Estimation of Panel Data Regression Models  
with Two-Sided Censoring or Truncation*  

Sule Alan†  Bo E. Honoré‡  Luojia Hu §  Søren Leth–Petersen ¶  

November 6, 2012  

Abstract  
This paper constructs estimators for panel data regression models with individual specific heterogeneity and two–sided censoring and truncation. Following Powell (1986) the estimation strategy is based on moment conditions constructed from re–censored or re–truncated residuals. While these moment conditions do not identify the parameter of interest, they can be used to motivate objective functions that do. We apply one of the estimators to study the effect of a Danish tax reform on household portfolio choice. The idea behind the estimators can also be used in a cross sectional setting.  
Key Words: Panel Data, Censored Regression, Truncated Regression.  
JEL Code: C20, C23, C24.  

*This research was supported by NSF Grant No. SES-0417895 to Princeton University, the Gregory C. Chow Econometric Research Program at Princeton University, and the Danish National Research Foundation, through CAM at the University of Copenhagen (Honoré) and the Danish Social Science Research Council (Leth–Petersen). We thank Christian Scheuer and numerous seminar participants for helpful comments. The opinions expressed here are those of the authors and not necessarily those of the Federal Reserve Bank of Chicago or the Federal Reserve System.  
†Faculty of Economics, University of Cambridge, Sidgwick Avenue, Cambridge, UK, CB3 9DD. Email: sule.alan@econ.cam.ac.uk.  
‡Department of Economics, Princeton University, Princeton, NJ 08544-1021. Email: honore@Princeton.edu.  
§Economic Research Department, Federal Reserve Bank of Chicago, 230 S. La Salle Street, Chicago, IL 60604. Email: lhu@frbchi.org.  
¶Department of Economics, University of Copenhagen, Øster Farimagsgade 5, Building 26, DK-1353 Copenhagen K. and SFI, The Danish National Centre for Social Research, Herluf Trolles Gade 11, DK-1052. Email : soren.leth-petersen@econ.ku.dk.
1 Introduction

This paper generalizes a class of estimators for truncated and censored regression models to allow for two–sided truncation or censoring. The class of estimators is based on pairwise comparisons and the proposed generalizations therefore apply to both panel data and cross sectional data.

Two–sided censored regression models emerge in a number of contexts, but they are especially useful when the dependent variable is a fraction. For example, Alan and Leth–Petersen (2006) estimate a portfolio share equation where the portfolio shares are between 0 and 1, with a significant number of observations on either of the limits. Other recent applications in economics of regression models with two–sided censoring include Lafontaine (1993), Petersen and Rajan (1994), Petersen and Rajan (1995), Houston and Ryngaert (1997), Fehr, Kirchler, Weichbold, and Gachter (1998), Huang and Hauser (1998), McMillan and Woodruff (1999), de Figueiredo and Tiller (2001), Huang and Hauser (2001), Fenn and Liang (2001), Poterba and Samwick (2002), Nickerson and Silverman (2003), Officer (2004), Charness, Frechette, and Kagel (2004), Andrews, Schank, and Simmons (2005) and Gifford and Bernard (2005). We formulate the two–sided censored regression model as observations of \((y; x; L; U)\) from the model

\[
y^* = x' \beta + \varepsilon
\]  

where \(y^*\) is unobserved, but we observe

\[
y = \begin{cases} 
L & \text{if } y^* < L \\
y^* & \text{if } L \leq y^* \leq U \\
U & \text{if } y^* > U 
\end{cases}
\]

and \(\beta\) is the parameter of interest. When \(y\) is a share, \(L\) and \(U\) will typically be 0 and 1, respectively.

We will focus on panel data settings so the observations are indexed by \(i\) and \(t\) where \(i = 1, \ldots, n\) and \(t = 1, \ldots, T_i\). This allows for unbalanced panels, but we will maintain the restrictive assumption that \(T_i\) is exogenous in the sense that it satisfies all the assumptions made on the explanatory variables. In a panel data setting, it is also important to allow for individual specific effects in the errors \(\varepsilon_{it}\). We will do this implicitly by making assumptions of the type that \(\varepsilon_{it}\) is stationary conditional on \((x_{i1} \ldots x_{iT_i})\) or that \(\varepsilon_{it}\) and \(\varepsilon_{is}\) are independent and identically distributed conditional on some unobserved component \(\nu_{ist}\). These will have the textbook specification \(\varepsilon_{it} = \nu_i + \epsilon_{it} \) \(\{\varepsilon_{it}\} \ i.i.d.\) as a special case. In section 4.3, we discuss how to apply the same ideas to construct estimators of the cross sectional version of the model.
For the two–sided truncated regression model, we assume observations of \((y, x, L, U)\) from the distribution of \((y^*, x, L, U)\) conditional on \(L \leq y^* \leq U\). Two–sided truncated regression models are less common than two–sided censored regression models, but they play a role in duration models. Suppose, for example, that one wants to study the effect of early life circumstances on longevity by linking the Social Security Administration’s Death Master File to the 1900–1930 U.S. censuses\(^1\). This will miss a substantial number of deaths: (1) individuals who died under age 65 (since they were less likely to be collecting social security benefits and thus their deaths were less likely to be captured); (2) individuals who died before the year 1965 (the beginning of the computerized Social Security files); and (3) individuals who died after the last year the data is available. This is a case of two-sided truncation because we observe an individual only if the dependent variable, age at death, is greater than 65 years and if the death occurs between 1965 and the last year the data is available. This is essentially the empirical setting in Ferrie and Rolf (2011), although they only consider data from the 1900 census, so right–truncation is unlikely to be an issue in their application.

Honoré (1992) constructed moment conditions for similar panel data models with one–sided truncation or censoring and showed how they can be interpreted as the first–order conditions for a population minimization problem that uniquely identifies the parameter vector, \(\beta\). This paper generalizes that approach to the case when the truncation or censoring is two–sided. The main contribution of the paper is to show that some of those moment conditions can be turned into a minimization problem that actually uniquely identifies \(\beta\) when there is two–sided censoring or truncation. This is an important step because the moment conditions that we derive do not identify the parameters of the model. This is a generic problem with constructing estimators based on moment conditions. For example, Powell (1986) constructed moment conditions for a related cross sectional truncated and censored regression models based on symmetry of the error distribution. He also pointed out that while these moment conditions did not identify the parameter of interest, minimization of an objective function based on them did lead to identification.

The rest of the paper is organized as follows. Section 2 derives the moment conditions and the associated objective function for models with two–sided censoring and two–sided truncation. Section 3 then discusses how these can be used to estimate the parameters of interest. Generalizations of the two models considered in Section 2 are discussed in Section 4. We present the empirical application in Section 5 and Section 6 concludes.

\(^1\)1930 is the latest year for which this linkage is feasible.
2 Identification: Moment Conditions and Objective Functions

The challenge in constructing moment conditions in models with censoring and truncation is that one typically starts with assumptions on $\varepsilon$ conditional on $x$ in (1). If one had a random sample of $(y^*, x)$ then these assumptions could be used immediately to construct moment conditions. For example if $E[\varepsilon x] = 0$, then one has the moment conditions $E[(y^* - x'\beta) x] = 0$. However with truncation or censoring, $y - x'\beta$ will not have the same properties as $\varepsilon$. The idea employed in Powell (1986), Honoré (1992) and Honoré and Powell (1994) is to apply additional censoring and truncation to $y - x'\beta$ in such a way that the the resulting re-censored or re-truncated residual satisfies the conditions assumed on $\varepsilon$. For example, Powell (1986) assumed that $\varepsilon$ is symmetric conditional on $x$ in a censored regression model with censoring from below at 0. If that is the case, then $y - x'\beta = \max \{\varepsilon, -x'\beta\}$ will clearly not be symmetric conditional on $x$, but the re-censored residuals, $\min \{y - x'\beta, x'\beta\}$ will be. This implies moment conditions of the type $E[\min \{y - x'\beta, x'\beta\} x] = 0$. Unfortunately, this moment condition will not in general identify $\beta$, but Powell (1986) was able to prove that the integral (as a function of $b$) of $E[\min \{y - x'b, x'b\} x]$ is uniquely minimized at $b = \beta$ under appropriate regularity assumptions. Honoré (1992) applied the same insight to a panel setting where the assumption was that $\varepsilon$ is stationary conditional on the entire sequence of explanatory variables. Again, censoring or truncation destroys this stationarity, but it can be restored for a pair of residuals by additional censoring. Honoré and Powell (1994) then applied the same idea to any pair of observations in a cross section, and Hu (2002) generalized it to allow for lagged latent dependent variables as covariates.

The contribution of this paper is to generalize the approach in Honoré (1992) to the case with two-sided censoring or truncation. As in Powell (1986), Honoré (1992) and Honoré and Powell (1994), it is straightforward to construct moment conditions based on some re-censored or re-truncated residuals. However, it is not clear that these moment conditions will identify the parameters of interest, and we therefore construct (population) objective functions from these moment conditions, and then explicitly verify that these objective functions are uniquely minimized at the parameter. It is the construction of the objective functions and verifying that they are uniquely minimized at the true parameter value that constitute the methodological contribution of the paper.

The general approach is to start with a comparison of two observations for a given individual in a panel. Based on these observations we will construct re-censored or re-truncated residuals
\( e_{its} (y_{it}, x_{it}, x_{is}, L_{it}, L_{is}, U_{it}, U_{is}, b) \) and \( e_{ist} (y_{is}, x_{is}, x_{it}, L_{it}, L_{is}, U_{it}, U_{is}, b) \) that have the same properties as \( e_{it} \) and \( e_{is} \) when \( b = \beta \). This will then imply that if \( e_{it} \) and \( e_{is} \) are identically distributed conditional on \((x_{it}, x_{is})\), then

\[
E \left[ (e_{its} (y_{it}, x_{it}, x_{is}, L_{it}, L_{is}, U_{it}, U_{is}, \beta) - e_{ist} (y_{is}, x_{is}, x_{it}, L_{it}, L_{is}, U_{it}, U_{is}, \beta)) \mid x_{it}, x_{is} \right] = 0. \tag{3}
\]

This will form the basis for construction of our estimators. Of course, once it has been established that \( e_{its} (y_{it}, x_{it}, x_{is}, L_{it}, L_{is}, U_{it}, U_{is}, \beta) \) and \( e_{ist} (y_{is}, x_{is}, x_{it}, L_{it}, L_{is}, U_{it}, U_{is}, \beta) \) are identically distributed then for any function, \( \xi (\cdot) \), we also have the moment condition

\[
E \left[ (\xi (e_{its} (y_{it}, x_{it}, x_{is}, L_{it}, L_{is}, U_{it}, U_{is}, \beta)) - \xi (e_{ist} (y_{is}, x_{is}, x_{it}, L_{it}, L_{is}, U_{it}, U_{is}, \beta))) \mid x_{it}, x_{is} \right] = 0. \tag{4}
\]

provided that the moment exists.

Moreover, if the errors \( e_{it} \) and \( e_{is} \) are also independent conditional on \((x_{it}, x_{is})\), then so are \( e_{its} (y_{it}, x_{it}, x_{is}, L_{it}, L_{is}, U_{it}, U_{is}, \beta) \) and \( e_{ist} (y_{is}, x_{is}, x_{it}, L_{it}, L_{is}, U_{it}, U_{is}, \beta) \). This implies that their difference is symmetrically distributed around 0, so for any odd function \( \xi (\cdot) \),

\[
E \left[ (\xi (e_{its} (y_{it}, x_{it}, x_{is}, L_{it}, L_{is}, U_{it}, U_{is}, \beta)) - \xi (e_{ist} (y_{is}, x_{is}, x_{it}, L_{it}, L_{is}, U_{it}, U_{is}, \beta))) \mid x_{it}, x_{is} \right] = 0. \tag{5}
\]

provided that the moment exists.\(^2\)

In this paper we will focus on (3) and the generalization (5). The reason for this is that in a linear model without censoring or truncation, (3) will correspond to OLS on the differenced data, whereas (5) will also accommodate least absolute deviation estimation on the differenced data as a special case. As already mentioned, the construction of the residuals \( e_{its} (y_{it}, x_{it}, x_{is}, L_{it}, L_{is}, U_{it}, U_{is}, \beta) \) is fairly straightforward, and the challenge is to show that although (3) and (5) may not identify \( \beta \), and unconditional version of them can be integrated to yield a population objective function that is uniquely minimized at \( b = \beta \). This will involve the integral of \( \xi (\cdot) \), which we will denote by \( \Xi (\cdot) \), and in addition to being odd, we will assume that \( \xi (\cdot) \) is also increasing, so \( \Xi (\cdot) \) is a convex symmetric function. The leading cases are \( \Xi (d) = |d| \) and \( \Xi (d) = d^2 \).

In order to simplify the exposition, we will first develop the case when \( L_{it} = 0 \) and \( U_{it} = 1 \). We will then demonstrate that the result can be adapted to the general case.

\(^2\)Of course one could combine the insight in (4) and (5) to get even more general moment conditions. See also the discussion in Arellano and Honoré (2001) and Honoré and Hu (2004).
2.1 Two-Sided Censoring

Consider first the situation with two-sided censoring. Consider an individual, $i$, in two time periods, $t$ and $s$, and assume that $\varepsilon_{it}$ and $\varepsilon_{is}$ are identically distributed. The distribution of $y_{it} - x'_{it}\beta$ will be the same as that of $\varepsilon_{it}$ except that the former is censored from below at $-x'_{it}\beta$ and from above at $1 - x'_{it}\beta$. Figure 1 illustrates this. The dotted line depicts the distribution of $\varepsilon_{it}$, while the solid line gives the distribution of $y_{it} - x'_{it}\beta$, which typically has point mass at $-x'_{it}\beta$ and $1 - x'_{it}\beta$ (illustrated by the latter vertical lines). Since $x'_{it}\beta$ will typically differ from $x'_{is}\beta$, the distributions of $y_{it} - x'_{it}\beta$ and $y_{is} - x'_{is}\beta$ (given $(x_{it}, x_{is})$) will differ even if $\{\varepsilon_{it}\}$ is stationary (given $(x_{it}, x_{is})$). However, it is clear that one could obtain identically distributed “residuals” by artificially censoring $y_{it} - x'_{it}\beta$ and $y_{is} - x'_{is}\beta$ from below at $\max \{-x'_{it}\beta, -x'_{is}\beta\}$ and from above at $\min \{1 - x'_{it}\beta, 1 - x'_{is}\beta\}$. See the dashed lines in Figure 1. One can then form moment conditions from the fact that the difference in these “re-censored” residuals will be orthogonal to functions of $(x_{it}, x_{is})$.\(^3\) Of course, this construction is only useful if $-1 < x'_{it}\beta - x'_{is}\beta < 1$, because otherwise, the supports of $y_{it} - x'_{it}\beta$ and $y_{is} - x'_{is}\beta$ will not overlap.

In order to proceed, we need explicit expressions for the difference in these “re-censored” residuals. Consider first the case when $x'_{it}\beta \geq x'_{is}\beta$. Then the difference in the artificially censored residuals for individual $i$ in periods $t$ and $s$ is

$$\begin{align*}
\max \{y_{it} - x'_{it}\beta, -x'_{is}\beta\} - \min \{y_{is} - x'_{is}\beta, 1 - x'_{it}\beta\} \\
= (\max \{y_{it} - (x'_{it}\beta - x'_{is}\beta) , 0\} - x'_{is}\beta) - (\min \{y_{is}, 1 - (x'_{it}\beta - x'_{is}\beta)\} - x'_{is}\beta) \\
= \max \{y_{it} - (x'_{it}\beta - x'_{is}\beta) , 0\} - \min \{y_{is}, 1 - (x'_{it}\beta - x'_{is}\beta)\},
\end{align*}$$

and when $x'_{it}\beta \leq x'_{is}\beta$

$$\begin{align*}
\min \{y_{it} - x'_{it}\beta, 1 - x'_{is}\beta\} - \max \{y_{is} - x'_{is}\beta, -x'_{it}\beta\} \\
= (\min \{y_{it}, 1 + (x'_{it}\beta - x'_{is}\beta)\} - x'_{it}\beta) - (\max \{y_{is} + (x'_{it}\beta - x'_{is}\beta), 0\} - x'_{it}\beta) \\
= \min \{y_{it}, 1 + (x'_{it}\beta - x'_{is}\beta)\} - \max \{y_{is} + (x'_{it}\beta - x'_{is}\beta), 0\}.
\end{align*}$$

\(^3\)Clearly, one can also use the fact that differences in functions of the re-censored residuals will be orthogonal to functions for the explanatory variables. As discussed in Arellano and Honoré (2001), one can also construct moment conditions based on symmetry under the additional assumption that $(\varepsilon_{i1}, ..., \varepsilon_{iT})$ is exchangeable conditional on $(x_{i1}, ..., x_{iT})$. This is the motivation for the approach in Honoré (1992).
If we define

$$u(y_1, y_2, d) = \begin{cases} 
0 & \text{for } d < -1 \\
1 + d & \text{for } -1 < d < c_1 \\
\min \{1 - y_2, y_1\} & \text{for } c_1 < d < c_2 \\
y_1 - y_2 - d & \text{for } c_2 < d < c_3 \\
\max \{y_1 - 1, -y_2\} & \text{for } c_3 < d < c_4 \\
d - 1 & \text{for } c_4 < d < 1 \\
0 & \text{for } d > 1 
\end{cases}$$

(8)

where $c_1 = \min \{-y_2, y_1 - 1\}$, $c_2 = \max \{-y_2, y_1 - 1\}$, $c_3 = \min \{1 - y_2, y_1\}$ and $c_4 = \max \{1 - y_2, y_1\}$

then $u(y_{it}, y_{is}, x_{it} \beta - x_{is} \beta)$ will give the difference in the re-censored residuals discussed above (see Appendix 1). Hence the moment conditions are

$$E \left[ u(y_{it}, y_{is}, x_{it} \beta - x_{is} \beta) \mid x_{it}, x_{is} \right] = 0,$$

which implies the unconditional moments

$$E \left[ u(y_{it}, y_{is}, (x_{it} - x_{is}) \beta) (x_{it} - x_{is}) \right] = 0.$$  

(9)

Panel A of Figures 2–4 depict the contribution to the moment condition function $u(y_1, y_2, d)$ for pairs of uncensored observations (Figure 2), pairs with one censored and one uncensored
Figure 2: The Functions $u\left(y_1, y_2, \cdot \right)$ and $U\left(y_1, y_2, \cdot \right)$. Neither Observation Censored.

observation (Figure 3) and pairs with one observation censored from above and one from below (Figure 4).

Although the true parameter value will satisfy (9), it is not in general the unique solution to the moment condition. This is illustrated in Panel A of Figure 5. It considers the case when $y_{i1}^* \sim N(0.5, 1)$ and $y_{i2}^* \sim N(0.4, 1)$ and both are censored from below at 0 and from above at 1. Note that this is the data generation process that one would get with $\beta = 0.1$, $x_{i1} = 1$ and $x_{i2} = 0$ for all $i$.

It is clear from Figure 5 that the moment condition $E\left[ u\left(y_{it}, y_{is}, x_{it}^\prime b - x_{is}^\prime b\right) | x_{it}, x_{is} \right] = 0$ does not identify the parameter, $\beta$, of the model. The most obvious reason is that, as mentioned, only observations for which $-1 < (x_{i1} - x_{i2})^\prime b < 1$ will contribute. In this case $x_{i1} - x_{i2} = 1$ for all observations, so the moment condition is automatically satisfied when $|b| > 1$.

Following Powell (1986), we attempt to overcome the non–identification based on the moment condition by turning it into the first order condition for a minimization problem. It is easy to see that (9) is (half of minus) the first order condition for minimizing

$$E\left[ U\left(y_{it}, y_{is}, (x_{it} - x_{is})^\prime b\right) \right],$$

(10)

8
Figure 3: The Functions \( u(y_1, y_2, \cdot) \) and \( U(y_1, y_2, \cdot) \). One Observation Censored.

Figure 4: The Functions \( u(y_1, y_2, \cdot) \) and \( U(y_1, y_2, \cdot) \). Both Observations Censored.
Figure 5: The Functions $E[u(y_1, y_2, \cdot)]$ and $E[U(y_1, y_2, \cdot)]$.

where

$$U(y_1, y_2, d) = \begin{cases} 
1 + 2c_1 + c_1^2 - 2c_3c_1 + 2c_3c_2 + (y_1 - y_2 - c_2)^2 & \text{for } d < -1 \\
-2d - d^2 + 2c_1 + c_1^2 - 2c_3c_1 + 2c_3c_2 + (y_1 - y_2 - c_2)^2 & \text{for } -1 < d < c_1 \\
-2c_3d + 2c_3c_2 + (y_1 - y_2 - c_2)^2 & \text{for } c_1 < d < c_2 \\
(y_1 - y_2 - d)^2 & \text{for } c_2 < d < c_3 \\
-2c_2d + 2c_2c_3 + (y_1 - y_2 - c_3)^2 & \text{for } c_3 < d < c_4 \\
d^2 + 2d + c_4^2 - 2c_4 - 2c_2c_4 + 2c_2c_3 + (y_1 - y_2 - c_3)^2 & \text{for } c_4 < d < 1 \\
1 + c_4^2 - 2c_4 - 2c_2c_4 + 2c_2c_3 + (y_1 - y_2 - c_3)^2 & \text{for } d > 1
\end{cases}$$

Panel B of Figures 2–4 depict the contribution to the objective function $U(y_1, y_2, d)$ for pairs of uncensored observations (Figure 2), pairs with one censored and one uncensored observation (Figure 3) and pairs with one observation censored from above and one from below (Figure 4). Like the estimator for the panel data one-sided censored regression model developed in Honoré (1992), the objective function is piecewise quadratic or linear. However, surprisingly, going from one-sided to two-sided censoring ruins the convexity of the objective function, and the shape of the function is more similar to the objective function in Powell (1986) although that estimator was developed for a cross section model with symmetrically distributed errors.

It is clear from the discussion above that the true $\beta$ will solve the first order condition for
minimizing the population objective function in (10). However, since $U$ is constant, linear, quadratic and convex, and quadratic and concave over different regions, it is not at all obvious that $\beta$ will be the unique solution to these first order conditions. The key step for establishing identification of $\beta$ is therefore to establish that the function in (10) is minimized at $\beta$. We establish this in Appendix 1 (Section 7.1), and the result is illustrated in Panel B of Figure 5.

As mentioned, it is also possible to construct moment conditions based on (5). Let $\Xi$ be convex and symmetric, and let $\xi(\cdot) = \Xi'(\cdot)$ (when it exists). When $\varepsilon_{it}$ and $\varepsilon_{is}$ are independent and identically distributed, we also have the moment conditions

$$E \left[ \xi \left( u(y_{it}, y_{is}, x_{it}' \beta - x_{is}' \beta) \right) \mid x_{it}, x_{is} \right] = 0$$

which imply the unconditional moments

$$E \left[ \xi \left( u(y_{it}, y_{is}, (x_{it} - x_{is})' \beta) \right) (x_{it} - x_{is}) \right] = 0$$

where

$$\xi(u(y_1, y_2, d)) = \begin{cases} 
0 & \text{for } d < -1 \\
\xi(1 + d) & \text{for } -1 < d < \min \{-y_2, y_1 - 1\} = c_1 \\
\xi(c_3) & \text{for } c_1 < d < \max \{-y_2, y_1 - 1\} = c_2 \\
\xi(y_1 - y_2 - d) & \text{for } c_2 < d < \min \{1 - y_2, y_1\} = c_3 \\
\xi(c_2) & \text{for } c_3 < d < \max \{1 - y_2, y_1\} = c_4 \\
\xi(d - 1) & \text{for } c_4 < d < 1 \\
0 & \text{for } d > 1 
\end{cases}$$

Except for a multiplicative constant, (11) is the first order condition for minimizing

$$E \left[ U_{\Xi}(y_{it}, y_{is}, (x_{it} - x_{is})' b) \right]$$

where $U_{\Xi}$ is found by integrating $\xi(u(y_1, y_2, d))$ over each of the regions and insisting on continuity at the boundaries between the regions:

$$U_{\Xi}(y_1, y_2, d) = \begin{cases} 
\Xi(0) - \Xi(1 + c_1) - \xi(c_3) c_1 + \xi(c_3) c_2 + \Xi(y_1 - y_2 - c_2) & \text{for } d < -1 \\
\Xi(1 + d) - \Xi(1 + c_1) - \xi(c_3) c_1 + \xi(c_3) c_2 + \Xi(y_1 - y_2 - c_2) & \text{for } -1 < d < c_1 \\
-\xi(c_3) d + \xi(c_3) c_2 + \Xi(y_1 - y_2 - c_2) & \text{for } c_1 < d < c_2 \\
-\xi(c_3) d + \xi(c_3) c_2 + \Xi(y_1 - y_2 - c_2) & \text{for } c_2 < d < c_3 \\
\Xi(d - 1) - \Xi(c_4 - 1) - \xi(c_2) c_4 + \xi(c_2) c_3 + \Xi(y_1 - y_2 - c_3) & \text{for } c_4 < d < 1 \\
\Xi(0) - \Xi(c_4 - 1) - \xi(c_2) c_4 + \xi(c_2) c_3 + \Xi(y_1 - y_2 - c_3) & \text{for } d > 1 
\end{cases}$$

The appendix establishes that (12) is uniquely minimized at $b = \beta$. 

11
2.2 Two–sided Truncation

Mimicing the argument for the model with two–sided censoring, it is clear that if $\varepsilon_{it}$ and $\varepsilon_{is}$ are independent and identically distributed conditional on $(x_{it}, x_{is})$, then the observed errors, $y_{it} - x_{it}'\beta$ and $-x_{is}'\beta$ from below and $1 - x_{it}'\beta$ and $1 - x_{is}'\beta$ from above. We can then construct identically distributed residuals by artificially truncating $y_{it} - x_{it}'\beta$ at $-x_{is}'\beta$ from below and at $1 - x_{is}'\beta$ from above (and similarly for $y_{is} - x_{is}'\beta$). This yields many moment conditions, including

$$
E \left[ r(y_{it}, y_{is}, (x_{it} - x_{is})'\beta) \right] = E \left[ (y_{it} - y_{is} - (x_{it} - x_{is})'\beta) \cdot 1 \{ -x_{is}'\beta \leq y_{it} - x_{it}'\beta \leq 1 - x_{is}'\beta \} \cdot 1 \{ -x_{it}'\beta \leq y_{is} - x_{is}'\beta \leq 1 - x_{it}'\beta \} (x_{it} - x_{is}) \right] \quad (13) 
$$

It is an easy exercise to see that except for a multiplicative constant (13) is the first order condition for minimizing

$$
E [R(y_{it}, y_{is}, (x_{it} - x_{is})'b)] 
$$

where $R(y_1, y_2, d)$ is defined by

$$
\begin{cases} 
\frac{1}{2} (y_1 - y_2 - \max \{y_1 - 1, -y_2\})^2 & \text{if } \max \{y_1 - 1, -y_2\} > d \\
\frac{1}{2} (y_1 - y_2 - d)^2 & \text{if } \max \{y_1 - 1, -y_2\} \leq d \leq \min \{y_1, 1 - y_2\} \\
\frac{1}{2} (y_1 - y_2 - \min \{y_1, 1 - y_2\})^2 & \text{if } d > \min \{y_1, 1 - y_2\} 
\end{cases}
$$

Figure 6 depicts the function $R$ and its derivative, whereas Figure 7 shows their expectation when $y_1^* \sim N(0.5, 1)$ and $y_2^* \sim N(0.4, 1)$ and both are truncated from below at 0 and from above at 1.

More generally, again let $\Xi$ be convex and symmetric, and let $\xi(\cdot) = \Xi'(\cdot)$ (when it exists). Then

$$
E \left[ \xi(y_{it} - y_{is} - (x_{it} - x_{is})'\beta) \cdot 1 \{ -x_{is}'\beta \leq y_{it} - x_{it}'\beta \leq 1 - x_{is}'\beta \} \cdot 1 \{ -x_{it}'\beta \leq y_{is} - x_{is}'\beta \leq 1 - x_{it}'\beta \} (x_{it} - x_{is}) \right] = 0
$$
or

$$
E \left[ \xi(y_{it} - y_{is} - (x_{it} - x_{is})'\beta) \cdot 1 \{ 0 \leq y_{it} - (x_{it} - x_{is})'\beta \leq 1 \} \cdot 1 \{ 0 \leq y_{is} + (x_{it} - x_{is})'\beta \leq 1 \} (x_{it} - x_{is}) \right] = 0
$$
Figure 6: The Functions $r(y_1, y_2, \cdot)$ and $R(y_1, y_2, \cdot)$.

Figure 7: The Functions $E[r(y_1, y_2, \cdot)]$ and $E[R(y_1, y_2, \cdot)]$. 
or with \( \delta = (x_{it} - x_{is})' \beta \),

\[
0 = E \left[ \xi (y_{it} - y_{is} - \delta) \cdot 1 \{0 \leq y_{it} - \delta \leq 1\} \cdot 1 \{0 \leq y_{is} + \delta \leq 1\} (x_{it} - x_{is}) \right] \\
= E \left[ \xi (y_{it} - y_{is} - \delta) \cdot 1 \{0 \leq y_{it} - \delta \leq 1\} \cdot 1 \{y_{it} - \delta \leq 1\} \cdot 1 \{0 \leq y_{is} + \delta \leq 1\} \cdot 1 \{y_{is} + \delta \leq 1\} (x_{it} - x_{is}) \right] \\
= E \left[ \xi (y_{it} - y_{is} - \delta) \cdot 1 \{\delta \leq y_{it}\} \cdot 1 \{y_{it} - 1 \leq \delta\} \cdot 1 \{-y_{is} \leq \delta\} \cdot 1 \{\delta \leq 1 - y_{is}\} (x_{it} - x_{is}) \right] \\
= E \left[ \xi (y_{it} - y_{is} - \delta) \cdot 1 \{\max \{y_{it} - 1, -y_{is}\} \leq \delta \leq \min \{y_{it}, 1 - y_{is}\}\} (x_{it} - x_{is}) \right]
\]

This is minus the derivative of \( E \left[ R_{E} (y_{it}, y_{is}, (x_{it} - x_{is})' b) \right] \) evaluated at \( b = \beta \), where

\[
R_{E} (y_{1}, y_{2}, d) = \begin{cases} 
\Xi (\max \{y_{1} - 1, -y_{2}\}) & \text{for } d < \max \{y_{1} - 1, -y_{2}\} \\
\Xi (y_{1} - y_{2} - d) & \text{for } \max \{y_{1} - 1, -y_{2}\} \leq d \leq \min \{y_{1}, 1 - y_{2}\} \\
\Xi (\min \{y_{1}, 1 - y_{2}\}) & \text{for } d > \min \{y_{1}, 1 - y_{2}\}
\end{cases}
\]

As was the case for the censored model, the argument above only establishes that the true \( \beta \) will solve the first order condition for minimizing \( E \left[ R_{E} (y_{it}, y_{is}, (x_{it} - x_{is})' b) \right] \). On the other hand, it is clear that without additional strong assumptions, \( \beta \) will not be the unique minimizer. The reason is that we know that, in general, the truncated regression model will not be identified with exponentially distributed errors. As a result, assumptions must be added that rule out the exponential distribution. In Appendix 1 (section 7.2), we show that \( \beta \) is the unique minimizer of \( E \left[ R (y_{it}, y_{is}, (x_{it} - x_{is})' b) \right] \), and more generally of \( E \left[ R_{E} (y_{it}, y_{is}, (x_{it} - x_{is})' b) \right] \), provided that the errors have a log–concave probability distribution. In his paper on censored and truncated cross sectional models with symmetric errors, Powell (1986) assumed that the errors have a strictly unimodal distribution in order to establish identification in the truncated case. As explained in Honoré (1992), one can think of log–concavity as a bivariate version of this: if two continuous random variables are independent and identically distributed, then the distribution of the difference is (strictly) unimodal conditional on the sum if, and only if, each has a (strictly) log–concave density. Honoré and Powell (1994) and Abrevaya (1999b) also assumed log–concavity to achieve identification; see also the discussion in Chen (forthcoming).

### 3 Estimation

The arguments leading to identification of the parameters of interest, \( \beta \), above were based on comparing two observations for the same individual and we showed that \( \beta \) could be expressed as
the unique minimizer of an expectation of the form \( E \left[ Q \left( y_{it}, y_{is}, (x_{it} - x_{is})' b \right) \right] \) for some function \( Q \). This suggests estimating \( \beta \) by minimizing a sample analog of this such as

\[
\hat{\beta} = \arg \min_b \frac{1}{n} \sum_{i=1}^{n} \left( \frac{T_i}{2} \right)^{-1} \sum_{1 \leq s < t \leq T_i} Q \left( y_{it}, y_{is}, (x_{it} - x_{is})' b \right)
\]

In this aggregation, observations get different weight depending on the number of observations for a given individual. Alternatively, one could also use objective functions of the type

\[
\arg \min_b \frac{1}{n} \sum_{i=1}^{n} \sum_{1 \leq s < t \leq T_i} w_{ist} Q \left( y_{it}, y_{is}, (x_{it} - x_{is})' b \right)
\]

where the \( w_{ist} \)'s are exogenous weights. In particular, with unbalanced panels, one might want \( w_{ist} \) to depend on \( T_i \), the number of time periods for individual \( i \). For example, one can think of the usual fixed effects estimator in a linear regression model as minimizing

\[
\frac{1}{n} \sum_{i=1}^{n} \sum_{1 \leq s < t \leq T_i} \left( y_{isa} - y_{ita} - (x_{isa} - x_{ita})' b \right)^2,
\]

so a simple natural choice for \( w_{ist} \) could be \( \frac{1}{T_i} \).

For two–sided censoring, the resulting estimator is

\[
\hat{\beta} = \arg \min_b \sum_{i=1}^{n} \sum_{1 \leq s < t \leq T_i} w_{ist} U \left( y_{it}, y_{is}, (x_{it} - x_{is})' b \right)
\]

or more generally

\[
\hat{\beta} = \arg \min_b \sum_{i=1}^{n} \sum_{1 \leq s < t \leq T_i} w_{ist} U_{*} \left( y_{it}, y_{is}, (x_{it} - x_{is})' b \right)
\]

where the functions \( U \) and \( U_{*} \) are defined in Section 2.1. Standard arguments\(^4\) yield

**Theorem 1** Consider a random sample of size \( n \) from \( \left( T_i, \{ y_{it}, x_{it} \}_{t=1}^{T_i} \right) \). If

1. \( y_{it} = \begin{cases} 
0 & \text{if } x'_{it} \beta + \varepsilon_{it} < 0 \\
 x'_{it} \beta + \varepsilon_{it} & \text{if } 0 \leq x'_{it} \beta + \varepsilon_{it} \leq 1 \\
1 & \text{if } x'_{it} \beta + \varepsilon_{it} > 1
\end{cases} \)

2. \( (\varepsilon_{i1}, \varepsilon_{i2}, \cdots, \varepsilon_{iT_i}) \) is continuously distributed conditional on \( \left( T_i, \{ x_{it} \}_{t=1}^{T_i} \right) \) with a density that is continuous and positive everywhere,

\(^4\) Consistency follows from Theorem 4.1.1 of Amemiya (1985) and asymptotic normality from, for example, Theorem 3.3 of Pakes and Pollard (1989).
3. the sequence $\varepsilon_{i1}, \varepsilon_{i2}, \ldots, \varepsilon_{iT_i}$ is stationary conditional on $\left(T_i, \{x_{it}\}_{t=1}^{T_i}\right)$, and for any $s, t \leq T_i$

there exists a random variable, $\nu_{ist}$, such that $\varepsilon_{is}$ and $\varepsilon_{it}$ are independent conditional on $\nu_{ist}$,

4. the matrix

$$E \left[ (x_{is} - x_{it})(x_{is} - x_{it})' \right] - 1 < (x_{is} - x_{it})' \beta < 1$$

has full rank

then

$$\sqrt{n} \left( \hat{\beta} - \beta \right) \xrightarrow{d} N \left( 0, \Gamma^{-1}V\Gamma^{-1} \right)$$

where $\hat{\beta}$ is defined in (15) and

$$\Gamma = \left. \frac{dE \left[ \sum_{s<t} w_{i,t-s} \xi \left( u \left( y_{is}, y_{is}, (x_{it} - x_{is})' \beta \right) \right) (x_{it} - x_{is}) \right] }{\beta} \right|_{\beta = \beta}$$

and

$$V = E \left[ v_i v_i' \right]$$

with

$$v_i = \sum_{s<t} w_{i,t-s} \xi \left( u \left( y_{is}, y_{is}, (x_{it} - x_{is})' \beta \right) \right) (x_{it} - x_{is}) \cdot$$

These assumptions are consistent with a “fixed effects” model in which $\varepsilon_{it} = \alpha_i + \varepsilon_{it}$ with $\alpha_i$
unrestricted and the sequence $\{\varepsilon_{it}\}_{t=1}^{T_i}$ independent and identically distributed. The assumptions
also allow for some correlation in the $\varepsilon_{it}$’s. For example if $(\varepsilon_{is}, \varepsilon_{it})$ is bivariate normal with the
same variance, then they can be written as $\varepsilon_{is} = Z_{is} + Q_i$ and $\varepsilon_{it} = Z_{it} + Q_i$ where $Z_{it}, Z_{is}$ and
$Q_i$ are independent normals. So $\varepsilon_{is}$ and $\varepsilon_{it}$ are independent conditional on $Q_i$. The assumption
that $(\varepsilon_{i1}, \varepsilon_{i2}, \ldots, \varepsilon_{iT_i})$ is continuously distributed is necessary if one wants to allow $\Xi$ to be non-
differentiable (i.e., $\Xi \left( d \right) = |d|$). Without it, the derivative in the expression for $\Gamma$ might not exist.

When $\Xi \left( d \right) = d^2$, condition 3 can be reduced to assuming that the sequence $\varepsilon_{i1}, \varepsilon_{i2}, \ldots, \varepsilon_{iT_i}$ is
stationary conditional on $\left(T_i, \{x_{it}\}_{t=1}^{T_i}\right)$, and condition 2 is not necessary. In that case the terms
in the asymptotic variance reduce to

$$\Gamma = E \left[ \sum_{s<t} w_{i,t-s} \xi \left\{ -1 < (x_{is} - x_{it})' \beta < 1 \right\} \right.$$

$$\left( 1 \{ -1 < (x_{is} - x_{it})' \beta < y_{is} - 1 \} - 1 \{ 0 < (x_{is} - x_{it})' \beta < y_{is} \} - 1 \{ -y_{it} < (x_{is} - x_{it})' \beta < 0 \} + 1 \{ 1 - y_{it} < (x_{is} - x_{it})' \beta < 1 \} \right) (x_{is} - x_{it}) (x_{is} - x_{it})'$$
\[ V = E \left[ v_i v'_i \right] \]

with

\[ v_i = \sum_{s < t} w_{i,t-s} u \left( y_{is}, (x_{is} - x_{it})' \beta \right) (x_{is} - x_{it}) \]

Following standard arguments, these are consistently estimated by

\[ \hat{\beta} = \arg \min_b \frac{1}{n} \sum_{i=1}^n \sum_{s < t} w_{i,t-s} \left\{ -1 < (x_{is} - x_{it})' \hat{\beta} < y_{is} - 1 \right\} - \left\{ 0 < (x_{is} - x_{it})' \hat{\beta} < y_{is} \right\} - \left\{ -y_{it} < (x_{is} - x_{it})' \hat{\beta} < 0 \right\} + \left\{ 1 - y_{it} < (x_{is} - x_{it})' \hat{\beta} < 1 \right\} (x_{is} - x_{it}) (x_{is} - x_{it})' \]

and

\[ \hat{V} = \frac{1}{n} \sum_{i=1}^n \hat{v}_i \hat{v}'_i \]

with

\[ \hat{v}_i = \sum_{s < t} w_{i,t-s} u \left( y_{is}, (x_{is} - x_{it})' \hat{\beta} \right) (x_{is} - x_{it}) \]

For two-sided truncation the resulting estimator is

\[ \hat{\beta} = \arg \min_b \frac{1}{n} \sum_{i=1}^n \sum_{1 \leq s < t \leq T_i} w_{ist} R \left( y_{it}, y_{is}, (x_{it} - x_{is})' b \right) \]

or more generally

\[ \hat{\beta} = \arg \min_b \frac{1}{n} \sum_{i=1}^n \sum_{1 \leq s < t \leq T_i} w_{ist} R_{xz} \left( y_{it}, y_{is}, (x_{it} - x_{is})' b \right) \]

where the functions \( R \) and \( R_{xz} \) are defined in Section 2.2.

We have

**Theorem 2** Consider a random sample of size \( n \) from \( \left( T_i, \{ y_{it}, x_{it} \}_{i=1}^{T_i} \right) \). If

1. \( y_{it} \) is drawn from the distribution of \( x_{it}' \beta + \varepsilon_{it} \) conditional on \( 0 \leq x_{it}' \beta + \varepsilon_{it} \leq 1 \)

2. \( (\varepsilon_{i1}, \varepsilon_{i2}, \ldots, \varepsilon_{iT_i}) \) is continuously distributed conditional on \( \left( T_i, \{ x_{it} \}_{t=1}^{T_i} \right) \) with a density that is continuous and positive everywhere

3. the sequence \( \varepsilon_{i1}, \varepsilon_{i2}, \ldots, \varepsilon_{iT_i} \) is stationary conditional on \( \left( T_i, \{ x_{it} \}_{t=1}^{T_i} \right) \), and for any \( s, t \leq T_i \) there exists a random variable, \( \nu_{ist} \), such that \( \varepsilon_{is} \) and \( \varepsilon_{it} \) are independent and have log-concave density conditional on \( \nu_{ist} \),
4. the matrix
\[ E \left[ (x_{it} - x_{it}) (x_{is} - x_{it})' \right] - 1 < (x_{is} - x_{it})' \beta < 1 \]
has full rank

then
\[ \sqrt{n} \left( \hat{\beta} - \beta \right) \xrightarrow{d} N \left( 0, \Gamma^{-1} \Gamma^{-1} \right) \]
where \( \hat{\beta} \) is defined in (17) and
\[
\Gamma = \frac{dE \left[ \sum_{s < t} w_{i,t-s} \xi \left( r \left( y_{it}, y_{is}, (x_{it} - x_{is})' \beta \right) \right) (x_{it} - x_{is}) \right]}{db'} \bigg|_{b=\beta}
\]
and
\[ V = E \left[ v_i v_i' \right] \]
with
\[
v_i = \sum_{s < t} w_{i,t-s} \xi \left( r \left( y_{it}, y_{is}, (x_{it} - x_{is})' \beta \right) \right) (x_{it} - x_{is})
\]

4 Extensions

4.1 Mixed Censored/Truncation

Having considered models with two–sided censoring or truncation, it is natural to also consider a regression model with censoring from one side and truncation from the other:

\[
y_{it}^* = x_{it}' \beta + \varepsilon_{it}
\]
\[(y_{it}, x_{it}) = (\min \{ y_{it}^*, U_{it} \}, x_{it}) \text{ conditional on } L_{it} \leq y_{it} \quad (18)
\]

To simplify the notation, we again focus on the case where \( L_{it} = 0 \) and \( U_{it} = 1 \). In this case the moment condition based on the same logic as above is

\[
0 = E \left[ 1 \left\{ y_{it} - x_{it}' \beta > -x_{is}' \beta \right\} \cdot \min \left\{ y_{it} - x_{it}' \beta, 1 - x_{is}' \beta \right\} \right] \cdot E \left[ 1 \left\{ y_{is} - x_{is}' \beta > -x_{it}' \beta \right\} \cdot \min \left\{ y_{is} - x_{is}' \beta, 1 - x_{it}' \beta \right\} \right] \bigg| x_{it}, x_{is}
\]

\[
= E \left[ 1 \left\{ y_{it} > (x_{it} - x_{is})' \beta \right\} \cdot \min \left\{ y_{is}, 1 - (x_{it} - x_{is})' \beta \right\} \right] \bigg| x_{it}, x_{is}
\]

\[
= E \left[ t \left( y_{it}, y_{is} (x_{it} - x_{is})' \beta \right) \bigg| x_{it}, x_{is} \right].
\]
where we have assumed that \( \varepsilon_{it} \) and \( \varepsilon_{is} \) are independent and identically distributed conditional on \((x_{it}, x_{is})\).

This implies the unconditional moment condition

\[
E \left[ t \left( y_{it}, y_{is} (x_{it} - x_{is})' \beta \right) (x_{it} - x_{is}) \right] = 0
\]

where

\[
t (y_1, y_2, d) = \begin{cases} 
0 & \text{for } d < \bar{c}_1 \\
1 - y_2 & \text{for } \bar{c}_1 < d < c_2 \\
(y_1 - y_2 - d) & \text{for } c_2 < d < c_3 \\
y_1 - 1 & \text{for } c_3 < d < \bar{c}_4 \\
0 & \text{for } \bar{c}_4 < d
\end{cases}
\]

where \( \bar{c}_1 = -y_2, \) \( c_2 = \max \{y_1 - 1, -y_2\}, \) \( c_3 = \min \{1 - y_2, y_1\} \) and \( \bar{c}_4 = y_1. \) Note that \( c_2 \) and \( c_3 \) are defined as before, but \( \bar{c}_1 \) and \( \bar{c}_4 \) differ from \( c_1 \) and \( c_4. \)

Let

\[
T (y_1, y_2, d) = \begin{cases} 
-2 (1 - y_2) \bar{c}_1 + 2 (1 - y_2) c_2 + (y_1 - y_2 - c_2)^2 & \text{for } d < \bar{c}_1 \\
-2 (1 - y_2) d + 2 (1 - y_2) c_2 + (y_1 - y_2 - c_2)^2 & \text{for } \bar{c}_1 < d < c_2 \\
(y_1 - y_2 - d)^2 & \text{for } c_2 < d < c_3 \\
-2 (y_1 - 1) d + 2 (y_1 - 1) c_3 + (y_1 - y_2 - c_3)^2 & \text{for } c_3 < d < \bar{c}_4 \\
-2 (y_1 - 1) \bar{c}_4 + 2 (y_1 - 1) c_3 + (y_1 - y_2 - c_3)^2 & \text{for } \bar{c}_4 < d
\end{cases}
\]

We then define the estimator of \( \beta \) by minimizing

\[
\sum_i \sum_{t < s} w_{its} T (y_{it}, y_{is}, (x_{it} - x_{is})' b)
\]

The function \( T \) and its derivative are depicted in Figures 8 and 9 for a pair of uncensored observations and for a pair with one censored and one uncensored observation, respectively.

Figure 10 shows the moment condition and the expected value of the objective function when \( y_1^* \sim N (0.5, 1) \) and \( y_2^* \sim N (0.4, 1) \) and both are truncated from below at 0 and censored from above at 1.

As before, we also have

\[
E \left[ \xi \left( t \left( y_{it}, y_{is} (x_{it} - x_{is})' \beta \right) \right) (x_{it} - x_{is}) \right] = 0
\]

where \( \Xi \) is convex and symmetric, and \( \xi (\cdot) = \Xi' (\cdot) \) (when it exists)
Figure 8: The Functions $t(y_1, y_2, \cdot)$ and $T(y_1, y_2, \cdot)$. Neither Observation Censored.

Figure 9: The Functions $t(y_1, y_2, \cdot)$ and $T(y_1, y_2, \cdot)$. One observation censored.
Figure 10: The Functions $E[t(y_1, y_2, \cdot)]$ and $E[T(y_1, y_2, \cdot)]$.

Let

\[
T_\Xi(y_1, y_2; d) = \begin{cases} 
-\xi (1 - y_2) \bar{c}_1 + \xi (1 - y_2) c_2 + \Xi (y_1 - y_2 - c_2) & \text{for } d < \bar{c}_1 \\
-\xi (1 - y_2) d + \xi (1 - y_2) c_2 + \Xi (y_1 - y_2 - c_2) & \text{for } \bar{c}_1 < d < c_2 \\
\Xi (y_1 - y_2 - d) & \text{for } c_2 < d < c_3 \\
-\xi (y_1 - 1) d + \xi (y_1 - 1) c_3 + \Xi (y_1 - y_2 - c_3) & \text{for } c_3 < d < \bar{c}_4 \\
-\xi (y_1 - 1) \bar{c}_4 + \xi (y_1 - 1) c_3 + \Xi (y_1 - y_2 - c_3) & \text{for } \bar{c}_4 < d
\end{cases}
\]

We then define the estimator of $\beta$ by minimizing

\[
\sum_i \sum_{t<s} w_{its} T_\Xi (y_{it}, y_{is}, (x_{it} - x_{is})' b).
\]

### 4.2 General Censoring Points.

It is easy to generalize the results above to the case where the truncation and censoring points are not all 0 and 1. For example, consider the model with two-sided censoring

\[
y_{it} = \begin{cases} 
L_i & \text{if } y_{it}^* < L_i \\
x_{it}' \beta + \varepsilon_{it} & \text{if } L_i \leq y_{it}^* \leq U_i \\
U_i & \text{if } y_{it}^* > U_i
\end{cases}
\]
then

\[
\begin{cases}
\frac{y_{it} - L_i}{U_i - L_i} = 0 & \text{if } y_{it}^* < L_i \\
\left( \frac{x_{it} - L_i}{U_i - L_i} \right)' + \frac{\varepsilon_{it} - L_i}{U_i - L_i} & \text{if } L_i \leq y_{it}^* \leq U_i \\
1 & \text{if } y_{it}^* > U_i
\end{cases}
\]

and we could estimate \( \beta \) by

\[
\hat{\beta} = \arg\min_b \sum_i \sum_{t < s} w_{its} U \Xi \left( \frac{y_{it} - L_i}{U_i - L_i}, \frac{y_{is} - L_i}{U_i - L_i}, \left( \frac{x_{it} - x_{is}}{U_i - L_i} \right)' b \right)
\]  

(19)

This simple approach does not work when the censoring points are time-varying, because then \( \frac{\varepsilon_{it} - L_i}{U_i - L_i} \) is not stationary.

In order to proceed, we need explicit expressions for the difference in these “re–censored” residuals. We first note that only pairs for which the support of the re–censored residuals overlap can play a role in the moment conditions leading to the objective function. These pairs are characterized by

\[
L_{it} - U_{is} < x_{it}' \beta - x_{is}' \beta < U_{it} - L_{is}
\]

and for such pairs, the difference in the artificially censored residuals for individual \( i \) in periods \( t \) and \( s \) is

\[
\text{mami} \left\{ L_{is} - x_{is}' \beta, y_{it} - x_{it}' \beta, U_{is} - x_{is}' \beta \right\} - \text{mami} \left\{ L_{it} - x_{it}' \beta, y_{is} - x_{is}' \beta, U_{it} - x_{it}' \beta \right\} = \text{mami} \left\{ L_{is}, y_{it} - (x_{it}' \beta - x_{is}' \beta), U_{is} \right\} - \text{mami} \left\{ L_{it}, y_{is} + (x_{it}' \beta - x_{is}' \beta), U_{it} \right\} + (x_{it}' \beta - x_{is}' \beta)
\]

where we use the notation \( \text{mami} \{a, x, b\} = \max \{a, \min \{x, b\} \} \).

If we define

\[
k(L, U, y, d) = \begin{cases}
U & \text{for } d < y - U \\
y - d & y - U < d < y - L \\
L & \text{for } d > y - L
\end{cases}
\]

and

\[
u(y_1, y_2, d, L_1, L_2, U_1, U_2) = 1 \left\{ L_1 - U_2 < d < U_1 - L_2 \right\} (k(L_2, U_2, y_1, d) - k(L_1, U_1, y_2, -d) + d)
\]

then

\[
E \left[ u \left( y_{it}, y_{is}, (x_{it}' \beta - x_{is}' \beta), L_{it}, L_{is}, U_{it}, U_{is} \right) \left( x_{it}, x_{is} \right) \right] = 0
\]

and hence

\[
E \left[ u \left( y_{it}, y_{is}, (x_{it}' \beta - x_{is}' \beta), L_{it}, L_{is}, U_{it}, U_{is} \right) \left( x_{it} - x_{is} \right) \right] = 0.
\]

These will be the moment conditions that lead to the estimator in this case.
Also define

\[
K(L, U, y, d) = \begin{cases} 
2yU - 2dU - U^2 & \text{for } d < y - U \\
(y - d)^2 & \text{for } y - U < d < y - L \\
2yL - 2dL - L^2 & \text{for } d > y - L 
\end{cases}
\]

\[
S(y_1, y_2, d, L_1, L_2, U_1, U_2) = K(L_2, U_2, y_1, d) + K(L_1, U_1, y_2, -d) - d^2
\]

and

\[
V(y_1, y_2, d, L_1, L_2, U_1, U_2) = \begin{cases} 
S(y_1, y_2, L_1 - U_2, L_1, L_2, U_1, U_2) & \text{for } d < L_1 - U_2 \\
S(y_1, y_2, d, L_1, L_2, U_1, U_2) & \text{for } L_1 - U_2 < d < U_1 - L_2 \\
S(y_1, y_2, U_1 - L_2, L_1, L_2, U_1, U_2) & \text{for } d > U_1 - L_2 
\end{cases}
\]

and the estimator for \( \beta \) is then defined by

\[
\arg\min_b \sum_{i=1}^{n} \sum_{t<s} w_{its} V(y_{it}, y_{is}, (x_{it} - x_{is})' b, L_{it}, L_{is}, U_{it}, U_{is})
\]

A version of this can be developed for a general loss function.

All of these extensions assume that the censoring and truncation points are exogenous in the sense that one must make assumptions on the error terms conditional on them. In a recent paper, Khan, Ponomareva, and Tamer (2011) consider a (one-sided) censored regression model with endogenous censoring. Their approach only leads to partial identification, but it would be interesting to generalize it to more general versions of the models considered here.

### 4.3 Pairwise Difference Versions

If we can estimate a panel data model with two observations per unit, then we can apply the same idea to any two observations in a cross section, treating the constant in the cross sectional model as an individual-specific effect. This idea was explicitly used in Honoré and Powell (1994) to construct estimators for the parameters of cross sectional (one-sided) censored and truncated regression models based on the panel data estimator in Honoré (1992). Among others, this also characterizes the relationship between the estimators in Manski (1987) and Han (1987)\(^5\) and between the estimators in Kyriazidou (1997) and Powell (1987). The same idea can be applied to the

---

\(^5\)Strictly speaking, Manski (1987) only considers the binary choice model, while Han (1987) deals with a more general model. However, Abrevaya (1999a) has demonstrated how the approach in Manski (1987) can be generalized to deal with models like the ones in Han (1987).
To illustrate this, consider a generalized accelerated failure time model of the type\(^6\),

\[
\Lambda(T_i; \tau) = \phi(z_i; \beta) + \nu_i
\]

where \(\Lambda\) and \(\phi\) are known functions, \(\tau\) and \(\beta\) are parameters to be estimated, \(z_i\) is a vector of covariates and \(\nu_i\) is a random variable which is independent of \(z_i\), but whose distribution is unspecified. For simplicity, function \(\Lambda\) is assumed to be strictly increasing in its first argument.

With stock sampling, one only observes \(T_i\) conditional on \(T_i\) being greater than some duration \(a_i\). Moreover \(T_i\) may be censored at \(b_i\). For the accelerated failure time model, \(\Lambda(T_i; \tau) = \log(T_i)\), and a combination of the insights in Sections 4.1 and 4.2 can therefore be used to estimate \(\beta\). More generally, (20) implies the conditional moment conditions.

\[
0 = E \left[ \min \left\{ \Lambda(T_i; \tau) - \phi(z_i; \beta) > \Lambda(a_j; \tau) - \phi(z_j; \beta) \right\} 1 \left\{ \Lambda(T_j; \tau) - \phi(z_j; \beta) > \Lambda(a_i; \tau) - \phi(z_i; \beta) \right\} \right] z_i, z_j
\]

Whether or not this is sufficient to identify \(\tau\) and \(\beta\) will depend on the functions \(\Lambda\) and \(\phi\).

5 Empirical Application

In this section we apply the estimator in Section 2.1 to analyze the portfolio-reshuffling effect of a tax reform that increased the after-tax capital income on bonds relative to stocks in Denmark in 1987. The dependent variable is the portfolio share of bonds, which is censored at 0 from below and at 1 from above. This motivates the use of the estimation procedures proposed in this paper.

We use a panel data set constructed from administrative records covering two years before and after the reform to estimate a portfolio share equation for bonds as a function of marginal tax rates on capital income. The analysis presented here follows the literature on taxation and portfolio structure, e.g., Feldstein (1976), Hubbard (1985), King and Leape (1998), Samwick (2000), Poterba and Samwick (2002), Poterba (2002) and Alan, Crossley, Atalay, and Jeon (2010). These papers analyze (repeated) cross sections of households.\(^7\) Here the analysis is extended by using panel data and controlling for time-invariant correlated heterogeneity, i.e., fixed effects. Controlling for


\(^7\)Bakija (2000) uses the limited panel module of the American Survey of Consumer Finances (SCF) to study portfolio changes around the 1988 tax reform. However, his data set is very small (984 households) and unrepre-
correlated unobserved fixed factors is likely to be important in this context, since the portfolio composition of a household is likely to be influenced by time–invariant factors such as risk aversion and time discounting.

In the next subsection we give a brief overview over the tax reform. After this, we introduce the data and present the results.

5.1 The Tax Reform

The tax reform, announced in 1985 and implemented in 1987, broke the link between the marginal tax rates on earned income and capital income. Before the reform, all income was taxed at the same marginal tax rate. With the reform the tax rate on positive capital income for high-income households was decreased from 73 percent to 56 percent. The reform thereby increased the after-tax return on interest-bearing assets and therefore encouraged households to shift their portfolios toward such assets. The reform also changed the tax value of interest deductions from 73 to about 50 percent, and this substantially increased the cost of debt, primarily mortgages, for leveraged high-income households. For such households the reform effectively brought a negative wealth shock, giving them a strong incentive to lower their debt burden.8

The exact changes, however, differed across municipalities. The Danish income tax system is built around a proportional local government tax and a progressive tax collected by the central government. While the progressive schedule is the same for everybody in Denmark, the local government tax rates vary across municipalities. A tax ceiling, however, insured that the marginal tax rate could be at the maximum 73 percent. After the reform the tax ceiling on earned income was reduced to 68 percent in the highest bracket9 and 56 percent in the middle bracket. Capital income was now taxed at the same rate independently of the level of earned income. The marginal tax rates across tax brackets before and after the reform are summarized in Table 1 (see Appendix 2)

8 Alan and Leth-Petersen (2006) document that the reduced value of the interest deduction led households to liquidate financial assets to lower their mortgage debt. This was possible because pre-payment of mortgage debt is not restricted in Denmark.

9 Approximately 20 percent of the population belong in the top bracket.
The application of a tax ceiling together with the heterogeneous local government tax rates implies that the reform had differential effects on people living in different municipalities. Figures 11 and 12 illustrate the changes in marginal tax rates due to the reform for a high-tax and a low-tax municipality, respectively.

For a high-income person living in the municipality with the high local government tax, the marginal tax rate on positive net capital income falls by 14.5 percentage points and the marginal tax rate on negative net capital income falls by 20.5 percentage points. For a similar person living in the municipality with the low local government tax rate, the marginal tax rate on positive net capital income falls by 16.1 percentage points and negative net capital income falls by 22.1 percentage points. It is these differences in changes of marginal tax rates that we will exploit for identifying the effect of changes in marginal tax rates on the portfolio allocation when using the fixed effects estimator.

The marginal tax rates on capital income refer to income received in the form of dividends on stocks and interest payments from interest bearing accounts and bonds. Both before and after the reform, realized capital gains/losses associated with trading assets were generally not taxed. The exemption from this rule is capital gains from corporate stocks held for less than three years. Such
capital gains are taxed as earnings. Dividend payments were low relative to interest received from bonds.\(^{10}\) This suggests that lowering the marginal tax rate on positive capital income affected bonds and stocks differentially, favoring mainly income from bonds. In the empirical analysis we therefore focus on reshuffling between bonds and stocks.

### 5.2 Data

The data set is drawn from a random sample of 10 percent of the Danish population observed in the years 1984 to 1988. Information on portfolio allocations, income, wealth and demographics is collected and merged from different public administrative registers for all adult members of the household that the sampled person belongs to. Portfolio and income information is obtained from the income tax register. The portfolio information exists because Denmark had a wealth tax that required all wealth holdings to be reported to the tax authorities. This information allows us to break the wealth of each household into holdings of stocks and bonds. “Stocks” includes all

\(^{10}\)The median household in the sample holding stocks received dividends corresponding to 2 percent of the value of the stocks. The median household in the sample holding bonds received interest payments from these corresponding to 10 percent of the value of the bonds.
holdings of publicly and privately traded stocks, and “bonds” includes government and corporate bonds. The holdings of stocks and bonds are self-reported through the tax return and then audited by the tax authority.

5.2.1 Sample selection

For our analysis we exclude observations if one of the household members is self-employed, since register data are not likely to contain a good measure of own business wealth and because taxable income is quite volatile for those individuals. Sampled individuals younger than 18 or older than 60 are dropped as are students and individuals living together with his/her parents or living in a common household, i.e., a household with more than one family. To keep the focus on the importance of tax incentives, we include only stable couples, i.e., couples where the partner is the same in 1984 through 1988. On the same grounds we also exclude couples moving in the sample period. For the purpose of the analysis we require that households entering the sample be observed in all years in the period 1984-1988 so that we have a balanced panel.

Our objective is to investigate whether households reshuffle their portfolios in response to a change in tax incentives. As in most industrialized countries many Danish households have fairly undiversified portfolios. Since the decrease in the value of interest deductions generated a large negative wealth shock, clearly, these households are not likely to engage in portfolio reshuffling and hence cannot give us a clean answer regarding portfolio readjustments. We therefore construct a sub-sample of households holding positive amounts of stocks or bonds of at least 5,000 DKK in 1984. We also require households to hold a positive amount of either stocks or bonds throughout the rest of the observation period. This selection is introduced because we want to focus on households with a potential to reshuffle between stocks and bonds. We also drop renters because there are few renters with diversified portfolios.\footnote{For assessing portfolio reshuffling renters could have been included. We have chosen to leave them out of this analysis because there are only a few renters (898) with positive financial wealth of at least 5000 DKK in 1984. Moreover, renters generally do not provide a good comparison group for homeowners, since different preference parameters may govern their behavior.} This results in a final subsample which includes 8,577 households.\footnote{See Alan and Leth-Petersen (2006) for a more detailed analysis.}
5.3 Results

In this section we investigate if households reshuffled their portfolio of bonds and stocks as a response to the changes in relative after-tax returns on assets brought about by the 1987 tax reform. To do this, we employ the estimator in equation (15) with $\Xi(d) = d^2$, and estimate a portfolio share equation where the fraction of bonds in financial wealth, defined as the sum of bonds and stocks, is regressed on the marginal tax rate on positive capital income and some control variables.

The distinguishing feature of our data set is the panel dimension. This facilitates estimating portfolio share equations allowing for correlated unobserved time–invariant heterogeneity. This is important because we believe that unobserved time–invariant factors, such as risk aversion and time preferences, are correlated with wealth. High risk aversion may, for example, lead to a higher portfolio share of safe assets, such as bonds, for a given level of wealth.

Before the reform, capital income and earnings were lumped together and taxed according to a progressive tax scheme. This implies that households choose their tax bracket when choosing their portfolios and that the marginal tax rate on capital income is likely to be an endogenous regressor. We address this by calculating the marginal tax rate on capital income based on the household’s income in 1984, the year before the reform was announced, but using current year rules. In this way the individual level tax bracket is allocated based on information that was predetermined relative to the portfolio response to the reform.

We regress portfolio shares on the marginal tax rate on positive capital income, the log of total financial assets, i.e., assets held in stocks and bonds, and a set of year dummies. Tax rate changes vary across municipalities, but most of the change in tax rates is common across municipalities. Year dummies control for the effect of this common part, thereby also removing the major part of the wealth effect brought about by the reform. Effectively, by introducing year dummies, the coefficients on marginal tax rates are identified by differences in changes of marginal tax rates. Year dummies may also pick up common effects related to fluctuations in assets. Financial assets
control for any remaining wealth effect that might be present.\footnote{An alternative identification strategy could be based on comparing the behavior of households in different tax brackets. Households in the lowest tax bracket faced only very small changes in marginal tax rates on capital income, and households in the middle tax bracket faced different changes in marginal tax rates than households in the highest tax bracket. In our case this is not a natural approach to follow. High- and low-income people are different in terms of wealth levels and portfolio composition and possibly different with respect to preference parameters such as the discount rate and the level of risk aversion. Households in lower tax brackets therefore do not represent a natural control group for high-income households.}

Table 2 of Appendix 2 presents the parameter estimates from estimating random effects Tobit and fixed effects censored regression models for the portfolio share of bonds in financial wealth. The estimated parameter on the marginal tax rate is negative.\footnote{As explained in Honoré (2008), the parameter estimates for both the random effects and the fixed effects models can be converted to marginal effects by multiplying them by the fraction of observations that are not censored.} If year dummies and financial assets pick up the wealth effect related to the reform, in particular the effect of the reduction in the value of interest deduction that led households to liquidate financial assets, then this is exactly what economic theory predicts. Households should substitute from stocks toward bonds, whose relative after-tax return increased, and this is what the results indicate.

Considering the corresponding random effects estimates, we can see that the parameter estimates on financial assets and on year dummies are quite different, and the test of equality of all the parameters in the random effects and fixed effects specifications rejects.

### 6 Concluding Remarks

This paper constructs estimators for panel data regression models with individual specific heterogeneity and two-sided censoring and truncation. Following Powell (1986) the estimation strategy is based on moment conditions constructed from re-censored or re-truncated residuals. While these moment conditions do not identify the parameter of interest, they can be used to motivate objective functions that do. This part is the main methodological contribution of the paper. We apply one of the estimators to study the effect of a Danish tax reform on household portfolio choice. We find that a random effects specification can be rejected in favor of the “fixed” effects specification studied here, although both models yield the same sign of the key parameter that one would anticipate from economic theory.
References


7 Appendix 1: Proofs and Derivations

7.1 Two–Sided Censoring

This section provides justification for the statements about the estimators for the model with two-sided censoring. We first verify that \( u(y_{it}, y_{is}, x_{it}^\prime \beta - x_{is}^\prime \beta) \) in equation (8) does indeed yield the difference in the re-censored residuals defined in (6) and (7). Write \( d = x_{it}^\prime \beta - x_{is}^\prime \beta \), and consider first the case where \( d > 0 \). In this case, the difference in the re-censored residuals is

\[
\max\{y_{it} - d, 0\} - \min\{y_{is}, 1 - d\}
\]

which is most easily analyzed by considering a number of cases for \( d \) between 0 and 1. As mentioned earlier, the contribution to the moment condition should be 0 when \( d \) is outside the interval between \(-1 \) and \( 1 \), which is consistent with the definition of \( u \).

There are four cases based on combinations of whether \( y_{it} - d \geq 0 \) and \( y_{is} \geq 1 - d \):\(^{15}\)

- **Case 1** (\( y_{it} - d \geq 0 \) and \( y_{is} \geq 1 - d \)): In this case, \( \max\{y_{it} - d, 0\} - \min\{y_{is}, 1 - d\} = (y_{it} - d) - (1 - d) = y_{it} - 1 \).
- **Case 2** (\( y_{it} - d \leq 0 \) and \( y_{is} \geq 1 - d \)): In this case, \( \max\{y_{it} - d, 0\} - \min\{y_{is}, 1 - d\} = 0 - (1 - d) = d - 1 \).
- **Case 3** (\( y_{it} - d \geq 0 \) and \( y_{is} \leq 1 - d \)): In this case, \( \max\{y_{it} - d, 0\} - \min\{y_{is}, 1 - d\} = (y_{it} - d) - y_{is} = y_{it} - y_{is} - d \).
- **Case 4** (\( y_{it} - d \leq 0 \) and \( y_{is} \leq 1 - d \)): In this case, \( \max\{y_{it} - d, 0\} - \min\{y_{is}, 1 - d\} = 0 - y_{is} = -y_{is} \).

Case 3 corresponds to values of \( d \) close to (or at) \( 0 \). Specifically, the region for Case 3 is \((0, \min\{y_{it}, 1 - y_{is}\}) = (0, c_3)\). Noting that \( c_2 \leq 0 \), it is clear that \( \max\{y_{it} - d, 0\} - \min\{y_{is}, 1 - d\} = y_{it} - y_{is} - d \) is consistent with the definition of \( u \) in equation (8).

Case 2 corresponds to values of \( d \) close to (or at) \( 1 \), specifically the region \((\max\{y_{it}, 1 - y_{is}\}, 1) = (c_4, 1)\), and it is again clear that \( u \) delivers \( \max\{y_{it} - d, 0\} - \min\{y_{is}, 1 - d\} \).

The region that defines the other two cases is \((\min\{y_{it}, 1 - y_{is}\}, \max\{y_{it}, 1 - y_{is}\}) = (c_3, c_4)\). Cases 1 and 4 give different expressions depending on whether \( y_{it} \) or \( 1 - y_{is} \) is larger, but these

\(^{15}\)Since both the difference in the re-censored residuals and \( u \) are continuous in \( d \), it is not necessary to distinguish between closed and open intervals in the following discussion.

35
expressions correspond exactly to the two cases for the max in the definition of $u$ in (8) over the interval $(c_3, c_4)$.

The case $d < 0$ is dealt with in exactly the same manner.

We now turn to the question of why

$$E[ U(y_{it}, y_{is}, (x_{it} - x_{is})' b) ]$$

is uniquely minimized at the true $\beta$. This is the key for consistency of the proposed estimator.

The result follows from the following lemma

**Lemma 3** Suppose

$$y_{i1} = \text{mami} \{0, \delta + \varepsilon_{i1}, 1\}$$

and

$$y_{i2} = \text{mami} \{0, \varepsilon_{i2}, 1\}$$

where $\varepsilon_{i1}$ and $\varepsilon_{i2}$ are identically distributed random variables with support on the whole real line.

Then

$$\arg \max_{d \in [-1, 1]} E[ U(y_{i1}, y_{i2}, d) ] = \begin{cases} -1 & \text{if } \delta \leq -1 \\ \delta & \text{if } -1 < \delta < 1 \\ 1 & \text{if } \delta \geq 1 \end{cases}$$

**Proof:**

For $1 \geq d \geq 0$

$$E[ u(y_{i1}, y_{i2}, d) ] = E[ \max \{ y_{i1} - d, 0 \} - \min \{ y_{i2} + d, 1 - d \} ]$$

$$= E[ \max \{ \text{mami} \{0, \delta + \varepsilon_{i1}, 1\} - d, 0 \} - \min \{ \text{mami} \{0, \varepsilon_{i2}, 1\} + d, 1 - d \} ]$$

$$= E[ \max \{ \text{mami} \{0 - d, \delta - d + \varepsilon_{i1}, 1 - d\} , 0 \} - \text{mami} \{0, \varepsilon_{i2}, 1 - d\} ]$$

$$= E[ \text{mami} \{0, \delta - d + \varepsilon_{i1}, 1 - d\} - \text{mami} \{0, \varepsilon_{i2}, 1 - d\} ]$$

If $\varepsilon_{i1}$ (and $\varepsilon_{i2}$) have full support, then this is negative for $d > \delta$ and positive for $d < \delta$.

For $-1 \leq d \leq 0$

$$E[ u(y_{i1}, y_{i2}, d) ] = E[ \min \{ y_{i1}, 1 + d \} - \max \{ y_{i2} + d, 0 \} ]$$

$$= E[ \min \{ \text{mami} \{0, \delta + \varepsilon_{i1}, 1\} + d \} - \max \{ \text{mami} \{0, \varepsilon_{i2}, 1\} + d, 0\} ]$$

$$= E[ \text{mami} \{0, \delta + \varepsilon_{i1}, 1 + d\} - \max \{ \text{mami} \{d, \varepsilon_{i2} + d, 1 + d\} , 0\} ]$$

$$= E[ \text{mami} \{0, \delta + \varepsilon_{i1}, 1 + d\} - \text{mami} \{0, \varepsilon_{i2} + d, 1 + d\} ]$$

36
If \( \varepsilon_{i1} \) (and \( \varepsilon_{i2} \)) have full support, then this is negative for \( d > \delta \) and positive for \( d < \delta \).

Since

\[
E \left[ U' \left( y_{i1}, y_{i2}, d \right) \right] = E \left[ u \left( y_{i1}, y_{i2}, d \right) \right]
\]

the argument above shows that

\[
\arg \max_{d \in [-1,1]} E \left[ U \left( y_{i1}, y_{i2}, d \right) \right] = \begin{cases} 
-1 & \text{if } \delta \leq -1 \\
\delta & \text{if } -1 < \delta < 1 \\
1 & \text{if } \delta \geq 1
\end{cases}
\]

**Corollary 4** Consider the model

\[
y_{it}^* = x_{it}' \beta + \varepsilon_{it}
\]

\[
y_{it} = \begin{cases} 
0 & \text{if } y_{it}^* < 0 \\
y_{it}^* & \text{if } 0 \leq y_{it}^* \leq 1 \\
1 & \text{if } y_{it}^* > 1
\end{cases}
\]

for \( t = 1,2. \) If \( \varepsilon_{it} \) is stationary conditional on \((x_{i1}, x_{i2})\) with support on the whole real line, then the set of solutions to

\[
\max_b E \left[ U \left( y_{i1}, y_{i2}, \text{mami} \left\{ -1, (x_{i1} - x_{i2})' b, 1 \right\} \right) \right]
\]

is

\[
\{ b : P \left( \text{mami} \left\{ -1, (x_{i1} - x_{i2})' b, 1 \right\} = \text{mami} \left\{ -1, (x_{i1} - x_{i2})' \beta, 1 \right\} \right) = 1 \}
\]

The Corollary above requires that the errors are stationary conditional on the regressors. This is much more general than the usual assumption that the individual—specific effect and the contemporaneous errors interact additively. To see that \( E \left[ U_\Xi \left( y_{it}, y_{is}, (x_{it} - x_{is})' b \right) \right] \) in (12) is uniquely minimized, it is convenient to assume that \( \varepsilon_{it} \) and \( \varepsilon_{is} \) are independent and identically distributed conditional on some individual—specific effect, \( v_i \). With this assumption, the argument for why \( E \left[ U_\Xi \left( y_{it}, y_{is}, (x_{it} - x_{is})' b \right) \right] \) is uniquely minimized follows essentially the same logic as above. Specifically, when \( 1 \geq d \geq 0 \)

\[
\xi \left( u \left( y_{it}, y_{is}, d \right) \right) = \xi \left( \text{mami} \left\{ 0, \delta - d + \varepsilon_{it}, 1 - d \right\} - \text{mami} \left\{ 0, \varepsilon_{is}, 1 - d \right\} \right)
\]

Let \( \delta = (x_{it} - x_{is})' \beta. \)
If \( \varepsilon_{it} \) and \( \varepsilon_{is} \) are independent and identically distributed conditional on \( v_i \), then

\[
E [\xi (u (y_{it}, y_{is}, \delta))] | x_{it}, x_{is} = E [E [\xi (u (y_{it}, y_{is}, \delta))] | v_i, x_{it}, x_{is}] | x_{it}, x_{is}]
\]

\[
= E [E [\xi (\text{mami} \{0, \varepsilon_{it}, 1 - \delta\} - \text{mami} \{0, \varepsilon_{is}, 1 - \delta\})] | v_i, x_{it}, x_{is}] | x_{it}, x_{is} = 0
\]

because \( \text{mami} \{0, \varepsilon_{it}, 1 - \delta\} - \text{mami} \{0, \varepsilon_{is}, 1 - \delta\} \) is symmetrically distributed conditional on \( v_i \), and \( \xi (\cdot) \) is an odd function.

For \( d > \delta \)

\[
\text{mami} \{0, \delta - d + \varepsilon_{it}, 1 - d\} - \text{mami} \{0, \varepsilon_{is}, 1 - d\} 
\]

\[
\leq \text{mami} \{0, \varepsilon_{it}, 1 - d\} - \text{mami} \{0, \varepsilon_{is}, 1 - d\}
\]

with probability 1 (conditional on \( v_i \)), and since \( \xi \) is increasing

\[
E [\xi (\text{mami} \{0, \delta - d + \varepsilon_{it}, 1 - d\} - \text{mami} \{0, \varepsilon_{is}, 1 - d\})] | v_i, x_{it}, x_{is}] 
\]

\[
\leq E [\xi (\text{mami} \{0, \varepsilon_{it}, 1 - d\} - \text{mami} \{0, \varepsilon_{is}, 1 - d\})] | v_i, x_{it}, x_{is} = 0
\]

The line of argument is the same when \( d < \delta \). Strict inequalities follow from a full support assumption on \( \varepsilon_{i1} \) (and \( \varepsilon_{i2} \)).

We therefore have that \( E[U_{\Xi} (y_{it}, y_{is}, d) | x_{it}, x_{is}] \) is decreasing to the left of \((x_{it} - x_{is})'\beta \) and increasing to the right. Hence it is minimized at \( d = (x_{it} - x_{is})'\beta \). Subject to a rank condition, this implies that \( E[U_{\Xi} (y_{it}, y_{is}, (x_{it} - x_{is})'b) | x_{it}, x_{is}] \) is minimized at \( b = \beta \).

### 7.2 Two–Sided Truncation

The following Lemma (combined with the obvious rank–condition) establishes that minimization of \( E[R_{\Xi} (y_{i1}, y_{i2}, (x_{11} - x_{21})'b)] \) will identify \( \beta \) if the distribution of \( \varepsilon \) is log–concave. This is the assumption that was made in a number of other papers (including Honoré (1992), Honoré and Powell (1994) and Abrevaya (1999b); see also the discussion in Chen (forthcoming)). In Section 2.2, we only consider the case with (two–sided) truncation at 0 and 1. It is just as easy to prove identification for general individual– and time–specific truncation points. In the following we therefore denote the truncation points by \( L \) and \( U \).

**Lemma 5** Let \( (L, U) \) be a vector of random variables such that \( L < U \) with probability 1. Assume that \( \varepsilon \) is independent of \( (L, U) \) and has a continuous, log–concave distribution with support on the whole real line. Let \( y_{it} = \delta_{it} + \varepsilon_{it} \) for some real number, \( \delta_{it} \), and consider two draws \((y_{i1}, L_{i1}, U_{i1})\) and
(y_{i2}, L_{i2}, U_{i2}) from the distribution of (y, L, U) conditional on \( L < y < U \). then \( E[R_\Xi(y_{i1}, y_{i2}, d)] \) is uniquely minimized at \( d = \delta_{i1} - \delta_{i2} \).

### 7.3 Proof of Lemma 5.

Let \( E \) denote expectation conditional on truncation and \( E^* \) in population.

The moment condition can then be written as

\[
E[x (\Delta y_i - \Delta d_i) 1 \{L_{i2} - d_{i2} \leq y_{i1} - d_{i1} \leq U_{i2} - d_{i2} \} 1 \{L_{i1} - d_{i1} \leq y_{i2} - d_{i2} \leq U_{i1} - d_{i1} \}] \\
= E[x (\Delta y_i - \Delta d_i) 1 \{L_{i2} + \Delta d_i \leq y_{i1} \leq U_{i2} + \Delta d_i \} 1 \{L_{i1} - \Delta d_i \leq y_{i2} \leq U_{i1} - \Delta d_i \}] \\
= E[x (\Delta \varepsilon_i + \Delta \delta_i - \Delta d_i) 1 \{L_{i2} - (\Delta \delta_i - \Delta d_i) \leq \varepsilon_{i1} + \delta_{i2} \leq U_{i2} - (\Delta \delta_i - \Delta d_i) \} 1 \{L_{i1} + (\Delta \delta_i - \Delta d_i) \leq \varepsilon_{i2} + \delta_{i1} \leq U_{i1} + (\Delta \delta_i - \Delta d_i) \}]
\]

where \( \Delta a_i = a_{i1} - a_{i2} \). Letting \( \kappa_i = (\Delta \delta_i - \Delta d_i) \)

\[
= E[x (\Delta \varepsilon_i + \kappa_i) 1 \{L_{i2} - \delta_{i2} - \kappa_i \leq \varepsilon_{i1} \leq U_{i2} - \delta_{i2} - \kappa_i \} 1 \{L_{i1} - \delta_{i1} + \kappa_i \leq \varepsilon_{i2} \leq U_{i1} - \delta_{i1} + \kappa_i \}] \\
= E^*[x (\Delta \varepsilon_i + \kappa_i) 1 \{L_{i2} - \delta_{i2} - \kappa_i \leq \varepsilon_{i1} \leq U_{i2} - \delta_{i2} - \kappa_i \} 1 \{L_{i1} - \delta_{i1} + \kappa_i \leq \varepsilon_{i2} \leq U_{i1} - \delta_{i1} + \kappa_i \} \\
1 \{L_{i1} \leq y_{i1} \leq U_{i1} \} 1 \{L_{i2} \leq y_{i2} \leq U_{i2} \} 1 \{L_{i1} \leq y_{i1} \leq U_{i1}, L_{i2} \leq y_{i2} \leq U_{i2} \}] \\
1 \{L_{i1} \leq y_{i1} \leq U_{i1}, L_{i2} \leq y_{i2} \leq U_{i2} \}.
\]

It suffices to show that this is nonpositive for \( \kappa_i < 0 \), strictly negative in a neighborhood to the left of 0, 0 for \( \kappa_i = 0 \), nonnegative for \( \kappa_i > 0 \), and strictly positive in a neighborhood to the right of 0. Now consider the term

\[
E^*[x (\Delta \varepsilon_i + \kappa_i) 1 \{L_{i2} - \delta_{i2} - \kappa_i \leq \varepsilon_{i1} \leq b_j - \delta_{i2} - \kappa_i \} 1 \{L_{i1} - \delta_{i1} + \kappa_i \leq \varepsilon_{i2} \leq U_{i1} - \delta_{i1} + \kappa_i \} \\
1 \{L_{i1} - \delta_{i1} \leq \varepsilon_{i1} \leq U_{i1} - \delta_{i1} \} 1 \{L_{i2} - \delta_{i2} \leq \varepsilon_{i2} \leq b_j - \delta_{i2} \}].
\]

Defining \( w_{i1} = \frac{1}{2} (\varepsilon_{i1} - \varepsilon_{i2}) \) and \( w_{i2} = \frac{1}{2} (\varepsilon_{i1} + \varepsilon_{i2}) \) (so \( \varepsilon_{i1} = (w_{i1} + w_{i2}) \) and \( \varepsilon_{i2} = w_{i2} - w_{i1} \) and \( \Delta \varepsilon_i = 2w_{i1} \))
\[ E^*[\xi(\Delta \varepsilon_i + \kappa_i)] 1 \{ L_{i2} - \delta_{i2} - \kappa_i \leq \varepsilon_{i1} \leq U_{i2} - \delta_{i2} - \kappa_i \} 1 \{ L_{i1} - \delta_{i1} + \kappa_i \leq \varepsilon_{i2} \leq U_{i1} - \delta_{i1} + \kappa_i \} \\
1 \{ L_{i1} - \delta_{i1} \leq \varepsilon_{i1} \leq U_{i1} - \delta_{i1} \} 1 \{ L_{i2} - \delta_{i2} \leq \varepsilon_{i2} \leq U_{i2} - \delta_{i2} \} \\
= E^*[\xi(2w_{i1} + \kappa_i)] 1 \{ L_{i2} - \delta_{i2} - \kappa_i \leq w_{i1} + w_{i2} \leq U_{i2} - \delta_{i2} - \kappa_i \} \\
1 \{ L_{i1} - \delta_{i1} + \kappa_i \leq w_{i2} - w_{i1} \leq U_{i1} - \delta_{i1} + \kappa_i \} \\
1 \{ L_{i1} - \delta_{i1} \leq w_{i1} + w_{i2} \leq U_{i1} - \delta_{i1} \} \\
1 \{ L_{i2} - \delta_{i2} \leq w_{i2} - w_{i1} \leq U_{i2} - \delta_{i2} \} \\
= E^*[\xi(2w_{i1} + \kappa_i)] \\
1 \left\{ \max \left\{ L_{i2} - \delta_{i2} - w_{i2} - \frac{1}{2} \kappa_i, -U_{i1} + \delta_{i1} + w_{i2} - \frac{1}{2} \kappa_i, L_{i1} - \delta_{i1} - w_{i2} + \frac{1}{2} \kappa_i, -U_{i2} + \delta_{i2} + w_{i2} + \frac{1}{2} \kappa_i \right\} \right\} \\
\leq w_{i1} + \frac{1}{2} \kappa_i \leq \\
\min \left\{ \frac{\varepsilon_{i2} - \delta_{i2} - \frac{1}{2} \kappa_i, -U_{i1} + \delta_{i1} + w_{i2} - \frac{1}{2} \kappa_i, U_{i1} - \delta_{i1} - w_{i2} + \frac{1}{2} \kappa_i, -L_{i2} + \delta_{i2} + w_{i2} + \frac{1}{2} \kappa_i \right\} \\
= E^*[\xi(2w_{i1} + \kappa_i)] 1 \left\{ -c_i \leq w_{i1} + \frac{1}{2} \kappa_i \leq c_i \right\} \\
= E^*[E^*[\xi(2w_{i1} + \kappa_i)] 1 \left\{ -c_i \leq w_{i1} + \frac{1}{2} \kappa_i \leq c_i \right\} w_{i2}, L_{i1}, U_{i1}, L_{i2}, U_{i2}] \\
\text{where } c_i = \min \left\{ U_{i2} - \delta_{i2} - w_{i2} - \frac{1}{2} \kappa_i, -L_{i1} + \delta_{i1} + w_{i2} - \frac{1}{2} \kappa_i, U_{i1} - \delta_{i1} - w_{i2} + \frac{1}{2} \kappa_i, -L_{i2} + \delta_{i2} + w_{i2} + \frac{1}{2} \kappa_i \right\}. \\
\text{Strict log–concavity of } \varepsilon_{it}, \text{ implies that } w_{i1} \text{ is strictly unimodal and symmetric conditional on } w_{i2}. \\
40
It therefore follows that

\[
E^* \left[ E^* \left[ \xi(2w_{i1} + \kappa_i) 1 \left\{ -c_i \leq w_{i1} + \frac{1}{2} \kappa_i \leq c_i \right\} \right| w_{i2}, L_{i1}, U_{i1}, L_{i2}, U_{i2} \right] \right]
\]

\[
\begin{align*}
\leq 0 & \text{ if } \kappa_i < 0 \\
< 0 & \text{ if } \kappa_i < 0, \quad P(-c_i \leq w_{i1} + \frac{1}{2} \kappa_i \leq c_i) > 0 \\
= 0 & \text{ if } \kappa_i = 0 \\
\geq 0 & \text{ if } \kappa_i > 0 \\
> 0 & \text{ if } \kappa_i > 0, \quad P(-c_i \leq w_{i1} + \frac{1}{2} \kappa_i \leq c_i) > 0.
\end{align*}
\]

Since the \(\varepsilon\)'s (and hence the \(w_i\)'s) are continuous, the condition that \(P(-c_i \leq w_{i1} + \frac{1}{2} \kappa_i \leq c_i) > 0\) will be satisfied if \(P(c_i > 0) > 0\).

We will next show that for \(\kappa_i \in (0, k)\), \(P(c_i > 0) > 0\) for some \(k > 0\). Note that

\[
1 \{c_i > 0\} = 1 \left\{ -L_{i1} + \delta_{i1} + w_{i2} - \frac{1}{2} \kappa_i > 0 \right\} \cdot 1 \left\{ U_{i1} - \delta_{i1} - w_{i2} + \frac{1}{2} \kappa_i > 0 \right\}
\]

\[
\cdot 1 \left\{ -L_{i2} + \delta_{i2} + w_{i2} + \frac{1}{2} \kappa_i > 0 \right\} \cdot 1 \left\{ U_{i2} - \delta_{i2} - w_{i2} - \frac{1}{2} \kappa_i > 0 \right\}
\]

\[
= 1 \left\{ w_{i2} > L_{i1} - \delta_{i1} + \frac{1}{2} \kappa_i \right\} \cdot 1 \left\{ U_{i1} - \delta_{i1} + \frac{1}{2} \kappa_i > w_{i2} \right\}
\]

\[
\cdot 1 \left\{ w_{i2} > L_{i2} - \delta_{i2} - \frac{1}{2} \kappa_i \right\} \cdot 1 \left\{ U_{i2} - \delta_{i2} - \frac{1}{2} \kappa_i > w_{i2} \right\}
\]

\[
= 1 \left\{ U_{i1} - \delta_{i1} + \frac{1}{2} \kappa_i > w_{i2} > L_{i1} - \delta_{i1} + \frac{1}{2} \kappa_i \right\} \cdot 1 \left\{ U_{i2} - \delta_{i2} - \frac{1}{2} \kappa_i > w_{i2} > L_{i2} - \delta_{i2} - \frac{1}{2} \kappa_i \right\}.
\]

This will have positive probability provided that

\[
P\left( L_{i1} - \delta_{i1} + \frac{1}{2} \kappa_i < U_{i2} - \delta_{i2} - \frac{1}{2} \kappa \right) > 0
\]

which follows from Lemma 6.

**Lemma 6** If \((U_1, V_1)\) and \((U_2, V_2)\) are two independent draws of a random vector \((U, V)\) with \(P(U < V) = 1\), then there exists a \(k > 0\) such that for \(0 < \kappa < k\), \(P(U_1 < V_2 - \kappa) > 0\).

**Proof.** Since \(V - U > 0\) with probability 1, there exists an \(m > 0\) such that \(P(V - U > m) \geq \frac{1}{2}\). Now consider the space \(\{(u, v) : v - u > m\}\). This can be divided into a countable number of regions, \(A_k\), such that for \((u_1, v_1), (u_2, v_2) \in A_k\), \(|u_1 - u_2| < \frac{m}{2}\) and \(|v_1 - v_2| < \frac{m}{2}\). Since \(P(V - U > m) = \frac{1}{2}\), at least one of these regions, \(A_k\), must have positive probability. Hence there is positive probability that \((U_1, V_1)\) and \((U_2, V_2)\) in the statement of the lemma both belong to this \(A_k\), in which case \(U_1 < V_1 - m < V_2 - \frac{m}{2}\).
## 8 Appendix 2: Empirical Results

Table 1: Marginal tax rates before and after implementation of the 1987 tax reform

<table>
<thead>
<tr>
<th>Before Reform</th>
<th>After Reform</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tax bracket</td>
<td>Tax bracket</td>
</tr>
<tr>
<td>Earnings + Cap inc</td>
<td>Earnings inc &lt; 0</td>
</tr>
<tr>
<td>0-113</td>
<td>0 – 130</td>
</tr>
<tr>
<td>113-186</td>
<td>130 – 200</td>
</tr>
<tr>
<td>186-</td>
<td>200–</td>
</tr>
<tr>
<td>Tax ceiling</td>
<td>73</td>
</tr>
</tbody>
</table>

Note: M is the local government tax rate. Threshold values for the tax brackets are given in 1000 DKK. Thresholds are adjusted yearly. Threshold values used in the table are for 1986 (before the reform) and 1987 (after the reform). The marginal tax rates refer to personal income (as opposed to household income).

(1) The tax brackets for positive net capital income refer to the sum of earnings and positive net capital income.

After the reform positive capital income is taxed progressively up to the first threshold, 130,000 DKK. For a married couple the progression threshold is 260,000 based on the sum of their joint positive net capital income and earnings.

(2) The 68 percent tax ceiling applies only to the top bracket.
Table 2: Random and Fixed Effects Censored Regression Estimates of the Portfolio Share of Bonds in Financial Wealth.

<table>
<thead>
<tr>
<th></th>
<th>Fixed Effects</th>
<th>Random Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>MTR capital income</td>
<td>−0.130</td>
<td>−0.206</td>
</tr>
<tr>
<td></td>
<td>(0.071)</td>
<td>(0.045)</td>
</tr>
<tr>
<td>Ln(Financial Assets)</td>
<td>0.177</td>
<td>0.085</td>
</tr>
<tr>
<td></td>
<td>(0.008)</td>
<td>(0.003)</td>
</tr>
<tr>
<td>D85</td>
<td>−0.214</td>
<td>−0.157</td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.006)</td>
</tr>
<tr>
<td>D86</td>
<td>−0.314</td>
<td>−0.248</td>
</tr>
<tr>
<td></td>
<td>(0.009)</td>
<td>(0.006)</td>
</tr>
<tr>
<td>D87</td>
<td>−0.318</td>
<td>−0.285</td>
</tr>
<tr>
<td></td>
<td>(0.013)</td>
<td>(0.008)</td>
</tr>
<tr>
<td>D88</td>
<td>−0.383</td>
<td>−0.331</td>
</tr>
<tr>
<td></td>
<td>(0.013)</td>
<td>(0.008)</td>
</tr>
<tr>
<td>Constant</td>
<td>—</td>
<td>0.047</td>
</tr>
</tbody>
</table>

| # households/observations | 8,577 / 42,885 |
| # left/right censored obs | 7,529 / 15,655 |
| Test of Parameter Equality (d.f.) | 167 (6) |

Standard errors in parenthesis