OPTIMALLY COMBINING CENSORED AND UNCENSORED DATASETS

PAUL J. DEVEREUX AND GAUTAM TRIPATHI

Abstract. Economists and other social scientists often encounter data generating mechanisms (dgm's) that produce censored or truncated observations. These dgm'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. These additional observations, or refreshment samples as they are sometimes called, can often be obtained by creatively matching existing datasets. To illustrate our results in an empirical setting, we show how to estimate the effect of changes in compulsory schooling laws on age at first marriage, a variable that is censored for younger individuals. We also demonstrate how refreshment samples for this application can be created by matching cohort information across census datasets.

1. Introduction

In applied research, economists often face situations in which they have access to two datasets that they can use but one set of data suffers from censoring or truncation. In some cases, especially if the censored sample is larger, researchers use it and attempt to deal with the problem of partial observation in some manner[1]. In other cases, economists simply use the clean sample, i.e., the dataset not subject to censoring or truncation, and ignore the censored one so as to avoid biases. It is rarely the case that researchers utilize both datasets. Instead, they have to choose between the two mainly because they lack guidance about how to combine them. In this paper, we develop a methodology that allows the censored and uncensored datasets to be combined in a tractable manner so that resulting estimators are consistent and use all information optimally. When the censored sample, henceforth

Date: February 24, 2004. **Preliminary and incomplete version. Comments welcome.**

Key words and phrases. Censoring, Empirical likelihood, GMM, Refreshment samples, Truncation.

An earlier version of this paper was titled "Combining datasets to overcome selection caused by censoring and truncation in moment based models". We thank Moshe Buchinsky, Ken Couch, Graham Elliott, Jin Hahn, Jim Heckman, Yuichi Kitamura, Arthur Lewbel, Geert Ridder, Yixiao Sun, Allan Timmermann, Hal White, and participants at several seminars for helpful suggestions and conversations. Financial support from the NSF (Devereux) and the University of Connecticut graduate school (Tripathi) is gratefully acknowledged.

[1]A comprehensive survey of the econometric literature on censoring and truncation is beyond the scope of our paper. Readers interested in this should see, among others, Hausman and Wise (1976, 1977), Heckman (1976, 1979), Maddala (1983), Powell (1983, 1984, 1986a, 1986b, 1994), Amemiya (1984), Chamberlain (1986), Duncan (1986), Fernandez (1986), Horowitz (1986, 1988), Newey (1988), Manski (1989, 1995), Newey and Powell (1990), Lee (1993a, 1993b), Honoré and Powell (1994), Buchinsky and Hahn (1998), Vella (1998), Chen and Khan (2001), Khan and Powell (2001), Khan and Lewbel (2003), and the many references therein.

referred to as the *master* sample, is much bigger than the clean or the *refreshment* sample, one can think of the addition of the clean sample as providing identification where it is otherwise absent. In contrast, when the datasets are of similar sizes so the clean dataset could typically be used alone, our methodology can be interpreted as using information from the censored sample to increase efficiency. In fact, we show that using the clean sample alone leads to estimators that are asymptotically inefficient revealing that there is information in censored or truncated samples that can be exploited to enable more efficient estimation. Note that the existence of clean or refreshment samples should not be regarded as being an overly restrictive requirement. As we illustrate in Section 6, they can often be constructed by creatively *matching* existing datasets.

Let the triple (Z^*, f^*, μ^*) describe the *target* population, i.e., the population for which inference is to be drawn, where Z^* is a random vector in $\mathbb{R}^d$ (following usual mathematical convention, "vector" means a column vector) that denotes an observation from the target population, and f^* the unknown density of Z^* with respect to some dominating measure $\mu^* = \otimes_{i=1}^d \mu_i^*$; since Z^* is allowed to have discrete components, the μ_i^*'s need not all be Lebesgue measures. Similarly, let (Z, f, μ) represent the *realized* population, i.e., the observed data, where Z denotes the resulting observation and f its density with respect to a dominating measure $\mu = \otimes_{i=1}^d \mu_i$. The basic inferential problem is that *the model comes from f^* but the data comes from f.*

In this paper, f is different from f^* because some or all coordinates of Z^* are censored, or, truncated. As stated earlier, our objective is to develop new techniques that will allow economists to do efficient semiparametric inference by supplementing the partially observed master sample with some additional data that are complete, i.e., not subject to the mechanism that created the censored or truncated observations. The idea of using supplementary samples in incomplete, biased sampling, or measurement error models has of course been recognized earlier; see, e.g., Manski and Lerman (1977), Titterington (1983), Titterington and Mill (1983), Vardi (1985), Ridder (1992), Carroll, Ruppert, and Stefanski (1995), Wansbeek and Meijer (2000), Hirano, Imbens, Ridder, and Rubin (2001), Chen, Hong, and Tamer (2004), Tripathi (2004), and the references therein. However, the use of matching to facilitate efficient moment based inference in overidentified models with incomplete 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.

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 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). 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), and the 2002 special issue of the Journal of Business and Economics Statistics.

This paper illustrates several advantages of using refreshment samples to handle incomplete data. For instance, many partially observed or missing data models in applied work are identified by assuming that the selection probabilities (i.e., the probability that an observation is fully observed)

— or propensity scores, as they are often called — are completely known or can be parametrically modelled; see, e.g., Robins, Rotnitzky, and Zhao (1994), Robins, Hsieh, and Newey (1995), Nan, Emond, and Wellner (2002), and the references therein. However, as noted by Horowitz and Manski (1998), these assumptions lead to stochastic restrictions that are usually not testable. In contrast, since we rely upon a refreshment sample to provide identifying power, the selection probabilities in this paper are fully nonparametric (i.e., completely unrestricted) and, hence, we can avoid the above mentioned non-testability problems. More generally, in this paper we demonstrate how a refreshment sample allows standard GMM and EL based inference, including tests of overidentifying restrictions, with censored or truncated data to go through *without* imposing parametric, independence, symmetry, quantile, or "special regressor" restrictions as done in the existing literature.

Furthermore, there is no need to worry about issues relating to heteroscedasticity (a major concern for parametric models) because GMM and EL automatically produce correct standard errors and, unlike quantile restriction models, there is no need to restrict attention to applications where only scalar-valued continuously distributed random variables are censored or truncated, or use any nonparametric smoothing procedures to estimate asymptotic variances. Extension to the case where more than one random variable (discrete or continuous) is censored or truncated is straightforward and the usual analogy principle that delivers standard errors for GMM or EL works here as well.

The treatment proposed here is general enough to handle censoring and truncation of some or all coordinates of both endogenous and exogenous variables and the results obtained here are applicable to a large class of potentially overidentified models which nest linear regression as a special case; e.g., the ability to handle IV models permits semiparametric inference in Box-Cox type transformation models using censored or truncated data without imposing parametric or quantile restrictions.

The rest of the paper is organized as follows. In Section 2 we set up the moment based model and describe how to deal with the censoring or truncation of random vectors in a multivariate framework, several examples of which are given throughout the paper (including Section 3). Section 2 also describes how we model he data combination process. Section 4 shows how censored data can be combined with a refreshment sample to do efficient moment based inference and Section 5 does the same with truncated data. Section 6 contains an interesting application where refreshment samples are obtained by matching census datasets. Section 7 concludes by addressing some topics for future research. Proofs, tables, and figures are all in the appendices.

2. Setup

The econometric models we consider can be expressed in terms of moment conditions as follows. Let Θ be a subset of $\mathbb{R}^p$ such that

$$\mathrm{H}_0: \quad \mathbb{E}_{f^*}\{g(Z^*, \theta^*)\} = 0 \quad \text{for some } \theta^* \in \Theta, \tag{2.1}$$

where g is a $q \times 1$ vector of known functions with $q \geq p$ and $\mathbb{E}_{f^*}$ denotes expectation with respect to the pdf f^*. Models specified via moment conditions are particularly important for structural estimation: Since economic theory attributes outcomes at the micro level to optimizing behavior on the part of

firms or individuals, moment based models arise naturally in microeconometrics as solutions to the first order conditions of the stochastic optimization problems economic agents are assumed to solve.

Well-known examples of (2.1) include linear regression models $Y^* = X^{*\prime}\theta^* + \varepsilon^*$, where the error term is uncorrelated with the regressors; i.e., $\mathbb{E}_{f^*}\{X^*\varepsilon^*\} = 0$. Here $g(Z^*, \theta^*) = X^*\{Y^* - X^{*\prime}\theta^*\}$ and $Z^* = (Y^*, X^*)$, where Y^* denotes the endogenous variable and X^* the vector of explanatory variables. This can be extended to nonlinear regression models of the form $Y^* = \psi(X^*, \theta^*) + \varepsilon^*$, where ψ is known up to θ^* and $\mathbb{E}_{f^*}\{\varepsilon^*\partial\psi(X^*, \theta^*)/\partial\theta\} = 0$. Other generalizations include linear and nonlinear regression models with endogenous regressors and multivariate simultaneous equations models where the error terms are uncorrelated with some or all of the explanatory variables.

The class of models defined in (2.1) also contains instrumental variables (IV) models. Suppose we have the conditional mean restriction $\mathbb{E}_{Y^*|X^*}\{\tilde{g}(Y^*, X^*, \theta^*)|X^*\} = 0$ w.p.1, where $\tilde{g}$ is a $k \times 1$ vector of known functions and Y^* the vector of endogenous variables. Letting $A(X^*)$ denote a $q \times k$ matrix of instruments, this yields unconditional moment restrictions of the form $\mathbb{E}_{f^*}\{g(Z^*, \theta^*)\} = 0$, where $g(Z^*, \theta^*) = A(X^*)\tilde{g}(Y^*, X^*, \theta^*)$.

2.1. **Censoring.** If the target variable Z^* is fully observed, then (2.1) is easily handled; see, e.g., Newey and McFadden (1994). 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. So 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

$$Z^{(i)} = \begin{cases} Z^{*(i)} & \text{if } Z^{*(i)} < c^{(i)} \\ c^{(i)} & \text{otherwise} \end{cases} \qquad \text{for } i = 1, \ldots, d$$

and $c = (c^{(1)}, \ldots, c^{(d)})$ is a $d \times 1$ vector of known constants[2].

We allow for the possibility that some components of Z^* may not be censored: If, say, the ith coordinate of Z^* is not subject to the censoring mechanism, simply set $c^{(i)} = \infty$; if the ith and jth 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 Z^* can be decomposed into endogenous and exogenous parts as (Y^*, X^*), we can handle situations where only Y^* is censored (pure endogenous censoring), or only X^* is censored (pure exogenous censoring)[3], or only some coordinates of either variables are censored. Left censoring of, say, the ith, jth, and kth coordinates can also be accommodated by replacing $Z^{*(i,j,k)}$ with $-Z^{*(i,j,k)}$ and $c^{(i,j,k)}$ with $-c^{(i,j,k)}$.

[2]The results obtained in this paper continue to hold in a more general *fixed* censoring framework where the censoring point is modelled as a random variable C with unknown distribution such that C is observed for censored as well as uncensored observations; see, e.g., the application in Section 6.

[3]The term "exogenous" is, strictly speaking, an abuse of terminology since (2.1) does not involve any conditioning although, as mentioned earlier in the introduction, (2.1) does nest IV models based on conditional moment restrictions. Therefore, the careful reader may want to substitute "censoring (resp. truncation) based on explanatory variables" for "exogenous censoring (resp. truncation)" whenever the latter is encountered.

Let S^* denote the survival function induced by f^*, i.e., $S^*(\xi) = \Pr_{f^*}(\cap_{i=1}^d \{Z^{*(i)} > \xi^{(i)}\})$, and δ_c the Dirac measure at c, i.e., $\delta_c(A) = \mathbb{I}(c \in A)$, where $\mathbb{I}$ is the indicator function. To keep matters simple, *we assume that μ^* does not have an atom at c.* This assumption, which can be relaxed at the cost of greater mathematical complexity, is weaker than requiring μ^* to be a Lebesgue measure (the usual assumption made for censored regression models).

If $d = 1$, the density of Z with respect to the dominating measure $\mu = \mu^* + \delta_c$ is given by

$$f(z) = f^*(z)\mathbb{I}(z < c) + S^*(c)\mathbb{I}(z = c). \tag{2.2}$$

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 coordinates of Z^* that remain after the ith, jth, and kth 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 $f^*_{i,j,k|-(i,j,k)}$, the conditional density of $Z^{*(i,j,k)}|Z^{*-(i,j,k)}$. 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 $\mu_i = \mu_i^* + \delta_c^{(i)}$, is given by

$$f(z) = f^*(z)\mathbb{I}(z \overset{elt}{<} c) + \sum_{r=1}^{d-1} \sum_{i_1=1}^{d-r+1} \sum_{i_2=i_1+1}^{d-r+2} \cdots \sum_{i_r=i_{r-1}+1}^{d} S^*_{i_1,\ldots,i_r|-(i_1,\ldots,i_r)}(c^{(i_1,\ldots,i_r)}) f^*_{-(i_1,\ldots,i_r)}(z^{-(i_1,\ldots,i_r)})$$
$$\times \mathbb{I}(z^{(i_1,\ldots,i_r)} = c^{(i_1,\ldots,i_r)}, z^{-(i_1,\ldots,i_r)} \overset{elt}{<} c^{-(i_1,\ldots,i_r)}) + S^*(c)\mathbb{I}(z = c), \tag{2.3}$$

where $\overset{elt}{<}$ denotes element-by-element strict inequality; i.e., $\mathbb{I}(z \overset{elt}{<} c) = \prod_{i=1}^d \mathbb{I}(z^{(i)} < c^{(i)})$. Of course, $z = c$ still denotes element-by-element equality; i.e., $\mathbb{I}(z = c) = \prod_{i=1}^d \mathbb{I}(z^{(i)} = c^{(i)})$. Note that f has support $(-\infty, c^{(1)}] \times \ldots \times (-\infty, c^{(d)}]$ with an atom at c.

2.2. **Truncation.** Sometimes censoring is so severe that the target variable is completely unobserved outside a certain region. This phenomenon is called truncation; e.g., in many job training programs subjects are allowed entry only if their household income falls below a certain level. 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 μ^* is given by

$$f(z) = \frac{f^*(z)\mathbb{I}(z \in T)}{\int_T f^*(z)\, d\mu^*}. \tag{2.4}$$

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 $I_i = I_j = I_k = \mathbb{R}$.

2.3. **Data combination.** Since f^* and, hence, f are completely unknown, censoring or truncation of Z^* creates 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 (2.2)–(2.3) and (2.4), the target density $f^*(z)$ cannot be expressed in terms of $f(z)$ for all $z \in \mathbb{R}^d$. In other words, f^* cannot be fully recovered from f; i.e., f^* is not identified. But if there is no way of going from the realized density (loosely speaking, the "reduced form") to the target density (the "structural form"), then statistical inference about f^* and, hence, the target cdf $F^*(\xi) = \Pr_{f^*}(Z^* \overset{elt}{\leq} \xi)$ is impossible[4]. As shown later, combining the master and refreshment samples allows us to overcome this problem.

We model the data combination process as follows: Let Z denote an observation from the combined sample. Along with Z we observe a dummy variable R that indicates whether Z comes from the refreshment sample or the master sample; i.e., $R = 1$ if Z is from the refreshment sample whereas $R = 0$ if Z belongs to the master sample. Hence, the conditional density of $Z|R = r$ is[5]

$$f_{Z|R=r}(z) = \begin{cases} f^*(z)\mathbb{I}(z \overset{elt}{\neq} c)r + f(z)(1-r) & \text{if } Z^* \text{ is censored} \\ f^*(z)r + f(z)(1-r) & \text{if } Z^* \text{ is truncated,} \end{cases} \tag{2.5}$$

where $r \in \{0,1\}$, $\mathbb{I}(z \overset{elt}{\neq} c) = \prod_{i=1}^d \mathbb{I}(z^{(i)} \neq c^{(i)})$, and, depending on whether Z^* is censored or truncated, f is given by (2.2)–(2.3) or (2.4), respectively. If Z^* is censored, then $f_{Z|R=r}$ is a conditional density with respect to μ and has an atom at c. On the other hand, if Z^* is truncated, then $f_{Z|R=r}$ is a conditional density with respect to μ^*.

Assume that $R \overset{d}{\sim} \text{Bernoulli}(K_0)$, where $K_0 \in (0,1]$ is unknown and will be estimated along with the parameters of interest. Loosely speaking, K_0 is the probability of sampling from the target population without subjecting the observations to censoring or truncation. Therefore, using (2.5), the joint density of Z and R is given by

$$f_e(z,r) = \begin{cases} K_0 f^*(z)\mathbb{I}(z \overset{elt}{\neq} c)r + (1-K_0)f(z)(1-r) & \text{if } Z^* \text{ is censored} \\ K_0 f^*(z)r + (1-K_0)f(z)(1-r) & \text{if } Z^* \text{ is truncated.} \end{cases} \tag{2.6}$$

Henceforth, let n denote the size of the *enriched* sample; i.e., the master and refreshment samples combined together. Throughout the paper, all limits are taken as $n \uparrow \infty$. Observations $(Z_1, R_1), \ldots, (Z_n, R_n)$ from the enriched dataset are regarded as iid draws from f_e, which is a density

[4]Since distribution functions characterize random variables, estimating $F^*(\xi)$ at each $\xi \in \mathbb{R}^d$ determines the probabilistic behavior of Z^*. Efficient estimation of the target cdf is important for bootstrapping from the target population. Brown and Newey (2002) note that when prior information about the target population is available, merely using a consistent estimator of F^* can lead to poor inference from the bootstrap. They recommend that resampling be done using $\hat{F}^*$, an efficient estimator of F^* that incorporates restrictions imposed by the model (2.1). Estimating the realized cdf under (2.1) is also useful because comparing it with $\hat{F}^*$ can help reveal the extent of bias induced by censoring or truncation. Of course, $\hat{F}^*$ can always be compared with the empirical cdf of the observed data. But since the latter does not take the model into account, it will be less precise (though more robust) than an estimator of the realized cdf under (2.1).

[5]Since f^* is a density with respect to μ^*, it is only identified up to sets of μ^*-measure zero. Therefore, if Z^* is censored then $f^*(z)\mathbb{I}(z \overset{elt}{\neq} c)$ is a μ^*-version of f^* and, hence, $\mathbb{E}_{f^*}\{g(Z^*, \theta^*)\} = 0$ if and only if $\mathbb{E}_{f^*}[g(Z^*, \theta^*)\mathbb{I}(Z^* \overset{elt}{\neq} c)] = 0$.

with respect to $\mu \otimes \kappa$, where κ denotes the counting measure on $\{0, 1\}$. In Sections 4 and 5 we show how data from this enriched density can be used to fully recover f^* and estimate and test (2.1).

3. Examples

In this section we look at some examples of censoring and truncation in a multivariate framework. The primary aim is to illustrate what happens in linear models when only the master sample is used for estimation; examples 3.2 and 3.4 are particularly instructive. Note that since no refreshment sample is used in this section, n here just denotes the master sample size. This minor notational ambiguity should not cause any confusion.

Example 3.1 (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 $Z_1^*, \ldots, Z_n^*$ from the target density f^* we have the master random sample $Z_1, \ldots, Z_n$ from the realized density f defined in (2.2) or (2.3). Therefore, the naive estimator $\sum_{j=1}^n Z_j/n$ will not consistently estimate θ^* because $\sum_{j=1}^n Z_j/n \xrightarrow{p} \mathbb{E}_f\{Z\}$ by the weak law of large numbers (WLLN), but

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

where, for any function $h(\cdot)$, the symbol

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

denotes h evaluated at exactly r censored coordinates and $Z^*[i_1, \ldots, i_r]$ stands for Z^* with its i_1, …, i_rth coordinates replaced by $c^{(i_1)}, \ldots, c^{(i_r)}$, respectively, and the remaining coordinates unchanged; i.e., $Z^*[i_1, \ldots, i_r] = Z^*\big|_{Z^{*(i_1,\ldots,i_r)}=c^{(i_1,\ldots,i_r)}}$. □

Example 3.2 (Censored linear regression)**.** Let $Y^* = X^{*\prime}\theta^* + \varepsilon^*$, 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 (2.3). If we ignore censoring and simply regress Y on X, then θ^* cannot be consistently estimated by the least squares estimator $\hat{\theta}_M = (\sum_{j=1}^n X_j X_j')^{-1} \sum_{j=1}^n X_j Y_j$. To see this, observe that the probability limit of $\hat{\theta}_M$ is given by

$$\begin{aligned}(\mathbb{E}_f XX')^{-1}(\mathbb{E}_f XY) = (\mathbb{E}_{f^*}\{X^*X^{*\prime}\mathbb{I}(Y^* < c^{(1)}, X^* \overset{elt}{<} c^{-(1)}) + \sum_{r=1}^{d-1}(X^*X^{*\prime})_r + c^{-(1)}c^{-(1)\prime}S^*(c)\})^{-1} \\ \times \mathbb{E}_{f^*}\{X^*Y^*\mathbb{I}(Y^* < c^{(1)}, X^* \overset{elt}{<} c^{-(1)}) + \sum_{r=1}^{d-1}(X^*Y^*)_r + c^{-(1)}c^{(1)}S^*(c)\}, \quad (3.1)\end{aligned}$$

where $d = p + 1$ and, in the notation introduced in Example 3.1,

$$(X^*Y^*)_r = \sum_{i_1=1}^{d-r+1} \sum_{i_2=i_1+1}^{d-r+2} \cdots \sum_{i_r=i_{r-1}+1}^{d} (X^*Y^*)\big|_{(Y^*,X^*)^{(i_1,\ldots,i_r)}=c^{(i_1,\ldots,i_r)}} \mathbb{I}(Z^{*(i_1,\ldots,i_r)} \overset{elt}{>} c^{(i_1,\ldots,i_r)}, Z^{*-(i_1,\ldots,i_r)} \overset{elt}{<} c^{-(i_1,\ldots,i_r)}), \quad (3.2)$$

and $(X^*X^{*\prime})_r$ is obtained by replacing Y^* in (3.2) with $X^{*\prime}$. Hence, $\text{plim}(\hat{\theta}_M) \neq (\mathbb{E}_{f^*} X^* X^{*\prime})^{-1}(\mathbb{E}_{f^*} X^* Y^*)$.

The special case of pure endogenous censoring, frequently called the tobit or limited dependent variable model in the econometrics literature, is obtained by letting $c^{-(1)} = (\infty, \ldots, \infty)$ and using the convention that $0 \cdot \infty = 0$. Doing so, (3.1) implies that

$$\text{plim}(\hat{\theta}_M) = \theta^* - \{\mathbb{E}_{f^*} X^* X^{*\prime}\}^{-1} \mathbb{E}_{f^*}\{X^*(Y^* - c^{(1)})\mathbb{I}(Y^* > c^{(1)})\} \neq \theta^*,$$

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 (3.1),

$$\text{plim}(\hat{\theta}_M) = \{\mathbb{E}_{f^*}[X^* X^{*\prime} \mathbb{I}(X^* \overset{elt}{<} c^{-(1)}) + \sum_{r=1}^{d-1} (X^* X^{*\prime})_r]\}^{-1} \mathbb{E}_{f^*}\{X^* Y^* \mathbb{I}(X^* \overset{elt}{<} c^{-(1)}) + \sum_{r=1}^{d-1} (X^* Y^*)_r\} \neq \theta^*,$$

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

$$\sum_{i_1=2}^{d-r+1} \sum_{i_2=i_1+1}^{d-r+2} \cdots \sum_{i_r=i_{r-1}+1}^{d} (X^*Y^*)\big|_{(Y^*,X^*)^{(i_1,\ldots,i_r)}=c^{(i_1,\ldots,i_r)}} \times \mathbb{I}(X^{*(i_1-1,\ldots,i_r-1)} \overset{elt}{>} c^{(i_1,\ldots,i_r)}, X^{*-(i_1-1,\ldots,i_r-1)} \overset{elt}{<} c^{-(1,i_1,\ldots,i_r)}) \quad (3.3)$$

and $(X^*X^{*\prime})_r$ follows by replacing Y^* in (3.3) with $X^{*\prime}$. Hence, pure exogenous censoring cannot be ignored here. In fact, pure exogenous censoring may not be ignorable even if $\mathbb{E}_{f^*}\{X^*\varepsilon^*\} = 0$ is replaced by the stronger condition $\mathbb{E}_{Y^*|X^*}\{\varepsilon^*|X^*\} = 0$ w.p.1[6]. However, as shown in Example 3.4, the situation changes if X^* is truncated instead of censored. □

Example 3.3 (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 $\mathbb{E}_f\{Z\} = \mathbb{E}_{f^*}[Z^*\mathbb{I}(Z^* \in T)]/\int_T f^*(z)\, d\mu^*$, as in Example 3.1 the naive estimator is not consistent for $\mathbb{E}_{f^*}\{Z^*\}$. □

Example 3.4 (Truncated linear regression)**.** Consider the linear model of Example 3.2, 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

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

[6]As an illustration, let $Y^* = X^*\theta^* + \varepsilon^*$ where X^* is scalar, $\mathbb{E}_{Y^*|X^*}\{\varepsilon^*|X^*\} = 0$ w.p.1, and $c = (\infty, c^{(2)})_{2\times 1}$. Then

$$\text{plim}(\hat{\theta}_M) = \frac{\mathbb{E}_f\{XY\}}{\mathbb{E}_f\{X^2\}} \overset{(3.1)}{=} \frac{\mathbb{E}_{f^*}\{X^*Y^*\mathbb{I}(X^* < c^{(2)}) + Y^* c^{(2)} \mathbb{I}(X^* > c^{(2)})\}}{\mathbb{E}_{f^*}\{X^{*2}\mathbb{I}(X^* < c^{(2)}) + [c^{(2)}]^2 \mathbb{I}(X^* > c^{(2)})\}} = \frac{\mathbb{E}_{f^*}\{X^{*2}\mathbb{I}(X^* < c^{(2)}) + X^* c^{(2)} \mathbb{I}(X^* > c^{(2)})\}}{\mathbb{E}_{f^*}\{X^{*2}\mathbb{I}(X^* < c^{(2)}) + [c^{(2)}]^2 \mathbb{I}(X^* > c^{(2)})\}} \theta^*,$$

where the last equality follows because $\mathbb{E}_{Y^*|X^*}\{Y^*|X^*\} = X^*\theta^*$ w.p.1. Therefore, provided $c^{(2)} \neq 0$, $\tilde{\theta}_M$ remains inconsistent under pure exogenous censoring even when ε^* is mean independent of X^*.

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

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

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

$$\text{plim}(\hat{\theta}_M) = \{\mathbb{E}_{f^*} X^* X^{*\prime} \mathbb{I}(X^* \in T_2)\}^{-1} \mathbb{E}_{f^*}\{X^* Y^* \mathbb{I}(X^* \in T_2)\} \neq \theta^*. \tag{3.4}$$

Therefore, even pure exogenous truncation is not ignorable. But, unlike Example 3.2, if the identifying assumption $\mathbb{E}_{f^*}\{X^*\varepsilon^*\} = 0$ is replaced by $\mathbb{E}_{Y^*|X^*}\{\varepsilon^*|X^*\} = 0$ w.p.1, then from (3.4) it is easy to see that ignoring pure exogenous truncation does not make the least squares estimator inconsistent. □

4. Inference with censored data

4.1. **Efficient estimation.** Since $\int_{r\in\{0,1\}} f_e(z,r)\, d\kappa(r) = K_0 f^*(z)\mathbb{I}(z \overset{elt}{\neq} c) + (1-K_0) f(z)$ by (2.6), from (2.2)–(2.3) it follows that

$$f^*(z)\mathbb{I}(z \overset{elt}{\neq} c) = \frac{\mathbb{I}(z \overset{elt}{\neq} c)}{K_0 + (1-K_0)\mathbb{I}(z \overset{elt}{<} c)} \int_{r\in\{0,1\}} f_e(z,r)\, d\kappa(r). \tag{4.1}$$

Hence, a μ^*-version of f^* can be recovered in terms of f_e alone. Next, since $\mathbb{E}_{f^*}\{g(Z^*,\theta^*)\} = 0$ if and only if $\mathbb{E}_{f^*}[g(Z^*,\theta^*)\mathbb{I}(Z^* \overset{elt}{\neq} c)] = 0$, use (4.1) to rewrite (2.1) in terms of the enriched density as[7]

$$\mathbb{E}_{f_e}\{g(Z,\theta^*)\mathbb{I}(Z \overset{elt}{\neq} c)/a(Z,K_0)\} = 0, \tag{4.2}$$

where $a(Z,K_0) = K_0 + (1-K_0)\mathbb{I}(Z \overset{elt}{<} c)$. However, (4.1) also implies that

$$\mathbb{E}_{f_e}\{\mathbb{I}(Z \overset{elt}{\neq} c)/a(Z,K_0)\} = 1 \iff \mathbb{E}_{f_e}\{\mathbb{I}(Z \overset{elt}{\neq} c)\mathbb{I}(Z \overset{elt}{<} c)^{\sim} - K_0\mathbb{I}(Z \overset{elt}{<} c)^{\sim}\} = 0, \tag{4.3}$$

where $(Z \overset{elt}{<} c)^{\sim}$ is the set-complement of the event $(Z \overset{elt}{<} c)$. Furthermore, since $\mathbb{E}_{f_e}\{R - K_0\} = 0$, efficient estimation of θ^* must account for this restriction as well.

So, for a fixed (θ, K), if we let

$$\rho(Z,R,\theta,K) = \begin{bmatrix} g(Z,\theta)\mathbb{I}(Z \overset{elt}{\neq} c)/a(Z,K) \\ \mathbb{I}(Z \overset{elt}{\neq} c)\mathbb{I}(Z \overset{elt}{<} c)^{\sim} - K\mathbb{I}(Z \overset{elt}{<} c)^{\sim} \\ R - K \end{bmatrix} \overset{def}{=} \begin{bmatrix} \rho_1(Z,\theta,K) \\ \rho_2(Z,K) \\ \rho_3(R,K) \end{bmatrix}_{(q+2)\times 1}, \tag{4.4}$$

the optimal GMM estimator of (θ^*, K_0) is given by $(\tilde{\theta}_{gmm}, \tilde{K}_{gmm}) = \text{argmin}_{(\theta,K)\in\Theta\times[0,1]} \text{GMM}(\theta,K)$, where

$$\text{GMM}(\theta,K) = \hat{\rho}'(\theta,K)\hat{\Upsilon}^{-1}\hat{\rho}(\theta,K), \tag{4.5}$$

$\hat{\rho}(\theta,K) = \sum_{j=1}^n \rho(Z_j,R_j,\theta,K)/n$ and $\hat{\Upsilon} = \sum_{j=1}^n \rho(Z_j,R_j,\theta^\dagger,K^\dagger)\rho'(Z_j,R_j,\theta^\dagger,K^\dagger)/n$ is the optimal weighting matrix constructed using some preliminary estimator $(\theta^\dagger, K^\dagger)$.

---

[7] Hence, a sufficient condition for θ^* to be *locally* identified is that $\mathbb{E}_{f_e}\{[\mathbb{I}(Z \overset{elt}{\neq} c)/a(Z,K_0)]\partial g(Z,\theta^*)/\partial\theta\}$, the $q \times p$ Jacobian matrix, exists and is of full column rank. In some models, especially those linear in θ^*, this may be simpler than employing Chamberlain (1986) type "identification at infinity" arguments or imposing Powell (1984, 1986a) type "sign" restrictions on the regression function to identify θ^*.

The following assumption, which is maintained throughout the paper, ensures that the GMM and EL estimators (described subsequently) are consistent and asymptotically normal. Let $\|\cdot\|$ denote the Euclidean norm and $B(\theta,\delta)$ an open ball with center θ and radius δ.

Assumption 4.1. (i) $K_0 \in (0,1)$*;* (ii) Θ *is compact;* (iii) $\theta^* \in int(\Theta)$ *is the unique root of* (2.1)*;* (iv) $g(Z,\theta)$ *is continuous on* Θ *w.p.1; There exist* $\eta > 0$ *and* $\delta > 0$ *such that the following conditions hold:* (v) $\mathbb{E}_{f^*}\{\sup_{\theta\in\Theta}\|g(Z,\theta)\|^{2(1+\eta)}\} < \infty$*;* (vi) $g(Z,\theta)$ *is twice continuously differentiable on* $B(\theta^*,\delta)$ *w.p.1;* (vii) $\mathbb{E}_{f^*}\{\sup_{\theta\in B(\theta^*,\delta)}\|\partial g(Z,\theta)/\partial\theta\|\} < \infty$*;* (viii) $\mathbb{E}_{f^*}\{\sup_{\theta\in B(\theta^*,\delta)}|\partial^2 g^{(i)}(Z,\theta)/\partial\theta^{(j)}\partial\theta^{(k)}|\} < \infty$ *for* $i = 1,\ldots,q$ *and* $j,k = 1,\ldots,p$.

(i) rules out the uninteresting case since if $K_0 = 1$ the censoring or truncation problem vanishes. (iii) ensures that θ^* in (2.1) is globally identified. (i)–(viii) 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).

Let $D = \mathbb{E}_{f_e}\{\partial\rho_1(Z,\theta^*,K_0)/\partial\theta\}$, $V_1 = \mathbb{E}_{f_e}\{\rho_1(Z,\theta^*,K_0)\rho_1(Z,\theta^*,K_0)'\}$, $V_2 = \mathbb{E}_{f_e}\{\rho_2^2(Z,K_0)\}$, $V_3 = \mathbb{E}_{f_e}\{\rho_3^2(R,K_0)\}$, $\Sigma_{12} = \mathbb{E}_{f_e}\{\rho_1(Z,\theta^*,K_0)\rho_2(Z,K_0)\}$, $\Sigma_{13} = \mathbb{E}_{f_e}\{\rho_1(Z,\theta^*,K_0)\rho_3(R,K_0)\}$. Then, defining $\varepsilon = \rho_1(Z,\theta^*,K_0) - \mathrm{Proj}_{f_e}\{\rho_1(Z,\theta^*,K_0)\big|1,\rho_2(Z,K_0),\rho_3(R,K_0)\}$ to be the residual from orthogonally projecting $\rho_1(Z,\theta^*,K_0)$ onto the linear span of $\{1,\rho_2(Z,K_0),\rho_3(Z,R,K_0)\}$ using the inner product $\langle a,b\rangle_{f_e} = \mathbb{E}_{f_e}\{a'b\}$, we can show that[8]

Theorem 4.1. $\begin{bmatrix} n^{1/2}(\tilde\theta_{gmm}-\theta^*) \\ n^{1/2}(\tilde K_{gmm}-K_0)\end{bmatrix} \xrightarrow{d} \mathrm{N}(0_{(p+1)\times 1}, \begin{bmatrix} (D'\Omega^{-1}D)^{-1} & 0_{p\times 1} \\ 0'_{p\times 1} & K_0(1-K_0)\end{bmatrix})$*, where* $\Omega = \mathbb{E}_{f_e}\{\varepsilon\varepsilon'\}$.

Since it is well known that optimally weighted GMM estimators are efficient as the sample size goes to infinity (Chamberlain 1987), it follows that $\tilde\theta_{gmm}$ and $\tilde K_{gmm}$ are asymptotically efficient. In Theorem 4.3, we show that $(D'\Omega^{-1}D)^{-1}$ is strictly smaller (in the positive definite sense) than the asymptotic variance of the GMM estimator obtained by using the refreshment sample alone. Hence, efficiency gains from combining censored and uncensored datasets do not come from the latter alone and it makes sense to use *both* the master and the refreshment samples for estimating θ^*.

There is a simpler version of (4.4) that still leads to an asymptotically efficient estimator of θ^*; i.e., an estimator whose asymptotic variance is equal to $(D'\Omega^{-1}D)^{-1}$. This follows because

$$\mathrm{Proj}_{f_e}\{\rho_1(Z,\theta^*,K_0)\big|1,\rho_2(Z,K_0),\rho_3(R,K_0)\} \overset{\text{Lemma A.1}}{=} \mathrm{Proj}_{f_e}\{\rho_1(Z,\theta^*,K_0)\big|1,\rho_2(Z,K_0)\}, \tag{4.6}$$

i.e., $\rho_3(R,K_0)$ is redundant once $\rho_2(Z,K_0)$ is controlled for, suggesting that the asymptotic variance of the GMM estimator of θ^* given in Theorem 4.1 is not affected if only $\rho_1(Z,\theta^*,K_0)$ and $\rho_2(Z,K_0)$ are used for estimation; i.e., even if we ignore the information regarding whether Z comes from the refreshment or the master sample. *Therefore, for the remainder of Section* 4 *we assume that* θ^* *and* K_0 *are estimated using*

$$\rho(Z,\theta,K) = \begin{bmatrix} g(Z,\theta)\mathbb{I}(Z \overset{elt}{\neq} c)/a(Z,K) \\ \mathbb{I}(Z \overset{elt}{\neq} c)\mathbb{I}(Z \overset{elt}{<} c)^\sim - K\mathbb{I}(Z \overset{elt}{<} c)^\sim\end{bmatrix} = \begin{bmatrix}\rho_1(Z,\theta,K) \\ \rho_2(Z,K)\end{bmatrix}_{(q+1)\times 1}. \tag{4.7}$$

[8] We use $0_{k\times 1}$ to denote a $k\times 1$ vector of zeros; $0'_{k\times 1}$ is its transpose.

This leads to the following result.

Theorem 4.2. *Let* $(\hat{\theta}_{gmm}, \hat{K}_{gmm})$ *denote the GMM estimator of* (θ^*, K_0) *using* (4.7). *Then*[9]

$$\begin{bmatrix} n^{1/2}(\hat{\theta}_{gmm} - \theta^*) \\ n^{1/2}(\hat{K}_{gmm} - K_0) \end{bmatrix} \xrightarrow{d} \mathrm{N}(0_{(p+1)\times 1}, \begin{bmatrix} (D'\Omega^{-1}D)^{-1} & 0_{p\times 1} \\ 0'_{p\times 1} & K_0(1-K_0)/[1-F^*(c)] \end{bmatrix}.$$

The asymptotic variance of $\hat{\theta}_{gmm}$ is still $(D'\Omega^{-1}D)^{-1}$ although dropping $\rho_3(R, K_0)$ increases the asymptotic variance of $\hat{K}_{gmm}$ as compared to $\tilde{K}_{gmm}$. This is not surprising since $\rho_3(R, K_0)$ provides information about K_0 and does not matter in practice since K_0 is a nuisance parameter. Another practical advantage of dropping $\rho_3(R, K_0)$ is that since (4.7) just identifies K_0, using the $\rho(Z, \theta, K)$ defined in (4.7) ensures that $\mathbb{E}_{f^*}\{g(Z^*, \theta^*)\} = 0$ if and only if $\mathbb{E}_{f_e}\{\rho(Z, \theta^*, K_0)\} = 0$. Therefore, as shown in Section 4.3, we can also base a specification test for (2.1) on (4.7). This cannot not be done with (4.4) since there K_0 is overidentified. Incidentally, since (4.6) implies that ε is now the residual from projecting $\rho_1(Z, \theta^*, K_0)$ onto the span of $\{1, \rho_2(Z, K_0)\}$, it follows that $\Omega = V_1 - \Sigma_{12}\Sigma'_{12}/V_2$.

Now, let $\hat{\theta}_R$ denote the optimal GMM estimator of θ^* obtained using *only* the refreshment sample; i.e., $\hat{\theta}_R$ is based on the moment condition

$$\mathbb{E}_{f_e}\{g(Z, \theta^*)|R = 1\} = 0 \iff \mathbb{E}_{f_e}\{g(Z, \theta^*)R\} = 0. \tag{4.8}$$

The next result shows that $\hat{\theta}_R$ is asymptotically inefficient relative to $\hat{\theta}_{gmm}$. Therefore, as stressed earlier, it makes sense to estimate θ^* using the enriched sample.

Theorem 4.3. *Let* $D_* = \mathbb{E}_{f^*}\{\partial g(Z, \theta^*)/\partial\theta\}$ *and* $V_* = \mathbb{E}_{f^*}\{g(Z, \theta^*)g'(Z, \theta^*)\}$. *Then*

$$n^{1/2}(\hat{\theta}_R - \theta^*) \xrightarrow{d} \mathrm{N}(0_{p\times 1}, (D'_* V_*^{-1} D_*)^{-1}/K_0)$$

and $\mathrm{asvar}(\hat{\theta}_R) > \mathrm{asvar}(\hat{\theta}_{gmm})$, *where "*asvar*" is shorthand for "asymptotic variance".*

The inflation factor $1/K_0$ appearing in the asymptotic variance of $\hat{\theta}_R$ is not surprising since $\hat{\theta}_R$ only makes use of a fraction of the enriched sample.

Before proceeding any further, we now provide some intuition behind how transforming the moment condition allows us to impute the censored values. So, using the fact that, by (4.3), $K_0 = \mathbb{E}_{f_e}\{\mathbb{I}(Z \overset{elt}{\neq} c)\mathbb{I}(Z \overset{elt}{<} c)^\sim\}/\mathbb{E}_{f_e}\{\mathbb{I}(Z \overset{elt}{<} c)^\sim\}$, notice that we can decompose

$$\begin{aligned}\mathbb{E}_{f_e}\{\rho_1(Z, \theta^*, K_0)\} = \mathbb{E}_{f_e}\{g(Z, \theta^*)|Z \overset{elt}{<} c\}\mathrm{Pr}_{f_e}(Z \overset{elt}{<} c) \\ + \mathbb{E}_{f_e}\{g(Z, \theta^*)|(Z \overset{elt}{\neq} c) \cap (Z \overset{elt}{<} c)^\sim\}\mathrm{Pr}_{f_e}(\{Z \overset{elt}{<} c\}^\sim).\end{aligned} \tag{4.9}$$

Therefore, the moment condition in (4.2) can be expressed as a weighted sum of the best predictors of $g(Z^*, \theta^*)|(Z^*$ is uncensored) and $g(Z^*, \theta^*)|(Z^*$ is censored), with the weights being equal to the

[9]For the sake of completeness, note that if $(\check{\theta}, \check{K})$ is the GMM estimator of (θ^*, K_0) based on $\rho_1(Z, \theta^*, K_0)$ and $\rho_3(R, K_0)$, then it is easy to show that asymptotically $n^{1/2}(\check{\theta} - \theta^*)$ and $n^{1/2}(\check{K} - K_0)$ are jointly normal with mean zero and variance $\begin{bmatrix} (D'\Gamma^{-1}D)^{-1} & 0_{p\times 1} \\ 0'_{p\times 1} & K_0(1-K_0) \end{bmatrix}$, where $\Gamma = V_1 - \Sigma_{13}\Sigma'_{13}/V_3$. From Lemma A.2, (A.10), and (A.11) we know that $V_2 = K_0(1-K_0)[1-F^*(c)]$ and $\Sigma_{13} = \Sigma_{12}$. Hence, since $V_3 = K_0(1-K_0)$, it follows that $\Gamma \geq \Omega$ (i.e.,$\Gamma - \Omega$ is positive semidefinite). Therefore, $(D'\Gamma^{-1}D)^{-1} \geq (D'\Omega^{-1}D)^{-1}$ implying that asymptotically $\hat{\theta}_{gmm}$ is better than $\check{\theta}$.

probability that Z^* is uncensored or censored, respectively. The estimators proposed in Theorem 4.2 use the enriched sample to automatically replace $g(Z^*, \theta^*)$ with its best predictor when observations are censored and then consistently and efficiently estimate these best predictors and selection probabilities; see Example 4.1 for a nice illustration.

Next, we show how to estimate θ^* by EL. As for GMM, we base EL estimation on (4.7). So let p_j denote the probability mass placed at the jth observation by a discrete distribution that has support on the realized observations. For fixed (θ, K) concentrate out the p_j's by solving the nonparametric maximum likelihood problem $\max_{p_1,\ldots,p_n} \sum_{j=1}^n \log p_j$ subject to the constraints that the p_j's are nonnegative, $\sum_{j=1}^n p_j = 1$, and $\sum_{j=1}^n \rho(Z_j, \theta, K) p_j = 0$. The solution to this problem is given by $\hat{p}_j(\theta, K) = n^{-1}\{1 + \lambda'(\theta, K)\rho(Z_j, \theta, K)\}^{-1}$ for $j = 1, \ldots, n$, where the Lagrange multiplier $\lambda(\theta, K)$ satisfies $\sum_{j=1}^n \rho(Z_j, \theta, K)\{1 + \lambda'(\theta, K)\rho(Z_j, \theta, K)\}^{-1} = 0$. We define the empirical likelihood estimator of (θ^*, K_0) as $(\hat{\theta}_{el}, \hat{K}_{el}) = \text{argmax}_{(\theta,K)\in\Theta\times[0,1]} \text{EL}(\theta, K)$, where

$$\text{EL}(\theta, K) = \sum_{j=1}^n \log \hat{p}_j(\theta, K) = -\sum_{j=1}^n \log\{1 + \lambda'(\theta, K)\rho(Z_j, \theta, K)\} - n \log n. \tag{4.10}$$

Consistency of $\hat{\theta}_{el}$ and $\hat{K}_{el}$ follows from Kitamura (1997, Theorem 1). By standard EL theory as in Qin and Lawless (1994), EL and GMM estimators have the same asymptotic distribution. Hence, the EL estimator of θ^* is also asymptotically efficient. In finite samples the GMM and EL estimators will be different, though the two coincide if θ^* is just identified (i.e., $q = p$) because then the EL probabilities $\hat{p}_j(\theta, K) = 1/n$ for each j, θ, and K.

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 (2.1) 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 $F^*(\xi)$ and $F_e(\xi) = \Pr_{f_e}(Z \overset{elt}{\leq} \xi)$, the target and realized cdf's, respectively, are easily estimated. Let $\hat{F}^*(\xi) = \sum_{j=1}^n \hat{p}_j(\hat{\theta}_{el}, \hat{K}_{el}) \mathbb{I}(Z_j \overset{elt}{\leq} \xi) \mathbb{I}(Z_j \overset{elt}{\neq} c)/a(Z_j, \hat{K}_{el})$ and $\hat{F}(\xi) = \sum_{j=1}^n \hat{p}_j(\hat{\theta}_{el}, \hat{K}_{el}) \mathbb{I}(Z_j \overset{elt}{\leq} \xi)$, where the $\hat{p}_j(\hat{\theta}_{el}, \hat{K}_{el})$'s denote EL probabilities evaluated at $(\hat{\theta}_{el}, \hat{K}_{el})$. Then, letting $M_\Omega = \Omega^{-1} - \Omega^{-1}D(D'\Omega^{-1}D)^{-1}D'\Omega^{-1}$, $I(Z, \xi) = \mathbb{I}(Z \overset{elt}{\leq} \xi) - F^*(\xi)$, and $u = I(Z, \xi)\mathbb{I}(Z \overset{elt}{\neq} c)/a(Z, K_0) - \text{Proj}_{f_e}\{I(Z, \xi)\mathbb{I}(Z \overset{elt}{\neq} c)/a(Z, K_0) \big| 1, \rho_2(Z, K_0)\}$, we have that

Theorem 4.4. *$n^{1/2}\{\hat{F}^*(\xi) - F^*(\xi)\}$ is asymptotically normal with mean zero and variance*

$$\mathbb{E}_{f_e}\{u^2\} - \mathbb{E}_{f_e}\{\varepsilon' u\} M_\Omega \mathbb{E}_{f_e}\{\varepsilon u\}.$$

Theorem 4.5. *$n^{1/2}\{\hat{F}(\xi) - F_e(\xi)\}$ is asymptotically normal with mean zero and variance*

$$F_e(\xi)\{1 - F_e(\xi)\} - \mathbb{E}_{f_e}\{\varepsilon' \mathbb{I}(Z \overset{elt}{\leq} \xi)\} M_\Omega \mathbb{E}_{f_e}\{\varepsilon \mathbb{I}(Z \overset{elt}{\leq} \xi)\}.$$

Hence, 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 4, let $(\hat{\theta}, \hat{K})$ denote the GMM or EL estimator of (θ^*, K_0). The asymptotic variances of $\hat{\theta}$, $\hat{F}(\xi)$, and $\hat{F}^*(\xi)$ can be estimated in the obvious manner by replacing D and Ω with consistent estimators $\hat{D} = n^{-1}\sum_{j=1}^n \partial\rho_1(Z_j, \hat{\theta}, \hat{K})/\partial\theta$ and $\hat{\Omega} = \hat{V}_1 - \hat{\Sigma}_{12}\hat{\Sigma}_{12}'/\hat{V}_2$, where $\hat{V}_1 = \sum_{j=1}^n \rho_1(Z_j, \hat{\theta}, \hat{K})\rho_1'(Z_j, \hat{\theta}, \hat{K})/n$, $\hat{\Sigma}_{12} = \sum_{j=1}^n \rho_1(Z_j, \hat{\theta}, \hat{K})\rho_2(Z_j, \hat{K})/n$, and $\hat{V}_2 = \sum_{j=1}^n \rho_2^2(Z_j, \hat{K})/n$. Equivalently, $\hat{\Omega} = \sum_{j=1}^n \hat{\varepsilon}\hat{\varepsilon}'/n$, where $\hat{\varepsilon}$ is the residual from regressing $\rho_1(Z, \hat{\theta}, \hat{K})$ element-by-element on a constant and $\rho_2(Z, \hat{K})$.

Example 4.1 (Example 3.1 contd.)**.** Here $\rho_1(Z, \theta, K) = (Z - \theta)\mathbb{I}(Z \overset{elt}{\neq} c)/a(Z, K)$ and, since there are no overidentifying restrictions, $(\hat{\theta}, \hat{K})$ solve $\sum_{j=1}^n \rho_1(Z_j, \hat{\theta}, \hat{K}) = 0$ and $\sum_{j=1}^n \rho_2(Z_j, \hat{K}) = 0$; i.e.,

$$\hat{\theta} = n^{-1}\sum_{j=1}^n \frac{Z_j \mathbb{I}(Z_j \overset{elt}{\neq} c)}{a(Z_j, \hat{K})} \quad \text{and} \quad \hat{K} = \frac{\sum_{j=1}^n \mathbb{I}(Z_j \overset{elt}{\neq} c)\mathbb{I}(Z_j \overset{elt}{<} c)^\sim}{\sum_{j=1}^n \mathbb{I}(Z_j \overset{elt}{<} c)^\sim}. \tag{4.11}$$

To gain further insight into $\hat{\theta}$, notice that for $d = 1$ we can write

$$\hat{\theta} = n^{-1}\sum_{j=1}^n \mathbb{I}(Z_j < c) \times \frac{\sum_{j=1}^n Z_j \mathbb{I}(Z_j < c)}{\sum_{j=1}^n \mathbb{I}(Z_j < c)} + n^{-1}\sum_{j=1}^n \mathbb{I}(Z_j \geq c) \times \frac{\sum_{j=1}^n Z_j \mathbb{I}(Z_j > c)}{\sum_{j=1}^n \mathbb{I}(Z_j > c)}.$$

In light of (4.9), it comes as no surprise that $\hat{\theta}$ is a convex combination of the sample means of uncensored and censored observations in the enriched dataset with the weights being the fraction of uncensored and censored observations in the enriched sample. By Theorem 4.2, $n^{1/2}(\hat{\theta} - \theta^*)$ converges in distribution to a normal random variable with mean zero and variance Ω. Since there are no overidentifying restrictions in this example (and the next one), $\hat{F}(\xi) = n^{-1}\sum_{j=1}^n \mathbb{I}(Z_j \overset{elt}{\leq} \xi)$ is just the empirical cdf of the realized observations and $\hat{F}^*(\xi) = n^{-1}\sum_{j=1}^n \mathbb{I}(Z_j \overset{elt}{\leq} \xi)\mathbb{I}(Z_j \overset{elt}{\neq} c)/a(Z_j, \hat{K})$. By Theorem 4.4, $n^{1/2}\{\hat{F}^*(\xi) - F^*(\xi)\}$ has asymptotic variance $\mathbb{E}_{f_e}\{u^2\}$. Similarly, by Theorem 4.5, the asymptotic variance of $n^{1/2}\{\hat{F}(\xi) - F_e(\xi)\}$ is simply $F_e(\xi)\{1 - F_e(\xi)\}$. □

Example 4.2 (Example 3.2 contd.)**.** Here, $\rho_1(Z, \theta, K) = X(Y - X'\theta)\mathbb{I}(Z \overset{elt}{\neq} c)/a(Z, K)$. Hence, $\hat{\theta} = (\sum_{j=1}^n \hat{X}_j X_j')^{-1}(\sum_{j=1}^n \hat{X}_j Y_j)$, where $\hat{X}_j = X_j \mathbb{I}(Z_j \overset{elt}{\neq} c)/a(Z_j, \hat{K})$ and $\hat{K}$ is given in (4.11). Notice that this tantamounts to replacing the original regressors in Example 3.2 with new instruments $\hat{X}$. If censoring is purely endogenous or purely exogenous, then $a(Z, K) = K + (1 - K)\mathbb{I}(Y_j < c^{(1)})$ or $a(Z, K) = K + (1 - K)\mathbb{I}(X_j \overset{elt}{<} c^{-(1)})$, respectively, and the expression for $\hat{\theta}$ simplifies. □

Example 4.3 (Censored linear regression with endogenous regressors)**.** Let $Y^* = X^{*\prime}\theta^* + \varepsilon^*$ such that some or all of the regressors are correlated with ε^*. We have a $q \times 1$ vector of instrumental variables W^* that are uncorrelated with ε^*; 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, $Z^* = (Y^*, X^*, \tilde{W}^*)$ and $g(Z^*, \theta^*) = W^*(Y^* - X^{*\prime}\theta^*)$. If the dependent variable, regressors, and instruments are all censored, then $\rho_1(Z, \theta, K) = W(Y - X'\theta)\mathbb{I}(Z \overset{elt}{\neq} c)/a(Z, K)$. On the other hand, the endogenous tobit model where only Y^* is censored and X^* is endogenous is very important for applications and follows by letting $c^{-(1)} = (\infty, \ldots, \infty)$ so that $\rho_1(Z, \theta, K) = W(Y - X'\theta)\mathbb{I}(Y \neq c^{(1)})/a(Y, K)$, where $a(Y, K) = K + (1-K)\mathbb{I}(Y < c^{(1)})$. In either case, θ^* can be estimated by GMM or EL as described earlier and the asymptotic distribution of $\hat{\theta}$ follows readily from Theorem 4.2. □

Example 4.4 (Censoring and IV)**.** Suppose $Y_1^* = X_1^{*\prime}\theta_1^* + \varepsilon_1^*$ and $Y_2^* = X_2^{*\prime}\theta_2^* + \varepsilon_2^*$, where $\varepsilon^* = (\varepsilon_1^*, \varepsilon_2^*)_{2\times 1}$ satisfies the conditional moment restriction $\mathbb{E}_{(Y_1^*, Y_2^*)|X^*}(\varepsilon^*|X^*) = 0$ w.p.1 and X^* denotes a vector containing the exogenous coordinates of $(X_1^*, X_2^*)_{(p_1+p_2)\times 1}$ and other instruments. Hence, $\mathbb{E}_{f^*}\{A(X^*)\begin{bmatrix} Y_1^* - X_1^{*\prime}\theta_1^* \\ Y_2^* - X_2^{*\prime}\theta_2^* \end{bmatrix}\} = 0$, where $A(X^*)$ is a $q \times 2$ matrix of instrumental variables such that $q \geq p_1 + p_2$. If $Z^* = (Y_1^*, Y_2^*, X^*)$ is censored, then (4.7) can be used to estimate $\theta^* = (\theta_1^*, \theta_2^*)$ by GMM or EL as proposed earlier. Although this simultaneous equations model has been studied before, see, e.g., Smith and Blundell (1986) and Blundell and Smith (1993), the treatment here is more general because we do not assume that ε^* is Gaussian and allow for the possibility that besides Y_1^* and Y_2^* the instruments may also be censored. Censoring of $Y^* = (Y_1^*, Y_2^*)$ alone means that $c^{-(1,2)} = (\infty, \ldots, \infty)$. Hence, $\rho_1(Z, \theta, K) = A(X)\begin{bmatrix} Y_1 - X_1'\theta_1 \\ Y_2 - X_2'\theta_2 \end{bmatrix}\mathbb{I}(Y_1^* \neq c^{(1)}, Y_2^* \neq c^{(2)})/a(Y, K)$, where $a(Y, K) = K + (1-K)\mathbb{I}(Y_1 < c^{(1)}, Y_2 < c^{(2)})$, and θ^* can be estimated by GMM or EL as before. □

Example 4.5. 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. Moment based auxiliary information about f^* can be easily incorporated in our framework by first stacking $g(Z^*, \theta^*)$ and $m(Z^*)$ to produce an augmented vector of moment conditions and then proceeding as before. 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). However, Imbens and Lancaster (1994) and Hellerstein and Imbens (1999) assume that Z^* is fully observed. Nevo (2003) allows Z^* to be entirely missing (due to attrition) but not censored. He also restricts attention to the case where the parameter of interest is just identified. In addition, he assumes that the selection probability is known up to a finite dimensional parameter and imposes an identification condition that rules out truncated Z^*'s as well. By contrast, we allow (2.1) to be overidentified and the selection probabilities for censoring or truncation of Z^* to be fully unknown. None of these papers discuss efficient estimation of the target or realized cdf's. □

4.2. **Hypothesis tests and confidence regions.** Suppose we want to test the parametric restriction $H(\theta^*) = 0$ against the alternative that it is false, where H is a $h \times 1$ vector of twice continuously differentiable functions such that $\partial H(\theta^*)/\partial\theta$ has rank $h \leq p$. Since $n^{1/2}(\hat{\theta} - \hat{\theta}^*) \xrightarrow{d} \mathrm{N}(0, (D'\Omega^{-1}D)^{-1})$, the Wald statistic $\mathrm{W} = nH'(\hat{\theta})\{\frac{\partial H(\hat{\theta})}{\partial\theta}(\hat{D}'\hat{\Omega}^{-1}\hat{D})^{-1}\frac{\partial H'(\hat{\theta})}{\partial\theta}\}^{-1}H(\hat{\theta}) \xrightarrow{d} \chi_h^2$ under the null hypothesis.

Another alternative is to use a distance metric test. So let $(\bar{\theta}, \bar{K})$ denote the GMM and EL estimators under the null; i.e., $(\bar{\theta}_{gmm}, \bar{K}_{gmm}) = \text{argmin}_{\{(\theta,K)\in\Theta\times[0,1]:R(\theta)=0\}} \text{GMM}(\theta, K)$ and $(\bar{\theta}_{el}, \bar{K}_{el}) = \text{argmax}_{\{(\theta,K)\in\Theta\times[0,1]:R(\theta)=0\}} \text{EL}(\theta, K)$. Next, let $\text{DM} = n\{\text{GMM}(\bar{\theta}_{gmm}, \bar{K}_{gmm}) - \text{GMM}(\hat{\theta}_{gmm}, \hat{K}_{gmm})\}$ and $\text{LR} = 2\{\text{EL}(\hat{\theta}_{el}, \hat{K}_{el}) - \text{EL}(\bar{\theta}_{el}, \bar{K}_{el})\}$, where DM denotes the GMM based distance metric statistic and LR the EL based likelihood ratio test statistic. A test for $H(\theta^*) = 0$ can also 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 $\text{DM} \xrightarrow{d} \chi^2_h$ and $\text{LR} \xrightarrow{d} \chi^2_h$ under the null.

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; e.g., although all three statistics can be inverted to obtain asymptotically valid confidence regions, the DM and LR based regions are invariant to the formulation of $\tilde{\text{H}}_0$ 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.

4.3. **Specification tests.** For the remainder of this section, assume that $q > p$. Since inference based on the estimated θ^* is sensible only if (2.1) 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 $n\text{GMM}(\hat{\theta}_{gmm}, \hat{K}_{gmm}) \xrightarrow{d} \chi^2_{q-p}$ under H_0. Therefore, a test for overidentifying restrictions (usually called the J-test) in (2.1) 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 $\hat{\theta}$ and $\hat{K}$ denote $n^{1/2}$-consistent preliminary estimators of θ^* and K_0; e.g., $\hat{\theta}$ and $\hat{K}$ can be the GMM or EL estimators defined previously. The restricted (i.e., under H_0) EL can be written as $\text{EL}^r = \sum_{j=1}^n \log \hat{p}_j(\hat{\theta}, \hat{K})$. 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 $\text{EL}^{ur} = -n \log n$. Now let $\text{ELR} = 2(\text{EL}^{ur} - \text{EL}^r) = 2\sum_{j=1}^n \log\{1 + \lambda'(\hat{\theta}, \hat{K})\rho(Z_j, \hat{\theta}, \hat{K})\}$. Then ELR can be regarded as an analog of the usual parametric likelihood ratio test statistic; i.e., H_0 is rejected if ELR is large enough. By Qin and Lawless (1994, Corollary 4), $\text{ELR} \xrightarrow{d} \chi^2_{q-p}$ under H_0. Hence, critical values for ELR are easily obtained.

5. Inference with truncated data

5.1. **Efficient estimation.** We now show how enriched data can be used to efficiently estimate models where variables are truncated. So let $b^* = \int_T f^*(z)\, d\mu^* \in (0, 1)$ denote the probability that Z^* is observed. Although b^* is unknown, we can use the refreshment sample to identify it via the moment condition

$$b^* = \mathbb{E}_{f_e}\{\mathbb{I}(Z \in T)|R = 1\} \iff \mathbb{E}_{f_e}\{\mathbb{I}(Z \in T) - b^*\}R = 0.$$

Next, since (2.4) and (2.6) imply that

$$f^*(z) = \frac{\int_{r\in\{0,1\}} f_e(z,r)\, d\kappa(r)}{K_0 + (1-K_0)\mathbb{I}(z \in T)/b^*}, \tag{5.1}$$

we can rewrite (2.1) in terms of the enriched density as

$$\mathbb{E}_{f_e}\{\frac{g(Z,\theta^*)}{K_0 + (1-K_0)\mathbb{I}(Z \in T)/b^*}\} = 0.$$

To estimate (θ^*, b^*, K_0), let $a(Z,b,K) = K + (1-K)\mathbb{I}(Z \in T)/b$ and define

$$\rho(Z,R,\theta,b,K) = \begin{bmatrix} g(Z,\theta)/a(Z,,b,K) \\ [\mathbb{I}(Z \in T) - b]R \\ R - K \end{bmatrix} \stackrel{def}{=} \begin{bmatrix} \rho_1(Z,\theta,b,K) \\ \rho_2(Z,R,b) \\ \rho_3(R,K) \end{bmatrix}_{(q+2)\times 1}. \tag{5.2}$$

By (5.1), $\mathbb{E}_{f^*}\{g(Z^*,\theta^*)\} = 0$ if and only if $\mathbb{E}_{f_e}\{\rho(Z,R,\theta^*,b^*,K_0)\} = 0$. Hence, (θ^*, b^*, K_0) can be jointly and efficiently estimated by using the latter moment condition. In particular, the GMM and EL estimators are given by $(\hat\theta_{gmm}, \hat b_{gmm}, \hat K_{gmm}) = \text{argmin}_{(\theta,b,K)\in\Theta\times[0,1]\times[0,1]}\, \text{GMM}(\theta,b,K)$ and $(\hat\theta_{el}, \hat b_{el}, \hat K_{el}) = \text{argmax}_{(\theta,b,K)\in\Theta\times[0,1]\times[0,1]}\, \text{EL}(\theta,b,K)$, respectively, where the objective functions are obtained analogously from (4.5) and (4.10) by replacing K with (b,K).

Since $\Pr_{f_e}\{Z \in T\} = K_0 b^* + 1 - K_0$, the logic behind the imputation mechanism becomes clear upon observing that

$$\mathbb{E}_{f_e}\{\rho_1(Z,\theta^*,b^*,K_0)\} = b^*\mathbb{E}_{f_e}\{g(Z,\theta^*)|Z \in T\} + (1-b^*)\mathbb{E}_{f_e}\{g(Z,\theta^*)|Z \notin T\}; \tag{5.3}$$

i.e., the transformed moment condition combines the best predictors of $g(Z^*,\theta^*)|(Z^* \text{ is not truncated})$ and $g(Z^*,\theta^*)|(Z^* \text{ is truncated})$ weighted by probabilities of the corresponding events. As in the case of censoring, the estimators we propose automatically carry out this imputation in the enriched sample to efficiently estimate the parameters of interest.

Let $\varepsilon = \rho_1(Z,\theta^*,b^*,K_0) - \text{Proj}_{f_e}\{\rho_1(Z,\theta^*,b^*,K_0)\big|1,\rho_2(Z,R,b^*),\rho_3(R,K_0)\}$, $\alpha^* = K_0 b^* + 1 - K_0$, and $v = \varepsilon + (\alpha^*/b^*)\text{Proj}_{f_e}\{\rho_1(Z,\theta^*,b^*,K_0)\big|1,\rho_2(Z,R,b^*)\}$[10]. Analogous to the notation in Section 4.1, define $\Omega = \mathbb{E}_{f_e}\{\varepsilon\varepsilon'\}$, $D = \mathbb{E}_{f_e}\{\partial\rho_1(Z,\theta^*,b^*,K_0)/\partial\theta\}$, $V_2 = \mathbb{E}_{f_e}\{\rho_2^2(Z,R,b^*)\}$, $\Sigma_{12} = \mathbb{E}_{f_e}\{\rho_1(Z,\theta^*,b^*,K_0)\rho_2(Z,R,b^*)\}$, and let $(\hat\theta,\hat b,\hat K)$ denote the GMM or EL estimator of (θ^*,b^*,K_0) for the remainder of the paper. Then, letting $V = \mathbb{E}_{f_e}\{vv'\}$ and $\gamma = (K_0^2/V_2) + (K_0/V_2^2)(\alpha^*/b^*)\Sigma_{12}'\Omega^{-1}\Sigma_{12}$,

Theorem 5.1. *$n^{1/2}(\hat\theta - \theta^*)$, $n^{1/2}(\hat b - b^*)$, and $n^{1/2}(\hat K - K_0)$ converge jointly in distribution to a $(p+2)\times 1$ normal random vector with mean zero and variance-covariance matrix*

$$\begin{bmatrix} (D'V^{-1}D)^{-1} & -(K_0/V_2\gamma)(\alpha^*/b^*)(D'V^{-1}D)^{-1}D'\Omega^{-1}\Sigma_{12} & 0_{p\times 1} \\ -(K_0/V_2\gamma)(\alpha^*/b^*)\Sigma_{12}'\Omega^{-1}D(D'V^{-1}D)^{-1} & \{(K_0^2/V_2) + (K_0/V_2)^2(\alpha^*/b^*)^2\Sigma_{12}'M_\Omega\Sigma_{12}\}^{-1} & 0 \\ 0_{p\times 1}' & 0 & K_0(1-K_0) \end{bmatrix}.$$

Since ε is the residual from an orthogonal projection and $\Sigma_{23} = \mathbb{E}_{f_e}\{\rho_2(Z,R,b^*)\rho_3(R,K_0)\} = 0$, $\Omega = V_1 - \Sigma_{12}\Sigma_{12}'/V_2 - \Sigma_{13}\Sigma_{13}'/V_3$ and $V = \Omega + (\alpha^*/b^*)^2\Sigma_{12}\Sigma_{12}'/V_2$, where V_1, Σ_{13}, and V_3 are defined as in Section 4.1[11]; i.e., $V_1 = \mathbb{E}_{f_e}\{\rho_1(Z,\theta^*,b^*,K_0)\rho_1'(Z,\theta^*,b^*,K_0)\}$, $\Sigma_{13} = \mathbb{E}_{f_e}\{\rho_1(Z,\theta^*,b^*,K_0)\rho_3(R,K_0)\}$,

[10] The second term in v is an adjustment for the fact that b^* is being estimated.

[11] Further simplifications reveal that $V = V_1 + (1-K_0)(\alpha^*/b^*)\Sigma_{12}\Sigma_{12}'/K_0 b^{*2}$; see, e.g., the proof of Theorem 5.2.

and $V_3 = \mathbb{E}_{f_e}\{\rho_3^2(R, K_0)\}$. Following Chamberlain (1987), $\hat{\theta}$ and $\hat{b}$ are asymptotically efficient. Also, since $\hat{b}$ is obtained by using the refreshment sample alone just as $\hat{\theta}_R$ was, its asymptotic variance when there is no overidentification is given by $b^*(1-b^*)/K_0$ because $V_2 = K_0 b^*(1-b^*)$. Hence, as expected, overidentification leads to a better estimator of b^*.

The next result shows that $\hat{\theta}$, the GMM or EL estimator of θ^* proposed in Theorem 5.1, is asymptotically better than $\hat{\theta}_R$. Hence, even in the case of truncation, efficiency gains do not come solely from the refreshment sample; i.e., truncated datasets possess information that we exploit to increase efficiency.

Theorem 5.2. *As in Section* 4.1*, let $\hat{\theta}_R$ denote the estimator of θ^* obtained by using the refreshment sample alone; i.e., $\hat{\theta}_R$ is based on the moment condition in* (4.8)*. Then,* $\text{asvar}(\hat{\theta}_R) > \text{asvar}(\hat{\theta})$.

Now we consider cdf estimation. So let $\hat{F}^*(\xi) = \sum_{j=1}^n \hat{p}_j(\hat{\theta}_{el}, \hat{b}_{el}, \hat{K}_{el}) \mathbb{I}(Z_j \overset{elt}{\leq} \xi)/a(Z_j, \hat{b}_{el}, \hat{K}_{el})$ and $\hat{F}(\xi) = \sum_{j=1}^n \hat{p}_j(\hat{\theta}_{el}, \hat{b}_{el}, \hat{K}_{el}) \mathbb{I}(Z_j \overset{elt}{\leq} \xi)$. Then, letting $I(Z, \xi) = \mathbb{I}(Z \overset{elt}{\leq} \xi) - F^*(\xi)$ and $u = I(Z, \xi)/a(Z, b^*, K_0) - \text{Proj}_{f_e}\{I(Z, \xi)/a(Z, b^*, K_0) \big| 1, \rho_2(Z, R, b^*), \rho_3(R, K_0)\}$, we can show that

Theorem 5.3. $n^{1/2}\{\hat{F}^*(\xi) - F^*(\xi)\}$ *is asymptotically normal with mean zero and variance*

$$\mathbb{E}_{f_e}\{w^2\} - \mathbb{E}_{f_e}\{v'w\} M_V \mathbb{E}_{f_e}\{vw\},$$

where $w = u + (\alpha^*/b^*)\text{Proj}_{f_e}\{I(Z, \xi) \big| 1, \rho_2(Z, b^*)\}$ *and* $M_V = V^{-1} - V^{-1}D(D'V^{-1}D)^{-1}D'V^{-1}$.

Theorem 5.4. $n^{1/2}\{\hat{F}(\xi) - F_e(\xi)\}$ *is asymptotically normal with mean zero and variance*

$$F_e(\xi)\{1 - F_e(\xi)\} - \mathbb{E}_{f_e}\{v'\mathbb{I}(Z \overset{elt}{\leq} \xi)\} M_V \mathbb{E}_{f_e}\{v\mathbb{I}(Z \overset{elt}{\leq} \xi)\}.$$

These results show that if the model is overidentified then imposing it leads to an efficiency gain in estimating the target and realized cdf's. Asymptotic optimality of EL implies that $\hat{F}(\xi)$ and $\hat{F}^*(\xi)$ are also asymptotically efficient. As in Section 4.1, the asymptotic variances of $\hat{\theta}$, $\hat{b}$, $\hat{K}$, $\hat{F}^*$, and $\hat{F}$ can be estimated by applying the analogy principle.

Example 5.1 (Example 3.3 contd.)**.** Since here $\rho_1(Z, \theta, b, K) = (Z - \theta)/a(Z, b)$ and there are no overidentifying restrictions, $(\hat{\theta}, \hat{b}, \hat{K})$ are obtained by solving $\sum_{j=1}^n \rho(Z_j, \hat{\theta}, \hat{b}, \hat{K}) = 0$. Therefore, $\hat{b} = \sum_{j=1}^n \mathbb{I}(Z_j \in T) R_j / \sum_{j=1}^n R_j$ is the fraction of observations in the refreshment sample that are not truncated, $\hat{K} = \sum_{j=1}^n R_j/n$ the size of the refreshment sample relative to the enriched sample, and $\hat{\theta} = n^{-1} \sum_{j=1}^n Z_j/a(Z_j, \hat{b}, \hat{K})$ since $\sum_{j=1}^n 1/a(Z_j, \hat{b}, \hat{K}) = n$. Using the fact that $\mathbb{I}(Z_j \notin T)(1 - R_j) = 0$, which follows by the definition of R_j, a little algebra shows that we can express $\hat{\theta}$ more intuitively as

$$\hat{\theta} = \hat{b} \times \frac{\sum_{j=1}^n Z_j \mathbb{I}(Z_j \in T)}{\sum_{j=1}^n \mathbb{I}(Z_j \in T)} + (1 - \hat{b}) \times \frac{\sum_{j=1}^n Z_j \mathbb{I}(Z_j \notin T) R_j}{\sum_{j=1}^n \mathbb{I}(Z_j \notin T) R_j},$$

which is what we would expect from (5.3). By Theorem 5.1, $n^{1/2}(\hat{b} - b^*)$ has asymptotic variance $b^*(1-b^*)/K_0$. The asymptotic variance of $\hat{\theta}$ can be similarly obtained. The absence of overidentifying restrictions implies that $\hat{F}(\xi) = n^{-1} \sum_{j=1}^n \mathbb{I}(Z_j \overset{elt}{\leq} \xi)$ and $\hat{F}^*(\xi) = n^{-1} \sum_{j=1}^n \mathbb{I}(Z_j \overset{elt}{\leq} \xi)/a(Z_j, \hat{b}, \hat{K})$ in this example and the next one. Therefore, by Theorem 5.3, $n^{1/2}\{\hat{F}^*(\xi) - F^*(\xi)\}$ is asymptotically

normal with mean zero and variance $\mathbb{E}_{f_e}\{w^2\}$. Similarly, by Theorem 5.4, $n^{1/2}\{\hat{F}(\xi) - F_e(\xi)\}$ is asymptotically normal with mean zero and variance $F_e(\xi)\{1 - F_e(\xi)\}$. □

Example 5.2 (Example 3.4 contd.)**.** Since $\rho_1(Z, \theta, b, K) = X(Y - X'\theta)/a(Z, b, K)$ and $a(Z, b, K) = K + (1 - K)\mathbb{I}(Y \in T_1, X \in T_2)/b$, we have $\hat{\theta} = \{\sum_{j=1}^n X_j X_j'/a(Z_j, \hat{b}, \hat{K})\}^{-1}\{\sum_{j=1}^n X_j Y_j/a(Z_j, \hat{b}, \hat{K})\}$ and $\hat{b}$ as in the previous example. Hence, $n^{1/2}(\hat{\theta} - \theta^*) \xrightarrow{d} \mathrm{N}(0, D^{-1}VD^{-1})$ by Theorem 5.1, where $D = -\mathbb{E}_{f_e}\{XX'/a(Z, b^*)\}$ and V as defined earlier. If truncation is purely endogenous then $\hat{\theta} = \{\sum_{j=1}^n X_j X_j'/a(Y_j, \hat{b}, \hat{K})\}^{-1}\{\sum_{j=1}^n X_j Y_j/a(Y_j, \hat{b}, \hat{K})\}$ where $a(Y_j, \hat{b}, \hat{K}) = \hat{K} + (1 - \hat{K})\mathbb{I}(Y_j \in T_1)/\hat{b}$. Similarly, under pure exogenous truncation, $\hat{\theta} = \{\sum_{j=1}^n X_j X_j'/a(X_j, \hat{b}, \hat{K})\}^{-1}\{\sum_{j=1}^n X_j Y_j/a(X_j, \hat{b}, \hat{K})\}$, where $a(X_j, \hat{b}, \hat{K}) = \hat{K} + (1 - \hat{K})\mathbb{I}(X_j \in T_2)/\hat{b}$. Asymptotic variance of $\hat{\theta}$ under pure endogenous (exogenous) truncation are given by setting $T_2 = \mathbb{R}^p$ ($T_1 = \mathbb{R}$) in the expressions for D and V. □

Example 5.3 (Truncated linear regression with endogenous regressors)**.** Consider the setup of Example 4.3, but now suppose that the target variable Z^* is truncated outside $T_1 \times T_2 \times T_3$. If the response, regressors, and instruments are all truncated then $\rho_1(Z, \theta, b, K) = W(Y - X'\theta)/a(Y, X, \tilde{W}, b, K)$, where $a(Y, X, \tilde{W}, b, K) = K + (1 - K)\mathbb{I}(Y \in T_1, X \in T_2, \tilde{W} \in T_3)/b$. Hence, θ^* and b^* can be estimated by GMM or EL and their asymptotic distributions obtained using Theorem 5.1. Of course, the pure endogenous truncation case can be handled by setting $T_2 = \mathbb{R}^p$ and $T_3 = \mathbb{R}^{q-p_1}$. □

Example 5.4 (Truncation and IV)**.** Again, consider the simultaneous equations model of Example 4.4, 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. θ^* and b^* can then be estimated by GMM or EL upon noting that $\rho_1(Z, \theta, b, K) = A(X)\begin{bmatrix} Y_1 - X_1'\theta_1 \\ Y_2 - X_2'\theta_2 \end{bmatrix}/a(Y, b, K)$ and $a(Y, b, K) = K + (1 - K)\mathbb{I}(Y_1 \in T_1, Y_2 \in T_2)/b$. □

5.2. **Hypothesis and specification testing.** Hypotheses of the form $H(\theta^*) = 0$ can be tested using the Wald, DM, or LR statistics as described in Section 4.2 by basing the test on $\rho(Z, R, \theta^*, b^*, K_0)$. In each case the test statistic is asymptotically distributed as a χ_h^2 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 $\rho(Z, R, \theta^*, b^*, K_0)$; details are analogous to those in Section 4.3. In either case, the test statistic is asymptotically χ_{q-p}^2 under H_0.

6. Application

Our application studies the effects of changes in compulsory schooling laws on age at first marriage. While the primary purpose of the application is to demonstrate the methodology developed in this paper, this is also a topic of some substantive importance. Our data are 1% samples from the Public Use Files of the U.S. Census of Population for the years 1960, 1970, and 1980.

Understanding the determinants of age at first marriage is considered to be important for several reasons. In recent years, age at first marriage has risen. Much literature suggests that a rising age at first marriage may be socially undesirable because marriage may encourage good behavior and outcomes. For example, Akerlof (1998) provides evidence that marriage has a beneficial effect on male behavior, leading to a decrease in socially undesirable activities such as alcoholism, drug abuse,

and violence. Also, Korenman and Neumark (1991) find that in the cross-section, married men earn about 11% more than observationally equivalent unmarried men. When they utilize panel data and estimate a fixed effects model, the marriage effect is about 2/3 the size of the cross-sectional estimate. Thus, it appears that there is a direct effect of being married on male earnings. However, in other work, they find that marriage reduces female participation and does not positively impact their wage rates (Korenman and Neumark 1992). Second, there is a great deal of concern about the effects of out-of-wedlock childbearing on single parents and their children. If rising age at first marriage is not accompanied by postponed childbearing, this problem becomes more severe. Relatedly, it has long been known, see, e.g., Coale (1971), that age at first marriage is an important determinant of fertility. However, rising age at first marriage may also have socially beneficial effects (Goldin and Katz 2002) because they have been linked to greater opportunities for young people, especially women, to obtain education and develop a professional career.

Theoretically, the effects of increased education on age of marriage are unclear. Koball (1998) describes the "economic provider" hypothesis that men are less likely to marry until they are securely employed. Because more education leads to higher earnings, it may lead to earlier marriage through this channel. The "adult transition" hypothesis proposes that events that delay the transition to adulthood will also delay marriage. More education will tend to delay marriage through this channel. Empirically, there is a positive relationship between education and age of marriage and rising education may be related to increased age at first marriage in recent decades. However, the correlation between education and age at first marriage may reflect the fact that young people with low ability and poor labor market prospects choose both to marry early and to drop out of school early rather than a causal relationship between education and age at first marriage. One way to examine this issue is to look at the effects of changes in policy that led to increased education. In particular, we study whether increased mandatory educational attainment (through compulsory schooling legislation) encourages people to defer marriage. If so, these factors should be considered when evaluating the benefits of this type of legislation.

We use variation in compulsory schooling laws across states and over time. Changes in these laws had a significant impact on education and indeed have been used as instruments for education in other contexts by Acemoglu and Angrist (2001), Lochner and Moretti (2004), and Lleras-Muney (2002). Since the history of compulsory schooling laws in the U.S. is by now well documented (see, in particular, Lleras-Muney (2001) and Goldin and Katz (2003)), we will not describe them in great detail here. Essentially, there were five possible restrictions on educational attendance: (i) maximum age by which a child must be enrolled, (ii) minimum age at which a child may drop out, (iii) minimum years of schooling before dropping out, (iv) minimum age for a work permit, and (v) minimum schooling required for a work permit. In the years relevant to our sample, 1939 to 1958, states changed compulsory attendance laws many times, usually upwards but sometimes downwards. Papers on the topic have used a variety of combinations of these restrictions as measures of compulsory schooling. We use required years of schooling, defined as the difference between the minimum dropout age and the maximum enrollment age following Lleras-Muney and Goldin and Katz. We follow Acemoglu and Angrist (2001) and Lochner and Moretti (2004) in assigning compulsory attendance laws to people

on the basis of state of birth and the year when the individual was 14 years old (with the exception that the enrollment age is assigned based on the laws in place when the individual was 7 years old). Also, we follow them in creating four indicator variables, depending on whether years of compulsory schooling are 8 or less, 9, 10, and 11 or more.

Our sample is composed of men and women born between 1925 and 1944. We choose this group of cohorts for two reasons. First, many of the changes in compulsory schooling laws were enacted between 1939 and 1958 and so had a major impact on this group. Secondly, the question on age at first marriage is not asked in the Census prior to 1960 or after 1980 so we are limited in terms of which cohorts we can study.

The empirical model can be written as

$$\log(Y_j^*) = X_j^{*\prime}\theta^* + \varepsilon_j^*, \quad j = 1, \ldots, n, \tag{6.1}$$

where Y_j^* denotes age at first marriage for the jth individual in the sample, X_j^* is a vector of explanatory variables including a constant, compulsory schooling law variables, year of birth dummies, state dummies, and a race dummy, and ε_j^* an unobserved error term that is uncorrelated with the regressors. There are 3 included compulsory schooling law variables describing the level of compulsory schooling: CA9 (9 years), CA10 (10 years), and CA11 (11 or 12 years). The omitted category is 8 years or less.

There are a few points to note about (6.1): First, it contains fixed cohort effects and state effects. The cohort effects are necessary to allow for secular changes in age at first marriage that may be completely unrelated to compulsory schooling laws. The state effects allow for the fact that variation in the timing of the law changes across states may not have been exogenous to the marriage market (for example, states with strict compulsory schooling laws may be states where people tend to marry late in any case).

The major problem in running this regression is that Y^* is censored for *younger* individuals because we observe age at first marriage only for ever *married* individuals; i.e., for each person we only observe

$$Y_j = \begin{cases} Y_j^* & \text{if } M_j = 1 \\ C_j & \text{if } M_j = 0, \end{cases} \tag{6.2}$$

where C_j is the chronological age and M_j an indicator for ever being married.

There are two elements of the censoring problem: (i) people who do get married at some point in their life but who have never been married at the time of interview, and (ii) people who never get married. Our goal is to address the first problem.[12] The usual approach to dealing with (i) is to restrict the sample to older men and women (e.g., Bergstrom and Schoeni (1996) restrict the sample to persons aged 40–60). This is obviously not a satisfactory solution because it replaces the censoring problem with a truncation problem. In contrast, our approach is to use both young and old persons, acknowledging that age at first marriage is significantly censored for younger women and men. As

[12]We cannot solve the second problem as, by definition, it is impossible to construct a refreshment sample for the group that will never marry.

discussed above we use the 1925–1944 cohorts, and these people are aged 16–35 in 1960, and 26–45 in 1970. Clearly, age at first marriage is censored for many of these persons. To deal with this problem, we need a refreshment sample that is not censored and is from the same population as our master sample (aged 16–35 in 1960 and 26–45 in 1970). We obtain this by using individuals from the same cohort: A 16 year old woman in 1960 is considered to be from the same population as a 26 year old woman in 1970, and a 36 year old in 1980. Hence, for women who were between 16–35 in 1960 and 26–45 in 1970, the refreshment sample consists of women aged 36–55 in 1980.

For the group of people aged 36–55 in 1980 to be a suitable refreshment sample, it must possess two characteristics. First, it must be a draw from the same population as the master sample. We consider this to be a reasonable assumption in this case because: (a) they are from the exact same birth cohorts as persons in the master sample; (b) we only use individuals born in the U.S. so immigration is not a problem; (c) we do not include individuals aged more than 55 (and these cohorts were not involved in World War 2 or Vietnam) so mortality is not a major consideration. We report descriptive statistics for our sample in tables 1 and 2 for women and men, respectively. Note that the percentage white, average year-of-birth, and the proportions affected by each compulsory schooling law regime are very similar across Census samples. This is as we would expect given we are tracking a population as they age. On the other hand, the average values of age at first marriage and education differ greatly by Census due to censoring. To further corroborate that we are following samples from the same population, in figure 1 we also present QQ plots for age at first marriage of men and women aged at least 26 that were married before they were 26 years old[13]. The linearity of the plots is strong evidence that the uncensored observations in these samples indeed come from the same population.

The second characteristic of a refreshment sample is that it should not have a censoring problem. We examine this issue in table 3. In this table, we track each birth cohort over time, and list the percentage who have never been married. For women, we see that the proportion never married flattens out as women reach their early 30's and it appears that very few women marry for the first time after age 35. Thus, it appears that the refreshment sample for women is approximately free of censoring bias. Men tend to marry at later ages and so there does appear to be some censoring in the refreshment sample for men. However, it impacts a very small proportion of cases; it appears that about 6% of men never marry, and very few cohorts in the refreshment sample have more than 6% of censored observations in 1980. Despite the evidence that there may be some censoring in the 1980 sample, in estimation we treat it as a refreshment sample that has no censored observations.

As mentioned above, we cannot address the second type of censoring (people who never get married) using a refreshment sample approach. Instead, we have taken a few different ad hoc approaches and verify that our results are not very sensitive to the exact method used. The approaches we have tried are (i) impute age at first marriage as equal to current age for never married individuals in the refreshment sample, and (ii) impute age at marriage for all cases where individuals are not married by 35 (we have tried imputing the age to 55 and 65). We have found that our GMM

[13]All individuals are aged at least 26 in the 1970 and 1980 samples. To compare 1960 to 1980, we restricted the sample to the oldest 10 cohorts i.e. persons aged at least 26 in 1960 and at least 46 in 1980. This trades off the number of cohorts included against the number of uncensored marriage ages observed.

estimates (which are identical to EL estimates since here θ^* is just identified) are reasonably robust to the imputation method used and so we report the results using method (i).

We report the following GMM estimates of the coefficients of the compulsory schooling variables and the white dummy in table 4: GMM60, obtained by matching the 1960 master sample with the 1980 refreshment sample to create the enriched dataset, and GMM70, the GMM estimator when the 1970 and 1980 samples are matched. Estimates for men and women are reported separately. Following the procedure described in Section 4.1 (see Example 4.2 for an illustration), both estimators were based on (4.7) and implemented in the GAUSS programming language. Since the consistency of our estimators does not depend upon the extent to which the data are censored, we also expect GMM60 and GMM70 to give similar estimates in finite samples even though censoring is less of a problem in 1970. This is borne out by the evidence summarized in table 4.

An enriched dataset has to, by definition, contain some observations that are not subject to the censoring mechanism. Since age at first marriage is censored from above by chronological age in this application, an enriched dataset here must contain some observations for which age at first marriage is *greater* than chronological age; i.e., loosely speaking, we must have some counterfactual observations for whom we can "look into the future" at the time of interview and see when they first get married. To construct such an enriched dataset by matching, say, the 1960 and 1980 samples, we first create a new variable $\tilde{C}_j = C_j \mathbb{I}(j \in 1960) + (C_j - 20)\mathbb{I}(j \in 1980)$ that represents the chronological age of the jth individual in 1960. The enriched observations used to construct GMM60 are then obtained by replacing C_j in (6.2) with $\tilde{C}_j$. GMM70 is obtained in a similar manner by matching the 1970 and 1980 datasets.

To contrast our GMM estimators with some competing estimators, we also report OLS60, OLS70, TOBIT60, and TOBIT70, the OLS and tobit estimates for each year. Another estimator we consider is OLS80, obtained by doing least squares on just the 1980 sample. It is consistent because the refreshment sample is not censored. Therefore, GMM70 and OLS80 both serve as consistency checks for GMM60. Incidentally, note that although age at first marriage is a continuously distributed random variable, in the data it is recorded in discrete units (years). Therefore, we cannot do censored quantile regression in this application.

First, consider the compulsory schooling estimates for women. The GMM estimates for both 1960 and 1970 are quite similar and suggest that moving from less than 9 years of compulsory schooling to 9 years increases log age at first marriage by about 0.01, implying age at first marriage increases by approximately 1%. The effects for 10 years of compulsory schooling is about 1.5%, and the effects of 11 or more is about 2%. Not surprisingly, these effects are about the same size as one obtains using just the refreshment sample (the 1980 data) because the refreshment sample does not suffer from censoring bias. Note, however, that the GMM estimates are more precisely estimated than the OLS estimates from 1980, as GMM is optimally using additional information from the 1960 and 1970 samples. The gain in efficiency is bigger for GMM70 than for GMM60, presumably because the 1970 data has less of a censoring problem and hence is more informative[14]. The OLS estimates from 1960

[14]The difference in the standard errors between OLS80 and the GMM estimators is not that big in this application. We have experimented with reducing the size of the refreshment sample by taking random 10% and 20% subsamples

and 1970 show signs of bias due to censoring. In particular, the 1960 estimates indicate very large effects of the compulsory schooling laws on age at first marriage. The final two columns in table 4 report tobit estimates. The tobit estimates of the compulsory schooling laws are typically lower than that of the GMM estimators. Also, there is a substantial difference between the tobit estimates for 1960 and the equivalent estimates for 1970, indicating that tobit is performing poorly in this situation.

The estimate of the white dummy for women is also in table 4. The GMM estimates both indicate that whites tend to marry at younger ages than non-whites – the point estimates imply the difference is about 8–9%. Once again, OLS estimates for 1960 and 1970 are very different, suggesting that censoring bias is serious for these samples. The two tobit estimates are again very different from the GMM estimates.

The compulsory schooling and white estimates for men are also in table 4. They differ from the female results in that the GMM estimates only suggest significant effects of 10 years of required schooling (9 years is marginally significant for GMM70). In contrast, the OLS estimates for 60 and 70 show strong significant effects of all the laws on age of first marriage. As in the female sample, the GMM estimates of the white coefficient imply a difference of about 8-9%. Once again, OLS and tobit estimates for 1980 and 1990 are very different, suggesting that censoring bias is serious.

Cohort and state fixed effects were also included in the specification. The estimated cohort effects show how the conditional mean of log(age at first marriage) varies by birth cohort. The oldest cohort (persons born in 1925) is the excluded dummy in the regression, so the estimate for this group is normalized to zero. Rather than report the coefficients of the cohort dummies, we plot them for women and men in figures 2 and 3, respectively. Not surprisingly, the cohort effects for OLS60 are radically different from the rest. The cohort effects for the rest of the estimators are quite similar to each other.

In summary, we find positive effects of the compulsory schooling laws on age at first marriage. However, the magnitude of the effects are much smaller than would be inferred from ignoring the censoring problem in the 1960 and 1970 data. In contrast, we find large racial differences that are largely obscured in the censored data. Taken together, these demonstrate the importance in this application of using an approach that takes account of censoring. The similarity of the GMM estimates from 1960 and 1970 to each other and to the OLS estimates from 1980 also demonstrates our theoretical result that the proposed estimators are consistent irrespective of the extent of censoring.

7. Concluding remarks and some topics for future research

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 refreshment sample can be combined with the original dataset of

and found a much bigger gain in precision for the GMM estimators over OLS80 (although the resulting estimates are all much less precise than those reported in table 4).

censored or truncated observations to efficiently correct for the effects of partial observation so that all standard GMM and EL based inference goes through. To illustrate our results in an empirical setting, we show how to estimate the effect of changes in compulsory schooling laws on age at first marriage, a variable that is censored for younger individuals, and also demonstrate how refreshment samples in this application can be created by matching cohort information across census datasets.

Of course, the methods developed in this paper are readily applicable in many other applied contexts[15]. For example, an important potential application is to the estimation of unemployment durations and re-employment wages subsequent to job displacement. U.S. analyses of the consequences of job displacement have predominantly relied on the Displaced Worker Supplement (DWS) to the Current Population Survey (CPS). However, serious problems arise because many individuals have not become re-employed by the time of the CPS survey so that unemployment durations are censored and re-employment wages are truncated. By using panel data sets such as the Panel Study of Income Dynamics (PSID), one can augment the CPS with a sample that does not have these censoring problems (as individuals are followed for years after displacement) and consistently estimate parameters of interest. We intend to examine this application in future research. The theory developed here can be extended to handle binary response, ordered response, and models involving interval censored or missing data as well. Research on all these topics is also in progress and will be presented in subsequent papers.

Appendix A. Proofs of the results in section 4

For notational convenience, let $\hat{\beta} = (\hat{\theta}, \hat{K})_{(p+1)\times 1}$ and $\beta^* = (\theta^*, K_0)$.

Proof of Theorem 4.1. From standard GMM theory we know that $n^{1/2}(\hat{\beta}_{gmm} - \beta^*)$ is asymptotically normal with mean zero and variance $(D_{f_e}' V_{f_e}^{-1} D_{f_e})^{-1}$, where $D_{f_e} = \mathbb{E}_{f_e}\{\partial\rho(Z, R, \beta^*)/\partial\beta\}$ and $V_{f_e} = \mathbb{E}_{f_e}\{\rho(Z, R, \beta^*)\rho'(Z, R, \beta^*)\}$. Then, letting $\Sigma_{q\times 2} = [\,\Sigma_{12}\ \Sigma_{13}\,]$ and $\Sigma_{23} = \mathbb{E}_{f_e}\{\rho_2(Z, K_0)\rho_3(R, K_0)\}$,

$$V_{f_e} = \begin{bmatrix} V_1 & \Sigma \\ \Sigma' & V_{-1} \end{bmatrix}_{(q+2)\times(q+2)} \quad \text{where} \quad V_{-1} = \begin{bmatrix} V_2 & \Sigma_{23} \\ \Sigma_{23} & V_3 \end{bmatrix}_{2\times 2}.$$

Hence, by the partitioned inverse formula,

$$V_{f_e}^{-1} = \begin{bmatrix} \Omega^{-1} & -\Omega^{-1}\Sigma V_{-1}^{-1} \\ -V_{-1}^{-1}\Sigma'\Omega^{-1} & V_{-1} + V_{-1}^{-1}\Sigma'\Omega^{-1}\Sigma V_{-1}^{-1} \end{bmatrix}, \tag{A.1}$$

[15]Applications where refreshment samples are relatively straightforward to construct seem to be those where censoring or truncation can in some sense be regarded as nuisance processes, i.e., where the underlying economic outcomes are not restricted but their measured or recorded versions are. In contrast, it seems hard, at least to us, to non-experimentally construct refreshment samples by matching datasets in applications where censoring or truncation are thought of as being behavioral in origin, i.e., where there are fundamental constraints that bind economic behavior such as those in models of female labor supply or household demand for durable goods.

where $\Omega = V_1 - \Sigma V_{-1}^{-1}\Sigma'$. Since ε is the residual from an orthogonal projection of $\rho_1(Z,\theta^*,K_0)$ onto the linear span of $\{1,\rho_2(Z,K_0),\rho_3(R,K_0)\}$, it is immediate that $\mathbb{E}_{f_e}\{\varepsilon\varepsilon'\} = \Omega$. Furthermore, since

$$V_{-1} \overset{\text{Lemma A.2}}{=} \begin{bmatrix} K_0(1-K_0)[1-F^*(c)] & K_0(1-K_0)[1-F^*(c)] \\ K_0(1-K_0)[1-F^*(c)] & K_0(1-K_0) \end{bmatrix}, \tag{A.2}$$

V_{-1}^{-1} is easily obtained. Next, observe that

$$D_{f_e} = \begin{bmatrix} D & \mathbb{E}_{f_e}\{\partial\rho_1(Z,\theta^*,K_0)/\partial K\} \\ 0'_{p\times 1} & -\mathbb{E}_{f_e}\{\mathbb{I}(Z \overset{elt}{<} c)^{\sim}\} \\ 0'_{p\times 1} & -1 \end{bmatrix} \overset{\text{Lemma A.3}}{=} \begin{bmatrix} D & -\Sigma_{12}/K_0(1-K_0) \\ 0'_{p\times 1} & -[1-F^*(c)] \\ 0'_{p\times 1} & -1 \end{bmatrix}_{(q+2)\times(p+1)}. \tag{A.3}$$

Therefore, using (A.1)–(A.3), straightforward matrix multiplication shows that

$$D'_{f_e}V_{f_e}^{-1}D_{f_e} = \begin{bmatrix} D'\Omega^{-1}D & 0_{p\times 1} \\ 0'_{p\times 1} & 1/K_0(1-K_0) \end{bmatrix}.$$

The desired result follows. □

Proof of Theorem 4.2. Same as the proof of Theorem 4.1, the only difference being that since estimation here is based on (4.7) we now have

$$D_{f_e} = \begin{bmatrix} D & -\Sigma_{12}/K_0(1-K_0) \\ 0'_{p\times 1} & -[1-F^*(c)] \end{bmatrix} \quad \text{and} \quad V_{f_e} = \begin{bmatrix} V_1 & \Sigma_{12} \\ \Sigma'_{12} & K_0(1-K_0)[1-F^*(c)] \end{bmatrix}.$$

Therefore,

$$D'_{f_e}V_{f_e}^{-1}D_{f_e} = \begin{bmatrix} D'\Omega^{-1}D & 0_{p\times 1} \\ 0'_{p\times 1} & [1-F^*(c)]/K_0(1-K_0) \end{bmatrix}$$

and the desired result follows. □

Proof of Theorem 4.3. Since $\hat{\theta}_R$ is the optimal GMM estimator based on $\mathbb{E}_{f_e}\{g(Z,\theta^*)R\} = 0$, we know that $n^{1/2}(\hat{\theta}_R - \theta^*)$ is asymptotically normal with mean zero and variance $(D'_R V_R^{-1} D_R)^{-1}$, where $D_R = \mathbb{E}_{f_e}\{\partial g(Z,\theta^*)R/\partial\theta\}$ and $V_R = \mathbb{E}_{f_e}\{g(Z,\theta^*)g'(Z,\theta^*)R\}$. But,

$$D_R = \mathbb{E}_{f_e}\{\partial g(Z,\theta^*)R/\partial\theta\} = K_0\mathbb{E}_{f_e}\{\partial g(Z,\theta^*)/\partial\theta | R=1\} \overset{(2.5)}{=} K_0\mathbb{E}_{f^*}\{\partial g(Z,\theta^*)/\partial\theta\} = K_0 D_*.$$

Similarly, we can show that $V_R = K_0 V_*$. Hence, $(D'_R V_R^{-1} D_R)^{-1} = (D'_* V_*^{-1} D_*)^{-1}/K_0$. Next, observe that $D_* = D$ by (4.1) and the fact that $\mu^*(\{c\}) = 0$. Hence, to prove $\text{asvar}(\hat{\theta}_R) > \text{asvar}(\hat{\theta}_{gmm})$ it suffices to show that $V_*/K_0 > \Omega$; i.e., $V_*/K_0 - \Omega$ is positive definite. So, by (4.1) and $\mu^*(\{c\}) = 0$, observe that $V_1 = \mathbb{E}_{f^*}\{g(Z,\theta^*)g'(Z,\theta^*)\mathbb{I}(Z \overset{elt}{<} c)\} + \mathbb{E}_{f^*}\{g(Z,\theta^*)g'(Z,\theta^*)\mathbb{I}(Z \overset{elt}{<} c)^{\sim}\}/K_0$. Hence,

$$\Omega = V_1 - \Sigma_{12}\Sigma'_{12}/V_2 = V_*/K_0 - [(1/K_0 - 1)\mathbb{E}_{f^*}\{g(Z,\theta^*)g'(Z,\theta^*)\mathbb{I}(Z \overset{elt}{<} c)\} + \Sigma_{12}\Sigma'_{12}/V_2].$$

Therefore, $\Omega < V_*/K_0$ since $K_0 \in (0,1)$. The desired result follows. □

Proof of Theorem 4.4. Since[16] $\sum_{j=1}^n \hat{p}_j(\hat{\beta}_{el})\mathbb{I}(Z_j \overset{elt}{\neq} c)/a(Z,\hat{K}_{el}) = 1$, we have $\hat{F}^*(\xi) - F^*(\xi) = \sum_{j=1}^n \hat{p}_j(\hat{\beta}_{el})I(Z_j,\xi)\mathbb{I}(Z_j \overset{elt}{\neq} c)/a(Z_j,\hat{K}_{el})$. But, w.p.a.1,

$$\frac{I(Z_j,\xi)\mathbb{I}(Z_j \overset{elt}{\neq} c)}{a(Z_j,\hat{K}_{el})} = \frac{I(Z_j,\xi)\mathbb{I}(Z_j \overset{elt}{\neq} c)}{a(Z_j,K_0)} + \frac{\partial}{\partial K}\{\frac{I(Z_j,\xi)\mathbb{I}(Z_j \overset{elt}{\neq} c)}{a(Z_j,K_0)}\}(\hat{K}_{el}-K_0) + O(|\hat{K}_{el}-K_0|^2),$$

where the $O(|\hat{K}_{el}-K_0|^2)$ term does not depend upon j. Hence, since $n^{1/2}(\hat{K}_{el}-K_0) = O_p(1)$,

$$\begin{aligned} n^{1/2}\{\hat{F}^*(\xi) - F^*(\xi)\} = n^{1/2}\sum_{j=1}^n \hat{p}_j(\hat{\beta}_{el})\frac{I(Z_j,\xi)\mathbb{I}(Z_j \overset{elt}{\neq} c)}{a(Z_j,K_0)} \\ + \{\sum_{j=1}^n \hat{p}_j(\hat{\beta}_{el})\frac{\partial}{\partial K}\frac{I(Z_j,\xi)\mathbb{I}(Z_j \overset{elt}{\neq} c)}{a(Z_j,K_0)}\}\, n^{1/2}(\hat{K}_{el}-K_0) + O_p(n^{-1/2}). \quad \text{(A.4)} \end{aligned}$$

Now, $\hat{K}_{el} = \sum_{j=1}^n \hat{p}_j(\hat{\beta}_{el})\mathbb{I}(Z_j \overset{elt}{\neq} c)\mathbb{I}(Z_j \overset{elt}{<} c)^\sim / \sum_{j=1}^n \hat{p}_j(\hat{\beta}_{el})\mathbb{I}(Z_j \overset{elt}{<} c)^\sim$ by (4.7). Hence,

$$n^{1/2}(\hat{K}_{el}-K_0) = n^{1/2}\sum_{j=1}^n \hat{p}_j(\hat{\beta}_{el})\rho_2(Z_j,K_0) / \sum_{j=1}^n \hat{p}_j(\hat{\beta}_{el})\mathbb{I}(Z_j \overset{elt}{<} c)^\sim. \quad \text{(A.5)}$$

Since $\mathbb{E}_{f_e}\{\rho_2(Z,K_0)\} = 0$, the numerator of (A.5) is $O_p(n^{-1/2})$. Furthermore, we can also verify that

$$\begin{aligned} \sum_{j=1}^n \hat{p}_j(\hat{\beta}_{el})\frac{\partial}{\partial K}\frac{I(Z_j,\xi)\mathbb{I}(Z_j \overset{elt}{\neq} c)}{a(Z_j,K_0)} &= \mathbb{E}\{\frac{\partial}{\partial K}\frac{I(Z,\xi)\mathbb{I}(Z \overset{elt}{\neq} c)}{a(Z,K_0)}\} + o_p(1) \\ &= -\mathbb{E}_{f_e}\{\frac{I(Z,\xi)\mathbb{I}(Z \overset{elt}{\neq} c)}{a(Z,K_0)}\rho_2(Z,K_0)\}/K_0(1-K_0) + o_p(1), \quad \text{(A.6)} \\ \sum_{j=1}^n \hat{p}_j(\hat{\beta}_{el})\mathbb{I}(Z_j \overset{elt}{<} c)^\sim &= \mathbb{E}_{f_e}\{\mathbb{I}(Z \overset{elt}{<} c)^\sim\} + o_p(1) = [1-F^*(c)] + o_p(1), \quad \text{(A.7)} \end{aligned}$$

where (A.6) can be shown as in the proof of Lemma A.3(i) and (A.7) follows by Lemma A.3(ii). Combining (A.5)–(A.7) and using the fact that $V_2 = K_0(1-K_0)[1-F^*(c)]$, (A.4) becomes

$$n^{1/2}\{\hat{F}^*(\xi) - F^*(\xi)\} = n^{1/2}\sum_{j=1}^n \hat{p}_j(\hat{\beta}_{el})u_j + o_p(1).$$

Therefore, since $n^{1/2}\{\hat{F}(\xi) - F_e(\xi)\} = n^{1/2}\sum_{j=1}^n \hat{p}_j(\hat{\beta}_{el})\{\mathbb{I}(Z_j \overset{elt}{\leq} \xi) - \mathbb{E}_{f_e}\mathbb{I}(Z \overset{elt}{\leq} \xi)\}$, the asymptotic distribution of $n^{1/2}\{\hat{F}^*(\xi) - F^*(\xi)\}$ is easily obtained by replacing $\mathbb{I}(Z_j \overset{elt}{\leq} \xi) - \mathbb{E}_{f_e}\mathbb{I}(Z \overset{elt}{\leq} \xi)$ in the proof of Theorem 4.5 with u_j. The desired result follows. □

---

[16] A little algebra shows that $\rho_2(Z,K) = K\{\mathbb{I}(Z \overset{elt}{\neq} c)/a(Z,K) - 1\}$. Therefore, since $\sum_{j=1}^n \hat{p}_j(\hat{\beta}_{el})\rho_2(Z_j,\hat{K}_{el}) = 0$ and the $\hat{p}_j$'s sum to one, it follows that $\sum_{j=1}^n \hat{p}_j(\hat{\beta}_{el})\mathbb{I}(Z_j \overset{elt}{\neq} c)/a(Z_j,\hat{K}_{el}) = 1$.

Proof of Theorem 4.5. From Qin and Lawless (1994, Theorem 1) we know that $n^{1/2}\{\hat{F}(\xi) - F_e(\xi)\}$ is asymptotically normal with mean zero and variance

$$\mathbb{E}_{f_e}\{\mathbb{I}(Z \overset{elt}{\leq} \xi) - F_e(\xi)\}^2 - \mathbb{E}_{f_e}\{\rho'(Z, \beta^*)\mathbb{I}(Z \overset{elt}{\leq} \xi)\} M_{f_e} \mathbb{E}_{f_e}\{\rho(Z, \beta^*)\mathbb{I}(Z \overset{elt}{\leq} \xi)\}, \tag{A.8}$$

where $M_{f_e} = V_{f_e}^{-1} - V_{f_e}^{-1} D_{f_e} (D'_{f_e} V_{f_e}^{-1} D_{f_e})^{-1} D'_{f_e} V_{f_e}^{-1}$. Using the expressions for D_{f_e}, $V_{f_e}^{-1}$, and $D'_{f_e} V_{f_e}^{-1} D_{f_e}$ given in the proof of Theorem 4.2, some straightforward calculations show that

$$M_{f_e} = \begin{bmatrix} M_\Omega & -M_\Omega \Sigma_{12}/V_2 \\ -\Sigma'_{12} M_\Omega / V_2 & \Sigma'_{12} M_\Omega \Sigma_{12} / V_2^2. \end{bmatrix}.$$

A little algebra then reveals that the second term in (A.8) equals $\mathbb{E}_{f_e}\{\varepsilon' \mathbb{I}(Z \overset{elt}{\leq} \xi)\} M_\Omega \mathbb{E}_{f_e}\{\varepsilon \mathbb{I}(Z \overset{elt}{\leq} \xi)\}$, and the desired result follows. □

Lemma A.1. $\text{Proj}_{f_e}\{\rho_1(Z, \theta^*, K_0) \big| 1, \rho_2(Z, K_0), \rho_3(R, K_0)\} = \text{Proj}_{f_e}\{\rho_1(Z, \theta^*, K_0) \big| 1, \rho_2(Z, K_0)\}$.

Proof of Lemma A.1. To prove this result, it suffices to show that

$$\mathbb{E}_{f_e}\{[\rho_1(Z, \theta^*, K_0) - \text{Proj}_{f_e}\{\rho_1(Z, \theta^*, K_0) | 1, \rho_2(Z, K_0)\}]\rho_3(R, K_0)\} = 0. \tag{A.9}$$

But $\text{Proj}_{f_e}\{\rho_1(Z, \theta^*, K_0) | 1, \rho_2(Z, K_0)\} = \Sigma_{12}\rho_2(Z, K_0)/V_2$. Hence, by Lemma A.2, (A.9) holds if and only if $\Sigma_{13} = \Sigma_{12}$. Now, $\Sigma_{13} = K_0 \mathbb{E}_{f_e}\{\rho_1(Z, \theta^*, K_0) | R = 1\} = K_0 \mathbb{E}_{f^*}\{\rho_1(Z, \theta^*, K_0)\}$ by (2.5) and $\mathbb{E}_{f^*}\{\rho_1(Z, \theta^*, K_0)\} = \mathbb{E}_{f^*}\{g(Z, \theta^*)[\mathbb{I}(Z \overset{elt}{<} c) + \mathbb{I}(Z \overset{elt}{<} c)^\sim]/a(Z, K_0)\}$ since $\mu^*(\{c\}) = 0$. Hence,

$$\Sigma_{13} = -(1 - K_0)\mathbb{E}_{f^*}\{g(Z, \theta^*)\mathbb{I}(Z \overset{elt}{<} c)\}. \tag{A.10}$$

Next, $\Sigma_{12} = (1 - K_0)\mathbb{E}_{f_e}\{g(Z, \theta^*)\mathbb{I}(Z \overset{elt}{\neq} c)\mathbb{I}(Z \overset{elt}{<} c)^\sim / a(Z, K_0)\} = (1 - K_0)\mathbb{E}_{f^*}\{g(Z, \theta^*)\mathbb{I}(Z \overset{elt}{<} c)^\sim\}$, where the second equality follows by (4.1) and the assumption that $\mu^*(\{c\}) = 0$. Hence,

$$\Sigma_{12} = -(1 - K_0)\mathbb{E}_{f^*}\{g(Z, \theta^*)\mathbb{I}(Z \overset{elt}{<} c)\}. \tag{A.11}$$

The desired result follows by (A.10) and (A.11). □

Lemma A.2. *(i)* $\Sigma_{23} = K_0(1 - K_0)[1 - F^*(c)]$ *and (ii)* $V_2 = K_0(1 - K_0)[1 - F^*(c)]$.

Proof of Lemma A.2. We only show (i) since the proof of (ii) is very similar. So, note that

$$\Sigma_{23} = \mathbb{E}_{f_e}\{\rho_2(Z, K_0)R\} = K_0 \mathbb{E}_{f_e}\{\rho_2(Z, K_0) | R = 1\} \overset{(2.5)}{=} K_0 \mathbb{E}_{f^*}\{\rho_2(Z, K_0)\}.$$

Hence, (i) follows since $\mathbb{E}_{f^*}\{\rho_2(Z, K_0)\} = (1 - K_0)\mathbb{E}_{f^*}\{\mathbb{I}(Z \overset{elt}{<} c)^\sim\} = (1 - K_0)[1 - F^*(c)]$. □

Lemma A.3. *(i)* $\mathbb{E}_{f_e}\{\partial\rho_1(Z, \theta^*, K_0)/\partial K\} = -\Sigma_{12}/K_0(1 - K_0)$ *and (ii)* $\mathbb{E}_{f_e}\{\mathbb{I}(Z \overset{elt}{<} c)^\sim\} = 1 - F^*(c)$.

Proof of Lemma A.3. Now $\partial\rho_1(Z, \theta^*, K_0)/\partial K = -g(Z, \theta^*)\mathbb{I}(Z \overset{elt}{\neq} c)\mathbb{I}(Z \overset{elt}{<} c)^\sim / a^2(Z, K_0)$. Hence, $\mathbb{E}_{f_e}\{\partial\rho_1(Z, \theta^*, K_0)/\partial K\} = \mathbb{E}_{f^*}\{g(Z, \theta^*)\mathbb{I}(Z \overset{elt}{<} c)\}/K_0$ by (4.1) and (i) follows by (A.11). Next,

$$\mathbb{E}_{f_e}\{\mathbb{I}(Z \overset{elt}{<} c)^\sim\} = 1 - \mathbb{E}_{f_e}\{\mathbb{I}(Z \overset{elt}{\neq} c)\mathbb{I}(Z \overset{elt}{<} c)\} = 1 - \mathbb{E}_{f^*}\{\mathbb{I}(Z \overset{elt}{\neq} c)\mathbb{I}(Z \overset{elt}{<} c)a(Z, K_0)\} = 1 - F^*(c)$$

since $\mu^*(\{c\}) = 0$ by assumption. □

Appendix B. Proofs of the results in section 5

To be written.

Appendix C. Tables and Figures

Table 1. Descriptive statistics for women by year

	Mean	Std. Dev.	Min	Max
		1960 (220730 observations)		
Birth Cohort	1934.62	5.98	1925	1944
Age	25.38	5.98	16	35
Age at First Marriage	19.68	3.59	14	35
Never Married	0.29	0.45	0	1
White	0.88	0.32	0	1
≤ 8 Years of Schooling Required	0.19	0.39	0	1
9 Years of Schooling Required	0.66	0.47	0	1
10 Years of Schooling Required	0.08	0.27	0	1
≥ 11 Years of Schooling Required	0.07	0.26	0	1
		1970 (216036 observations)		
Birth Cohort	1934.69	5.94	1925	1944
Age	35.31	5.94	26	45
Age at First Marriage	21.23	5.19	14	45
Never Married	0.07	0.25	0	1
White	0.88	0.32	0	1
≤ 8 Years of Schooling Required	0.19	0.39	0	1
9 Years of Schooling Required	0.66	0.47	0	1
10 Years of Schooling Required	0.08	0.27	0	1
≥ 11 Years of Schooling Required	0.07	0.26	0	1
		1980 (223903 observations)		
Birth Cohort	1934.73	5.95	1925	1944
Age	45.28	5.95	36	55
Age at First Marriage	22.07	7.01	12	55
Never Married	0.05	0.22	0	1
White	0.88	0.33	0	1
≤ 8 Years of Schooling Required	0.19	0.39	0	1
9 Years of Schooling Required	0.66	0.47	0	1
10 Years of Schooling Required	0.08	0.27	0	1
≥ 11 Years of Schooling Required	0.07	0.26	0	1

TABLE 2. Descriptive statistics for men by year

	Mean	Std. Dev.	Min	Max
		1960 (213184 observations)		
Birth Cohort	1934.69	6.00	1925	1944
Age	25.31	6.00	16	35
Age at First Marriage	21.36	3.95	14	35
Never Married	0.42	0.49	0	1
White	0.89	0.31	0	1
≤ 8 Years of Schooling Required	0.19	0.39	0	1
9 Years of Schooling Required	0.66	0.47	0	1
10 Years of Schooling Required	0.08	0.27	0	1
≥ 11 Years of Schooling Required	0.07	0.26	0	1
		1970 (207129 observations)		
Birth Cohort	1934.71	5.94	1925	1944
Age	35.29	5.94	26	45
Age at First Marriage	23.82	5.26	14	45
Never Married	0.10	0.30	0	1
White	0.90	0.30	0	1
≤ 8 Years of Schooling Required	0.19	0.39	0	1
9 Years of Schooling Required	0.66	0.47	0	1
10 Years of Schooling Required	0.08	0.27	0	1
≥ 11 Years of Schooling Required	0.07	0.26	0	1
		1980 (212244 observations)		
Birth Cohort	1934.80	5.93	1925	1944
Age	45.20	5.93	36	55
Age at First Marriage	24.95	7.26	12	55
Never Married	0.07	0.25	0	1
White	0.89	0.31	0	1
≤ 8 Years of Schooling Required	0.19	0.39	0	1
9 Years of Schooling Required	0.66	0.47	0	1
10 Years of Schooling Required	0.08	0.27	0	1
≥ 11 Years of Schooling Required	0.07	0.26	0	1

TABLE 3. Proportion censored by cohort and year

age in	% of women censored			% of men censored		
1960	1960	1970	1980	1960	1970	1980
16	94	13	7	99	20	9
17	88	11	6	98	17	8
18	75	10	6	95	15	8
19	59	9	6	87	13	8
20	46	8	6	75	12	7
21	34	7	5	62	10	7
22	25	7	5	50	10	6
23	19	7	5	40	9	6
24	15	6	5	32	8	7
25	13	6	5	27	9	6
26	11	5	5	22	8	6
27	9	6	4	19	7	6
28	9	5	5	16	7	5
29	9	5	5	15	7	6
30	8	5	4	13	7	6
31	7	5	4	12	7	6
32	6	5	4	11	7	6
33	6	5	5	11	7	6
34	6	5	4	10	7	6
35	6	5	5	9	6	6
36	6	5	4	8	6	6
37	6	6	5	8	6	6
38	5	5	5	8	6	6
39	6	5	5	8	6	6
40	6	6	5	7	6	6

TABLE 4. Effects of compulsory schooling laws and race on log(age at first marriage). Also included in the specification, but not reported in this table, are a constant, year-of-birth indicators, and state dummies.

Women	OLS60	OLS70	OLS80	GMM60	GMM70	TOBIT60	TOBIT70
9 Years Schooling Reqd.	.0157*	.0080*	.0096*	.0102*	.0094*	.0029	.0077*
	(.0016)	(.0022)	(.0025)	(.0024)	(.0022)	(.0018)	(.0021)
10 Years Schooling Reqd.	.0232*	.0112*	.0146*	.0150*	.0129*	.0075*	.0103*
	(.0021)	(.0030)	(.0035)	(.0034)	(.0031)	(.0026)	(.0031)
11+ Years Schooling Reqd.	.0456*	.0317*	.0184*	.0188*	.0223*	.0157*	.0299*
	(.0038)	(.0052)	(.0061)	(.0060)	(.0053)	(.0049)	(.0057)
White	−.0261*	−.0476*	−.0827*	−.0927*	−.0808*	−.0393*	−.0534*
	(.0012)	(.0018)	(.0021)	(.0020)	(.0019)	(.0014)	(.0016)
Men	OLS60	OLS70	OLS80	GMM60	GMM70	TOBIT60	TOBIT70
9 Years Schooling Reqd.	.0114*	.0073*	.0031	.0046	.0061*	−.0028	.0062*
	(.0015)	(.0022)	(.0026)	(.0024)	(.0023)	(.0019)	(.0022)
10 Years Schooling Reqd.	.0205*	.0120*	.0130*	.0131*	.0137*	.0052	.0109*
	(.0019)	(.0029)	(.0035)	(.0033)	(.0031)	(.0029)	(.0031)
11+ Years Schooling Reqd.	.0359*	.0152*	.0049	.0028	.0070	.0055	.0121*
	(.0035)	(.0053)	(.0063)	(.0060)	(.0056)	(.0053)	(.0057)
White	−.0156*	−.0444*	−.0792*	−.0826*	−.0792*	−.0301*	−.0515*
	(.0011)	(.0018)	(.0021)	(.0020)	(.0019)	(.0016)	(.0017)

Standard errors in parenthesis. A "*" superscript denotes that effect is significant at 5% level.

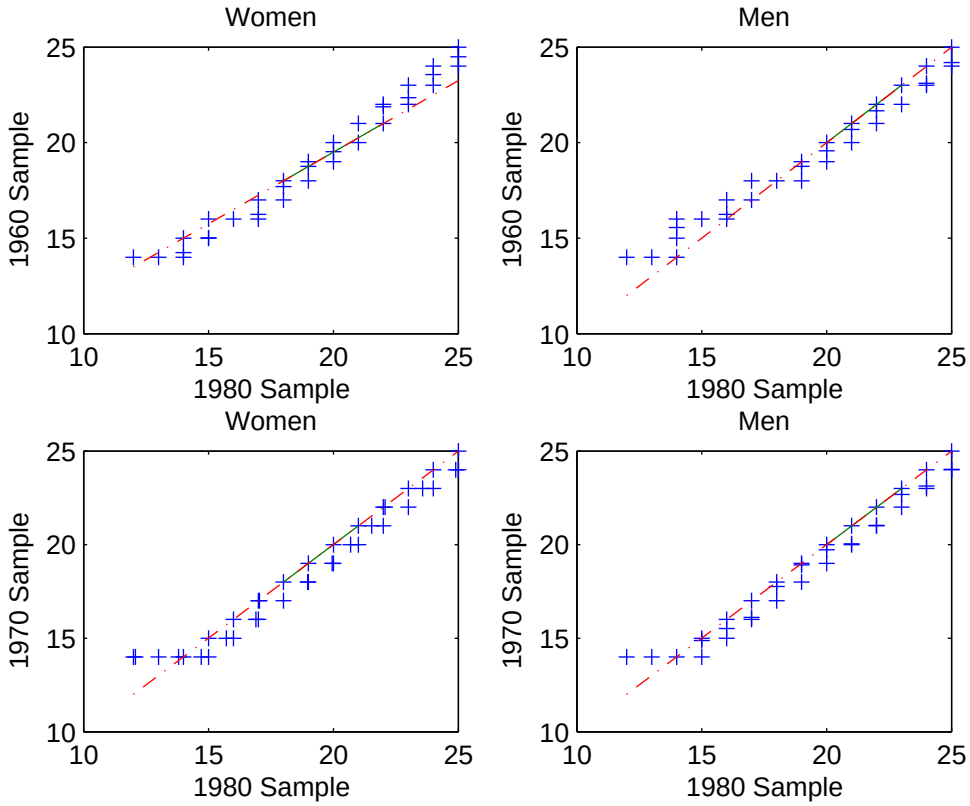

FIGURE 1. QQ plots of age at first marriage for individuals aged at least 26 that are uncensored; i.e., those who married before age 26.

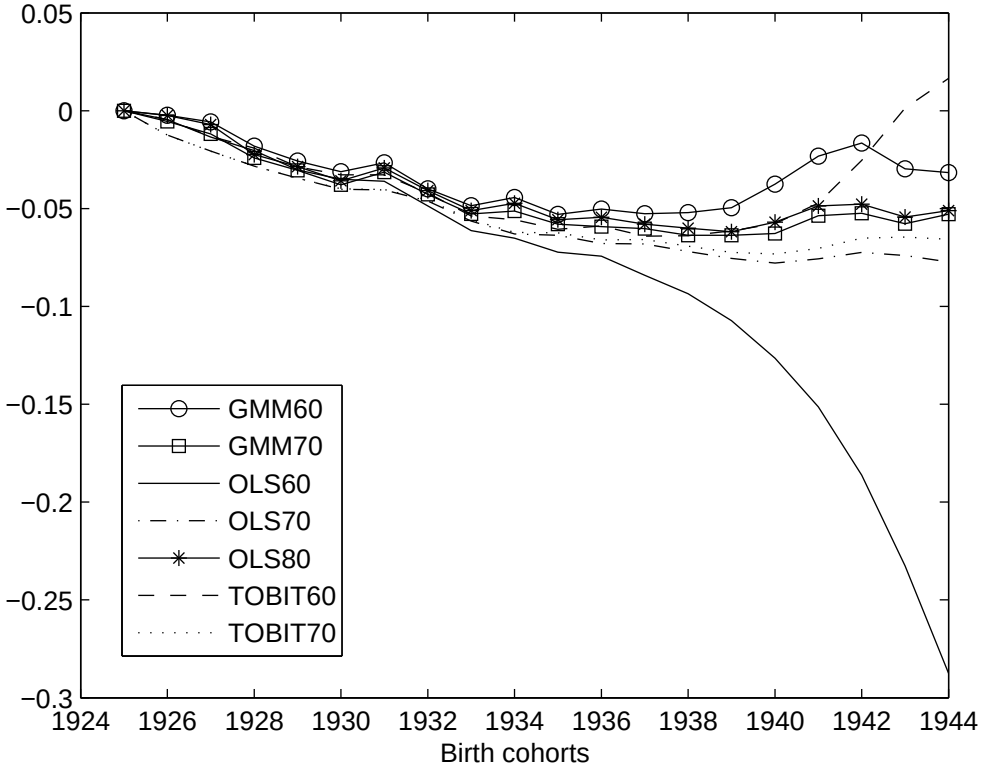

FIGURE 2. Cohort effects for women.

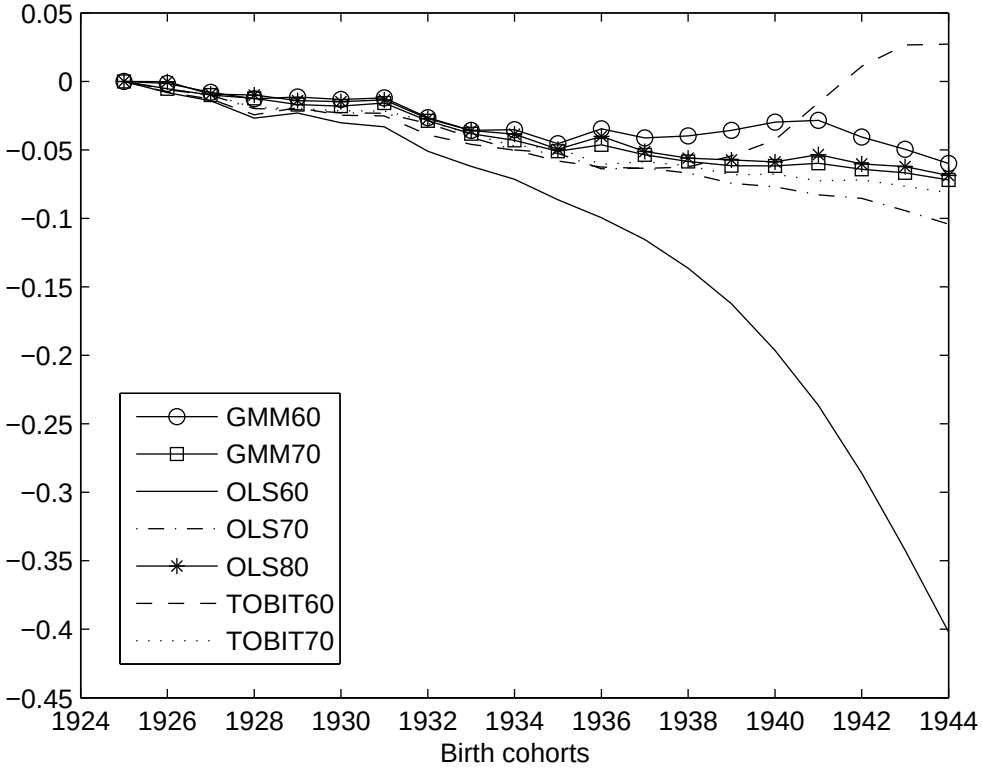

FIGURE 3. Cohort effects for men.

References

Acemoglu, D., and J. Angrist (2001): "How large are human capital externalities? Evidence from compulsory schooling laws," in *NBER Macroeconomics Annual 2000*, ed. by B. S. Bernanke, and K. Rogoff. MIT Press.

Akerlof, G. A. (1998): "Men without children," *Economic Journal*, 108, 287–309.

Amemiya, T. (1984): "Tobit models: A survey," *Journal of Econometrics*, 4, 3–61.

Bergstrom, T., and R. F. Schoeni (1996): "Income prospects and age-at-marriage," *Journal of Population Economics*, 9, 115–130.

Blundell, R. W., and R. J. Smith (1993): "Simultaneous microeconometric models with censored or qualitative dependent variables," in *Econometrics*. Amsterdam: North-Holland, vol. 11 of *Handbook of Statistics*, pp. 117–143.

Brown, B. W., and W. K. Newey (2002): "GMM, efficient bootstrapping, and improved inference," *Journal of Business and Economic Statistics*, 20, 507–517.

Buchinsky, M., and J. Hahn (1998): "An alternative estimator for the censored quantile regression model," *Econometrica*, 66, 653–672.

Carroll, R., D. Ruppert, and L. Stefanski (1995): *Measurement error in nonlinear models*. Chapman & Hall/CRC.

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

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

Chen, S., and S. Khan (2001): "Estimation of a partially linear censored regression model," *Econometric Theory*, 17, 567–590.

Chen, X., H. Hong, and E. Tamer (2004): "Measurement error models with auxiliary data," *Review of Economic Studies*, Forthcoming.

Coale, A. (1971): "Age patterns of marriage," *Population Studies*, 25, 193–214.

Duncan, G. (1986): "A semiparametric censored regression estimator," *Journal of Econometrics*, 32, 5–34.

Fernandez, L. (1986): "Nonparametric maximum likelihood estimation of censored regression models," *Journal of Econometrics*, 32, 35–57.

Goldin, C., and L. F. Katz (2002): "The power of the Pill: Oral contraceptives and womens career and marriage decisions," *Journal of Political Economy*, 110, 730–770.

——— (2003): "Mass secondary schooling and the State: The role of state compulsion in the high school movement," NBER Working Paper 10075.

Hansen, L. P. (1982): "Large sample properties of generalized methods of moments estimators," *Econometrica*, 50, 1029–1054.

Hausman, J. A., and D. A. Wise (1976): "The evaluation of results from truncated samples: The New Jersey negative income tax experiment," *Annals of Economic and Social Measurement*, 5, 421–445.

——— (1977): "Social experimentation, truncated distributions, and efficient estimation," *Econometrica*, 45, 919–938.

Heckman, J. J. (1976): "The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models," *Annals of Economic and Social Measurement*, 5/4, 475–492.

——— (1979): "Sample selection bias as a specification error," *Econometrica*, 47, 153–161.

Hellerstein, J., and G. W. Imbens (1999): "Imposing moment restrictions from auxiliary data by weighting," *Review of Economics and Statistics*, 81, 1–14.

Hirano, K., G. W. Imbens, G. Ridder, and D. B. Rubin (2001): "Combining panel data sets with attrition and refreshment samples," *Econometrica*, 69, 1645–1659.

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

Horowitz, J. L. (1986): "A distribution-free least squares estimator for censored linear regression models," *Journal of Econometrics*, 32, 59–84.

——— (1988): "Semiparametric M-estimation of censored linear regression models," *Advances in Econometrics*, 7, 45–83.

Horowitz, J. L., and C. F. Manski (1998): "Censoring of outcomes and regressors due to survey nonresponse: Identification and estimation using weights and imputations," *Journal of Econometrics*, 84, 37–58.

Imbens, G. W. (1997): "One-step estimators for over-identified generalized method of moments models," *Review of Economic Studies*, 64, 359–383.

Imbens, G. W., and T. Lancaster (1994): "Combining micro and macro data in microeconomic models," *Review of Economic Studies*, 61, 655–680.

Khan, S., and A. Lewbel (2003): "Weighted and two stage least squares estimation of semiparametric truncated regression models," Manuscript.

Khan, S., and J. L. Powell (2001): "Two step estimation of semiparametric censored regression models," *Journal of Econometrics*, 103, 73–110.

Kitamura, Y. (1997): "Empirical likelihood methods with weakly dependent processes," *Annals of Statistics*, 25, 2084–2102.

——— (2001): "Asymptotic optimality of empirical likelihood for testing moment restrictions," *Econometrica*, 69, 1661–1672.

Koball, H. (1998): "Have African American men become less committed to marriage? Explaining the twentieth century racial cross-over in mens marriage timing," *Demography*, 35, 251–258.

Korenman, S., and D. Neumark (1991): "Does marriage really make men more productive," *Journal of Human Resources*, 26, 282–307.

——— (1992): "Marriage, motherhood, and wages," *Journal of Human Resources*, 27, 233–255.

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

——— (1993b): "Winsorized mean estimator for censored regression model," *Econometric Theory*, 8, 368–382.

Lleras-Muney, A. (2001): "Were compulsory attendance and child labor laws effective? An analysis from 1915 to 1939," *Journal of Law and Economics* (forthcoming).

——— (2002): "The relationship between education and adult mortality in the United States," NBER working paper 8986.

Lochner, L., and E. Moretti (2004): "The effect of education on crime: Evidence from prison inmates, arrests, and self-reports," *American Economic Review*, 94, 155–189.

Maddala, G. (1983): *Limited-dependent and qualitative variables in econometrics*. Cambridge University Press.

Manski, C. F. (1989): "Anatomy of the selection problem," *Journal of Human Resources*, 24, 343–360.

——— (1995): *Identification Problems in the Social Sciences*. Harvard University Press.

Manski, C. F., and S. R. Lerman (1977): "The estimation of choice probabilities from choice based samples," *Econometrica*, 45, 1977–1988.

Nan, B., M. Emond, and J. A. Wellner (2002): "Information bounds for Cox regression models with missing data," Manuscript.

Nevo, A. (2003): "Using weights to adjust for sample selection when auxiliary information is available," *Journal of Business and Economic Statistics*, 21, 43–52.

Newey, W. K. (1988): "Efficient estimation of Tobit models under symmetry," in *Nonparametric and semiparametric methods in econometrics and statistics. Proceedings of the fifth international symposium in Economic Theory and Econometrics*, ed. by W. A. Barnett, J. Powell, and G. Tauchen, pp. 291–336. Cambridge University Press.

Newey, W. K., and D. McFadden (1994): "Large sample estimation and hypothesis testing," in *Handbook of Econometrics, vol. IV*, ed. by R. Engle, and D. McFadden. Elsevier Science B.V., pp. 2111–2245.

Newey, W. K., and J. L. Powell (1990): "Efficient estimation of linear and type I censored regression models under conditional quantile restrictions," *Econometric Theory*, 6, 295–317.

Newey, W. K., and R. J. Smith (2003): "Higher order properties of GMM and generalized empirical likelihood estimators," *Econometrica*, Forthcoming.

Owen, A. (1988): "Empirical likelihood ratio confidence intervals for a single functional," *Biometrika*, 75(2), 237–249.

——— (2001): *Empirical likelihood*. Chapman and Hall/CRC.

Powell, J. L. (1983): "The asymptotic normality of two stage least absolute deviations estimators," *Econometrica*, 51, 1569–1575.

——— (1984): "Least absolute deviations estimation for the censored regression model," *Journal of Econometrics*, 25, 303–325.

——— (1986a): "Censored regression quantiles," *Journal of Econometrics*, 32, 143–155.

——— (1986b): "Symmetrically trimmed least squares estimation for tobit models," *Econometrica*, 54, 1435–1460.

——— (1994): "Estimation of semiparametric models," in *Handbook of Econometrics, vol. IV*, ed. by R. Engle, and D. McFadden. Elsevier Science B.V., pp. 2443–2521.

Qin, J., and J. Lawless (1994): "Empirical likelihood and general estimating equations," *Annals of Statistics*, 22, 300–325.

Ridder, G. (1992): "An empirical evaluation of some models for non-random attrition in panel data," *Structural Change and Economic Dynamics*, 3, 337–355.

Robins, J. M., F. Hsieh, and W. Newey (1995): "Semiparametric efficient estimation of a conditional density with missing or mismeasured covariates," *Journal of the Royal Statistical Society, Series B*, 57, 409–424.

Robins, J. M., A. Rotnitzky, and L. P. Zhao (1994): "Estimation of regression coefficients when some regressors are not always observed," *Journal of the American Statistical Association*, 89, 846–866.

Smith, R. J. (1997): "Alternative semi-parametric likelihood approaches to generalized method of moments estimation," *Economic Journal*, 107, 503–519.

Smith, R. J., and R. W. Blundell (1986): "An exogeneity test for a simultaneous equation tobit model with an application to labor supply," *Econometrica*, 54, 679–685.

Titterington, D. (1983): "Kernel-based density estimation using censored, truncated or grouped data," *Communications in Statistics, Theory and Methods*, 12, 2151–2167.

Titterington, D., and G. Mill (1983): "Kernel-based density estimates from incomplete data," *Journal of the Royal Statistical Society, Series B*, 45, 258–266.

Tripathi, G. (2004): "Moment based inference with stratified data," Manuscript. Department of Economics, University of Connecticut-Storrs.

Vardi, Y. (1985): "Empirical distributions in selection biased models," *Annals of Statistics*, 13, 178–203.

Vella, F. (1998): "Estimating models with sample selection bias: A survey," *Journal of Human Resources*, 33, 127–172.

Wansbeek, T., and E. Meijer (2000): *Measurement error and latent variables in econometrics*. North-Holland, Amsterdam.

Department of Economics, University of California, Los Angeles, CA-90095.
E-mail address: devereux@econ.ucla.edu
URL: www.econ.ucla.edu/devereux

Department of Economics, University of Connecticut, Storrs, CT-06269.
E-mail address: gautam.tripathi@uconn.edu
URL: web.uconn.edu/tripathi