squares estimator inconsistent. \( \square \)

Since in this paper we assume that \( f^* \) and, hence, \( f \) are completely unknown, censoring and truncation of \( Z^* \) create a fundamental identification problem. To see this, first note that since \( Z \) is the observed version of \( Z^* \), the realized density \( f \) is identified by definition. However, as is evident from (1.2), (1.3), and (1.7), the target density \( f^* \) cannot be expressed in terms of \( f \) for all \( z \in \mathbb{R}^d \). In other words, \( f^* \) cannot be fully recovered from \( f \); i.e., \( f^* \) is not identified. But, as noted in Example 1.1, this is problematic because our model (1.1) is specified in terms of \( f^* \). If there is no way of going from the realized density to the target density, then statistical inference about \( f^* \) and, hence, the target...
CDF $F^*(\xi) = \Pr_{f^*}\{Z^* \leq \xi\}$ is impossible\(^4\). So the first task is to overcome this lack of identification. Basically, this is done by requiring some additional information about $f^*$.

The information needed to ensure identification of $f^*$ can come in different guises. For instance, the classical solution is to simply assume that $f^*$ is known up to a finite dimensional parameter. Hence, identification of the finite dimensional parameter ensures that $f^*$ is also identified. There is a substantial literature on showing how to do maximum likelihood or “two-step” estimation with incomplete observations if a parametric model for $f^*$ is acceptable. See, for example, Hausman and Wise (1976, 1977), Heckman (1976, 1979), Dempster, Laird, and Rubin (1977), Goldberger (1981), Maddala (1983), Amemiya (1985), Hajivassiliou and Ruud (1994), Vella (1998), and the many references therein. If, however, it is not feasible to make such strong distributional assumptions, then other approaches that do not require a parametric specification for $f^*$ have to be used instead.

The real challenge in working with censored or truncated data arises when $f^*$ is completely unknown and we want to use these incomplete observations to do semiparametric inference. Most of the literature in this area seems to have focused on the tobit model, with different authors using different assumptions to identify the finite dimensional parameters. For e.g., Bhattacharya, Chernoff, and Yang (1983), Chamberlain (1986), Duncan (1986), Fernandez (1986), Horowitz (1986, 1988), Cossetti (1987, 1991), Tsui, Jewell, and Wu (1988), Moon (1989), and Honoré and Powell (1994) assume that $\varepsilon$ and $X^*$ are independent; Powell (1983, 1984, 1986a), Nawata (1990), Newey and Powell (1990), Buchinsky and Hahn (1998), Chen and Khan (2001), Khan and Powell (2001), and Hong and Tamer (2003) assume knowledge of some quantile of the conditional distribution of $\varepsilon$ given $X^*$; Powell (1986b), Lee (1993a, 1993b), and Newey (1988) assume that the distribution of $\varepsilon|X^*$ is symmetric. Miller and Halperin (1982) also discuss semiparametric estimation of the censored regression model. See Powell (1994) for generalizations. Newey (2001) has results about the kind of restrictions on the distribution of $\varepsilon|X^*$ that lead to $n^{1/2}$-consistent estimation of the regression coefficients. A different set of assumptions involving the existence of a “special regressor” that is continuously distributed, has large support, and satisfies a certain exclusion restriction, have been used by Khan and Lewbel (2003) to identify slope coefficients in a truncated regression model.

Although independence, symmetry, quantile, or special regressor restrictions are sufficient to identify finite dimensional parameters, they cannot ensure identification of the target density itself. This means that these restrictions are not strong enough to permit consistent estimation of the target CDF $F^*(\xi)$ at each $\xi \in \mathbb{R}^d$. In contrast, we ensure identification of $f^*$ and, hence, $F^*$ by assuming that in addition to the incomplete data we also have a set of supplementary observations that are

\(^4\)Since distribution functions characterize random variables, estimating the target CDF $F^*(\xi)$ at each $\xi \in \mathbb{R}^d$ determines the probabilistic behavior of $Z^*$. Efficient estimation of $F^*$ is also important if one wants to bootstrap from the target population. Brown and Newey (2002) note that when prior information about target population in the form of (1.1) is available, just using a consistent estimator of $F^*$ can lead to poor inference from the bootstrap. They recommend that in order to obtain an efficient bootstrapping procedure, resampling be done using $\hat{F}^*$, an efficient estimator of $F^*$ that incorporates restrictions imposed by the model. Estimating the realized CDF under (1.1) is also useful because comparing it with $\hat{F}^*$ can help reveal the extent of bias induced by censoring or truncation. Of course, we can also compare $\hat{F}^*$ with the empirical CDF of the observed data. But because the latter does not take the model into account, it will be less precise than the estimator of the realized CDF under (1.1).
unaffected by censoring or truncation. To be precise, we assume henceforth that the realized random variable $Z$ has density

$$f_e(z) = K_0 f^*(z) + (1 - K_0) f(z),$$  

(1.8)

where $K_0 \in (0, 1]$ is a known constant and, depending on whether $Z^*$ is censored or truncated, $f$ is given by (1.2) and (1.3), or (1.7). Loosely speaking, $K_0$ denotes the probability of sampling from the whole target population without subjecting the observations to censoring or truncation. As a practical example, consider a dataset of 100 observations, 90 of which belong to the original sample of censored observations and the remaining 10 constitute the uncensored supplementary sample. Hence, we can let $K_0 = 0.10$ and regard this dataset as a random sample from $f_e$. The no censoring or truncation case corresponds to setting $K_0 = 1$. In Sections 3.1 and 4.1 we show how this “enriched” density can be used to identify $f^*$. It also leads to simpler identification conditions for the finite dimensional parameters. E.g., we do not have to employ Chamberlain (1986) type “identification at infinity” arguments to identify $\theta^*$. The well known Rothenberg (1971) approach that works for identifying parametric and GMM models does the job here as well.

The idea of using additional samples to ensure identification in incomplete, biased sampling, or measurement error models has, of course, been recognized earlier. See, e.g., Titterington (1983), Titterington and Mill (1983), Vardi (1985), Carroll, Ruppert, and Stefanski (1995), Tripathi (2003), and the references therein. However, the use of a supplementary sample to do inference in general overidentified moment based models with censored or truncated data seems to be new to the literature, and the results obtained in this paper cannot be found in any of the references cited here. This approach is most reasonable in applications where censoring or truncation happens due to administrative reasons and the cost of collecting the supplementary sample is relatively small. For e.g., when dealing with top-coded data, the custodial agency in charge of the data may be persuaded to release a small supplementary sample where the observations are not subject to censoring. Sometimes, a small survey can be carried out to create the supplementary sample. This, for e.g., has been suggested by Manski and Lerman (1977) to deal with the problem of unknown aggregate shares in stratified sampling, and used by Ridder (1992) and Hirano, Imbens, Ridder, and Rubin (2001) in the context of missing data in panel data models.

Once questions regarding identification are settled, we can proceed with estimating and testing (1.1). Our inferential approach is based on the generalized method of moments (GMM) proposed by Hansen (1982) and empirical likelihood (EL) proposed by Owen (1988). We focus on GMM because

5If $Z^*$ is censored, then $f_e$ is a density with respect to the dominating measure $\mu$ introduced earlier and has an atom at $\{c\}$. On the other hand, if $Z^*$ is truncated then $f_e$ is a density with respect to $\mu^*$.

6Although (1.8) essentially shuts down censoring or truncation for part of the observation, our identification approach is not related to the identification at infinity argument because the latter, unlike (1.8), relies upon manipulating the supports of underlying random variables to achieve identification.

7Nevo (2003) does allow for $Z^*$ to be partially observed, but restricts attention to the case where the moment condition model is just identified. He also assumes that the probability of observing $Z^*$ is known up to a finite dimensional parameter. In contrast, we allow (1.1) to be overidentified and let $Pr_{f^*} \{Z^* < c\}$ (if $Z^*$ is censored) or $Pr_{f^*} \{Z^* \in T\}$ (if $Z^*$ is truncated) be fully unknown. Furthermore, although Nevo also uses a supplementary sample to estimate finite dimensional parameters, unlike us he does not consider identification and estimation of the target or realized cdf’s.
its unifying approach and wide applicability has made it the method of choice for estimating and testing nonlinear economic models. Its availability in canned software packages has also added to its popularity with applied economists. An excellent exposition on GMM can be found in Newey and McFadden (1994), Hayashi (2000), and Wooldridge (2001). We also look at EL because it has lately begun to emerge as a serious contender to GMM. See, e.g., Qin and Lawless (1994), Imbens (1997), Kitamura (1997), Smith (1997), Owen (2001), the 2002 special issue of the Journal of Business and Economics Statistics, and Newey and Smith (2003).

Provided a supplementary sample is available, we show that standard GMM and EL based inference for (1.1) using censored or truncated data goes through without imposing parametric, independence, symmetry, quantile, or special regressor restrictions on the underlying random variables and without any major computational problems. Also, there is no need to worry about issues relating to heteroscedasticity (which are a major concern for censored and truncated data) because GMM and EL automatically produce correct standard errors. Furthermore, unlike quantile restriction models, we do not have to restrict attention to scalar error terms nor do we need to use any nonparametric smoothing procedures to estimate asymptotic variances: Extension to the multivariate case is straightforward and the usual analogy principle that delivers standard errors for GMM or EL works here as well. The statistical procedures developed in this paper are applicable to a large class of potentially overidentified models which nests linear regression as a special case: e.g., the ability to handle IV models allows us to do semiparametric inference in Box-Cox type models using incomplete data without imposing parametric or quantile restrictions. Our treatment is general enough to handle censoring and truncation of both exogenous and endogenous variables. It also allows us to test the validity of the overidentifying restrictions in (1.1) with censored or truncated data.

The rest of the paper is organized as follows. Section 2 lists the assumptions under which our large sample results hold. Section 3 shows how we can combine censored data with a supplementary sample to do efficient moment based inference, and Section 4 does the same with truncated data. Section 5 concludes.

2. Regularity conditions

The following assumption is maintained throughout the paper. It ensures that the GMM and EL estimators described subsequently are consistent and asymptotically normal. Let \( \| \cdot \| \) denote the Euclidean norm, \( \frac{\partial g(z, \theta)}{\partial \theta} \) the \( q \times p \) Jacobian matrix of partial derivatives of \( g(z, \theta) \) with respect to \( \theta \), and \( B(\theta, \delta) \) an open ball with center \( \theta \) and radius \( \delta \).

**Assumption 2.1.** (i) \( \Theta \) is compact; (ii) \( \theta^* \in \text{int}(\Theta) \) is the unique root of (1.1); (iii) \( g(z, \theta) \) is continuous on \( \Theta \) w.p.1; There exist \( \eta > 0 \) and \( \delta > 0 \) such that: (iv) \( \mathbb{E}_{f^*} \{ \sup_{\theta \in \Theta} \| g(z, \theta) \|^{2(1+\eta)} \} < \infty \); (v) \( g(z, \theta) \) is twice continuously differentiable on \( B(\theta^*, \delta) \) w.p.1; (vi) \( \mathbb{E}_{f^*} \{ \sup_{\theta \in B(\theta^*, \delta)} \| \frac{\partial g(z, \theta)}{\partial \theta} \| \} < \infty \); (vii) \( \mathbb{E}_{f^*} \{ \sup_{\theta \in B(\theta^*, \delta)} \| \frac{\partial^2 g(z, \theta)}{\partial \theta_i \partial \theta_j} \| \} < \infty \) for \( i = 1, \ldots, q \) and \( j, k = 1, \ldots, p \).

(i)–(vii) are used to prove the consistency and asymptotic normality of EL estimators as in Kitamura (1997) and Qin and Lawless (1994), although GMM estimators can be shown to be consistent and asymptotically normal under slightly weaker conditions. See, e.g., Newey and McFadden (1994).
Inference with censored data

3.1. Identification. Let \( a(z) = K_0 + (1 - K_0)\mathbb{I}\{z < c\} \) and note that, by (1.2) and (1.8),

\[
f^*(z) = a^{-1}(z)[f_c(z) - (1 - K_0)S^*(c)\mathbb{I}\{z = c\}]\]

for \( d = 1 \). Similarly, we can use (1.3) and (1.8) to show that for \( d > 1 \),

\[
f^*(z) = a^{-1}(z)[f_c(z) - \frac{1}{K_0}\sum_{r=1}^{d-1} \sum_{i_1=1}^{d-r+1} \sum_{i_2=i_1+1}^{d-r+2} \cdots \sum_{i_r=i_{r-1}+1}^{d} S^*_{z_{(i_1,\ldots,i_r)}=(i_1,\ldots,i_r)}(c^{(i_1,\ldots,i_r)})f^*_{-z_{(i_1,\ldots,i_r)}=c^{(i_1,\ldots,i_r)}}(z^{-(i_1,\ldots,i_r)})\mathbb{I}\{z^{(i_1,\ldots,i_r)} = c^{(i_1,\ldots,i_r)}, z^{-(i_1,\ldots,i_r)} < c^{-(i_1,\ldots,i_r)}\} - (1 - K_0)S^*(c)\mathbb{I}\{z = c\})\]

But since \( \mathbb{E}_{f_c}[\mathbb{I}\{Z = c\}] = (1 - K_0)S^*(c) \) and

\[
\int_{\{z^{(i_1,\ldots,i_r)} > c^{(i_1,\ldots,i_r)}\}} f_c(z)\,d\mu(z^{(i_1,\ldots,i_r)}) = K_0S^*_{z_{(i_1,\ldots,i_r)}=(i_1,\ldots,i_r)}(c^{(i_1,\ldots,i_r)})f^*_{-z_{(i_1,\ldots,i_r)}=c^{(i_1,\ldots,i_r)}}(z^{-(i_1,\ldots,i_r)}),
\]

we can write

\[
f^*(z) = a^{-1}(z)[f_c(z) - \mathbb{E}_{f_c}[\mathbb{I}\{Z = c\}]\mathbb{I}\{z = c\}] \quad (3.1)
\]

for \( d = 1 \), and

\[
f^*(z) = a^{-1}(z)[f_c(z) - \frac{1}{K_0}\sum_{r=1}^{d-1} \sum_{i_1=1}^{d-r+1} \sum_{i_2=i_1+1}^{d-r+2} \cdots \sum_{i_r=i_{r-1}+1}^{d} \int_{\{z^{(i_1,\ldots,i_r)} > c^{(i_1,\ldots,i_r)}\}} f_c(z)\,d\mu(z^{(i_1,\ldots,i_r)})\mathbb{I}\{z^{(i_1,\ldots,i_r)} = c^{(i_1,\ldots,i_r)}, z^{-(i_1,\ldots,i_r)} < c^{-(i_1,\ldots,i_r)}\} - \mathbb{E}_{f_c}[\mathbb{I}\{Z = c\}]\mathbb{I}\{z = c\}] \quad (3.2)
\]

for \( d > 1 \). Thus, by (3.1) and (3.2), \( f^* \) is identified since it can be expressed in terms of \( f_c \) alone.

Once the target density is identified, a sufficient condition for local identification of \( \mu^* \) is easy to formulate: Since \( \mu^* \) does not have an atom at \( \{c\} \), the Radon-Nikodym derivative \( d\mu^*/d\mu = 1 \). Hence, using (3.1) and (3.2), a little algebra shows that \( \mathbb{E}_{f^*}[g(Z^*, \theta^*)] = \mathbb{E}_{f_c}[\rho(Z, \theta^*)] \), where

\[
\rho(Z, \theta) = \begin{cases} 
\frac{g(Z, \theta)}{a(Z)} - \frac{\mathbb{I}\{Z = c\}}{K_0}g(c, \theta) & \text{if } d = 1 \\
\frac{g(Z, \theta)}{a(Z)} - \frac{(1 - K_0)}{K_0} \sum_{r=1}^{d-1} g_r(Z, \theta) - \frac{\mathbb{I}\{Z = c\}}{K_0}g(c, \theta) & \text{if } d > 1
\end{cases}
\]

and \( g_r(Z, \theta) \) is defined according to the notation introduced earlier in Example 1.1. Therefore,

\[
\mathbb{E}_{f^*}[g(Z^*, \theta^*)] = 0 \iff \mathbb{E}_{f_c}[\rho(Z, \theta^*)] = 0, \quad (3.3)
\]
and inference can be based on the latter moment condition without any loss of information. Hence, following Rothenberg (1971), it is straightforward to show that \( \theta^* \) is locally identified if the matrix \( D'\Omega^{-1}D \) exists and is nonsingular, where \( D = \mathbb{E}_{f_c} \{ \frac{\partial g(Z, \theta^\prime)}{\partial \theta} \} \) and \( \Omega = \mathbb{E}_{f_c} \{ \rho(Z, \theta^*)\rho'(Z, \theta^*) \} \).

3.2. Efficient estimation. Let \( Z_1, \ldots, Z_n \) denote iid random vectors with \( \mu \)-density \( f_c \). Using (3.3), we can estimate \( \theta^* \) by two-step GMM as follows. Define \( \hat{\rho}(\theta) = \sum_{j=1}^n \rho(Z_j, \theta)/n \). In the first step, use the preliminary estimator \( \tilde{\theta} = \arg\min_{\theta \in \Theta} \hat{\rho}'(\theta)\hat{\rho}(\theta) \) to construct the optimal weight matrix \( \tilde{W} = \sum_{j=1}^n \rho(Z_j, \tilde{\theta})\rho'(Z_j, \tilde{\theta})/n \). In the second step, obtain the GMM estimator \( \hat{\theta}_{gmm} = \arg\min_{\theta \in \Theta} \text{GMM}(\theta) \), where the objective function \( \text{GMM}(\theta) = \hat{\rho}'(\theta)\tilde{W}^{-1}\hat{\rho}(\theta) \). By standard GMM theory (see, e.g., Newey and McFadden (1994)), \( \hat{\theta}_{gmm} \) is consistent and asymptotically normal with limiting distribution given in Theorem 3.1. Since it is well known that optimally weighted GMM estimators are efficient as the sample size goes to infinity (see, e.g., Chamberlain (1987) or Severini and Tripathi (2001)), it follows that \( \hat{\theta}_{gmm} \) is asymptotically efficient.

**Theorem 3.1.** \( n^{1/2}(\hat{\theta}_{gmm} - \theta^*) \overset{d}{\to} N(0, (D'\Omega^{-1}D)^{-1}) \).

Next, we show how to estimate \( \theta^* \) by EL. Let \( p_j \) denote the probability mass placed at \( Z_j \) by a discrete distribution that has support on the realized observations \( Z_1, \ldots, Z_n \). For a fixed \( \theta \), concentrate out the \( p_j \)'s by solving the nonparametric maximum likelihood problem

\[
\max_{p_1, \ldots, p_n} \sum_{j=1}^n \log p_j \quad \text{s.t.} \quad p_j \geq 0, \quad \sum_{j=1}^n p_j = 1, \quad \sum_{j=1}^n \rho(Z_j, \theta)p_j = 0. \tag{3.4}
\]

If the convex hull of \( \rho(Z_1, \theta), \ldots, \rho(Z_n, \theta) \) contains the origin, then (3.4) can be solved by Lagrange multipliers. It is easy to verify that the solution to (3.4) is given by

\[
\hat{\rho}_j(\theta) = \frac{1}{n \{1 + \lambda'(\theta)\rho(Z_j, \theta)\}} \quad \text{for} \quad j = 1, \ldots, n, \tag{3.5}
\]

where the multiplier \( \lambda(\theta) \) satisfies

\[
\sum_{j=1}^n \rho(Z_j, \theta) \frac{1}{1 + \lambda'(\theta)\rho(Z_j, \theta)} = 0. \tag{3.6}
\]

Letting \( \text{EL}(\theta) = \sum_{j=1}^n \log \hat{\rho}_j(\theta) = -\sum_{j=1}^n \log \{1 + \lambda'(\theta)\rho(Z_j, \theta)\} - n \log n \), define the empirical likelihood estimator of \( \theta^* \) as \( \hat{\theta}_{el} = \arg\max_{\theta \in \Theta} \text{EL}(\theta) \). Since \( \mathbb{E}_{f_c} \{ \rho(Z, \theta^*) \} = 0 \), consistency of \( \hat{\theta}_{el} \) follows from Kitamura (1997, Theorem 1). By standard EL theory as in Qin and Lawless (1994), \( n^{1/2}(\hat{\theta}_{el} - \theta^*) \) and \( n^{1/2}(\hat{\theta}_{gmm} - \theta^*) \) have the same asymptotic distribution. Hence, the EL estimator is also asymptotically efficient. Note that in finite samples the GMM and EL estimators \( \hat{\theta}_{gmm} \) and \( \hat{\theta}_{el} \) will be different. However, it is easy to see that the two will coincide if \( \theta^* \) is just identified; i.e., \( q = p \).

Although GMM and EL based statistical inference is asymptotically first order equivalent, recent research by Newey and Smith (2003) has shown that under certain regularity conditions EL has better second order properties than GMM. For instance, they show that, unlike GMM, the second order bias of EL does not depend upon the number of moment conditions. This makes EL very attractive for estimating models with large \( q \) (e.g., panel data models with long time dimension) where GMM is known to perform poorly in small samples. As far as testing is concerned, Kitamura (2001)
has demonstrated that an EL based specification test for (1.1) is asymptotically optimal in terms of a Hoeffding type large deviation criterion. Brown and Newey (2002) show that EL is intimately connected with the theory of efficient bootstrapping. Another advantage of EL is that efficient estimators of the realized cdf \( F_m \) and the target cdf \( F^*(\xi) \) are easily obtained. So let \( \hat{F}(\xi) = \sum_{j=1}^n \hat{p}_j(\hat{\theta}_d)I\{Z_j \leq \xi\} \), and \( F^*(\xi) = \sum_{j=1}^n \hat{p}_j(\hat{\theta}_d)I\{Z_j \leq \xi\}/a(Z_j) \), where \( \hat{p}_j(\hat{\theta}_d) \) denote the EL probability weights in (3.5) evaluated at \( \hat{\theta}_d \). Then, by Qin and Lawless (1994, Theorem 1),

**Theorem 3.2.** \( n^{1/2}\{\hat{F}(\xi) - F_e(\xi)\} \) is asymptotically normal with mean zero and variance

\[
F_e(\xi)\{1 - F_e(\xi)\} - \mathbb{E}_{f_e}\{\rho'(Z, \theta^*)\}I\{Z \leq \xi\}\{(\Omega^{-1} - \Omega^{-1}D(D'\Omega^{-1}D)^{-1}D'\Omega^{-1})\mathbb{E}_{f_e}\{\rho(Z, \theta^*)\}I\{Z \leq \xi\}\}.
\]

Next, letting

\[
\beta(Z) = \frac{\mathbb{I}\{Z \leq \xi\}}{a(Z)} - \mathbb{I}\{d > 1\}\left(\frac{1 - K_0}{K_0}\sum_{r=1}^{d-1} \sum_{i_1=1}^{d-r+1} \sum_{i_2=i_1+1}^{d-r+2} \ldots \sum_{i_r=1}^{d} \mathbb{I}\{Z[i_1, \ldots, i_r] \leq \xi\} \right)
\]

\[
\times \mathbb{I}\{Z^{(i_1, \ldots, i_r)} > c^{(i_1, \ldots, i_r)}, Z^{-(i_1, \ldots, i_r)} < c^{-(i_1, \ldots, i_r)}\} - \frac{\mathbb{I}\{c \leq \xi\}}{K_0} \mathbb{I}\{Z = c\},
\]

a linearization argument reveals that\(^8\)

**Theorem 3.3.** \( n^{1/2}\{\hat{F}^*(\xi) - F^*(\xi)\} \) is asymptotically normal with mean zero and variance

\[
\text{var}_{f_e}\{\beta(Z)\} - \mathbb{E}_{f_e}\{\rho'(Z, \theta^*)\beta(Z)\}\{(\Omega^{-1} - \Omega^{-1}D(D'\Omega^{-1}D)^{-1}D'\Omega^{-1})\mathbb{E}_{f_e}\{\rho(Z, \theta^*)\}\beta(Z)\}.
\]

Theorems 3.2 and 3.3 show that imposing the overidentified model leads to an efficiency gain in estimating \( F_e \) and \( F^* \). Asymptotic optimality of EL implies that \( \hat{F}(\xi) \) and \( \hat{F}^*(\xi) \) are also asymptotically efficient. For the remainder of Section 3, let \( \hat{\theta} \) denote the GMM or EL estimator of \( \theta^* \). The asymptotic variances of \( \hat{\theta} \), \( \hat{F}(\xi) \), and \( \hat{F}^*(\xi) \) can be estimated in the obvious manner by replacing \( D \) and \( \Omega \) with consistent estimators \( \hat{D} = \sum_{j=1}^n \frac{\partial p(Z_j, \hat{\theta})}{\partial \theta} / n \) and \( \hat{\Omega} = \sum_{j=1}^n \rho(Z_j, \hat{\theta}) \rho'(Z_j, \hat{\theta}) / n \), respectively.

**3.3. Examples.** Before concluding Section 3, let us look at some useful examples.

**Example 3.1** (Example 1.1 contd.). Since there are no overidentifying restrictions and

\[
\rho(Z, \theta) = \begin{cases} 
\frac{Z - \theta}{a(Z)} - \frac{\mathbb{I}\{Z = c\}}{K_0} & \text{if } d = 1 \\
\frac{Z - \theta}{a(Z)} - \frac{(1 - K_0)}{K_0} \sum_{r=1}^{d-1} (Z - \theta)_r - \frac{\mathbb{I}\{Z = c\}}{K_0} & \text{if } d > 1,
\end{cases}
\]

\( \hat{\theta} \) is obtained by setting the sample analog of \( \mathbb{E}_{f_e}\{\rho(Z, \theta)\} \) to zero; i.e.,

\[
\hat{\theta} = \begin{cases} 
\sum_{j=1}^n \mathbb{I}\{Z_j - c\}/a(Z_j) & \text{if } d = 1 \\
\sum_{j=1}^n \mathbb{I}\{Z_j - c\}/a(Z_j) - K_0^{-2} (1 - K_0) \sum_{j=1}^n \sum_{r=1}^{d-1} Z_{j,r} & \text{if } d > 1,
\end{cases}
\]

where \( 1_r(Z) = \sum_{i_1=1}^{d-r+1} \sum_{i_2=i_1+1}^{d-r+2} \ldots \sum_{i_r=1}^{d} \mathbb{I}\{Z^{(i_1, \ldots, i_r)} > c^{(i_1, \ldots, i_r)}, Z^{-(i_1, \ldots, i_r)} < c^{-(i_1, \ldots, i_r)}\} \). By the WLLN, it is straightforward to show that \( \hat{\theta} \) is consistent for \( \mathbb{E}_{f^*}\{Z^*\} \). Furthermore, by Theorem 3.1 or by a direct application of the central limit theorem (CLT), it can be seen that \( n^{1/2}(\hat{\theta} - \theta^*) \) converges

---

\(^8\)The proof is omitted since it is very similar to the proof of Theorem 4.10 in Tripathi (2003).
in distribution to a normal random variable with mean zero and variance $E_{f_e}\{\rho(Z, \theta^*)\rho'(Z, \theta^*)\}$. Note that since here $q = p$, i.e., there are no overidentifying restrictions, the EL weights $\hat{\rho}_j(\theta) = 1/n$ for $j = 1, \ldots, n$. Hence, in this example (and the next one), $\hat{F}(\xi) = n^{-1}\sum_{j=1}^n 1\{Z_j \leq \xi\}$ is just the empirical cdf of the realized observations and $\hat{F}^*(\xi) = n^{-1}\sum_{j=1}^n 1\{Z_j \leq \xi\}/a(Z_j)$. By Theorem 3.2 or by the CLT, the asymptotic variance of $n^{1/2}\{\hat{F}(\xi) - F_e(\xi)\}$ is simply $F_e(\xi)\{1 - F_e(\xi)\}$. Similarly, by Theorem 3.3, $n^{1/2}\{\hat{F}^*(\xi) - F^*(\xi)\}$ has asymptotic variance $\text{var}_{f_e}\{b_\xi(Z)\}$. □

Example 3.2 (Example 1.2 contd.). Here $d > 1$ and $g(Z, \theta) = X(Y - X')$. Hence,

$$\rho(Z, \theta) = \frac{X(Y - X')}{{a(Y, X)}},$$

where

$$a(Y, X) = K_0 + (1 - K_0)\{Y < c(1), X < c(1)\},$$

and $(XY)_r, (XX)_r$ are as defined in (1.5). Since there are no overidentifying restrictions, $\hat{\theta}$ is obtained by solving $\sum_{j=1}^n \rho(Z_j, \hat{\theta}) = 0$; i.e.,

$$\hat{\theta} = \left(\sum_{j=1}^n \frac{X_jX'_j}{{a(Y_j, X_j)}} - \frac{(1 - K_0)}{K_0^2} \sum_{r=1}^{d-1} (X_jX'_j)_r - K_0^{-1}\{Y_j = c(1), X_j = c(1)\}c(1)c(1)\right)^{-1}$$

and

$$\times \left(\sum_{j=1}^n \frac{X_jY_j}{{a(Y_j, X_j)}} - \frac{(1 - K_0)}{K_0^2} \sum_{r=1}^{d-1} (X_jY_j)_r - K_0^{-1}\{Y_j = c(1), X_j = c(1)\}c(1)c(1)\right). \quad (3.7)$$

By Theorem 3.1, $n^{1/2}(\hat{\theta} - \theta^*)$ is asymptotically normal with mean zero and variance $D^{-1}\Omega D^{-1}$, where

$$D = -E_{f_e}\{\rho(Z, \theta^*)\rho'(Z, \theta^*)\}$$

$$\Omega = E_{f_e}\{\rho(Z, \theta^*)\rho'(Z, \theta^*)\}^{-1}$$

$$\{X(X'X')_r - K_0^{-1}\{Y = c(1), X = c(1)\}c(1)c(1)'\}. \quad (3.8)$$

If censoring is purely endogenous, i.e., $c(1) = (\infty, \ldots, \infty)$, then (3.7) implies that

$$\hat{\theta} = \left(\sum_{j=1}^n \frac{X_jX'_j}{{K_0 + (1 - K_0)\{Y_j < c(1)\}}} - \frac{(1 - K_0)}{K_0^2}\{Y_j > c(1)\}c(1)c(1)'\right)^{-1}$$

$$\times \left(\sum_{j=1}^n \frac{X_jY_j}{{K_0 + (1 - K_0)\{Y_j < c(1)\}}} - \frac{(1 - K_0)}{K_0^2}\{Y_j > c(1)\}X_jc(1)\right).$$

The asymptotic variance of $\hat{\theta}$ can be obtained by setting $c(1) = (\infty, \ldots, \infty)$ in (3.8).

Similarly, under pure exogenous censoring, i.e., $c(1) = \infty$,

$$\hat{\theta} = \left(\sum_{j=1}^n \frac{X_jX'_j}{{K_0 + (1 - K_0)\{X_j < c(1)\}}} - K_0^{-2}(1 - K_0)\sum_{r=1}^{d-1} (X_jX'_j)_r\right)^{-1}$$

$$\times \left(\sum_{j=1}^n \frac{X_jY_j}{{K_0 + (1 - K_0)\{X_j < c(1)\}}} - K_0^{-2}(1 - K_0)\sum_{r=1}^{d-1} (X_jY_j)_r\right),$$

where $(X_jY_j)_r$ and $(X_jX'_j)_r$ are as defined in (1.6). The asymptotic variance of $\hat{\theta}$ is given by letting $c(1) = \infty$ in (3.8). □
Example 3.3 (Censored linear regression with endogenous regressors). Let \( Y^* = X^*\theta^* + \varepsilon \) such that some or all of the regressors are correlated with \( \varepsilon \). We have a \( q \times 1 \) vector of instrumental variables \( W^* \) that are uncorrelated with \( \varepsilon \); i.e., \( W^* \) satisfies the moment condition \( \mathbb{E}_f \{ W^*\varepsilon \} = 0 \). Let \( W^* = (X_1^*, \tilde{W}^*) \), where \( X_1^* \) denotes the \( p_1 \times 1 \) vector of exogenous coordinates of \( X^* \), and \( \tilde{W}^* \) the \((q-p_1) \times 1 \) vector of instruments for the endogenous coordinates of \( X^* \). Hence, in this example, \( Z^* = (Y^*, X^*, W^*) \) and \( g(Z^*, \theta^*) = W^*(Y^* - X^*\theta^*) \).

First, let us consider the case where the dependent variable, regressors, and instruments are all subject to censoring. Here

\[
\rho(Z, \theta) = \frac{W(Y - X^*\theta)}{a(Z)} - \frac{(1 - K_0)}{K_0^2} \sum_{r=1}^{d-1} [(WY)_r - (WX^*)_r, \theta] - \frac{I\{ Z = \varepsilon \}}{K_0} g(c, \theta), \tag{3.9}
\]

and \( \theta^* \) can be estimated by GMM or EL as described earlier. Note that since \( \rho(Z, \theta) \) is linear in \( \theta \), no optimization is needed to obtain the GMM estimator.

The endogenous tobit model, where only \( Y^* \) is censored and \( X^* \) is endogenous, is very important for applications. It follows by letting \( \varepsilon^{-1} = (\infty, \ldots, \infty) \) in (3.9); i.e., in this case

\[
\rho(Z, \theta) = \frac{W(Y - X^*\theta)}{K_0 + (1 - K_0)I\{ Y < c^1 \}} - \frac{(1 - K_0)}{K_0^2} W(c^1 - X^*\theta)I\{ Y > c^1 \}. \tag{3.10}
\]

By Theorem 3.1, under pure endogenous censoring \( n^{1/2} (\hat{\theta} - \theta^*) \) converges in distribution to a normal random vector with mean zero and variance \( (D\Omega^{-1}D)^{-1} \), where

\[
\Omega = \mathbb{E}_f \{ \rho(Z, \theta^*)\rho'(Z, \theta^*) \}, \text{ and } \rho(Z, \theta) \text{ is defined in (3.10).} \]

\[\square\]

Example 3.4 (Censoring and IV). Consider the simultaneous equations model

\[
Y_1^* = X_1^*\theta_1^* + \varepsilon_1 \quad \text{and} \quad Y_2^* = X_2^*\theta_2^* + \varepsilon_2,
\]

where \( \varepsilon = (\varepsilon_1, \varepsilon_2)_{2 \times 1} \) satisfies \( \mathbb{E}_{f_0}(Y_{1*}, Y_{2*}|X^*, \varepsilon|X^*) = 0 \) w.p.1. If none of the regressors is endogenous, then \( X^* = (X_1^*, X_2^*)_{(p_1 + p_2) \times 1} \). Otherwise, \( X^* \) contains the exogenous coordinates of \( (X_1^*, X_2^*) \) and other instruments. \( \mathbb{E}_{Y^*|X^*}(\varepsilon|X^*) = 0 \) w.p.1 implies that \( \mathbb{E}_f \{ A(X^*) \begin{bmatrix} Y_{1*}^* - X_1^*\theta_1^* \\ Y_{2*}^* - X_2^*\theta_2^* \end{bmatrix} \} = 0 \), where \( A(X^*) \) is a \( q \times 2 \) matrix of instrumental variables such that \( q \geq p_1 + p_2 \). This can be conveniently expressed as \( \mathbb{E}_f \{ g(Z^*, \theta^*) \} = 0 \), where \( Z^* = (Y_1^*, Y_2^*, X^*), \theta^* = (\theta_1^*, \theta_2^*), \text{ and } g(Z^*, \theta^*) = A(X^*) \begin{bmatrix} Y_{1*}^* - X_1^*\theta_1^* \\ Y_{2*}^* - X_2^*\theta_2^* \end{bmatrix} \). If instead of \( Z^* \) we observe its censored version \( Z \), then (3.3) can be used to estimate \( \theta^* \) by GMM or EL as described earlier in this section.

The simultaneous equations system in this example has been studied before. See, e.g., Smith and Blundell (1986) and Blundell and Smith (1993). However, the treatment here is more general because unlike these papers we do not assume that \( \varepsilon \) is Gaussian and also allow for the possibility that some other coordinates, besides \( Y_1^* \) and \( Y_2^* \), of \( Z^* \) may also be censored. For the reader’s convenience, we now examine the case when only \( Y_1^* \) and \( Y_2^* \) are censored. This illustration can be used to compare
our treatment with the aforementioned papers. In our notation, censoring of $Y_1^*$ and $Y_2^*$ alone can be easily handled by setting $c^{(1,2)} = (\infty, \ldots, \infty)$. This implies that

$$
\rho(Z, \theta) = A(X)[\frac{1}{K_0 + (1 - K_0)\mathbb{I}\{Y_1^* < c^{(1)}, Y_2^* < c^{(2)}\}} \cdot \begin{bmatrix} Y_1 - X_1^* \theta_1 \\ Y_2 - X_2^* \theta_2 \end{bmatrix}
- \frac{(1 - K_0)}{K_0^2} \cdot \begin{bmatrix} c^{(1)} - X_1^* \theta_1 \\ c^{(2)} - X_2^* \theta_2 \end{bmatrix} \mathbb{I}\{Y_1 > c^{(1)}, Y_2 < c^{(2)}\} + \begin{bmatrix} Y_1 - X_1^* \theta_1 \\ c^{(2)} - X_2^* \theta_2 \end{bmatrix} \mathbb{I}\{Y_2 > c^{(2)}, Y_1 < c^{(1)}\}
+ \begin{bmatrix} c^{(1)} - X_1^* \theta_1 \\ c^{(2)} - X_2^* \theta_2 \end{bmatrix} \mathbb{I}\{Y_1 > c^{(1)}, Y_2 > c^{(2)}\}],
$$

and $\theta^*$ can be estimated by GMM or EL as before.

Example 3.5. Sometimes we may possess auxiliary information about a feature of the target density; e.g., we may know beforehand that the mean of the target population is zero. In general, suppose it is known a priori that $\mathbb{E}_f, \{m(Z^*)\} = 0$, where $m$ is a vector of known functions. These types of models, which are a special case of the general unconditional moment restrictions model examined in this paper, have been investigated by Imbens and Lancaster (1994), Hellerstein and Imbens (1999), and Nevo (2003) (described earlier in Section 1). However, unlike us, Imbens and Lancaster (1994) and Hellerstein and Imbens (1999) assume that $Z^*$ is fully observed. Moreover, none of these papers discuss efficient estimation of target and realized cdf’s. Following our approach, moment based auxiliary information about $f^*$ is easily incorporated by first stacking up $g(Z^*, \theta^*)$ and $m(Z^*)$ to produce an augmented vector of moment conditions and then proceeding as before.

3.4. Hypothesis tests and confidence regions. Suppose we want to test the parametric restriction $H_0 : R(\theta^*) = 0$ against the alternative $H_1 : R(\theta^*) \neq 0$, where $R$ is a $\bar{r} \times 1$ vector of twice continuously differentiable functions such that $\frac{\partial R(\theta^*)}{\partial \theta}$ has rank $\bar{r} \leq p$. Since $n^{-1/2}(\hat{\theta}_{gmm} - \theta^*) \overset{d}{\rightarrow} \mathcal{N}(0, (D'\Omega^{-1}D)^{-1})$, the GMM based Wald statistic $W = nR'(\hat{\theta}_{gmm}) (D'\hat{\Omega}^{-1}D)^{-1} \frac{\partial R(\theta_{gmm})}{\partial \theta}^{-1}R(\theta_{gmm}) \overset{d}{\rightarrow} \chi^2_{\bar{r}}$. Hence, a test for $H_0$ against $H_1$ can be based upon $W$. Another alternative is to use a distance metric test. So let $\hat{\theta}_{gmm} = \arg\max_{\theta \in \mathcal{R}, R(\theta)=0} \text{GMM}(\theta)$ and $\tilde{\theta}_{el} = \arg\max_{\theta \in \mathcal{R}, R(\theta)=0} \text{EL}(\theta)$ denote the GMM and EL estimators under $H_0$. Next, define $\text{DM} = n\big\{\text{GMM}(\hat{\theta}_{gmm}) - \text{GMM}(\hat{\theta}_{gmm})\big\}$ and $\text{LR} = 2\{\text{EL}(\tilde{\theta}_{el}) - \text{EL}(\tilde{\theta}_{el})\}$. Here, DM denotes the GMM based distance metric statistic and LR the EL based likelihood ratio test statistic. A test for $H_0$ against $H_1$ can be based upon DM or EL. To obtain the critical values for these tests we can use Newey and McFadden (1994, Theorem 9.2) and Qin and Lawless (1994, Theorem 2) to see that $\text{DM} \overset{d}{\rightarrow} \chi^2_{\bar{r}}$ and $\text{LR} \overset{d}{\rightarrow} \chi^2_{\bar{r}}$ under $H_0$.

Since $W$, DM, and LR are asymptotically equivalent, the decision to use a particular test statistic in practice usually depends upon computational and other considerations. For e.g., though all three statistics can be inverted to obtain asymptotically valid confidence regions, the DM and LR based regions are invariant to the formulation of $H_0$ and automatically satisfy natural range restrictions. Furthermore, unlike $W$ and DM, the likelihood ratio statistic LR is internally studentized; i.e., it does not require preliminary estimation of any variance terms. This guarantees that confidence regions based on LR are also invariant to nonsingular transformations of the moment conditions.
3.5. **Specification tests.** For the remainder of this section, assume that \( q > p \). Since inference based on the estimated \( \hat{\theta}^* \) is sensible only if (1.1) is true, it is important to test \( H_0 \) against the alternative that it is false. In this section, we describe two ways of testing \( H_0 \). The first approach is easy: GMM theory tells us that \( n \text{GMM}(\hat{\theta}_{gmm}) \xrightarrow{d} \chi^2_{q-p} \) under \( H_0 \). Therefore, a test for overidentifying restrictions (usually called the J-test) in (1.1) can be based on this result.

An EL based specification test for \( H_0 \) can also be developed. Besides being internally studentized and invariant to nonsingular and algebraic transformations of the moment conditions, this test has been shown by Kitamura (2001) to be optimal in terms of a large deviations criterion. So let \( \hat{\theta} \) denote a \( n^{1/2} \)-consistent preliminary estimator of \( \theta \). E.g., \( \hat{\theta} \) can be the GMM or EL estimators defined previously. Following (3.4), the restricted (i.e., under \( H_0 \)) EL can be written as

\[
\text{EL}_r = \sum_{j=1}^{n} \log \hat{p}_j(\hat{\theta}), \quad \text{where } \hat{p}_j(\hat{\theta}) \text{ is given by (3.5).}
\]

Next, consider the unrestricted problem where the model is not imposed. It is well known that the nonparametric maximum likelihood estimator of \( f_e \) in the absence of any auxiliary information puts mass \( 1/n \) at each realized observation and is zero elsewhere. Therefore, the unrestricted nonparametric likelihood is given by

\[
\text{EL}_{ur} = -n \log f_e. \quad \text{Now let } \text{ELR} = 2(\text{EL}_{ur} - \text{EL}_r) = 2 \sum_{j=1}^{n} \log \left\{ 1 + \lambda'(\hat{\theta}) \rho(z,j,\hat{\theta}) \right\}, \text{where } \lambda(\hat{\theta}) \text{ solves (3.6). Then ELR can be regarded as an analog of the usual parametric likelihood ratio test statistic; i.e., } H_0 \text{ is rejected if ELR is large enough. By Qin and Lawless (1994, Corollary 4), } \text{ELR} \xrightarrow{d} \chi^2_{q-p} \text{ Under } H_0. \text{ Hence, critical values for ELR are easily obtained.}
\]

4. **Inference with truncated data**

We now show how to do efficient inference using truncated data. As before, we begin by demonstrating that \( f^* \) is identified if the realized observations are drawn from the enriched mixture density defined in (1.8).

4.1. **Identification.** Let \( b^* = \int_T f^*(z) \, d\mu^* \) denote the probability that \( Z^* \) is observed. Throughout the paper we assume that \( b^* \in (0,1) \) is unknown. Since \( b^* \) can be explicitly written as

\[
b^* = \frac{\mathbb{E}_{f_e} \{I\{Z \in T\}\} - (1 - K_0)}{K_0}, \quad (4.1)
\]

it follows that \( b^* \) is identified. Next, by (1.7) and (1.8),

\[
f^*(z) = \frac{f_e(z)}{K_0 + (1 - K_0)\mathbb{I}\{z \in T\}/b^*}, \quad (4.2)
\]

for all \( z \in \mathbb{R}^d \), where \( b^* \) is given by (4.1). Hence, \( f^* \) is also identified. A sufficient condition for local identification of \( \theta^* \) can be obtained as in Section 3.1.

4.2. **Efficient estimation.** Estimation is also straightforward although now there is an additional parameter (\( b^* \)) to estimate. So let

\[
\rho(Z, \theta, b) = \left[ \frac{\mathbb{I}\{Z \in T\} + (1 - K_0)\mathbb{I}\{Z \in T\}^{\theta} \mathbb{I}\{Z \in T\}^{\theta - b}}{K_0 - (1 - K_0)\mathbb{I}\{Z \in T\}^{\theta - b}} \right]_{(q+1) \times 1}.
\]

Since, by (4.2),

\[
\mathbb{E}_{f^*}\{g(Z^*, \theta^*)\} = 0 \iff \mathbb{E}_{f_e}\{\rho(Z, \theta^*, b^*)\} = 0, \quad (4.3)
\]
where  and  can be jointly and efficiently estimated by using the latter moment condition. In particular, the GMM and EL estimators of  and  are given by  and , respectively, where the objective functions are obtained by replacing  with  in Section 3.2 with  and .

Let  and  where and  are similar to those in Tripathi (2003, Theorem 4.8). This intuition can be made mathematically precise by using the approach of Tripathi (2003). In particular, we can show from first principles that the efficiency bounds for estimating  and  coincide with the asymptotic variance of  for the remainder of the paper. Then, by standard GMM or EL theory,  and  are asymptotically efficient.

Some straightforward but tedious matrix algebra can be used to show that

\[ 
\sum_{j=1}^{n} \hat{g}(Z, \theta) + \frac{1 - K_0}{K_0 b^*} \mathbb{I}\{Z \in T\} \mathbb{E}_{\theta, b} \left[ \frac{g(Z, \theta^* \mathbb{I}\{Z \in T\})}{a(Z, b^*)} \right]
\]

and  are asymptotically normal with mean zero and variance .

\[ 
\Sigma_{12} = -(K_0 b^*)^{-1}(K_0 b^* + 1 - K_0)(D'(\Omega^{-1} D)^{-1}) G
\]

\[ 
\Sigma_{22} = K_0^{-2} \variance \left[ \frac{g(Z, \theta^*)}{a(Z, b^*)} \right] \mathbb{I}\{Z \in T\}
\]

and  can be estimated by replacing the  with consistent estimators: We can use  and  and  and  for the number of observations falling in  for the remainder of the paper. Then, by standard GMM or EL theory, we can estimate  by (4.3) or (4.4) as

\[ 
\hat{\Sigma}_{12} = -K_0^{-1}(K_0 b^* + 1 - K_0)(D'(\Omega^{-1} D)^{-1}) G
\]

\[ 
\hat{\Sigma}_{22} = K_0^{-2} \variance \left[ \frac{g(Z, \theta^*)}{a(Z, b^*)} \right] \mathbb{I}\{Z \in T\}
\]

As mentioned earlier, from (4.3) it is intuitively clear that  and  are asymptotically efficient. This intuition can be made mathematically precise by using the approach of Tripathi (2003). In particular, we can show from first principles that the efficiency bounds for estimating  and  coincide with the asymptotic variance of  and  in Theorem 4.1. We omit the details since they are very similar to those provided for the proof of Theorem 4.9 in Tripathi (2003).

Next, we efficiently estimate  and  . So let  and  and  and  are the EL probability weights evaluated at  and  . Then, from Qin and Lawless (1994, Theorem 1), we know that  is asymptotically normal with mean zero and variance

\[ 
\hat{F}(\xi) = \sum_{j=1}^{n} \hat{\rho}_j(\hat{\theta}_{el}, \hat{b}_{el}) \mathbb{I}\{Z_j \leq \xi\}
\]

where  and  are the EL probability weights evaluated at  and  . Then, from Qin and Lawless (1994, Theorem 1), we know that  is asymptotically normal with mean zero and variance

\[ 
\hat{F}(\xi) \left( 1 - \hat{F}(\xi) \right) - \mathbb{E}_{\theta, b} \left[ \frac{g(Z, \theta^*, b^*)}{a(Z, b^*)} \mathbb{I}\{Z \leq \xi\} \right] M_f \mathbb{E}_{\theta, b} \left[ \frac{g(Z, \theta^*, b^*)}{a(Z, b^*)} \mathbb{I}\{z \leq \xi\} \right]
\]
Theorem 4.2. \( n^{1/2} \{ \hat{F}(\xi) - F_e(\xi) \} \) is asymptotically normal with mean zero and variance
\[
F_e(\xi) \{ 1 - F_e(\xi) \} \cdot \text{cov}_{f_e} \left\{ \frac{\hat{g}(Z, \theta^*)}{a(Z, b^*)}, I(Z \leq \xi) \right\} \left( \Omega^{-1} - \Omega^{-1} D (D' \Omega^{-1} D)^{-1} D' \Omega^{-1} \right) \text{cov}_{f_e} \left\{ \frac{\hat{g}(Z, \theta^*)}{a(Z, b^*)}, I(Z \leq \xi) \right\}.
\]

Also, letting
\[
m_\xi(Z) = \mathbb{I} \{ Z \leq \xi \} + \frac{(1 - K_0)}{K_0 b^*} \mathbb{I} \{ Z \in T \} \mathbb{E}_{f_e} \left[ \mathbb{I} \{ Z \leq \xi \} \mathbb{I} \{ Z \in T \} \right] a(Z, b^*)
\]
and following the linearization argument described in the proof of Theorem 4.10 in Tripathi (2003), it is easy to show that

Theorem 4.3. \( n^{1/2} \{ \hat{F}^*(\xi) - F^*(\xi) \} \) is asymptotically normal with mean zero and variance
\[
\text{var}_{f_e} \left\{ \frac{m_\xi(Z)}{a(Z, b^*)} \right\} \cdot \text{cov}_{f_e} \left\{ \frac{\hat{g}(Z, \theta^*)}{a(Z, b^*)}, m_\xi(Z) \right\} \left( \Omega^{-1} - \Omega^{-1} D (D' \Omega^{-1} D)^{-1} D' \Omega^{-1} \right) \text{cov}_{f_e} \left\{ \frac{\hat{g}(Z, \theta^*)}{a(Z, b^*)}, m_\xi(Z) \right\}.
\]

As before, Theorems 4.2 and 4.3 show that if the model is overidentified, then imposing it leads to an efficiency gain in estimating \( F_e \) and \( F^* \). Asymptotic optimality of EL implies that \( \hat{F}(\xi) \) and \( \hat{F}^*(\xi) \) are also asymptotically efficient. Alternatively, this can be proved directly by using the approach of Tripathi (2003) to first calculate the efficiency bounds for estimating \( F_e(\xi) \) and \( F^*(\xi) \) and then comparing these bounds with the asymptotic variances in Theorems 4.2 and 4.3. We omit the details to keep our presentation concise.

4.3. Examples. Let us reexamine some of the examples in Sections 1 and 4.3.

Example 4.1 (Example 1.3 contd.). Since here
\[
\rho(Z, \theta, b) = \left[ \frac{Z - \theta}{\mathbb{I} \{ Z \in T \} - (1 - K_0)} - b \right]
\]
and there are no overidentifying restrictions, \( \hat{\theta} \) and \( \hat{b} \) are obtained by solving \( \sum_{j=1}^n \rho(Z_j, \hat{\theta}, \hat{b}) = 0 \).
Hence, \( \hat{b} = [n_T/n - (1 - K_0)]/K_0 \) and \( \hat{\theta} = \left( \sum_{j=1}^n 1/a(Z_j, \hat{b}) \right)^{-1} \left\{ \sum_{j=1}^n Z_j/a(Z_j, \hat{b}) \right\} \). By the weak law of large numbers, it is straightforward to show that \( \hat{\theta} \) and \( \hat{b} \) are consistent. Furthermore, by the CLT,
\[
n^{1/2} (\hat{\theta} - \theta^*) \overset{d}{\rightarrow} \mathcal{N}(0, \text{var}_{f_e} \left\{ \frac{Z - \theta^*}{a(Z, b^*)}, (1 - K_0) \mathbb{E}_{f_e} \left[ \mathbb{I} \{ Z \in T \} \right] \right\}).
\]

Of course, this matches with the result in Theorem 4.1. The absence of overidentifying restrictions implies that in this example, and the next one, \( \hat{F}(\xi) = n^{-1} \sum_{j=1}^n I \{ Z_j \leq \xi \} \) and \( \hat{F}^*(\xi) = n^{-1} \sum_{j=1}^n I \{ Z_j \leq \xi \}/a(Z_j, b_0) \). Therefore, by Theorem 4.2 or by a direct application of the CLT, \( n^{1/2} \{ \hat{F}(\xi) - F_e(\xi) \} \) is asymptotically normal with mean zero and variance \( F_e(\xi) \{ 1 - F_e(\xi) \} \). Similarly, by Theorem 4.3, \( n^{1/2} \{ \hat{F}^*(\xi) - F^*(\xi) \} \) is asymptotically normal with mean zero and variance \( \text{var}_{f_e} \left\{ \frac{m_\xi(Z)}{a(Z, b^*)} \right\} \).

Example 4.2 (Example 1.4 contd.). Again, since there are no overidentifying restrictions and
\[
\rho(Z, \theta, b) = \left[ \frac{X - \theta}{\mathbb{I} \{ Y \in T \} - (1 - K_0)} - b \right],
\]
where \( a(Y, X, b) = K_0 + (1 - K_0) \mathbb{I}\{Y \in T_1, X \in T_2\}/b \), we have \( \hat{b} = [n_T/n - (1 - K_0)]/K_0 \) and

\[
\hat{\theta} = \left( \sum_{j=1}^{n} \frac{X_jX_j'}{a(Y_j, X_j, b)} \right)^{-1} \sum_{j=1}^{n} \frac{X_jY_j}{a(Y_j, X_j, b)}.
\]  

(4.5)

Hence, \( n^{1/2}(\hat{\theta} - \theta^*) \overset{\mathcal{D}}{\to} N(0, D^{-1}\Omega D^{-1}) \) by Theorem 4.1, where

\[
D = -\mathbb{E}_f\{ \frac{XX'}{a(Y, X, \theta^*)} \},
\]

\[
\Omega = \text{var}_f\{ \frac{X(Y - X'\theta^*)}{a(Y, X, \theta^*)} + \frac{(1 - K_0)}{K_0b^2} \mathbb{I}\{Y \in T_1, X \in T_2\} \mathbb{E}_f\left[ \frac{X(Y - X'\theta^*)\mathbb{I}\{Y \in T_1, X \in T_2\}}{a(Y, X, \theta^*)} \right] \}.
\]

Notice that if truncation is purely endogenous, then (4.5) implies that

\[
\hat{\theta} = \left( \sum_{j=1}^{n} \frac{X_jX_j'}{K_0 + (1 - K_0)\mathbb{I}\{Y_j \in T_1\}/b} \right)^{-1} \sum_{j=1}^{n} \frac{X_jY_j}{K_0 + (1 - K_0)\mathbb{I}\{Y_j \in T_1\}/b}.
\]

The asymptotic variance of \( \hat{\theta} \) can be obtained by setting \( T_2 = \mathbb{R}^p \) in (4.6). Similarly, under pure exogenous truncation,

\[
\hat{\theta} = \left( \sum_{j=1}^{n} \frac{X_jX_j'}{K_0 + (1 - K_0)\mathbb{I}\{X_j \in T_2\}/b} \right)^{-1} \sum_{j=1}^{n} \frac{X_jY_j}{K_0 + (1 - K_0)\mathbb{I}\{X_j \in T_2\}/b},
\]

and its asymptotic variance is given by substituting \( T_1 = \mathbb{R} \) in (4.6).

\[\Box\]

**Example 4.3** (Truncated linear regression with endogenous regressors). Consider the setup of Example 3.3 but now suppose that, instead of being censored, \( Z^* \) is truncated outside \( T_1 \times T_2 \times T_3 \). If the dependent variable, regressors, and instruments are all subject to truncation, then

\[
\rho(Z, \theta, b) = \begin{bmatrix}
W(Y - X'\theta^*) \\
K_0 + (1 - K_0)\mathbb{I}\{Y \in T_1, X \in T_2, W \in T_3\}/b \\
\mathbb{I}\{Y \in T_1, X \in T_2, W \in T_3\} - (1 - K_0) - b
\end{bmatrix}^T.
\]

Hence, \( \theta^* \) and \( b^* \) can be estimated by GMM or EL as described earlier.

Under pure endogenous truncation (i.e., \( T_2 = \mathbb{R}^p \) and \( T_3 = \mathbb{R}^{q-p} \)), by Theorem 4.1 it follows that \( n^{1/2}(\hat{\theta} - \theta^*) \) is asymptotically normal with mean zero and variance \( (D'\Omega^{-1}D)^{-1} \), where

\[
D = -\mathbb{E}_f\{ \frac{WX'}{K_0 + (1 - K_0)\mathbb{I}\{Y \in T_1\}/b^*} \}
\]

and

\[
\Omega = \text{var}_f\{ \frac{W(Y - X'\theta^*)}{K_0 + (1 - K_0)\mathbb{I}\{Y \in T_1\}/b^*} + \frac{(1 - K_0)}{K_0b^2} \mathbb{I}\{Y \in T_1\} \mathbb{E}_f\left[ \frac{W(Y - X'\theta^*)\mathbb{I}\{Y \in T_1\}}{K_0 + (1 - K_0)\mathbb{I}\{Y \in T_1\}/b^*} \right] \}.
\]

A similar result can be derived for pure exogenous truncation as well. \[\Box\]
Example 4.4 (Truncation and IV). Again, consider the simultaneous equations model of Example 3.4, but now assume that \( Z^* \) is truncated instead of being censored. To further simplify the exposition, suppose that only \( Y_1^* \) and \( Y_2^* \) are truncated outside \( T_1 \) and \( T_2 \), respectively. In this case

\[
\rho(Z, \theta, b) = \begin{bmatrix}
A(X) \left[ \frac{Y_1 - X_1'\theta_1}{K_0 + (1-K_0)1\{Y_1 \in T_1, Y_2 \in T_2\} - (1-K_0)} \right] \\
1\{Y_1 \in T_1, Y_2 \in T_2\}
\end{bmatrix},
\]

and \((\theta^*, b^*)\) can be estimated by GMM or EL as described earlier in the section.

4.4. Hypothesis and specification tests. Hypotheses of the form \( R(\theta^*) = 0 \) vs \( R(\theta^*) \neq 0 \) can be tested using the Wald, DM, or LR statistics as described in Section 3.4 by basing the test on \( \rho(Z, \theta^*, b^*) \). In each case the test statistic is asymptotically distributed as a \( \chi^2_r \) random variable under the null hypothesis. Similarly, if \( q > p \) then a test for overidentifying moment restrictions can be done using the GMM based \( J \)-statistic or the ELR statistic on the moment vector \( \rho(Z, \theta^*, b^*) \). The details are analogous to those in Section 3.5. In either case, the test statistic has an asymptotic \( \chi^2_{q-p} \) distribution under \( H_0 \).

5. Conclusion

This paper develops efficient semiparametric inference for models with unconditional moment restrictions when the target population is subject to censoring or truncation. Instead of imposing parametric, independence, symmetry, quantile, or special regressor restrictions on the distributions of the underlying random variables, we solve the identification problem created due to the incompleteness of data by using a supplementary sample of observations that are not subject to censoring or truncation. We show how this additional sample can be combined with the original dataset of censored or truncated observations to efficiently correct for the effects of partial observation so that all standard GMM and EL based inference goes through. No special software is required to implement the procedures developed in this paper. Any computer package that can do GMM or EL will be able to do the same with censored or truncated data. The techniques developed here can be extended to handle missing data as well. Research on this topic is in progress and will be presented in a subsequent paper.

References


DEPARTMENT OF ECONOMICS, UNIVERSITY OF WISCONSIN-MADISON, MADISON, WI-53706.

E-mail address: gtripath@ssc.wisc.edu