University of Connecticut

Department of Economics Working Paper Series

Optimally Combining Censored and Uncensored Datasets

Paul J. Devereux
UCLA

Gautam Tripathi
University of Connecticut

Working Paper 2005-10R

April 2005, revised October 2007

---

341 Mansfield Road, Unit 1063
Storrs, CT 06269–1063
Phone: (860) 486–3022
Fax: (860) 486–4463
http://www.econ.uconn.edu/

Abstract

Economists and other social scientists often face situations where they have access to two datasets that they can use but one set of data suffers from censoring or truncation. If the censored sample is much bigger than the uncensored sample, it is common for researchers to use the censored sample alone and attempt to deal with the problem of partial observation in some manner. Alternatively, they simply use only the uncensored sample and ignore the censored one so as to avoid biases. It is rarely the case that researchers use both datasets together, mainly because they lack guidance about how to combine them. In this paper, we develop a simple semiparametric framework for combining the censored and uncensored datasets so that the resulting estimators are consistent, asymptotically normal, and use all information optimally. No nonparametric smoothing is required to implement our estimators. 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 combining cohort information across census datasets. Results from a small simulation experiment suggest that the estimator proposed in this paper can work very well in finite samples.

Journal of Economic Literature Classification: C14, C24, C34, C51

Keywords: Censoring, Empirical Likelihood, GMM, Refreshment samples, Truncation

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, they simply use the uncensored sample and ignore the censored one so as to avoid biases. It is rare 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 based on the generalized method of moments (GMM) that allows the censored and uncensored samples — henceforth referred to as the "master" and "refreshment" samples, respectively — to be combined in a tractable manner so that the resulting estimators are consistent, asymptotically normal, and use all information optimally. In fact, we show that using the refreshment 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. The existence of refreshment samples should not be regarded as being an overly restrictive requirement; as we show in Section 6, they can often be constructed by creatively combining existing datasets.

Semiparametric inference with censored or truncated data thus far seems to have focused mainly on linear regression models where only the response variable is censored or truncated. The present work extends the literature in a significant manner to include nonlinear models and multiple censored or truncated variables. In particular, we demonstrate how efficiently combining two datasets allows standard moment based inference with censored or truncated data to go through without imposing parametric, independence, symmetry, quantile, or "special regressor" restrictions as done in the existing literature, and without doing any nonparametric smoothing.[2] The biggest appeal is the simplicity of our estimators. For instance, 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 is censored or truncated is straightforward and the usual analogy principle that delivers standard errors for GMM works here as well. The treatment proposed here is general enough to handle censoring and truncation of some or all coordinates of both endogenous

---

[1] A comprehensive review of the econometric literature on censoring and truncation is beyond the scope of our paper; see, e.g., the surveys by Amemiya (1984), Blundell and Smith (1993), and Powell (1994).

[2] Although the basic idea of combining datasets has been explored earlier in other contexts, see, for instance, Angrist and Kreuger (1992), Arellano and Meghir (1992), Hirano, Imbens, Ridder, and Rubin (2001), Hu and Ridder (2003), Ridder and Moffitt (2003), Chen, Hong, and Tarozzi (2004), Chen, Hong, and Tamer (2005), Ichimura and Martinez-Sanchis (2005), and the references therein, its use to facilitate efficient moment based inference in overidentified models with censored or truncated data seems to be new to the literature.

and exogenous variables and our results are applicable to a large class of models which nest linear regression as a special case; e.g., the ability to handle instrumental variables (IV) models permits semiparametric inference in Box-Cox type models using censored or truncated data without imposing parametric or quantile restrictions.[3] Access to the refreshment sample also means that incompleteness of the data does not complicate identification conditions.

The paper is organized as follows. In Section 2 we set up the problem of censoring or truncation of random vectors in a moment based framework. Section 3 models the data combination process, and Section 4 shows how censored data can be combined with a refreshment sample to do efficient semiparametric inference (Section 5 does the same with truncated data). Section 6 contains an interesting application where refreshment samples are obtained by combining census datasets, and in Section 7 we describe the results of a small Monte-Carlo experiment to study the finite sample properties of our estimator. We conclude by addressing some topics for future research in Section 8.

2. Censoring and truncation in a moment based framework

Let (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 $\mathbb{R}^d$ that denotes an observation from the target population (following usual mathematical convention, "vector" means a column vector), and f^* is the unknown density of Z^* w.r.t some dominating measure $\mu^* := \times_{i=1}^d \mu_i^*$; since Z^* can have discrete components, the μ_i^*'s need not all be Lebesgue measures. Similarly, (Z, f, μ) represents the "realized" population, i.e., the observed data, with Z the resulting observation and f its density w.r.t a dominating measure $\mu := \times_{i=1}^d \mu_i$. In this paper, f is different from f^* because some or all coordinates of Z^* are censored, or, truncated.

The econometric models we consider can be expressed as moment conditions in the target population.[4] So let Θ be a subset of $\mathbb{R}^p$ such that

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

where g is a $q \times 1$ vector of known functions with $q \geq p$ and $\mathbb{E}_{f^*}$ denotes expectation w.r.t f^*. Well-known examples of (2.1) include linear and nonlinear regression models and multivariate simultaneous equations models. The class of models defined in (2.1) also contains IV models derived from conditional moment restrictions in the target population.

[3]In Appendix C we show that the GMM based results continue to hold for the empirical likelihood approach proposed by Owen (1988) that is rapidly gaining popularity in econometrics.

[4]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 that economic agents are assumed to solve. Hence, such models are particularly important for structural estimation.

2.1. **Censoring.** If Z^* is fully observed, then (2.1) is easily handled. But in many cases economists cannot fully observe Z^*. For instance, variables often get censored due to administrative reasons; e.g., government agencies routinely "top-code" income data before releasing it for public use or 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;[5] i.e., instead of observing $Z^* := (Z^{*(1)}, \ldots, Z^{*(d)})_{d\times 1}$ we observe the random vector $Z := (Z^{(1)}, \ldots, Z^{(d)})_{d\times 1}$, where

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

and $c := (c^{(1)}, \ldots, c^{(d)})$ is a $d \times 1$ vector of known constants.[6]

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 censoring, 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), 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 $-c^{(i,j,k)}$; e.g., if $Z^{*(i)}$ is left censored by $c^{(i)}$, then simply redefine the corresponding realization to be $-Z^{*(i)}$ if $-Z^{*(i)} < -c^{(i)}$, and $-c^{(i)}$ otherwise. The case we do not cover in this paper is that of interval censoring where the same coordinate is subject to left and right censoring simultaneously.

Let $S^*(c) := \Pr_{f^*}(Z^{*(1)} > c^{(1)}, \ldots, Z^{*(d)} > c^{(d)})$ denote the probability that all coordinates of Z^* are censored and δ_c be the Dirac measure at c, i.e., $\delta_c(A) := \mathbb{1}(c \in A)$, where $\mathbb{1}$ is the indicator function. To keep matters simple, we assume that μ^* does not place any mass at c; consequently, applied researchers should not use this methodology when a censored variable happens to be discrete and the censoring threshold lies in its support. 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 w.r.t the dominating measure $\mu := \mu^* + \delta_c$ is given by

$$f(z) := f^*(z)\mathbb{1}(z < c) + S^*(c)\mathbb{1}(z = c). \tag{2.2}$$

[5]The definitions of censoring and truncation we use in this paper are standard in the literature; see, e.g., Hajivassiliou and Ruud (1994).

[6]The results obtained in this paper continue to hold in a more general fixed censoring framework where the censoring threshold is modeled 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.

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 $f^*_{i,j,k|-(i,j,k)}$ the conditional density of $Z^{*(i,j,k)}$ given $Z^{*-(i,j,k)}$. Then, letting $S^*_{i,j,k|-(i,j,k)}(c^{(i,j,k)})$ denote the conditional probability that $Z^{*(i,j,k)}$ are censored given $Z^{*-(i,j,k)}$, it is easy to show that for $d > 1$ the density of Z w.r.t $\mu := \times_{i=1}^{d}\mu_i$, where $\mu_i := \mu_i^* + \delta_{c^{(i)}}$, is given by

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

where $\overset{\text{elt}}{<}$ denotes element-by-element strict inequality, i.e., $\mathbb{1}(z \overset{\text{elt}}{<} c) := \prod_{i=1}^{d} \mathbb{1}(z^{(i)} < c^{(i)})$, and $z = c$ is element-by-element equality, i.e., $\mathbb{1}(z = c) = \prod_{i=1}^{d} \mathbb{1}(z^{(i)} = c^{(i)})$. Note that the realized density f has support $(-\infty, c^{(1)}] \times \ldots \times (-\infty, c^{(d)}]$ with a mass point 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 Z^* we observe

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

where T is a known region in $\mathbb{R}^d$ such that Z^* lies in T with positive probability. In this case, the density of Z w.r.t μ^* is given by

$$f(z) := \frac{f^*(z)\mathbb{1}(z \in T)}{\int_T f^*(z)\, d\mu^*}. \quad (2.4)$$

Note that f has support T. 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 intervals. If, say, $Z^{*(i,j,k)}$ are not truncated, then simply let $I_i = I_j = I_k = \mathbb{R}$.

2.3. **Examples.** We now look at some examples of censoring and truncation in a multivariate framework. The primary aim is to illustrate the behavior of least squares estimators in linear regression models when only the master sample is used for estimation and more than one variable is censored or truncated; examples 2.2 and 2.4 are particularly instructive. Since no refreshment sample is used in this section, n here is just the master sample size.

Example 2.1 (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 $Z_1, \dots, Z_n$ 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, but

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

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

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

denotes h evaluated at exactly r censored coordinates, and $Z^*[i_1, \dots, i_r]$ stands for Z^* with its i_1, …, i_rth coordinates replaced by $c^{(i_1)}, \dots, c^{(i_r)}$, respectively, and the remaining coordinates unchanged. Hence, $\mathbb{E}_f\{Z\} \neq \mathbb{E}_{f^*}\{Z^*\}$. □

Example 2.2 (Censored linear regression)**.** Let $Y^* = X^{*\prime}\theta^* + \varepsilon^*$, where $\mathbb{E}_{f^*}\{X^*\varepsilon^*\} = 0$. Therefore, $\theta^* = (\mathbb{E}_{f^*} X^* X^{*\prime})^{-1}(\mathbb{E}_{f^*} X^* Y^*)$. Suppose 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

$$(\mathbb{E}_f XX')^{-1}(\mathbb{E}_f XY) = (\mathbb{E}_{f^*}\{X^* X^{*\prime} \mathbb{1}(Y^* < c^{(1)}, X^* \overset{\text{elt}}{<} c^{-(1)}) + \sum_{r=1}^{d-1} (X^* X^{*\prime})_{[r]} + c^{-(1)} c^{-(1)\prime} S^*(c)\})^{-1}$$
$$\times \mathbb{E}_{f^*}\{X^* Y^* \mathbb{1}(Y^* < c^{(1)}, X^* \overset{\text{elt}}{<} c^{-(1)}) + \sum_{r=1}^{d-1} (X^* Y^*)_{[r]} + c^{-(1)} c^{(1)} S^*(c)\}, \quad (2.5)$$

where $d = p + 1$. Hence, $\text{plim}(\hat{\theta}_M) \neq \theta^*$.

The special case of pure endogenous censoring, called the tobit or limited dependent variable model in econometrics, is obtained by letting $c^{-(1)} = (\infty, \dots, \infty)$ and using the convention that $0 \cdot \infty = 0$. Doing so, (2.5) implies that

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

as is well known from tobit theory.

However, a fact that does not seem to be as widely known is that the least squares estimator remains inconsistent even if censoring is purely exogenous; see Rigobon and Stoker (2003) for more on this. In particular, by letting $c^{(1)} = \infty$ in (2.5), we can see that

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

Thus 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. To see this, consider the simple linear regression model $Y^* = \theta^{*(1)} + X^*\theta^{*(2)} + \varepsilon^*$, where X^* is scalar and $\mathbb{E}_{Y^*|X^*}\{\varepsilon^*|X^*\} = 0$ w.p.1. Since Y^* and the constant regressor are not censored, $c = (\infty, \infty, c^{(3)})_{3\times 1}$. Hence, by (2.5),

$$\begin{aligned}\text{plim}(\hat{\theta}_M^{(2)}) = \frac{\text{cov}_f(Y,X)}{\text{var}_f(X)} &= \frac{\text{cov}_{f^*}(Y^*, X^*\mathbb{1}(X^* < c^{(3)}) + c^{(3)}\mathbb{1}(X^* > c^{(3)}))}{\text{var}_{f^*}(X^*\mathbb{1}(X^* < c^{(3)}) + c^{(3)}\mathbb{1}(X^* > c^{(3)}))} \\ &= \frac{\text{cov}_{f^*}(X^*, X^*\mathbb{1}(X^* < c^{(3)}) + c^{(3)}\mathbb{1}(X^* > c^{(3)}))}{\text{var}_{f^*}(X^*\mathbb{1}(X^* < c^{(3)}) + c^{(3)}\mathbb{1}(X^* > c^{(3)}))}\theta^{*(2)},\end{aligned}$$

where the last equality follows because $\mathbb{E}_{Y^*|X^*}\{Y^*|X^*\} = \theta^{*(1)} + X^*\theta^{*(2)}$ w.p.1. Therefore, $\hat{\theta}_M$ is inconsistent under pure exogenous censoring although ε^* is mean independent of X^*. However, as shown in Example 2.4, the situation changes if X^* is truncated instead of censored. □

Example 2.3 (Truncated mean)**.** Suppose we want to estimate the mean of the target population but now Z^* is truncated outside region T. Since $\mathbb{E}_f\{Z\} = \mathbb{E}_{f^*}\{Z^*\mathbb{1}(Z^* \in T)\}/\int_T f^*(z)\,d\mu^*$, as in Example 2.1 the naive estimator is not consistent for $\mathbb{E}_{f^*}\{Z^*\}$. □

Example 2.4 (Truncated linear regression)**.** Consider the linear model of Example 2.2 but now suppose that, instead of being censored, Z^* is truncated outside $T := T_1 \times T_2$. Since now

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

$\hat{\theta}_M$ is not consistent for θ^*. Under pure endogenous truncation, i.e., $T_2 = \mathbb{R}^p$, we get that

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

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

$$\text{plim}(\hat{\theta}_M) = \{\mathbb{E}_{f^*}X^*X^{*\prime}\mathbb{1}(X^* \in T_2)\}^{-1}\mathbb{E}_{f^*}\{X^*Y^*\mathbb{1}(X^* \in T_2)\} \neq \theta^*. \tag{2.6}$$

Therefore, even pure exogenous truncation is not ignorable. But, unlike Example 2.2, 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 (2.6) it is easy to see that ignoring pure exogenous truncation does not make the least squares estimator inconsistent. □

3. Data combination

As in Tripathi (2007), 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 or the master sample; i.e., $R = 1$ if Z is from the refreshment sample and $R = 0$ if Z belongs to the master sample. Hence, for $r \in \{0,1\}$, the conditional density of $Z|R = r$ is given by[7]

$$f_{Z|R=r}(z) := \begin{cases} f^*(z)\mathbb{1}(z \overset{\text{elt}}{\neq} c)r + f(z)(1-r) & \text{if } Z^* \text{ is censored} \\ f^*(z)r + f(z)(1-r) & \text{if } Z^* \text{ is truncated,} \end{cases} \tag{3.1}$$

where $\mathbb{1}(z \overset{\text{elt}}{\neq} c) := \prod_{i=1}^{d} \mathbb{1}(z^{(i)} \neq c^{(i)})$ 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 w.r.t μ and has a mass point at c. On the other hand, if Z^* is truncated then $f_{Z|R=r}$ is a conditional density w.r.t μ^*.

Let $R \overset{d}{\sim} \text{Bernoulli}(K_0)$, where $K_0 \in (0,1)$ is an unknown nuisance parameter that will be estimated along with the parameters of interest. Therefore, using (3.1), the joint density of Z and R is given by

$$f_e(z,r) := \begin{cases} K_0 f^*(z)\mathbb{1}(z \overset{\text{elt}}{\neq} c)r + (1-K_0)f(z)(1-r) & \text{if } Z^* \text{ is censored} \\ K_0 f^*(z)r + (1-K_0)f(z)(1-r) & \text{if } Z^* \text{ is truncated.} \end{cases} \tag{3.2}$$

Henceforth, let n denote the size of the "enriched" sample, i.e., the master and refreshment samples combined together. Observations $(Z_1, R_1), \ldots, (Z_n, R_n)$ from the enriched dataset are regarded as iid draws from f_e — a density w.r.t $\mu \times \kappa$, where κ is the counting measure on $\{0,1\}$ — and all limits are taken as $n \uparrow \infty$. 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).

A technical remark: Introducing the refreshment dummy R allows the combined sample to be treated as a collection of iid draws from the enriched density f_e, which greatly simplifies the mathematical treatment (because an iid setting makes it easier to calculate efficiency bounds, apply standard statistical arguments to prove our results, etc.) although it makes the refreshment sample size $\sum_{j=1}^{n} R_j$ a random variable. However, as shown later in Sections 4 and 5, inference about θ^* is actually conditional on the observed value of $\sum_{j=1}^{n} R_j$ because we estimate θ^* jointly and efficiently with K_0. Therefore, our results coincide with those obtained in a setting where the size of the refreshment sample is nonstochastic and observations from the combined sample are regarded as being independent but not identically distributed.

4. Inference with censored data

The marginal density of Z in the enriched sample is, by (3.2),

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

[7] Since f^* is a density w.r.t μ^*, it is only identified up to sets of μ^*-measure zero. Thus if Z^* is censored then $f^*(z)\mathbb{1}(z \overset{\text{elt}}{\neq} c)$ is a μ^*-version of f^*. Hence, $\mathbb{E}_{f^*}\{g(Z^*,\theta^*)\} = 0$ if and only if $\mathbb{E}_{f^*}\{g(Z^*,\theta^*)\mathbb{1}(Z^* \overset{\text{elt}}{\neq} c)\} = 0$.

Hence, letting $a(z, K_0) := K_0 + (1 - K_0)\mathbb{1}(z \overset{\text{elt}}{<} c)$, by (2.2) and (2.3) it follows that

$$f^*(z)\mathbb{1}(z \overset{\text{elt}}{\neq} c) = \int_{r\in\{0,1\}} f_e(z,r)\mathbb{1}(z \overset{\text{elt}}{\neq} c)\, d\kappa/a(z,K_0). \tag{4.1}$$

Therefore, since $\mathbb{E}_{f^*}\{g(Z^*, \theta^*)\} = 0$ if and only if $\mathbb{E}_{f^*}\{g(Z^*, \theta^*)\mathbb{1}(Z^* \overset{\text{elt}}{\neq} c)\} = 0$, we can use (4.1) to write (2.1) in terms of the enriched density as

$$\mathbb{E}_{f_e}\{g(Z,\theta^*)\mathbb{1}(Z \overset{\text{elt}}{\neq} c)/a(Z,K_0)\} = 0. \tag{4.2}$$

However, (4.1) also implies that[8]

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

where $(Z \overset{\text{elt}}{<} c)^{\sim}$ denotes the set-complement of the event $(Z \overset{\text{elt}}{<} c)$. Furthermore, since

$$\mathbb{E}_{f_e}\{R - K_0\} = 0, \tag{4.4}$$

efficient estimation of θ^* must account for this restriction as well.

For notational convenience, define $\beta^* := (\theta^*, K_0)_{(p+1)\times 1}$ and

$$\rho(Z,R,\beta) := \begin{bmatrix} g(Z,\theta)\mathbb{1}(Z \overset{\text{elt}}{\neq} c)/a(Z,K) \\ \mathbb{1}(Z \overset{\text{elt}}{\neq} c)\mathbb{1}(Z \overset{\text{elt}}{<} c)^{\sim} - K\mathbb{1}(Z \overset{\text{elt}}{<} c)^{\sim} \\ R - K \end{bmatrix} := \begin{bmatrix} \rho_1(Z,\beta) \\ \rho_2(Z,K) \\ \rho_3(R,K) \end{bmatrix}_{(q+2)\times 1}. \tag{4.5}$$

The two-step optimal GMM estimator of β^* is then given by $\tilde{\beta} := \operatorname{argmin}_{\beta\in\mathcal{B}} \hat{\rho}'(\beta)\breve{V}_\rho^{-1}\hat{\rho}(\beta)$, where $\mathcal{B} := \Theta \times [0,1]$, $\hat{\rho}(\beta) := \sum_{j=1}^n \rho(Z_j, R_j, \beta)/n$, and $\breve{V}_\rho := \sum_{j=1}^n \rho(Z_j, R_j, \breve{\beta})\rho'(Z_j, R_j, \breve{\beta})/n$ is an estimator of the optimal weighting matrix using a preliminary estimator $\breve{\beta}$.

Let $\|\cdot\|$ denote the Euclidean norm. The following standard regularity conditions ensure that GMM estimators are consistent and asymptotically normal.

Assumption 4.1. (i) *$\beta^* \in \mathcal{B}$ is the unique solution to $\mathbb{E}_{f_e}\{\rho(Z,R,\beta)\} = 0$;* (ii) *$\mathcal{B}$ is compact;* (iii) *$\rho(Z,R,\beta)$ is continuous at each $\beta \in \mathcal{B}$ w.p.1;* (iv) *$\mathbb{E}_{f_e}\{\sup_{\beta\in\mathcal{B}} \|\rho(Z,R,\beta)\|^2\} < \infty$;*

---

[8]Since $\{\mathbb{1}(Z \overset{\text{elt}}{<} c) + \mathbb{1}(Z \overset{\text{elt}}{<} c)^{\sim}\}/a(Z,K_0) = \mathbb{1}(Z \overset{\text{elt}}{<} c) + \mathbb{1}(Z \overset{\text{elt}}{<} c)^{\sim}/K_0$ and $\mathbb{1}(Z \overset{\text{elt}}{\neq} c)\mathbb{1}(Z \overset{\text{elt}}{<} c) = \mathbb{1}(Z \overset{\text{elt}}{<} c)$,

$$\begin{aligned} \mathbb{E}_{f_e}\{\mathbb{1}(Z \overset{\text{elt}}{\neq} c)/a(Z,K_0)\} = 1 &\iff \mathbb{E}_{f_e}\{\mathbb{1}(Z \overset{\text{elt}}{\neq} c)\mathbb{1}(Z \overset{\text{elt}}{<} c)\} + \mathbb{E}_{f_e}\{\mathbb{1}(Z \overset{\text{elt}}{\neq} c)\mathbb{1}(Z \overset{\text{elt}}{<} c)^{\sim}\}/K_0 = 1 \\ &\iff \mathbb{E}_{f_e}\{\mathbb{1}(Z \overset{\text{elt}}{\neq} c)\mathbb{1}(Z \overset{\text{elt}}{<} c)^{\sim}\} = K_0\mathbb{E}_{f_e}\{\mathbb{1}(Z \overset{\text{elt}}{<} c)^{\sim}\} \\ &\iff \mathbb{E}_{f_e}\{\mathbb{1}(Z \overset{\text{elt}}{\neq} c)\mathbb{1}(Z \overset{\text{elt}}{<} c)^{\sim} - K_0\mathbb{1}(Z \overset{\text{elt}}{<} c)^{\sim}\} = 0. \end{aligned}$$

Hence, the equivalence in (4.3) holds.

(v) $\mathbb{E}_{f_e}\{\rho(Z,R,\beta^*)\rho'(Z,R,\beta^*)\}$ *is nonsingular;* (vi) $\beta^* \in \text{int}(\mathcal{B})$*;* (vii) $\rho(Z,R,\beta)$ *is continuously differentiable in a neighborhood* $\mathcal{N}$ *of* β^* *and* $\mathbb{E}_{f_e}\{\sup_{\beta\in\mathcal{N}}\|\partial\rho(Z,R,\beta)/\partial\beta\|\} < \infty$*;* (viii) $\mathbb{E}_{f_e}\{\partial\rho(Z,R,\beta^*)/\partial\beta\}$ *is of full column rank.*

(i)–(v) can be used to prove consistency and (vi)–(viii) to prove the asymptotic normality of GMM estimators as in Newey and McFadden (1994). Note that the consistency of our estimators does not depend upon the extent to which the data are censored.

Now let $D := \mathbb{E}_{f_e}\{\partial\rho_1(Z,\beta^*)/\partial\theta\}$, $V_1 := \mathbb{E}_{f_e}\{\rho_1(Z,\beta^*)\rho_1(Z,\beta^*)'\}$, $V_2 := \mathbb{E}_{f_e}\{\rho_2^2(Z,K_0)\}$, $V_3 := \mathbb{E}_{f_e}\{\rho_3^2(R,K_0)\}$, $\Sigma_{12} := \mathbb{E}_{f_e}\{\rho_1(Z,\beta^*)\rho_2(Z,K_0)\}$, $\Sigma_{13} := \mathbb{E}_{f_e}\{\rho_1(Z,\beta^*)\rho_3(R,K_0)\}$, and $\Omega := \mathbb{E}_{f_e}\{\varepsilon\varepsilon'\}$, where $\varepsilon := \rho_1(Z,\beta^*) - \text{Proj}\{\rho_1(Z,\beta^*)\big|1,\rho_2(Z,K_0),\rho_3(R,K_0)\}$ is the residual from the linear projection (under f_e) of $\rho_1(Z,\beta^*)$ onto the span of 1, $\rho_2(Z,K_0)$, and $\rho_3(R,K_0)$. The next result is shown in Appendix A.

Theorem 4.1. *Let Assumption* 4.1 *hold with the moment function* $\rho(Z,R,\beta)$ *defined in* (4.5)*. Then, letting* $0_{k\times1}$ *denote the* $k\times1$ *vector of zeros and* $0'_{k\times1}$ *its transpose,*

$$\begin{bmatrix} n^{1/2}(\tilde\theta-\theta^*) \\ n^{1/2}(\tilde K-K_0)\end{bmatrix} \xrightarrow{d} \mathrm{N}(0_{(p+1)\times1}, \begin{bmatrix}(D'\Omega^{-1}D)^{-1} & 0_{p\times1} \\ 0'_{p\times1} & K_0(1-K_0)\end{bmatrix}).$$

In Theorem A.1 of Appendix A we show that $(D'\Omega^{-1}D)^{-1}$ is the efficiency bound for estimating θ^*; therefore, $\tilde\theta$ is asymptotically efficient. Furthermore, Theorem 4.3 shows 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 (4.5) 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 is because

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

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 $\rho_1(Z,\beta^*)$ 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 the moment function

$$\rho(Z,\beta) := \begin{bmatrix} g(Z,\theta)\mathbb{1}(Z \overset{\text{elt}}{\neq} c)/a(Z,K) \\ \mathbb{1}(Z \overset{\text{elt}}{\neq} c)\mathbb{1}(Z \overset{\text{elt}}{<} c)^\sim - K\mathbb{1}(Z \overset{\text{elt}}{<} c)^\sim \end{bmatrix} := \begin{bmatrix}\rho_1(Z,\beta) \\ \rho_2(Z,K)\end{bmatrix}_{(q+1)\times1}. \tag{4.7}$$

This leads to the following result.

Theorem 4.2. *Let Assumption* 4.1 *hold with the moment function* $\rho(Z,\beta)$ *defined in* (4.7) *and let* $\hat{\beta} := (\hat{\theta}, \hat{K})_{(p+1)\times 1}$ *denote the GMM estimator of* β^* *using* (4.7). *Then,*

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

The asymptotic variance of $\hat{\theta}$ is still $(D'\Omega^{-1}D)^{-1}$ although dropping $\rho_3(R, K_0)$ increases the asymptotic variance of $\hat{K}$ as compared to $\tilde{K}$.[9] 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. Since (4.6) implies that ε is just the residual from projecting $\rho_1(Z,\beta^*)$ onto the span of 1 and $\rho_2(Z, K_0)$, it follows that $\Omega = V_1 - \Sigma_{12}\Sigma_{12}'/V_2$. The asymptotic variance of $\hat{\theta}$ can be estimated by replacing D and Ω with consistent estimators $\hat{D} := n^{-1}\sum_{j=1}^n \partial\rho_1(Z_j,\hat{\beta})/\partial\theta$ and $\hat{\Omega} := \hat{V}_1 - \hat{\Sigma}_{12}\hat{\Sigma}_{12}'/\hat{V}_2$, where $\hat{V}_1 := \sum_{j=1}^n \rho_1(Z_j,\hat{\beta})\rho_1'(Z_j,\hat{\beta})/n$, $\hat{\Sigma}_{12} := \sum_{j=1}^n \rho_1(Z_j,\hat{\beta})\rho_2(Z_j,\hat{K})/n$, and $\hat{V}_2 := \sum_{j=1}^n \rho_2^2(Z_j,\hat{K})/n$; equivalently, $\hat{\Omega} = \sum_{j=1}^n \hat{\varepsilon}_j\hat{\varepsilon}_j'/n$, where $\hat{\varepsilon}$ is the residual from regressing $\rho_1(Z,\hat{\beta})$ element-by-element on a constant and $\rho_2(Z,\hat{K})$.

To get some intuition about why transforming the moment condition works, note that since $K_0 \overset{(4.3)}{=} \mathbb{E}_{f_e}\{\mathbb{1}(Z \overset{\text{elt}}{\neq} c)\mathbb{1}(Z \overset{\text{elt}}{<} c)^{\sim}\}/\mathbb{E}_{f_e}\{\mathbb{1}(Z \overset{\text{elt}}{<} c)^{\sim}\}$, we can decompose

$$\begin{aligned} \mathbb{E}_{f_e}\{\rho_1(Z,\beta^*)\} = {} & \mathbb{E}_{f_e}\{g(Z,\theta^*)|Z \overset{\text{elt}}{<} c\}\mathrm{Pr}_{f_e}(Z \overset{\text{elt}}{<} c) \\ & + \mathbb{E}_{f_e}\{g(Z,\theta^*)|(Z \overset{\text{elt}}{\neq} c) \cap (Z \overset{\text{elt}}{<} c)^{\sim}\}\mathrm{Pr}_{f_e}(\{Z \overset{\text{elt}}{<} c\}^{\sim}). \end{aligned} \quad (4.8)$$

Therefore, the moment function 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 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 weights; see Example 4.1 for a nice illustration.

Efficiently estimating θ^* jointly with K_0 ensures that $\hat{\theta}$ and $\sum_{j=1}^n R_j$ are asymptotically independent. To see this, we can use the proof of Theorem 4.2 to show that $\hat{\theta}$ is asymptotically linear with influence function $-(D'\Omega^{-1}D)^{-1}D'\Omega^{-1}\varepsilon$; i.e., we can show that $n^{1/2}(\hat{\theta}-\theta^*) = n^{-1/2}\sum_{j=1}^n -(D'\Omega^{-1}D)^{-1}D'\Omega^{-1}\varepsilon_j + o_p(1)$. But, by the Cramér-Wold device and the central limit theorem, $n^{1/2}(\hat{\theta}-\theta^*)$ and $n^{-1/2}\sum_{j=1}^n(R_j - K_0)$ are jointly asymptotically

[9] For the sake of completeness, note that if $\check{\beta}$ is the GMM estimator of β^* based on $\rho_1(Z,\beta^*)$ 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 $\left[\begin{smallmatrix} (D'\Gamma^{-1}D)^{-1} & 0_{p\times 1} \\ 0'_{p\times 1} & K_0(1-K_0) \end{smallmatrix}\right]$, where $\Gamma := V_1 - \Sigma_{13}\Sigma_{13}'/V_3$. From Lemma A.2, (A.9), 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 $\Gamma \geq \Omega$. Therefore, $(D'\Gamma^{-1}D)^{-1} \geq (D'\Omega^{-1}D)^{-1}$ implying that asymptotically $\hat{\theta}$ is better than $\check{\theta}$.

normal. Therefore, since ε is orthogonal to $\rho_3(R, K_0)$,[10] it follows that $\hat{\theta}$ and $\sum_{j=1}^n R_j$ are asymptotically independent. Consequently, as mentioned at the end of Section 3, inference based on the asymptotic distribution of $\hat{\theta}$ is equivalent to inference based on the asymptotic conditional distribution of $\hat{\theta}$ given $\sum_{j=1}^n R_j$.

Finally, 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

$$\mathbb{E}_{f_e}\{g(Z, \theta^*)|R = 1\} = 0 \iff \mathbb{E}_{f_e}\{g(Z, \theta^*)R\} = 0. \tag{4.9}$$

The next result shows that $\hat{\theta}_R$ is asymptotically inefficient relative to $\hat{\theta}$. 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 $\mathrm{asvar}(\hat{\theta}_R) > \mathrm{asvar}(\hat{\theta})$, *where "*asvar*" is shorthand for "asymptotic variance".*

The inflation factor $1/K_0$ 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. In Remark A.1 after the proof of Theorem 4.3, we show that Ω is a decreasing (in the positive definite sense) function of K_0. Hence, we can expect the finite sample performance of $\hat{\theta}$ to improve as the refreshment sample gets larger.

Although $\hat{\theta}_R$ is asymptotically inefficient for estimating θ^*, it can be used in applied work for specification testing. In particular, by contrasting components of $\hat{\theta}_R$ and $\hat{\theta}$ that may be of particular empirical interest, or even $\hat{\theta}_R$ and $\hat{\theta}$ themselves, we can do a Hausman test of the hypothesis that the master and refreshment samples are drawn from the same population.[11] So let B be a diagonal matrix of ones and zeroes that picks out the coordinates of $\hat{\theta}_R$ and $\hat{\theta}$ to be contrasted (if B is the $p \times p$ identity matrix then $\hat{\theta}_R$ is compared with $\hat{\theta}$), $\hat{V}_{B\hat{\theta}} := B(\hat{D}'\hat{\Omega}^{-1}\hat{D})^{-1}B'/n$ denote the estimated asymptotic variance of $B\hat{\theta}$ and, following the proof of Theorem 4.3, let $\hat{V}_{B\hat{\theta}_R} := B(\hat{D}_R'\hat{\Omega}_R^{-1}\hat{D}_R)^{-1}B'/n$ be the estimated asymptotic variance of $B\hat{\theta}_R$, where $\hat{D}_R := n^{-1}\sum_{j=1}^n \partial g(Z_j, \hat{\theta}_R)R_j/\partial\theta$ and $\hat{V}_R := n^{-1}\sum_{j=1}^n g(Z_j, \hat{\theta}_R)g'(Z_j, \hat{\theta}_R)R_j$. It is then straightforward to show that the Hausman statistic $n(B\hat{\theta}_R - B\hat{\theta})'(\hat{V}_{B\hat{\theta}_R} - \hat{V}_{B\hat{\theta}})^{-1}(B\hat{\theta}_R - B\hat{\theta})$ is asymptotically $\chi^2_{\dim(B\theta^*)}$ under the null hypothesis that the master and refreshment samples come from the same population. Hence, it can be used to test this hypothesis; see the application in Section 6.

[10] This follows immediately from (A.8) in the proof of Lemma A.1.

[11] This is because $\hat{\theta}_R$ is consistent for θ^* even if the master and refreshment samples are from different populations — recall that $\hat{\theta}_R$ is based on the refreshment sample alone — whereas $\hat{\theta}$ is a consistent and efficient estimator of θ^* only when the master and refreshment samples are from the same population.

Example 4.1 (Example 2.1 contd.)**.** Here $\rho_1(Z,\beta) = (Z-\theta)\mathbb{1}(Z \overset{\text{elt}}{\neq} c)/a(Z,K)$ and no overidentifying restrictions. Hence, $(\hat{\theta},\hat{K})$ solve $\sum_{j=1}^{n}\rho_1(Z_j,\hat{\beta}) = 0$ and $\sum_{j=1}^{n}\rho_2(Z_j,\hat{K}) = 0$; i.e.,

$$\hat{\theta} = n^{-1}\sum_{j=1}^{n}\frac{Z_j\mathbb{1}(Z_j \overset{\text{elt}}{\neq} c)}{a(Z_j,\hat{K})} \quad \text{and} \quad \hat{K} = \frac{\sum_{j=1}^{n}\mathbb{1}(Z_j \overset{\text{elt}}{\neq} c)\mathbb{1}(Z_j \overset{\text{elt}}{<} c)^{\sim}}{\sum_{j=1}^{n}\mathbb{1}(Z_j \overset{\text{elt}}{<} c)^{\sim}}. \tag{4.10}$$

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

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

In light of (4.8), 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. □

Example 4.2 (Example 2.2 contd.)**.** Here $\rho_1(Z,\beta) = X(Y - X'\theta)\mathbb{1}(Z \overset{\text{elt}}{\neq} c)/a(Z,K)$. Hence, $\hat{\theta} = (\sum_{j=1}^{n}\hat{X}_jX_j')^{-1}(\sum_{j=1}^{n}\hat{X}_jY_j)$, where $\hat{X}_j := X_j\mathbb{1}(Z_j \overset{\text{elt}}{\neq} c)/a(Z_j,\hat{K})$ and $\hat{K}$ is given in (4.10); i.e., $\hat{\theta}$ is the IV estimator with instruments $\hat{X}$. If censoring is purely endogenous or purely exogenous, then $a(Z,K) = K + (1-K)\mathbb{1}(Y_j < c^{(1)})$ or $a(Z,K) = K + (1-K)\mathbb{1}(X_j \overset{\text{elt}}{<} c^{-(1)})$, respectively, and the expression for $\hat{\theta}$ simplifies accordingly. □

Example 4.3 (Endogenous censored regression)**.** Let $Y^* = X^{*\prime}\theta^* + \varepsilon^*$ such that some or all regressors are correlated with ε^*. Let W^* be the vector of instruments, i.e., $\mathbb{E}_{f^*}\{W^*\varepsilon^*\} = 0$. Hence, $g(Z^*,\theta^*) = W^*(Y^* - X^{*\prime}\theta^*)$ and $\rho_1(Z,\beta) = W(Y - X'\theta)\mathbb{1}(Z \overset{\text{elt}}{\neq} c)/a(Z,K)$. Endogenous tobit, where X^* is endogenous and only Y^* is censored, is important for applications and follows by letting $\rho_1(Z,\beta) = W(Y - X'\theta)\mathbb{1}(Y \neq c^{(1)})/a(Y,K)$, where $a(Y,K) = K + (1-K)\mathbb{1}(Y < c^{(1)})$. The asymptotic distribution of $\hat{\theta}$ follows readily from Theorem 4.2. □

Example 4.4 (Simultaneous equations)**.** Let $Y_1^* = X_1^{*\prime}\theta_1^* + \varepsilon_1^*$ and $Y_2^* = X_2^{*\prime}\theta_2^* + \varepsilon_2^*$, where ε_1^* and ε_2^* are mean independent of X^*, the vector of instruments. Hence, $\mathbb{E}_{f^*}\{A(X^*)\left[\begin{smallmatrix} Y_1^* - X_1^{*\prime}\theta_1^* \\ Y_2^* - X_2^{*\prime}\theta_2^* \end{smallmatrix}\right]\} = 0$, where $A(X^*)$ is a matrix of instrumental variables and (4.7) can be used to estimate θ_1^* and θ_2^*. Although this model has been studied earlier, see, e.g., Blundell and Smith, our treatment is more general because we do not assume that ε_1^* and ε_2^* are Gaussian and allow for the possibility that other variables besides Y_1^* and Y_2^* may also be censored. Censoring of $Y^* := (Y_1^*, Y_2^*)$ alone implies that $\rho_1(Z,\beta) = A(X)\left[\begin{smallmatrix} Y_1 - X_1'\theta_1 \\ Y_2 - X_2'\theta_2 \end{smallmatrix}\right]\mathbb{1}(Y_1 \neq c^{(1)}, Y_2 \neq c^{(2)})/a(Y,K)$, where $a(Y,K) = K + (1-K)\mathbb{1}(Y_1 < c^{(1)}, Y_2 < c^{(2)})$. □

Example 4.5 (Auxiliary information)**.** Sometimes we may possess 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 stacking $g(Z^*, \theta^*)$ and $m(Z^*)$. These types of models have been investigated by Imbens and Lancaster (1994), Hellerstein and Imbens (1999), and Nevo (2003). However, Imbens and Lancaster, as well as Hellerstein and Imbens, assume that Z^* is fully observed. Nevo 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 probabilities of censoring or truncation of Z^* to be fully unknown. □

We end this section with a brief remark about hypothesis and specification tests. 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 $h \le p$. As described in Newey and McFadden (1994, Theorem 9.2), a variety of statistics based on the moment function in (4.7) can be used to test this hypothesis. In each case, the test statistic is asymptotically χ^2_h under the null. Confidence regions can be obtained by inverting these test statistics. Next, assume that $q > p$. Since inference based on the estimated θ^* is sensible only if (2.1) is true, it is important to test it against the alternative that it is false. The standard approach is to use a criterion function based statistic usually called Hansen's J-test: GMM theory tells us that $n\hat{\rho}'(\hat{\beta})\breve{V}_\rho^{-1}\hat{\rho}(\hat{\beta}) \xrightarrow{d} \chi^2_{q-p}$ under the null hypothesis that (2.1) is true, where ρ is the moment function in (4.7). Therefore, a test for overidentifying restrictions in (2.1) can be based on this result.

5. Inference with truncated data

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, the refreshment sample identifies it via the moment condition

$$b^* = \mathbb{E}_{f_e}\{\mathbb{1}(Z \in T)|R = 1\} \iff \mathbb{E}_{f_e}\{[\mathbb{1}(Z \in T) - b^*]R\} = 0. \tag{5.1}$$

Next, (2.4) and (3.2) imply that $f^*(z) = \int_{r\in\{0,1\}} f_e(z,r)\, d\kappa / a(z, b^*, K_0)$, where $a(z, b^*, K_0) := K_0 + (1 - K_0)\mathbb{1}(z \in T)/b^*$ and $\int_{r\in\{0,1\}} f_e(z,r)\, d\kappa$ is the marginal density of Z in the enriched sample. Hence, we can rewrite (2.1) in terms of the enriched density as

$$\mathbb{E}_{f_e}\{g(Z, \theta^*)/a(Z, b^*, K_0)\} = 0. \tag{5.2}$$

Finally, as before,

$$\mathbb{E}_{f_e}\{R - K_0\} = 0. \tag{5.3}$$

To estimate $\beta^* := (\theta^*, b^*, K_0)_{(p+2)\times 1}$, let[12]

$$\rho(Z,R,\beta) := \begin{bmatrix} g(Z,\theta)/a(Z,b,K) \\ [\mathbb{1}(Z\in T)-b]R \\ R-K \end{bmatrix} := \begin{bmatrix} \rho_1(Z,\beta) \\ \rho_2(Z,R,b) \\ \rho_3(R,K) \end{bmatrix}_{(q+2)\times 1}. \tag{5.4}$$

Since (5.1) and (5.3) just identify b^* and K_0, by (5.2) it follows that $\mathbb{E}_{f^*}\{g(Z^*,\theta^*)\} = 0$ if and only if $\mathbb{E}_{f_e}\{\rho(Z,R,\beta^*)\} = 0$. Hence, β^* can be efficiently estimated by using the latter moment condition. Using notation introduced in Section 4, the GMM estimator is given by $\hat\beta := \mathrm{argmin}_{\beta\in\mathcal{B}}\hat\rho'(\beta)\breve V_\rho^{-1}\hat\rho(\beta)$ with ρ defined in (5.4) and $\mathcal{B} := \Theta\times[0,1]\times[0,1]$. Since $\Pr_{f_e}(Z\in T) = K_0b^* + 1 - K_0$,

$$\mathbb{E}_{f_e}\{\rho_1(Z,\beta^*)\} = b^*\mathbb{E}_{f_e}\{g(Z,\theta^*)|Z\in T\} + (1-b^*)\mathbb{E}_{f_e}\{g(Z,\theta^*)|Z\notin T\}; \tag{5.5}$$

i.e., the transformed moment function combines best predictors of $g(Z^*,\theta^*)|(Z^*$ is not truncated) and $g(Z^*,\theta^*)|(Z^*$ is truncated) weighted by probabilities of the corresponding events. As in the case of censoring, this procedure is automatically carried out in the enriched sample to efficiently estimate the parameters of interest; see Example 5.1 for a nice illustration.

Let $\alpha^* := K_0b^* + 1 - K_0$ and $v := \varepsilon + (\alpha^*/b^*)\mathrm{Proj}\{\rho_1(Z,\beta^*)\big|1,\rho_2(Z,R,b^*)\}$, where $\varepsilon := \rho_1(Z,\beta^*) - \mathrm{Proj}\{\rho_1(Z,\beta^*)\big|1,\rho_2(Z,R,b^*),\rho_3(R,K_0)\}$. Analogous to the notation in Section 4, define $\Omega := \mathbb{E}_{f_e}\{\varepsilon\varepsilon'\}$, $D := \mathbb{E}_{f_e}\{\partial\rho_1(Z,\beta^*)/\partial\theta\}$, $V_2 := \mathbb{E}_{f_e}\{\rho_2^2(Z,R,b^*)\}$, and $\Sigma_{12} := \mathbb{E}_{f_e}\{\rho_1(Z,\beta^*)\rho_2(Z,R,b^*)\}$. Letting $V := \mathbb{E}_{f_e}\{vv'\}$ and $M_V := V^{-1} - V^{-1}D(D'V^{-1}D)^{-1}D'V^{-1}$, we can then show the following result.

Theorem 5.1. *Let Assumption* 4.1 *hold with the moment function* $\rho(Z,R,\beta)$ *defined in* (5.4). *Then,* $n^{1/2}(\hat\theta-\theta^*)$, $n^{1/2}(\hat b-b^*)$, *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*

$$\begin{bmatrix} (D'V^{-1}D)^{-1} & -(\alpha^*/K_0b^*)(D'V^{-1}D)^{-1}D'V^{-1}\Sigma_{12} & 0_{p\times 1} \\ -(\alpha^*/K_0b^*)\Sigma_{12}'V^{-1}D(D'V^{-1}D)^{-1} & V_2/K_0^2 - (\alpha^*/K_0b^*)^2\Sigma_{12}'M_V\Sigma_{12} & 0 \\ 0_{p\times 1}' & 0 & K_0(1-K_0) \end{bmatrix}.$$

Since $\Sigma_{23} := \mathbb{E}_{f_e}\{\rho_2(Z,R,b^*)\rho_3(R,K_0)\} = 0$ and ε is the residual from an orthogonal projection, we have $\Omega = V_1 - \Sigma_{12}\Sigma_{12}'/V_2 - \Sigma_{13}\Sigma_{13}'/V_3$ and $V = \Omega + (\alpha^*/b^*)^2\Sigma_{12}\Sigma_{12}'/V_2$,[13] where $V_1 := \mathbb{E}_{f_e}\{\rho_1(Z,\beta^*)\rho_1'(Z,\beta)\}$, $\Sigma_{13} := \mathbb{E}_{f_e}\{\rho_1(Z,\beta)\rho_3(R,K_0)\}$, and $V_3 := \mathbb{E}_{f_e}\{\rho_3^2(R,K_0)\}$. In Theorem B.1 of Appendix B we show that $(D'V^{-1}D)^{-1}$ and $V_2/K_0^2 - (\alpha^*/K_0b^*)^2\Sigma_{12}'M_V\Sigma_{12}$ coincide with the efficiency bounds for estimating θ^* and b^*, respectively. Therefore, $\hat\theta$ and $\hat b$ are asymptotically efficient. Since $\hat b$ is obtained by using the refreshment sample alone (as is

[12]Using the fact that $\mathbb{1}(Z\notin T)(1-R) = 0$, it is easy to show that $\rho_3(R,K_0)(1-b^*) - \rho_2(Z,R,b^*) = [K_0\alpha^*/(1-K_0)]\{1/a(Z,b^*,K_0)-1\}$. Thus $1/a(Z,b^*,K_0)-1$ can be written as a linear combination of $\rho_2(Z,R,b^*)$ and $\rho_3(R,K_0)$. Hence, the moment condition $\mathbb{E}_{f_e}\{1/a(Z,b^*,K_0)\} = 1$ is automatically satisfied.

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

$\hat{\theta}_R$), its asymptotic variance when $q = p$ is given by $b^*(1-b^*)/K_0$ because $V_2 = K_0 b^*(1-b^*)$. Hence, overidentification of θ^* leads to a better estimator of b^*.

We can use the proof of Theorem 5.1 to show that $\hat{\theta}$ is asymptotically linear with influence function $-(D'V^{-1}D)^{-1}D'V^{-1}v$. But since v is orthogonal to $\rho_3(R, K_0)$,[14] an application of the Cramér-Wold device and the central limit theorem reveals that $\hat{\theta}$ and $\sum_{j=1}^n R_j$ are asymptotically independent. Therefore, as for censoring, inference using the asymptotic distribution of $\hat{\theta}$ is the same as inference based on the asymptotic distribution of $\hat{\theta}$ given $\sum_{j=1}^n R_j$.

The next result shows that $\hat{\theta}$ 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 also possess information that can be exploited to increase efficiency.

Theorem 5.2. *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* (4.9)*. Then,* $\operatorname{asvar}(\hat{\theta}_R) > \operatorname{asvar}(\hat{\theta})$*.*

As in Section 4, parametric hypotheses can be tested by basing the test on the moment function defined in (5.4). Similarly, if $q > p$ then a test for overidentifying moment restrictions can be done using the J-statistic as described earlier.

Example 5.1 (Example 2.3 contd.)**.** Here $\rho_1(Z, \beta) = (Z - \theta)/a(Z, b, K)$ and no overidentifying restrictions. Thus $\hat{\beta}$ solves $\sum_{j=1}^n \rho(Z_j, \hat{\beta}) = 0$. Hence, $\hat{b} = \sum_{j=1}^n \mathbb{1}(Z_j \in T)R_j / \sum_{j=1}^n R_j$ 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 $\mathbb{1}(Z_j \notin T)(1 - R_j) = 0$, which follows by the definition of R_j, a little algebra shows that we can express $\hat{\theta}$ more revealingly as

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

which is exactly what we would expect from (5.5). □

Example 5.2 (Example 2.4 contd.)**.** Let $\hat{X}_j := X_j/a(Z_j, \hat{b}, \hat{K})$ with $\hat{b}$ and $\hat{K}$ as in Example 5.1 and $a(Z, b, K) = K + (1-K)\mathbb{1}(Y \in T_1, X \in T_2)/b$. Then $\hat{\theta} = \{\sum_{j=1}^n \hat{X}_j X_j'\}^{-1}\{\sum_{j=1}^n \hat{X}_j Y_j\}$. For pure endogenous or exogenous truncation, $a(Z_j, \hat{b}, \hat{K})$ is either $\hat{K} + (1-\hat{K})\mathbb{1}(Y_j \in T_1)/\hat{b}$ or $\hat{K} + (1-\hat{K})\mathbb{1}(X_j \in T_2)/\hat{b}$, respectively, and $\hat{\theta}$ simplifies accordingly. By Theorem 5.1, $n^{1/2}(\hat{\theta} - \theta^*)$ is asymptotically normal with mean zero and variance $D^{-1}VD^{-1\prime}$. Truncated versions of the endogenous regression and simultaneous equations models in Examples 4.3 and 4.4 can also be estimated using our approach. □

[14] By its definition, ε is orthogonal to $\rho_3(R, K_0)$. Moreover, $\rho_2(Z, R, b^*)$ and $\rho_3(R, K_0)$ are also orthogonal. Therefore, since $v = \varepsilon + (\alpha^*/b^*)\text{Proj}\{\rho_1(Z, \beta^*)\big|1, \rho_2(Z, R, b^*)\}$, it follows that v is orthogonal to $\rho_3(R, K_0)$.

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, it is also a topic of some substantive importance.

There is a large recent literature on the impacts of changes in compulsory schooling laws both in the United States and abroad and this is a topic of much interest to economists and policy makers. Researchers have studied the effects of these laws on educational attainment, probability of teenage childbearing, and a host of adult outcomes including earnings, wealth, happiness, health, and mortality risk; see, e.g., Oreopoulos (2007), Black, Devereux, and Salvanes (2007), and the references therein. Our application to the effects of compulsory schooling laws on age at first marriage adds to this literature.

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 2/3rd 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 it has been linked to greater opportunities for young people, especially women, to obtain education and develop a professional career.

Theoretically, the effects of increased years of compulsory schooling 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 compulsory schooling 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 compulsory schooling will tend to delay marriage through this channel. Our goal is to study whether compulsory schooling legislation encourages people to defer marriage. If so, these factors should be considered when evaluating the benefits of this type of legislation.

Our data are 1% samples from the Public Use Files of the U.S. Census of population for the years 1960, 1970, and 1980 (Ruggles et al., 2004), and 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. We use variation in compulsory schooling laws across states and over time.

The empirical model can be written as

$$\log(Y_j^*) = X_j^{*\prime}\theta^* + \varepsilon_j^*, \tag{6.1}$$

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 three included compulsory schooling law dummy variables describing the level of compulsory schooling: one for 9 years of schooling, another for 10 years, and the third for 11 or 12 years of schooling; the omitted category is 8 years or less of schooling. We measure age and age at first marriage in units of a quarter of a year. Additional details about the compulsory schooling variables and age variables are provided in Appendix D. Note that (6.1) 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; e.g., states with strict compulsory schooling laws may be states where people tend to marry late in any case.

Note that our research design is similar to other papers in the literature. It is standard when studying the effects of U.S. compulsory schooling laws to include control variables for race and a full set of cohort and state dummies. The assumption is that conditional on cohort and state, the level of compulsory schooling is exogenous. The rationale is that changes in these laws tend to occur for reasons unrelated to the marriage ages of people who happen to be in the particular cohorts impacted in the particular state. Because we use the same specification as other studies, our estimates are comparable to the literature and fit naturally within it.

The major problem in running this regression is that Y^* is censored for younger individuals because census records report age at first marriage for only those individuals who married before the census interview took place; otherwise, they simply report the individuals chronological age at the time of interview. Hence, for each person we can only observe

$$Y_j := \begin{cases} Y_j^* & \text{if } Y_j^* < C_j \\ C_j & \text{otherwise,} \end{cases} \tag{6.2}$$

where C_j denotes chronological age at the time of interview.

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.[15] 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 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 that we are tracking a population as they age. On the other hand, the average values of age at first marriage differ greatly by census due to censoring.

To statistically corroborate that we are following samples from the same population, we also carry out Hausman tests described in Section 4 to take advantage of the fact that GMM with the enriched sample is more efficient than OLS on the refreshment sample alone. We perform two variants of the Hausman test. The first, labeled "Hausman statistic" in Tables 4 and 5, restricts the test to the three compulsory schooling dummies that are of primary interest in this application; the second, not reported here, tests the equality of all coefficients including the compulsory schooling dummies, race dummy, cohort effects, and state dummies. In all specifications, we pass the first test and fail the second. We are not surprised that we fail the

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

second test because we have over 400,000 observations in the combined samples. With sample sizes this large, even tiny differences in coefficients can be statistically significant. We find the fact that we pass the test for our coefficients of interest to be very reassuring. For visual support, we also plot estimates of the cdf of Y^* using the refreshment and enriched samples; i.e., since $\Pr_{f^*}(Y^* \leq t) = \mathbb{E}_{f_e}\{\mathbb{1}(Y \leq t)\mathbb{1}(Y \neq c)/a(Y, K_0)\}$ for all $t \in \mathbb{R}$ by (4.1), which itself is based on the assumption that the master and refreshment samples are from the same population, we plot sample analogs of the left and right hand sides using the refreshment and enriched samples, respectively, over a grid in $\mathbb{R}$. As can be seen from Figure 1, estimates of $\Pr_{f^*}(Y^* \leq t)$ from the refreshment sample alone are virtually identical to those from the enriched sample over a range of values for t. Therefore, the Hausman test and these plots provide strong statistical and visual evidence that the observations in our master and refreshment 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; the results are displayed in Table 5). We find that our GMM estimates are reasonably robust to the imputation method used and so in Table 4 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 (note that since θ^* here is just identified, its GMM and empirical likelihood estimators are identical): GMM60, obtained by combining 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 combined. Estimates for men and women are reported separately. Following the procedure described in Section 4, 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 expect GMM60 and GMM70 to give similar estimates in finite samples even though censoring is less of a problem in 1970. This is a good check of robustness and 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 combining, say, the 1960 and 1980 samples, we first create a new variable $\tilde{C}_j = C_j \mathbb{1}(j \in 1960) + (C_j - 20)\mathbb{1}(j \in 1980)$ 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 similarly by combining 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 (quarters); 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 1.3%, 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. The OLS estimates from 1960 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%. 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 predominantly suggest significant effects of 10 years of required schooling (9 years is statistically 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%. The OLS80 and tobit estimates are again very different, suggesting that censoring bias is severe for the tobit estimates.

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. By 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. Simulation

In this section we describe the results of a small experiment to study the finite sample properties of $\hat{\theta}$, our estimator for the target population mean $\theta^* := \mathbb{E}_{f^*}\{Y^*\}$ when Y^* is censored from the right by c.

Letting $n := n_M + n_R$, where n_M and n_R are the master and refreshment sample sizes, we first generated $Y_i^* := \theta^* + \varepsilon_i^*$, $i = 1, \ldots, n$, where each ε_i^* is an equiprobable mixture of

$\mathrm{N}(\mu_1, \sigma_1^2)$ and $\mathrm{N}(\mu_2, \sigma_2^2)$ random variables. Next, we created the master sample

$$Y_i := \begin{cases} Y_i^* & \text{if } Y_i^* < c \\ c & \text{otherwise,} \end{cases} \qquad\qquad i = 1, \ldots, n_M,$$

the refreshment sample $(Y^*_{n_M+1}, \ldots, Y^*_n)$, and the enriched sample $(Y_1, \ldots, Y_{n_M}, Y^*_{n_M+1}, \ldots, Y^*_n)$.

Data was generated by letting $(\mu_1, \sigma_1) = (-2, 1)$ and $(\mu_2, \sigma_2) = (2, 1)$ so that Y^* has a bimodal hence, non-normal, distribution centered at θ^*; see Figure 4. Therefore, under this specification, tobit is inconsistent for θ^*. We set $\theta^* = 0$, which simplifies presentation by making the point estimate equal to the bias, and consider two specifications for c; one where c is fixed and another where $c_i \overset{d}{\sim} \mathrm{N}(\mu_c, 1)$ for each i. By letting c (resp. μ_c) take the value -2 or 2, we obtain master samples that are 75% or 25% censored. Finally, the number of observations in the master sample $n_M \in \{100, 500\}$ and n_R/n_M is either 20% or 80%, leading to the sixteen combinations of censoring probabilities, total sample sizes, and n_R/m_M ratios, in Tables 6–9.[16]

Simulation results averaged across 10,000 replications are reported in these tables for the following estimators: $\bar{Y}$, the mean of the master sample; "TOBIT", the tobit estimator of θ^* using the master sample; $\hat{\theta}_R$, the mean of the refreshment sample; and $\hat{\theta}$, the estimator of θ^* described in Example 4.1. Keep in mind that $\bar{Y}$ and TOBIT are inconsistent, whereas $\hat{\theta}_R$ is consistent but inefficient. Only $\hat{\theta}$ is both consistent and efficient. To make efficiency comparisons easier to read, the last column in Tables 6–9 treats the mean squared error (MSE) of $\hat{\theta}$ as the numeraire.

For the case where c is fixed, Tables 6 and 8 clearly show that $\bar{Y}$ and TOBIT are inconsistent since their point estimates basically stay the same as n increases. By contrast, $\hat{\theta}_R$ and $\hat{\theta}$ are very close to the truth although $\hat{\theta}$ far outperforms $\hat{\theta}_R$ in terms of MSE even when the refreshment sample is 80% of the master sample; note that these results hold whether censoring is high (75%) or low (25%). Tables 7 and 9 indicate that essentially the same is true even when the censoring threshold is random. In short, the performance of $\hat{\theta}$ appears to be remarkably robust to various combinations of censoring probabilities and relative size of the refreshment and master samples. Therefore, even for the simple design considered in this experiment, these simulation results demonstrate very effectively the power of optimally combining the master and refreshment samples and thus lend additional support to the empirical results obtained earlier in Section 6.

[16]Simulations were done in MATLAB. Code for the application and simulations is available on request from the authors.

8. Conclusion

The methods developed in this paper are relevant in many other applied contexts.[17] 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.

Acknowledgements

We thank an associate editor and two anonymous referees for comments that greatly improved this paper. We also thank participants at several seminars, including those at the 9th world congress of the econometric society, for helpful suggestions and conversations. Financial support from the NSF (Devereux) and the University of Connecticut graduate school (Tripathi) is gratefully acknowledged.

Appendix A. Proofs of the results in section 4

Proof of Theorem 4.1. From standard GMM theory we know that $n^{1/2}(\tilde{\beta}-\beta^*)$ is asymptotically normal with mean zero and variance $(D_\rho' V_\rho^{-1} D_\rho)^{-1}$, where $D_\rho := \mathbb{E}_{f_e}\{\partial\rho(Z,R,\beta^*)/\partial\beta\}$ and $V_\rho := \mathbb{E}_{f_e}\{\rho(Z,R,\beta^*)\rho'(Z,R,\beta^*)\}$. Letting $\Sigma := [\Sigma_{12}\ \Sigma_{13}]$ and $\Sigma_{23} := \mathbb{E}_{f_e}\{\rho_2(Z,K_0)\rho_3(R,K_0)\}$, we can write $V_\rho = \begin{bmatrix} V_1 & \Sigma \\ \Sigma' & V_{-1}\end{bmatrix}$, where $V_{-1} := \begin{bmatrix} V_2 & \Sigma_{23} \\ \Sigma_{23} & V_3\end{bmatrix}$. Hence, by the partitioned inverse formula,

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

[17]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 combining 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 $\rho_1(Z,\beta^*)$ 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

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

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

$$D'_\rho V_\rho^{-1} D_\rho = \begin{bmatrix} D'\Omega^{-1}D & 0_{p\times 1} \\ 0'_{p\times 1} & 1/K_0(1-K_0) \end{bmatrix}. \tag{A.4}$$

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 the moment function $\rho(Z,\beta)$ defined in (4.7), we now have

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

Therefore,

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

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,

$$D_R = K_0\mathbb{E}_{f_e}\{\partial g(Z,\theta^*)/\partial\theta | R=1\} \overset{(3.1)}{=} K_0\mathbb{E}_{f^*}\{\partial g(Z^*,\theta^*)/\partial\theta\} = K_0 D_*.$$

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 $\text{asvar}(\hat\theta_R) > \text{asvar}(\hat\theta)$ it suffices to show that $V_*/K_0 > \Omega$; i.e., $V_*/K_0 - \Omega$ is positive definite. So, by (4.1), $\mu^*(\{c\}) = 0$, and the fact

$$\{\mathbb{1}(Z \overset{\text{elt}}{<} c) + \mathbb{1}(Z \overset{\text{elt}}{<} c)^\sim\}/a(Z,K_0) = \mathbb{1}(Z \overset{\text{elt}}{<} c) + \mathbb{1}(Z \overset{\text{elt}}{<} c)^\sim/K_0, \tag{A.5}$$

we can write V_1 as

$$V_1 = \mathbb{E}_{f^*}\{g(Z^*,\theta^*)g'(Z^*,\theta^*)\mathbb{1}(Z^* \overset{\text{elt}}{<} c)\} + \mathbb{E}_{f^*}\{g(Z^*,\theta^*)g'(Z^*,\theta^*)\mathbb{1}(Z^* \overset{\text{elt}}{<} c)^\sim\}/K_0. \tag{A.6}$$

Hence, we have that

$$\Omega = V_1 - \Sigma_{12}\Sigma_{12}'/V_2$$
$$= V_*/K_0 - [(1/K_0 - 1)\mathbb{E}_{f^*}\{g(Z^*,\theta^*)g'(Z^*,\theta^*)\mathbb{1}(Z^* \overset{\text{elt}}{<} c)\} + \Sigma_{12}\Sigma_{12}'/V_2]. \quad \text{(A.7)}$$

Therefore, $V_*/K_0 > \Omega$ since $K_0 \in (0,1)$. □

Remark A.1. For notational convenience, let $\Delta_1 := \text{var}_{f^*}\{g(Z^*,\theta^*)|(Z^* \overset{\text{elt}}{<} c)^\sim\}$ and

$$\Delta_2 := \mathbb{E}_{f^*}\{g(Z^*,\theta^*)g'(Z^*,\theta^*)\mathbb{1}(Z^* \overset{\text{elt}}{<} c)\}$$
$$+ \mathbb{E}_{f^*}\{g(Z^*,\theta^*)\mathbb{1}(Z^* \overset{\text{elt}}{<} c)^\sim\}\mathbb{E}_{f^*}\{g'(Z^*,\theta^*)\mathbb{1}(Z^* \overset{\text{elt}}{<} c)^\sim\}/(1 - F^*(c)).$$

Then, using (A.6), (A.10), Lemma A.2(ii), and Lemma A.3(ii), a little algebra shows that

$$\Omega = V_1 - \Sigma_{12}\Sigma_{12}'/V_2 = [(1 - F^*(c))/K_0]\Delta_1 + \Delta_2.$$

Therefore, since Δ_1 and Δ_2 do not depend upon K_0, it follows that Ω is a decreasing (in the positive definite sense) function of K_0. Furthermore, by (A.10) and Lemma A.2(ii), we can write (A.7) as

$$V_*/K_0 - \Omega = [(1 - K_0)/K_0]\Delta_2.$$

Since $K_0 \mapsto (1-K_0)/K_0$ is monotonically decreasing on $(0,1)$, the gap $V_*/K_0 - \Omega$ is also a decreasing function of K_0. □

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

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

$$\mathbb{E}_{f_e}\{[\rho_1(Z,\beta^*) - \text{Proj}\{\rho_1(Z,\beta^*)|1, \rho_2(Z,K_0)\}]\rho_3(R,K_0)\} = 0. \quad \text{(A.8)}$$

But $\text{Proj}\{\rho_1(Z,\beta^*)|1,\rho_2(Z,K_0)\} = \Sigma_{12}\rho_2(Z,K_0)/V_2$. Hence, by Lemma A.2, we have that (A.8) holds if and only if $\Sigma_{13} = \Sigma_{12}$. Now, by (3.1),

$$\Sigma_{13} = K_0\mathbb{E}_{f_e}\{\rho_1(Z,\beta^*)|R = 1\} = K_0\mathbb{E}_{f^*}\{\rho_1(Z^*,\beta^*)\}.$$

Moreover, since $\mu^*(\{c\}) = 0$,

$$\mathbb{E}_{f^*}\{\rho_1(Z^*,\beta^*)\} = \mathbb{E}_{f^*}\{g(Z^*,\theta^*)[\mathbb{1}(Z^* \overset{\text{elt}}{<} c) + \mathbb{1}(Z^* \overset{\text{elt}}{<} c)^\sim]/a(Z^*,K_0)\}.$$

Hence, using (A.5), we obtain that

$$\Sigma_{13} = -(1 - K_0)\mathbb{E}_{f^*}\{g(Z^*,\theta^*)\mathbb{1}(Z^* \overset{\text{elt}}{<} c)\}. \quad \text{(A.9)}$$

Next, observe that

$$\Sigma_{12} = (1 - K_0)\mathbb{E}_{f_e}\{g(Z,\theta^*)\mathbb{1}(Z \overset{\text{elt}}{\neq} c)\mathbb{1}(Z \overset{\text{elt}}{<} c)^\sim/a(Z,K_0)\}$$
$$= (1 - K_0)\mathbb{E}_{f^*}\{g(Z^*,\theta^*)\mathbb{1}(Z^* \overset{\text{elt}}{<} c)^\sim\} \quad \text{(A.10)}$$
$$= -(1 - K_0)\mathbb{E}_{f^*}\{g(Z^*,\theta^*)\mathbb{1}(Z^* \overset{\text{elt}}{<} c)\}, \quad \text{(A.11)}$$

where the second equality follows by (4.1) and the assumption that $\mu^*(\{c\}) = 0$. Therefore, the desired result follows by (A.9) and (A.11). $\square$

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. Note that

$$\Sigma_{23} = \mathbb{E}_{f_e}\{\rho_2(Z,K_0)R\} = K_0\mathbb{E}_{f_e}\{\rho_2(Z,K_0)|R=1\} \overset{(3.1)}{=} K_0\mathbb{E}_{f^*}\{\rho_2(Z^*,K_0)\}.$$

Hence, (i) follows since

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

To show (ii), observe that

$$\mathbb{E}_{f_e}\{\rho_2^2(Z,K_0)\} = \mathbb{E}_{f_e}\{\mathbb{1}(Z \overset{\text{elt}}{\neq} c)\mathbb{1}(Z \overset{\text{elt}}{<} c)^\sim\} + K_0^2\mathbb{E}_{f_e}\{\mathbb{1}(Z \overset{\text{elt}}{<} c)^\sim\} - 2K_0\mathbb{E}_{f_e}\{\mathbb{1}(Z \overset{\text{elt}}{\neq} c)\mathbb{1}(Z \overset{\text{elt}}{<} c)^\sim\}.$$

But using (4.1) and the assumption that $\mu^*(\{c\}) = 0$, it is easy to show that

$$\mathbb{E}_{f_e}\{\mathbb{1}(Z \overset{\text{elt}}{\neq} c)\mathbb{1}(Z \overset{\text{elt}}{<} c)^\sim\} = K_0[1-F^*(c)].$$

Therefore, (ii) follows by Lemma A.3(ii). $\square$

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

Proof of Lemma A.3. First, note that

$$\begin{aligned}\partial\rho_1(Z,\beta^*)/\partial K &= -g(Z,\theta^*)\mathbb{1}(Z \overset{\text{elt}}{\neq} c)\mathbb{1}(Z \overset{\text{elt}}{<} c)^\sim/a^2(Z,K_0)\\ &= -g(Z,\theta^*)\mathbb{1}(Z \overset{\text{elt}}{\neq} c)\mathbb{1}(Z \overset{\text{elt}}{<} c)^\sim/[a(Z,K_0)K_0],\end{aligned}$$

where the second equality is due to (A.5). Hence, by (4.1) and $\mu^*(\{c\}) = 0$,

$$\mathbb{E}_{f_e}\{\partial\rho_1(Z,\beta^*)/\partial K\} = \mathbb{E}_{f^*}\{g(Z^*,\theta^*)\mathbb{1}(Z^* \overset{\text{elt}}{<} c)\}/K_0.$$

Therefore, (i) follows by (A.11). Next, since $\mathbb{1}(Z \overset{\text{elt}}{<} c) = \mathbb{1}(Z \overset{\text{elt}}{\neq} c)\mathbb{1}(Z \overset{\text{elt}}{<} c)$,

$$\begin{aligned}\mathbb{E}_{f_e}\{\mathbb{1}(Z \overset{\text{elt}}{<} c)^\sim\} &= 1-\mathbb{E}_{f_e}\{\mathbb{1}(Z \overset{\text{elt}}{<} c)\} = 1-\mathbb{E}_{f_e}\{\mathbb{1}(Z \overset{\text{elt}}{\neq} c)\mathbb{1}(Z \overset{\text{elt}}{<} c)\}\\ &\overset{(4.1)}{=} 1-\mathbb{E}_{f^*}\{\mathbb{1}(Z^* \overset{\text{elt}}{\neq} c)\mathbb{1}(Z^* \overset{\text{elt}}{<} c)a(Z^*,K_0)\} = 1-F^*(c)\end{aligned}$$

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

A.1. **Efficiency bounds under censoring.** We use the methodology of Severini and Tripathi (2001) to calculate the efficiency bounds. Begin by writing the enriched density of Z and R as $f_e(z,r) = \phi_0^2(z,r)$. This ensures that ϕ_0 lies in $L_2(z,r)$, the set of real-valued functions on $\mathbb{R}^d \times \{0,1\}$ square-integrable with respect to $\mu \times \kappa$. Now, suppose that we want to calculate the efficiency bound for estimating $\eta(\phi_0)$, a pathwise differentiable functional of ϕ_0 (see Severini and Tripathi (2001) for technical definitions and details). We proceed as follows. Let $t \mapsto \phi_t$ be a curve from an interval containing zero into the unit ball of $L_2(z,r)$ such that $\phi_t|_{t=0} = \phi_0$. Since the observed loglikelihood for t in this submodel is $\log \phi_t^2(z,r)$, the Fisher information for a single observation is given by $i_F := 4\int_{\mathbb{R}^d\times\{0,1\}} \dot{\phi}^2(z,r)\,d\mu\,d\kappa$, where $\dot{\phi}$ denotes the tangent vector to ϕ_t at $t=0$; i.e., $\dot{\phi}$ is an element of the tangent space $\mathcal{T} := \{\dot{\phi} \in L_2(z,r) : \int_{\mathbb{R}^d\times\{0,1\}} \phi_0(z,r)\dot{\phi}(z,r)\,d\mu\,d\kappa = 0\}$. Note that i_F is induced by the Fisher inner-product $\langle \dot{\phi}_1, \dot{\phi}_2\rangle_F := 4\int_{\mathbb{R}^d\times\{0,1\}} \dot{\phi}_1(z,r)\dot{\phi}_2(z,r)\,d\mu\,d\kappa$. Thus $i_F = \|\dot{\phi}\|_F^2$, where $\|\cdot\|_F$ denotes the norm generated by the Fisher inner-product.

Since (2.1) is equivalent to $\mathbb{E}_{f_e}\{g(Z,\theta^*)\mathbb{1}(Z \overset{\text{elt}}{\neq} c)/a(Z,K_0)\} = 0$, we have to use the additional information in (4.2) when calculating the efficiency bound for estimating $\eta(\phi_0)$. So let $t \mapsto (\theta_t, K_t)$ denote a curve passing through (θ^*, K_0) at $t=0$ such that for all t in a neighborhood of zero

$$\int_{\mathbb{R}^d\times\{0,1\}} g(z,\theta_t)\mathbb{1}(z \overset{\text{elt}}{\neq} c)\phi_t^2(z,r)/a(z,K_t)\,d\mu\,d\kappa = 0, \tag{A.12}$$

where, by (4.3) and (4.4), K_t is defined via the moment conditions

$$\begin{aligned}\int_{\mathbb{R}^d\times\{0,1\}} (\mathbb{1}(z \overset{\text{elt}}{\neq} c)\mathbb{1}(z \overset{\text{elt}}{<} c)^{\sim} - K_t\mathbb{1}(z \overset{\text{elt}}{<} c)^{\sim})\phi_t^2(z,r)\,d\mu\,d\kappa = 0,\\ \int_{\mathbb{R}^d\times\{0,1\}} (r - K_t)\phi_t^2(z,r)\,d\mu\,d\kappa = 0.\end{aligned} \tag{A.13}$$

By (A.12), the tangent vectors $\dot{\phi}$, $\dot{\theta}$, and $\dot{K}$ must satisfy

$$D\dot{\theta} + 2\int_{\mathbb{R}^d\times\{0,1\}} \rho_1(z,\beta^*)\phi_0(z,r)\dot{\phi}(z,r)\,d\mu\,d\kappa + \mathbb{E}_{f_e}\{\partial\rho_1(Z,\beta^*)/\partial K\}\dot{K} = 0 \tag{A.14}$$

and from (A.13) we know that $\dot{K}$ solves

$$\begin{bmatrix} -[1-F^*(c)] \\ -1 \end{bmatrix}\dot{K} + 2\int_{\mathbb{R}^d\times\{0,1\}} \rho_{-1}(z,r,K_0)\phi_0(z,r)\dot{\phi}(z,r)\,d\mu\,d\kappa = 0, \tag{A.15}$$

where $\rho_{-1}(z,r,K_0) := (\rho_2(z,K_0), \rho_3(r,K_0))_{2\times 1}$. Therefore, stacking (A.14) and (A.15), we have that

$$D_\rho\dot{\beta} + 2\int_{\mathbb{R}^d\times\{0,1\}} \rho(z,r,\beta^*)\phi_0(z,r)\dot{\phi}(z,r)\,d\mu\,d\kappa = 0, \tag{A.16}$$

where D_ρ is given by (A.3) and $\dot{\beta} := (\dot{\theta}, \dot{K})_{(p+1)\times 1}$.

Now let W be a $(q+2)\times(q+2)$ symmetric positive-definite non-stochastic matrix. Premultiplying (A.16) by $(D_\rho' W D_\rho)^{-1}D_\rho' W$ and solving for $\dot{\beta}$, we obtain that

$$\dot{\beta} = -2(D_\rho' W D_\rho)^{-1}D_\rho' W\int_{\mathbb{R}^d\times\{0,1\}} \rho(z,r,\beta^*)\phi_0(z,r)\dot{\phi}(z,r)\,d\mu\,d\kappa. \tag{A.17}$$

Finally, substituting (A.17) in (A.16), we get that

$$(I_{(q+2)\times(q+2)} - D_\rho(D_\rho' W D_\rho)^{-1} D_\rho' W) \int_{\mathbb{R}^d\times\{0,1\}} \rho(z,r,\beta^*)\phi_0(z,r)\dot{\phi}(z,r)\,d\mu\,d\kappa = 0. \tag{A.18}$$

Since $x \mapsto D_\rho(D_\rho' W D_\rho)^{-1} D_\rho' W x$ is an orthogonal projection onto the column space of D_ρ using the weighted inner product $\langle x_1, x_2\rangle_W := x_1' W x_2$, it follows that (A.18) is satisfied by only those tangent vectors $\dot{\phi}$ for which $\int_{\mathbb{R}^d\times\{0,1\}} \rho(z,r,\beta^*)\phi_0(z,r)\dot{\phi}(z,r)\,d\mu\,d\kappa$ lies in the column space of D_ρ.

Let $\mathcal{T}_W$ denote the set of tangent vectors that satisfy (A.18). The efficiency bound for estimating $\eta(\phi_0)$ is given by $\sup_{W\in\mathcal{W}} \|\nabla\eta\|_W^2$, where $\mathcal{W}$ is the set of $(q+2)\times(q+2)$ symmetric positive-definite matrices, $\nabla\eta$ denotes the pathwise derivative of η, and $\|\nabla\eta\|_W := \sup_{\{\dot{\phi}\in\mathcal{T}_W:\dot{\phi}\neq 0\}} |\nabla\eta(\dot{\phi})|$ is the operator norm of $\nabla\eta$. To calculate the bound, we first employ a guess-and-verify strategy to find, for any $W \in \mathcal{W}$, a $\phi_W^* \in \mathcal{T}$ satisfying

$$\nabla\eta(\dot{\phi}) = \langle \dot{\phi}, \phi_W^*\rangle_F \quad \text{for all } \dot{\phi} \in \mathcal{T}_W. \tag{A.19}$$

Next, we pick a $W^* \in \mathcal{W}$ so that $\int_{\mathbb{R}^d\times\{0,1\}} \rho(z,r,\beta^*)\phi_0(z,r)\phi_{W^*}^*(z,r)\,d\mu\,d\kappa$ lies in the column space of D_ρ. This means that $\phi_{W^*}^* \in \mathcal{T}_{W^*}$ and we can use this fact to show that $\|\nabla\eta\|_{W^*} = \|\phi_{W^*}^*\|_F$.[18] Since W^* is determined uniquely up to scale, see, e.g., the proof of Theorem A.1, the efficiency bound for estimating $\eta(\phi_0)$ is therefore given by $\|\phi_{W^*}^*\|_F^2$.

We use this procedure in Theorem A.1 to obtain the efficiency bound for estimating an arbitrary linear combination of θ^* (so that the object of interest is a real valued functional). A comparison with the asymptotic variance in Theorem 4.2 then reveals that $\hat{\theta}$ is asymptotically efficient.

Theorem A.1. *The efficiency bound for estimating* θ^* *is given by* $(D'\Omega^{-1}D)^{-1}$.

Proof of Theorem A.1. Let $\xi \in \mathbb{R}^p$ be arbitrary. To obtain the efficiency bound for estimating $\eta(\phi_0) := \xi'\theta^*$, the tangent vectors $\dot{\phi}$ and $\dot{\theta}$ must satisfy $\nabla\eta(\dot{\phi}) = \zeta'\dot{\beta}$, where $\zeta := (\xi, 0)_{(p+1)\times 1}$. Hence, by (A.17), for any $W \in \mathcal{W}$ we have that

$$\nabla\eta(\dot{\phi}) = -2\zeta'(D_\rho' W D_\rho)^{-1} D_\rho' W \int_{\mathbb{R}^d\times\{0,1\}} \rho(z,r,\beta^*)\phi_0(z,r)\dot{\phi}(z,r)\,d\mu\,d\kappa.$$

By (A.19), we have to find a $\phi_W^* \in \mathcal{T}$ such that

$$\int_{\mathbb{R}^d\times\{0,1\}} \{\phi_W^*(z,r) + 0.5\zeta'(D_\rho' W D_\rho)^{-1} D_\rho' W \rho(z,r,\beta^*)\phi_0(z,r)\}\dot{\phi}(z,r)\,d\mu\,d\kappa = 0 \tag{A.20}$$

for all $\dot{\phi} \in \mathcal{T}_W$. We claim that

$$\phi_W^*(z,r) = -0.5\zeta'(D_\rho' W D_\rho)^{-1} D_\rho' W \rho(z,r,\beta^*)\phi_0(z,r).$$

It is easily verified that $\phi_W^* \in \mathcal{T}$ and satisfies (A.20) for all $\dot{\phi} \in \mathcal{T}_W$. Hence, we only have to determine W^* such that $\int_{\mathbb{R}^d\times\{0,1\}} \rho(z,r,\beta^*)\phi_0(z,r)\phi_{W^*}^*(z,r)\,d\mu\,d\kappa$ lies in the column space of D_ρ. But since

$$\int_{\mathbb{R}^d\times\{0,1\}} \rho(z,r,\beta^*)\phi_0(z,r)\phi_W^*(z,r)\,d\mu\,d\kappa = -0.5 V_\rho W D_\rho (D_\rho' W D_\rho)^{-1}\zeta,$$

[18] By (A.19), $\nabla\eta(\dot{\phi}) = \langle\dot{\phi}, \phi_{W^*}^*\rangle_F$ for all $\dot{\phi} \in \mathcal{T}_{W^*}$. Hence, $\|\nabla\eta\|_{W^*} \le \|\phi_{W^*}^*\|_F$ by Cauchy-Schwarz. But since $\phi_{W^*}^* \in \mathcal{T}_{W^*}$, we also have $\|\phi_{W^*}^*\|_F^2 = \nabla\eta(\phi_{W^*}^*) \le \|\nabla\eta\|_{W^*}\|\phi_{W^*}^*\|_F$; i.e., $\|\nabla\eta\|_{W^*} \ge \|\phi_{W^*}^*\|_F$.

it follows that $\int_{\mathbb{R}^d\times\{0,1\}} \rho(z,r,\beta^*)\phi_0(z,r)\phi^*_{W^*}(z,r)\,d\mu\,d\kappa$ lies in the column space of D_ρ if and only if $V_\rho W^* \propto I_{q\times q}$. Hence,

$$\phi^*_{W^*}(z,r) = -0.5\zeta'(D'_\rho V_\rho^{-1}D_\rho)^{-1}D'_\rho V_\rho^{-1}\rho(z,r,\beta^*)\phi_0(z,r)$$

and the efficiency bound for estimating $\xi'\theta^*$ is given by

$$4\int_{\mathbb{R}^d\times\{0,1\}}\{\phi^*_{W^*}(z,r)\}^2\,d\mu\,d\kappa = \zeta'(D'_\rho V_\rho^{-1}D_\rho)^{-1}\zeta \overset{(A.4)}{=} \xi'(D'\Omega^{-1}D)^{-1}\xi.$$

The desired result follows since ξ was arbitrary. □

Appendix B. Proofs of the results in section 5

The proofs below are very similar to those in Appendix A.

Proof of Theorem 5.1. As in the proof of Theorem 4.1, $n^{1/2}(\hat{\beta}-\beta^*)$ is asymptotically normal with mean zero and variance $(D'_\rho V_\rho^{-1}D_\rho)^{-1}$, where

$$D_\rho \overset{\text{Lemma B.1}}{=} \begin{bmatrix} D & (1-K_0)\Sigma_{12}/b^{*2} & \Sigma_{12}/K_0 b^* \\ 0'_{p\times 1} & -K_0 & 0 \\ 0'_{p\times 1} & 0 & -1 \end{bmatrix} \tag{B.1}$$

$$V_\rho^{-1} = \begin{bmatrix} \Omega^{-1} & -\Omega^{-1}\Sigma V_{-1}^{-1} \\ -V_{-1}^{-1}\Sigma'\Omega^{-1} & V_{-1}^{-1} + V_{-1}^{-1}\Sigma'\Omega^{-1}\Sigma V_{-1}^{-1} \end{bmatrix} \quad \text{and} \quad V_{-1} = \begin{bmatrix} V_2 & 0 \\ 0 & V_3 \end{bmatrix}.$$

Hence, letting $\gamma := (K_0^2/V_2) + (\alpha^* K_0/V_2 b^*)^2\Sigma'_{12}\Omega^{-1}\Sigma_{12}$,

$$D'_\rho V_\rho^{-1}D_\rho = \begin{bmatrix} D'\Omega^{-1}D & (\alpha^* K_0/V_2 b^*)D'\Omega^{-1}\Sigma_{12} & 0_{p\times 1} \\ (\alpha^* K_0/V_2 b^*)\Sigma'_{12}\Omega^{-1}D & \gamma & 0 \\ 0'_{p\times 1} & 0 & 1/V_3 \end{bmatrix}.$$

Now, since $V = \Omega + (\alpha^*/b^*)^2\Sigma_{12}\Sigma'_{12}/V_2$, by the Sherman-Morrison formula

$$\Omega^{-1} = [V - \frac{\Sigma_{12}\Sigma'_{12}}{V_2(b^*/\alpha^*)^2}]^{-1} = V^{-1} + \frac{V^{-1}\Sigma_{12}\Sigma'_{12}V^{-1}}{V_2(b^*/\alpha^*)^2 - \Sigma'_{12}V^{-1}\Sigma_{12}}.$$

Therefore, applying the partitioned inverse formula, it can be verified that $(D'_\rho V_\rho^{-1}D_\rho)^{-1}$ equals

$$\begin{bmatrix} (D'V^{-1}D)^{-1} & -(\alpha^*/K_0 b^*)(D'V^{-1}D)^{-1}D'V^{-1}\Sigma_{12} & 0_{p\times 1} \\ -(\alpha^*/K_0 b^*)\Sigma'_{12}V^{-1}D(D'V^{-1}D)^{-1} & V_2/K_0^2 - (\alpha^*/K_0 b^*)^2\Sigma'_{12}M_V\Sigma_{12} & 0 \\ 0'_{p\times 1} & 0 & K_0(1-K_0) \end{bmatrix}. \tag{B.2}$$

The desired result follows. □

Proof of Theorem 5.2. As in the proof of Theorem 4.3, we show that $V_*/K_0 > V$. Now, using Lemma B.2, a little algebra shows that $V = V_1 + (1-K_0)(\alpha^*/K_0 b^{*3})\Sigma_{12}\Sigma'_{12}$. But since

$$\begin{aligned} V_1 &= \mathbb{E}_{f_e}\{g(Z,\theta^*)g'(Z,\theta^*)[\mathbb{1}(Z\in T) + \mathbb{1}(Z\notin T)]/a^2(Z,b^*,K_0)\} \\ &= V_*/K_0 - (1-K_0)\mathbb{E}_{f^*}\{g(Z^*,\theta^*)g'(Z^*,\theta^*)\mathbb{1}(Z^*\in T)\}/K_0\alpha^*, \end{aligned}$$

we obtain that $V = V_*/K_0 - (1-K_0)\Delta/K_0$, where

$$\Delta := \mathbb{E}_{f^*}\{g(Z^*,\theta^*)g'(Z^*,\theta^*)\mathbb{1}(Z^* \in T)\}/\alpha^* - (\alpha^*/b^{*3})\Sigma_{12}\Sigma_{12}'.$$

Next, a little simplification reveals that $\Sigma_{12} = (b^*/\alpha^*)\mathbb{E}_{f^*}\{g(Z,\theta^*)\mathbb{1}(Z \in T)\}$. Hence,

$$\begin{aligned}\alpha^* b^* \Delta &= b^*\mathbb{E}_{f^*}\{g(Z^*,\theta^*)g'(Z^*,\theta^*)\mathbb{1}(Z^* \in T)\} - \mathbb{E}_{f^*}\{g(Z^*,\theta^*)\mathbb{1}(Z^* \in T)\}\mathbb{E}_{f^*}\{g'(Z^*,\theta^*)\mathbb{1}(Z^* \in T)\}\\ &= b^{*2}\mathrm{var}_{f^*}\{g(Z^*,\theta^*)|Z^* \in T\}.\end{aligned}$$

Therefore, $\Delta = (b^*/\alpha^*)\mathrm{var}_{f^*}\{g(Z^*,\theta^*)|Z^* \in T\}$ and we have

$$V = V_*/K_0 - [(1-K_0)b^*/K_0\alpha^*]\mathrm{var}_{f^*}\{g(Z^*,\theta^*)|Z^* \in T\}. \tag{B.3}$$

Hence, assuming $\mathrm{var}_{f^*}\{g(Z^*,\theta^*)|Z^* \in T\}$ is positive definite, the desired result follows. □

Remark B.1. From (B.3) we know that $V = [V_* - ((1-K_0)b^*/\alpha^*)\mathrm{var}_{f^*}\{g(Z^*,\theta^*)|Z^* \in T\}]/K_0$. But $K_0 \mapsto 1/K_0$ is a decreasing function of K_0 and, upon recalling the definition of α^*, it is easily seen that $K_0 \mapsto -(1-K_0)b^*/\alpha^*$ is increasing in K_0. Since V_* and $\mathrm{var}_{f^*}\{g(Z^*,\theta^*)|Z^* \in T\}$ do not depend upon K_0, it follows that V is the product of a decreasing function of K_0 with an increasing function of K_0. Hence, V may not be monotonically decreasing in K_0. However, since $K_0 \mapsto (1-K_0)b^*/K_0\alpha^*$ is a decreasing function of K_0, the gap $V_*/K_0 - V$ is monotonically decreasing in K_0. □

Lemma B.1. (i) $\mathbb{E}_{f_e}\{\partial\rho_1(Z,\beta^*)/\partial b\} = (1-K_0)\Sigma_{12}/b^{*2}$ *and* (ii) $\mathbb{E}_{f_e}\{\partial\rho_1(Z,\beta^*)/\partial K\} = \Sigma_{12}/K_0 b^*$.

Proof of Lemma B.1. First, note that

$$\mathbb{E}_{f_e}\{\partial\rho_1(Z,\beta^*)/\partial b\} = [(1-K_0)/b^{*2}]\mathbb{E}_{f_e}\{g(Z,\theta^*)\mathbb{1}(Z \in T)/a^2(Z,b^*,K_0)\}.$$

But $\mathbb{E}_{f_e}\{g(Z,\theta^*)\mathbb{1}(Z \in T)/a^2(Z,b^*,K_0)\}$ can be decomposed as

$$\mathbb{E}_{f_e}\{g(Z,\theta^*)[\mathbb{1}(Z \in T) - b^*]/a^2(Z,b^*,K_0)\} + b^*\mathbb{E}_{f_e}\{g(Z,\theta^*)/a^2(Z,b^*,K_0)\}.$$

Furthermore, it is easily seen that

$$\begin{aligned}\Sigma_{12} &= K_0\mathbb{E}_{f_e}\{g(Z,\theta^*)[\mathbb{1}(Z \in T) - b^*]/a^2(Z,b^*,K_0)\}\\ \Sigma_{13} &= K_0\mathbb{E}_{f_e}\{g(Z,\theta^*)/a^2(Z,b^*,K_0)\}.\end{aligned}$$

Therefore, (i) follows by Lemma B.2. The proof of (ii) is very similar and is omitted. □

Lemma B.2. $(1-K_0)\Sigma_{12} + b^*\Sigma_{13} = 0$.

Proof of Lemma B.2. Since $(1-K_0)\rho_2(Z,R,b^*)/b^* = \{a(Z,b^*,K_0) - 1\}R$ and $\mathbb{E}_{f_e}\{g(Z,\theta^*)R\} = 0$,

$$(1-K_0)\Sigma_{12}/b^* = \mathbb{E}_{f_e}\{g(Z,\theta^*)R\} - \mathbb{E}_{f_e}\{g(Z,\theta^*)R/a(Z,b^*,K_0)\} = -\Sigma_{13}.$$

The desired result follows. □

Theorem B.1. *The efficiency bounds for estimating* θ^* *and* b^* *are given by* $(D'V^{-1}D)^{-1}$ *and* $V_2/K_0^2 - (\alpha^*/K_0 b^*)^2\Sigma_{12}'M_V\Sigma_{12}$*, respectively.*

Proof of Theorem B.1. Following the procedure described earlier in Section A.1, we can show that analogous versions of (A.17) and (A.18) hold with $\dot{\beta} := (\dot{\theta}, \dot{b}, \dot{K})_{(p+2)\times 1}$ and D_ρ given by (B.1). Now let $\xi_1 \in \mathbb{R}^p$ and $\xi_2 \in \mathbb{R}$ be arbitrary, and $\zeta := (\xi_1, \xi_2, 0)_{(p+2)\times 1}$. Then, as in Theorem A.1 we can show that the efficiency bound for estimating $\eta(\phi_0) := \xi_1'\theta^* + \xi_2 b^* = \zeta'\beta^*$ is given by $\zeta'(D_\rho' V_\rho^{-1} D_\rho)^{-1}\zeta$. The desired result follows by (B.2) and the fact that ξ_1 and ξ_2 are arbitrary. □

Appendix C. Empirical likelihood based inference with censored and truncated data

We now briefly describe how the GMM based results obtained in Sections 4 and 5 also hold for the empirical likelihood (EL) approach that has lately begun to emerge as a serious contender to GMM; see, e.g., Qin and Lawless (1994), Imbens (1997), Kitamura (1997, 2001, 2006), Smith (1997, 2005), Imbens, Spady, and Johnson (1998), and Owen (2001). Although EL and GMM based inference is asymptotically equivalent up to a first order analysis, recent research by Newey and Smith (2004) has shown that under certain regularity conditions EL has better second order properties than GMM; e.g., unlike GMM, the second order bias of EL does not depend upon the number of moment conditions which makes it very attractive for estimating models with large q, such as panel data models with long time dimension, where GMM is known to perform poorly in small samples.

As for GMM with censored data, we use the moment function defined in (4.7) to do EL based estimation and testing.[19] 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 a fixed β, 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 $\sum_{j=1}^n \rho(Z_j, \beta) p_j = 0$. The solution is given by $\hat{p}_j(\beta) := n^{-1}[1 + \lambda'(\beta)\rho(Z_j, \beta)]^{-1}$, $j = 1, \ldots, n$, where the Lagrange multiplier $\lambda(\beta)$ satisfies $\sum_{j=1}^n \rho(Z_j, \beta)/[1 + \lambda'(\beta)\rho(Z_j, \beta)] = 0$. We define the empirical likelihood estimator of β^* as $\hat{\beta}_{el} := \operatorname{argmax}_{\beta \in \mathcal{B}} \mathrm{EL}(\beta)$, where $\mathrm{EL}(\beta) := \sum_{j=1}^n \log \hat{p}_j(\beta) = -\sum_{j=1}^n \log\{1 + \lambda'(\beta)\rho(Z_j, \beta)\} - n \log n$.

Under (i)–(v) of Assumption 4.1, consistency of $\hat{\beta}_{el}$ follows from Newey and Smith (2004, Theorem 3.1). Moreover, under (vi)–(viii) of Assumption 4.1, EL and GMM estimators have the same asymptotic distribution; see Theorem 3.2 of Newey and Smith and related discussion on p. 673 of Guggenberger and Smith (2005). Hence, $\hat{\beta}_{el}$ is also asymptotically efficient. Note that in finite samples the GMM and EL estimators are generally different although the two coincide if θ^* is just identified because then the EL probabilities $\hat{p}_j(\beta) = 1/n$ for each j and β.

Parametric restrictions of the form $H(\theta^*) = 0$ can be tested by using the empirical likelihood ratio test described in Qin and Lawless (1994, Theorem 2). An EL based specification test can also be developed if θ^* is overidentified. 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 $\hat{\beta}$ denote $n^{1/2}$-consistent preliminary estimator of β^*; e.g., $\hat{\beta}$ can be the GMM or EL estimator defined previously. The restricted, i.e., under (2.1), empirical likelihood can be written as $\mathrm{EL}^r := \sum_{j=1}^n \log \hat{p}_j(\hat{\beta})$. Next, consider the unrestricted problem where the model is not imposed. It is well known that the nonparametric maximum likelihood

[19] EL based inference under truncation is exactly the same; just use the moment function in (5.4).

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 $\mathrm{EL}^{ur} := -n \log n$. Now let $\mathrm{ELR} := 2(\mathrm{EL}^{ur} - \mathrm{EL}^{r}) = 2\sum_{j=1}^{n} \log\{1 + \lambda'(\hat{\beta})\rho(Z_j, \hat{\beta})\}$. Then ELR can be regarded as an analog of the usual parametric likelihood ratio test statistic; i.e., (2.1) is rejected if ELR is large enough. By Qin and Lawless (1994, Corollary 4), $\mathrm{ELR} \xrightarrow{d} \chi^2_{q-p}$ under the null; hence, critical values are easily obtained.

Appendix D. Data Appendix

Compulsory schooling law variables. 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). We also follow them in creating four indicator variables, depending on whether years of compulsory schooling are 8 or less, 9, 10, and 11 or more.

Age at first marriage. The Census dataset includes information on age in years and age at first marriage in years. It also provides information on quarter of birth and quarter of first marriage. We use these variables to calculate age and age at first marriage in quarters as follows. In 1960, 1970, and 1980, the Census took place on April 1st, i.e., right at the beginning of the second quarter. We assume that each individual's birthday took place in the middle of the quarter of birth. Thus, we can calculate age in quarters as being equal to

$$\text{Age (in quarters)} = \begin{cases} \text{age (in years)} + 0.125 & \text{if birth quarter} = 1 \\ \text{age (in years)} + 0.375 & \text{if birth quarter} = 4 \\ \text{age (in years)} + 0.625 & \text{if birth quarter} = 3 \\ \text{age (in years)} + 0.875 & \text{if birth quarter} = 2 \end{cases}$$

Similarly, we use information on quarter of birth and quarter of first marriage to calculate age at first marriage in quarters. If the marriage quarter is one quarter after the birth quarter, then we calculate age at first marriage as being age at first marriage (in years) plus 0.25. If the marriage quarter is two quarters after the birth quarter, then we calculate age at first marriage as being age at first marriage (in years) plus 0.5. If the marriage quarter is three quarters after the birth quarter, then we calculate age at first marriage as being age at first marriage (in years) plus 0.75. If the marriage quarter and the birth quarter coincide, we cannot tell whether the marriage date is before or after the birthday and so the detailed age at marriage could either be very close to the reported age at marriage or very close to the next age. In this case, we simply calculate detailed age at first marriage as being age at marriage plus 0.5.

Appendix E. 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	25.88	5.99	16.1	35.9
Age at First Marriage	20.18	3.59	14.3	35.9
Never Married	0.29	0.45	0	1
White	0.88	0.32	0	1
$\leq$ 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
$\geq$ 11 Years of Schooling Required	0.07	0.26	0	1
		1970 (216036 observations)		
Birth Cohort	1934.69	5.94	1925	1944
Age	35.81	5.94	26.1	45.9
Age at First Marriage	21.73	5.19	14.3	45.9
Never Married	0.07	0.25	0	1
White	0.88	0.32	0	1
$\leq$ 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
$\geq$ 11 Years of Schooling Required	0.07	0.26	0	1
		1980 (223903 observations)		
Birth Cohort	1934.73	5.95	1925	1944
Age	45.76	5.96	36.1	55.9
Age at First Marriage	22.57	7.01	12.3	55.9
Never Married	0.05	0.22	0	1
White	0.88	0.33	0	1
$\leq$ 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
$\geq$ 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	25.81	6.00	16.1	35.9
Age at First Marriage	21.85	3.96	14.3	35.9
Never Married	0.42	0.49	0	1
White	0.89	0.31	0	1
$\leq$ 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
$\geq$ 11 Years of Schooling Required	0.07	0.26	0	1
		1970 (207129 observations)		
Birth Cohort	1934.71	5.94	1925	1944
Age	35.79	5.95	26.1	45.9
Age at First Marriage	24.32	5.25	14.3	45.9
Never Married	0.10	0.30	0	1
White	0.90	0.30	0	1
$\leq$ 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
$\geq$ 11 Years of Schooling Required	0.07	0.26	0	1
		1980 (212244 observations)		
Birth Cohort	1934.80	5.93	1925	1944
Age	45.70	5.93	36.1	55.9
Age at First Marriage	25.44	7.25	12.3	55.9
Never Married	0.07	0.25	0	1
White	0.89	0.31	0	1
$\leq$ 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
$\geq$ 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 of 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.	.0155*	.0078*	.0094*	.0092*	.0090*	.0031	.0076*
	(.0016)	(.0021)	(.0025)	(.0022)	(.0021)	(.0017)	(.0021)
10 Years Schooling Reqd.	.0229*	.0109*	.0143*	.0128*	.0127*	.0075*	.0101*
	(.0020)	(.0029)	(.0034)	(.0030)	(.0029)	(.0026)	(.0030)
11+ Years Schooling Reqd.	.0449*	.0309*	.0180*	.0173*	.0217*	.0156*	.0292*
	(.0037)	(.0051)	(.0060)	(.0053)	(.0050)	(.0048)	(.0056)
White	−.0257*	−.0467*	−.0813*	−.0828*	−.0772*	−.0386*	−.0524*
	(.0012)	(.0017)	(.0021)	(.0017)	(.0017)	(.0014)	(.0016)
Hausman statistic	–	–	–	1.2776	3.8332	–	–
				[0.7345]	[0.2800]		
Men	**OLS60**	**OLS70**	**OLS80**	**GMM60**	**GMM70**	**TOBIT60**	**TOBIT70**
9 Years Schooling Reqd.	.0111*	.0073*	.0030	.0033	.0056*	−.0028	.0062*
	(.0015)	(.0021)	(.0025)	(.0022)	(.0021)	(.0018)	(.0021)
10 Years Schooling Reqd.	.0202*	.0119*	.0127*	.0117*	.0127*	.0053	.0108*
	(.0018)	(.0028)	(.0034)	(.0030)	(.0028)	(.0028)	(.0030)
11+ Years Schooling Reqd.	.0355*	.0150*	.0047	.0024	.0074	.0056	.0119*
	(.0034)	(.0052)	(.0062)	(.0055)	(.0051)	(.0052)	(.0056)
White	−.0153*	−.0435*	−.0778*	−.0753*	−.0743*	−.0295*	−.0505*
	(.0010)	(.0017)	(.0021)	(.0019)	(.0018)	(.0015)	(.0017)
Hausman statistic	–	–	–	1.6544	6.2699	–	–
				[0.6471]	[0.0992]		

Standard errors are in parentheses and p-values are in square brackets. An asterisk denotes that the effect is significant at 5% level of significance. The Hausman statistic tests the null hypothesis that the three compulsory schooling law estimates are the same as those estimated by OLS80.

Table 5. (Robustness check) Effects of compulsory schooling laws and race on log of age at first marriage when age at first marriage for unmarried individuals in the refreshment sample is imputed to be 55 or 65 years.

	55 years			65 years		
Women	OLS80	GMM60	GMM70	OLS80	GMM60	GMM70
9 Years Schooling Reqd.	.0099* (.0027)	.0097* (.0023)	.0095* (.0023)	.0105* (.0029)	.0103* (.0026)	.0101* (.0025)
10 Years Schooling Reqd.	.0130* (.0038)	.0116* (.0034)	.0115* (.0033)	.0137* (.0042)	.0122* (.0037)	.0121* (.0036)
11+ Years Schooling Reqd.	.0137* (.0065)	.0129* (.0058)	.0175* (.0055)	.0141* (.0071)	.0131* (.0063)	.0177* (.0061)
White	−.0957* (.0024)	−.0969* (.0021)	−.0910* (.0020)	−.1056* (.0026)	−.1065* (.0024)	−.1004* (.0023)
Hausman statistic	–	1.0152 [0.7976]	3.5848 [0.3099]	–	0.9352 [0.8169]	3.2461 [0.3552]
Men	OLS80	GMM60	GMM70	OLS80	GMM60	GMM70
9 Years Schooling Reqd.	.0017 (.0027)	.0019 (.0024)	.0045* (.0023)	.0018 (.0030)	.0021 (.0027)	.0047 (.0025)
10 Years Schooling Reqd.	.0107* (.0038)	.0096* (.0034)	.0108* (.0032)	.0113* (.0042)	.0102* (.0039)	.0113* (.0036)
11+ Years Schooling Reqd.	−.0018 (.0068)	−.0041 (.0061)	.0014 (.0058)	−.0026 (.0075)	−.0052 (.0068)	.0005 (.0064)
White	−.0915* (.0024)	−.0885* (.0022)	−.0873* (.0021)	−.1016* (.0027)	−.0981* (.0025)	−.0968* (.0024)
Hausman statistic	–	1.5428 [0.6724]	6.3554 [0.0955]	–	1.5533 [0.6700]	5.9269 [0.1152]

Standard errors are in parentheses and p-values are in square brackets. An asterisk denotes that the effect is significant at 5% level of significance. The Hausman statistic tests the null hypothesis that the three compulsory schooling law estimates are the same as those estimated by OLS80.

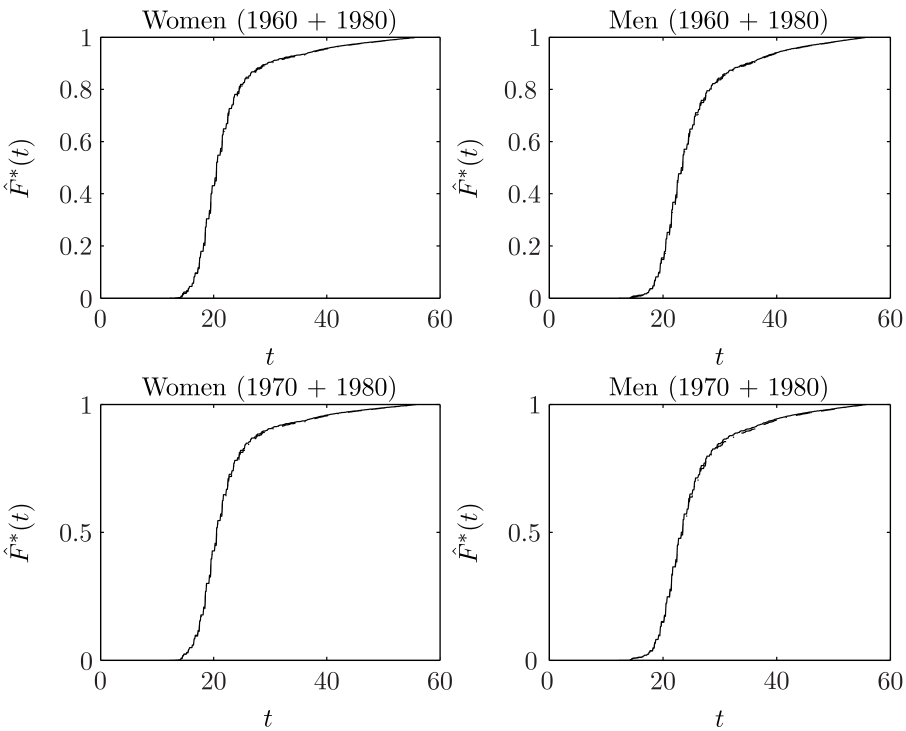

FIGURE 1. Estimates of $F^*(t) := \Pr_{f^*}(Y^* \leq t)$ using the refreshment sample alone are plotted using a dashed line and estimates using the enriched sample are drawn with a solid line. However, they are virtually indistinguishable.

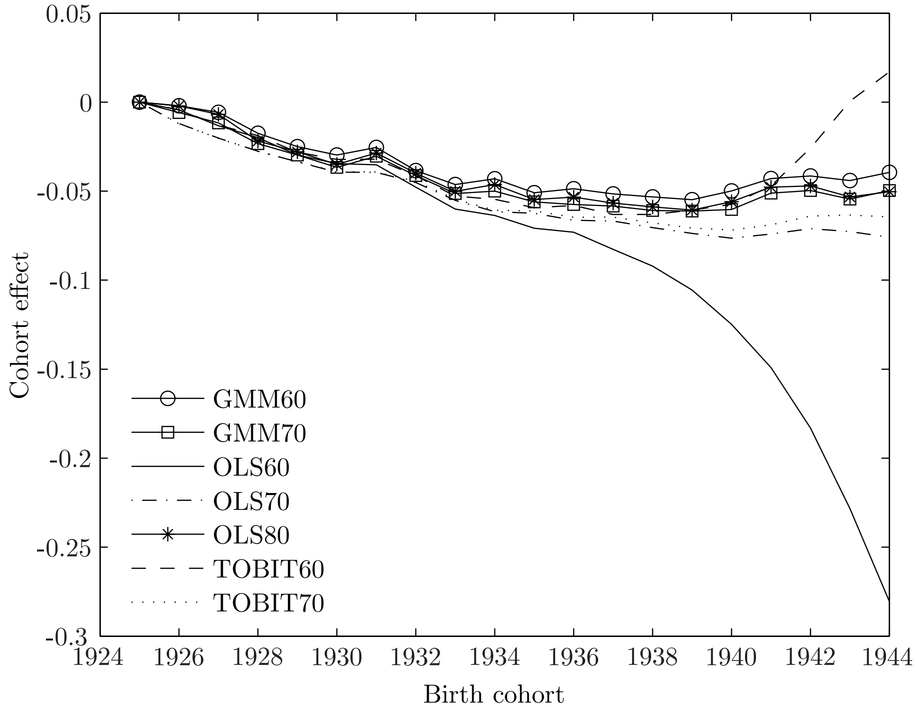

FIGURE 2. Cohort effects for women.

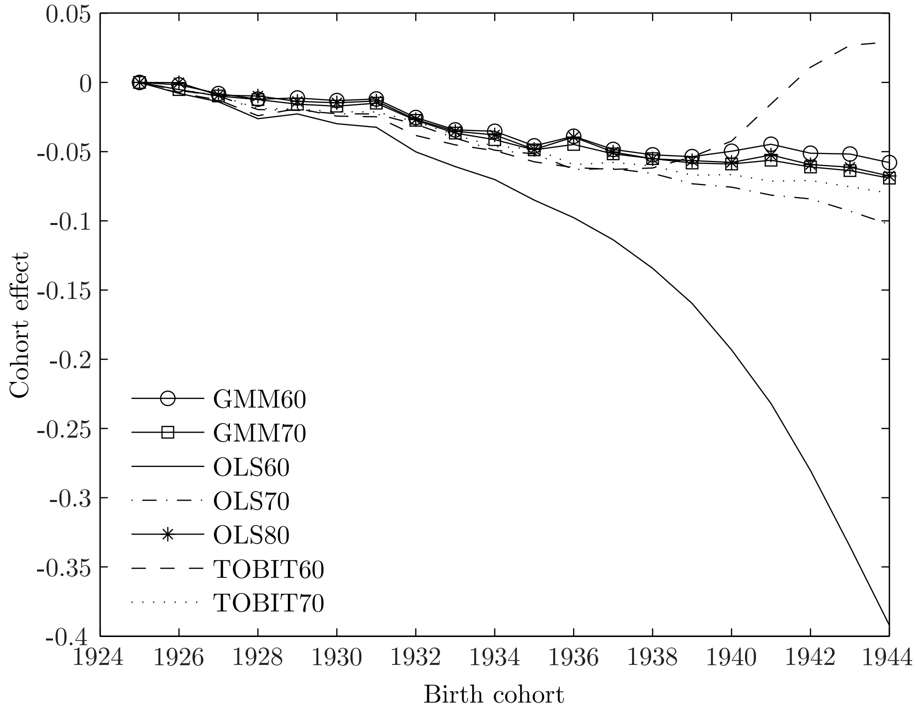

FIGURE 3. Cohort effects for men.

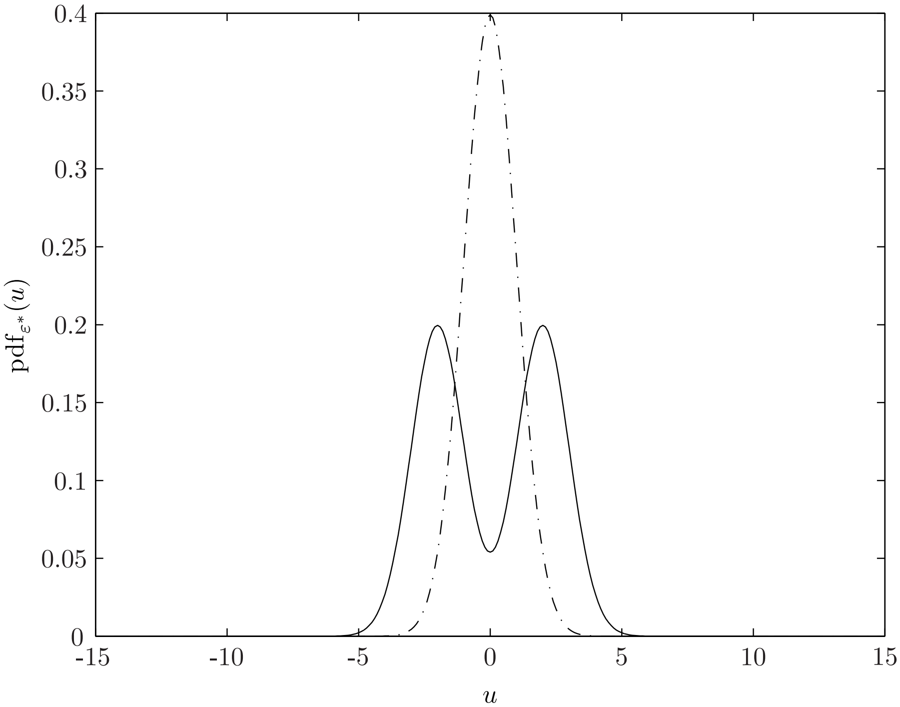

FIGURE 4. The density of ε^*, an equiprobable mixture of $\mathrm{N}(-2, 1)$ and $\mathrm{N}(2, 1)$ random variables, is represented by the solid line; the standard normal density, drawn for reference, is the dashed line.

Table 6. Simulation results when c is fixed and $n_R/n_M = 20\%$.

Censoring		$n = 120$				$n = 600$		
High (75%)	Bias	Variance	MSE	MSE/MSE($\hat{\theta}$)	Bias	Variance	MSE	MSE/MSE($\hat{\theta}$)
$\bar{Y}$	-2.1990	0.0021	4.8376	34.3925	-2.1996	0.0004	4.8387	176.4706
TOBIT	-1.1195	0.0722	1.3255	9.4236	-1.1291	0.0134	1.2882	46.9820
$\hat{\theta}_R$	0.0004	0.2504	0.2504	1.7805	0.0033	0.0503	0.0503	1.8335
$\hat{\theta}$	0.0036	0.1406	0.1407	1.0000	0.0007	0.0274	0.0274	1.0000
Low (25%)								
$\bar{Y}$	-0.2014	0.0391	0.0796	1.7395	-0.2000	0.0078	0.0478	5.2541
TOBIT	0.1690	0.0752	0.1038	2.2673	0.1663	0.0147	0.0424	4.6587
$\hat{\theta}_R$	-0.0000	0.2461	0.2461	5.3759	0.0014	0.0494	0.0494	5.4256
$\hat{\theta}$	-0.0012	0.0458	0.0458	1.0000	-0.0003	0.0091	0.0091	1.0000

Table 7. Simulation results when c is random and $n_R/n_M = 20\%$.

Censoring		$n = 120$				$n = 600$		
High (75%)	Bias	Variance	MSE	MSE/MSE($\hat{\theta}$)	Bias	Variance	MSE	MSE/MSE($\hat{\theta}$)
$\bar{Y}$	-2.2841	0.0093	5.2262	34.1797	-2.2832	0.0019	5.2147	169.4479
TOBIT	-0.7532	0.1084	0.6757	4.4188	-0.7654	0.0200	0.6058	19.6852
$\hat{\theta}_R$	0.0015	0.2499	0.2499	1.6346	-0.0008	0.0507	0.0507	1.6464
$\hat{\theta}$	0.0023	0.1529	0.1529	1.0000	-0.0019	0.0308	0.0308	1.0000
Low (25%)								
$\bar{Y}$	-0.2830	0.0379	0.1180	2.3276	-0.2829	0.0076	0.0877	8.7884
TOBIT	0.1110	0.0700	0.0823	1.6239	0.1077	0.0141	0.0257	2.5749
$\hat{\theta}_R$	-0.0024	0.2442	0.2442	4.8176	0.0023	0.0497	0.0497	4.9833
$\hat{\theta}$	-0.0025	0.0507	0.0507	1.0000	0.0005	0.0100	0.0100	1.0000

TABLE 8. Simulation results when c is fixed and $n_R/n_M = 80\%$.

Censoring	$n = 180$				$n = 900$			
High (75%)	Bias	Variance	MSE	MSE/MSE($\hat{\theta}$)	Bias	Variance	MSE	MSE/MSE($\hat{\theta}$)
$\bar{Y}$	-2.1996	0.0021	4.8404	107.1511	-2.1996	0.0004	4.8388	558.0203
TOBIT	-1.1229	0.0753	1.3362	29.5806	-1.1277	0.0133	1.2851	148.1967
$\hat{\theta}_R$	0.0010	0.0638	0.0638	1.4133	-0.0004	0.0125	0.0125	1.4405
$\hat{\theta}$	-0.0003	0.0452	0.0452	1.0000	0.0005	0.0087	0.0087	1.0000
Low (25%)								
$\bar{Y}$	-0.1990	0.0392	0.0787	2.7364	-0.1998	0.0079	0.0478	8.4238
TOBIT	0.1704	0.0758	0.1049	3.6440	0.1668	0.0149	0.0427	7.5292
$\hat{\theta}_R$	-0.0008	0.0633	0.0633	2.2013	-0.0001	0.0124	0.0124	2.1857
$\hat{\theta}$	-0.0005	0.0288	0.0288	1.0000	-0.0002	0.0057	0.0057	1.0000

TABLE 9. Simulation results when c is random and $n_R/n_M = 80\%$.

Censoring	$n = 180$				$n = 900$			
High (75%)	Bias	Variance	MSE	MSE/MSE($\hat{\theta}$)	Bias	Variance	MSE	MSE/MSE($\hat{\theta}$)
$\bar{Y}$	-2.2827	0.0091	5.2199	111.0475	-2.2827	0.0018	5.2127	571.0471
TOBIT	-0.7524	0.1140	0.6801	14.4692	-0.7618	0.0202	0.6005	65.7898
$\hat{\theta}_R$	0.0021	0.0633	0.0633	1.3466	0.0005	0.0123	0.0123	1.3517
$\hat{\theta}$	0.0005	0.0470	0.0470	1.0000	0.0007	0.0091	0.0091	1.0000
Low (25%)								
$\bar{Y}$	-0.2824	0.0382	0.1179	3.9918	-0.2837	0.0078	0.0883	15.1844
TOBIT	0.1107	0.0695	0.0817	2.7660	0.1067	0.0142	0.0256	4.4095
$\hat{\theta}_R$	-0.0005	0.0632	0.0632	2.1393	-0.0009	0.0124	0.0124	2.1379
$\hat{\theta}$	-0.0004	0.0295	0.0295	1.0000	-0.0013	0.0058	0.0058	1.0000

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.

Angrist, J. D. and A. B. Kreuger (1992): "The effect of age at school entry on educational attainment: An application of instrumental variables with moments from two samples," *Journal of the American Statistical Association*, 87, 328–336.

Arellano, M. and C. Meghir (1992): "Female labor supply and on-the-job search: An empirical model estimated using complementary data sets," *Review of Economic Studies*, 59, 537–557.

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

Black, S. E., P. J. Devereux, and K. G. Salvanes (2007): "Out of the classroom and into the maternity ward? The effects of compulsory schooling laws on teenage births," *Economic Journal*, forthcoming.

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*, 117–143.

Chen, X., H. Hong, and E. Tamer (2005): "Measurement error models with auxiliary data," *Review of Economic Studies*, 72, 343–366.

Chen, X., H. Hong, and A. Tarozzi (2004): "Semiparametric efficiency in GMM models of nonclassical measurement errors, missing data and treatment effects," Manuscript.

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

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.

Guggenberger, P. and R. J. Smith (2005): "Generalized empirical likelihood estimators and tests under partial, weak, and strong identification," *Econometric Theory*, 21, 667–709.

Hajivassiliou, V. A. and P. A. Ruud (1994): "Classical estimation methods for LDV models using simulation," in *Handbook of Econometrics, vol. IV*, ed. by R. Engle and D. McFadden, Elsevier Science B.V., 2383–2441.

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.

Hu, Y. and G. Ridder (2003): "Estimation of nonlinear models with measurement error using marginal information," Manuscript.

Ichimura, H. and E. Martinez-Sanchis (2005): "Identification and estimation of GMM models by a combination of two data sets," Manuscript.

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.

Imbens, G. W., R. H. Spady, and P. Johnson (1998): "Information theoretic approaches to inference in moment condition models," *Econometrica*, 66, 333–357.

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.

——— (2006): "Empirical Likelihood Methods in Econometrics: Theory and Practice," Invited symposium on Weak Instruments and Empirical Likelihood at the 9th World Congress of the Econometric Society.

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.

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

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.

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. 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., 2111–2245.

Newey, W. K. and R. J. Smith (2004): "Higher order properties of GMM and generalized empirical likelihood estimators," *Econometrica*, 72, 219–255.

Oreopoulos, P. (2007): "Do dropouts drop out too soon? Wealth, health, and happiness from compulsory schooling," *Journal of Public Economics*, forthcoming.

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

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

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

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

Ridder, G. and R. Moffitt (2003): “The econometrics of data combination,” Manuscript.

Rigobon, R. and T. M. Stoker (2003): “Censored regressors and expansion bias,” Working paper 4451-03. MIT Sloan School of Management.

Ruggles, S., M. Sobek, T. Alexander, C. A. Fitch, R. Goeken, P. K. Hall, M. King, and C. Ronnander (2004): *Integrated Public Use Microdata Series: Version* 3.0, Minnesota Population Center, Minneapolis, MN.

Severini, T. A. and G. Tripathi (2001): “A simplified approach to computing efficiency bounds in semiparametric models,” *Journal of Econometrics*, 102, 23–66.

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

——— (2005): “Weak Instruments and Empirical Likelihood: A Discussion of the Papers by D.W.K. Andrews and J. H. Stock and Y. Kitamura,” Invited discussion of the symposium on Weak Instruments and Empirical Likelihood at the 9th World Congress of the Econometric Society.

Tripathi, G. (2007): “Moment based inference with stratified data,” *Econometric Theory*, forthcoming.

Department of Economics, University College Dublin, Dublin, Ireland.
E-mail address: devereux@ucd.ie

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