Estimation with Censored Regressors: Basic Issues

Roberto Rigobon Thomas M. Stoker*

Final revision, June 2007

Abstract

We study issues that arise for estimation of a linear model when a regressor is censored. We discuss the efficiency losses from dropping censored observations, and illustrate the losses for bound censoring. We show that the common practice of introducing a dummy variable to 'correct for' censoring does not correct bias or improve estimation. We show how censored observations generally have zero semiparametric information, and we discuss implications for estimation. We derive the likelihood function for a parametric model of mixed bound-independent censoring, and apply that model to the estimation of wealth effects on consumption.

KEYWORDS: Top-coding, linear regression, bias, mismeasured data, maximum likelihood

JEL Classification Codes: C13, C24, C42, C81

1. Introduction

It is easy to argue that the development of models of discrete elements in economic data is the most important evolution in econometrics in the latter part of the 20th century. Many decisions by consumers and producers are inherently discrete; what brand to buy, what mode of transportation to take, what house to choose, etc. More subtle but no less important is how

* Sloan School of Management, MIT, 50 Memorial Drive, Cambridge, MA 02142 USA. We thank Jonathan Parker for his generosity in sharing with us not only his data but all the programs used to reproduce the results in his NBER Macro Annual paper. We have received valuable comments from James Heckman, Vincent Hogan, Dale Jorgenson, Richard Blundell, Charles Manski, Orazio Attanasio, James Banks, Costas Meghir, Alberto Abadie, Jerry Hausman, Gary Chamberlain, Peter Bickel, Jin Hahn, Elie Tamer and especially Whitney Newey.

discrete criterion can affect the selected nature of data samples: only women who choose to work have measured market wages, only consumers who choose to redeem a coupon get the discount, etc. The modeling of discrete elements had the further consequence of requiring substantial understanding of nonlinear models in econometrics, as discrete actions or choices are not well represented by standard linear regression models. This began with a thorough development of the econometrics of parametric discrete choice and selection models, and continued with development of more flexible semiparametric and nonparametric econometric models. There is no more important name in this development than Daniel McFadden. His work has had an enormous impact on the work of several generations of econometricians, both on the theoretical side and the practical side. No one else comes close.

In this paper, we look at an aspect of discreteness in econometrics that has been largely overlooked, where regressors in an economic model are discretely censored. The well-known problem covered in the literature is when a dependent variable is censored or selected. That problem has the impact of altering the effective sample for estimation and, for instance, causes biases in OLS estimates of linear model coefficients. This has stimulated a great deal of work on consistent estimators of coefficients when there is a censored dependent variable.

As such, it seems surprising that very little attention has been paid to the implication of having a censored regressor, or independent variable, in the estimation of a linear model. Indeed, it would seem that researchers encounter censored regressors as often or even more often than situations of censored dependent variables. Consider how often variables are observed in ranges, including unlimited top and bottom categories. For instance, observed household income is often recorded in increments of one thousand or five thousand dollars, but would have a top-coded response of, say, "$100,000 and above." Another example is where imperfect measurement has the impact of censoring; for instance, in measuring components of household wealth, there may

be many zero values, some of which are genuine zeros but many represent nonreporting or other mismeasurement These are just a couple examples of where censoring appears, but it seems clear that censored regressors are a common phenomena in empirical work.

If one ignores the censored nature of a regressor, one can induce a particularly insidious practical problem, namely estimates that are too large. This phenomena, which we term *expansion bias*,[1] can give a spurious impression of the importance of a regressor. To see how this can arise, Figure 1 shows a scatterplot where a regressor is top-coded and bottom-coded, or double bound censored. The small circles are the resulting data points when the regressor is censored at upper and lower bounds. The estimated regression using the censored regressor clearly has a steeper slope that the one using the uncensored regressor. Expansion bias arises because of the "pile-up" of observations at each limit. Obviously, expansion bias would arise if there was only top-coding or bottom-coding alone.

One approach would be to just view any observation with censoring as bad, and drop them for estimation. This is called *complete case* analysis. However, this has the potential to introduce further bias from selecting the sample in an endogenous way. This is not a problem under *exogenous censoring* as we define below; in that situation, complete case analysis provides consistent parameter estimates.

We consider several aspects of model estimation with a censored regressor. We assume exogenous censoring, which reflects censoring that is not connected to the dependent variable under study. Our interest is in issues of estimation with the full data sample. We illustrate the efficiency loss due to censoring, highlighting non-independent censoring such as top-coding. We show that the common practice of including a dummy variable for censored observations is not

[1] Expansion bias is the opposite of *attenuation bias*, familiar from problems such as errors-in-variables or censored dependent variable models.

advisable – the procedure eliminates bias only under strong restrictions, and otherwise, no useful information is gained from including the censored observations in this way. We then establish this feature more broadly, by showing that there is zero semiparametric information for the parameters of interest in the censored data, when there are no restrictions on the distribution of the censored observations. We discuss general estimation under the assumption that a proxy equation is appropriate for the censored regressor. For our empirical application, we specify a parametric model of mixed independent and bound censoring. We derive the likelihood function to facilitate maximum likelihood estimation of our mixed censoring model.[2]

We discuss how censored regressors arise in the estimation of wealth effects on consumption. We show how extensive censoring can be, when components of wealth are taken into consideration. We apply our mixed censoring model to analyze household consumption and wealth data, and compare the estimation results to those obtained from linear regression that ignores the censoring. We show how expansion bias manifests in simple regression models, and how the the size and precision of wealth and income effects is changed when wealth censoring is taken into account. Our discussion is intended to give concrete illustration to the ideas, and we plan to carry out further applications as part of future research.

As mentioned above, there is relatively little literature on censored regressors in econometrics. An exception is Manski and Tamer (2002), who study identification and consistent estimation with interval data. The statistical literature on missing data problems covers some situations of censored regressors, with most results applicable to data missing at random. See the surveys by Little (1992) and Little and Rubin (2002) for coverage of this large literature, and Ridder and Moffit (2003) for survey of the related literature on data combination.[3] Top-coding and bottom-

[2] Our discussion focuses on exogenous censoring, but the derivation of the likelihood function includes the situation where the censored regressor is endogenous and instruments are observed.

[3] A valuable early contribution is Ai (1997). Recent contributions include Chen, Hong and Tamer (2005),

coding are nonignorable data coarsenings in the sense of Heitjan and Rubin (1990, 1991).[4] Also related is recent work on partially identified econometric models, which often include situations of censored regressors; see Chernozhukov, Hong and Tamer (2004) and Shaik (2005) among others. Related discussions on information and efficiency can be found in Horowitz and Manski (1998, 2000), Robins and Rotnitzky (1995) and Rotnitzky, Robins and Scharfstein (1998).

This paper is part of a series on the problems raised by censored regressors. The bias that arises from censored regressors is studied in great detail in Rigobon and Stoker (2006a), including results for bias in OLS estimators, bias in IV estimators when the censored regressor is endogenous, and bias transmission in situations of multiple regressors and with extreme 0-1 censoring. Testing for the presence of bias from censored regressors is covered in Rigobon and Stoker (2006b). This amounts to testing whether potentially censored values are, in fact, correctly measured.[5] Under exogenous censoring as introduced below, straightforward chi square tests are available.

The paper is organized as follows: Section 2 discusses the basic results of using censored regressors in the estimation of linear models. Section 3 discusses consistent estimation with the full data sample, presents our model of mixed independent and bound censoring, and derives the likelihood function for that model. Section 4 illustrates the biases that arise in an application where the marginal propensity of consumption out of wealth is estimated. Finally, Section 5 concludes.

Chen, Hong and Tarozzi (2004), Liang, Wang, Robins and Carroll (2004), Tripathi (2003,2004), Mahajan (2004) and Ichimura and Martinez-Sanchis (2005).

[4] Some recent work has shown how data heaping in duration data (censoring or rounding due to memory effects) data can be accomodated in estimation of survival models. See Torelli and Trivellato (1993) and Petoussis, Gill and Zeelenberg (2004), among others.

[5] For instance, suppose a variable is bounded below by 0. The question is whether the observed 0 values are correct observations or censored values.

2. Basic Issues of Estimation with Censored Regressors

2.1. Framework: Linear Model and Censoring

We consider the impact of censoring in a linear regression framework. We assume that the true model is an equation of the form

$$y_i = \alpha + \beta x_i + \phi' w_i + \varepsilon_i \quad i = 1, ..., n \tag{1}$$

where x_i is a single regressor of interest, and w_i is a k-vector of other regressors. We assume that the distribution of $\left(x_i, w_i', \varepsilon_i\right)$ is nonsingular and has finite second moments. We assume that the model is a properly specified regression model, with $E\left(\varepsilon_i | x_i, w_i\right) = 0$.

We do not observe x_i for all observations, but rather a censored version of it. Suppose that the indicator d_i describes the censoring process, with $d_i = 0$ denoting an uncensored observation and $d_i = 1$ a censored one, for which we observe the value ξ. That is, we observe

$$x_i^{cen} = \left(1 - d_i\right) x_i + d_i \xi \tag{2}$$

where x_i^{cen} is the censored version of x_i. The probability of censoring is denoted as $p = \Pr\{d = 1\}$, and we assume that $0 < p < 1$.

Our model of censoring includes most of the common types of censoring found in practice. The process for d_i can be quite general, but we assume $p < 1$, so that some correct (uncensored) values of x_i are observed. Another restriction is to censoring to a single value ξ. This is a convenience, and many of the points we make will apply to censoring to two or more different values.

Our framework includes single value *bound censoring*. For instance, *top-coding* involves observing x_i only when it is less than a bound ξ, namely

$$d_i = 1\left[x_i > \xi\right] \tag{3}$$

and the bound ξ is the censoring value. *Bottom-coding* involves observing x_i only when it is above a bound ξ, with

$$d_i = 1\left[x_i < \xi\right] \tag{4}$$

Double bounding, where there is both top-coding and bottom-coding at the same time, is two-value censoring. This case would not change our analysis meaningfully.

The processes (3) and (4) have d_i determined by x_i, and our framework includes cases where d_i is a stochastic censoring process or a more complicated deterministic censoring process. For instance, independent censoring refers to where d_i is statistically independent of x_i, w_i and ε_i. Random processes that involve dependence of virtually any kind can be included.[6] One important omission from our discussion is 0-1 censoring. This refers to where a dummy variable is observed in place of x_i, for instance

$$x_i^{cen} = 1\left[x_i \geq \xi\right] \tag{5}$$

where x_i^{cen} indicates whether x_i is above the threshold ξ. This is two-value censoring, but the problem is that no true values of x_i are observed. Every observation is censored, which is a violation of $p < 1$. Estimation in this case will involve some different considerations that those

[6] In the parlance of the missing data literature (c.f. Little and Rubin (2002)), our notion of independent censoring is analogous to "missing completely at random," or MCAR. Top-coding and bottom-coding involve censoring determined by the value of the regressor, so that they are analogous to "not missing at random" processes, or NMAR, where in addition, the censoring threshold is given by the censoring value ξ.

we discuss here.[7]

It is worth mentioning that, in linear regression analysis, the main problem that censoring causes is bias. Namely, if we ignore that x_i^{cen} is not x_i and estimate the model

$$y_i = a + bx_i^{cen} + f' w_i + u_i \quad i = 1, ..., n, \tag{6}$$

then the estimates $\hat{a}$, $\hat{b}$, $\hat{f}$ are asymptotically biased estimators of α, β, ϕ.[8] The bias can easily be seen to depend on the censoring process as well as the censoring value ξ.

2.2. Selection and Exogenous Censoring

Since we observe x_i for a fraction of the sample, why not just estimate with those full observations? To consider this, suppose that the sample is ordered with the $n_0 = \sum_{i=1}^{n} (1 - d_i)$ uncensored observations first, $i = 1, ..., n_0$, followed by the $n_1 = \sum_{i=1}^{n} d_i$ censored observations, $i = n_0 + 1, ...n_0 + n_1$. Therefore, consider estimating the equation

$$y_i = \alpha + \beta x_i + \phi' w_i + \varepsilon_i \quad i = 1, ..., n_0. \tag{7}$$

This is referred to as *Complete Case (CC)* regression analysis.

This raises issues that are familiar to students of selection problems. The question is how the distribution of ε_i is altered by restricting attention to observations with $d_i = 0$. When the mean of ε_i varies with d_i, then CC analysis induces biases from truncation. This is the same problem as with traditional models of (bound) censored dependent variables, where $d_i = 1[y_i < \zeta]$, and

[7] See Rigobon and Stoker (2006a) for a discussion of OLS bias with 0-1 censoring.

[8] Zero asymptotic bias occurs only in very unusual situations. One case arises with three conditions holding simultaneously: (a) d_i is independent of x_i, w_i and ε_i, (b) censoring is to the mean $\xi = E(x)$ and (c) x_i is independent of w_i.

a CC regression must be adjusted for the truncated nature of the CC data.[9] That is, in full generality, a censored regressor can induce problems similar to those caused by censoring of the dependent variable.

For our discussion of basic issues raised by censored regressors, we assume *exogenous censoring*, namely

$$E\left(\varepsilon_i | d_i, x_i, w_i\right) = 0 \tag{8}$$

This assumes away the standard problems of censoring or truncating the dependent variable. Under exogenous censoring, the CC model (7) is a well-specified regression model. CC regression analysis gives consistent estimators of α, β and ϕ. For part of our discussion, we will need a stronger version of exogeneity. In particular, we define *strict exogenous censoring* as statistical independence of ε_i from d_i conditional on x_i and w_i. Thus the distribution of ε_i conditional on x_i and w_i is the same when further conditioned by $d_i = 0$, which clearly includes (8).

Under exogenous censoring, the basic estimation issues concern how best to employ the censored observations to improve estimation. We now turn to those issues. In passing, it is worth mentioning that exogenous censoring also provides the foundation for straightforward Hausman tests of the absence of bias from censored regressors. Under exogenous censoring, estimation of (7) with complete cases gives consistent estimates of α, β and ϕ, and under the null hypothesis of no bias, estimation of (6) with the full sample gives efficient estimates. Chi-square tests based on the difference of these estimators are developed and illustrated in Rigobon and Stoker (2006b).

[9] That is, one adds

$$E\left(\varepsilon_i | d_i = 0, x_i, w_i\right) = E\left(\varepsilon_i | \varepsilon_i \geq \zeta - \alpha - \beta x_i - \phi' w_i\right)$$

to the regression. When ε_i is normally distributed, this expectation is the inverse Mill's ratio, which is added to the regression equation to facilitate consistent estimation.

2.3. Efficiency Loss from Censored Regressors

Recall that the model is

$$y_i = \alpha + \beta x_i + \phi' w_i + \varepsilon_i \quad i = 1, ..., n \tag{9}$$

applying to the full sample, and for simplicity, we now assume homoskedasticity of ε_i,

$$Var\left(\varepsilon_i | x_i, w_i\right) = \sigma^2. \tag{10}$$

We assume strict exogenous censoring. Therefore, CC analysis is consistent; the OLS estimates $\hat{\alpha}_0$, $\hat{\beta}_0$, $\hat{\phi}_0$ and $\hat{\sigma}_0^2$ of

$$y_i = \alpha + \beta x_i + \phi' w_i + \varepsilon_i \quad i = 1, ..., n_0 \tag{11}$$

are consistent for α, β, ϕ and σ^2, respectively.

We are interested in how valuable the censored observations are to estimation. The regression model appropriate for the censored observations is

$$y_i = \alpha + \beta g_1\left(w_i\right) + \phi' w_i + u_i \quad i = n_0 + 1, ..., n_0 + n_1 \tag{12}$$

where

$$g_1\left(w_i\right) = E\left(x_i | w_i, d_i = 1\right) \tag{13}$$

The disturbance

$$u_i = \beta\left(x_i - g_1\left(w_i\right)\right) + \varepsilon_i \tag{14}$$

has $E\left(u_i | w_i, d_i = 1\right) = 0$ and $\sigma_u^2\left(w_i\right) = Var\left(u_i | w_i, d_i = 1\right) = \beta^2 Var\left(x_i | w_i, d_i = 1\right) + \sigma^2$. In essence, since x_i is not available, the best possible situation is where you know the value $g_1\left(w_i\right)$

of the conditional expectation for each i, and $\sigma_u^2(w_i)$ for each i. Then one could do an efficient pooled estimation of (11)-(12), estimating with the whole sample. For the ith observation of the censored data, this amounts to using $g_1(w_i)$ in place of x_i^{cen}, and weighting by $1\Big/\sqrt{\sigma_u^2(w_i)}$. Clearly, this is the best regression procedure given that $x_i - g_1(w_i)$ is not observed. This leads us to two efficiency comparisons to gauge the loss from censoring. First is the relative efficiency of CC analysis with estimation with the full uncensored sample (with x_i observed). Second is the efficiency of the pooled estimation of (11)-(12) described above with $g_1(\cdot)$ known, relative to estimation with the full uncensored sample.

For interpretation, consider the case of where d_i is independent of x_i and w_i. Censoring of $p = 20\%$ of the observations coincides with efficiency of $1 - p = 80\%$ of CC analysis relative to estimation with the full uncensored sample, since the censoring alters nothing but the sample size. If $g_1(w_i)$ is known, there are more efficiency gains the more highly correlated x_i is with w_i, as the unobserved term $x_i - g_1(w_i)$ will have smaller variance.

With independence, the censored observations have the same distribution as uncensored observations. Alternatively, consider bound censoring, or top-coding in particular. Top-coding does not resemble independent censoring; it involves censoring the upper tail, which contains some of the most influential observations for estimating the regression parameters.[10] CC analysis with 20% top-coding will involve a lower efficiency than 80%.

How much lower? Table 1 presents efficiencies for normally distributed regressors for different amounts of censoring from top-coding.[11] The bivariate column uses a model with no w_i, so that the conditional mean is a constant $g_1 = E(x_i|d_i = 1)$ and $\sigma_u^2 = \beta^2 Var(x_i|d_i = 1) + \sigma^2$. With

[10] "Influential" is used here in the same sense as in the literature on regression diagnostics or experimental design: see Belsley, Kuh and Welsch (1980) among many others.

[11] We set $\alpha = 1$, $\beta = 1$, and $\gamma = 1$, took the variances of x and z to be the same and equal to half the value of the variance of ε.

one additional regressor w_i, the mean $g_1(w_i)$ and variance $\sigma_u^2(w_i)$ are computed for the bivariate normal regressors. We see that for the bivariate model, the relative efficiency of CC analysis is much lower than it would be with random sampling: 47% efficiency with 20% top-coding, 25% efficiency with 40% top-coding, etc. Table 1 also addresses how valuable it is to know the mean of the top-coded data. Notice how a great deal of the efficiency loss can be eliminated when $g_1(w_i)$ is known.

When there is an additional regressor, the efficiency loss in estimating β is less that in the bivariate case, and improves with higher correlation between x and w. When the conditional mean of the top-coded data is known, the efficiency improves for each coefficient, but not to the same extent as with the bivariate model. Finally, we notice that the improvements in efficiency for β and ϕ (from knowing the mean) are more balanced with higher correlation.[12]

As such, it appears that with substantial censoring, the efficiency losses from CC analysis can be large, depending on the nature of the censoring. Even with top coding, these losses could be recovered to a large degree if the conditional means and variances for the censored observations are known.

2.4. Ineffectiveness of Dummy Variable Methods

We now take a slight detour about an empirical technique that will, in fact, lead us back to our discussion of efficiency. A common practice in empirical work is to regress y_i on a constant, x_i^{cen}, w_i and the censoring dummy d_i. Here we discuss the practice of including d_i to empirically 'correct' for the censoring.

[12] These calculations are done with optimal (GLS) weighting, but we did not find that the results were very sensitive to whether weighting was done or not.

The true model (1) written with the censored regressor is

$$y_i = \alpha + \beta x_i^{cen} + \phi w_i + \beta \left(g_1 \left(w_i \right) - \xi \right) \cdot d_i + u_i \quad i = 1, ..., n \tag{15}$$

where $u_i = \beta \left(x_i - g_1 \left(w_i \right) \right) d_i + \varepsilon_i$, and we take $g_1 \left(\cdot \right)$ as unknown. Thus, the true 'coefficient' of d_i varies with w_i, which is a potentially very serious misspecification. Unless $g_1 \left(\cdot \right)$ is constant, $g_1 \left(w_i \right) = g_1$, or approximately so,[13] all the coefficient estimates will be biased. It is not clear whether they will be more or less biased than the coefficients obtained from regressing y_i on a constant, x_i^{cen} and w_i – or ignoring the original censoring. In general, the inclusion of the censoring dummy is not advisable.

Consider where the constancy assumption is valid by construction, namely in the bivariate model where there is no additional variable w_i. Now the true model is

$$y_i = \alpha + \beta x_i^{cen} + \beta \left(g_{1_i} - \xi \right) \cdot d_i + u_i \quad i = 1, ..., n \tag{16}$$

where $g_1 = E \left(x | d = 1 \right)$. This is a well specified regression model including the intercept, x_i^{cen} and censoring indicator d_i. However, there is another issue. For the complete cases ($i = 1, ..., n_0$), the model is linear with intercept α and slope β. For the censored data, the model is a constant, with value $\alpha + \beta g_1$. If g_1 is not known, then there is no parameter restriction between the complete cases and the censored data.[14] Therefore, the estimate of β from model (16) is exactly the same as the estimate from CC analysis, or estimating with complete cases only, and it has the same variance. There is no gain from including the censored observations together with the censoring indicator.

[13] For example, if x were income and z a demographic variable, then constancy implies that the mean of top-coded income is the same for all demographic groups indicated by z.

[14] This includes the variances as well.

The same remarks apply to the related procedure of including interactions with d_i. That is, from (15), one might consider approximating $g_1(w_i)$ by a general linear function in w_i. For this, one would regress y_i on an intercept, x_i^{cen}, w_i, d_i and $d_i w_i$. It is easy to see that if $g_1(w_i)$ were linear, then this would be a well specified model. But this parametrization has the same effect as discussed for (16); namely there is no parameter restriction between the complete cases and the censored data. As before, with the mean $g_1(w_i)$ unknown, this procedure yields no gain over CC analysis.

Similar issues arise for the practice of imputing tail means with bound censoring. For instance, if observed income is top-coded at \$100,000, the practice is to replace all top-coded values with an imputed mean of incomes over \$100,000. In view of (16), this practice will adjust for censoring bias in bivariate regression.[15] But when there are additional regressors w_i, this practice is only correct when $g_1(w_i)$ is constant; namely when x_i is mean-independent of w_i given $d_i = 1$. That is, the appropriate imputation would be to replace top-coded values by their conditional expectation on all other regressors, $g_1(w_i)$; doing that correctly could bring the efficiency gains for available when $g_1(\cdot)$ is known.

2.5. The Semiparametric Information in Censored Observations

The fact that dummy variable methods fail to uncover new information about the parameters is ominous, and indicative of a more general issue for flexible approaches to estimation. How much information about the parameters of interest – α, β, ϕ and σ^2 – is available in the censored data?

[15] For bivariate regression, censoring bias is given as plim $\hat{b} = \beta(1+\Lambda)$, where

$$\Lambda = .p(1-p) \cdot \frac{(E(x|d=1)-\xi)(\xi - E(x|d=0))}{Var(x^{cen})}$$

Imputation sets $\xi = E(x|d=1)$, which zeros the bias.

We now answer this question by appealing to the concept of semiparametric information.[16].

The structure we seek is clear from the following example:

Example 1. *For the model (12)-(14) for censored data, assume*

$$\varepsilon \sim \mathcal{N}\left(0, \sigma^2\right) \quad \text{and} \quad \beta\left(x_i - g_1\left(w_i\right)\right) \sim \mathcal{N}\left(0, \sigma_{\beta x}^2\right).$$

Suppose that η is a vector of nuisance parameters, parameterizing the conditional expectation $g^\eta(w) = E(x|w, d = 1, \eta)$, where by construction $\eta = 0$ coincides with the true function $g_1(w) = g^0(w)$. Under these assumptions, the density of y for the censored data is

$$\ln f\left(y|w, \alpha, \beta, \phi, \eta\right) = -\ln\sqrt{2\pi} - (1/2)\ln\left(\sigma_{\beta x}^2 + \sigma^2\right) - (1/2)\frac{\left(y - \alpha - \beta g^\eta(w) - \phi w\right)^2}{\left(\sigma_{\beta x}^2 + \sigma^2\right)} \tag{17}$$

Denoting $\varepsilon = y - \alpha - \beta g^\eta(w) - \phi w$, the scores of the parameters of interest are

$$\ell_\alpha = \frac{\partial \ln f}{\partial \alpha} = \frac{\varepsilon}{\left(\sigma_{\beta x}^2 + \sigma^2\right)}, \tag{18}$$

$$\ell_\beta = \frac{\partial \ln f}{\partial \beta} = \frac{\varepsilon}{\left(\sigma_{\beta x}^2 + \sigma^2\right)} \cdot g_1(w), \tag{19}$$

$$\ell_\phi = \frac{\partial \ln f}{\partial \phi} = \frac{\varepsilon}{\left(\sigma_{\beta x}^2 + \sigma^2\right)} \cdot w \tag{20}$$

and

$$\ell_{\sigma^2} = \frac{\partial \ln f}{\partial \sigma^2} = -\frac{1}{2\left(\sigma_{\beta x}^2 + \sigma^2\right)} + \left(\frac{1}{2}\right) \cdot \frac{\varepsilon^2}{\left(\sigma_{\beta x}^2 + \sigma^2\right)^2}. \tag{21}$$

[16] See Newey (1990) for the definition of semiparametric information and the semiparametric variance bound.

The scores of the nuisance parameters are

$$\ell_\eta = \frac{\partial \ln f}{\partial \eta} = \frac{\varepsilon}{\left(\sigma_{\beta x}^2 + \sigma^2\right)} \cdot \frac{\partial g^\eta(w)}{\partial \eta} \tag{22}$$

$$\ell_{\sigma_{\beta x}^2} = \frac{\partial \ln f}{\partial \sigma_{\beta x}^2} = -\frac{1}{2\left(\sigma_{\beta x}^2 + \sigma^2\right)} + \left(\frac{1}{2}\right) \cdot \frac{\varepsilon^2}{\left(\sigma_{\beta x}^2 + \sigma^2\right)^2} \tag{23}$$

The semiparametric information on α, β, ϕ and σ^2 is the variance of their scores, after projection onto subspace orthogonal to that spanned by the scores of the nuisance parameters. When $g_1(w)$ is unrestricted, then a sufficiently rich parameterization $g^\eta(w)$ can be found such that linear combinations of $\left\{\partial g^\eta(w)/\partial \eta_j\right\}$ will approximate a constant, w and $g_1(w)$ arbitrarily well. Therefore, the projection of ℓ_α, ℓ_β, ℓ_ϕ, ℓ_{σ^2} onto the subspace orthogonal to the span of ℓ_η, $\ell_{\sigma_{\beta x}^2}$ will be arbitrarily small. Consequently, the semiparametric information on α, β, ϕ and σ^2 is zero.

It is clear that for more general settings – in particular, general densities of ε and of x given w – we have the same conclusion[17]

Proposition 2. *If there are no restrictions on the conditional expectation $g_1(w) = E(x|w, d = 1)$, then the semiparametric information on α, β, ϕ and σ^2 from the censored data, is zero. The semiparametric variance bound for the estimation of α, β, ϕ and σ^2 using complete cases only is the same as the semiparametric variance bound using the complete cases together with the censored data.*

Thus, the phenomena discussed with regard to dummy variable methods above applies more generally. There is no gain in estimation from using the censored data, unless restrictions can

[17] The semiparametric variance bound is the inverse of the semiparametric information.

be applied to the conditional expectation $g_1(w)$.[18] We now discuss estimation with this in mind.

3. Estimation with the Full Data Sample

There are a number of approaches for estimation which include the censored observations, but all must add information beyond the basic regression model.[19] We now discuss these issues in the context of (corrected) regression estimators.

3.1. Use of a Proxy Equation

We consider pooled estimation using the complete cases, with model

$$y_i = \alpha + \beta x_i + \phi' w_i + \varepsilon_i \quad i = 1, ..., n_0, \tag{24}$$

together with the censored observations. As noted above, a correctly specified regression model for the censored observations is

$$y_i = \alpha + \beta g_1(w_i) + \phi' w_i + u_i \quad i = n_0 + 1, ..., n_0 + n_1 \tag{25}$$

where $E(u_i|w_i) = 0$, since

$$g_1(w_i) = E(x_i|w_i, d_i = 1) \tag{26}$$

[18] Similar structure is discussed in Horowitz and Manski (1998,2000). See also Robins and Rotnitzky (1995).

[19] There are also likely to be approaches based on partial information and bounds. We do not pursue this year, but note is as a potentially fruitful area of future research.

is the appropriate proxy for x_i, for the censored observations. The question is how to estimate $g_1(w_i)$, in a way that will be valuable for the estimation of α, β, ϕ and σ^2.

With the uncensored observations, we can identify and estimate the conditional expectation of x_i given w_i, namely

$$g_0(w_i) = E(x_i|w_i, d_i = 0) \tag{27}$$

This raises one immediate approach to identifying $g_1(w_i)$ that has received much attention in the econometrics literature, namely independent censoring. If d_i is independent of x_i and w_i, then

$$g_1(w_i) = g_0(w_i) \tag{28}$$

Estimation with the full sample can proceed as follows: form the estimate $\hat{g}_0(\cdot)$ with the complete cases, and then use $\hat{g}_0(w_i)$ in place of x_i^{cen} for the censored observations.[20]

When (28) is not valid, we need some other structure that bridges the censored and uncensored observations. Perhaps the most natural is to assume the existence of a proxy equation for x_i that applies in the full sample. A regression proxy is based on the model

$$x_i = G(w_i) + v_i \tag{29}$$

where $E(v_i|w_i) = 0$; namely $G(w_i) = E(x_i|w_i)$ is the regression applicable to the full sample. When the proxy $G(w_i)$ can be estimated, then we have a method of estimating $g_1(w_i)$. Namely, we estimate $g_0(\cdot)$ with the complete cases, and with the full data sample, we estimate the

[20] For instance, Arellano and Meghir (1992) propose using the best linear predictor of x on z as a proxy, which can be estimated using the complete cases only when the censoring is independent, or doesn't introduce bias. Much recent methodological work relies on the "censoring at random" or "missing at random" structure – see Chen, Hong and Tamer (2005), Chen, Hong and Tarozzi (2004) and Liang, Wang, Robins and Carroll (2004) among others.

conditional probability of censoring

$$p(w_i) = E(d_i|w_i). \tag{30}$$

Therefore, we can estimate $g_1(w_i)$ by plugging those estimates into the identity

$$g_1(w_i) = \frac{G(w_i) - (1 - p(w_i))\, g_0(w_i)}{p(w_i)} \tag{31}$$

The most flexible versions of this approach will require significant regularity conditions; for instance, if a nonparametric estimator of $p(\cdot)$ is used in the denominator of (31), then trimming or some other method will be needed.[21]

The model (29) often will permit estimation of $G(\cdot)$ with the complete cases. If d_i represents bound censoring, say with $d = 1[x > \xi]$, then (29) restricted to complete cases is a truncated regression model.[22] If $G(w)$ is linear, then a variety of semiparametric procedures can be applied to estimate the coefficients. Depending upon the structure assumed, index model estimators, or quantile estimators would be applicable. Here, we implement a fully parametric model of censoring, in part because we are interested in the structure of censoring in our empirical application. However, there is no reason to use that much structure, in principle.

[21] It is useful to note that there are other estimation approaches based on $G(\cdot)$. For instance, one could discard x_i in the complete cases, and estimate using the proxy G for the entire sample, fitting

$$y_i = \alpha + \beta \cdot G(w_i) + \phi' w_i + U_i \quad ; \quad i = 1, ..., n_0 + n_1$$

This idea would seem valuable only in unlikely settings, such as where the complete cases were a tiny fraction of the full data sample, but for some reason $G(w_i)$ is a terrific proxy, capturing almost all of the variation in x_i. Then the loss of $x_i - G(w_i)$ for the complete cases would involve a small loss in estimation efficiency.

[22] One might consider estimating (24) as a reverse regression, but that will not work in our framework. While the reverse regression has a censored dependent variable (x_i^{cen}), it is not well specified, because the regressor (y_i) is correlated with the error term, and part of that correlation is due to the censoring of x_i that we are studying here. An instrument or other additional information would be required.

3.2. A Normal Mixed-Censoring Model

Our application focuses on the wealth effects on consumption. Log wealth is bounded below and censored to 0, but it is not obvious that the censoring follows a natural pattern for censoring from bottom-coding alone. That is, it is not obvious that low wealth values are more likely to be censored than high wealth values. We now propose a parametric censoring model that allows us to examine this issue together with the impact on the estimation of wealth effects. We retain our notation above, where later x_i will be log wealth and $\xi = 0$.

We add to the basic equation (1) by assuming that the proxy $G(w_i) = E(x_i|w_i)$ of x_i is linear

$$G(w_i) = \delta_0 + \delta_1' w_i, \tag{32}$$

and we assume that v_i of (29) is normally distributed and homoskedastic

$$v_i \sim \mathcal{N}\left(0, \sigma_v^2\right). \tag{33}$$

This assumption facilitates modeling bottom-coding with formulae familiar from censored normal regression models.[23]

In our application, we implement a more general censoring model, that allows a mixture of independent censoring and bottom-coding. The approach is to model bottom-coding together with (conditionally independent) censoring of probability $R(w_i)$ for observations that are not bottom-coded. Let

$$d_{1i} = 1\left[v_i < \xi - \left(\delta_0 + \delta_1' w_i\right)\right] \tag{34}$$

[23] See, for instance, Ruud (2000), Green (2003) or Davidson and McKinnon (2004). Analogous formulae are available for top-coding.

and

$$d_{2i} = 1\left[s_i < -\left(\eta_0 + \eta_1^{'} w_i\right)\right] \tag{35}$$

represent the two sources of censoring. We assume $v_i \sim \mathcal{N}(0, \sigma_v^2)$ and $s_i \sim \mathcal{N}(0,1)$, and that v_i and s_i are conditionally independent given w_i.

The overall censoring indicator d is now defined as

$$d_i = d_{1i} + d_{2i} - d_{1i} d_{2i} \tag{36}$$

This reflects bottom-coding, plus a probability of

$$R(w_i) = \Phi\left(-\left(\eta_0 + \eta_1^{'} w_i\right)\right) \tag{37}$$

of (non-bottom-coded) observations being randomly censored to the same value ξ, with Φ the normal c.d.f.. To simplify the formulae that follow, denote the probability of bottom-coding as

$$P(w_i) \equiv \Phi\left(\frac{\xi - \left(\delta_0 + \delta_1^{'} w_i\right)}{\sigma_v}\right) \tag{38}$$

To compute the required regression formulae, note first that $d = 0$ if and only if $d_1 = 0$ and $d_2 = 0$. Therefore, by conditional independence,

$$\Pr\{d = 0 | w_i\} = [1 - P(w_i)][1 - R(w_i)] \tag{39}$$

so that the overall probability of censoring is

$$p\left(w_i\right) = \Pr\left\{d = 1 | w_i\right\} = P\left(w_i\right) + R\left(w_i\right) - P\left(w_i\right) \cdot R\left(w_i\right) \tag{40}$$

For the regression of x on w in the complete cases, we have

$$\begin{aligned} g_0\left(w_i\right) &= E\left(x_i | \ w_i, d = 0\right) \\ &= E(x_i \mid w_i, \ d_{1i} = 0 \text{ and } d_{2i} = 0) \\ &= \delta_0 + \delta_1^{'} w_i + E\left(v_i \mid w_i, \ v_i < \xi - \left(\delta_0 + \delta_1^{'} w_i\right) \text{ and } s_i < -\left(\eta_0 + \eta_1^{'} w_i\right)\right) \\ &= \delta_0 + \delta_1^{'} w_i + E\left(v_i \mid w_i, \ v_i < \xi - \left(\delta_0 + \delta_1^{'} w_i\right)\right) \end{aligned} \tag{41}$$

where the last equality follows from the conditional independence of v_i and s_i given w_i. Therefore, $g_0\left(\cdot\right)$ is given by the following formula (which is also appropriate for bottom-coding only)

$$g_0\left(w_i\right) = \delta_0 + \delta_1^{'} w_i + \sigma_v \cdot \lambda_0 \left(\frac{\xi - \left(\delta_0 + \delta_1^{'} w_i\right)}{\sigma_v}\right) \tag{42}$$

Here $\lambda_0\left(\cdot\right) \equiv \phi\left(\cdot\right) / \left[1 - \Phi\left(\cdot\right)\right]$, with ϕ the normal density function.

The regression of x_i on w_i for the censored data is found by applying (31) using (40) and (41). The result is

$$g_1\left(w_i\right) = \delta_0 + \delta_1^{'} w_i - \Psi\left(w_i\right) \cdot \sigma_v \cdot \lambda_1 \left(\frac{\xi - \left(\delta_0 + \delta_1^{'} w_i\right)}{\sigma_v}\right) \tag{43}$$

where $\lambda_1(\cdot) \equiv \phi(\cdot)/\Phi(\cdot)$.and

$$\Psi(w_i) = \frac{P(w_i)[1 - R(w_i)]}{R(w_i) + P(w_i)[1 - R(w_i)]} \tag{44}$$

The correction term Ψ is easily seen to be $\Psi(w_i) = (p(w_i) - R(w_i))/p(w_i)$, the relative probability of bottom-coding in the mixed censoring.

It is worth pointing out that all the parameters of the model are identified. In brief, the linear regression (1) applied to the complete cases identifies α, β, ϕ and σ_ε, and the normal truncated regression (41) applied to the complete cases identifies δ_0, δ_1 and σ_v. Finally, with δ_0, δ_1 and σ_v, the (scaled) probit model (39) applied to the full sample identifies η_0 and η_1. We could consider various estimation approaches using the moment restrictions implied by the various regressions above, but instead we derive the likelihood function for consistent and efficient estimation.

3.3. The Likelihood Function for the Normal Mixed Censoring Model

We derive the likelihood function for a slightly more general model than above, allowing for the possibility of separate instruments for the (uncensored) regressor. In brief, the model is

$$y_i = \beta x_i + W_i'\gamma + \varepsilon_i$$
$$x_i = Z_i'\theta + v_i$$
$$x_i^{cen} = 1(x_i \geq 0) \cdot 1\left(Z_i'\eta + s_i \geq 0\right) \cdot x_i$$

where W_i, Z_i need not coincide, and each may contain a constant. We assume the normal parametric specification

$$\begin{pmatrix} \varepsilon_i \\ v_i \\ s_i \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{bmatrix} \sigma_\varepsilon^2 & 0 & 0 \\ 0 & \sigma_v^2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right)$$

We use a censoring value of $\xi = 0$ in the following, without loss of generality.

We construct the likelihood function following Ruud (2000, Chapter 18), by first deriving the joint c.d.f. of (y, x^{cen}) conditional on W and Z, namely[24]

$$F(c_1, c_2) = \Pr\{y \leq c_1, x^{cen} \leq c_2\}. \tag{45}$$

We then derive the likelihood by differentiating with respect to c_1, c_2 where possible, and differencing where not. First, for the case where $c_2 < 0$, we have

$$F(c_1, c_2) = 0, \quad c_2 < 0 \tag{46}$$

For $c_2 > 0$, we have that

$$y \leq c_1 \;\Leftrightarrow\; \varepsilon \leq c_1 - \beta x^{cen} - W'\gamma$$

and

$$x^{cen} \leq c_2 \;\Leftrightarrow v \leq c_2 - Z'\theta$$

[24] We suppress the dependence of F on W and Z in the notation, which hopefully will not cause any confusion.

where this condition is sufficient regardless of the value of s. Therefore

$$F(c_1, c_2) = \Phi\left(\frac{c_1 - \beta x^{cen} - W'\gamma}{\sigma_\varepsilon}\right) \cdot \Phi\left(\frac{c_2 - Z'\theta}{\sigma_v}\right), \quad c_2 > 0 \tag{47}$$

The final case, $c_2 = 0$, requires some calculation. Begin by writing y in terms of the errors as

$$y = \beta Z'\theta + W'\gamma + \beta v + \varepsilon$$

Therefore, $F(c_1, 0) = \Pr\{y \leq c_1, x^{cen} \leq 0\}$ is the probability that

$$I: \qquad \beta v + \varepsilon \leq c_1 - \beta Z'\theta - W'\gamma$$

and that either

$$II: \qquad v \leq -Z'\theta \quad \text{or} \quad III: \quad s \leq -Z'\eta$$

holds. Denoting II' and III' as the opposite condition to II and III respectively, we have

$$\begin{aligned} F(c_1, 0) &= \Pr\{I \text{ and } (II \text{ or } III)\} \\ &= \Pr\{I\} - \Pr\left\{I \text{ and } II' \text{ and } III'\right\} \\ &= \Pr\{I\} - \Pr\left\{I \text{ and } II'\right\} \Pr\left\{III'\right\} \\ &= \Pr\{I\} - [\Pr(I) - \Pr\{I \text{ and } II\}] \Pr\left\{III'\right\} \\ &= \Pr\{I\} \Pr\{III\} + (1 - \Pr\{III\}) \Pr\{I \text{ and } II\} \end{aligned} \tag{48}$$

where the third equality is by independence of s and ε, v. Clearly, we have that

$$\Pr\{I\} = \Phi\left(\frac{c_1 - \beta Z'\theta - W'\gamma}{\sqrt{\beta^2\sigma_v^2 + \sigma_\varepsilon^2}}\right)$$

and

$$\Pr\{III\} = \Phi\left(-Z'\eta\right) = 1 - \Phi\left(Z'\eta\right)$$

We complete the ingredients of (48). by noting that

$$\Pr\{I \text{ and } II\} = \int_{-\infty}^{-Z'\theta}\int_{-\infty}^{c_1-\beta Z'\theta - W'\gamma} \phi_{biv}\left(\begin{pmatrix} \beta v + \varepsilon \\ v \end{pmatrix}, \Sigma\right) d\left(\beta v + \varepsilon\right) dv$$

where ϕ_{biv} is the bivariate normal density, with covariance matrix

$$\Sigma = \begin{bmatrix} \beta^2\sigma_v^2 + \sigma_\varepsilon^2 & \beta\sigma_v^2 \\ \beta\sigma_v^2 & \sigma_v^2 \end{bmatrix}$$

In summary

$$\begin{aligned} F(c_1, 0) &= \left(1 - \Phi\left(Z'\eta\right)\right)\cdot\Phi\left(\frac{c_1 - \beta Z'\theta - W'\gamma}{\sqrt{\beta^2\sigma_v^2 + \sigma_\varepsilon^2}}\right) + \\ &\quad \Phi\left(Z'\eta\right)\cdot\int_{-\infty}^{-Z'\theta}\int_{-\infty}^{c_1-\beta Z'\theta - W'\gamma} \phi_{biv}\left(\begin{pmatrix} \beta v + \varepsilon \\ v \end{pmatrix}, \Sigma\right) d\left(\beta v + \varepsilon\right) dv \end{aligned} \tag{49}$$

Now, to compute the components of the likelihood function, we differentiate/difference the c.d.f.. For $c_2 < 0$, we have that

$$\frac{\partial F(c_1, c_2)}{\partial c_1 \partial c_2} = 0 \tag{50}$$

For $c_2 > 0$, we have that

$$\frac{\partial F(c_1, c_2)}{\partial c_1 \partial c_2} = \frac{1}{\sigma_\varepsilon}\phi\left(\frac{c_1 - \beta x^{cen} - W'\gamma}{\sigma_\varepsilon}\right) \cdot \frac{1}{\sigma_v}\phi\left(\frac{c_2 - Z'\theta}{\sigma_v}\right) \tag{51}$$

For $c_2 = 0$, we differentiate w.r.t c_1 as

$$\begin{aligned}\frac{\partial F(c_1, 0)}{\partial c_1} &= \left(1 - \Phi\left(Z'\eta\right)\right) \cdot \frac{1}{\sqrt{\beta^2\sigma_v^2 + \sigma_\varepsilon^2}}\phi\left(\frac{c_1 - \beta Z'\theta - W'\gamma}{\sqrt{\beta^2\sigma_v^2 + \sigma_\varepsilon^2}}\right) + \\ &\quad \Phi\left(Z'\eta\right)\frac{\partial}{\partial c_1}\left(\int_{-\infty}^{-Z'\theta}\int_{-\infty}^{c_1 - \beta Z'\theta - W'\gamma}\phi_{biv}\left(\begin{pmatrix}\beta v + \varepsilon \\ v\end{pmatrix}; \Sigma\right) d(\beta v + \varepsilon)\, dv\right)\end{aligned}$$

The final derivative is solved for explicitly using the fact that if $u \equiv v - \rho(\beta v + \varepsilon)$ is independent of $\beta v + \varepsilon$, where

$$\rho = \frac{\beta\sigma_v^2}{\beta^2\sigma_v^2 + \sigma_\varepsilon^2},$$

and that the variance of u is

$$\sigma_u^2 = \sigma_v^2\left(1 - \frac{\beta^2\sigma_v^2}{\beta^2\sigma_v^2 + \sigma_\varepsilon^2}\right)$$

Now, we have

$$
\begin{aligned}
& \frac{\partial}{\partial c_1}\left(\int_{-\infty}^{-Z'\theta}\int_{-\infty}^{c_1-\beta Z'\theta-W'\gamma}\phi_{biv}\left(\begin{pmatrix}\beta v+\varepsilon\\ v\end{pmatrix};\Sigma\right)d\left(\beta v+\varepsilon\right)dv\right)\\
= & \int_{-\infty}^{-Z'\theta}\phi_{biv}\left(\begin{pmatrix}c_1-\beta Z'\theta-W'\gamma\\ v\end{pmatrix};\Sigma\right)dv\\
= & \phi\left(c_1-\beta Z'\theta-W'\gamma;\beta^2\sigma_v^2+\sigma_\varepsilon^2\right)\cdot\int_{-\infty}^{-Z'\theta}\phi\left(v-\rho\left(c_1-\beta Z'\theta-W'\gamma\right);\sigma_u^2\right)dv\\
= & \frac{1}{\sqrt{\beta^2\sigma_v^2+\sigma_\varepsilon^2}}\phi\left(\frac{c_1-\beta Z'\theta-W'\gamma}{\sqrt{\beta^2\sigma_v^2+\sigma_\varepsilon^2}}\right)\cdot\Phi\left(\frac{-Z'\theta-\rho\left(c_1-\beta Z'\theta-W'\gamma\right)}{\sigma_u}\right)\\
= & \frac{1}{\sqrt{\beta^2\sigma_v^2+\sigma_\varepsilon^2}}\phi\left(\frac{c_1-\beta Z'\theta-W'\gamma}{\sqrt{\beta^2\sigma_v^2+\sigma_\varepsilon^2}}\right)\cdot\\
& \Phi\left(-\frac{\sqrt{\beta^2\sigma_v^2+\sigma_\varepsilon^2}}{\sigma_v\sigma_\varepsilon}Z'\theta-\frac{\beta\sigma_v}{\sqrt{\beta^2\sigma_v^2+\sigma_\varepsilon^2}\sigma_\varepsilon}\left(c_1-\beta Z'\theta-W'\gamma\right)\right)
\end{aligned}
$$

In summary, we have

$$
\begin{aligned}
\frac{\partial F\left(c_1,0\right)}{\partial c_1} = & \frac{1}{\sqrt{\beta^2\sigma_v^2+\sigma_\varepsilon^2}}\phi\left(\frac{c_1-\beta Z'\theta-W'\gamma}{\sqrt{\beta^2\sigma_v^2+\sigma_\varepsilon^2}}\right)\cdot \\
& \left(1+\Phi\left(Z'\eta\right)\left[\Phi\left(-\frac{\sqrt{\beta^2\sigma_v^2+\sigma_\varepsilon^2}}{\sigma_v\sigma_\varepsilon}Z'\theta-\frac{\beta\sigma_v}{\sqrt{\beta^2\sigma_v^2+\sigma_\varepsilon^2}\sigma_\varepsilon}\left(c_1-\beta Z'\theta-W'\gamma\right)\right)-1\right]\right)
\end{aligned}
\tag{52}
$$

These calculations allow us to write the log-likelihood function directly. Recall that $d_i =$

$1\left[x_i^{cen}=0\right]$ indicates an observation with a censored regressor. We have

$$\begin{aligned}\ln\mathcal{L} &= C+\sum_{i=1}^{n}(1-d_i)\left(-\ln\sigma_\varepsilon-\frac{1}{2}\frac{\left(y_i-\beta x_i^{cen}-W_i^{'}\gamma\right)^2}{\sigma_\varepsilon^2}\right) \\ &+\sum_{i=1}^{n}(1-d_i)\left(-\ln\sigma_v-\frac{1}{2}\frac{\left(x_i^{cen}-Z_i^{'}\theta\right)^2}{\sigma_v^2}\right) \\ &+\sum_{i=1}^{n}d_i\left(-\frac{1}{2}\ln\left(\beta^2\sigma_v^2+\sigma_\varepsilon^2\right)-\frac{1}{2}\frac{\left(y_i-\beta Z_i^{'}\theta-W_i^{'}\gamma\right)^2}{\beta^2\sigma_v^2+\sigma_\varepsilon^2}\right) \\ &+\sum_{i=1}^{n}d_i\ln\left(1+\Phi\left(Z_i^{'}\eta\right)\left[\Phi\left(-\frac{\sqrt{\beta^2\sigma_v^2+\sigma_\varepsilon^2}}{\sigma_v\sigma_\varepsilon}Z_i^{'}\theta-\frac{\beta\sigma_v}{\sqrt{\beta^2\sigma_v^2+\sigma_\varepsilon^2}\sigma_\varepsilon}\left(y_i-\beta Z_i^{'}\theta-W_i^{'}\gamma\right)\right)-1\right]\right)\end{aligned} \tag{53}$$

The terms are easy to interpret; the first three are normal log-likelihoods for regressing y on x^{cen} and W in the complete cases, for regressing x^{cen} on Z in the complete cases, and for regressing y on $Z^{'}\theta$ and W in the censored data, respectively. This final term corrects for selection on y induced by the censoring of x. As such, this log-likelihood has natural similarity to the log-likelihood for normal selection models.

It is not difficult to establish the conditions for consistency and asymptotic normality of maximum likelihood, as laid out in Newey and McFadden (1994). For consistency – Newey and McFadden Theorem 2.5 – we create a compact parameter space by bounding all parameters. We assume variances have a small positive lower bound and a large upper bound, and other parameters have (large) negative lower bounds and positive upper bounds. Continuity is apparent, and the bounding condition is clear for all four terms above (for instance, the last term is bounded above by $\ln(1)=0$). For asymptotic normality – Newey and McFadden Theorem 3.3 – the log-likelihood is clearly twice continuously differentiable, and the remaining regularity conditions follow for the first three terms from standard normal linear regression and for the last term from the linear forms within the normal c.d.f.(as with a probit model).

It is worth remarking that the authors have failed to discover general conditions under which this log-likelihood displays gobal concavity. However, since the first three components are very well behaved (and globally concave themselves), it is natural to suspect that some situations exist where overall global concavity can be shown. Then, maximum likelihood estimation would be as well behaved as for some other censoring problems, such as a normal regression model with a censored dependent variable.

4. The Effects of Wealth on Consumption

4.1. General Discussion

In recent years, many developed countries have witnessed tremendous expansion in consumption expenditures at the same time as substantial increases in household wealth levels.[25] This has fueled great interest in the measurement of the effects of wealth on consumption decisions.

One encounters many types of censoring when studying consumption at the household level. Income is typically top-coded, by survey design. Wealth is nonnegative, in part because of survey bounding but more because of a failure to observe negative wealth components such as household debt. Thus, our analysis of log wealth as censored may be incomplete, as we will take positive wealth observations as correct. That is, there may be further mismeasurement issues applicable to our 'complete cases.'

Published estimates of the elasticity of consumption with regard to financial wealth seem unusually large. With aggregate data, estimates in the range of 4% but up to 10% can be

[25] During the 1990's there were multiyear expansions in consumption in the US and the UK (among others). During the same time, the total wealth of Americans grew more than 15 trillion dollars, with a 262% increase in corporate equity and a 14% increase in housing and other tangible assets (see Poterba (2000) for an excellent survey). Housing prices increased in both countries as well.

found, varying with the type of asset included and the time period under consideration.[26] With individual data, estimates tend to be larger,[27] such as 8%. We are interested in whether the censored character of income and wealth can help account for the magnitude of these estimates?[28]

It is worth mentioning that estimates of wealth effects are of substantial interest to economic policy. A key issue of monetary policy is how much aggregate demand is affected by changes in interest rates. Interest rates affect consumption directly, but also housing wealth as well as financial wealth. A substantial impact of wealth on consumption, either through enhanced borrowing or cashing out of capital gains, will be a big part of whether interest rates have a real impact or not, and thus are relevant for the design of effective monetary policy.[29]

4.2. Application to Consumption Data

We now study the impact of censored regressors in an application to household consumption and wealth.[30] The data includes consumption, current income and a computed permanent component of consumption that depends on the cohort in which the household belongs, characteristics of the household (such as retirement status, family size, etc.), and financial information. By construction, the income variables are *not* censored – the observations with top-coded income variables of the original survey have been dropped. That is, our data is already a set of 'com-

[26] Laurence Meyer and Associates (1994) find an elasticity of 4.2 percent, Brayton and Tinsley (1996) find 3 percent, Ludvigson and Steindel (1999) estimate an overall elasticity of 4 percent (as well as some estimates as high as 10 percent).

[27] See Parker (1999), Juster, Lupton, Smith and Stafford (1999) and Starr-McCluer (1999).

[28] Similarly, large effects of housing wealth on consumption are estimated by Aoki, et. al. (2002a, 2002b) and Attanasio, et. al. (1994), among others. Somewhat smaller estimates are given in Engelhardt (1996) and Skinner (1996).

[29] See Muellbauer and Murphy (1990), King (1990), Pagano (1990) Attanasio and Weber (1994) and Attanasio et. al. (2003), for various arguments on the connection between consumption and housing prices. In terms of whether assets prices should be targeted as part of monetary policy, see Bernanke and Gertler (1999,2001), Cecchetti et. al. (2000) and Rigobon and Sack (2003).

[30] We thank Jonathan Parker for his tremendous help and support in providing us not only with the data but with valuable suggestions.

plete cases' in terms of income. The only censored variables are the financial wealth variables. There are three sorts of wealth variables that interest us: total wealth, housing wealth, and stock market wealth.

In Table 2 we show the proportion of the variables that are at the censoring bound in the data. There is a moderate proportion of total wealth observations at the bound (27 %) but this increases to 43 % for housing, to 76 % for stock market wealth, and to 81% when one or more wealth variables is at the bound.[31] This raises substantial concerns for the estimation of consumption impacts from different types of wealth.[32]

To see a coarse impact of censoring, Table 3 gives estimates of linear regressions of log consumption on log wealth and wealth components, without any additional regressors.[33] If only total wealth is included, there are 8,735 complete (uncensored) cases, and when all three wealth variables are included, there are 2,272 complete cases. With the bivariate regression of log consumption on log wealth, there is an expansion bias of 29%, namely (.181/.140) - 1. With the components included, using all data gives a total wealth elasticity of .202, whereas the complete cases give a total elasticity of .136, which reflects a 48% expansion bias. There are some huge relative shifts; in particular a much larger housing wealth effect in the complete case data.

To apply our model of wealth censoring, we focus on the total wealth effect, using a log-form

[31] For consistency among the components, total wealth is censored when it is less than $5,000, housing wealth when it is less than $4,000 and stock market wealth when it is less than $1,000. This gives slightly higher censoring than when all levels are censored at zero, but facilitates taking logarithms.

[32] An issue we have not highlighted here is that for some of the wealth observations, a value at the bound may be the correct wealth value. That is, zero wealth may be zero wealth, as opposed to a censored nonzero wealth. In our estimates, this is partly accomodated for by allowing for random censoring. But a more full treatment, this possibility could be more fully modelled. We did carry out the test of Rigobon and Stoker (2005b), and rejected that censoring bias was zero.

[33] Heteroskedasticity consistent (White) standard errors are presented in parentheses.

regression equation similar to that estimated by Parker (1999).

$$\ln C_{it} = \alpha + \beta \cdot \ln W_{it} + \phi_1 \ln PINC_{it} + \phi_2 \ln INC_{it} + \phi_3^{'} Controls_{it} + \varepsilon_{it} \tag{54}$$

where $C_{i,t}$ is consumption of household i at time t. W_{it} is total wealth, which is censored.[34] There are two income variables; $PINC_{it}$ is a constructed permanent component of income and $INC_{i,t}$ is the current income, which are uncensored regressors in our data. These are uncensored regressors in our data. $Controls_{it}$ are variables accounting for retirement status, family size, cohorts, time, etc. For a detailed description of the data and the definition of the variables, see Parker (1999).

Table 4 presents our estimation results. The first two columns give OLS estimates of wealth and income effects from estimating (54) over the full data and over the complete cases. The third and fourth columns give maximum likelihood estimates of the model with bound censoring and independent random censoring. The third column has only the income variables as additional regressors, setting $\phi_3 = 0$ in (54), and including only the income variables in the equations for wealth bound censoring and for independent random censoring. Finally, the fourth column gives maximum likelihood estimates where all controls are included in (54) and in the equations for bound censoring and for random censoring.

With all controls, there is not a great deal of difference between the OLS estimates for the full sample, and those for the complete cases. The maximum likelihood estimates display a larger wealth elasticity than the OLS estimates (roughly 14%). Moreover, the effects of the income variables have much smaller standard errors. The larger wealth elasticity is a bit surprising, but since we had many controls and two types of censoring, it was not clear what type of impact one

[34] To include housing and stock market wealth, we would need to model the joint censoring process of all wealth components. We focus on total wealth only just to keep things simple here.

should expect. Some lowering of the standard errors was expected, since we are now including all of the censored data into the estimation in a consistent fashion.

We did encounter one problem in estimation, that did not seem to impact the estimates presented in Table 4. We estimated a very small independent probability of censoring beyond the bound censoring of wealth. The coefficient estimates for this probability were very imprecise, which makes sense since they appear in the likelihood in the tail of the normal c.d.f. As such, we checked for robustness of the main wealth and income effects by setting different values of the independent censoring probability; this exercise uncovered no substantial differences in the main estimates. In any event, this aspect of the estimation merits further study.[35]

5. Conclusion

The fact that censoring of regressors can routinely generate expansion bias was a surprise to both authors. We noticed the phenomena for bound censoring in some simulations, and were able to understand the source pretty easily. In fact, it is a straightforward point, as Figure 1 can be explained to students with only rudimentary knowledge of econometric methods. Nevertheless, we don't feel that it is a minor problem for practical applications. Quite the contrary, we feel that problems of censored regressors are likely as prevalent or more prevalent than problems of censored dependent variables in typical econometric applications. Some evidence of this is the development of the faulty empirical practice of including a dummy variable for the censored data, or that of replacing censored values with imputed tail means.

We feel we have made some progress in understanding the estimation issues posed by censored regressors. The use of a censoring dummy as a "fix" for censoring bias is not advisable, and

[35] All results and estimation details are available from the authors.

even the use of tail imputations is only advisable for bound censoring with very simple models. By establishing that there is zero semiparametric information in the censored observations, we have verified that there is no fully nonparametric "fix" for the censoring or the regressor, and that some additional structure (or side information from another data set, etc.) is required. This is true even in the simplest case of exogenous censoring, that has been our focus. We did not address whether there is partial identification from the censored data; for instance, whether top-coded data provides some additional bound information on the parameters. This would be a useful future direction to pursue.

We have illustrated the extent of censoring in an application to household consumption and wealth. We developed a normal parametric model as well as its likelihood function for estimation. Our maximum likelihood estimates had a larger wealth elasticity than OLS estimates, with greater precision of the income effects, because of using the censored data in a consistent fashion with the uncensored data. We found that it was difficult to estimate the exact structure of the normal censoring processes, although that didn't have a strong impact on our estimates of the wealth and income effects on consumption. While this conclusion is dependent on our specific model, we are very optimistic that semiparametric procedures can be developed in future research.

Censoring bias of a similar type arises in instrumental variables estimators when there are censored regressors, although there are some important differences with the case of OLS regression.[36] We cover some results of this kind in Rigobon and Stoker (2006a). Moreover, we have developed specification tests for the presence of censoring bias in Rigobon and Stoker (2006b), which would serve as a useful precursor to a discussion of how to incorporate the censored data

[36] For instance, IV estimators with random censoring of an endogenous variable display expansion bias, whereas OLS estimators display attenuation bias.

in estimation. In any case, our goal is to develop a sufficient set of empirical tools for a researcher to check for bias problems from censored regressors, and then appropriately estimate parameters using all available data.

References

[1] Ai, Chunrong (1997), "An Improved Estimator for Models with Randomly Missing Data," *Nonparametric Statistics*, 7, 331-347.

[2] Aoki, K., J. Proudman and G. Vlieghe (2002a) "House prices, consumption, and monetary policy: a financial accelerator approach" *Bank of England Quarterly Bulletin.*

[3] Aoki, K., J. Proudman and G.Vlieghe (2002b) "Houses as collateral: has the link between house prices and consumption in the UK changed?', *Economic Policy* Review Vol. 8 (1), Federal Reserve Bank of New York,

[4] 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, 537-559.

[5] Attanasio, O., L. Blow, R. Hamilton, and A. Leicester (2003) "Consumption, House Prices, and Expectations" Institute for Fiscal Studies, Mimeo.

[6] Attanasio, O., and G. Weber (1994) "The UK Consumption Boom of the late 1980s: aggregate implications of microeconomic evidence" in *The Economic Journal*, Vol. 104, Issue 427, November, pp 1269-1302.

[7] Belsley, D.A., E. Kuh and R. E. Welsch (1980), *Regression Diagnostics: Identifying Influential Data and Sources of Multicollinearity*, Wiley, New York.

[8] Bernanke, B., and M. Gertler (1999), "Monetary Policy and Asset Price Volatility," *Federal Reserve Bank of Kansas City Economic Review*, LXXXIV, 17–51.

[9] Bernanke, B., and M. Gertler(2001), "Should Central Banks Respond toMovements in Asset Prices?" *American Economic Review Papers and Proceedings*, XCI , 253–257.

[10] Borjas, G. (1994) "Long-Run Convergence of Ethnic Skill Differentials" NBER 4641.

[11] Brayton, F. and P. Tinsley. (1996). "A Guide to the FRB/US: A Macroeconomic Model of the United States.' Federal Reserve Board of Governors, Washington DC, Working Paper 1996-42.

[12] Card, D., J. DiNardo, and E. Estes (1998) "The More Things Change: Immigrants and the Children of Immigrants in the 1940s, the 1970s, and the 1990s" NBER 6519

[13] Cecchetti, S. G., H. Genberg, J. Lipsky, and S. Wadhwani, *Asset Prices and Central Bank Policy* (London: International Center for Monetary and Banking Studies, 2000).

[14] Chen, X., H. Hong and E. Tamer (2005), "Measurement Error Models with Auxiliary Data," *Review of Economic Studies*, 72, 343-366.

[15] Chen, X., H. Hong and A. Tarozzi (2004), "Semiparametric Efficiency in GMM Models of Nonclassical Measurement Errors, Missing Data and Treatment Effects," Working Paper, November.

[16] Chernozhukov, V., H. Hong and E. Tamer (2004), "Inference on Parameter Sets in Econometric Models," working paper, Duke University.

[17] Davidson, R. and J. D. McKinnon (2004), *Econometric Theory and Methods*, Oxford University Press, New York.

[18] Engelhardt, G. (1996). “House Prices and Home Owner Saving Behavior," *Regional Science and Urban Economics,* 26, pp. 313–36.

[19] Green, W.H. (2003). *Econometric Analysis*, 5th ed. New Jersey: Prentice Hall.

[20] Heitjan, D. F. and D. B. Rubin (1990), "Inference from Coarse Data Via Multiple Imputation With Application to Age Heaping," *Journal of the American Statistical Association*, 85, 304-314.

[21] Heitjan, D. F. and D. B. Rubin (1991), "Ignorability and Coarse Data," *Annals of Statistics*, 19, 2244-2253.

[22] Horowitz, J. and C. F. Manski (1998), "Censoring of Outcomes and Regressors Due to Survey Nonresponse: Identification and Estimation Using Weights and Imputations," *Journal of Econometrics*, 84, 37-58.

[23] Horowitz, J. and C. F. Manski (2000), "Nonparametric Analysis of Randomized Experiments with Missing Covariate and Outcome Data," *Journal of the American Statistical Association*, 95, 77-84.

[24] Ichimura, H. and E. Martinez-Sanchis (2005), "Identification and Estimation of GMM Models by Combining Two Data Sets," CEMMAP Working Paper, IFS, London, March.

[25] Juster, F. T., Joseph L., J. P. Smith, and F. Stafford. (1999). “Savings and Wealth: Then and Now.’ Mimeo, University of Michigan, Institute for Survey Research..

[26] King, M. (1990) “Discussion” in *Economic Policy*, Vol. 11, pp 383-387.

[27] Lawrence H. Meyer and Associates. (1994). *The WUMM Model Book*. St. Louis: L. H. Meyer and Associates.

[28] Liang, H, S. Wang, J.M. Robins and R.J. Carroll (2004), "Estimation in Partially Linear Models with Missing Covariates," *Journal of the American Statistical Association*, 99, 357-367.

[29] Little, R. J. A. (1992), "Regression with Missing X's: A Review," *Journal of the American Statistical Association*, 87, 1227-1237.

[30] Little, R. J. A. and D. B. Rubin (2002), *Statistical Analysis with Missing Data*, *2nd edition*, John Wiley and Sons, Hoboken, New Jersey.

[31] Ludvigson, S. and C. Steindel. (1999). "How Important is the Stock Market Effect on Consumption?' *Federal Reserve Bank of New York Economic Policy Review*. July, 5:2, pp. 29–52.

[32] Mahajan, A. (2004), "Identification and Estimation of Single Index Models with Misclassified Regressors," Working Paper, Stanford University, July.

[33] Manski, C.F. and E. Tamer (2002), "Inference on Regressions with Interval Data on a Regressor or Outcome," *Econometrica*, 70, 519-546.

[34] Muellbauer, J. and A. Murphy (1990) "Is the UK balance of payments sustainable?" in *Economic Policy*, Vol. 11, pp 345-383.

[35] Newey, W. K. (1990), "Semiparametric Efficiency Bounds," *Journal of Applied Econometrics*, 5, 99-135.

[36] Newey, W.K. and D. L. McFadden (1994), "Large Sample Estimation and Hypothesis Testing," Chapter 36 in R.F. Engle and D.L. McFadden, eds., *Handbook of Econometrics, Volume 4*, Elsevier, Amsterdam.

[37] Newey, W.K. and T. M. Stoker (1993), "Efficiency of Weighted Average Derivative Estimators and Index Models," *Econometrica*, 61, 1199-1223.

[38] Pagano C.(1990) "Discussion" in *Economic Policy*, Vol. 11, pp 387-390.

[39] Parker, J. (1999). "Spendthrift in America? On Two Decades of Decline in the U.S. Saving Rate?' in *NBER Macroeconomics Annual 1999*. B. Bernanke and J. Rotemberg, eds. Cambridge: MIT Press.

[40] Petoussis, K., R. D. Gill and C. Zeelenberg (2004), "Statistical Analysis of Heaped Duration Data," draft, Department of Psychology, Vrije Universiteit Amsterdam, February.

[41] Poterba, J. M. (2000) "Stock Market Wealth and Consumption," *Journal of Economic Perspectives,* Volume 14, Number 2, Spring 2000, pp. 99–118.

[42] Ridder, G. and R. Moffit (2003), "The Econometrics of Data Combination," chapter for *Handbook of Econometrics, Volume 6*, forthcoming.

[43] Rigobon, R and T. M. Stoker (2006a) "Bias from Censored Regressors,", MIT Working Paper, revised September.

[44] Rigobon, R and T. M. Stoker (2006b) "Testing for Bias from Censored Regressors", MIT Working Paper, revised February.

[45] Robins, J. M. and A. Rotnitzky (1995), "Semiparametric Efficiency in Multivariate Regression Models with Missing Data," *Journal of the American Statistical Association*, 90, 122-129.

[46] Robinson, P.M. (1988). "Root-N-Consistent Semiparametric Regression," *Econometrica*, 56, 931-954.

[47] Rotnitzky, A., J. M. Robins and D. O. Scharfstein (1998), "Semiparametric Regression for Repeated Outcomes With Nonignorable Response," *Journal of the American Statistical Association*, 93, 1321-1339.

[48] Ruud, P.A. (2000), *An Introduction to Classical Econometric Theory*, Oxford University Press, Oxford.

[49] Schmalensee, R. and T. M. Stoker (1999), "Household Gasoline Demand in the United States," *Econometrica*, 67, 645-662.

[50] Shaik, A. (2005), "Inference for Partially Identified Econometric Models," working paper, Stanford University.

[51] Skinner, J. (1996). "Is Housing Wealth a Sideshow?' in *Advances in the Economics of Aging*. D. Wise, ed. Chicago: University of Chicago Press, pp. 241–68.

[52] Starr-McCluer, M. (1999). "Stock Market Wealth and Consumer Spending.' Mimeo, Federal Reserve Board of Governors.

[53] Tripathi, G. (2003), "GMM and Empirical Likelihood with Imcomplete Data," Working Paper, December.

[54] Tripathi, G. (2004), "Moment Based Inference with Incomplete Data," Working Paper, June.

[55] Torelli, N. and U. Trivellato (1993). "Modeling Inaccuracies in Job-Search Duration Data," *Journal of Econometrics*, 59, 185-211.

[56] Yatchew, A. (2003). *Semiparametric Regression for the Applied Econometrician*, Cambridge: Cambridge University Press.

		Bivariate	One Additional Regressor			
	Procedure		Correlation .5		Correlation .9	
Truncation		Eff $\hat{\beta}$	Eff $\hat{\beta}$	Eff ϕ	Eff $\hat{\beta}$	Eff ϕ
20%	CC	47 %	52 %	80 %	71 %	80 %
	Known Mean	88 %	62 %	86 %	76 %	83 %
40%	CC	25 %	30 %	60 %	48 %	60 %
	Known Mean	78 %	46 %	71 %	58 %	65 %
60%	CC	13 %	15 %	40 %	28 %	40 %
	Known Mean	66 %	36 %	56 %	42 %	45 %

Table 1: Efficiency Relative to Complete Sample

	Percentage Censored	Total Observations	Not Censored
Total Wealth	26.6%	11,903	8,735
Housing	43.4%	11,903	6,737
Stock Market	76.5%	11,903	2,797
One or More Censored	80.9%	11,903	2,272

Table 2: Proportion of Censoring in Total Wealth, Housing, and Stock Market Wealth

	OLS			
	All Data	CC	All Data	CC
Sample Size	11,903	8,735	11,903	2,272
Total Wealth	0.181	0.140	0.149	0.055
	(.0029)	(.0038)	(.0054)	(.0135)
Housing Wealth			0.020	0.069
			(.0053)	(.0132)
Stock Market Wealth			0.033	0.012
			(.0032)	(.0069)

Table 3: Log Consumption Results, Simple Models

	OLS		Maximum Likelihood with Censoring	
	All Data	CC	No Controls.	All Controls.
Sample Size	11,903	8,735	11,903	11,903
Total Wealth	0.052	0.054	0.064	0.062
	(.0045)	(.0062)	(.0035)	(.0040)
Current Income	0.180	0.165	0.202	0.179
	(.0117)	(.0137)	(.0049)	(.0048)
Permanent Income	0.175	0.177	0.214	0.180
	(.0160)	(.0208)	(.0065)	(.0068)

Table 4: Log Consumption Results

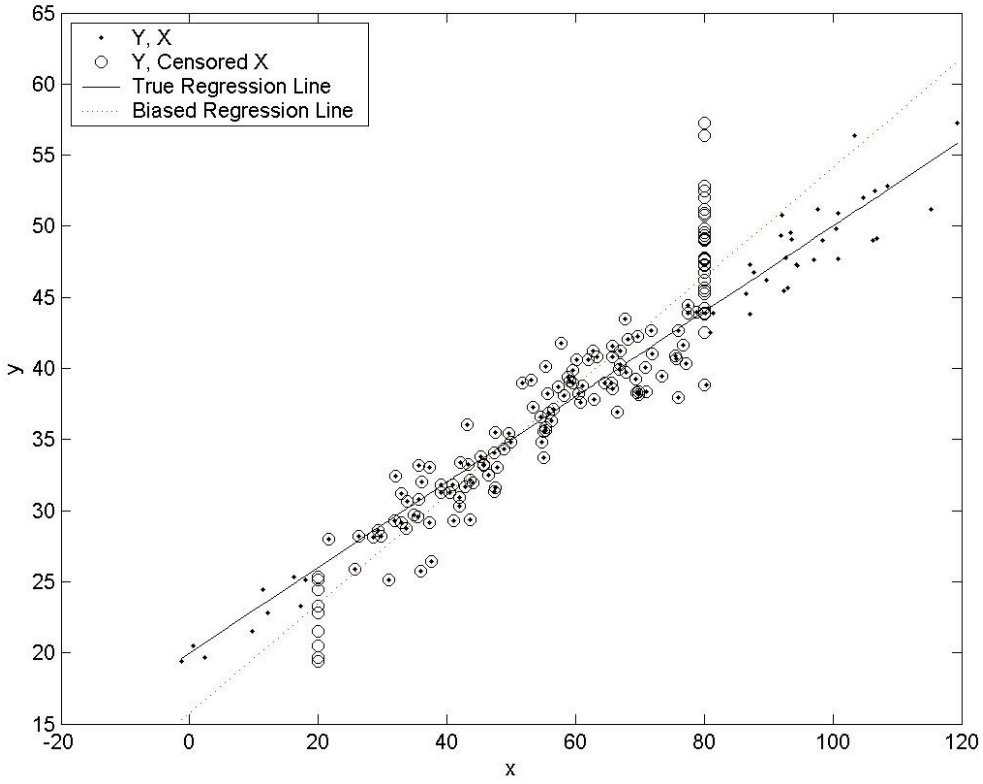

Figure 1: Expansion Bias