cemmap
centre for microdata methods and practice

Set identification with Tobin regressors

Victor Chernozhukov
Roberto Rigobon
Thomas M. Stoker

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP12/09

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

An ESRC Research Centre

SET IDENTIFICATION WITH TOBIN REGRESSORS

VICTOR CHERNOZHUKOV, ROBERTO RIGOBON, AND THOMAS M. STOKER

Abstract. We give semiparametric identification and estimation results for econometric models with a regressor that is endogenous, bound censored and selected, called a Tobin regressor. First, we show that true parameter value is set identified and characterize the identification sets. Second, we propose novel estimation and inference methods for this true value. These estimation and inference methods are of independent interest and apply to any problem where the true parameter value is point identified conditional on some nuisance parameter values that are set-identified. By fixing the nuisance parameter value in some suitable region, we can proceed with regular point and interval estimation. Then, we take the union over nuisance parameter values of the point and interval estimates to form the final set estimates and confidence set estimates. The initial point or interval estimates can be frequentist or Bayesian. The final set estimates are set-consistent for the true parameter value, and confidence set estimates have frequentist validity in the sense of covering this value with at least a prespecified probability in large samples. We apply our identification, estimation, and inference procedures to study the effects of changes in housing wealth on household consumption. Our set estimates fall in plausible ranges, significantly above low OLS estimates and below high IV estimates that do not account for the Tobin regressor structure.

Date: February 2009.

The authors wish to thank the participants of seminars at UCL and Cambridge University, and CEMMAP conference in June 2007 for helpful comments. We particularly thank Richard Blundell, Andrew Chesher, Whitney Newey, Richard Smith and Elie Tamer.

1. Introduction

In economic surveys, financial variables are often mismeasured in nonrandom ways. The largest values of household income and wealth are often eliminated by top-coding above prespecified threshold values. Income and wealth are also typically reported as nonnegative, which may neglect large transitory income losses, large debts (negative components of wealth), or other aspects that could be modeled as bottom-coding below a prespecified threshold value. In addition to mismeasurement problems related to upper and lower bounds, income and wealth are often missing due to nonresponse.[1]

These measurement problems are particularly onerous when they obscure key features of the economic process under study. For instance, suppose one is studying the impact of liquidity constraints on consumption spending using data from individual households. It is a widespread practice to drop all household observations when there are top-coded income values. However, that practice seemingly eliminates households that are the least affected by liquidity constraints, which would provide the most informative depiction of baseline consumption behavior. Likewise, if one is studying the household demand for a luxury good, the most informative data is from rich households, who, for confidentiality reasons, often won't answer detailed questions about their income and wealth situations.

These problems can be compounded when the observed financial variable is itself an imperfect proxy of the economic concept of interest. For instance, suppose one is studying the impact of the availability of cash on a firm's investment decisions. Only imperfect proxies of "cash availability" are observed in balance sheet data, such

[1] For many surveys, extensive imputations are performed to attempt to "fill in" mismeasured or unrecorded data often in ways that are difficult to understand. For instance, in the U.S.Consumer Expenditure (CEX) survey, every component of income is top-coded;namely wages, interest, gifts, stock dividends and gains, retirement income,transfers, bequests, etc., and there is no obvious relation between the top-coding on each component and the top-coding on total income. The CEX makes extensive use of ad hoc multiple imputation methods to fill in unrecorded income values.

as whether the firm has recently issued dividends. The mismeasurement of those proxies is not random; positive dividends indicate positive cash availability but zero dividends can indicate either mild cash availability or severe cash constraints. Thus, observed dividends represent a censored (bottom coded at zero) version of the cash availability status of a firm.

The study of mismeasurement due to censoring and selection was initiated by the landmark work of Tobin(1958). In the context of analyzing expenditures on durable goods, Tobin showed how censoring of a dependent variable induced biases, and how such bias could be corrected in a parametric framework This work has stimulated an enormous literature on parametric and semiparametric estimation with censored and selected dependent variables. The term "Tobit Model" is common parlance for a model with a censored or truncated dependent variable.

We study the situation where an endogenous regressor is censored or selected. This also causes bias to arise in estimation; bias whose sign and magnitude varies with the mismeasurement process as well as the estimation method used (Rigobon and Stoker (2006a)). With reference to the title, we use the term "Tobin regressor" to refer to a regressor that is bound censored, selected and endogenous.

When the mismeasurement of the regressor is exogenous to the response under study,[2] consistent estimation is possible by using only "complete cases," or estimating with a data sample that drops any observations with a mismeasured regressor.[3] When endogenous regressors are censored or selected, the situation is considerably more complicated. Dropping observations with a mismeasured regressor creates a selected sample for the response under study. Standard instrumental variables methods are

[2]That is, both the correctly measured regressor and the censoring/selection process is exogenous.

[3]The existence of consistent estimates allows for tests of whether bias is evident in estimates computed from the full data sample. See Rigobon and Stoker (2006b) for regression tests and Nicholetti and Peracchi (2005) for tests in a GMM framework.

biased when computed from the full data sample, and are also biased and inconsistent when computed using the "complete cases" only.

In this paper, we provide a full identification analysis and estimation solution for situations with Tobin regressors. We show how the true parameter value is set identified and characterize the identification sets. Second, we propose novel estimation and inference methods for this true value. These estimation and inference methods are of independent interest and apply to any problem where the true parameter value is point identified conditional on some nuisance parameter values that are set-identified. Indeed, fixing the nuisance parameter value in some suitable region, we can proceed with regular point and interval estimation. Then, we take the union over nuisance parameter values of the point and interval estimates to form the final set estimates and confidence set estimates. The initial point or interval estimates can be frequentist or Bayesian. The final set estimates are set-consistent for the true parameter value, and confidence set estimates have frequentist validity in the sense of covering this value with at least a prespecified probability in large samples.

Our approach is related to several contributions in the literature. Without censoring or selection, our framework is in line with work on nonparametric estimation of endogeous model with non-additivity, as developed by Altonji and Matzkin (2005), Chesher (2003), Imbens and Newey (2005) and Chernozhukov and Hansen (2005), among others. Our accommodation of endogeneity uses the control function approach, as laid out by Blundell and Powell (2003). In terms of dealing with censoring, we follow Powell's (1984) lead in using monotonicity assumptions together with quantile regression methods (see Koenker's(2005) excellent review of quantile regression). Our inference results complement the inferential procedures proposed in Chernozhukov, Hong and Tamer (2007) and other literature.

There is a great deal of literature on mismeasured data, some focused on regressors. Foremost is Manski and Tamer (2002), who use monotonicity restrictions to propose consistent estimation with interval data. For other contributions in econometrics, see Ai (1997), Chen, Hong and Tamer (2005), Chen, Hong and Tarozzi (2004), Liang, Wang, Robins and Carroll (2004), Tripathi (2004), among many others, which are primarily concerned with estimation when data is missing at random. The large literature in statistics on missing data is well surveyed by Little and Rubin (2002), and work focused on mismeasured regressors is surveyed by Little (1992). The exposition proceeds by introducing our approach within a simple framework, in Section 2. Section 3 gives our general framework and a series of generic results on identification and estimation. Section 4 contains an empirical application, where we show how accommodating censoring and selection gives rise to a much larger estimates of the impact of housing wealth on consumption.

2. A Simplified Framework and the Basic Identification Approach

2.1. **A Linear Model with a Tobin Regressor.** We introduce the ideas in a greatly simplified setting, where a linear model is the object of estimation. We do this in order the highlight the main concepts of our approach. Our main results do not rely on a linear model, but are based on a very general (parametric or nonparametric) framework. We spell this out in in Section 3.

Consider the estimation of a linear model with a (potentially) endogenous regressor:

$$Y = X^*\alpha + U^* \tag{2.1}$$

$$X^* = Z'\gamma + V^* \tag{2.2}$$

$$U^* = \beta V^* + \varepsilon \tag{2.3}$$

where

$$\varepsilon \text{ is mean (or median or quantile) independent of } (V^*, X^*), \tag{2.4}$$

$$V^* \text{ is median independent of } Z \tag{2.5}$$

Here X^* is the uncensored regressor, which is endogenous when $\beta \neq 0$, and Z represents valid instruments (without censoring or selection). We make no further assumption on the distribution of ε or V^*.

The regressor X^* is not observed. Rather, we observe a censored version of X^*:

$$X = I\{R=1\}I\{X^*>0\}X^* \tag{2.6}$$

$$R = \begin{cases} 1 & \text{with prob} \quad 1-\pi \\ 0 & \text{with prob} \quad \pi \end{cases}, \text{ independent of } Z. \tag{2.7}$$

The observed X matches X^* unless one of two sources of censoring arises, in which case $X = 0$. . The first source is bound censoring, which occurs when $X^* \leq 0$ or $V^* \leq -Z'\gamma$. The second source is an independent censoring method, which selects $X = 0$ when $R = 0$ (or equivalently, selects to observe positive X^* when $R = 1$). We sometimes refer to the probability π as the "selection" probability in the following; it would be more complete to refer to π as "the probability of independent selection for censoring."

In this sense, the observed regressor X is *censored*, *selected* and *endogenous*, which we refer to as a *Tobin regressor*. There exists an instrument Z for the uncensored regressor X^*, but that instrument will typically be correlated with $X-X^*$. Therefore, Z will not be a valid instrument if X is used in place of X^* in the response equation (2.1).

In our simplified setup, censoring is modeled with the lower bound (bottom-coding) of 0, but top-coding or different bound values are straightforward to incorporate. The selection probability π is taken as constant here, but is allowed to vary with covariates in our general framework. We assume that $P\left[Z'\gamma > 0\right] > 0$, which is both convenient and empirically testable.

It is also straightforward to include additional controls in the response equation (2.1). With that in mind, we develop some examples for concreteness.

Example 1. Income and Consumption: Suppose X is income and Y is household consumption expenditure. X is typically endogenous, top-coded and missing for various households.. Bound censoring arises for large income values, and selection refers to missing values, possibly due to households declining to report their income. For instance, if one is estimating a permanent income model of consumption, then X would be observed permanent income (or wealth). If one is investigating excess sensitivity (or liquidity constraints), then X would be observed current income (and the equation would include lagged consumption). Here, the instruments Z could include unanticipated income shocks, lagged income values, and demographic variables that are not included in the consumption equation. Finally, the same censoring issues can arise in an Engel curve analysis, where Y is the expenditure on some commodity and X is total expenditures on all commodities.

Example 2: Dividends and Firm Investment: Suppose X is declared dividends and Y is investment, for individual firms. Here X^* is the level of cash availability (or opposite of cash constraints). Positive dividends X indicate positive cash availability, but zero dividends arises with either mild or severe cash constraints (small or large negative X^*). The instruments Z could include exogenous variables that affect the cost of debt, such as foreign exchange fluctuations.

Example 3: Day Care Expenditures and Female Wages: Suppose you are studying the economic situation faced by single mothers, where Y is expenditure on day care and X is the observed wage rate. X is potentially endogenous (work more to pay for higher quality day care), and is selected due to the labor participation choice. Here, the instruments Z could include job skills and other factors affecting the labor productivity of single mothers, as well as exogenous household income shocks.

2.2. **Basic Identification and Estimation Ideas.** The strategy for identification is to set the amount of selection first, which allows the rest of the model to be identified. That is, suppose we set a value π^* for $\pi = Pr[R = 0]$. The following steps give identification:

1) Note that the conditional median curve $Q_{X^*}(\frac{1}{2}|Z) = Z'\gamma$ is partially identified from the estimable curve

$$Q_X\left(\frac{1}{2}(1-\pi^*) + \pi^*|Z\right) = \max[Z'\gamma, 0], \tag{2.8}$$

since $Z'\gamma > 0$ with positive probability.

2) Given γ, we can estimate the control function

$$V^* = X^* - Z'\gamma = X - Z'\gamma, \tag{2.9}$$

whenever $X > 0$.

3) Given the control function V^*, we can recover the regression function of interest (mean, median or quantile) for the sub-population where $X > 0$ and $Z'\gamma > 0$. For instance, if ε is mean independent of (V^*, X^*), we can estimate the mean regression

$$E[Y|X, V^*] = X'\alpha + \beta V^*. \tag{2.10}$$

If ε is quantile independent of (V^*, X^*), we can estimate the quantile regression

$$Q_Y(\tau|X, V^*) = X'\alpha + \beta V^*. \tag{2.11}$$

4) All of the above parameters depend on the value π^*. We recognize this functional dependence by writing $\alpha(\pi^*)$, $\beta(\pi^*)$, $\gamma(\pi^*)$ for solutions of steps 1), 2), and 3). For concreteness, suppose the particular value $\alpha_0 = \alpha(\pi_0)$ is of interest. If we can determine the set $\mathcal{P}_0$ of all feasible values of π^*, the set

$$\mathcal{A}_0 = \{\alpha(\pi), \pi \in \mathcal{P}_0\}$$

clearly contains α_0. Likewise, if we denote $\theta(\pi) = \{\alpha(\pi), \beta(\pi), \gamma(\pi)\}$, then $\theta_0 = \theta(\pi_0)$ is contained in the set

$$\Theta_0 = \{\theta(\pi), \pi \in \mathcal{P}_0\}.$$

5) It remains to characterize the set $\mathcal{P}_0$. In the absence of further information, this set is given by:

$$\mathcal{P}_0 = [0, \inf_{z \in \text{support}(Z)} \Pr[D > 0 | Z = z]]. \tag{2.12}$$

where

$$D \equiv 1 - I\{R = 1\} I\{X^* > 0\}$$

is the index of observations that are censored. In words, $\mathcal{P}_0$ is the interval containing all values between 0 and the smallest probability of censoring in the population.

This outlines the basic identification strategy. It is clear that point identification is achieved if π is a known value. For instance, if there is only bound censoring (no R term in (2.6)), then $\pi = 0$. Then estimation (step 1) uses median regression to identify the (single) control function needed.

Estimation proceeds by the analogy principle: empirical curves are used in place of the population curves above to form estimators . In Section 3 we justify this for our general framework. We also show how inference is possible by simply constructed confidence sets. That is, suppose α_0 is of interest and the set $\mathcal{P}_0$ is known. Given π, a standard confidence region for $\alpha(\pi)$ is

$$CR_{1-\alpha}(\alpha(\pi)) = [\hat{\alpha}(\pi) \pm c_{1-\alpha} s.e.(\hat{\alpha}(\pi))].$$

We note that a $1 - \alpha$-confidence region for $\alpha_0 = \alpha(\pi_0)$ is simply

$$\cup_{\pi \in \mathcal{P}_0} CR_{1-\alpha}(\alpha(\pi)). \tag{2.13}$$

Reporting such a confidence region is easy, by reporting its largest and smallest elements.

We also discuss adjustments that arise because of the estimation of $\mathcal{P}_0$, the range of selection probabilities. In particular, a confidence region for $\mathcal{P}_0$ can be developed, as well as adjustments for the level of significance of the parameter confidence regions such as (2.13).

2.3. **A Geometric View of Identification and Estimation.** We illustrate the basic idea of identification through a sequence of figures that illustrate a simple one-regressor version of our empirical example. In the first step, we fix a set of values of π in a set from 0 to .08 (the range for the true π_0) and fit a family of censored conditional quantile estimates. Thus, we obtain a family of "first stage" estimates, shown in Figure 1, indexed by the admissible values of π. In the second step, we form a control function V_π using the results of the first step, and then we run mean regressions of Y on X and V_π. The results are indexed by the values of $\pi \in [0, 0.08]$. Thus, we obtain a family of "second stage" estimates, shown in Figure 2, indexed by the admissible values of π. Finally, Figure 3 shows the construction of a conservative though consistent upper bound on π. The illustration here corresponds to roughly to the results obtained in the empirical section of the paper. The panels of Figure 3 show the fitted probabilities of missing data on X. The top panel shows a naive plug-in upper bound on π. The bottom panel shows the upper bound of π adjusted up by the two times standard error times a logarithmic factor in the sample size.

We now turn to our general framework and main results.

3. Generic Set Identification and Inference

3.1. **The General Framework.** The stochastic model we consider is given by the system of quantile equations:

$$Y = Q_Y(U|X^*, W, V) \tag{3.1}$$

$$X^* = Q_{X^*}(V|W, Z) \tag{3.2}$$

where Q_Y is the conditional quantile function of Y given X^*, W, V and Q_{X^*} is the conditional quantile function of X^* given Z. Here U is Skorohod disturbance such that $U \sim U(0,1)|X^*, W, V$, and V is Skorohod disturbance such that $V \sim U(0,1)|W, Z$. The latent true regressor is X^*, which is endogenous when V enters the first equation nontrivially. Z represents "instruments" for X^* and W represents covariates.

The observed regressor X, the *Tobin regressor*, is given by the equation

$$X = I\{R = 1\} I\{X^* > 0\} X^* \tag{3.3}$$

where

$$R = \begin{cases} 1 & \text{with probability} \quad 1 - \pi(W) \\ 0 & \text{with probability} \quad \pi(W) \end{cases} \tag{3.4}$$
$$\text{conditional on } W, Z, V.$$

There are two sources of censoring of X^* to 0. First there is bound censoring, occurring when $X^* \leq 0$. Second is independent selection censoring, which occurs when $R = 0$. As such, X is endogenous, censored and selected.

The model (3.1), (3.2) is quite general, encompassing a wide range of nonlinear models with an endogenous regressor. The primary structural restriction is that the system is triangular; that is, V can enter both (3.1) and (3.2) but U does not enter (3.2). The Skohorod disturbances U and V index the conditional quantiles of Y and X^*. We have by definition that

$$\begin{aligned} U &= F_Y(Y|X^*, W, V) \\ V &= F_{X^*}(X^*|W, Z) \end{aligned}$$

where F_Y is the conditional distribution function of Y given X^*, W, V and F_{X^*} is the conditional distribution function of X^* given W, Z. The random variables U and V provide an equivalent parameterization to the stochastic model as would additive disturbances or other (more familiar) ways capturing randomness. For example, the linear model (2.1)-(2.2) is written in the form of (3.1), (3.2) as

$$\begin{aligned} Y &= X^*\alpha + Q_{U^*}(U|V) \\ X^* &= Z'\gamma + Q_{V^*}(V|Z) \end{aligned}$$

where the additive disturbances U^* and V^* have been replaced by U and V through the equivalent quantile representations $U^* = Q_{U^*}(U|V)$ and $V^* = Q_{V^*}(V|Z)$.

The primary restriction of the *Tobin regressor* is that selection censoring is independent of bound censoring, (conditional on W, Z and V). We have left the selection probability in the general form $\pi(W)$, which captures many explicit selection models. For instance, we could have selection based on threshold crossing. In that case, the selection mechanism is $R \equiv 1\left[W'\delta + \eta \geq 0\right]$, with η an independent disturbance, which implies the selection probability is $\pi(W) = \Pr\left\{\eta < -W'\delta\right\}$.

3.2. **Set Identification without Functional Form Assumptions.** We now state and prove our first main result. We require the following assumption

Assumption 1: *We assume that the systems of equations (3.1)-(3.4) and independence assumptions hold as specified above, and that $v \mapsto Q_{X^*}(v|W, Z)$ is strictly increasing in $v \in (0, 1)$ almost surely.*

Our main identification result is

Proposition 1. *The identification regions for $Q_Y(\cdot|X^*, V, W)$ and $F_Y(\cdot|X^*, V, W)$ on the subregion of the support of (X^*, V, W) implied by $X > 0$ are given by*

$$\mathcal{Q} = \{Q_Y(\cdot|X, V_\pi, W), \pi \in \mathcal{P}\}$$

and

$$\mathcal{F} = \{F_Y(\cdot|X, V_\pi, W), \pi \in \mathcal{P}\}$$

where when $X > 0$

$$V_\pi = \frac{F_X(X|Z,W) - \pi(W)}{1 - \pi(W)}, \tag{3.5}$$

or equivalently when $X > 0$

$$V_\pi = \int_0^1 1\left\{Q_X\left((\pi(W) + (1 - \pi(W))v|W,Z\right) \leq X\right\} dv. \tag{3.6}$$

Finally,

$$\mathcal{P} = \left\{\pi(\cdot) \text{ } measurable : 0 \leq \pi(W) \leq \min_{z \in supp(Z)|W} F_X(0|W,z) \text{ } a.s\right\}. \tag{3.7}$$

Proposition 1 says that given the level of the selection probability $\pi(W)$, we can identify the quantile function of Y with respect to X^* by using the (identified) quantile function of Y with respect to the observed Tobin regressor X, where we shift the argument V to V_π of (3.5,3.6). The identification region is comprised of the quantile functions for all possible values of $\pi(W)$. The proof is constructive, including indicating how the quantiles with respect to X^* and to X are connected. It also makes clear that point identification of the functions is possible where $X > 0$ and $\pi(W)$ is known (or point identified), including the no selection case with $\pi(W) = 0$.

In empirical applications, one is typically not interested in (nonparametrically) estimating the full conditional distribution of Y given X, V, W, but rather in more interpretable or parsimonious features. That is, one wants to estimate

$$\theta(\pi) = \theta\Big(Q(\cdot;\pi)\Big) , \tag{3.8}$$

a functional of π taking values in Θ, where the quantile Q can either be the conditional quantile Q_Y or Q_{X^*}, or equivalently

$$\theta(\pi) = \theta^*\Big(F(\cdot;\pi)\Big) \tag{3.9}$$

where the conditional distribution F is either F_Y or F_{X^*} .The functional $\theta(\pi)$ can represent parameters of a model of Q or F, average policy effects, average derivatives, local average responses and other features (including representing the full original functions Q or F). The following corollary establishes identification of such aspects of interest.

Corollary 1. *The identification region for the functional* $\theta(\pi_0)$ *is*

$$\{\theta(\pi), \pi \in \mathcal{P}\}.$$

Given Proposition 1, the proof of this Corollary is immediate. We now give the proof of our first main result.

Proof of Proposition 1 (Identification):. We follow the logic of the identification steps outlined in the previous section. Suppose we first set a value $\pi(W)$ for $Pr[R = 0|W]$. For $x > 0$, we have that

$$\begin{aligned}
\Pr[X \leq x|W, Z] &= \Pr[R = 0|W, Z] + \Pr[R = 1 \text{ and } X^* \leq x|W, Z] \\
&= \Pr[R = 0|W, Z] + \Pr[R = 1|W, Z] \cdot \Pr[X^* \leq x|W, Z]
\end{aligned}$$

That is,

$$F_X[x|W, Z] = \pi(W) + (1 - \pi(W))F_{X^*}[x|W, Z]$$

In terms of distributions, whenever $X > 0$,

$$V_\pi = F_{X^*}[X|W, Z] = \frac{F_X(X|Z, W) - \pi(W)}{1 - \pi(W)}$$

Thus V_π is identified from the knowledge of $F_X(X|Z, W)$ and $\pi(W)$ whenever $X > 0$. In addition

$$X^* = X.$$

when $X > 0$.

In terms of quantiles,

$$Q_{X^*}(V_\pi|W,Z) = Q_X\left(\frac{F_X(X|Z,W)-\pi(W)}{1-\pi(W)}\middle| W,Z\right) \tag{3.10}$$

$$= Q_X\left((\pi(W)+(1-\pi(W))V_\pi|W,Z\right).$$

This implies that for any $X>0$

$$\begin{aligned} V_\pi &= \int_0^1 1\{Q_{X^*}(v|W,Z)\leq X^*\}dv \\ &= \int_0^1 1\{Q_X\left((\pi(W)+(1-\pi(W))v|W,Z\right)\leq X\}dv \end{aligned}$$

Thus V_π is identified from the knowledge of $Q_X(\cdot|Z,W)$ whenever $X>0$.

Inserting

$$X, V_\pi \text{ for cases } X>0$$

into the outcome equation we have a point identification of the quantile functional

$$Q_Y(\cdot|X,V_\pi,W)$$

over the region implied by the condition $X>0$. This functional is identifiable from the quantile regression of Y on X, V_π, W.

Likewise, we have the point identification of the distributional functional

$$F_Y(\cdot|X,V_\pi,W)$$

over the region implied by the condition $X>0$. This functional is identified either by inverting the quantile functional or by the distributional regression of Y on X, V_π, W.

Now, since the (point) identification of the functions depends on the value $\pi(W)$, by taking the union over all $\pi(\cdot)$ in the class $\mathcal{P}$ of admissible conditional probability functions of W, we have the following identified sets for both quantities:

$$\{Q_Y(\cdot|X,V_\pi,W), \pi(\cdot)\in\mathcal{P}\}$$

and

$$\{F_Y(\cdot|X, V_\pi, W), \pi(\cdot) \in \mathcal{P}\}.$$

The quantities above are sets of functions or correspondences.

It remains to characterize the admissible set $\mathcal{P}$. From the relationship

$$F_X[0|W, Z] = \pi(W) + (1 - \pi(W)) \cdot F_{X^*}[0|W, Z]$$

we have

$$0 \leq \pi(W) = \frac{F_X\,[0|W, Z] - F_{X^*}[0|W, Z]}{1 - F_{X^*}[0|W, Z]} \leq F_X(0|W, Z),$$

where the last observation is by the equalities

$$0 = \min_{0 \leq x \leq F} \left(\frac{F - x}{1 - x} \right) \leq \max_{0 \leq x \leq F} \left(\frac{F - x}{1 - x} \right) = F.$$

Taking the best bound over z, we have

$$0 \leq \pi(W) \leq \min_{z \in \mathcal{Z}|W} F_X(0|W, z),$$

Hence

$$\mathcal{P} = \left\{ \pi(\cdot) \text{ measurable} : \ 0 \leq \pi(W) \leq \min_{z \in \mathcal{Z}|W} F_X(0|W, z) \text{ a.s} \right\}.$$

which demonstrates Proposition 1. $\square$

3.3. **Estimation and Inference.** Our constructive derivation of identication facilitates a general treatment of estimation. Here we present the general results. In the following section, we discuss some particulars of estimation as well as related results in the literature.

Estimation can be based on the analogy principle. Here we consider a plug-in estimator

$$\widehat{\theta}(\pi) = \theta\Big(\widehat{Q}(\cdot;\pi)\Big) \text{ or } \theta^*\Big(\widehat{F}(\cdot;\pi)\Big).$$

where the true quantile or distribution function is replaced by an estimator. We assume that the model structure is sufficiently regular to support a Central Limit

Theorem for $\hat{\theta}(\pi)$ for each value of π, and that estimates of confidence intervals are available for for each π. This is summarized in the following two assumptions.

Assumption 2.1 For each $\pi \in P$, suppose an estimate $\widehat{Q}$ or $\widehat{F}$ is available such that

$$\mathcal{Z}_n(\pi) := A_n(\pi)\Big(\widehat{\theta}(\pi) - \theta(\pi)\Big) \Rightarrow \mathcal{Z}_\infty(\pi);\ \text{for each } \pi \in \mathcal{P}$$

where convergence occurs in some metric space $(B, \|\cdot\|_B)$, where $A_n(\pi)$ is a sequence of scalers, possibly data dependent.

Assumption 2.2 Let

$$c(1-\alpha,\pi) :=\ \alpha\text{-quantile of }\ \|\mathcal{Z}_\infty(\pi)\|_B$$

and suppose that the distribution function of $\|\mathcal{Z}_\infty(\pi)\|_B$ is continuous at $c(1-\alpha, P)$. Estimates are available such that $\widehat{c}(1-\alpha,\pi) \rightarrow_p c(1-\alpha,\pi)$ for each π.

With these assumptions, we can show the following generic result. This shows how to construct confidence intervals when the set $\mathcal{P}$ is known. The answer is simple; construct the confidence intervals for all values of $\pi \in \mathcal{P}$ and take their union.

Proposition 2. *Let*

$$C_{1-\alpha}(\pi) := \Big\{\theta \in \Theta : \Big\|A_n(\pi)\Big(\widehat{\theta}(\pi) - \theta\Big)\Big\|_B \leq \widehat{c}(1-\alpha,\pi)\Big\}.$$

Let

$$CR_{1-\alpha} := \bigcup\nolimits_{\pi\in\mathcal{P}} C_{1-\alpha}(\pi).$$

Then

$$\liminf_{n\to\infty} P\Big\{\theta(\pi_0) \in CR_{1-\alpha}\Big\} \geq 1-\alpha.$$

Proof of Proposition 2. We have that

$$\begin{aligned} P\Big\{\theta(\pi_0) \ &\in\ CR_{1-\alpha}\Big\} \geq P\Big\{\theta(\pi_0) \in CR_{1-\alpha}(\pi_0)\Big\} = P\Big\{\|\mathcal{Z}_\infty(\pi)\|_B \leq \hat{c}(1-\alpha,\pi_0)\Big\} \\ &=\ P\Big\{\|\mathcal{Z}_\infty(\pi)\|_B \leq c(1-\alpha,\pi_0)\Big\} + o\,(1) = 1-\alpha + o\,(1) \end{aligned}$$

where we have used the continuity of the map $c \longmapsto P\Big\{\|\mathcal{Z}_\infty(\pi)\|_B \le c\Big\}$ at $c = c(1-\alpha, \pi_0)$ and the consistency property $\hat{c}(1-\alpha, \pi_0) = c(1-\alpha, \pi_0) + o_p(1)$. $\square$

In applications, the set $\mathcal{P}$ is a nuisance parameter that needs so be estimated, and the above confidence intervals need to be adjusted for that estimation. Estimation of $\mathcal{P}$ poses some new challenges. From (3.7), estimation of $\mathcal{P}$ is equivalent to estimation of the boundary function:

$$\ell(W) = \min_{z\in\mathcal{Z}} F_X[0|W, z].$$

Let $\widehat{\ell}(W)$ be a suitable estimate of this function. One example is

$$\widehat{\ell}(W) = \min_{z\in\mathcal{Z}} \widehat{F}_X[0|W, z]. \tag{3.11}$$

We require that the precision of $\widehat{\ell}(W)$ can be approximated. In particular, we assume that the model structure is sufficiently regular to justify the following assumption:

Assumption. 2.3 Let $\widehat{\kappa}_n(1-\beta)$ and the known scaler $B_n(W)$ to be such that

$$\ell(W) - \widehat{\ell}(W) \le B_n(W)\widehat{\kappa}_n(1-\beta)$$

for all W with probability at least $1-\beta$.

Conservative forms of confidence regions of this type are available from the literature on simultaneous confidence bands. For instance, Assumption 2.3 holds for $\widehat{\ell}(W) = \min_{z\in\mathcal{Z}} \widehat{F}_X[0|W, z]$. if we set

$$\hat{z} = arg\min_{z\in\mathcal{Z}} \widehat{F}_X[0|W, z]$$

and

$$B_n(W) := \Big[\text{s.e.}(\widehat{F}(W, z))\Big]_{z=\hat{z}_0(W)}; \qquad \kappa_n(1) = 2\sqrt{\log n}.$$

Sharper confidence regions for minimized functions are likely available, but their construction is relatively unexplored. For some initial results of this type, see Chernozhukov, Lee and Rosen (2008).

Let $\pi(W)$ belong to the parameter set $\mathcal{P}$. From Assumption 2.3, the confidence region for $\pi(W)$ is given by

$$CR'_{1-\beta} = \{\pi \in \Pi : \pi(W) - \widehat{l}(W) \leq B_n(W)\widehat{\kappa}_n(1-\beta)\} \tag{3.12}$$

We combine this with Proposition 2 to obtain:

Proposition 3. *Let*

$$CR_{1-\alpha} := \bigcup\nolimits_{\pi \in CR'_{1-\beta}} C_{1-\alpha}(\pi).$$

Then

$$\liminf_{n\to\infty} P\Big\{\theta(\pi_0) \in CR_{1-\alpha}\Big\} \geq 1-\alpha-\beta.$$

Proof of Proposition 3. We have that

$$\begin{aligned}
P\Big\{\theta(\pi_0) \ \in \ CR_{1-\alpha}\Big\} &\geq P\Big\{\theta(\pi_0) \in CR_{1-\alpha}(\pi_0) \cap \pi_0 \in CR'_{1-\beta}\Big\} \\
&\geq P\Big\{\theta(\pi_0) \in CR_{1-\alpha}\Big\} - P\Big\{\pi_0 \notin CR'_{1-\beta}\Big\} \\
&\geq P\Big\{\theta(\pi_0) \in CR_{1-\alpha}\Big\} \ - P\Big\{\pi_0(W) \geq \widehat{l}(W) + B_n(W)\widehat{\kappa}_n(1-\beta)\}\Big\}
\end{aligned}$$

By the proof of Proposition 2, the lower limit of the first term is bounded below by $1-\alpha$ and by construction the lower limit of the second term is bounded below by $-\beta$. $\square$

Thus, we construct parameter confidence intervals by taking the union of confidence intervals for all $\pi(W)$ in the confidence interval $CR'_{1-\beta}$, which is an expanded version of the estimate of the parameter set $\mathcal{P}$. More conservative parameter intervals are obtained by choosing a larger confidence set $CR'_{1-\beta}$.

This completes our general estimation results. We now discuss some specifics features, as well as the related literature.

3.4. **Some Estimation Specifics.** At this point it is useful to summarize the steps in estimation, and relate our general results to the literature.

The first step is to estimate the allowable values of selection probabilities $\mathcal{P}$, using a boundary estimator such as (3.11). Then we widen the set to accomodate for estimation, obtaining the confidence set $CR'_{1-\beta}$ of (3.12). This set gives the values of selection probabilities $\pi(W)$ to be used in the subsequent estimation steps. That is, if π is a scalar parameter, then we choose values in a grid $\{\pi_k^*, k = 1, ..., K\}$ representing $CR'_{1-\beta}$. If $\pi(W)$ is modeled to depend nontrivially on covariates W, then the range of values of $\pi(W)$ are represented; for instance, if $\pi(W)$ depends on a vector of parameters, then a grid over the possible parameter values could be used. We summarize the grid as $\{\pi_k^*(W), k = 1, ..., K\}$ in the following.

The second step is to estimate the control function for the Tobin regressor for each value $\pi_k^*(W)$. With an estimator $\hat{F}_X(\cdot|Z, W)$ of the distribution $F_X(\cdot|Z, W)$, we compute (3.5) for $X > 0$ as:

$$\hat{V}_{\pi,k} = \frac{\hat{F}_X(X|Z, W) - \pi_k(W)}{1 - \pi_k(W)} \tag{3.13}$$

Alternatively, with an estimator $\hat{Q}_X(\cdot|W, Z)$ of the quantile function $Q_X(\cdot|W, Z)$, we compute (3.6) for $X > 0$ as:

$$\hat{V}_{\pi,k} = \int_0^1 1\left\{\hat{Q}_X\left((\pi_k(W) + (1 - \pi_k(W))v|W, Z\right) \leq X\right\} dv. \tag{3.14}$$

Either approach to estimating the control function can be used. If the model is restricted, then simpler methods of estimating the control function may be applicable. For instance, under the linear model discussed in Section 2, the general formulae (3.13) or (3.14) can be replaced by the simpler linear version (2.9):

$$\hat{V}_{\pi,k} = X - Z'\hat{\gamma}_k$$

where $\hat{\gamma}_k$ is from the estimation of (2.8) with $\pi = \pi_k$.

The third step estimates the response model for each control function estimate $\hat{V}_{\pi,k}$. This may involve estimating the conditional distribution $F_Y[\cdot|X, W, \hat{V}_{\pi,k}]$ or the conditional quantile function $Q_Y[\cdot|X, W, \hat{V}_{\pi,k}]$, either under a structural parameterization or using a nonparametric procedure. Or, this could involve estimating the mean regression $E[Y|X, W, \hat{V}_{\pi,k}]$ or some other interpretable function such as local policy effects, average derivatives or local average responses; again with either a parametric model or nonparametric procedure. Using our notation for the functional of interest, this step results in the estimate $\hat{\theta}_k = \hat{\theta}(\pi_k(W))$ of the parameter of interest. This step also yields an estimate of the confidence interval $C_k = C_{1-\alpha}(\pi_k(W))$ for each component of θ. For expositional ease, now we suppose that θ is a scalar parameter, so that $\hat{\theta}_k$ is a scalar and C_k is its estimated confidence interval in the following.

The final step is to assemble the results for all the grid values $\{\pi_k(W), k = 1, ..., K\}$ into the final estimates. That is, the set estimate for θ is formed as the interval

$$\left[\min_k \theta_k, \max_k \theta_k\right].$$

The confidence interval for θ is given as the union of the confidence intervals over all $\pi_k(W)$ values, namely

$$CR := \bigcup\nolimits_{k=1}^{K} C_k.$$

For a vector-valued θ, we would compute set estimates and confidence intervals for each component, and for function-valued θ, we could do the same for functional aspects of interest. This completes the procedure we have justified by our general derivations and results.

It is important to stress that our constructive identification approach relates set identification and inference to results for point identification and inference. We were brief in describing details for our third step - estimation of the response model given $\pi_k(W)$ and estimated control – because most of the necessary properties for estimation and inference are established in the existing literature.The foremost reading here is Imbens and Newey (2005), who discuss nonparametric series estimation in triangular equation systems with estimated regressors, and give detailed coverage to the properties needed for estimating many common functions such as policy effects and average derivatives. For parametric response models with an estimated regressor, much of the theory is available in Newey, Powell and Vella (2004), as well as in the classic Newey and McFadden (1994). Turning to censored quantile regression in a parametric framework (such as (2.8) here), see Powell (1984) and Chernozhukov and Han (2002), and for quantile regression with an estimated regressor, see Koenker and Ma (2006) and Lee (2006). Nonparametric quantile regression with estimated regressors is covered in Chaudhuri (1991), Chaudhuri, Doskum, and Samarov (1997), Belloni and Chernozhukov (2007), and Lee (2006). Finally, for estimation of the conditional distribution function of the response, see Hall, Wolff, Yao (1997) and Chernozhukov and Belloni (2007), among others.

We now turn to a substantive empirical application to illustrate our method including inference.

4. The Marginal Propensity to Consume out of Housing Wealth

Recent experience in housing markets has changed the composition of household wealth. In many countries such as the United States, housing prices have increased over a long period, followed by substantial softening. The market for housing debt, especially the risky subprime mortgage market, has experienced liquidity shortages that first resulted in increased volatility in many financial markets and later led to the collapse of credit markets.

In terms of economic growth, much interest centers on the impact of changes in housing wealth on household consumption. That is, if housing prices are permanently lower in the future, will household consumption, and therefore aggregate demand, be substantially lower as well? For this, one requires an assessment of the marginal propensity to consume out of housing wealth.

Surprisingly, the literature does not agree on the "right" measure of the marginal propensity of consumption out of housing wealth. Some papers find marginal propensities of 15 to 20 percent (e.g. Benjamin, Chinloy and Jud (2004)) while others report relatively low estimates of 2 percent in the short run and 9 percent in the long run (e.g.Carroll, Otsuka and Slacalek (2006)). Research in this area is very active, but no concensus has arisen about the impacts.[4]

One of the problems of estimation is the fact that variables such as income and housing wealth are endogenous and, in most surveys, also censored. The literature typically drops the censored observations, and tries to estimate the relationship by incorporating some non-linearities. As we have discussed, this is likely to bias the results, and therefore could have a role in why there is no agreement on a standard set of estimates. We feel that the estimation of the marginal propensity of consumption out of housing wealth is a good situation for using the methodologies developed here to shed light an a reasonable range of parameter values applicable to the design of policy.

We have data on U.S. household consumption and wealth from Parker (1999). These data are constructed by imputing consumption spending for observed households in the Panel Survey of Income Dynamics (PSID), using the Consumer Expenditure Survey (CEX). Income data is preprocessed — original observations on income are top-coded, but all households with a top-coded income value have been dropped in the construction of our data.

[4]This impact of housing wealth is of primary interest for the world economy, not just the US. See, for instance, Catte, Girouard, Price and Andre (2004) and Guiso, Paiella and Visco (2005) for European estimates in the range of 3.5 percent. Asian estimates are in a simlar range; see Cutler (2004) for estimates of 3.5 percent for Hong Kong.

We estimate a 'permanent income' style of consumption model:

$$\ln C_{it} = \alpha + \beta_{PY} \ln PY_{it} + \beta_H \ln H_{it}/W_{it} + \beta_W \ln W_{it} + \beta_Y \ln Y_{it} + U^*_{it} \quad (4.1)$$

$$= \alpha + \beta_{PY} \ln PY_{it} + \beta_H \ln H_{it} + (\beta_W - \beta_H) \ln W_{it} + \beta_Y \ln Y_{it} + U^*_{it}$$

Here C_{it} is consumption spending, PY_{it} is a constructed permanent component of income (human capital), H_{it} is housing wealth, W_{it}, is total wealth and Y_{it} is current income. Our focus is the elasticity β_H, the propensity to consume out of housing wealth.

Log housing wealth takes on many zero values, which we model as the result of bound censoring and selection. These features arise first by the treatment of mortgage debt (we do not observe negative housing wealth values) and by the choice of renting versus owning of a household's residence. We view the composition of wealth between housing and other financial assets as endogenous, being chosen as a function of household circumstances and likely jointly with consumption decisions. Thus, we model $\ln H_{it}$ is a *Tobin regressor*, and treat current income, permanent income and total wealth as exogenous. For instruments, we use lagged values of of the exogenous regressors..

One implication of the *Tobin regressor* structure is that all standard OLS and IV estimates are biased; including estimates that take into account either censoring or endogeneity, but not both. In Table 1, we present OLS and IV estimates for various subsamples of the data. The OLS estimates are all low; 2.7% for all data, 3.3% for households with observed lag values, and 5.3% for the "complete cases, " or households with nonzero housing values. The IV estimate for the complete cases is roughly a four-fold increase, namely 21.3%.

Estimation begins with establishing a range for the selection probability by studying the probability of censoring. Once the range is set, the estimates are computed

Table 1. Basic Estimates of Housing Effects

	All Households	Households with observed IV	Nonzero Housing Wealth (Complete Cases)
Sample Size	8,735	3,771	2,961
OLS	.027	0.033	0.053
	(.004)	(0.006)	(0.010)
IV (TSLS)			0.213
			(0.030)

in two steps. First, we compute quantile regressions of the Tobin regressor using censored LAD as in (2.8), for different values of the selection probability, and then estimate the control function for each probability value. Second, we estimate the model (4.1), including the estimated control function, as in (2.10) or (2.11). Our set estimates coincide with the range of coefficients obtained for all the different selection probability values. Their confidence intervals are given by the range of upper and lower confidence limits for coefficient estimates. All estimates were computed using Stata 10.0, and the code is available from the authors.

To set the range for the selection probability, we estimated the probability that $\ln H = 0$ given values of PY, Y and W, and found its minimum over the range of our data, as in (3.11). Specifically, we used a probit model, including linear and quadratic terms in all regressors. The minimum values were small, and, as a result we chose a rather low yet conservative value of $\widehat{\ell}(W) = .04$. After adjusting by two times standard error times a log factor, the upper bound estimate became .08. (We have illustrated this calculation graphically in Figure 3). Thus, we set the range for the selection probability to $\pi \in [0, .08]$.[5] Specifically, each estimation step is done over the grid of values 0, .008, .016, ...08

For the first estimation step, we implement the censored LAD estimation algorithm of Chernozhukov and Hong (2002). This requires (again) estimating the probability of censoring, and the performing standard quantile regression on samples with low censoring probability. These estimates are used to construct the control function V_π for each grid value. For the second estimation step, we computed mean regression, median regression, and quantile regression for the 10% and 90% quantiles.

We present some representative estimates in Figures 4 and 5. Figure 4 displays the different estimates of the housing effect, and Figure 5 gives the estimates of all coefficients for median regression. Each figure plots the estimates for each grid value of π, as well as the associated confidence interval, obtained by bootstrapping (denote "bci" in the legend). The set estimates are the projections of those curves onto the left axis.

Overall, there is very little variation in the estimates with π, the selection probability. The housing effect increases over quantiles, and there are other level differences not displayed. The bootstrap confidence interval values are fairly wide, reflecting variation from censored LAD estimation (as well as the selection probability) as well as the second step regressions. For what they are worth, Figure 5 includes the confidence interval estimates from the second step only (denoted "ci"); so that the difference with the bootstrap intervals gives a sense of the impact of the censored LAD estimation and control function construction.

The results on housing effects are summarized in Table 2. Interval estimates are fairly tight, evidencing the lack of sensitivity with the selection probability. We

[5] We tried several variations, including using a probit model with Cauchy tails (as in Koenker and Yoon (2007)), without changing out conclusion that $[0, .08]$ is a conservative (wide) choice of range.

Table 2. Confidence Sets for Housing Effects

	Set Estimate	Bootstrap Confidence Region
Housing coefficient β_H		
Mean Outcome	$[.133, .161]$	$[-.015, .312]$
Median Outcome	$[.151, .162]$	$[.026, .293]$
90% Quantile	$[.172, .196]$	$[.015, .357]$
10% Quantile	$[.120, .141]$	$[-.077, .341]$
Selection Probability π	$[0, .04]$	$[0, .08]$

note that all results are substantially larger than the OLS estimates (2.7%-5.3%), which ignore endogeneity. All results are substantially smaller than the IV estimate of 21.3%, which ignores censoring. Relative to the policy debate on the impact of housing wealth, our interval estimates fall in a very plausible range. However, the bootstrap confidence intervals are too wide to discriminate well among these ranges of values. We do see that bootstrap confidence intervals are smaller for median regression than mean regression, and much smaller than for the 10% and 90% quantiles, as expected.

5. Summary and Conclusion

We have presented a general set of identification and estimation results for models with a Tobin regressor, a regressor that is endogenous and mismeasured by bound censoring and (independent) selection. Tobin regressor structure arises very commonly with observations on financial variables, and our results are the first to deal with endogeneity and censoring together. As such, we hope our methods provide a good foundation for understanding of how top-coding, bottom-coding and selection distort the estimated impacts of changes in income, wealth, dividends and other financial variables.

Our results are restricted to particular forms of censoring. It is not clear how to get around this issue, because endogeneity requires undoing the censoring, and undoing the censoring (seemingly) requires understanding its structure. Here we separate selection and bound censoring with independence, use quantile regression to address bound censoring, and identify parameter sets for the range of possible selection probability values.

We have developed estimation and inference methods for set identified parameters. In particular, our results apply to any problem where the parameter value of interest is point identified conditional on the values of some nuisance parameters that are set identified. The procedure is quite simple: by fixing the nuisance parameter value in some suitable region, one first proceeds with regular point and interval estimation. Then, take the union over nuisance parameter values of the point and interval estimates to form the final set estimates and confidence set estimates. The final set estimates are set-consistent for the true parameter value, and the confidence set estimates cover this value with at least a prespecified probability in large samples.

One essential feature of our framework is that the censoring is not complete, namely that some true values of the censored variable are observed. Such "complete cases" provide the data for our estimation of the main equation of interest. However, not all forms of censoring involve observing complete cases. Suppose, for instance, that we were studying household data where all that we observe is whether the household is poor or not; or that their income falls below the poverty line threshold. In that case, using the "poor" indicator is a severely censored form of income, and no complete cases (income values) are observed. Our methods do not apply in this case, although it is of substantial practical interest.

References

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

[2] Altonji, J. and R. Matzkin (2005), "Cross Section and Panel Data Estimators for Nonseparable Models with Endogneous Regressors," *Econometrica*, 73, 1053-1103.

[3] Belloni, A., and V. Chernozhukov (2007): "Conditional Quantile and Probability Processes under Increasing Dimension," preprint.

[4] Benjamin, J.D., P. Chinloy and G.D. Donald (2004) "Why Do Households Concentrate Their Wealth in Housing?" *Journal of Real Estate Research*, Oct-Dec.

[5] Blundell, R., and J.L. Powell (2003): "Endogeneity in Nonparametric and Semiparametric Regression Models," chapter 8 in M. Dewatripont, L. Hansen, and S. Turnovsky (eds.), *Advances in Economics and Econometrics*, Cambridge University Press, Cambridge, 312-357.

[6] Carroll, C., M. Otsuka and J. Slacalek (2006), "How Large Is the Housing Wealth Effect? A New Approach," NBER Working Paper.

[7] Catte, P., N. Girouard, R. Price and C. André (2004), "Housing Markets, Wealth and the Business Cycle," OECD Working Paper no. 394.

[8] Chaudhuri, P.; K. Doksum and A. Samarov (1997), "On Average Derivative Quantile Regression," *Annals of Statistics*,. 25 (2), , 715–744.

[9] Chaudhuri, S. (1991), "Nonparametric Estimates of Regression Quantiles and their Local Bahadur Representation," *Annals of Statistics,* 19(2), 760–777.

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

[11] 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.

[12] Chernozhukov, V., and C. Hansen (2005), "An IV Model of Quantile Treatment Effects," *Econometrica*, 73, 245-261.

[13] Chernozhukov, V. and H. Hong (2002), "Three-step Censored Quantile Regression and Extramarital Affairs,". *Journal of the American Statistical Association*, 97, 872–882.

[14] Chernozhukov, V., H. Hong and E. Tamer, (2007), "Estimation and Confidence Regions for Parameter Sets in Econometric Models," *Econometrica*, 75, 1243-1284.

[15] Chernozhukov, V., I. Fernadez-Val and B. Melly (2007). “Inference on Counterfactual Distributions,” preprint.

[16] Chernozhukov, V, S. Lee and A. Rosen (2008)

[17] Chesher, A. (2003), “Identification in Nonseparable Models,” *Econometrica* 71, 1405-1441.

[18] Cutler, J. (2004), “The Relationship between Consumption, Income and Wealth in Hong Kong,” HKIMR Working Paper No.1/2004.

[19] Guiso, L. M. Paiella and I. Visco (2005), “Do capital gains affect consumption? Estimates of wealth effects from Italian households’ behavior” Economic Working Paper no. 555, Bank of Italy.

[20] Hall, P., R. Wolff, and Q. Yao (1999), “Methods for estimating a conditional distribution function,” *Journal of the American Statistical Association* 94, pp. 154–163.

[21] He, X. and Q-M Shao (2000), “On Parameters of Increasing Dimensions,” *Journal of Multivariate Analysis* 73(1) , 120–135.

[22] Imbens, G.W. and W.K. Newey, (2005), “Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity,” MIT Working Paper, revised September.

[23] Keonker, R. (2005), *Quantile Regression*, Cambridge University Press, Cambridge.

[24] Koenker R. and L. Ma (2006), “Quantile Regression Methods for Recursive Structural Equation Models”, forthcoming *Journal of Econometrics*.

[25] Koenker R. and J. Yoon (2007), “Parametric Links for Binary Response Models”, forthcoming, *Journal of Econometrics*.

[26] Lee, S. (2007), "Endogeneity in Quantile Regression Models: A Control Function Approach", *Journal of Econometrics*, forthcoming.

[27] 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.

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

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

[30] Manski, C.F. (2003), *Partial Identification of Probability Distributions*, Springer, Springer Series in Statistics.

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

[32] Newey, W. K, J. L. Powell and F.Vella (1999), "Nonparametric Estimation of Triangular Simultaneous Equations Models,". *Econometrica* 67, 565–603.

[33] Newey, W. and D. McFadden (1994), "Large Sample Estimation and Hypothesis Testing," in *Handbook of Econometrics, vol. IV*, North-Holland, Amsterdam, 2111–2245.

[34] Nicoletti, C. and F. Peracchi (2005), "The Effects of Income Imputations on Micro Analyses: Evidence from the ECHP," Working Paper, University of Essex, August.

[35] Powell, J.L. (1984), "Least Absolute Deviations Estimation for the Censored Regression Model," *Journal of Econometrics*, 25, 303-325.

[36] Rigobon, R and T. M. Stoker (2006a) "Bias from Censored Regressors", forthcoming *Journal of Business and Economic Statistics*.

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

[38] Tripathi, G. (2004), "GMM and Empirical Likelihood with Incomplete Data," Working Paper.

[39] Tobin, J. (1958), "Estimation of Relationships for Limited Dependent Variables," *Econometrica*, 26, 24-36.

Current address:, Department of Economics, MIT, 50 Memorial Drive, Cambridge, MA 02142 USA, vchern@mit.edu;, Sloan School of Management, MIT, 50 Memorial Drive, Cambridge, MA 02142 USA, rigobon@mit.edu;, Sloan School of Management, MIT, 50 Memorial Drive, Cambridge, MA 02142 USA, tstoker@mit.edu.,

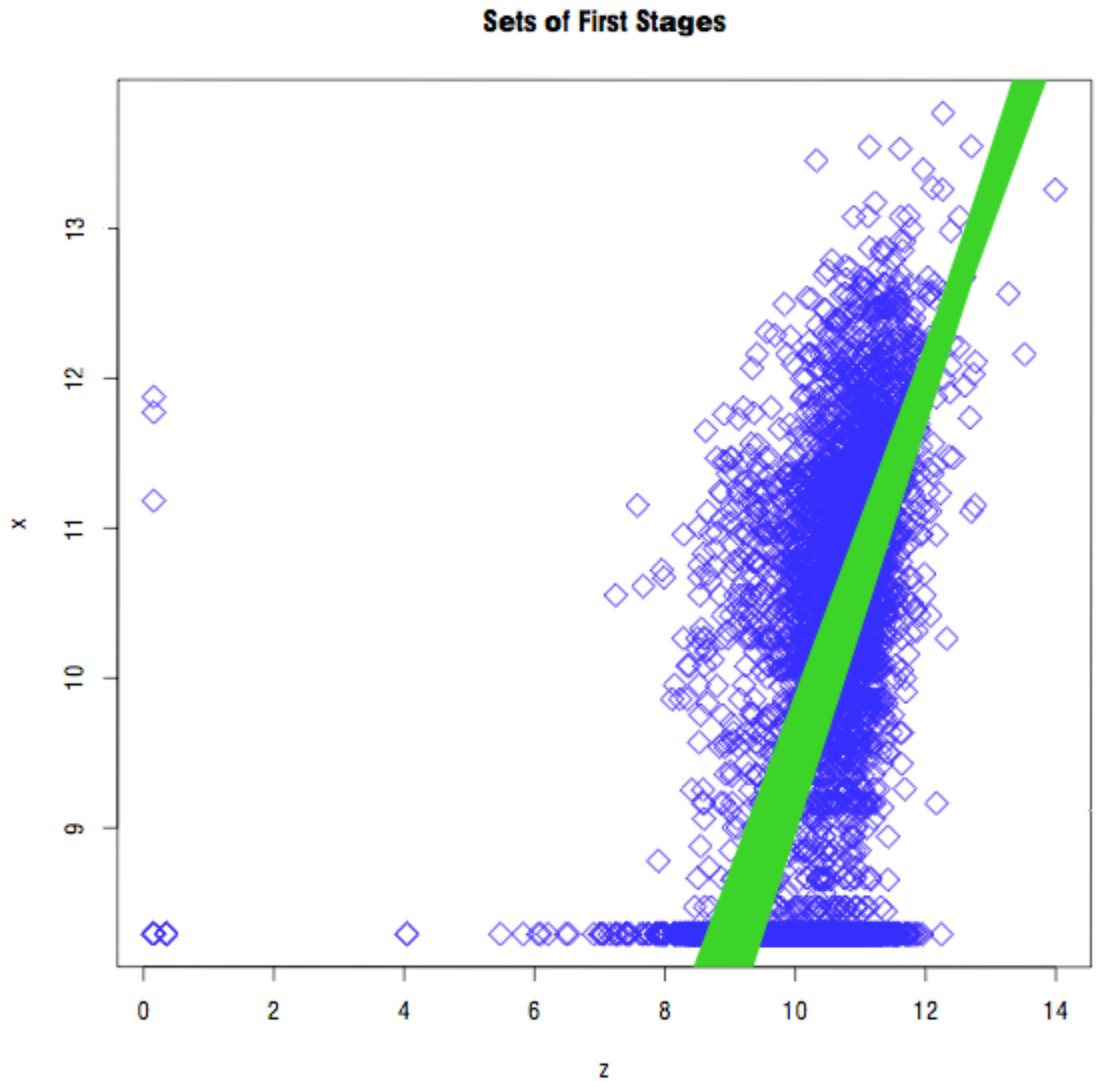

Figure 1

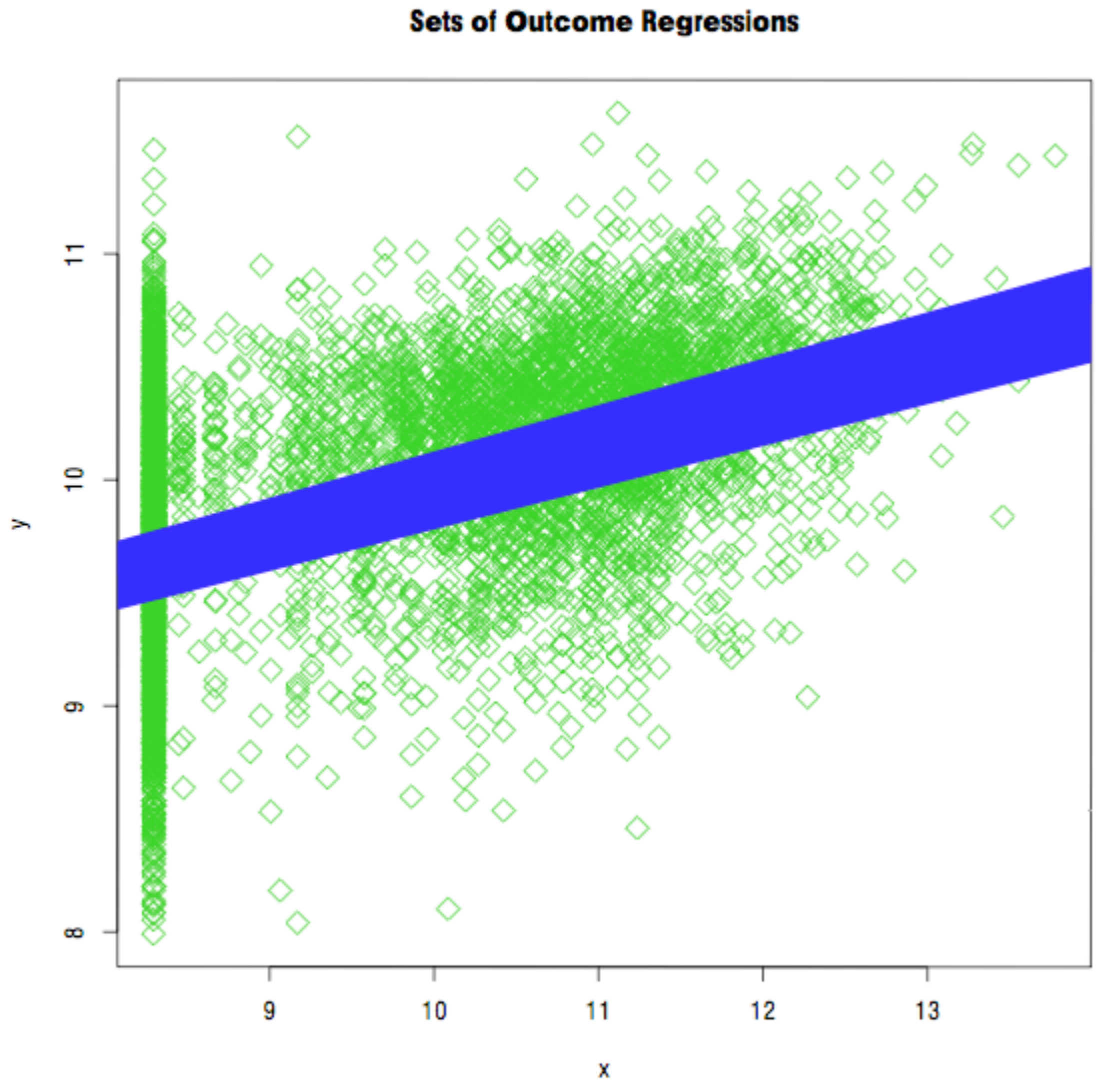

Figure 2

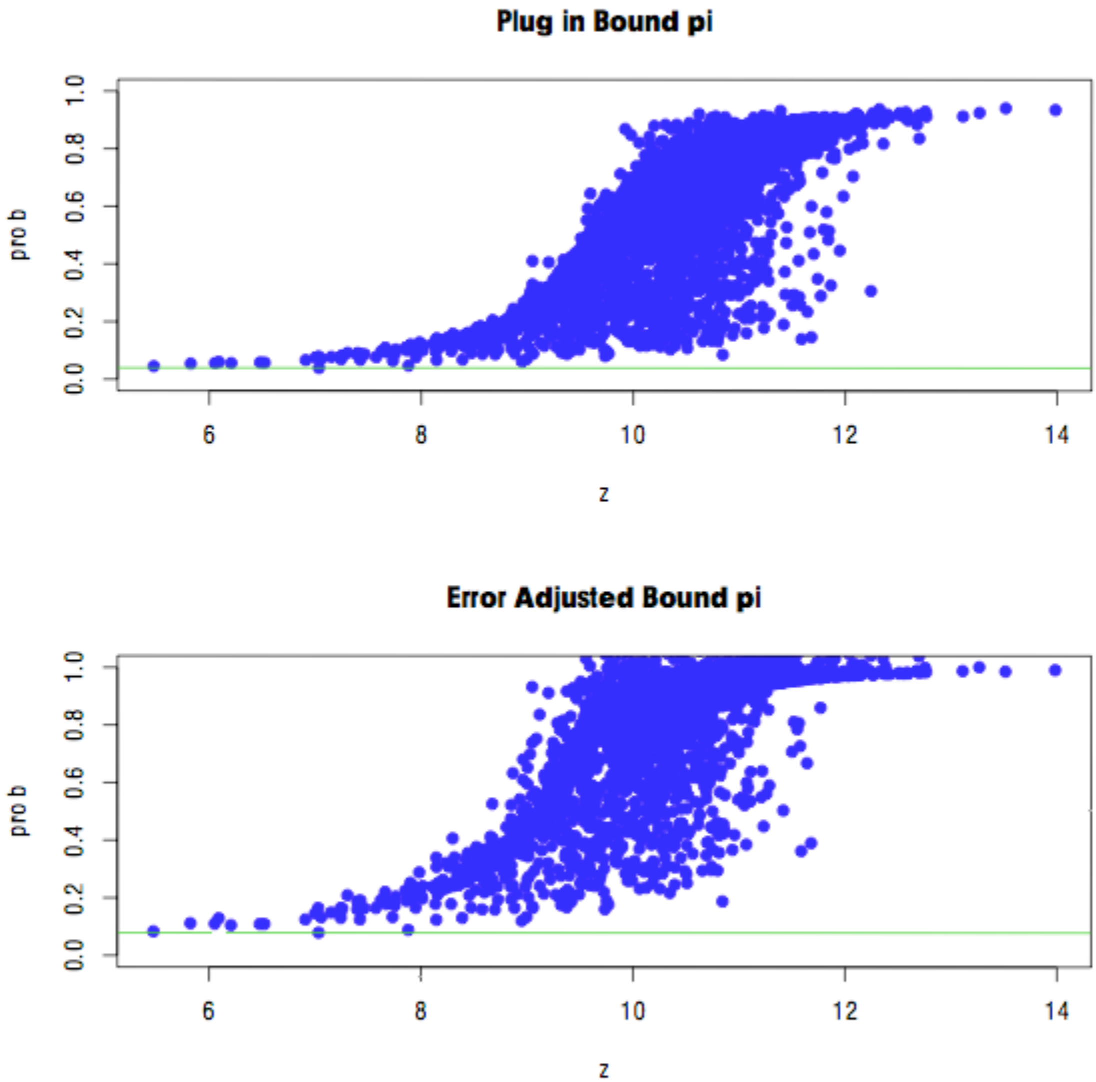
Plug in Bound pi
pro b
0.0
0.2
0.4
0.6
0.8
1.0
6
8
10
12
14
z
Error Adjusted Bound pi
pro b
0.0
0.2
0.4
0.6
0.8
1.0
6
8
10
12
14
z

Figure 3

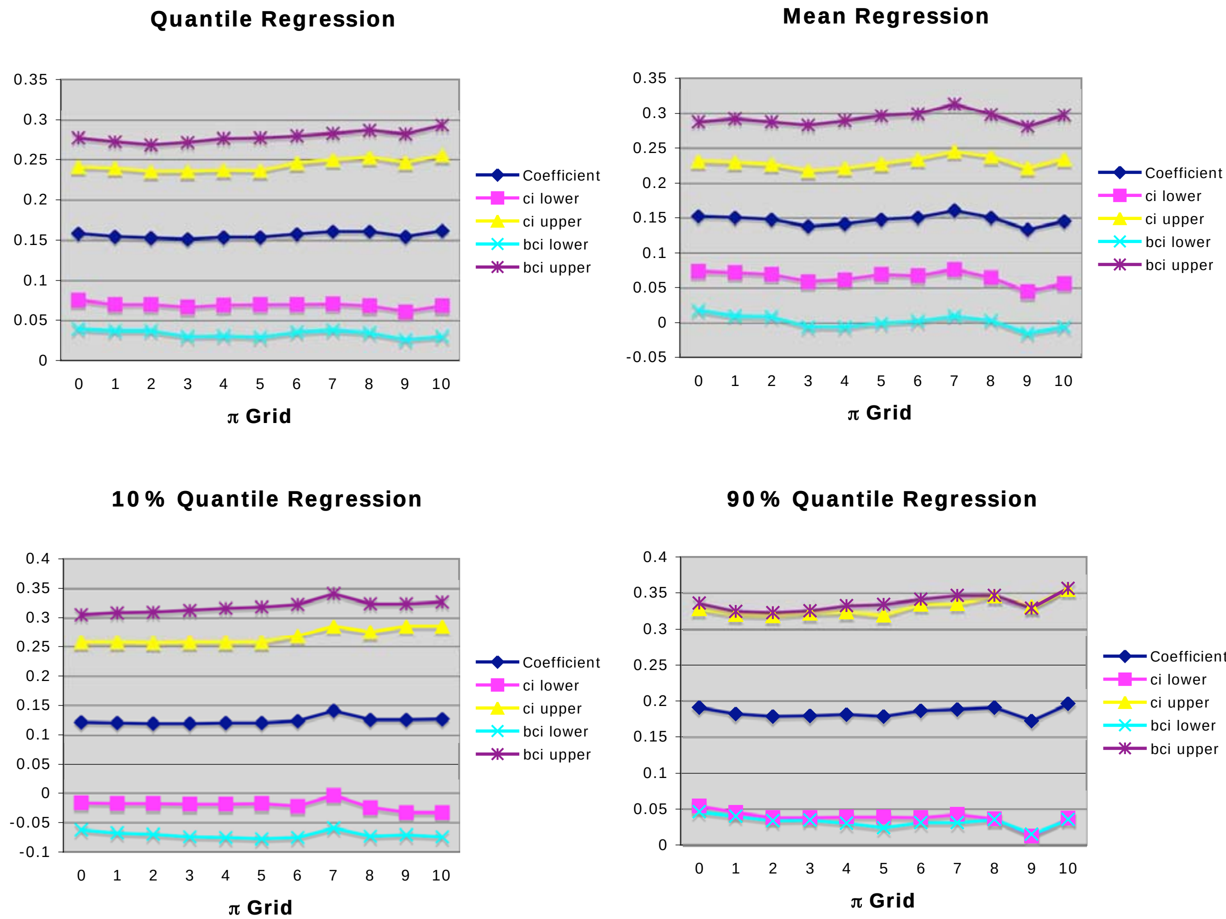

Figure 4: Housing Coefficient Estimates

Note: π Grid from 0.0 to 0.08

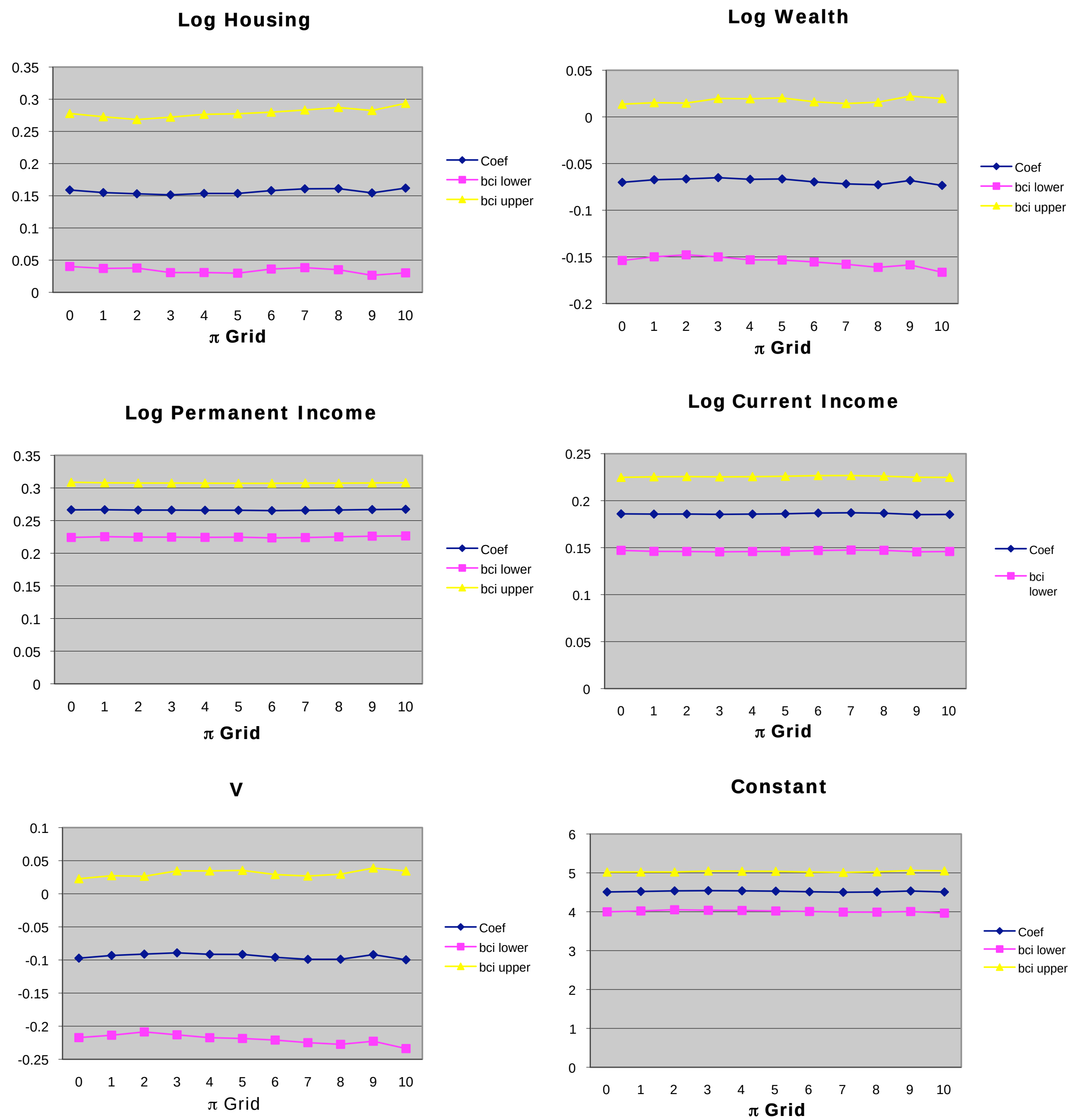

Figure 5: Quantile Regression Estimates

Note: π Grid from 0.0 to 0.08