cemmap

centre for microdata methods and practice

IV Methods for Tobit Models

Andrew Chesher
Dongwoo Kim
Adam M. Rosen

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP16/22

Economic and Social Research Council

IV Methods for Tobit Models*

Andrew Chesher† — UCL and CeMMAP
Dongwoo Kim‡ — Simon Fraser University
Adam M. Rosen§ — Duke University and CeMMAP

October 3, 2022

Abstract

This paper studies models of processes generating censored outcomes with endogenous explanatory variables and instrumental variable restrictions. Tobit-type left censoring at zero is the primary focus in the exposition. Extension to stochastic censoring is sketched. The models do not specify the process determining endogenous explanatory variables and they do not embody restrictions justifying control function approaches. Consequently, they can be partially or point identifying. Identified sets are characterized and it is shown how inference can be performed on scalar functions of partially identified parameters when exogenous variables have rich support. In an application using data on UK household tobacco expenditures inference is conducted on the coefficient of an endogenous total expenditure variable with and without a Gaussian

*We thank Xinyue Bei for providing excellent research assistance, and participants at several conferences and seminars for helpful discussions and comments. Financial support from the UK Economic and Social Research Council through grants (RES-589-28-0001, RES-589-28-0002 and ES/P008909/1) to the ESRC Centre for Microdata Methods and Practice (CeMMAP) is gratefully acknowledged. The first results on this topic were presented at the University of Cambridge on May 31, 2014 at a Festschrift Conference in honor of Richard J. Smith in a presentation "New Directions for IV".

†Address: Department of Economics, University College London, Gower Street, London WC1E 6BT, United Kingdom. Email: andrew.chesher@ucl.ac.uk.

‡Address: Department of Economics, Simon Fraser University, West Mall Centre, SFU, 8888 University Drive Burnaby, BC Canada, V5A 1S6. Email: dongwook@sfu.ca.

§Address: Department of Economics, Duke University, 213 Social Sciences, Box 90097 Durham, NC 27708, United States. Email: adam.rosen@duke.edu.

distributional restriction on the unobservable and compared with the results obtained using a point identifying complete triangular model.
Keywords: censored outcomes, endogeneity, incomplete models, instrumental variables, partial identification, stochastic censoring.

JEL classification: C10, C24, C26, C51.

1 Introduction

This paper develops and applies results on the identifying power of models for censored continuous outcomes, focussing on cases with left censoring at a fixed known value. In econometrics the Tobit model, Tobin (1958), is the iconic example. A simple example of the models considered here is the Tobit model with a linear index, endogenous variables $Y \equiv (Y_1, Y_2')'$, and exogenous variables Z and U such that

$$Y_1 = \max(0, Y_1^*) \qquad Y_1^* = \alpha Y_2 + \beta Z + U \tag{1}$$

with U unobserved and α, β parameter row vectors conformable with Y_2 and Z, respectively.

In the classical Tobit model there are no endogenous variables, Y_2, equivalently $\alpha = 0$. Here we study cases in which some explanatory variables may be endogenous. There are instrumental variables, components of Z, with restricted effect on Y_1^*, delivered for example by exclusion restrictions requiring elements of β to be zero. The instrumental variables (IVs) are restricted to be distributed to some degree independently of the unobserved variable in the structural equation for the censored outcomes. We consider mean, quantile, and full stochastic independence restrictions, with both parametric and nonparametric specifications of the distribution of unobserved variables. In this simple example Y_1^* is a linear index. Our results are derived under a nonlinear index specification and there can be a parametric, semiparametric or nonparametric specification.

The IV models employed here are incomplete[1] in the sense that there is no specification of the determination of the endogenous explanatory variables. If a complete model is used, obtained by adding an equation for Y_2, there is the danger of misspecification of the genesis of endogenous explanatory variables. Incomplete IV models provide a robust alternative.

[1] Complete models place restrictions on the determination of endogenous variables such that at every value of observed and unobserved exogenous variables a unique value of the endogenous variables eventuates. Models that do not possess this property are incomplete. The terminology is due to Koopmans, Rubin and Leipnik (1950).

STATA's `ivtobit` command,[2] despite its name, uses a complete triangular model assuming Gaussian unobserved variables to compute estimates of Tobit models with endogenous explanatory variables. That attack does not deliver consistent estimates when endogenous variables are discrete[3] or determined by multiple sources of heterogeneity or when there is simultaneity which breaks the triangular restriction. By contrast, the IV approach employed here can be used when endogenous variables are discrete and when there is simultaneous determination of endogenous variables, and it places no restrictions on the way in which endogenous explanatory variables' values are generated. The IV approach is also a useful alternative to control function approaches which require endogeneity to be absent once there is conditioning on identifiable functions of observed variables. This paper shows how inferences can be drawn using censored data and weakly-restrictive, robust IV models.

The models studied here fall in the class of Generalized Instrumental Variable (GIV) models introduced in Chesher and Rosen (2017) (CR). We use techniques developed in that paper to carry out identification analysis, and we show how to implement the high level results given in CR in models for censored outcomes.[4] The models may be partially identifying but they can be point identifying. Often it is not possible to determine identification status using the data available in applications so a cautious approach allowing, as here, for the possibility of partial identification is sensible.

We conduct inference on scalar functions or subvectors of potentially partially identified parameters using a "minmax" statistic as in Belloni, Bugni, and Chernozhukov (2018) with a self-normalized critical value from Chernozhukov, Chetverikov, and Kato (2019), appropriate for inference based on a very large number of moment inequalities as we find here. The approach applies regardless of identification status, and we propose a method for implementation when exogenous variables have rich support. We illustrate with an application to UK household survey data recording tobacco expenditure shares in which around 70% of households record zero expenditures.

The focus here is on IV Tobit models with left censoring at zero, with and without a Gaussian distributional restriction on the scalar unobserved variable, with and without a parametric specification of the index in the model. It is trivial to extend to cases with right censored outcomes and straightforward to extend to cases in which the censoring value is

[2] StataCorp (2019).

[3] The triangular model is not point identifying when Y_2 is discrete, see Chesher (2005).

[4] An interesting feature of the problem here is that outcomes are continuous rather than discrete as is the case in many other applications of GIV such as Berry and Compiani (forthcoming), Chesher, Rosen, and Siddique (2019), Chesher and Rosen (2020b), and Chesher and Rosen (2020a).

stochastic as shown in Section 6.[5] Leading examples with right censoring arise when the censored outcome is the time until an event occurs, so the methods developed here find application in models of durations admitting endogenous explanatory variables.[6]

Much of the prior literature studying models with censored outcomes and endogenous explanatory variables relies on a complete model for the determination of endogenous variables. Examples include the fully parametric specifications studied in Heckman (1978), Nelson and Olson (1978), Amemiya (1979), Smith and Blundell (1986), Newey (1987), and Blundell and Smith (1989) that made early contributions to the study of limited dependent variable models (including Tobit models in particular) permitting endogenous explanatory variables and enabling consistent estimation by way of control function approaches and marginal or conditional maximum likelihood procedures.[7] Control function approaches for semiparametric triangular models for censored outcomes are provided in Das (2002), Blundell and Powell (2007), and Chernozhukov, Fernandez-Val, and Kowalski (2015). These papers do not require parametric distributional restrictions on unobservable heterogeneity, with Das (2002) employing symmetry restrictions and Blundell and Powell (2007) and Chernozhukov, Fernandez-Val, and Kowalski (2015) using conditional quantile restrictions.

The approach taken here is in the spirit of Manski and Tamer (2002), which pioneered the use of incomplete models for censored outcomes or covariates and developed partial identification analysis. That paper characterized identified sets and proposed consistent set estimators for a variety of models with censored variables and exogenous explanatory variables, in which the censoring process is not specified. In the IV Tobit models studied here, the censoring process is specified, but endogenous explanatory variables are permitted and it is the lack of a specification of their determination that renders the models incomplete.

In this respect the models studied here have similarities to the incomplete models studied in Hong and Tamer (2003), Honore and Hu (2004), Chen and Wang (2020) and Wang and Chen (2021), which also do not impose a specification of the process determining values of endogenous explanatory variables. Hong and Tamer (2003) employs conditional quantile restrictions with a censored outcome variable, and focuses on settings in which the parameters are point identified. Sufficient conditions for point identification are proposed along

[5]In ongoing research we study cases in which the censoring variable is a function of endogenous variables as arises, for example, in competing risk models with endogenous variables affecting risks.

[6]See for example Lancaster and Chesher (1984), Lancaster (1985), Olsen and Farkas (1989), Frandsen (2015), and Wrenn, Klaiber, and Newburn (2017).

[7]Comparisons of the efficiency of different procedures are provided in Newey (1987) and Blundell and Smith (1989).

with an estimator, and the asymptotic properties of the estimator are characterized. The support conditions shown to deliver point identification are strong. In many cases they will not be satisfied in the sample to hand. Honore and Hu (2004) focus primarily on panel data models with a censored outcome, but in Section 4 moment conditions are derived for cross-sectional settings under different conditions on the distribution of unobservables, such as independence, symmetry, and quantile restrictions. Honore and Hu (2004) investigate the performance of a moment-based point estimator in Monte Carlo analyses, but they do not provide conditions for point identification. Chen and Wang (2020) and Wang and Chen (2021) propose moment-based estimators for models with censored outcomes under independence and quantile restrictions, respectively. They provide high-level conditions on the distribution of observable variables under which parameters are point-identified, but these are not always easy to verify in practice.

We characterize sharp identified sets for parameters applicable whether or not such additional conditions hold, and we employ inference that is robust to the possibility of partial identification. The analysis we provide thus does not require additional assumptions be made on the process being in studied in order to guarantee point identification. If, however, the distribution of observable variables is sufficiently rich to achieve point identification, the sharp identified set delivered by our analysis will, by construction, be a singleton set.

There is also research that considers the different but important problems of *endogenous censoring* of explanatory variables. This includes Khan and Tamer (2009), Khan, Ponomareva, and Tamer (2011), and Section 7 of Chesher and Rosen (2020b).

This paper makes contributions to the literature on models of censoring with endogenous explanatory variables and instrumental variable restrictions, as follows.

1. Most previous papers on this topic consider models that, unlike those considered here, require a complete specification for the determination of endogenous explanatory variables. Our analysis shows what can be learned when the specification of the genesis of endogenous explanatory variables is dropped.

2. Our analysis is robust to the possibility of partial identification, and is thus applicable when data are not compatible with conditions that ensure point identification.

3. We consider the use of more or less demanding restrictions on the distribution of unobservable heterogeneity conditional on instruments. This can be used to assess how robust empirical findings are to, for example, relaxation from a full stochastic independence restriction to selected conditional quantile independence restrictions.

4. We conduct inference on functions or subvectors of parameters partially identified by moment inequalities in these IV Tobit models, using recent developments in Chernozhukov, Chetverikov, and Kato (2019) and Belloni, Bugni, and Chernozhukov (2018) which allow for a large number of moment inequalities relative to the sample size, as encountered in our application.

5. We show how quantile independence restrictions at multiple quantiles can be incorporated, extending results in Chesher and Rosen (2017). This enables investigation of the increase in identifying power as one moves from invoking a single conditional quantile restriction to successively more quantile restrictions, approaching full stochastic independence as more such restrictions are imposed.

6. We show how to determine the identifying power of an IV model for censored outcomes under a stochastic independence restriction with a nonparametric specification of the distribution of the unobservable variable in the structural equation for the censored outcome.

The paper proceeds as follows. In the next section we present the class of IV Tobit models studied here. In Section 3 we characterize the identified set of structures – combinations of structural functions and distributions of unobserved heterogeneity – that are compatible with the censored outcome model. The identified set is shown to comprise those structures that lie in the intersection of two sets, each defined by a collection of conditional moment inequalities. We show how under some circumstances certain subsets of the inequalities reduce to moment equalities, and we show how exclusion restrictions can be incorporated into the characterization of the identified set. In Section 4 we analyze the identifying content of various restrictions on the joint distribution of exogenous variables and unobservable heterogeneity, such as conditional mean, conditional quantile, and stochastic independence restrictions. In Section 5 we describe how inference is carried out using results from Chernozhukov, Chetverikov, and Kato (2019) and Belloni, Bugni, and Chernozhukov (2018), and we propose a practical approach for application of the inference method when an identified set is characterized by conditional moment inequalities with continuous conditioning variables. We illustrate with an application in which we focus on conducting inference on the effect of total household nondurable expenditure on the share of expenditure spent on tobacco, previously considered using a control function approach in Adams, Blundell, Browning, and Crawford (2019). Section 6 sketches an extension to models with stochastic censoring. All proofs and figures are provided in the Appendix.

2 The IV Tobit model

Scalar endogenous outcome Y_1, possibly endogenous vector Y_2, exogenous vector $Z \in \mathcal{R}_Z$, and unobserved scalar $U \in \mathbb{R}$ satisfy:

$$Y_1 = \max(Y_1^*, 0), \qquad Y_1^* = m(Y_2, Z, U) \tag{2}$$

where the function m is continuous and strictly increasing in its third argument (U) and for all y_2 and z, $m(y_2, z, -\infty) \leq 0$. Define $Y \equiv (Y_1, Y_2)$. There is the inverse function $m^{-1}(y_2, z, y_1^*)$ such that for all y_1^*, y_2 and z

$$m(y_2, z, m^{-1}(y_2, z, y_1^*)) = y_1^*.$$

Example. *In a Gaussian Linear Tobit model (GLT model) the function m and its inverse are linear*

$$m(y_2, z, u) = \alpha y_2 + \beta z + u$$

$$m^{-1}(y_2, z, y_1^*) = y_1^* - \alpha y_2 - \beta z$$

and $U \sim N(0, \sigma^2)$ is restricted to be independent of Z. The model studied in *Tobin (1958) has no endogenous explanatory variables ($\alpha = 0$). The GLT model will be used as a running example and it features in the empirical application. Models that maintain this linear specification for $m(y_2, z, u)$ with alternative distributional restrictions to Gaussianity of U and independence of U and Z are referred to as linear Tobit models.*

We cast the problem into the GIV framework set out in CR in which a structure, $(m, \mathcal{G}_{U|Z})$ comprises two components.

1. A structural function determines which combinations of (Y, Z, U) can jointly occur. In the IV Tobit model this is fully determined by the function m which may be parametrically specified or nonparametrically specified, perhaps with shape restrictions.[8]

2. The second component of a structure is a collection of conditional distributions of U given Z, denoted

$$\mathcal{G}_{U|Z} \equiv \left\{ G_{U|Z}(\cdot|z) : z \in \mathcal{R}_Z \right\}$$

[8] In the notation of CR $h : \mathcal{R}_{YZU} \to \mathbb{R}$ is used to denote a structural function defined on the support of (Y, Z, U) such that $h(Y, Z, U) = 0$ with probability one. In the IV Tobit model there is a unique mapping from values of (Y_2, Z, U) to values of Y_1 and it is more transparent to work with the function, m, directly. In the notation of CR, one would define $h(Y, Z, U) \equiv Y_1 - \max(0, m(Y_2, Z, U))$.

where for any set $\mathcal{S} \subseteq \mathbb{R}$

$$G_{U|Z}(\mathcal{S}|z) \equiv \mathbb{P}[U \in \mathcal{S}|Z = z].$$

A model, say A, comprises a list of restrictions on structures, defining a set containing the admissible structures, $\mathcal{M}_A$, which satisfy the restrictions. A model's restrictions may limit the dependence between U and Z and may require that the function m satisfies conditions additional to those so far imposed, for example functional form and exclusion restrictions. In most applications exclusion restrictions will be prominent. When U and Z are stochastically independent a collection $\mathcal{G}_{U|Z}$ is a singleton $\{G_U\}$ where G_U is the marginal distribution of U. A model can additionally impose parametric restrictions on the distribution of U.

Example (continued). *In the GLT model the structural function is characterized by parameters* (α, β)*. The collection of distributions is a singleton,* $\{G_U\}$*, where* G_U *is a zero mean Gaussian distribution with variance* σ^2*. Each structure is characterized by a value of* $\theta \equiv (\alpha, \beta, \sigma)$*.*

Throughout the paper a model is referred to as an *IV Tobit Model* if it satisfies the following definition.

Definition 1 *An* **IV Tobit Model**, *A, comprises a set of restrictions on the process generating observed variables Y and Z such that (i) $Y_1 = \max(m(Y_2, Z, U), 0)$ as specified in (2) for some unobservable variable U residing on the same probability space as (Y, Z), and (ii) the function m and conditional distributions of U given Z are restricted to belong to some set, $\mathcal{M}_A$, of admissible pairs $(m, \mathcal{G}_{U|Z})$.*

A variety of IV Tobit Models are considered, imposing different restrictions on the pair $(m, \mathcal{G}_{U|Z})$. Our analysis here focuses on settings in which the censored variable Y_1^* is continuously distributed with censoring probability strictly between zero and one conditional on any realizations of Y_2 and Z.[9] Throughout the paper we assume $\mathcal{M}_A$ is such that $m(y_2, z, u)$ is restricted strictly increasing in u. Where convenient, notation $\tilde{G}_{U|Z}(t|z) \equiv G_{U|Z}((-\infty, t]\,|z)$ is used to denote the conditional cumulative distribution function of U given $Z = z$ associated with $G_{U|Z}(\cdot|z)$, and $\tilde{g}_{U|Z}(\cdot|z)$ is used to denote the corresponding conditional density. Notation $\mathcal{M}_A^0$ is used to denote the collection of structural functions m allowed by model $\mathcal{M}_A$.

[9]This affords simplification in the statement of some of our results, but is unnecessary for application of the identification analysis from Chesher and Rosen (2017).

3 Identification

3.1 Characterizations of identified sets

CR gives characterizations of identified sets of structures from which identified sets of structural features are obtained by projection. The characterizations make use of *residual sets* associated with a structure $(m, \mathcal{G}_{U|Z})$. A residual set $\mathcal{U}(y, z, m)$ is the set of values of unobserved U that, for a structural function m, are compatible with observed $Y = y$ and $Z = z$. In the IV Tobit model the residual sets are singleton when $y_1 > 0$ and semi-infinite intervals when $y_1 = 0$, as follows.

$$\mathcal{U}(y, z, m) = \begin{cases} (-\infty, m^{-1}(y_2, z, 0)] & , \quad y_1 = 0 \\ \{m^{-1}(y_2, z, y_1)\} & , \quad y_1 > 0 \end{cases}$$

Example (continued). *In the GLT model the residual sets are as follows.*

$$\mathcal{U}(y, z, \theta) = \begin{cases} (-\infty, -\alpha y_2 - \beta z] & , \quad y_1 = 0 \\ \{y_1 - \alpha y_2 - \beta z\} & , \quad y_1 > 0 \end{cases}$$

The parameter vector θ appears in place of m here because in the GLT model structures are characterized by the value of θ.

Let $\mathcal{F}_{Y|Z} \equiv \{F_{Y|Z}(\cdot|z) : z \in \mathcal{R}_Z\}$ denote the collection of conditional distributions of Y given Z, where for any set $\mathcal{Y} \subseteq \mathcal{R}_Y$

$$F_{Y|Z}(\mathcal{Y}|z) \equiv \mathbb{P}[Y \in \mathcal{Y}|Z = z].$$

Data is informative about this collection of distributions, which is assumed to be identified. Further define

$$B(z, t, m) \equiv \mathbb{P}[0 < Y_1 \leq m(Y_2, Z, t)|Z = z],$$

$$D(z, t, m) \equiv \mathbb{P}[Y_1 = 0 \wedge 0 \leq m(Y_2, Z, t)|Z = z],$$

$$C(z, t, m) \equiv \mathbb{P}[Y_1 \leq m(Y_2, Z, t)|Z = z] = D(z, t, m) + B(z, t, m),$$

$$\Delta(z, t_1, t_2, m) \equiv B(z, t_2, m) - B(z, t_1, m),$$

all of which are identified for each (z, t, m) given knowledge of $\mathcal{F}_{Y|Z}$.

The following Proposition characterizes the identified set of structures.

Proposition 1 *The identified set of structures delivered by an IV Tobit model, A, and a collection of conditional distributions of Y given Z,* $\mathcal{F}_{Y|Z} \equiv \{F_{Y|Z}(\cdot|z) : z \in \mathcal{R}_Z\}$ *is the intersection of two sets of structures*

$$\mathcal{I}(\mathcal{F}_{Y|Z}, \mathcal{R}_Z, A) = \mathcal{I}_1(\mathcal{F}_{Y|Z}, \mathcal{R}_Z, A) \cap \mathcal{I}_2(\mathcal{F}_{Y|Z}, \mathcal{R}_Z, A), \tag{3}$$

where $\mathcal{I}_1(\mathcal{F}_{Y|Z}, \mathcal{R}_Z, A)$*, associated with semi-infinite intervals, is:*

$$\mathcal{I}_1(\mathcal{F}_{Y|Z}, \mathcal{R}_Z, A) \equiv \left\{(m, \mathcal{G}_{U|Z}) \in \mathcal{M}_A : \forall t \in \mathbb{R}, G_{U|Z}((-\infty, t]|z) \geq C(z,t,m) \ \text{a.e. } z \in \mathcal{R}_Z\right\}, \tag{4}$$

and $\mathcal{I}_2(\mathcal{F}_{Y|Z}, \mathcal{R}_Z, A)$*, associated with finite intervals, is:*

$$\begin{aligned}\mathcal{I}_2(\mathcal{F}_{Y|Z}, \mathcal{R}_Z, A) \equiv \big\{(m, \mathcal{G}_{U|Z}) \in \mathcal{M}_A : \forall [t_1, t_2] \subset \mathbb{R},& \\ G_{U|Z}([t_1, t_2]|z) \geq \Delta(z, t_1, t_2, m) \ \text{a.e. } z \in \mathcal{R}_Z\big\}.&\end{aligned} \tag{5}$$

The proof of the Proposition, given in the Appendix, follows from application of the inequality

$$G_{U|Z}(\mathcal{S}|z) \geq \mathbb{P}[\mathcal{U}(Y, Z, m) \subseteq \mathcal{S}|Z = z] \tag{6}$$

to sets $\mathcal{S}$ comprising certain intervals on the real line.

The set

$$\mathcal{A}(\mathcal{S}, z, m) \equiv \{y : \mathcal{U}(y, z, m) \subseteq \mathcal{S}\} \tag{7}$$

contains the values of Y that for structural function m only occur when U takes a value in S. Accordingly $Y \in \mathcal{A}(\mathcal{S}, z, m) \implies U \in \mathcal{S}$ from which the inequality (6) follows directly.[10]

The probability $\mathbb{P}[\mathcal{U}(Y, Z, h) \subseteq \mathcal{S}|Z = z]$, is known as a *containment probability*.[11] For intervals $(-\infty, t]$ that are unbounded below:

$$\mathbb{P}[\mathcal{U}(Y, Z, m) \subseteq (-\infty, t]\,|Z = z] = C(z, t, m). \tag{8}$$

[10] Other characterizations of identified sets are available. One such will be employed when we consider the force of the restriction that unobserved U is mean independent of Z. All of the characterizations follow from the result that a structure $(m, G_{U|Z})$ is in the identified set if and only if for all z in the support of Z the distribution $G_{U|Z}(\cdot|z)$ is selectionable with respect to the conditional distribution of the random set $\mathcal{U}(Y, Z; m)$ given $Z = z$ induced by the distribution of Y given $Z = z$ delivered by the process under study. Definitions and details are in CR.

[11] See Molchanov and Molinari (2018).

For intervals $[t_1, t_2]$, with $t_1 > -\infty$ but with $t_2 \geq t_1$ unrestricted:

$$\mathbb{P}[\mathcal{U}(Y, Z, m) \subseteq [t_1, t_2] | Z = z] = \Delta(z, t_1, t_2, m), \tag{9}$$

from which it also follows that for intervals $[t_1, \infty)$ with $t_1 > -\infty$:

$$\mathbb{P}[\mathcal{U}(Y, Z, m) \subseteq [t_1, \infty) | Z = z] = P[Y_1 > 0 | z] - B(z, t_1, m). \tag{10}$$

It will be useful later to have an expression for the containment probability which applies for intervals $[t_1, t_2]$ with t_1 finite or infinite, as follows.

$$\begin{aligned} \mathbb{P}[\mathcal{U}(Y, Z, m) \subseteq [t_1, t_2] | Z = z] = 1[t_1 = -\infty] \times \mathbb{P}[Y_1 = 0 \wedge 0 \leq m(Y_2, Z, t_2) | Z = z] + \\ \mathbb{P}[Y_1 > 0 \wedge m(Y_2, Z, t_1) \leq Y_1 \leq m(Y_2, Z, t_2) | Z = z] \end{aligned} \tag{11}$$

Example (continued). *In the GLT model there is*

$$B(z, t, \theta) = \mathbb{P}[0 < Y_1 \leq \alpha Y_2 + \beta z + t | Z = z]$$

$$D(z, t, \theta) = \mathbb{P}[Y_1 = 0 \wedge 0 \leq \alpha Y_2 + \beta z + t | Z = z]$$

which can be estimated using realizations of observable variables (Y, Z)*, and*

$$G_{U|Z}([t_1, t_2] | z) = \Phi\left(\frac{t_2}{\sigma}\right) - \Phi\left(\frac{t_1}{\sigma}\right). \tag{12}$$

Throughout the paper Φ *and* ϕ *denote respectively the standard Gaussian distribution and density function.*

3.2 Characterizations using singly-infinite systems of moment inequalities

The sets $\mathcal{I}_1(\mathcal{F}_{Y|Z}, \mathcal{R}_Z, A)$ and $\mathcal{I}_2(\mathcal{F}_{Y|Z}, \mathcal{R}_Z, A)$ are determined by systems of respectively singly- and doubly-infinite moment inequalities. Under additional restrictions that imply that $B(z, t, m)$ is everywhere differentiable in t the doubly-infinite system can be replaced with an equivalent singly-infinite system. This can have computational advantages.

With U continously distributed conditional on Z and $\widetilde{G}_{U|Z}(\cdot | z)$ denoting the conditional cumulative distribution function of U given $Z = z$ the condition

$$G_{U|Z}([t_1, t_2]|z) \geq \Delta(z, t_1, t_2, m)$$

that appears in $\mathcal{I}_2(\mathcal{F}_{Y|Z}, \mathcal{R}_Z)$ can be expressed as

$$\widetilde{G}_{U|Z}(t_2|z) - \widetilde{G}_{U|Z}(t_1|z) \geq \Delta(z, t_1, t_2, m).$$

From this, with a differentiability restriction, Proposition 2 provides a singly-infinite moment inequality characterization for the set $\mathcal{I}_2(\mathcal{F}_{Y|Z}, \mathcal{R}_Z, A)$ originally defined in (5).

Proposition 2 *If $m(y_2, z, t)$ is everywhere differentiable with respect to t for all values of (y_2, z) and U is continuously distributed given Z with conditional density $\widetilde{g}_{U|Z}(\cdot|z)$ then*

$$\mathcal{I}_2(\mathcal{F}_{Y|Z}, \mathcal{R}_Z, A) = \left\{(m, \mathcal{G}_{U|Z}) \in \mathcal{M}_A : \forall t \in \mathbb{R},\ \widetilde{g}_{U|Z}(t|z) \geq b(z, t, m)\ \text{a.e. } z \in \mathcal{R}_Z\right\}, \quad (13)$$

where $b(z, t, m) \equiv \nabla_t B(z, t, m)$ is the partial derivative of $B(z, t, m)$ with respect to t.

Example (continued). *In the GLT model, $\widetilde{g}_{U|Z}(t|z) = \sigma^{-1}\phi(\frac{t}{\sigma})$ and $b(z, t, m)$ is the density function of $Y_1 - \alpha Y_2 - \beta z$ conditional on $Z = z$.*

3.3 Restrictions on the influence of exogenous Z on the structural function

In applications there will be restrictions on the influence of exogenous Z on the structural function. Consider Restriction ZD which requires that m depends on z solely through the variation in a possibly vector-valued function, $w(z)$, that arises as z varies across the support of Z. In a commonly occurring case with $z = (z_1, z_2)$, and z_2 excluded from the structural function, $w(z) = z_1$. Index restrictions, for example requiring that there exists γ such that for all z, $w(z) = \gamma z$ are also commonly employed.

Restriction ZD: Restricted Z dependence

$$\exists w(\cdot) \quad s.t. \quad \forall (z, z') \in \mathcal{R}_Z \times \mathcal{R}_Z, \forall (y_2, u), \quad w(z) = w(z') \implies m(y_2, z, u) = m(y_2, z', u)$$

Define the set of values that $w(z)$ can take as z varies across its support

$$\mathcal{W}(\mathcal{R}_Z) \equiv \{w(z) : z \in \mathcal{R}_Z\}$$

and for each element, w, of this set define the set of values of z such that $w(z) = w$:[12]

$$\mathcal{Z}(w, \mathcal{R}_Z) \equiv \{z \in \mathcal{R}_Z : w(z) = w\}.$$

In the case in which there is a *stochastic independence condition* so that $\mathcal{G}_{U|Z} = \{G_U\}$ with U continuously distributed the sets $\mathcal{I}_1(\mathcal{F}_{Y|Z}, \mathcal{R}_Z, A)$ and $\mathcal{I}_2(\mathcal{F}_{Y|Z}, \mathcal{R}_Z, A)$ are then as follows[13], where, recall, $\widetilde{G}_U(t)$ and $\widetilde{g}_U(t)$ are respectively the marginal distribution and density functions of U.

$$\mathcal{I}_1(\mathcal{F}_{Y|Z}, \mathcal{R}_Z, A) = \left\{ (m, \mathcal{G}_{U|Z}) : \forall w \in \mathcal{W}(\mathcal{R}_Z),\ t \in \mathbb{R},\ \widetilde{G}_U(t) \geq \sup_{z \in \mathcal{Z}(w, \mathcal{R}_Z)} C(z, t, m) \right\},$$

$$\mathcal{I}_2(\mathcal{F}_{Y|Z}, \mathcal{R}_Z, A) = \left\{ \left(m, \mathcal{G}_{U|Z}\right) : \forall w \in \mathcal{W}(\mathcal{R}_Z),\ t \in \mathbb{R},\ \widetilde{g}_U(t) \geq \sup_{z \in \mathcal{Z}(w, \mathcal{R}_Z)} b(z, t, m) \right\},$$

and the identified set of structures is as follows.

$$\mathcal{I}(\mathcal{F}_{Y|Z}, \mathcal{R}_Z, A) = \mathcal{I}_1(\mathcal{F}_{Y|Z}, \mathcal{R}_Z, A) \cap \mathcal{I}_2(\mathcal{F}_{Y|Z}, \mathcal{R}_Z, A)$$

3.4 Upper and lower bounds and moment equalities

The containment inequality (6) used to produce Proposition 1 provides a lower bound on the distribution of U. It is shown in CR that applying the inequality in (6) to the complement $\mathcal{S}^c$ of a set $\mathcal{S}$ delivers the following inequality satisfied by all structures in the identified set for all $z \in \mathcal{R}_Z$ and all closed sets $\mathcal{S}$ on the support of U.[14]

$$G_{U|Z}(\mathcal{S}|z) \leq \mathbb{P}[\mathcal{U}(Y, Z, m) \cap \mathcal{S} \neq \emptyset | Z = z] \tag{14}$$

So, for all structures in the identified set $\mathcal{I}(\mathcal{F}_{Y|Z}, \mathcal{R}_Z, A)$ the inequalities

$$\mathbb{P}[\mathcal{U}(Y, Z, m) \cap \mathcal{S} \neq \emptyset | Z = z] \geq G_{U|Z}(\mathcal{S}|z) \geq \mathbb{P}[\mathcal{U}(Y, Z, m) \subseteq \mathcal{S} | Z = z] \tag{15}$$

[12] When Z is entirely excluded from the structural function define $w(z) = c$ where c is some constant, for example $c = 0$. In this case $\mathcal{W}(\mathcal{R}_Z) = \{c\}$ and for all w, $\mathcal{Z}(w, \mathcal{R}_Z) = \mathcal{R}_Z$.

[13] In order to deal with possibilities of zero measure sets and conditions required to hold almost everywhere, here and throughout the paper the sup and inf operators are to be understood to mean "essential supremum" and "essential infimum" when applied to functions of realizations of random variables. So for instance $\sup_{z \in \mathcal{Z}} f(z)$ indicates the smallest value c such that $f(Z) \leq c$ with probability one given $Z \in \mathcal{Z}$.

[14] The symbol $\emptyset$ denotes the empty set.

hold for all $z \in \mathcal{R}_Z$ and all intervals, $\mathcal{S}$, on the real line.

There are moment *equalities* in the characterization of the identified set of structures when there are z and $\mathcal{S}$ such that the probabilities on the left and right hand sides of (15) are equal. In a sufficiently restrictive model and for particular collections of distributions $\mathcal{F}_{Y|Z}$ and support $\mathcal{R}_Z$, there is the possibility that these moment inequalities deliver point identification.

In the IV Tobit model the upper bounding probability $\mathbb{P}[\mathcal{U}(Y,Z,m) \cap \mathcal{S} \neq \emptyset | Z = z]$ is as follows.

$$\begin{aligned}\mathbb{P}[\mathcal{U}(Y,Z,m) \cap [t_1,t_2] \neq \emptyset | Z = z] &= \mathbb{P}[(Y_1 = 0) \wedge (0 \geq m(Y_2,Z,t_1)) | Z = z] \\ &\quad + \mathbb{P}[(Y_1 > 0) \wedge (m(Y_2,Z,t_1) \leq Y_1 \leq m(Y_2,Z,t_2)) | Z = z]. \quad (16)\end{aligned}$$

Considering (8), (9), and (16), bounding probabilities in (15) are equal for semi-infinite intervals $(-\infty, t]$ when z and m are such that

$$P[Y_1 = 0 | Z = z] = P[Y_1 = 0 \wedge 0 \leq m(Y_2, Z, t) | Z = z]$$

and for finite intervals $[t_1, t]$ when z and m are such that

$$P[Y_1 = 0 \wedge 0 \geq m(Y_2, Z, t) | Z = z] = 0.$$

Both conditions are satisfied when $m(Y_2, Z, t) > 0$ almost surely conditional on $Z = z$. A leading case in which this can occur is when the endogenous explanatory variable Y_2 has bounded support and the function m is unbounded above as t becomes large.

Conditions such that the inequalities (15) reduce to equalities for some values of z, can be the basis for establishing sufficient conditions for point identification. For example, in models for censored outcomes with a conditional median restriction (Hong and Tamer, 2003, p. 908) provide support conditions under which certain resulting moment equalities can establish point identification and a $\sqrt{n}$-consistent and asymptotically normal estimator of model parameters when m is linear in parameters. Under the restriction that $med\,(U|Z) = 0$ and $m(Y_2, Z, 0) > 0$ almost surely conditional on $Z = z$ with m linear in parameters and additive U there is in our notation

$$med\,(Y_1 - \alpha Y_2 - \beta Z | Z = z) = 0,$$

which corresponds to the moment equality delivered by the inequalities (15) applied to the set $(-\infty, 0]$. A condition requiring that the set of values of $z \in \mathcal{R}_Z$ such that $m(Y_2, Z, 0) > 0$ almost surely conditional on $Z = z$ has positive measure, in conjunction with a condition requiring sufficient variation in included endogenous variables conditional on instruments, is then used to establish sufficient conditions for point identification in Lemma 2 of Hong and Tamer (2003).

4 The impact of restrictions on the dependence between U and Z

In this Section we consider the identifying power of a conditional mean independence restriction, a conditional quantile independence restriction focussing on median independence, a stochastic independence restriction with no parametric specification of the distribution G_U and stochastic independence restriction with U restricted to be Gaussian.

We present characterizations of identified sets for models in which

$$Y_1 = \max(0, m(Y_2, Z, U))$$

where there may be a parametric, semiparametric or nonparametric specification of the function m. As each type of restriction is considered we provide calculations of features of identified sets for the case in which models specify

$$m(Y_2, Z, U) = \beta + \alpha Y_2 + U.$$

In these calculations we use probabilities[15]

- Structure 1

$$\begin{aligned} Y_1 &= \max(0, b + aY_2 + U_1) \\ Y_2 &= d_0 + d_1 Z + U_2 \end{aligned}$$

[15] Probabilities delivered by Structure 1 are calculated using the `pmvnorm` function in the `mvtnorm` package Genz, Bretz, Miwa, Mi, Leisch, Scheipl, and Hothorn (2021) which references Genz and Bretz (2009) and is implemented in R, R Core Team (2022). Probabilities delivered by Structure 2 are calculated using the `quadinf` function in R's `pracma` package, Borchers (2022), delivered by two specific structures, as follows.

- Structure 2

$$\begin{aligned} Y_1 &= \max(0, b + aY_2 + U_1) \\ Y_2 &= k - \exp(-d_0 - d_1 Z - U_2) \end{aligned}$$

In both cases

$$U = \begin{bmatrix} U_1 \\ U_2 \end{bmatrix} \sim N_2 \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} s_{11} & s_{12} \\ s_{12} & s_{22} \end{bmatrix} \right)$$

with U and Z independently distributed and in all the calculations the support of scalar Z is

$$\mathcal{R}_Z = \{-1, -0.9, -0.8, \ldots, 0, \ldots, 0.8, 0.9, 1\}.$$

A variety of parameter values are employed in the illustrations. These are set out in Tables 1 and 2 which give information about the incidence and extent of censoring delivered by each structure and parameter constellation and projections of the identified sets (intervals) onto the space of α, the coefficient on endogenous Y_2, obtained under alternative restrictions on the distribution of U and Z.

It is important to understand that these are specifications of *complete structures*. Only for complete structures can the probabilities that appear in the characterizations of identified sets be calculated. The *incomplete models* that we consider do not employ all the restrictions embodied in the complete structures. None of the models specify a structural equation for Y_2 and in only one of the models is the unobserved variable in the structural equation for Y_1 restricted to be Gaussian.

4.1 Mean independence

Consider models in which U is restricted to be *mean independent* of the instrumental variables. Absent censoring such a model is a linear in parameters IV model which is point identifying under a suitable rank condition.

RESTRICTION MI - Mean Independence: *Let $\mathcal{G}_{U|Z}$ comprise all collections of conditional distributions for U given Z, $G_{U|Z}$, satisfying $E[U|z] = 0$, a.e. $z \in \mathcal{R}_Z$.*

Manski and Tamer (2002) uses mean independence restrictions conditional on included exogenous variables in models with censored outcomes or covariates but no endogenous explanatory variables. Here we impose an IV version of a conditional mean restriction, conditioning on values of excluded instruments and any included exogenous variables. Restriction MI appears in CR but here, to simplify, the value of the conditional expectation is restricted to be zero rather than a member of a specified set of values. Modifying Theorem 5 of CR delivers Proposition 3.

Proposition 3 *Under Restriction MI the identified set for the function m is*

$$\{m \in \mathcal{M}_A^0 : 0 \leq E[m^{-1}(Y_2, Z, Y_1)|Z = z] \ a.e. \ z \in \mathcal{R}_Z\}.$$

Example (continued). *In a linear index Tobit model in which $m(y_2, z, u) = \alpha y_2 + z\beta + u$ under Restriction MI there is*

$$E[m^{-1}(Y_2, Z, Y_1)|Z = z] = E[Y_1|z] - \alpha E[Y_2|z] - \beta z.$$

Write $Z = (Z_1, Z_2)$ and suppose elements in β are restricted to be zero so that the variables Z_2 are excluded from the structural function. Then Restriction ZD[16] *holds with $w(z) = z_1$ and there is $\mathcal{W}(\mathcal{R}_Z) \equiv \{z_1 : z \in \mathcal{R}_Z\}$ and $\mathcal{Z}(w, \mathcal{R}_Z) = \{z \in \mathcal{R}_Z : z_1 = w\}$. The identified set for (α, β) is*

$$\left\{(\alpha, \beta) : \forall w \in \mathcal{W}(\mathcal{R}_Z), \quad 0 \leq \inf_{z \in \mathcal{Z}(w, \mathcal{R}_Z)} (E[Y_1|z] - \alpha E[Y_2|z] - \beta z)\right\}$$

which is an intersection of linear half spaces and therefore a convex set. When, as in the illustrations presented here, there are no included exogenous variables in the structural equation for the censored outcome, so $\beta z = \beta$ is simply a scalar intercept and the identified set is as follows.

$$\left\{(\alpha, \beta) : \quad 0 \leq \inf_{z \in \mathcal{R}_Z} (E[Y_1|z] - \alpha E[Y_2|z] - \beta)\right\}. \tag{17}$$

The four panes of Figure 1 show identified sets for (α, β) for Cases 1-3 of Structure 1 (see Table 1) and Case 1 of Structure 2 (see Table 2) under a variety of independence restrictions. The value of a and b in the structure that generates the probabilities used in the calculations is the green plotted point. The identified sets under *mean* independence comprise the regions below all the blue straight lines. The projections of the sets onto the space of α is the entire real line.[17] The projections of the sets onto the space of β is the entire real line unless there

[16] See Section 3.3.

[17] This is the case for all the parameter values considered so projections under mean independence are not reported in Tables 1 and 2.

exist z and z' in $\mathcal{R}_Z$ such that $E[Y_2|z] \leq 0 \leq E[Y_2|z']$ (a condition which holds with strong inequalities in the cases pictured in Figure 1) when the projection is a semi-infinite interval with finite upper limit.

Clearly a conditional *expectation* restriction does not lead to informative identified sets in these examples. We next consider the identifying power of conditional *quantile* restrictions.

4.2 Quantile independence

Now consider the identifying power of quantile independence restrictions.

RESTRICTION QI - Quantile Independence: *Let $\mathcal{G}_{U|Z}$ comprise all collections of conditional distributions of U given Z satisfying $\mathbb{P}[U \leq q_j|z] = \lambda_j$, a.e. $z \in \mathcal{R}_Z$ for all $j \in \{1, ..., J\}$ where $\Lambda \equiv \{\lambda_1, ..., \lambda_J\}$ is a collection of specified known values, and some collection of values $\{q_1, ..., q_J\} \in \mathcal{Q}$ with λ_j and q_j both increasing in j and $\mathcal{Q}$ a specified set of possible values of $\{q_1, ..., q_J\}$.*

Restriction QI restricts the J conditional quantiles of U given $Z = z$ specified by Λ to be invariant with respect to z. $\mathcal{Q}$ is a known set of possible values for these conditional quantiles. For example, a conditional median restriction corresponds to $\Lambda = \{0.5\}$ and the usual normalization that this conditional median is zero is then captured by setting $\mathcal{Q} = \{0\}$. In this case $J = 1$. However, Restriction QI allows one to restrict the conditional distributions of U to be Z-invariant at multiple quantiles.[18] Unrestricted quantile values are added to the list of unknown model parameters. In the application to tobacco expenditure shares in Section 5.2 multiple quantile restrictions are employed.

Sharp characterization of the identified set of structures is obtained by considering test sets comprising intervals of the form $(-\infty, q_j]$ for all $j = 1, ..., J$ and $[q_j, q_{j+1}]$ for all $j = 0, ..., J$ where $q_0 = -\infty$ and $q_{J+1} \equiv \infty$.

Proposition 4 *Let Restriction QI hold. Then the identified set of structural functions delivered by the IV Tobit Model is the set of functions $m \in \mathcal{M}_A^0$ such that for some $\{q_1, ..., q_J\} \in \mathcal{Q}$, it holds that $\forall j \in \{0, 1, ..., J\}$: (1) $C(z, q_j, m) \leq \lambda_j$, and (2) $\Delta(z, q_j, q_{j+1}, m) \leq \lambda_{j+1} - \lambda_j$.*

In numerical illustrations considered now, Restriction QI is imposed employing a single median independence restriction such that $J = 1$ and $\lambda_1 = 0.5$ with the normalization $q_1 = 0$. Thus $\mathcal{Q} = \{0\}$, the test sets employed are simply $\{(-\infty, 0), (0, \infty)\}$. Employing the

[18]It is straightforward to impose a symmetry restriction.

inequalities (8) and (10) with $G_{U|Z}((-\infty,0]|z) = 0.5$ delivers the identified set of structural functions, m, under the median independence restriction, as follows.

$$\begin{aligned}\{m : \mathbb{P}[Y_1 = 0 \wedge 0 \leq m(Y_2, Z, 0)|Z = z] + \mathbb{P}[0 < Y_1 \leq m(Y_2, Z, 0)|Z = z]&\\ \leq 0.5 \leq&\\ \mathbb{P}[Y_1 = 0|z] + \mathbb{P}[0 < Y_1 \leq m(Y_2, Z, 0)|Z = z] \text{ a.e. } z \in \mathcal{R}_Z\}.&\end{aligned}$$

Example (continued). *In a linear index Tobit model with no included exogenous variables such that $m(y_2, z, u) = \alpha y_2 + \beta + u$, under Restriction QI with $\Lambda = \{0.5\}$ and $\mathcal{Q} = \{0\}$ the identified set of values of α and β is as follows.*

$$\begin{aligned}\{(\alpha, \beta) : \mathbb{P}[Y_1 = 0 \wedge 0 \leq \alpha Y_2 + \beta|Z = z] + \mathbb{P}[0 < Y_1 \leq \alpha Y_2 + \beta|Z = z]&\\ \leq 0.5 \leq&\\ \mathbb{P}[Y_1 = 0|z] + \mathbb{P}[0 < Y_1 \leq \alpha Y_2 + \beta|Z = z] \text{ a.e. } z \in \mathcal{R}_Z\}.&\end{aligned}$$

Identified sets of values of α and β under a zero median restriction are the pink shaded regions drawn in Figure 1.

Structure 1 Case 1 delivers a very small identified set. This is very substantially enlarged on moving to Structure 1 Case 2 and then to Structure 1 Case 3. This occurs because the instrument Z becomes a much less informative predictor of the value of Y_2 - the value of d_1 is reduced by 50% on moving to Case 2 and by a further 50% on moving to Case 3. The lower right pane shows the identified set delivered by Structure 2 Case 1 under a conditional median independence restriction.

In the cases illustrated in the lower two panes of Figure 1 the identified intervals for α are unbounded above. Projections of identified sets onto the space of α obtained under conditional median independence are shown for these cases and a variety of other cases in the final two columns of Tables 1 and 2.

Considering Structure 1 (Table 1) there are two cases (6 and 9) with extensive censoring in which the conditional median independence restriction is uninformative and the identified interval for α is the entire real line. The cases in which the interval is semi-finite, unbounded above (3, 4, 7 and 8), either have an instrument with low predictive power (Case 3, $d_1 = 0.25$) or extensive censoring (Cases 4, 7 and 8). Case 1 of Structure 2 delivers a censoring probability exceeding 50% on average and a semi-finite interval for α. The other cases of Structure 2 illustrate quite extreme situations with low censoring (Cases 2 and 3) or highly informative instruments (Cases 4 with $d_1 = 5$ and Case 5 with $s_{22} = 0.1$). In all these cases the conditional median independence restriction is highly informative, delivering very short identified intervals for α.

The median independence restriction can deliver substantially more informative identified sets than the mean independence restriction and the identified sets it delivers can be very small indeed but when there is extensive censoring or relatively uninformative instruments the identified intervals can be unbounded.

4.3 Stochastic independence

Now consider a restriction requiring U and Z to be independently distributed but with *no* parametric specification of the distribution of U. We shortly consider the effect of imposing additionally a Gaussian restriction.

RESTRICTION NPSI - Nonparametric Stochastic Independence: *Unobservable random variable U is continuously distributed and stochastically independent of Z such that $G_{U|Z}$ is the singleton set $\{G_U\}$.*

Recall the characterization of identified sets given in Section 3, repeated here for convenience with the restriction NPSI imposed. The identified set of structures for a model A, comprising restrictions $\mathcal{M}_A$, $\mathcal{I}(\mathcal{F}_{Y|Z}, \mathcal{R}_Z, A)$, is the intersection of two sets.

$$\mathcal{I}(\mathcal{F}_{Y|Z}, \mathcal{R}_Z, A) = \mathcal{I}_1(\mathcal{F}_{Y|Z}, \mathcal{R}_Z, A) \cap \mathcal{I}_2(\mathcal{F}_{Y|Z}, \mathcal{R}_Z, A) \tag{18}$$

where

$$\mathcal{I}_1(\mathcal{F}_{Y|Z}, \mathcal{R}_Z, A) = \left\{ (m, \{G_U\}) \in \mathcal{M}_A : \forall t \in \mathbb{R}, \widetilde{G}_U(t) \geq \sup_{z \in \mathcal{R}_Z} C(z, t, m) \right\} \tag{19}$$

$$\begin{aligned} \mathcal{I}_2(\mathcal{F}_{Y|Z}, \mathcal{R}_Z, A) \equiv \{ & (m, \{G_U\}) \in \mathcal{M}_A : \forall [t_1, t_2] \subset \mathbb{R}, \\ & \left. \tilde{G}_U(t_2) - \tilde{G}_U(t_1) \geq \sup_{z \in \mathcal{R}_Z} \Delta(z, t_1, t_2, m) \right\}. \end{aligned} \tag{20}$$

Here $\tilde{G}_U(t) \equiv G_U((-\infty, t])$ is the distribution function of U,

$$\Delta(z, t_1, t_2, m) \equiv B(z, t_2, m) - B(z, t_1, m)$$

$$B(z,t,m) \equiv \mathbb{P}[Y_1 > 0 \wedge Y_1 \leq m(Y_2, Z, t)|z]$$

and

$$C(z,t,m) \equiv \mathbb{P}[Y_1 = 0 \wedge 0 \leq m(Y_2, Z, t)|z] + \mathbb{P}[Y_1 > 0 \wedge Y_1 \leq m(Y_2, Z, t)|z].$$

In the linear case used in the illustrations in which there is no exogenous variable in the structural equation, these functions are as follows.

$$B(z,t,\theta) = \mathbb{P}[Y_1 > 0 \wedge Y_1 \leq \alpha Y_2 + \beta + t|z]$$

$$C(z,t,\theta) = \mathbb{P}[Y_1 = 0 \wedge 0 \leq \alpha Y_2 + \beta + t|z] + \mathbb{P}[Y_1 > 0 \wedge Y_1 \leq \alpha Y_2 + \beta + t|z].$$

In this case we use $\theta = (\alpha, \beta)$ instead of m as an argument of these functions. Absent additional restrictions on the distribution of U, the distribution of observable variables contains no information about the value of the constant term β, so, in determining the identified set of values of α, β can be normalized, set equal to an arbitrary value. It is set to zero in the numerical illustrations.

We now set out a method for calculating an outer set for the identified set of structures onto the space of structural functions under the NPSI restriction. For this purpose partition the support of U into N intervals: $(-\infty, t_1]$, $(t_1, t_2]$, $\ldots, (t_{N-1}, \infty)$ where N is large. For each $n \in \{1, ..., N\}$ define

$$p_n \equiv G_U\left((t_{n-1}, t_n]\right), \tag{21}$$

where it is understood that $(t_0, t_1]$ means $(-\infty, t_1]$ and $(t_{N-1}, t_N]$ means (t_{N-1}, ∞).

It follows from Proposition 1 that for any structural function m such that there exists a distribution G_U such that (m, G_U) is in the identified set, there must exist probabilities $p_1, \ldots, p_N$ each nonnegative and summing to one such that

$$\forall n = 1, ..., N: \quad \sum_{i=1}^{n} p_i \geq C(z, t_n, m), \tag{22}$$

$$\forall n = 1, ..., N: \quad p_n \geq B(z, t_n, m) - B(z, t_{n-1}, m), \tag{23}$$

for almost every $z \in \mathcal{R}_Z$.

Let $\mathcal{I}\left(\mathcal{F}_{Y|Z}, \mathcal{R}_Z, A, \mathcal{T}\right)$ denote the set of admissible functions m that satisfy these inequalities using the partition of intervals whose endpoints are consecutive elements of $\mathcal{T} \equiv \{t_0, t_1, ..., t_N\}$. Inequalities of the form (22) correspond to those defining $\mathcal{I}_1(\mathcal{F}_{Y|Z}, \mathcal{R}_Z, A)$ in the statement of Proposition 1 with $t = t_n$. Inequalities of the form (23) are those characterizing

$\mathcal{I}_2(\mathcal{F}_{Y|Z}, \mathcal{R}_Z, A)$ with t_{n-1} and t_n in place of t_1 and t_2, respectively.[19] Indeed, Proposition 1 implies that the identified set for m are those admissible structural functions that satisfy these inequalities for all t_n and t_{n-1} for some possible distribution G_U, with $G_U((t_{n-1}, t_n])$ replacing p_n, almost surely. Indeed, it is precisely the use of only a finite set of intervals $(t_{n-1}, t_n]$ that makes the resulting characterization non-sharp.

For any collection of N intervals the conditions determining when structural function m is in $\mathcal{I}\left(\mathcal{F}_{Y|Z}, \mathcal{R}_Z, A, \mathcal{T}\right)$ can be checked by solving a linear program as set out now. We use the method developed in Theorem 4 of Chesher and Rosen (2020a) which explores the identifying power of models of interdependent determination of discrete outcomes.

Define the following objects.[20]

$$\underset{1\times N}{A} \equiv \underbrace{[1, 1, \dots, 1]}_{N \text{ times}} \qquad \underset{N\times 1}{x} \equiv [p_1, \dots, p_N]' \qquad b = 1 \tag{24}$$

$$\underset{R(2N-1)\times N}{\mathbf{B}} = \begin{bmatrix} \mathbf{B}^* \\ \vdots \\ \mathbf{B}^* \end{bmatrix} \qquad \underset{R(2N-1)\times 1}{c(m)} = \begin{bmatrix} c(z_1, m) \\ \vdots \\ c(z_R, m) \end{bmatrix} \qquad R = \sharp(\mathcal{R}_Z) \tag{25}$$

$$\underset{(2N-1)\times N}{\mathbf{B}^*} = - \begin{bmatrix} 1 & 0 & \cdots & & 0 & 0 \\ 0 & 1 & \cdots & & 0 & 0 \\ \vdots & & \ddots & & \vdots & \vdots \\ \cdots & & & \ddots & & \\ 0 & 0 & \cdots & \cdots & 1 & 0 \\ 0 & 0 & \cdots & \cdots & 0 & 1 \\ 1 & 0 & 0 & \cdots & 0 & 0 \\ 1 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & & & \vdots & \vdots \\ 1 & 1 & 1 & \cdots & 0 & 0 \\ 1 & 1 & 1 & \cdots & 1 & 0 \end{bmatrix} \qquad \underset{(2N-1)\times 1}{c(z,m)} = \begin{bmatrix} B(z, t_0, m) - B(z, t_1, m) \\ B(z, t_1, m) - B(z, t_2, m) \\ \vdots \\ \vdots \\ B(z, t_{N-2}, m) - B(z, t_{N-1}, m) \\ B(z, t_{N-1}, m) - B(z, t_N, m) \\ -C(z, t_1, m) \\ -C(z, t_2, m) \\ \vdots \\ -C(z, t_{N-2}, m) \\ -C(z, t_{N-1}, m) \end{bmatrix} \tag{26}$$

Application of Theorem 4 of Chesher and Rosen (2020a) to the characterization of $\mathcal{I}\left(\mathcal{F}_{Y|Z}, \mathcal{R}_Z, A, \mathcal{T}\right)$ given by (22) and (23) then yields the following result.

[19] These are equivalently conditional containment probabilities applied to intervals of the form $(-\infty, t_n]$ and $(t_{n-1}, t_n]$, respectively.

[20] This is set out in the notation in Chesher and Rosen (2020a) which employs matrix B, here denoted $\mathbf{B}$, not to be confused with the function $B(z, t, m)$.

Proposition 5 *Let $\mathcal{T} \equiv \{t_0, t_1, ..., t_N\}$ be an increasing sequence of scalars with $t_N = -t_0 = \infty$. In an IV Tobit Model A restricting structural function m to the set $\mathcal{M}_A^0$ and in which Restriction NPSI holds, the set*

$$\mathcal{I}\left(\mathcal{F}_{Y|Z}, \mathcal{R}_Z, A, \mathcal{T}\right) = \left\{m \in \mathcal{M}_A^0 : v^*(m) \geq 0\right\}$$

comprises bounds on m, where $v^*(m)$ is the value attained by the linear program

$$\min_{s,t,v} v \tag{27}$$

subject to the constraints

$$sA + t\mathbf{B} \geq 0, \tag{28}$$

$$t \geq 0 \tag{29}$$

and

$$s + t \cdot c(m) \leq v, \tag{30}$$

where $s \in \mathbb{R}$, $t \in \mathbb{R}^K$, $v \in \mathbb{R}$, and $K = R(2N-1)$.[21]

The outer set approximation to the identified set of structural functions can be made closer to the sharp identified set by finer choice of $\mathcal{T}$.[22] Many points $t_1, ..., t_N$ can be used because large linear programs can be solved very quickly. In the illustrations $N = 100$ points are used.[23]

Turning to the illustrative calculations, Tables 1 and 2 show identified sets (intervals $[\underline{\alpha}, \overline{\alpha}]$) for α delivered by the NPSI restriction using probabilities delivered by various cases of Structures 1 and 2.[24] In all cases these sets, which are strictly speaking outer regions, are subsets of the identified set for α obtained under the median independence restriction. For Cases 6 and 9 of Structure 1 which have extensive censoring, the identified sets under NPSI are bounded below unlike the sets obtained under median independence which were

[21] The proof of this proposition given in Chesher and Rosen (2020a) makes use of a version of Farkas' Alternative. Again we stay with the notation in Chesher and Rosen (2020a) which employs a decision variable t in R2, not to be confused with t_n used to signify boundaries of intervals that partition the support of U.

[22] The outer set is the sharp identified set under the additional restriction that the distribution G has a piecewise constant density on each of the N intervals defined by successive points in $\mathcal{T}$.

[23] The partition used in the illustrative calculations has $\mathcal{T} = \left\{\Phi^{-1}(s) : s \in \{0, 1/N, 2/N, \ldots, 1\}\right\}$.

[24] Calculations to compute $\underline{\alpha}$ and $\overline{\alpha}$ were done using R's `uniroot` function with refinement via grid search, as well as R's `lpSolveAPI` package, Konis and Schwendinger (2020).

uninformative. Case 3 of Structure 1 in which the instrument is relatively weak delivers a semi-finite interval which is an improvement on the semi-finite identified set obtained under median independence. The NPSI restriction delivers a finite interval for Case 1 of Structure 2 which has quite extensive censoring, a considerable improvement on the semi-finite interval obtained under median independence. In all the other cases of Structure 1 and Structure 2 the NPSI restriction delivers very short identified intervals. It is interesting to see that close to point identification can be obtained under a stochastic independence restriction without imposing a parametric specification on the distribution of unobserved U.

4.4 Parametric restrictions - the Gaussian IV Tobit model

Now consider models in which in addition to a stochastic independence restriction there is the requirement that the distribution of U belongs to a parametric family of distributions. In the illustrative examples here a Gaussian distribution is employed as in the classic Tobit model.

RESTRICTION GaussSI - Stochastic Independence - Gaussian U: *The random variables U and Z are independently distributed; $\mathcal{G}_{U|Z}$ is a singleton set $\{G_U\}$ where G_U is a Gaussian distribution with variance σ^2.*

Tables 1 and 2 show projections of identified sets for (α, β, σ) onto the space of α under the GaussSI restriction. In every case these projections are subsets of the ones obtained under the our implementation of the NPSI restriction.[25]

Calculations are done using the characterization given in (18), (19) and (20). The structural function is written as

$$Y_1 = \max(0, \alpha Y_2 + \beta + \sigma U\}$$

where U is now restricted Gaussian with variance 1. Inequalities delivered by a finite collection of intervals are employed using a partition $\mathcal{T} = \{\Phi^{-1}(s) : s \in \{0, 1/N, 2/N, \ldots, 1\}$.[26]

In all the cases considered the identified intervals for α under the GaussSI restriction are finite and intervals are very short[27] except in three cases. These are cases 6 and 9

[25] If sharp identified sets were being calculated this subsetting would have to occur but outer sets are being calculated here, so subsetting cannot be guaranteed.

[26] Calculations for Structure 1 use values of N as large as 1000; Calculations for Structure 2 which are much more time consuming since they involve numerical integration, use values of N as large as 100. In every case $N \geq 100$.

[27] An identified interval reported as $[1.00, 1.00]$ should not be taken to indicate that the model is point identifying when the structure under consideration generates the probability distribution used to calculate the interval. There may be point identification, but to obtain the GaussSI intervals a sequence of values of α

of Structure 1 which have very extensive censoring, and Case 1 of Structure 2 which has considerable censoring and a weak instrument relative to Cases 4 and 5 where censoring is of a similar degree but identified intervals for α are short.

In the calculations done under the various independence restrictions, when censoring is not extensive and instruments are good predictors of endogenous variables the IV Tobit model can be close to point identifying even under a weak conditional median independence restriction. Imposing a stochastic independence restriction produces very short identified intervals except when censoring is extensive or the instrument is weak and in these cases additionally specifying a parametric distribution for the unobserved variable shortens identified intervals considerably.

5 Implementation

This section finishes with an application to a Tobit model of tobacco expenditure using UK survey data from the period 2000-2009. First, the method employed to calculate confidence regions on projections of the identified sets is set out.

5.1 Inference

We employ a test statistic proposed in Belloni, Bugni, and Chernozhukov (2018) (BBC18) to calculate 95% confidence regions for functions of partially identified parameter vectors, including subvectors, partially identified by a large number of moment inequalites. We use a self-normalized critical value, shown to be asymptotically valid in Chernozhukov, Chetverikov, and Kato (2019) (CCK19).[28]

In applications there will typically be many exogenous variables, Z, and some of these may be continuous. In this circumstance it is hard to make progress using conditional moment inequalities in which conditioning is on Z taking singleton values.[29] Instead we

around $\alpha = 1$ was considered, stepping up and down from 1 by increments of 0.01. Accordingly it should be understood that identified intervals that are any subinterval of $[0.995, 1.004\dot{9}]$ can be the identified interval for α when the identified interval in Tables 1 and 2 is reported as $[1.00, 1.00]$.

[28] Belloni, Bugni, and Chernozhukov (2018) additionally provide theoretical justification for alternative critical values using a bootstrap procedure that can further refine these confidence sets. We employ the self-normalized critical value for its computational simplicity.

[29] When Z has rich support it will be difficult to obtain accurate estimates of conditional probabilities. Kernel or sieve estimation would lead to estimated moment functions that are not simple means of contributions which is required when the BBC18 procedure is used.

conduct inference on outer regions obtained from moment inequalities derived by considering conditioning on the event that Z takes a value in a *set* of values, $\mathcal{Z}$.

In the models that we employ the moment inequalities arising when an interval $[t_1, t_2]$ is considered are

$$w(t_1, t_2, \theta, z) \leq p(t_1, t_2, \theta, z). \tag{31}$$

Here, employing (11),

$$\begin{aligned} w(t_1, t_2, \theta, z) \equiv 1[t_1 = -\infty] \times \mathbb{P}[Y_1 = 0 \wedge 0 \leq \alpha Y_2 + \beta Z + t_2)|Z = z] + \\ \mathbb{P}[Y_1 > 0 \wedge \alpha Y_2 + \beta Z + t_1 \leq Y_1 \leq \alpha Y_2 + \beta Z + t_2 | Z = z] \end{aligned}$$

and

$$p(t_1, t_2, \theta, z) \equiv G_{U|Z}([t_1, t_2]|z)$$

is the probability mass placed on the interval $[t_1, t_2]$ by the distribution of U given $Z = z$ admitted by the model. This may depend on the value of the parameter θ.

In the empirical work we estimate using three models in which the Gaussian restriction is dropped and quantile independence is imposed at 3, 5 or 7 quantiles associated with selected probabilities. In these three cases, as set out at the start of Section 4.2, t_1 and t_2 are unknown values of quantiles at the selected specified probabilities and $p(t_1, t_2, \theta, z)$ is the difference between those probabilities, independent of z. In these two cases the unknown values of the quantiles are elements of the parameter vector θ. In all the cases considered in the application $p(t_1, t_2, \theta, z)$ does not depend on z which we make explicit now by writing the probability as $p(t_1, t_2, \theta)$.

Let the distribution function of Z be $F_Z(z)$. If for some value of θ, t_1 and t_2 the moment conditions (31) hold for all z there is, for all sets $\mathcal{Z} \subseteq \mathcal{R}_Z$

$$\int_{\mathcal{Z}} w(t_1, t_2, \theta, z) dF_Z(z) \leq p(t_1, t_2, \theta) \int_{\mathcal{Z}} dF_Z(z)$$

and thus the moment conditions imply that, with

$$\begin{aligned} \tilde{w}(t_1, t_2, \theta, \mathcal{Z}) \equiv 1[t_1 = -\infty] \times \mathbb{P}[Y_1 = 0 \wedge 0 \leq \alpha Y_2 + \beta Z + t_2 \wedge Z \in \mathcal{Z}] + \\ \mathbb{P}[Y_1 > 0 \wedge \alpha Y_2 + \beta Z + t_1 \leq Y_1 \leq \alpha Y_2 + \beta Z + t_2 \wedge Z \in \mathcal{Z}] \end{aligned}$$

the inequality

$$\tilde{w}(t_1, t_2, \theta, \mathcal{Z}) \leq p(t_1, t_2, \theta)\mathbb{P}[Z \in \mathcal{Z}] \tag{32}$$

holds for that θ and and the interval $[t_1, t_2]$ and for all sets $\mathcal{Z} \subseteq \mathcal{R}_Z$.

Introducing the notation employed in BBC18, let the functions of moments required to be nonpositive at a parameter value θ in an identified set be denoted $m_j(\theta)$, $j \in \mathcal{J}$.

The test set employed in constructing $m_j(\theta)$ is the interval $[t_1^{k(j)}, t_2^{k(j)}]$ where $\{k(j) : j \in \mathcal{J}\}$ is a list of indexes identifying the endpoints of intervals and a set of values of Z, $\mathcal{Z}^{l(j)}$, is employed where $\{l(j) : j \in \mathcal{J}\}$ is a list of indexes identifying sets $\mathcal{Z}^{l(j)} \subseteq \mathcal{R}_Z$.[30]

The inequality (32) for each such interval $[t_1, t_2]$ and $\mathcal{Z}$ can therefore be represented by moment inequality $m_j(\theta) \leq 0$ using moment function $m_j(\theta)$ of the following form.

$$\begin{aligned}
m_j(\theta) &\equiv E\left[h\left(Y, Z, t_1^{k(j)}, t_2^{k(j)}, \mathcal{Z}^{l(j)}, \theta\right)\right], \\
h(Y, Z, t_1, t_2, \mathcal{Z}, \theta) &\equiv (h_1(Y, Z, t_1, t_2, \mathcal{Z}, \theta) + h_2(Y, Z, t_1, t_2, \mathcal{Z}, \theta)) \times 1[Z \in \mathcal{Z}], \\
h_1(Y, Z, t_1, t_2, \theta) &\equiv 1[t_1 = -\infty] \times 1[Y_1 = 0 \wedge 0 \leq \alpha Y_2 + \beta Z + t_2], \\
h_2(Y, Z, t_1, t_2, \theta) &\equiv 1[Y_1 > 0 \wedge \alpha Y_2 + \beta Z + t_1 \leq Y_1 \leq \alpha Y_2 + \beta Z + t_2] - p(t_1, t_2, \theta).
\end{aligned}$$

With data $\{(y_i, z_i) : i \in \{1, \ldots, N\}\}$ define the estimator $\hat{m}_j$ by the sample moment

$$\hat{m}_j(\theta) \equiv N^{-1}\sum_{i=1}^{N} m_{ji}(\theta), \qquad m_{ji}(\theta) \equiv h\left(y_i, z_i, t_1^{k(j)}, t_2^{k(j)}, \mathcal{Z}^{l(j)}, \theta\right).$$

A consistent estimator of the asymptotic variance of $N^{1/2}\hat{m}_j(\theta)$ is then

$$\hat{\sigma}_j^2(\theta) \equiv N^{-1}\sum_{i=1}^{N}(m_{ji}(\theta) - \hat{m}_j(\theta))^2.$$

Using the self-normalization-based critical value given in BBC18 and CCK19 the $100(1-\gamma)\%$ confidence region for the projection of the identified set onto the space of an element θ_k of θ is

$$CI(\theta_k, \gamma) \equiv \{r : T_{N,k}(r) \leq c_N(J, \gamma)\} \tag{33}$$

where

$$T_{N,k}(r) \equiv \inf_{\{\theta:\theta_k = r\}} \max_{j\in\{1,\ldots J\}} \left(N^{1/2}\frac{\hat{m}_j(\theta)}{\hat{\sigma}_j(\theta)}\right), \tag{34}$$

[30] When selected quantile independence is imposed the unique unknown elements in $\{t^{k(j)}\}_{j=1}^{J}$ are parameters, elements of θ.

and where $c_N(J,\gamma)$ is the critical value:

$$c_N(J,\gamma) \equiv \frac{\Phi^{-1}(1-\gamma/J)}{\sqrt{1-\Phi^{-1}(1-\gamma/J)^2/N}}.$$

5.2 Application: tobacco expenditure share

In this section, inspired by Adams, Blundell, Browning, and Crawford (2019) (ABBC), we present confidence regions and estimates of parameters of models for the share of household nondurable expenditure spent on tobacco.

ABBC estimate a Tobit model for tobacco expenditures expressed as a share of total nondurable expenditure with, as explanatory variables, log total expenditure on nondurables (potentially endogenous) and an OECD equivalence scale,[31] with log household disposable income as an excluded instrumental variable.

The data used in ABBC come from the UK Family Expenditure Survey (FES) 1980-2000 and are a sample of households with head of household aged 25-35 years in 1980. We do not aim to reproduce their analysis, and while we use the same explanatory variables and data source, we aim for reasonably large samples so we focus on households in the FES and its successor surveys from 2000-2009 with head of household aged 25-60 at the time of observation.[32,33] We conduct separate analysis of the periods 2000-04 and 2005-09 in which respectively 68% and 74% of households record zero tobacco expenditures in a two week diary.

ABBC take a quantile control function approach specifying a model in which nondurable expenditure is exogenous when there is conditioning on a control function which depends on nondurable expenditures, log household disposable income and the equivalence scale.[34] This control function restriction can arise in a complete triangular model in which there is a structural equation for log nondurable expenditure with, as explanatory variables, log disposable household income and the equivalence scale, and an unobserved variable which jointly with the unobserved variable in the structural equation for the tobacco expenditure

[31] The OECD equivalence scale is 1+0.7×the number of adults in excess of one + 0.5*×the number of children.

[32] In addition to the Family Expenditure Survey at the start of the sample period, the successor surveys are, from 2001, the Expenditure and Food Survey and from 2008, the Living Costs and Food Survey, Office for National Statistics (2002) and Office for National Statistics, Department for Environment, Food, and Rural Affairs (2010). These data are accessible by registration with the UK Data Service.

[33] We exclude households comprising more than one tax unit, and households with disposable weekly income recorded as £20 or less. All households have one or two adult members.

[34] Expenditures and income are recorded in UK pounds per week.

share is independent of the exogenous variables.

We employ the less restrictive, incomplete, single equation, IV model studied in this paper. The model has the same structural equation as in ABBC,

$$Y_1 = \max(0, \beta_0 + \alpha Y_2 + \beta_1 Z_1 + \sigma U_1)$$

where Y_1 denotes tobacco expenditure share, Y_2 denotes log nondurable expenditure and there are exogenous variables Z_1 (the OECD equivalence scale), and Z_2 (log household disposable income), the latter excluded from the structural equation for the tobacco expenditure share.[35] The model we use places no restriction on the process delivering Y_2.

We calculate confidence regions for the value of α using the method set out in Section 5.1.[36] We consider models imposing conditional quantile independence restrictions at three sets of quantile probabilities in turn, as follows,

QI3:	$(0.25, 0.5, 0.75)$
QI5:	$(0.167, 0.333, 0.5, 0.666, 0.833)$
QI7:	$(0.125, 0.25, 0.375, 0.5, 0.635, 0.75, 0.875)$

with the median normalized equal to zero in each case, and the values of remaining quantiles treated as parameters with unknown values. We also consider models (GQI3, GQI5, GQI7) in which the values of the quantiles are standard Gaussian quantiles scaled by the unknown value of the parameter σ.

We compare with estimates of α obtained using a classical Tobit model making no allowance for endogeneity and with estimates of α obtained using a complete, point identifying, triangular model in which there is the additional structural equation

$$Y_2 = \gamma_0 + \gamma_1 Z_1 + \gamma_2 Z_2 + U_2$$

and the restriction that (U_1, U_2) have a Gaussian distribution independent of $Z \equiv (Z_1, Z_2)$.

In the calculations using the IV model, sets of values of the exogenous variables are obtained as follows. For Z_1, the OECD equivalence scale, we define Z_1 to be the list of sets $\{1.0\}$,

[35] The partial correlation between log nondurable expenditure and log disposable household income given the OECD equivalence scale, $r_{Y_2 Z_2 . Z_1}$, is 0.64 in 2000-04 and 0.62 in 2005-09.

[36] Notice that in the equation for Y_1 the parameter σ multiplies U_1 which is restricted to be $N(0, 1)$ when the GaussSI restriction is imposed. By this device we are able to define intervals $[t_1, t_2]$ that span the effective range of U_1 at all values of the parameters.

$[1.5, 2.7]$, $[3.0, 5.7]$, $[1.0, 2.7]$, $[1.5, 5.7]$, $[3.0, 5.7]$.[37] For Z_2, log disposable household income, we take Z_2 to be the collection of all intervals with endpoints in $\{Q(0), Q(\frac{1}{8}), ..., Q(\frac{7}{8}), Q(1)\}$, where $Q(p)$ is the sample p-quantile of log income in the data. There are 36 such intervals. The sets $\mathcal{Z}$ employed in the calculations are the sets in the collection

$$\{\{Z : Z_1 \in \mathcal{Z}_1 \wedge Z_2 \in \mathcal{Z}_2\} : \mathcal{Z}_1 \in \mathsf{Z}_1, \mathcal{Z}_2 \in \mathsf{Z}_2\}.$$

In the calculations reported here there are $6 \times 36 = 216$ such sets.

Table 3 shows maximum likelihood estimates (MLE) of α, the coefficient on log non-durable expenditure, obtained using a classic Tobit model, in which Y_2 is restricted to be exogenous, and using a complete triangular model.[38] The Tobit model estimate is less than half the value obtained using the triangular model which allows Y_2 to be endogenous. The 95% confidence intervals (CI) are short, of length 0.017 in 2000-04 and 0.018 in 2005-09, giving a strong impression of accurate estimation

Table 4 shows 95% confidence regions for α obtained using the single equation IV model, which may be partially identifying, with quantile independence imposed at 3, 5 and 7 quantile probabilities. Figure 4 provides a convenient summary of the results.

Consider first the results for 2000-04 shown in the upper part of Figure 4. With 7 quantiles restricted independent of Z the length of the confidence region for α is 0.13, over 6 times longer than obtained using the complete triangular model.[39] The IV model is quite informative about the value of α but dropping the restrictions that complete the model substantially reduces the calculated precision. Restricting just 5 (3) quantiles independent of Z results in further reductions in precision delivering confidence regions with length 0.17 (0.57). Each confidence region is a subset of the regions delivered by less restrictive models. Also reported are 50% confidence regions, promoted as half-median-unbiased set estimates in Chernozhukov, Lee, and Rosen (2013) and Andrews and Shi (2013). A half-median-unbiased set estimate employs a conservative median-bias correction for inward bias which arises when using estimated intersection bounds. The correction is such that the upper (lower) endpoint

[37] Discrete Z_1 has no support on $(1.0, 1.5)$ and $(2.7, 3.0)$.

[38] Estimates calculated using STATA's `ivtobit` command.

[39] The QI3, QI5 and QI7 restrictions lead to identified sets defined by respectively 1944, 4320 and 7560 moment inequalities. In the estimations reported here *all connected unions* of the interquantile intervals were employed as test sets instead of just simple interquantile intervals. Additional test sets were included because it is possible that the functions of moments appearing in the inequalities delivered by some of the unions of core determining sets are more accurately estimated than the moment functions arising if only core determining test sets are employed.

of the interval is lower than (greater than) the upper (lower) bound of the interval for α delivered by a model with probability no greater than 1/2 asymptotically.

The lower part of Figure 4 shows results obtained using the 2005-09 data. The 2005-09 data deliver confidence regions which are unbounded below except in the case of the 50% region under the QI7 restriction. The data are less informative than the 2000-04 data with more zero expenditures (74%). However all the confidence regions suggest α is negative. For these data it is remarkable how much the restrictions of the complete triangular model deliver. Absent those restrictions the 2005-09 data has rather little to say other than to strongly suggest that the value of α is negative.

The analysis using quantile independence restrictions was conducted again requiring the values of the restricted quantiles to be, up to a scale factor σ, the values of standard Gaussian quantiles.[40] The results are reported in Table 5. In each case the confidence regions obtained for α were nearly identical to those obtained without the Gaussian restriction, shown in Table 4. So far as determining the value of α from these data the Gaussian restriction on the distribution of U delivers very little additional to the quantile independence restriction. By contrast the restrictions imposed by the complete triangular model are enormously influential.

6 Extensions

The models studied so far have left censoring at a known value. It is straightforward to extend the analysis to models in which there is more complex censoring. Consider for example models with stochastic censoring in which

$$Y_1 = \max(U_2, \alpha Y_2 + \beta Z + U_1)$$

and there is an observed censoring indicator, Y_3:

$$Y_3 = \begin{cases} 1 & , \quad U_2 > \alpha Y_2 + \beta Z + U_1, \\ 0 & , \quad U_2 < \alpha Y_2 + \beta Z + U_1, \end{cases}$$

[40] The corresponding inequalities are given by (12) implied under the GaussSI restriction with $t_1 = \sigma\Phi^{-1}(\lambda_j)$ and $t_2 = \sigma\Phi^{-1}(\lambda_{j'})$ for pairs $j' > j$.

with Y_3 taking either value 0 or 1 if $U_2 = \alpha Y_2 + \beta Z + U_1$.[41] Define $\theta \equiv (\alpha, \beta)$, $U \equiv (U_1, U_2)$, $Y \equiv (Y_1, Y_2, Y_3)$ and u and y similarly.

This model has residual sets as follows.

$$\begin{aligned}
\mathcal{U}((y_1, y_2, 1), z, \theta) &= \{u : u_1 \leq y_1 - \alpha y_2 - \beta z \wedge u_2 = y_1\}, \\
\mathcal{U}((y_1, y_2, 0), z, \theta) &= \{u : u_1 = y_1 - \alpha y_2 - \beta z \wedge u_2 \leq y_1\}.
\end{aligned}$$

These are shown in Figure 2 where the sets $\mathcal{U}((y_1, y_2, 1), z, \theta)$ and $\mathcal{U}((y_1, y_2, 0), z, \theta)$ comprise respectively the horizontal blue line and the vertical red line.[42]

The inequalities characterizing the identified set of structures delivered by this model arise on applying the inequality (6) to sets $\mathcal{S}$ which are connected unions of these residual sets. These take one of two forms, each determined by values of three real constants which are denoted $\left(u_1^L, u_1^H, a\right)$ when considering sets of the first type, $\mathcal{S}_1\left(u_1^L, u_1^H, a\right)$, and $\left(u_2^L, u_2^H, a\right)$ when considering sets of the second type, $\mathcal{S}_0(u_2^L, u_2^H, a)$. These are such that $u_j^L \leq u_j^H$ for each $j \in \{1, 2\}$. The sets are defined by:

$$\begin{aligned}
\mathcal{S}_0(u_1^L, u_1^H, a) &\equiv \{u : u_1 \in [u_1^L, u_1^H] \wedge u_2 - u_1 \leq a\}, \\
\mathcal{S}_1(u_2^L, u_2^H, a) &\equiv \{u : u_2 \in [u_2^L, u_2^H] \wedge u_2 - u_1 \geq a\}.
\end{aligned}$$

Figure 3 shows an example in which $\mathcal{S}_0(u_1^L, u_1^H, a)$ is oriented vertically and shaded pink and $\mathcal{S}_1(u_2^L, u_2^H, a)$ is oriented horizontally and shaded cyan.

Following the definition set out in (7), the values of Y that occur only when U takes a value in $\mathcal{S}_0(u_1^L, u_1^H, a)$ and $\mathcal{S}_1(u_2^L, u_2^H, a)$, respectively, are:

$$\begin{aligned}
\mathcal{A}\left(\mathcal{S}_0(u_1^L, u_1^H, a), z, \theta\right) &= \{y : y_1 - \alpha y_2 - \beta z \in [u_1^L, u_1^H] \wedge \alpha y_2 + \beta z \leq a \wedge y_3 = 0\}, \\
\mathcal{A}\left(\mathcal{S}_1(u_2^L, u_2^H, a), z, \theta\right) &= \{y : y_1 \in [u_2^L, u_2^H] \wedge \alpha y_2 + \beta z \geq a \wedge y_3 = 1\}.
\end{aligned}$$

Applying the inequality (6), the sharp identified set of structures comprises those struc-

[41] In models in which variables exhibit continuous variation the probability that $U_2 = \alpha Y_2 + \beta Z + U_1$ is zero and determination Y_3 in such cases is without consequence.

[42] In a model in which Y_3 is not observed, $y \equiv (y_1, y_2)$, and the residual sets are $\mathcal{U}((y_1, y_2, 1), z, \theta) \cup \mathcal{U}((y_1, y_2, 0), z, \theta)$. In this case the sets to be considered in obtaining a characterization of the sharp identified set of structures are $\mathcal{S}_1(u^L, u^H, a) \cup \mathcal{S}_0(u^L, u^H, a)$ obtained as constants $u^L \leq u^H$ and a vary over the real line, and all connected unions of such sets.

tures such that, for all $u_1^L \leq u_1^H$, $u_2^L \leq u_2^H$, $a \in \mathbb{R}$, and almost every $z \in \mathcal{R}_Z$:

$$\mathbb{P}\left[u_1^L + \beta Z \leq Y_1 - \alpha Y_2 \leq u_1^H + \beta Z \wedge \alpha Y_2 \leq a - \beta Z \wedge Y_3 = 0 | Z = z\right] \leq G_{U|Z=z}(\mathcal{S}_0(u_1^L, u_1^H, a)),$$

and

$$\mathbb{P}\left[u_2^L \leq Y_1 \leq u_2^H \wedge \alpha Y_2 \geq a - \beta Z \wedge Y_3 = 1 | Z = z\right] \leq G_{U|Z=z}(\mathcal{S}_1(u_2^L, u_2^H, a)).$$

In practice there will be restrictions on the dependence of U and Z, for example stochastic independence. Then, with a parametric specification of the distribution of U the probabilities on the right hand sides of these inequalities can be calculated. Inference can proceed as set out in Section 5.1. Under a nonparametric specification the support of U can be partitioned into cells, the right hand side probabilities can be expressed as sums of unknown cell probabilities and the method of Section 4.3 can be applied.

7 Concluding remarks

When working with censored data and endogenous explanatory variables the easy way to obtain estimates of structural parameters is to employ a complete triangular model like the Gaussian model underlying STATA's `ivtobit` command or to assume directly that a valid identifiable control function exists. When there is no economic rationale for such restrictions the IV model developed here provides a route to robust estimation. Even when more restrictive models are thought to be appropriate the IV model can deliver useful information regarding the force of additional restrictions. The IV model can signal misspecification of more restrictive models. It can deliver results when the complete model attack is not available, for example when endogenous variables are discrete or are determined by multiple sources of heterogeneity.

In the application to tobacco expenditures the IV results reveal the enormous power of the restrictions of the Gaussian triangular model in delivering not only point identification but apparently highly accurate estimation. In practice there will be many plausible ways to obtain a complete point identifying model. The unobserved variables could be non-Gaussian, there could be multiple sources of heterogeneity determining the values of the endogenous explanatory variable, in some applications there could be simultaneous not recursive determination of the endogenous variables. The identified sets delivered by the IV model contain all the values of parameters under all possible completions of the single equation IV model that can deliver the distribution of observed variables used to calculate the set. Accordingly the IV approach is a useful aid in sensitivity analysis.

The IV model can be partially or point identifying and it may not be possible to determine identification status in particular applications. So it is important to use methods for estimation and inference that deliver results regardless of whether there is point or partial identification, as has been done in the application presented here.

There is rarely a good economic argument for particular parametric restrictions on the distribution of the unobserved variable U in the structural equation for a censored outcome. Even so such restrictions are frequently imposed. We have shown how estimation and inference can be done using an IV Tobit model, dispensing with the commonly used Gaussian restriction, using multiple quantile independence restrictions, requiring the p-quantiles of U given instrumental variables, Z, to be independent of Z at selected quantile probabilities. The values of quantiles at the selected probabilities become parameters with unknown values and we calculate confidence regions on projections of the identified set of values of the augmented parameter vector onto the spaces of particular parameters of interest.

We have shown how to calculate outer regions for identified sets of structural parameter values and their projections onto the space of individual parameters under a restriction requiring U and Z to be stochastically independent with no further restriction on the distribution of U. A parameter value lies in the identified set if and only if the solution to a linear program is nonnegative. Although the program can involve a very large number of inequalities the solution is quick to calculate. Conducting inference using this method is a research challenge not addressed here, and there remain other challenges. For example, we have proposed and applied a procedure for conducting inference on a partially identified parameter capturing the marginal effect of an endogenous variable on an outcome of interest when instrumental variables are continuously distributed, in which one calculates joint probabilities of events defined by sets of values of endogenous variables and sets of values of instrumental variables. Finite sample performance will of course depend on the sets that are chosen and future research to help guide these choices is warranted. We have sketched an extension of our IV methods to a case with stochastic censoring. Further extension to competing risks models with potentially endogenous explanatory variables is a fruitful area for research.

References

Adams, A., R. Blundell, M. Browning, and I. Crawford (2019): "Prices versus Preferences: Taste Change and Revealed Preference," IFS Working Paper.

Amemiya, T. (1979): "The Estimation of a Simultaneous Equation Tobit Model," *International Economic Review*, 20(1), 169–181.

Andrews, D. W. K., and X. Shi (2013): "Inference Based on Conditional Moment Inequalities," *Econometrica*, 81(2), 609–666.

Belloni, A., F. Bugni, and V. Chernozhukov (2018): "Subvector Inference in PI Models with with Many Moment Inequalities," arXiv:1806.11466.

Berry, S., and G. Compiani (forthcoming): "An Instrumental Variable Approach to Dynamic Models," *Review of Economic Studies*.

Blundell, R., and J. L. Powell (2007): "Censored Regression Quantiles with Endogenous Regressors," *Journal of Econometrics*, 141(1), 65–83.

Blundell, R., and R. J. Smith (1989): "Estimation in a Class of Simultaneous Equation Limited Dependent Variable Models," *Review of Economic Studies*, 56(1), 37–58.

Borchers, H. W. (2022): *pracma: Practical Numerical Math Functions.* R package version 2.3.8.

Border, K. C. (2020): "Alternative Linear Inequalities," CalTech, version 2020.10.15::09.50.

Chen, S., and Q. Wang (2020): "Semiparametric Estimation of a Censored Regression Model with Endogeneity," *Journal of Econometrics*, 215(1), 239–256.

Chernozhukov, V., D. Chetverikov, and K. Kato (2019): "Inference on Causal and Structural Parameters Using Many Moment Inequalities," *Review of Economic Studies*, 86(5), 1867–1900.

Chernozhukov, V., I. Fernandez-Val, and A. Kowalski (2015): "Quantile Regression with Censoring and Endogeneity," *Journal of Econometrics*, 186, 201–221.

Chernozhukov, V., S. Lee, and A. M. Rosen (2013): "Intersection Bounds: Estimation and Inference," *Econometrica*, 81(2), 667–737.

Chesher, A., and A. Rosen (2017): "Generalized Instrumental Variable Models," *Econometrica*, 85(3), 959–989.

Chesher, A., and A. M. Rosen (2020a): "Econometric Modeling of Interdependent Discrete Choice," CeMMAP working paper CWP25/20.

——— (2020b): "Generalized Instrumental Variable Models, Methods, and Applications," in *The Handbook of Econometrics*, ed. by S. Durlauf, L. P. Hansen, J. J. Heckman, and R. Matzkin, vol. 7a. Elsevier.

Chesher, A., A. M. Rosen, and Z. a. Siddique (2019): "Estimating Endogenous Effects on Ordered Outcomes," CeMMAP working paper 66/19.

Das, M. (2002): "Estimation and Inference in a Censored Regression Model with Endogenous Covariates," Columbia Department of Economics Discussion Paper 0102-61.

Frandsen, B. R. (2015): "Treatment Effects with Censoring and Endogeneity," *Journal of the American Statistical Association*, 110(512), 1745–1752.

Genz, A., and F. Bretz (2009): *Computation of Multivariate Normal and t Probabilities*, Lecture Notes in Statistics. Springer-Verlag, Heidelberg.

Genz, A., F. Bretz, T. Miwa, X. Mi, F. Leisch, F. Scheipl, and T. Hothorn (2021): *mvtnorm: Multivariate Normal and t Distributions.* R package version 1.1-3.

Heckman, J. J. (1978): "Dummy Endogenous Variables in a Simultaneous Equation System," *Econometrica*, 46, 931–959.

Hong, H., and E. Tamer (2003): "Inference in Censored Models with Endogenous Regressors," *Econometrica*, 71(3), 905–932.

Honore, B., and L. Hu (2004): "Estimation of Cross Sectional and Panel Data Censored Regression Models with Endogeneity," *Journal of Econometrics*, 122(2), 293–316.

Khan, S., M. Ponomareva, and E. Tamer (2011): "Sharpness in Randomly Censored Linear Models," *Economics Letters*, 113(1), 23–25.

Khan, S., and E. Tamer (2009): “Inference on Endogenously Censored Regression Models Using Conditional Moment Inequalities,” *Journal of Econometrics*, 152(2), 104–119.

Konis, K., and F. Schwendinger (2020): *lpSolveAPI: R Interface to lp_ solve.* R package version 5.5.2.0-17.7.

Lancaster, T. (1985): “Simultaneous Equations Models in Applied Search Theory,” *Journal of Econometrics*, 28(1), 113–126.

Lancaster, T., and A. Chesher (1984): “Simultaneous Equations with Endogenous Hazards,” in *Studies in Labor Market Dynamics, Studies in Contemporary Economics, vol 11.*, ed. by G. Neumann, and N. Westergå rd Nielsen. Springer, Berlin, Heidelberg.

Manski, C. F., and E. Tamer (2002): “Inference on Regressions with Interval Data on a Regressor or Outcome,” *Econometrica*, 70(2), 519–546.

Molchanov, I., and F. Molinari (2018): *Random Sets in Econometrics.* Cambridge University Press, Cambridge, U.K.

Nelson, F., and L. Olson (1978): “Specification and Estimation of a Simultaneous Equation Model with Limited Dependent Variables,” *International Economic Review*, 19(3), 695–709.

Newey, W. K. (1987): “Efficient Estimation of Limited Dependent Variable Models with Endogenous Explanatory Variables,” *Journal of Econometrics*, 36(3), 231–250.

Office for National Statistics (2002): “Family Expenditure Survey, 2000–2001,” [data collection]. UK Data Service. Retrieved from `https://discover.ukdataservice.ac.uk/series/?sn=200016`.

Office for National Statistics, Department for Environment, Food, and Rural Affairs (2010): “Living Costs and Food Survey, 2001–2009,” [data collection]. *2nd Edition.* UK Data Service. Retrieved from `https://discover.ukdataservice.ac.uk/series/?sn=2000028`.

Olsen, R. J., and G. Farkas (1989): “Endogenous Covariates in Duration Models and the Effect of Adolescent Childbirth on Schooling,” *Journal of Human Resources*, 24(1), 39–53.

R Core Team (2022): *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria.

Smith, R. J., and R. Blundell (1986): "An Exogeneity Test for a Simultaneous Equation Tobit Model with an Application to Labor Supply," *Econometrica*, 54(3), 679–685.

StataCorp (2019): "Stata Statistical Software: Release 16," College Station, TX: StataCorp LLC.

Tobin, J. (1958): "Estimation of Relationships for Limited Dependent Variables," *Econometrica*, 26(1), 24–36.

Wang, Q., and S. Chen (2021): "Moment Estimation for Censored Quantile Regression," *Econometrics Reviews*, 40(9), 815–829.

Wrenn, D. H., H. A. Klaiber, and D. A. Newburn (2017): "Confronting Price Endogeneity in a Duration Model of Residential Subdivision Development," *Journal of Applied Econometrics*, 32(3), 661–682.

Appendix: Proofs and Figures

Proof of Proposition 1. Applying Theorem 1, Corollary 1 of CR gives the following characterization of the identified set of structures delivered by a model, A, a collection of conditional distributions of Y given Z and the support of Z:

$$\begin{aligned}\mathcal{I}(\mathcal{F}_{Y|Z},\mathcal{R}_Z,A)=\big\{&(m,\mathcal{G}_{U|Z})\in\mathcal{M}_A:\\ &\forall\mathcal{S}\in\mathsf{F}\left(\mathcal{R}_U\right),G_{U|Z}(\mathcal{S}|z)\geq\mathbb{P}[\mathcal{U}(Y,Z,m)\subseteq\mathcal{S}|Z=z]\text{ a.e. }z\in\mathcal{R}_Z\big\}\,,\end{aligned}\tag{35}$$

where $\mathsf{F}\left(\mathcal{R}_U\right)$ denotes the collection of closed subsets of $\mathcal{R}_U$. By Lemma 1 of CR it follows that the requirement that the inequality holds for all closed sets $\mathcal{S}$ can be replaced by the requirement that it holds for all $\mathcal{S}$ that are unions of sets on the support of $\mathcal{U}(Y,Z,m)$ conditional on $Z=z$. Each such set can be be written as a union of sets of the form $(-\infty,t]$ and $[t_1,t_2]$, where if $t_1=t_2=t$, the set $[t_1,t_2]$ is simply the point $\{t\}$. All such unions are themselves either of the form $\mathcal{S}=(-\infty,t]$ or $\mathcal{S}=[t_1,t_2]$. The collections $\mathcal{I}_1(\mathcal{F}_{Y|Z},\mathcal{R}_Z,A)$ and $\mathcal{I}_2(\mathcal{F}_{Y|Z},\mathcal{R}_Z,A)$ comprise those structures satisfying $G_{U|Z}(\mathcal{S}|z)\geq\mathbb{P}\left[\mathcal{U}(Y,Z,m)\subseteq\mathcal{S}|Z=z\right]$

for each of these two types of sets $\mathcal{S}$, respectively, completing the proof. ■

Proof of Proposition 2. Existence of $b(z,t,m) \equiv \nabla_t B(z,t,m)$ follows from differentiability of $m(y_2,z,t)$ with respect to t and the existence of the density $\widetilde{g}_{U|Z}(\cdot|z)$. The inequality

$$\widetilde{G}_{U|Z}(t_2|z) - \widetilde{G}_{U|Z}(t_1|z) \geq B(z,t_2,m) - B(z,t_1,m).$$

can then be expressed as

$$\int_{t_1}^{t_2} (\widetilde{g}_{U|Z}(t|z) - b(z,t,m))dt \geq 0,$$

which for any z holds for all $[t_1,t_2] \subseteq \mathbb{R}$ if and only if for all t, $\widetilde{g}_{U|Z}(t|z) \geq b(z,t,m)$. ■

Proof of Proposition 3. Theorem 5 of CR implies that identified set for structural function m comprises those functions m such that zero is an element of the Aumann expectation of $\mathcal{U}(Y,Z,m)$ conditional on $Z=z$ a.e. $z \in \mathcal{R}_Z$. Recall that the residual set in the model under study is

$$\mathcal{U}(y,z,m) = \begin{cases} (-\infty, m^{-1}(y_2,z,0)] & , \quad y_1 = 0 \\ \{m^{-1}(y_2,z,y_1)\} & , \quad y_1 > 0 \end{cases}$$

and let $\mathbb{E}[\mathcal{A}|z]$ denote the Aumann expectation of random set $\mathcal{A}$ conditional on $Z=z$.[43] There is

$$\mathbb{E}[\mathcal{U}(Y,Z,m)|z] = \mathbb{E}[\mathcal{U}(Y,Z,m)|z,Y_1=0]P[Y_1=0|z] + \mathbb{E}[\mathcal{U}(Y,Z,m)|z,Y_1>0]P[Y_1>0|z]$$

where the sum is a Minkowski sum.[44] Considering each term in turn there is[45]

$$\mathbb{E}[\mathcal{U}(Y,Z,m)|z,Y_1=0] = (-\infty, E[m^{-1}(Y_2,Z,0)|z,Y_1=0]]$$

[43] The Aumann expectation of a random set $\mathcal{A}$ is the set of expected values of all random variables A with finite expected values having the property that $A \in \mathcal{A}$ with probability one.

[44] The Minkowski sum of sets A and B is the set of values obtained by adding each element of A to each element of B.

$$A + B = \{a+b : a \in A, b \in B\}$$

[45] See Example 3.14 in Molchanov and Molinari (2018).

which is a semi-infinite interval and there is[46]

$$\mathbb{E}[\mathcal{U}(Y,Z,m)|z,Y_1>0]=\{E[m^{-1}(Y_2,Z,Y_1)|z,Y_1>0]\}$$

which is a singleton. The Minkowski sum of a semi-infinite interval and a singleton set is a semi-infinite interval. The result is that the conditional (on Z) Aumann expectation of the residual set is the semi-infinite interval

$$\mathbb{E}[\mathcal{U}(Y,Z,m)|z]=(-\infty,E[m^{-1}(Y_2,Z,Y_1)|z]],$$

which leads directly to the result of the Proposition. ∎

Proof of Proposition 4. From Proposition 1 the identified set for $\left(m,\mathcal{G}_{U|Z}\right)$ is

$$\mathcal{I}(\mathcal{F}_{Y|Z},\mathcal{R}_Z,A)=\mathcal{I}_1(\mathcal{F}_{Y|Z},\mathcal{R}_Z,A)\cap\mathcal{I}_2(\mathcal{F}_{Y|Z},\mathcal{R}_Z,A).$$

It will be shown that under Restriction QI the identified set for structural function m, which is the projection of $\mathcal{I}(\mathcal{F}_{Y|Z},\mathcal{R}_Z,A)$ onto $\mathcal{M}_A^0$, here denoted $\mathcal{M}^*$, is equivalent to the set of functions $m\in\mathcal{M}_A^0$ that satisfy conditions (1) and (2) in the statement of the Proposition.

Suppose first that $m\in\mathcal{M}^*$ such that for some $\mathcal{G}_{U|Z}$ satisfying Restriction QI $\left(m,\mathcal{G}_{U|Z}\right)\in\mathcal{I}(\mathcal{F}_{Y|Z},\mathcal{R}_Z,A)$. Conditions (1) and (2) then hold because these are implied by the inequalities that define the sets $\mathcal{I}_1(\mathcal{F}_{Y|Z},\mathcal{R}_Z,A)$ and $\mathcal{I}_2(\mathcal{F}_{Y|Z},\mathcal{R}_Z,A)$, respectively.

Now suppose that m satisfies conditions (1) and (2) for some $\{q_1,...,q_J\}\in\mathcal{Q}$. To show that $m\in\mathcal{M}^*$ it will be shown by construction that there exists a collection of conditional distributions $\mathcal{G}_{U|Z}(\cdot|z;m)$ with cumulative distribution functions $\tilde{G}(\cdot|z;m)=\tilde{G}_{U|Z}(\cdot|z;m)$ for each $z\in\mathcal{R}_Z$ satisfying Restriction QI such that $\left(m,\mathcal{G}_{U|Z}(\cdot|z;m)\right)\in\mathcal{I}(\mathcal{F}_{Y|Z},\mathcal{R}_Z,A)$. The inclusion of m in the notation $\mathcal{G}_{U|Z}(\cdot|z;m)$ signifies that the associated collection of conditional distributions $\mathcal{G}_{U|Z}(\cdot|z)=\mathcal{G}_{U|Z}(\cdot|z;m)$ in the construction will in general vary with m.

Specifically, it needs to be shown that for almost every $z\in\mathcal{R}_Z$ there exists a cumulative distribution function $\tilde{G}(\cdot|z;m)$ such that the following three conditions (36)–(38) hold. For all $j\in\{0,...,J\}$:

$$\tilde{G}(q_j|z;m)=\lambda_j. \tag{36}$$

[46] See Example 3.12 in Molchanov and Molinari (2018).

For all $t \in \mathbb{R}$:

$$\tilde{G}(t|z;m) \geq C(z,t,m). \tag{37}$$

For all $s \leq t$, each in $\mathbb{R}$:

$$\tilde{G}(t|z;m) - \tilde{G}(s|z;m) \geq \Delta(z,s,t,m). \tag{38}$$

Condition (36) ensures that Restriction QI holds and conditions (37) and (38) are the conditions defining the identified set for $(m, \mathcal{G}_{U|Z})$, as shown in Proposition 1. Note that for (38) it is equivalent to show that

$$\tilde{G}(t|z;m) - B(z,t,m) \tag{39}$$

is weakly increasing in t.

Construction of $\tilde{G}(t|z;m) : \mathbb{R} \to [0,1]$ for each z, m is as follows, divided into separate cases according to where argument t lies relative to $q_1, ..., q_J$.

1. **Case 1**: $t \in (-\infty, q_1]$. Define

$$\tilde{G}(t|z;m) \equiv C(z,t,m) + (\lambda_1 - C(z,q_1,m)) \exp(\eta(t - q_1)),$$

where $\eta > 0$ is arbitrary.[47] Since $\lim_{t\to-\infty} C(z,t,m) = 0$ and $\lim_{t\to-\infty} \exp(\eta(t-q_1)) = 0$ it follows that $\lim_{t\to-\infty} \tilde{G}(t|z;m) = 0$. There is also $\tilde{G}(q_1|z;m) = \lambda_1$. $\tilde{G}(t|z;m)$ is an increasing function of t because it is the sum of two increasing functions of t, $\tilde{G}(t|z;m) \geq C(z,t,m)$ by definition and

$$\tilde{G}(t|z;m) - B(z,t,m) = D(z,t,m) + (\lambda_2 - C(z,q_2,m)) \exp(\eta(t - q_2))$$

is an increasing function of t because it is the sum of two increasing functions of t.

2. **Case 2**: $t \in [q_j, q_{j+1}]$, each $j = 1, ..., J-1$. Define

$$L_j(z,t,m) \equiv B(z,t,m) + \lambda_j - B(z,q_j,m)$$

which is an increasing function of t with $L_j(z,t,m) - B(z,t,m)$ constant and $L_j(z,q_j,m) = \lambda_j$. Condition (2) ensures

$$\lambda_{j+1} - B(z,q_{j+1},m) \geq \lambda_j - B(z,q_j,m),$$

[47] Construction of $\tilde{G}(t|z;m)$ employing functions other than the exponential function could also be used.

from which it follows that $L_j(z, q_{j+1}, m) \le \lambda_{j+1}$. Define

$$M_j(z,t,m) \equiv C(z,t,m) + (\lambda_{j+1} - C(z,q_{j+1},m))\frac{(t-q_j)}{(q_{j+1}-q_j)}.$$

This is an increasing function of t with $M_j(z,t,m) \ge C(z,t,m)$, $M_j(z,q_j,m) = C(z,q_j,m)$, and $M_j(z,q_{j+1},m) = \lambda_{j+1}$. Define

$$\tilde{G}(t|z;m) = \max(L_j(z,t,m), M_j(z,t,m)).$$

There is

$$\tilde{G}(q_j|z;m) = \max(\lambda_j, C(z,q_j,m)) = \lambda_j,$$

and

$$\tilde{G}(q_{j+1}|z;m) = \max(L_j(z,q_{j+1},m), \lambda_{j+1}) = \lambda_{j+1},$$

because as just shown, $L_j(z,q_{j+1},m) \le \lambda_{j+1}$. This is an increasing function of t in the interval $[q_j, q_{j+1}]$ because it is the maximum of two increasing functions of t, and $\tilde{G}(t|z;m) \ge C(z,t,m)$ because $\tilde{G}(t|z;m)$ is the maximum of two functions one of which is at least equal to $C(z,t,m)$ in the interval under consideration. Moreover,

$$\begin{aligned}&\tilde{G}(t|z;m) - B(z,t,m) = \\ &\qquad \max\left(\lambda_j - B(z,q_j,m), D(z,t,m) + (\lambda_{j+1} - C(z,q_{j+1},m))\frac{(t-q_j)}{(q_{j+1}-q_j)}\right)\end{aligned}$$

which is increasing in t because it is the maximum of two increasing functions of t, verifying condition (39).

- **Case 3**: $t \in [q_J, \infty)$. Define

$$\tilde{G}(t|z;m) \equiv \max\left(C(z,t,m), B(z,t,m) - B(z,q_J,m) + \lambda_J\right).$$

There is

$$\tilde{G}(q_J|z;m) = \max(C(z,q_J,m), \lambda_J) = \lambda_J,$$

and then

$$\lim_{t\to\infty} \tilde{G}(t|z;m) = \max(1, B(z,\infty,m) - B(z,q_J,m) + \lambda_J) = 1$$

because from condition (2)

$$\lambda_J - B(z, q_J, m) \leq \lambda_{J+1} - B(z, q_{J+1}, m)$$

which implies

$$B(z, q_{J+1}, m) - B(z, q_J, m) + \lambda_J \leq \lambda_{J+1} = 1.$$

$\tilde{G}(t|z;m)$ is an increasing function of t because it is the maximum of two increasing functions of t. Finally

$$\tilde{G}(t|z;m) - B(z,t,m) = \max\left(D(z,t,m), -B(z,q_J,m) + \lambda_J\right)$$

which is an increasing function of t.

We have shown that for any $z \in \mathcal{R}_Z$ and any function m satisfying the conditions of the Proposition the piecewise function $\tilde{G}(t|z,m)$ defined above can be constructed and we have shown that it satisfies conditions (36), (37) and (38) Therefore any function m satisfying the conditions of the Proposition is contained in the identified set of structural functions delivered by the model. ■

Proof of Proposition 5. Suppose that m is in the projection of the identified set of structures $\left(m, \mathcal{G}_{U|Z}\right)$ delivered by the IV Tobit Model A under consideration. Then under Restriction NPSI there exists a distribution G such that $(m, \{G\}) \in \mathcal{I}(\mathcal{F}_{Y|Z}, \mathcal{R}_Z, A)$ and as explained in the text it follows from Proposition 1 that (22) and (23) hold with $p_1, ..., p_N$ as defined in (21) as a function of that distribution G. The existence of a vector of proper probabilities $\mathbf{p} = (p_1, ..., p_N)$ such that (22) and (23) hold almost surely is equivalent to the existence of $\mathbf{p} \in \mathbb{R}_N$ satisfying

$$\begin{aligned} A\mathbf{p} &= 1, \\ \mathbf{B}p &\leq c(m), \\ \mathbf{p} &\geq 0, \end{aligned}$$

a linear program in $\mathbf{p}$. Then, applying the same steps as in Chesher and Rosen (2020a), and in particular using the version of Farkas's Alternative provided in Border (2020) – see paragraph 12, Section 1.4 – such probabilities exist *if and only if* there is *no* solution for

(s,t), $s \in \mathbb{R}$, $t \in \mathbb{R}^K$, to the system

$$sA + t\mathbf{B} \geq 0,$$

$$t \geq 0,$$

$$s + t \cdot c(m) < 0,$$

which is equivalent there being a nonnegative value of the linear program (27) subject to (28), (29), and (30). ■

Figure 1: Identified sets of values of α and β in a linear index IV Tobit model delivered by Cases 1-3 of Structure 1 and Case 1 of Structure 2, see Tables 1 and 2. The area below all the blue lines is the identified set under a zero conditional mean restriction, $E[U|z] = 0$ for all z. Pink shaded regions are identified sets under a zero conditional median independence restriction. The value of a and b (1 and 0) generating the probabilities used to calculate each set is the point plotted in green. Note that the scale of the axes in the lower right pane differs from the scale of the axes on the other panes.

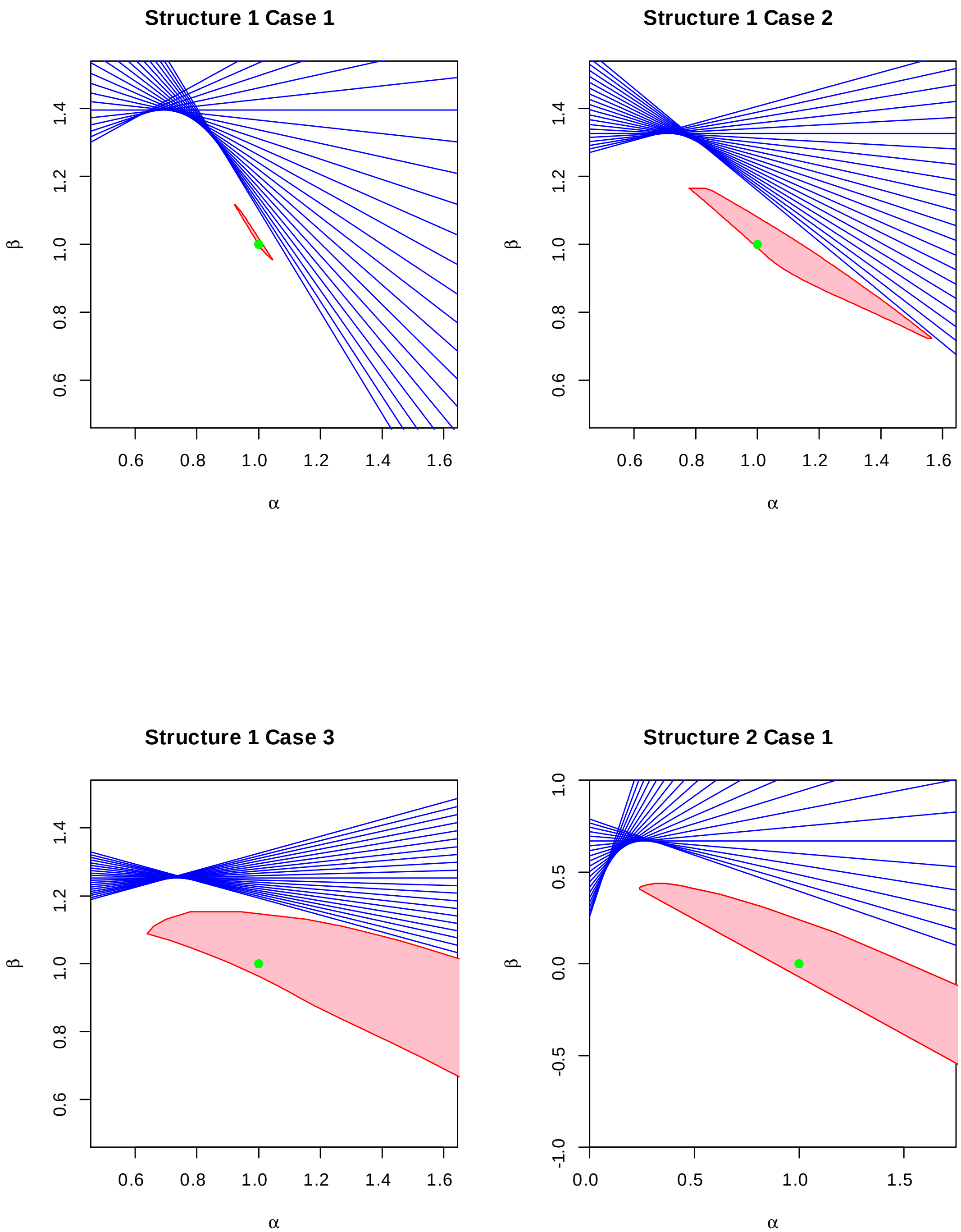

Figure 2: Sets of values of U delivering particular values y of Y for given values z of Z in a model with stochastic censoring. Above the dashed 45° line where $u_2 \geq \alpha y_2 + \beta z + u_1$ and $y_3 = 1$, lies the set $\mathcal{U}((y_1, y_2, 1), z, \theta)$ drawn blue. In this set $u_2 = y_1$ and $u_1 \leq y_1 - \alpha y_2 - \beta z$. Below the 45° line where $y_3 = 0$ lies the set $\mathcal{U}((y_1, y_2, 0), z, \theta)$ drawn red. In this set $u_1 = y_1 - \alpha y_2 - \beta z$ and $u_2 \leq y_1$.

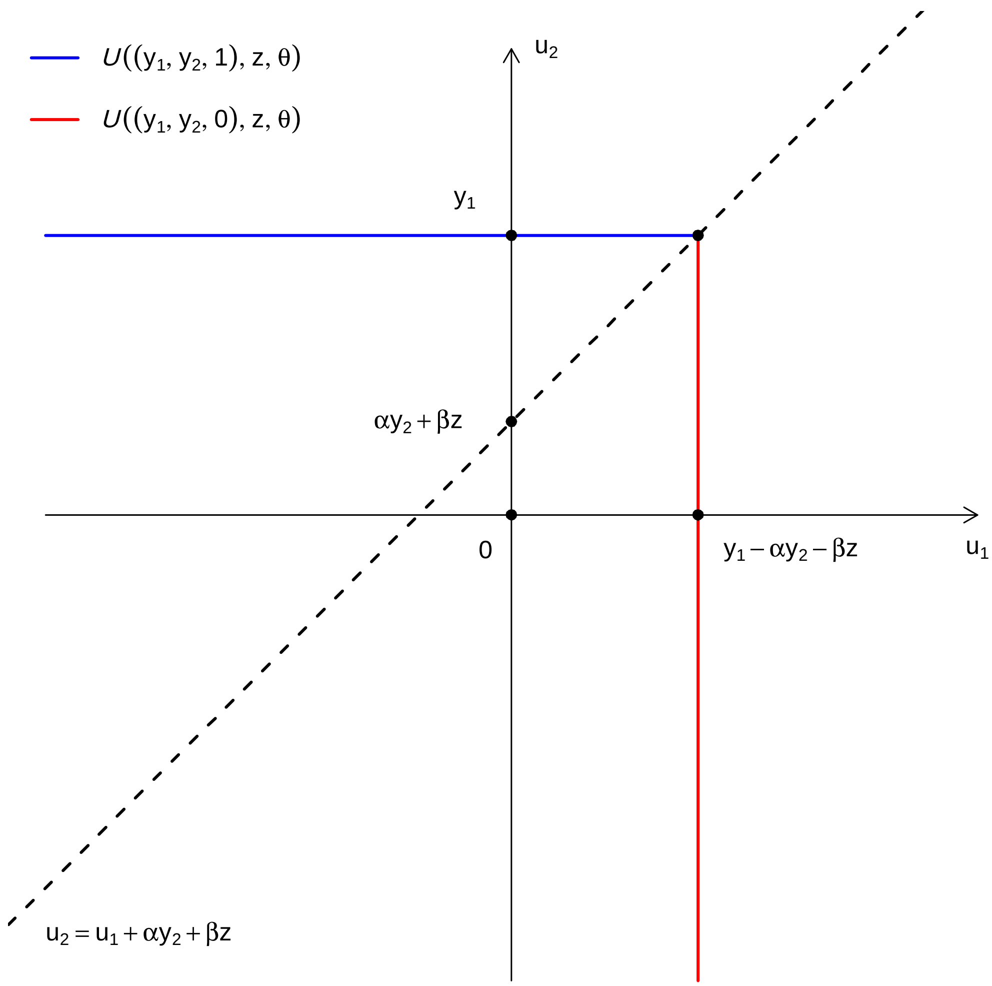

Figure 3: Sets of values of U determining moment inequalities in a model with stochastic censoring. The dashed line passing through the point $(0, a)$ is a 45° line. The sets are determined by the value of a and by intervals $[u_1^L, u_1^H]$ and $[u_2^L, u_2^H]$.

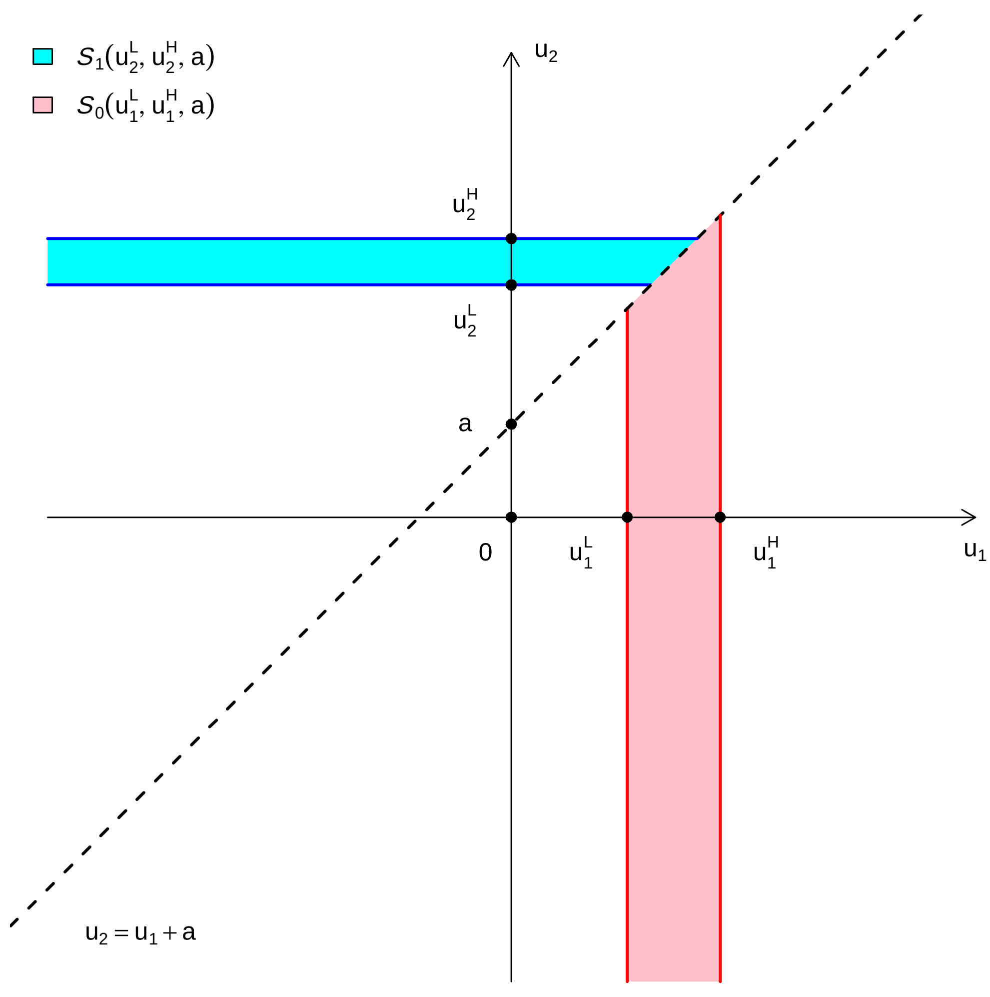

Figure 4: Confidence regions for, and estimates of α, the coefficient on endogenous log total expenditure on nondurables in the model for household tobacco expenditures. Results in the upper part of the graph are for 2000-2004, 18,473 households, 68% recording zero expenditure. Results in the lower part of the graph are for 2005-2009, 15,885 households, 74% recording zero expenditure.

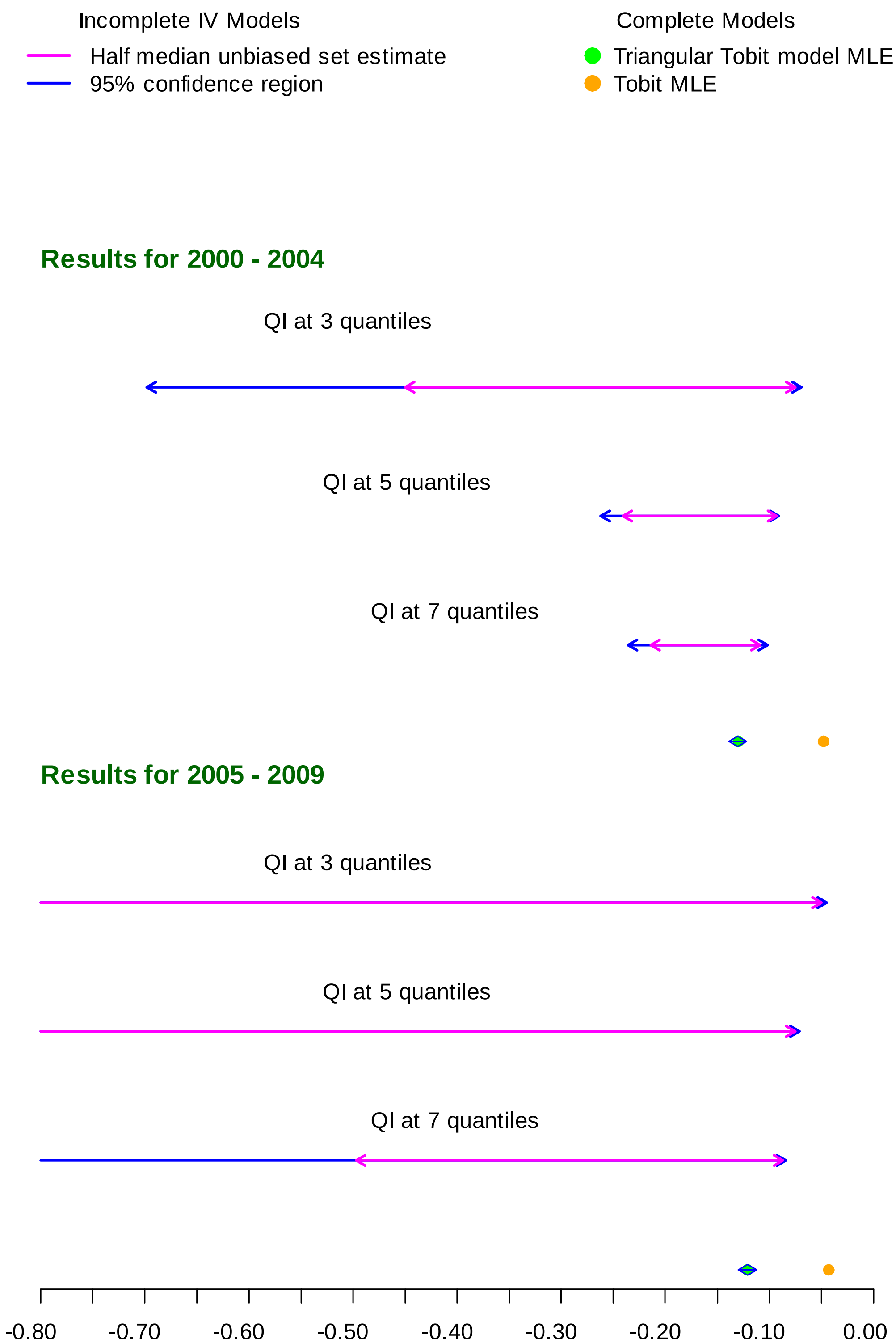

Table 1: Projections of the identified set onto the space of α under a median independence condition (MedI) and a stochastic independence restriction (NPSI) with a nonparametric specification of the distribution of U with the number of intervals $N = 100$ and under stochastic independence restriction with U restricted to be Gaussian (GaussSI). These results are for Structure 1. Censoring probabilities p_0^ℓ, $\overline{p_0}$, and p_0^u are defined as $\min_{z \in \mathcal{R}_Z} \mathbb{P}[Y_1 = 0 | Z = z]$, $\mathbb{P}[Y_1 = 0]$, and $\max_{z \in \mathcal{R}_Z} \mathbb{P}[Y_1 = 0 | Z = z]$, respectively.

Case	Structure 1 parameter values							Censoring Probabilities			Identified intervals for α					
											GaussSI		NPSI		MedI	
	b	a	d_0	d_1	s_{11}	s_{12}	s_{22}	p_0^ℓ	$\overline{p_0}$	p_0^u	$\underline{\alpha}$	$\overline{\alpha}$	$\underline{\alpha}$	$\overline{\alpha}$	$\underline{\alpha}$	$\overline{\alpha}$
1	1	1	0.5	1	2	0.5	1	0.11	0.24	0.40	1.00	1.00	1.00	1.00	0.92	1.05
2	1	1	0.25	0.5	1.5	0.375	1	0.17	0.25	0.34	1.00	1.00	0.96	1.01	0.77	1.59
3	1	1	0	0.25	1	0.25	1	0.21	0.26	0.32	1.00	1.00	0.90	∞	0.63	∞
4	0	1	0	1	1	0.5	1	0.28	0.50	0.72	1.00	1.00	0.93	1.01	0.51	∞
5	2	1	0	1	1	0.5	1	0.04	0.14	0.28	1.00	1.00	0.99	1.00	0.98	1.01
6	-1	1	0	1	1	0.5	1	0.50	0.71	0.88	0.99	4.57	0.84	∞	$-\infty$	∞
7	-1	1	0	2	1	0.5	1	0.28	0.68	0.96	1.00	1.00	0.97	1.00	0.51	∞
8	-1	1	0	1.5	1	0.5	1	0.39	0.69	0.93	1.00	1.02	0.91	1.02	0.28	∞
9	-1.5	1	0	1.5	1	0.5	1	0.50	0.78	0.96	1.00	2.35	0.86	∞	$-\infty$	∞

Table 2: Projections of the identified set onto the space of α under a median independence condition (MedI) and a stochastic independence restriction (NPSI) with a nonparametric specification of the distribution of U with the number of intervals $N = 100$ and under stochastic independence restriction with U restricted to be Gaussian (GaussSI. These results are for Structure 2. Censoring probabilities p_0^ℓ, $\overline{p_0}$, and p_0^u are defined as $\min_{z\in\mathcal{R}_Z} \mathbb{P}[Y_1 = 0|Z = z]$, $\mathbb{P}[Y_1 = 0]$, and $\max_{z\in\mathcal{R}_Z} \mathbb{P}[Y_1 = 0|Z = z]$, respectively.

												Identified intervals for α					
Case	Structure 2 parameter values								Censoring Probabilities			GaussSI		NPSI		MedI	
	b	a	k	d_0	d_1	s_{11}	s_{12}	s_{22}	p_0^ℓ	$\overline{p_0}$	p_0^u	$\underline{\alpha}$	$\overline{\alpha}$	$\underline{\alpha}$	$\overline{\alpha}$	$\underline{\alpha}$	$\overline{\alpha}$
1	0	1	1	0	1	1	0.5	1	0.34	0.54	0.76	0.76	1.26	0.38	1.77	0.23	∞
2	5	1	1	0	1	1	0.5	1	0.01	0.08	0.24	1.00	1.00	0.93	1.01	0.93	1.02
3	0	1	10	0	1	1	0.5	1	0.00	0.03	0.11	1.00	1.00	0.98	1.00	0.98	1.00
4	0	1	1	0	5	1	0.5	1	0.16	0.57	1.00	1.00	1.00	1.00	1.00	1.00	1.00
5	0	1	1	0	1	1	0.0	0.1	0.27	0.55	0.92	1.00	1.00	1.00	1.00	0.95	1.02

Table 3: Maximum likelihood estimates of α, the coefficient on log nondurable expenditure, using a classical Tobit model making no allowance for endogeneity and using a complete triangular model with Gaussian unobserved variables.

Years	N	% zero	-	Triangular model	Tobit model
2000-04	18473	68	*estimate*	-0.130	-0.048
			std err	(0.0043)	(0.0027)
			95% CI	$[-0.139, -0.122]$	$[-0.053, -0.043]$
2005-09	15885	74	*estimate*	-0.121	-0.043
			std err	(0.0046)	(0.0028)
			95% CI	$[-0.130, -0.112]$	$[-0.048, -0.038]$

Table 4: Confidence regions for α, the coefficient on log nondurable expenditure, obtained using an incomplete IV model with selected conditional quantiles of the unobservable required to be independent of Z.

Years	-	Quantile independence restricting:		
		3 quantiles	5 quantiles	7 quantiles
2000-04	*estimate*	$[-0.450, -0.075]$	$[-0.241, -0.093]$	$[-0.214, -0.109]$
	95% CI	$[-0.698, -0.069]$	$[-0.262, -0.091]$	$[-0.236, -0.102]$
2005-09	*estimate*	$(-\infty, -0.050]$	$(-\infty, -0.075]$	$[-0.497, -0.087]$
	95% CI	$(-\infty, -0.045]$	$(-\infty, -0.071]$	$(-\infty, -0.084]$

Table 5: Confidence regions for α, the coefficient on log nondurable expenditure, obtained using an incomplete IV model with selected conditional quantiles of the unobservable restricted Gaussian, required to be independent of Z.

Years	-	Gaussian quantile independence restricting:		
		3 quantiles	5 quantiles	7 quantiles
2000-04	*estimate*	$[-0.450, -0.075]$	$[-0.240, -0.093]$	$[-0.206, -0.109]$
	95% CI	$[-0.698, -0.069]$	$[-0.261, -0.092]$	$[-0.236, -0.103]$
2005-09	*estimate*	$(-\infty, -0.050]$	$(-\infty, -0.075]$	$[-0.496, -0.087]$
	95% CI	$(-\infty, -0.045]$	$(-\infty, -0.071]$	$(-\infty, -0.084]$