CONSTRUCTION OF STRUCTURAL FUNCTIONS IN NONSEPARABLE CONDITIONAL INDEPENDENCE MODELS

SAMI STOULI

JOB MARKET PAPER

Abstract. This paper considers identification and estimation in models imposing conditional independence restrictions and featuring a scalar disturbance. It is shown that for this class of models the disturbance is endowed with a specific structure that is highlighted and exploited to obtain full knowledge of the structural function. Structural effects of a policy or treatment are allowed to vary across subpopulations that can be located on the joint distribution of unobservables of the model. In nonseparable triangular models with continuous endogenous variables this approach delivers identification of structural functions conditional on values of the control variable. These results are obtained after a reanalysis of local identification in nonseparable triangular models, where the connection between Chesher (2003) and Imbens and Newey (2009) is made explicit. It is also shown that in the presence of nonmonotone continuous instruments, nonseparable triangular models are always overidentifying. A generic estimation framework is described, and an analog estimator based on a new regression method ("dual regression") is proposed. An empirical application illustrates the methodology by estimating gasoline demand functions in the United States.

Date: November 19, 2012. Email: s.stouli@ucl.ac.uk. I am grateful to Andrew Chesher and Lars Nesheim for their patience as well as their constant support and guidance. I also thank Roger Koenker, Whitney Newey and Richard Spady for numerous discussions at various stages of this work. Richard Blundell, Joel Horowitz and Matthias Parey kindly provided me with the data. I also benefited from comments and discussions with Josh Angrist and Frank Vella, and participants at the first joint UCL-LSE Doctoral Workshop in Econometrics, the 12th Doctoral Workshop in Economic Theory and Econometrics (MOOD 2012) in Rome, EIEF, the ESEM 2012 in Malaga, and the Labour & Applied Microeconometrics Seminar at the University of Oxford, where parts of this work were presented. I also gratefully acknowledge the financial support of the UK Economic and Social Research Council through a postgraduate grant (ES/G031223/1) and through a grant (RES-589-28-0001) to the ESRC Centre for Microdata Methods and Practice (CeMMAP) and of the Royal Economic Society through the Junior Fellowship Scheme.

1. **Introduction**

Knowledge of the *ceteris paribus* effect of an explanatory variable X on each point of the distribution of an outcome variable Y provides valuable information for policy analysis. It accounts for heterogeneity in microeconomic data and provides a very accurate understanding of the policy or treatment under study. When X is endogenous, however, full recovery of the underlying structural relationship between X and Y often requires imposing strong restrictions on the support of observables, or on the stochastic properties of the model, or even directly on the form of the relationship of interest. For instance, it is known that single-equation instrumental variables (SE-IV) models fail to point-identify the structural function relating X to Y when Y is discrete (Chesher (2010)). This paper provides weak conditions under which models imposing conditional independence restrictions and featuring a scalar disturbance preserve their point-identifying power.

Suppose a continuously distributed outcome variable Y admits the structural representation

$$Y = H(X, \varepsilon), \tag{1.1}$$

where the structural function H is strictly increasing in a scalar source of stochastic variations ε. Models featuring these properties are said to be nonseparable and allow for flexible modelling of the heterogeneity inherent to most populations of interest in applied work. Then, for this class of models, this paper shows that if a random variable V such that X is independent of ε conditional on V is available, all features of the structural relationship between X and Y can be recovered. In particular, the random variable V is exploited as an additional source of heterogeneity in order to obtain structural effects that vary across the joint distribution of ε and V. This is achieved without any parametric assumption on the structure of the model or the joint distribution of the data.

The main contribution of the paper is to highlight and exploit the specific structure of the disturbance stemming from the conditional independence restriction. A core feature of the methodology is the focus on the properties of the disturbance instead of on the structural function itself. This approach differs from existing identification strategies in nonseparable models which focus on identification of the structural function H. Here, the proposed construction can be characterized as dual in the sense that the structural function is a byproduct of the construction of a stochastic element satisfying specified properties. Incorporating all the information provided by knowledge of the structure of the disturbance reveals that the model preserves its point-identifying power under weak conditions.

The second contribution of this paper is to highlight that the construction of the structural function in conditional independence models can be decomposed into two distinct levels of analysis. Locally, at some specified points, values of the joint distribution of observables identify values of the joint distribution of Y, X and V. Globally, the structural function and the conditional quantile function of ε given V are constructed as functionals of the joint distribution of Y, X and V. The construction is said to be global when knowledge of the functions of interest is achieved at all points. An implication is that the global analysis of structural functions can be entirely formulated in terms of Y, X and V. This insight forms the basis of the dual construction that ensues and allows for incorporating information not used at the local level: strict monotonicity, the functional form of the disturbance and full conditional independence.

Three further contributions follow from the conducted local analysis. First, local identification of the structural function in nonseparable triangular models, as first considered in Chesher (2001, 2003), is revisited, describing how bi-dimensional unobserved heterogeneity inherent to these models can be exploited. Second, a fruitful connection with the global approach developed in Imbens and Newey (2009) is made explicit and the relationship between these two important papers is established. Therefore, this paper also contributes to the literature on nonseparable models by relating two approaches that have remained largely disconnected. Third, nonseparable triangular models with a continuous endogenous regressor are shown to be always overidentifying in the presence of a nonmonotone continuous instrument. This observation holds regardless of the dimensionality of the structural disturbance. In particular, the so-called Quantile Structural Function (QSF)[1] of Imbens and Newey (2009) is overidentified in that case, a fact that does not follow from any currently available result.

Finally, since the main identification result is constructive, estimation is not tied to a specific method and a generic estimation framework is described. The methodology is applied to a semiparametric triangular location-scale model and an estimator is introduced in order to illustrate the benefits of the dual construction advocated in this paper. The estimator builds on the dual regression methodology introduced in Spady and Stouli (2012). This regression framework constitutes an alternative to the quantile regression process for the global estimation of conditional distribution functions. Thus, identification and estimation are integrated, and all information available from the restrictions of the model is incorporated by the estimator.

[1]The QSF is a structural object that characterizes how the endogenous regressor affects the distribution of outcomes when the disturbance of the model is a random vector. See the discussion in Section 3.2.

Nonseparable models and control variable methods in economic applications. From an applied perspective, models incorporating nonseparability in the disturbances strongly weaken the *a priori* restrictions imposed by practitioneers and significantly reduce the risk of misspecification in models with endogenous regressors. In addition, models allowing for heterogenous effects across treatment levels (nonlinearity) and across the whole of the outcome variable distribution (nonseparability) allow applied researchers to characterize the effect of the policy or treatment under study across individuals in the population. This type of knowledge is essential for policymakers to learn about which part of the population may be harmed by a given policy and which part may actually benefit from it.

In this respect, control variable methods, as considered in this paper, can also be interpreted as including an additional dimension of heterogeneity to the analysis that can be exploited in order to model multiple sources of heterogeneity. Returns to schooling is the traditional example that serves to motivate nonseparable triangular models (Chesher (2003); Florens, Heckman, Meghir, and Vytlacil (2008); Jun (2009)). Beyond the standard use of control function methods, since the main result of this paper applies to any control variable, it is of direct interest to recent applied work estimating latent variables, for instance in labor economics and in the work of Heckman and coauthors (see Heckman, Stixrud, and Urzua (2006); Cunha, Heckman, and Schennach (2010)), or in item response theory (see Spady (2006, 2007)). Latent variables such as cognitive and noncognitive skills in the study of childhood development and adult outcomes could be used as control variables.

Control variable methods like the one introduced in this paper are also, more traditionally, of direct practical interest as a way of addressing endogeneity. They constitute a powerful yet simple methodology allowing for the recovery of structural effects and have been recently applied to address a variety of empirical questions such as welfare analysis and consumer demand (see Hoderlein and Vanhems (2011) and Hausman and Newey (2012)). Since endogeneity is a core issue in numerous fields of economics such as labor, health or development economics, applicability of control variable methods is broad and a general motivation is given in Imbens (2007) and Imbens and Newey (2009), who also illustrate their method with an analysis of Engel curves.

Related literature. This work takes as a starting point two important papers studying identification in nonseparable triangular models. The local analysis originates in Chesher (2003) who introduced a quantile-based identification methodology to show local identification of derivatives of structural functions in nonseparable triangular models. Second, that the structural function be treated as a functional of the conditional distribution function of Y given X and V is an insight put forward by Imbens and Newey (2009) in order to show that in the absence of a scalar disturbance, the QSF is identified. The dual construction proposed in this paper highlights the identifying power of the model when a scalar disturbance is considered instead, and builds on the framework of Spady and Stouli (2012) for the construction of conditional distribution functions.

A seminal paper in this literature is Blundell and Powell (2003) who introduced the Average Structural Function. Florens, Heckman, Meghir, and Vytlacil (2008) consider a nonseparable triangular framework, imposing polynomial restrictions on the structural equation. Recent work on identification in nonseparable triangular models with a scalar disturbance includes d'Haultfoeuille and Fevrier (2011) and Torgovitsky (2011). They obtained results on point identification of the structural function in the presence of discrete instruments.

Previous work on quantile-based estimation of triangular models includes Amemiya (1982) who first introduced a class of two-stage median regression estimators. More recently, Lee (2007) considered a semiparametric quantile regression version of a triangular model where the control function enters additively. Removing the separability intrinsic to location models, Ma and Koenker (2006) developed two general classes of estimators for a location-scale form of a parametric triangular model. Building on Chesher (2003), they construct an estimator which allows for a description of the entire stochastic relationship between the endogenous variable and the outcome. Following their approach Jun (2009) suggested a semiparametric estimator based on a random coefficients model. He considers a linear triangular model while allowing for nonseparability in the unobservables. Chernozhukov, Fernandez-Val, and Kowalski (2011) also consider a nonseparable triangular model and use a control function method to address endogeneity in quantile regression models with censoring.

An alternative approach to identification and estimation of nonseparable models in the presence of endogeneity is the SE-IV model, where the source of endogeneity is not specified and is left unrestricted. Although this model is more general, its estimation is also notoriously difficult due to ill-posedness of the associated inverse problem. For SE-IV models with a continuous outcome, see Chernozhukov and Hansen (2005); Horowitz and Lee (2007); Gagliardini and Scaillet (2012); Chen and Pouzo (2012) for developments based on conditional quantile restriction, and Spady and Stouli (2012) for a distributional approach via dual regression. The case of discrete outcomes is considered in Chesher (2010); Chesher, Rosen, and Smolinski (2011); Chesher and Smolinski (2012).

Organization of the paper. In the next section the model is described and structural features of interest are discussed. In Section 3, a local analysis, at some specified point, of the structural function is conducted. Section 4 considers global identification of the structural function both conditional on a specified value of V and unconditionally. Section 5 describes a generic estimation procedure and an analog estimator based on dual regression is introduced for a semiparametric triangular location-scale model. Section 6 illustrates the methodology with an application to gasoline demand in the United States.

2. A Nonseparable Conditional Independence Model

2.1. **The Framework.** Let Y be a scalar random variable with continuous support $\mathcal{Y}$ and X a scalar random variable with support $\mathcal{X}$. The model of interest features an outcome equation

$$Y = H(X, \varepsilon), \tag{2.1}$$

where X is said to be endogenous and is not assumed to be independent of the scalar disturbance ε, and the structural function $H(x, e)$ is restricted to be strictly increasing in e.[2] This setup corresponds to a traditional concern in econometrics when one is interested in discovering the structural relationship between an observed choice or treatment and a subsequent outcome of interest. Observational units may choose a value of X based on motives known only to them that are not independent of the disturbance ε determining the realized outcome; these motives effectively being the source of endogeneity of X.

In order to achieve knowledge of the structural relationship between Y and X, it is assumed that there exists a random variable V, typically unobserved, such that X and ε are independent once V has been conditioned on. Thus, conditional independence is a key identifying restriction of the model. This restriction specifies the source of endogeneity and requires the econometrician to have at her or his disposal an observed source of exogenous variations in X, for instance provided by an *instrumental variable*, in order to isolate and control for endogenous variations coming from V. Alternatively, the researcher may have an interpretation or some *a priori* knowledge about the source of endogeneity in the model allowing full recovery of V from this information. An example of which is given by the availability of *repeated measurements* of a latent variable believed to be causing the endogeneity of X.

[2]For notational simplicity, X is taken to be scalar and a random vector W, say, of observed discretely or continuously distributed characteristics can be included in the structural function H without altering the analysis. All results extend to vector X and additional covariates W.

In addition to conditional independence, the structural function is restricted to be strictly monotonic in a scalar source of stochastic variations ε.[3] This is the rank invariance assumption common in the literature on nonseparable models (Chernozhukov and Hansen (2005); Torgovitsky (2011))[4]. It can be interpreted as describing a world in which the ranking of observational units in the distribution of Y is invariant to the chosen value of X. That is to say, observational units faring well under a given policy or treatment value x, would also fare well would X be set to a different value.

2.2. **Structural features of interest.** As a starting point, it is identification of a value of the structural function at a specified value x^* of X for the subpopulation with value v^* of V, that is sought. In Equation (2.1), consider taking the conditional τ-quantile of Y given $X = x^*$ and $V = v^*$, denoted $Q_{Y|XV}(\tau|x^*, v^*)$. The equivariance property of quantiles under monotone transformation and local application of independence of X and ε conditional on V gives:

$$Q_{Y|XV}(\tau|x^*, v^*) = H(x^*, Q_{\varepsilon|V}(\tau|v^*)). \tag{2.2}$$

Knowledge of the structural function H evaluated at fixed values x^* and $Q_{\varepsilon|V}(\tau|v^*)$ of X and ε thus requires knowledge of $Q_{Y|XV}(\tau|x^*, v^*)$, and is the value of the structural function at different levels of two sources of heterogeneity, for a given value x^* of X. Indeed, $Q_{\varepsilon|V}(\tau|v^*)$ is the τ-quantile of ε for the subpopulation defined by $V = v^*$. Therefore, $H(x^*, Q_{\varepsilon|V}(\tau|v^*))$ ought to be interpreted as the value of the structural function for the subpopulation defined by $X = x^*$, $\varepsilon = Q_{\varepsilon|V}(\tau|v^*)$ *and* $V = v^*$.

Although unusual, this feature of the model is of interest in itself since it provides useful information on the value of the outcome Y for subpopulations defined by their location on the joint distribution of ε and V. In the classical Returns-to-Schooling example, one may actually be interested in the structural relationship between education (X) and earnings (Y) across ability levels (V)[5]. Besides, $H(x^*, Q_{\varepsilon|V}(\tau|v^*))$ *is* a structural feature of the model, since for the subpopulation defined by $V = v^*$, variations in X are independent of variations in ε.

[3]Hoderlein and Mammen (2007) consider the case of multivariate ε and exogenous X and discuss the implications of relaxing the strict monotonicity restriction.

[4]This assumption differs from the formalism adopted in the potential outcomes literature in which rankings across treatment values can be unrelated (see Imbens and Angrist (1994), Heckman and Vytlacil (2007), for instance).

[5]See Chesher (2003) and Jun (2009) for a detailed discussion of returns-to-schooling in the context of nonseparable models.

Three additional reasons motivate a detailed analysis of $H(x^*, Q_{\varepsilon|V}(\tau|v^*))$. First, all structural features of conditional independence models with a scalar disturbance can be expressed as functionals of $H(x, Q_{\varepsilon|V}(\tau|v))$. Therefore, a thorough understanding of $H(x^*, Q_{\varepsilon|V}(\tau|v^*))$ paves the way to a global construction of the structural function, at all points. Second, a careful analysis will demonstrate that, in nonseparable triangular models, for a fixed value v^* of V, nonmonotonicity of the relationship between X and the instrument leads to overidentification of $H(x^*, Q_{\varepsilon|V}(\tau|v^*))$. Last, one may sometimes want to restrict attention to local features of the model, for instance when full conditional independence seems too strong a restriction for the particular problem at hand.

The next step in the construction considers identification of the structural relationship between Y and X for a subpopulation of interest defined by a specified value v^* of V. That is, knowledge of $H(x, Q_{\varepsilon|V}(\tau|v^*))$ for all x in $\mathcal{X}$ and τ in $(0,1)$ will be sought next. This approach differs from the identification literature on nonseparable models that considers V to be a device in order to recover structural effects and seeks knowledge of $H(x, e)$ at unconditional values e of ε (with the exception of Chesher (2003)). On the other hand, the fact that V may itself be an explanatory variable of interest motivates the particular attention devoted in this paper to structural features of the model that depend on V.

2.3. **Main Conditions.** Let $\mathcal{X}_v = supp(X|V = v)$. The model is fully described by the following conditions and normalization.

Condition 1. *(DGP)* The data generating process for Y is given by $Y = H(X, \varepsilon)$, where Y and ε are continuously distributed scalar random variables and $H(x, e)$ is a function strictly increasing in e for all $x \in \mathcal{X}$.

Condition 2. *(Control variable)* There exists a known continuously distributed random variable V with bounded support $\mathcal{V} = [\underline{v}, \bar{v}]$, such that X and ε are independent conditional on V.

Normalization 1. For some $v^* \in \mathcal{V}$, let $\varepsilon^* = F_{\varepsilon|V}(\varepsilon|v^*)$. Then Y admits the structural representation $Y = H^*(X, \varepsilon^*)$, where for all $x \in \mathcal{X}$ the function $H^*(x, \tau)$ is strictly increasing in τ, and ε^* is normalized to be uniformly distributed on $[0, 1]$ conditional on $V = v^*$.

Condition 3. *(Support conditions)* (i) For all $v \in \mathcal{V}$, there is a value $x(v)$ of X in $\mathcal{X}_v$, such that $supp(Y|X = x(v), V = v) = \mathcal{Y}$. (ii) For $v^* \in \mathcal{V}$ specified in Normalization 1, $\mathcal{X}_{v^*} = \mathcal{X}$.

Condition 2 requires that the random variable V be known, that is V must either be directly observable or identified from data. In the leading example of nonseparable triangular models, it is well known that Condition 2 is satisfied by the availability of an instrumental variable Z such that the DGP of X is given by $X = G(Z, \eta)$, where $G(z, \eta)$ is a function strictly increasing in η, a scalar random variable independent of Z, for $Z \perp\!\!\!\perp (\varepsilon, \eta)$ (see Theorem 1 of Imbens and Newey (2009) or Kasy (2011)). Then the DGP admits a corresponding quantile regression representation

$$X = Q_{X|Z}(V|Z), \tag{2.3}$$

characterized by independence of the regression rank variable V and the regressor Z, and $V = F_{X|Z}(X|Z)$ serves as the control variable in that model.

Normalization 1 defines a scalar random variable ε^* as a strictly monotone transformation of ε and gives an equivalent structural representation for Y. ε^* satisfies $F_{\varepsilon^*|V}(\tau|v^*) = \tau$ since by definition of ε^* and strict monotonicity of $F_{\varepsilon|V}(\tau|v^*)$ in τ there is:

$$P[F_{\varepsilon|V}(\varepsilon|v^*) \leq \tau | V = v^*] = P[\varepsilon \leq Q_{\varepsilon|V}(\tau|v^*)|V = v^*] = \tau. \tag{2.4}$$

Therefore, ε^* is normalized to be uniformly distributed on the unit interval conditional on $V = v^*$, and the structural representation follows from the definition of ε^*:

$$Y = H(X, \varepsilon) = H(X, Q_{\varepsilon|V}(\varepsilon^*|v^*)) \equiv H^*(X, \varepsilon^*), \tag{2.5}$$

where (H, F_ε) and (H^*, F_{ε^*}) are observationally equivalent.

The normalization adopted is specific to the setup considered in this paper and differs from the distributional normalization suggested in Matzkin (2003) or used in Chernozhukov and Hansen (2005) and Torgovitsky (2011) where the marginal distribution of ε^* is normalized instead. Under Normalization 1 $H^*(x, \tau)$ is the τ-quantile of the counterfactual random variable $Y_x \equiv H(x, \varepsilon)$ conditional on $V = v^*$. Since the QSF introduced by Imbens and Newey (2009) is defined as the τ-quantile of Y_x, $H^*(x, \tau)$ can be interpreted as a local (to v^*) QSF.

Condition 3(i) is a common support assumption on the support of Y which is necessary for identification of $Q_{\varepsilon^*|V}$. Condition 3(i) effectively imposes that ε^* has full support conditional on V. Condition 3(ii) ensures that $H^*(x, \tau)$ is identified for all values of X. In the context of a triangular model, for X continuously distributed, Condition 3(ii) is satisfied if a continuously distributed instrumental variable is available. This condition imposes that for the subpopulation defined by $V = v^*$ the instrumental variable varies sufficiently for $Q_{X|Z}(v^*|Z)$ to have range the full support of X.

3. **Local Analysis of Structural Functions**

In this Section, identification of the value $H(x^*, Q_{\varepsilon|V}(\tau|v^*))$ of the structural function in nonseparable triangular models is revisited. This local analysis illustrates the methodological point that global construction of structural functions can be distinguished from local identification of $H(x^*, Q_{\varepsilon|V}(\tau|v^*))$. This value of the structural function is shown to be potentially locally overidentified by the distribution of observables, i.e $F_{Y|XZ}(Y|X, Z)$ in triangular models. Once known from data, $H(x^*, Q_{\varepsilon|V}(\tau|v^*))$ serves to achieve global knowledge of the structural function; this is illustrated by the construction of the QSF of Imbens and Newey (2009).

3.1. **Local Identification Under Instrumental Variable Availaibility.** In the context of triangular models, although implicitly understood from results given in Chesher (2003) and in Imbens and Newey (2009), the value $H(x^*, Q_{\varepsilon|V}(\tau|v^*))$ has not received a direct treatment in the literature. Instead, for specified values z^* and v^* of Z and V, under the local conditional independence assumption $Q_{\varepsilon|VZ}(\tau|v^*, z^*) = Q_{\varepsilon|V}(\tau|v^*)$, from Chesher (2003) it is known that

$$Q_{Y|XZ}(\tau|Q_{X|Z}(v^*|z^*), z^*) = H(Q_{X|Z}(v^*|z^*), Q_{\varepsilon|V}(\tau|v^*)), \tag{3.1}$$

where the value $H(Q_{X|Z}(v^*|z^*), Q_{\varepsilon|V}(\tau|v^*))$ of the structural function is directly identified from data by $Q_{Y|XZ}(\tau|Q_{X|Z}(v^*|z^*), z^*)$, the conditional τ-quantile of Y given $X = Q_{X|Z}(v^*|z^*)$ and $Z = z^*$. Two related difficulties with expression (3.1) are its interpretability and the construction of analog estimators[6]. The following discussion suggests a change of perspective in order to facilitate both interpretation and analog estimation by reformulating the local identification analysis in terms of *specified* values of X and V instead of values of X and V *induced* by values of Z and V, as in Equation (3.1).

A useful observation is that Equation (3.1) can be used to obtain knowledge of $H(x^*, Q_{\varepsilon|V}(\tau|v^*))$ instead. Indeed, from (3.1) it follows that for a value z of Z such that $x^* = Q_{X|Z}(v^*|z)$, the value $H(x^*, Q_{\varepsilon|V}(\tau|v^*))$ is actually identified by $Q_{Y|XZ}(\tau|Q_{X|Z}(v^*|z), z)$. However, in the absence of further restrictions, there may not be a unique value of Z such that $x^* = Q_{X|Z}(v^*|z)$, and therefore the value of interest of the structural function is potentially *over*identified in that context.

[6]These difficulties partly explain why several different proposals have been made in order to carry estimation based on (3.1) or the formula for the structural derivative of H with respect to x at a fixed value of V given in Chesher (2003) - see Ma and Koenker (2006); Lee (2007); Jun (2009).

In order to give a unified treatment of just- and over- identification of $H(x^*, Q_{\varepsilon|V}(\tau|v^*))$ accounting for a set of instrumental values such that $x^* = Q_{X|Z}(v^*|z)$ of arbitrary cardinality, the concept of *instrumental set* is now introduced. Consider the set of values of Z corresponding to a given value x of X for a given value v of V. Then, the instrumental set $\mathcal{Z}^*(x, v)$ is defined as follows:

$$\mathcal{Z}^*(x, v) = \{z \in \mathcal{Z} : \; Q_{X|Z}(v|z) = x\}. \tag{3.2}$$

This is the set of instrumental values such that observational units choose a value x of X, conditional on $V = v$. For $|\mathcal{Z}^*(x, v)|$ the number of elements of $\mathcal{Z}^*(x, v)$ and $j = 1, \ldots, |\mathcal{Z}^*(x, v)|$, denote by $z_j(x, v)$ the elements of that set. For $X = x^*$ and $V = v^*$, the abbreviated notation $z_j(x^*, v^*) \equiv z_j^*$ and $\mathcal{Z}^*(x^*, v^*) \equiv \mathcal{Z}^*$ is used. Then whenever $|\mathcal{Z}^*|$ is greater than one, this is an instance of *local* overidentification.

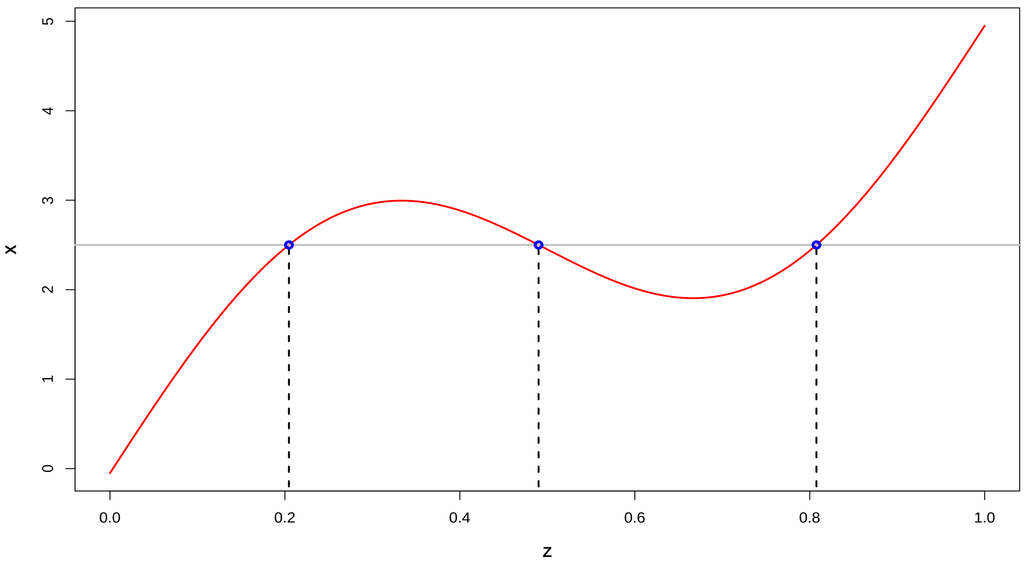
The case of a scalar instrument Z is of particular interest in practice. In that case, when $Q_{X|Z}(v^*|z)$ is nonconstant in z, the set $\mathcal{Z}^*$ has at most countably many elements. Note that if in addition Z has bounded support, then $\mathcal{Z}^*$ contains at most finitely many elements, a fact that guarantees that the instrumental set can easily be constructed. On the other hand, $\mathcal{Z}^*$ is a singleton when Z is scalar and $Q_{X|Z}(v^*|z)$ is strictly monotone in z.[7]

Figure 1a illustrates the construction of the instrumental set for a particular nonmonotone median function of X conditional on Z and for the pair of values $x^* = 2.5$ and $v^* = .5$. The elements of the instrumental set are given by the roots of the equation $Q_{X|Z}(.5|z) = 2.5$. The fact that there may be multiple relevant values of Z for $x^* = 2.5$ is apparent from Fig.1a that shows the three roots of the equation $Q_{X|Z}(.5|z) = 2.5$.

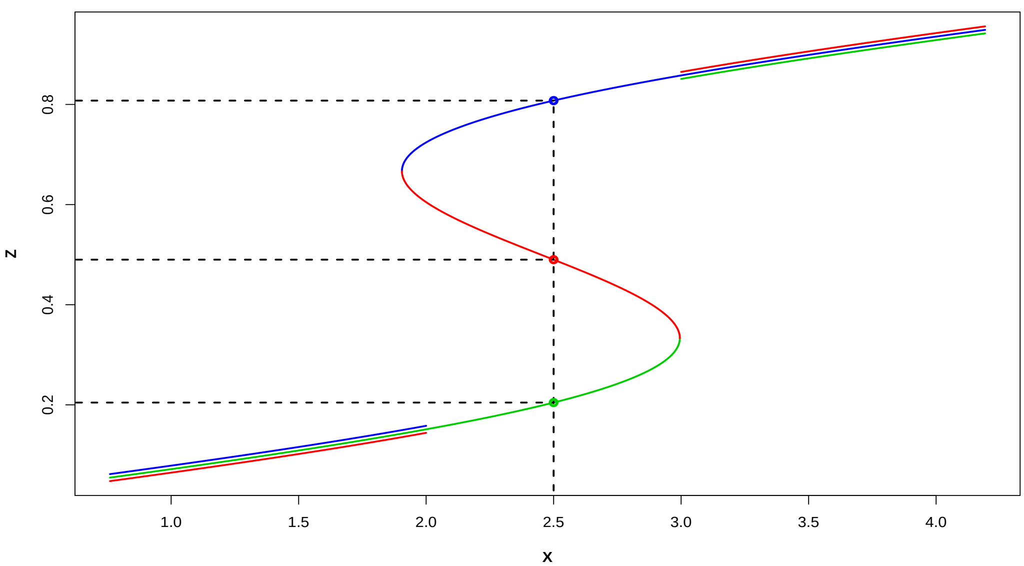
Formally, the set of instrumental values $\mathcal{Z}^*$ is the image of the inverse functions of $Q_{X|Z}(v^*|z)$ with respect to z. For all $v \in \mathcal{V}$, define the map $q_v : \mathcal{Z} \to \mathcal{X}_v$, the conditional v-quantile function of X conditional on Z, denoted $q_v(z)$, with range $\mathcal{X}_v$. By construction, the function $q_v(z)$ is a surjective function and thus is right invertible[8]. Also, the right inverse of q_v may only be piecewise monotone in x. To account for piecewise monotonicity, let $\mathcal{Z}_1, \ldots, \mathcal{Z}_{|\mathcal{Z}^*(x,v)|}$ be a partition of $\mathcal{Z}$ such that for each $j = 1, \ldots, |\mathcal{Z}^*(x, v)|$ the function $g_v^j : \mathcal{X}_v \to \mathcal{Z}_j$ is a right inverse of q_v and each g_v^j is one-to-one. Then $\{z_j(x, v)\}_{j=1}^{|\mathcal{Z}^*(x,v)|}$ is the collection of instrumental

[7]Local overidentification arising in the presence of multiple instrumental values is discussed in Chesher (2003, 2007) whereas the focus is here on highlighting nonmonotonicity of $Q_{X|Z}(v^*|z)$ in z (nonmonotone instruments) as a particular instance of local overidentification.

[8]Right invertibility of q_v means that there exists a function $g_v : \mathcal{X}_v \to \mathcal{Z}$ such that $q_v(g_v(x)) = x$ for every x. Note that when q_v is not monotone, a unique inverse function does not exist because q_v does not have a unique right inverse. Moreover, any right inverse g_v is not surjective. There exists $z \in \mathcal{Z}$ such that $g_v(q_v(z)) \neq z$.



(A)



(B)

FIGURE 1. A conditional median function (A), and its piecewise monotone right inverses $g_v^j(x)$, $j = 1, \ldots, 3$ (green, red and blue solid lines) (B), with $x^* = 2.5$ and $\{z_j(2.5, .5)\}_{j=1}^3 = \{0.21, .49, .81\}$.

values delivered by $\{g_v^j(x)\}_{j=1}^{|\mathcal{Z}^*(x,v)|}$, the elements of $\mathcal{Z}^*(x,v)$. Fig.1b shows the right inverses of the conditional median function shown in Fig.1a.

In order for $H(x^*, Q_{\varepsilon|V}(\tau|v^*))$ to be identified, the set $\mathcal{Z}^*$ must contain at least one element. Under this condition and local application of the conditional independence assumption in Condition 2, the value of the structural function $H(x^*, Q_{\varepsilon|V}(\tau|v^*))$ is then identified from data by the conditional quantile function $Q_{Y|XZ}(\tau|x^*, z_j^*)$, for $j \in \{1, \ldots, |\mathcal{Z}^*|\}$. This result is stated in Proposition 1.

Proposition 1. *Suppose that the instrumental set $\mathcal{Z}^*(x^*, v^*)$ is nonempty and Conditions 1 and 2 hold. Then for all $\tau \in (0,1)$, there is*

$$H(x^*, Q_{\varepsilon|V}(\tau|v^*)) = Q_{Y|XZ}(\tau|x^*, z_j^*) \qquad \forall j \in \{1, \ldots, |\mathcal{Z}^*(x^*, v^*)|\}. \tag{3.3}$$

When $|\mathcal{Z}^*(x^*, v^*)| > 1$, $H(x^*, Q_{\varepsilon|V}(\tau|v^*))$ is overidentified.

Equation (3.3) follows from Equation (2.2) and since by definition of $\mathcal{Z}^*$, for all $j \in \{1, \ldots, |\mathcal{Z}^*|\}$, conditioning on (x^*, z_j^*) is equivalent to conditioning on (x^*, v^*). Proposition 1 provides an expression for a value of the structural function that can be computed directly, thus it is constructive and is easily interpreted. Second, note that Proposition 1 implies that

$$Q_{Y|XZ}(\tau|x^*, z_j^*) = Q_{Y|XZ}(\tau|x^*, z_{j'}^*) \qquad \forall (j, j')/j \neq j', \tag{3.4}$$

since the value of the conditional τ-quantile of Y does not change for two different pairs of values (z_j^*, v^*) and $(z_{j'}^*, v^*)$ of Z and V - which is the essence of the exclusion restriction. This implication of Proposition 1 gives a basis for specification testing in nonseparable triangular models. Last, since for each j, $Q_{Y|XZ}(\tau|x^*, z_j^*) = Q_{Y|XV}(\tau|x^*, v^*)$, Proposition 1 provides a bridge to the Imbens and Newey (2009) methodology.

3.2. **Connection to the Quantile Structural Function.** Proposition 1 is a key step towards establishing the connection between the local identification approach developed in Chesher (2003) and the global approach of Imbens and Newey (2009). For scalar ε, V uniformly distributed on $[0,1]$ and $p \in (0,1)$, the QSF is defined as $H(x, Q_\varepsilon(p))$, the p-quantile of $H(x, \varepsilon)$, and is shown by Imbens and Newey (2009) to be given by

$$H(x, Q_\varepsilon(p)) = \inf\left\{y : \int F_{Y|XV}(y|x,v)dv \geq p\right\}, \tag{3.5}$$

where by Proposition 1 and definition of $F_{Y|XZ}(y|x,z)$, $F_{Y|XV}(y|x,v)$ can be expressed as

$$F_{Y|XV}(y|x,v) = \int 1\left(Q_{Y|XZ}(\tau|x,z_j(x,v)) \leq y\right) d\tau \qquad \forall j \in \{1,\ldots,|\mathcal{Z}^*(x,v)|\}. \tag{3.6}$$

Thus, for scalar ε the QSF is the structural function and is constructed as a functional of $Q_{Y|XZ}$. The main condition for identification of the QSF is that V has full support conditional on X (Assumption 2 of Imbens and Newey (2009)). This condition is clearly satisfied if $\mathcal{Z}^*(x,v)$ is nonempty for all v in $\mathcal{V}$ and all x in $\mathcal{X}$. The next Proposition states this result.

Proposition 2. *Suppose that Conditions 1 and 2 hold. If, for all $x \in \mathcal{X}$ and all v in $\mathcal{V}$, $\mathcal{Z}^*(x,v)$ is nonempty, then, for all $x \in \mathcal{X}$ and $p \in (0,1)$, $H(x, Q_\varepsilon(p))$ is given by*

$$H(x, Q_\varepsilon(p)) = \inf\left\{y : \int\int 1\left(Q_{Y|XZ}(\tau|x,z_j(x,v)) \leq y\right) d\tau dv \geq p\right\}, \tag{3.7}$$

for $j \in \{1,\ldots,|\mathcal{Z}^*(x,v)|\}$.

Proposition 2 has two main implications. First, upon using Proposition 1, this result relates the local-based identification methodology of Chesher (2003) and the global approach of Imbens and Newey (2009). The key observation in order to establish the link between the two methodologies was Proposition 1. Second, the right-hand side of (3.7) does not depend on the dimensionality of ε and corresponds to the QSF when ε is not scalar, instead of the structural function H. Since for each x in $\mathcal{X}$ and each v in $\mathcal{V}$ when $|\mathcal{Z}^*(x,v)|$ is greater than one $Q_{Y|XV}(\tau|x,v)$ is overidentified, Proposition 2 demonstrates that in the presence of nonmonotone continuous instruments, the QSF is always overidentified, independently of the dimensionality of ε.

3.3. **Discussion.** The main motivation for the perspective adopted in this Section is to design a local identification strategy that is both constructive and exhaustive, in the sense of incorporating all *local* information contained in the data. The methodology introduced via the instrumental set is a device allowing for a unified treatment of both the just- and overidentified cases. Since for a pair of values (x^*, v^*) of X and V there potentially correspond multiple pairs of values $\left\{(z_j^*, v^*)\right\}_{j=1}^{|\mathcal{Z}^*|}$ of Z and V, nonmonotonicity is a source of local overidentification in triangular models and it is desirable that the identification analysis accounts for the information provided by potential nonmonotonicity in the first stage, suggesting that $H(x^*, Q_{\varepsilon|V}(\tau|v^*))$ should be constructed locally when nonmonotone instruments are available.

The local anaysis conducted in this Section raises new questions regarding the analysis of nonseparable triangular models. Of particular interest will be the design of estimators that optimally combine additional information arising from local overidentification. Indeed, since $H(x^*, Q_{\varepsilon|V}(\tau|v^*))$ is identified by any element of the collection of values $\left\{Q_{Y|XZ}(\tau|x^*, z_j^*)\right\}_{j=1}^{|\mathcal{Z}^*|}$, $H(x^*, Q_{\varepsilon|V}(\tau|v^*))$ could be computed as any weighted combination of these values, or from either one of them.

Last, Proposition 1 and 2 together show that the structural function can be constructed globally as a functional of $H(x, Q_{\varepsilon|V}(u|v))$. However, the QSF construction is designed to allow for ε to have arbitrary dimensionality, suggesting that for ε known to be scalar an alternative construction exploiting this information is possible. In addition, the QSF does not describe the structural relationship between Y and X locally to a specified value of V. On the other hand, the quantile attack on the problem based on Proposition 1 requires constructing $H(x, Q_{\varepsilon|V}(\tau|v^*))$ one quantile index τ at a time, not exploiting the strict monotonicity and full conditional independence restrictions of the model[9]. The object of the next Section is to describe how to construct $H(x, Q_{\varepsilon|V}(\tau|v^*))$ for all quantile indices and values of X simultaneously, while exploiting knowledge of the dimensionality of ε.

4. **Dual Construction of Structural Functions**

Recalling that under Normalization 1 $H(x, Q_{\varepsilon|V}(\varepsilon^*|v^*)) \equiv H^*(x, \varepsilon^*)$, knowledge of H^* provides complete knowledge of the structural relationship between Y and X for the subpopulation defined by $V = v^*$, which was only considered at some specified point in the previous Section. Locally the analysis could be conducted in terms of the structural function H. Globally, imposing Normalization 1 is necessary in order to be able to separately identify H^* and $Q_{\varepsilon^*|V}$. This Section turns to identification of H^* and the conditional quantile function $Q_{\varepsilon^*|V}$ after having highlighted the particular structure of the disturbance ε^*. Once identification of H^* and $Q_{\varepsilon^*|V}$ is achieved, it is shown that knowledge of the QSF can also be recovered.

4.1. **A Structured Representation for ε^*.** For $H^*(x, \tau)$ strictly monotonic in τ, existence of a random variable V such that $\varepsilon^* \perp\!\!\!\perp X|V$ leads to an equivalent representation for Y:

$$Y = H^*(X, Q_{\varepsilon^*|V}(F_{Y|XV}(Y|X, V)|V)). \tag{4.1}$$

Representation (4.1) follows from the structural representation given in Normalization 1 upon taking the conditional u-quantile of Y given $X = x$ and $V = v$:

$$Q_{Y|XV}(u|x, v) = H^*(x, Q_{\varepsilon^*|XV}(u|x, v)), \tag{4.2}$$

[9]This is in contrast to Chesher (2003) where conditional quantile independence restrictions are central to the argument.

which for $u = F_{Y|XV}(y|x,v)$ and by conditional independence yields

$$y = H^*(x, Q_{\varepsilon^*|XV}(F_{Y|XV}(y|x,v)|x,v)) = H^*(x, Q_{\varepsilon^*|V}(F_{Y|XV}(y|x,v)|v)), \tag{4.3}$$

for all $(y,x,v) \in \mathcal{Y}\times\mathcal{X}\times\mathcal{V}$. Thus, representation (4.1) provides ε^* with a structure embedding the conditional independence restriction. Knowledge of the structure of ε^* is exploited in order to construct the structural function H^* and the conditional quantile function $Q_{\varepsilon^*|V}$.

4.2. **Main Result.** In order to proceed, first define the class Ψ^* of admissible conditional quantile functions ψ^* to which $Q_{\varepsilon^*|V}$ is assumed to belong.

Definition 1. For $v^* \in \mathcal{V}$ specified in Normalization 1 and all $\tau \in (0,1)$, each $\psi^* \in \Psi^*$ satisfies $\psi^{*-1}(\tau, v^*) = \tau$.

Second, define the random variable ε^{ψ^*}.

Definition 2. For $\psi^* \in \Psi^*$, the scalar random variable ε^{ψ^*} is defined by

$$\varepsilon^{\psi^*} = \psi^*(F_{Y|XV}(Y|X,V),V). \tag{4.4}$$

The main result of the paper is now stated.

Theorem 1. *Let $\psi^* \in \Psi^*$ and ε^{ψ^*} be as in Definition 2, and suppose that Conditions 1-3 hold and Normalization 1 is imposed. Then $\varepsilon^{\psi^*} \perp\!\!\!\perp X|V$ if, and only if, there exists a function $h^*(x,\tau)$ strictly increasing in τ such that for all $(y,x,v) \in \mathcal{Y}\times\mathcal{X}\times\mathcal{V}$*

$$y = h^*(x, \psi^*(F_{Y|XV}(y|x,v),v)), \tag{4.5}$$

where $\psi^*(u,v) = Q_{\varepsilon^*|V}(u|v)$ for all $(u,v) \in (0,1)\times\mathcal{V}$ and $h^*(x,\tau) = H^*(x,\tau)$ for all $x \in \mathcal{X}$ and $\tau \in (0,1)$.

The intuition behind the result stated in Theorem 1 follows from the local analysis of the previous Section: if for any subpopulation such that V is held fixed at a specified value, say v^*, variations in X are independent of variations in ε, then a structural function H^* that depends on the chosen value of V can be constructed from this subpopulation only. Given knowledge of H^*, the support condition in Condition 3(i) suffices for identification of $F_{\varepsilon^*|V}$ at all values of ε^* and V. Thus, knowledge of the structural function is obtained at all points by imposing all restrictions of the model simultaneously - conditional independence, scalar disturbance and strict monotonicity.

The function H^* describes the structural relationship between X and Y conditionally on a value of V, allowing the researcher to focus on a subpopulation of interest or on regions of the joint support of X and V where information from data is concentrated. In contrast with the recent literature on nonseparable models (see d'Haultfoeuille and Fevrier (2011) and Torgovitsky (2011)), the proposed methodology thus allows for conditioning on values of the control variable and accounting for unobserved heterogeneity in the model. The result does rely on global identifying restrictions that may not always hold in practice. Theorem 1 shows that H^* is point identified under a support condition local to v^* allowing for choosing a value of V such that Condition 3(ii) is satisfied, and if the support condition holds for all values of V then the QSF can be recovered as well, as shown below.

Applicability of Theorem 1 extends beyond nonseparable triangular models since it holds for any control variable. This suggests that fruitful connections can be established with the literature onmeasurement error and latent variable modelling. Also, Blundell and Matzkin (2010) show that under certain conditions simultaneous equations models admit an equivalent triangular model representation so that results of this paper apply to that case as well.

The result obtained is also directly constructive. Theorem 1 suggests constructing a composite *function* $\psi^*(F_{Y|XV}(y|x,v),v)$ strictly increasing in y and satisfying $\psi^{*-1}(\tau,v^*)=\tau$ for all $\tau\in(0,1)$ such that the *random variable* ε^{ψ^*} is independent of X conditional on V and such that $F_{\varepsilon^{\psi^*}|V}(\tau|v^*)=\tau$, for all $\tau\in(0,1)$. Alternatively, the following corollary provides Theorem 1 with a convenient interpretation, on which the estimation methodology in the next Section is based.

Corollary 1. *Let $\psi^*\in\Psi^*$ and $h^*(x,\tau)$ be a function strictly increasing in τ for all $x\in\mathcal{X}$, and suppose Conditions 1-3 hold and Normalization 1 is imposed. Then*

$$y=h^*(x,\psi^*(F_{Y|XV}(y|x,v),v)) \tag{4.6}$$

for all $(y,x,v)\in\mathcal{Y}\times\mathcal{X}\times\mathcal{V}$ if, and only if, $\psi^*(u,v)=Q_{\varepsilon^*|V}(u|v)$ for all $(u,v)\in(0,1)\times\mathcal{V}$ and $h^*(x,\tau)=H^*(x,\tau)$ for all $x\in\mathcal{X}$ and $\tau\in(0,1)$.

This corollary shows that given knowledge of $F_{Y|XV}$ and V, it is possible to recover H^* and $Q_{\varepsilon^*|V}$ from solving (4.6) for h^* and ψ^*. Estimation in Section 5 builds on this result.

4.3. **Connection to the Quantile Structural Function, revisited.** If Condition 3(ii) holds for all values of V, then the QSF can be recovered as well. Let

$$\varepsilon_v = F_{\varepsilon|V}(\varepsilon|v), \tag{4.7}$$

and define the function $h : \mathcal{X} \times [0,1] \times \mathcal{V} \to \mathcal{Y}$ such that

$$Y = H(X, Q_{\varepsilon|V}(\varepsilon_v|v)) \equiv h(X, \varepsilon_v, v). \tag{4.8}$$

$h(x,\tau,v)$ is the τ-quantile of $H(x,\varepsilon)$ conditional on $V = v$. If Condition 3(ii) holds for each $v \in \mathcal{V}$, then $h(\cdot,\cdot,v)$ is identified for each $v \in \mathcal{V}$ by Theorem 1. From (4.7) and (4.8), $F_{\varepsilon|V}(\varepsilon|v) = h^{-1}(X,Y,v)$ so that

$$F_\varepsilon(\varepsilon) \quad = \quad \int F_{\varepsilon|V}(\varepsilon|v) f_v(v)dv = \int h^{-1}(X,Y,v) f_v(v)dv. \tag{4.9}$$

By definition of the QSF there is

$$H(x, Q_\varepsilon(p)) = \inf \left\{ y : \int h^{-1}(x,y,v) f_v(v)dv \geq p \right\}. \tag{4.10}$$

Thus, Theorem 1 also provides an alternative way of estimating the structural function that could be compared to estimates of the QSF. It is conjectured that when the first stage is well specified these estimates coincide if and only if the disturbance is scalar. This is an important implication since there are no results on the testability of the dimensionality of error terms in models with endogenous regressors, and the interpretation of the object estimated and its derivatives depend of whether ε is a scalar or a vector.

The expression given in (4.10) for the QSF concludes the study of identification. The next Section is devoted to estimation.

5. **Estimation**

This Section describes a generic framework for estimation of four different functions: $F_{X|Z}$, $Q_{Y|XV}$, H^* and $Q_{\varepsilon^*|V}$. The method is generic in the sense that estimation of objects of interest such as conditional distribution functions can be done using any available estimation method. Estimation of the control variable $V = F_{X|Z}(X|Z)$ is of general interest and the proposed estimator for $Q_{Y|XV}$ can be seen as an alternative to quantile regression based methods for estimation of triangular models as in Ma and Koenker (2006), Jun (2009) and Chernozhukov, Fernandez-Val, and Kowalski (2011). Second, estimation of H^* and $Q_{\varepsilon^*|V}$ is considered building on Corollary 1. The estimation framework is illustrated by semiparametric estimation of a triangular location-scale model, the simplest example of a nonseparable model, via dual regression (Spady and Stouli (2012)). Basic elements of the method are described and its applicability to the problem at hand demonstrated.

5.1. **Generic estimation.** A generic framework for estimation of $Q_{Y|XV}$, H^* and $Q_{\varepsilon^*|V}$ is to implement the following two- or three- step estimation strategy:

(1) *(First stage) Estimate the control variable $\hat{V}$. In a triangular model, take*

$$\hat{V} = \hat{F}_{X|Z}(X|Z). \tag{5.1}$$

(2) *(Second Stage) Estimate the conditional distribution function $\hat{F}_{Y|XV}(Y|X, \hat{V})$. Then $\hat{Q}_{Y|XV}(u|x, v)$ is directly available at all $(u, x, v) \in (0, 1) \times \mathcal{X} \times \mathcal{V}$.*

(3) *(Structural functions) Estimate simultaneously $\hat{h}^*$ and $\hat{\psi}^*$ by solving*

$$\min_{(h^*,\psi^*)} \left\| Y - h^*(X, \psi^*(\hat{F}_{Y|XV}(Y|X, \hat{V}), \hat{V})) \right\|, \tag{5.2}$$

for some norm $\|\cdot\|$, where, for $v^* \in \mathcal{V}$, ψ^* satisfies

$$\psi^*(\hat{F}_{Y|XV}(Y|X, v^*), v^*) = \hat{F}_{Y|XV}(Y|X, v^*). \tag{5.3}$$

In order to estimate the conditional distribution functions in both the first and second steps kernel methods as described in Li and Racine (2011) are well-understood and easy to implement. Also, rearrangement and quantile regression can be used for the first two steps. Both linear and nonlinear quantile regression methods could be implemented. Detailed discussions of these methods can be found in Belloni and Fernandez-Val (2011); Chernozhukov, Fernandez-Val, and Melly (2009); Chernozhukov and Galichon (2010).

In the third step, a parametric specification can be adopted for h^* and ψ^*. Series based approximations constitute a nonparametric alternative which is no different than the parametric approach in terms of implementation. In order to exploit monotonicity of $h^*(x, \tau)$ in τ and $\psi^*(u, v)$ in u, shape-preserving splines could be used (see Chen (2007) for a review on sieves and shape-preserving sieves).

In the following, this estimation framework is applied to a triangular location-scale model and dual regression is used to estimate the conditional distribution functions in steps 1 and 2 above.

5.2. **Application: A semiparametric triangular location-scale model.** A convenient and flexible specification is to consider a location-scale model. Such a model forms the basis of the quantile regression representation since it allows for heterogenous slopes of the quantile regression coefficients of all covariates, including the endogenous regressor. This model can be made very flexible by taking transformations of the covariates, such as splines or trigonometric basis functions, an operation that does not alter the exposition below. Thus, the model considered is given by

$$y_i = \beta_{11}^* x_i + \beta_{12}^* \cdot w_i + (\beta_{21}^* x_i + \beta_{22}^* \cdot w_i)\varepsilon_i^* \tag{5.4}$$

$$x_i = \alpha_1 \cdot z_i + (\alpha_2 \cdot z_i)\eta_i \tag{5.5}$$

$$\varepsilon_i^* = \gamma_{11}^* + \gamma_{12}^* \eta_i + (\gamma_{21}^* + \gamma_{22}^* \eta_i) b_i, \tag{5.6}$$

where, for observation i, x_i is the endogenous regressor, w_i is a vector of additional regressors that includes an intercept and z_i is a vector of instrumental variables also including an intercept, and potentially additional covariates as well. η_i and b_i are disturbances normalized to have mean 0 and variance 1. Estimation of the structural function is considered for $v^* = .5$ and, as in the foregoing, the "$*$" notation indicates that a normalization is imposed. Thus, $\beta^* = (\beta_{11}^*, \beta_{12}^*, \beta_{21}^*, \beta_{22}^*)'$ is the vector of structural parameters for the subpopulation defined by $V = v^*$.

In this model Normalization 1 is imposed for $v^* = .5$ by setting $\gamma_{11}^* = 0$ and $\gamma_{21}^* = 1$. Combining these restrictions and the fact that b has mean 0 and variance 1 yields that ε^* has mean 0 and variance 1 conditionally on $\eta = 0$, which is the median value of η. This is equivalent to imposing that ε^* is uniformly distributed conditional on $V = v^*$ in a nonparametric framework but exploits the fact that the distribution of ε^* is restricted to have only nonzero first and second moments as follows from specifying a location-scale representation for ε^* given η in Equation (5.6) (see the discussion in Section 5.3.1). The correspondence between V and η is given by $v = F_\eta(\eta)$ and is discussed in detail in Section 5.3.1 below. Appendix C discusses identification and the normalization condition for other values of v^* in this triangular location-scale model.

Implementation of the three-step estimation procedure described above is now discussed in detail for model (5.4)-(5.6).

5.3. **First- and Second- Stage Estimation via Dual Regression (Spady and Stouli (2012)).**

5.3.1. *Control Variable Estimation.* The first step of the proposed estimation strategy is to estimate the control variable $V = F_{X|Z}(X|Z)$. Estimation of the control variable V can be done by dual regression, a regression technique that reformulates estimation of conditional distribution functions as the construction of a stochastic element $V = F_{X|Z}(X|Z)$ endowed with three properties: (i) V is uniformly distributed on the unit interval, (ii) V is independent of Z, and (iii) $F_{X|Z}(x|z)$ is strictly increasing in x for any value z of Z.

Given a sample of n points $\{z_i, x_i\}$, estimation of the n values $v_i = F_{X|Z}(x_i|z_i)$ using only the three defining properties of V is formulated as a sequence of mathematical programming problems that embodies these requirements and that generalizes the dual formulation of the quantile regression problem. Because the entire conditional distribution function is estimated simultaneously at all points, the resulting conditional quantile functions are largely free of the so-called quantile crossing problem. Thus, dual regression offers a powerful alternative to current methods available for estimation of the quantile regression process, while avoiding the need for *ex post* rearrangement.

The basic dual regression optimization problem adds $dim(Z)$ more constraints to the median dual quantile regression problem[10] and is given by

$$\max_{v} \{x'v| \begin{cases} Z^{\top}(v-\frac{1}{2}) & = 0 \\ Z^{\top}(v^2-\frac{1}{3}) & = 0 \end{cases},\ v \in [0,1]^n\}, \tag{5.8}$$

where x is now an $(n \times 1)$ vector of values of X and Z is an $(n \times dim(Z))$ matrix of instrumental variable values that includes an intercept (an $(n \times 1)$ vector of 1's) and potentially some additional explanatory variables. The set of $dim(Z)$ constraints on sample moments of v^2 would not appear in the dual quantile regression program, producing values of v that are largely 0 and 1, with $dim(Z)$ sample points being assigned v values that are neither 0 nor 1. In order to satisfy program (5.8), the v's have to be moved off $\{0\}, \{1\}$. Since Z contains an intercept, the sample moments of v and v^2 will be $\frac{1}{2}$ and $\frac{1}{3}$, and v and v^2 will be orthogonal to the components of Z, relations that are necessary but not sufficient for uniformity and independence.

It is equivalent to assign a value $\eta \in \mathbb{R}$ to each observation, where η obeys the independence and monotonicity requirement, but where η is given by $F_\eta^{-1}(v)$ for some distribution function F_η taken without loss of generality to correspond to a distribution with zero mean and unit variance. Such a η solution is transformed into a corresponding v solution by taking $v = F_\eta(\eta)$.

[10]From Koenker (2005) p. 87, Equation (3.12), the dual problem of the (linear) .5 quantile regression of x on Z is:

$$max\ \{x'v|Z^{\top}(v-\frac{1}{2})=0,\ v \in [0,1]^n\}, \tag{5.7}$$

where x is an $(n \times 1)$ vector of dependent variable values and Z is an $(n \times dim(Z))$ matrix of explanatory variable values that includes an intercept. This problem is dual to the more familiar problem of minimizing the average absolute loss function and solving for quantile regression coefficients.

Thus, the basic dual regression optimization problem is equivalently formulated as

$$(\mathbf{P.1}) \quad \max_{\eta} \{x'\eta| \begin{cases} Z^{\top}\eta & = 0 \\ Z^{\top}(\eta^2 - 1) & = 0 \end{cases},$$

where some simplification (particularly in computation) is obtained since η can take on any real value (whereas v is restricted to $[0, 1]$).

The solution to problem (P.1) is found from the Lagrangian

$$\mathscr{L} = \sum_{i=1}^{n} x_i\eta_i - \alpha_1 \sum_{i=1}^{n} z_i\eta_i - \frac{1}{2}\alpha_2 \sum_{i=1}^{n} z_i(\eta_i^2 - 1), \tag{5.9}$$

where $\alpha = (\alpha_1, \alpha_2)'$ is the $2 \times dim(Z)$ vector of Lagrange multipliers. Differentiating with respect to η_i, one obtains n first-order conditions:

$$\frac{\partial \mathscr{L}}{\partial \eta_i} = x_i - \alpha_1 \cdot z_i - (\alpha_2 \cdot z_i)\eta_i = 0, \tag{5.10}$$

which upon solving for η_i delivers $\eta_i = \frac{x_i - \alpha_1 \cdot z_i}{\alpha_2 \cdot z_i}$, where α can be seen to be the parameters of a data generating representation corresponding to (5.5):

$$x_i \quad = \quad \alpha_1 \cdot z_i + (\alpha_2 \cdot z_i)\eta_i. \tag{5.11}$$

These derivations make apparent the dual nature of the parameters α which are the Lagrange multipliers of an optimization problem whose n parameters are $\{\eta_i\}_{i=1}^{n}$. A key contribution of the dual regression approach is to show that constraints on the construction of the stochastic element η have 'shadow values' that are parameters of a data generating representation: a parameter of the DGP is the Lagrange multiplier of a specific constraint on the construction of the stochastic element.

This approach has numerous advantages, in particular in terms of monotonicity. In order to see this point, note that both the dual quantile regression and the basic dual regression programs impose monotonicity by simply correlating x and v, a criterion that suffices to impose monotonicity. However, the dual quantile regression program is dual to a linear program well-known to have solutions at which $dim(Z)$ observations are interpolated when $dim(Z)$ parameters are being estimated - i.e the hyperplanes obtained by regression quantiles must interpolate $dim(Z)$ observations. This is the source of finite-sample quantile crossing issues in well-specified quantile regression models, and is illustrated in the next Section.

Another difference between dual and quantile regression is made apparent by rewriting (5.10) as

$$x_i = \alpha_1 \cdot z_i + (\alpha_2 \cdot z_i)\eta_i \tag{5.12}$$

$$= (\alpha_1 + \alpha_2 F_\eta^{-1}(v_i)) \cdot z_i \tag{5.13}$$

$$\equiv \alpha(v_i) \cdot z_i, \tag{5.14}$$

a standard quantile regression representation, where the functional coefficient $\alpha(v_i)$ is parametrized by the finite-dimensional parameter α. This is in contrast with quantile regression which is semiparametric by construction. Thus, a fundamental property of dual regression is to approximate the functional coefficient $\alpha(v_i)$ by a linear combination of finite dimensional parameters. Besides, dual regression can be fully generalized to account for higher order moments (see Spady and Stouli (2012)), preserving the fact that $\alpha(v_i)$ can be parametrized by including additional terms.

5.3.2. *Estimation of* $F_{Y|XWV}(Y|X,W,V)$. Obtaining the vector of parameters η, respectively α, corresponds to Step 1 of the generic procedure above, and obtaining the parameter vector b, respectively λ, corresponds to Step 2. Given knowledge of η and proceding to the second step of the estimation procedure, there is the following representation for the outcome variable y_i (see Appendix C for details):

$$y_i = \lambda_1 \cdot d_i + (\lambda_2 \cdot d_i)\, b_i, \tag{5.15}$$

where $d_i = (x_i, w_i, x_i\eta_i, w_i\eta_i)$. Representation (5.15) serves as a basis for estimation in the second-stage. The outcome equation (5.15) also admits the following quantile regression representation:

$$y_i = \lambda(u_i) \cdot d_i, \tag{5.16}$$

with $\lambda(u_i) = \lambda_1 + \lambda_2 F_{nb}^{-1}(u_i)$, where u_i is given by

$$u_i = F_{nb}\left(\frac{y_i - \lambda_1 \cdot d_i}{\lambda_2 \cdot d_i}\right), \tag{5.17}$$

with F_{nb} denoting the empirical distribution of b. This is the conditional distribution function of y given x, w and v (for $v = F_\eta(\eta)$), and may be of interest in itself. For instance, it delivers the value of the conditional distribution of y given x, w and $v = .5$, by evaluating $F_{nb}\left(\frac{y-\lambda_1 \cdot d}{\lambda_2 \cdot d}\right)$ at $\eta = F_{n\eta}^{-1}(.5)$. The empirical distribution function given in (5.17) could be used to compute the QSF of Imbens and Newey (2009) which is a functional of u, although this is not pursued here.

Similarly to representation (5.11) obtained for x as the FOC of problem (P.1), representation (5.15) results from the first-order conditions of the following dual regression optimization problem:

$$(\mathbf{P.2}) \quad max_b \; \{y'b| \begin{cases} D^\top b & = 0 \\ D^\top (b^2 - 1) & = 0 \end{cases},$$

where D is the $(n \times (2 \times (1 + dim(W))))$ matrix of values of covariates in Step 2 of the estimation procedure.

An alternative to the two-step procedure is to add the FOCs of the lower level optimization problem (P.1) as a constraint to program (P.2) and obtain the following MPEC formulation (Su and Judd (2012)):

$$(\mathbf{P}) \quad max_{(b,\alpha)} \; \{y'b| \begin{cases} D^\top b & = 0 \\ D^\top (b^2 - 1) & = 0 \\ x_i - \alpha_1 \cdot z_i - (\alpha_2 \cdot z_i)\eta_i & = 0 \end{cases} \cdot$$

5.3.3. *Asymptotic properties.* The asymptotic distribution of the vector of parameters $\theta = (\alpha, \lambda)'$ obtains upon substituting $v(x_i, z_i; \alpha) \equiv \frac{x_i - \alpha_1 \cdot z_i}{\alpha_2 \cdot z_i}$ for η_i, letting $d_{vi} = (x_i, w_i, x_i v(x_i, z_i; \alpha), w_i v(x_i, z_i; \alpha))$ and noting that the following $2 \times (dim(D) + dim(Z))$ vector of moments is available

$$g(y, x, z, \theta) = \left(g_1(y, x, w, z, \theta)', g_2(y, x, w, z, \theta)', g_3(y, x, w, z, \theta)', g_4(y, x, w, z, \theta)'\right)',$$

where

$$g_1(y, x, w, z, \theta) = d_v \left(\frac{y - \lambda_1 \cdot d_v}{\lambda_2 \cdot d_v}\right) \tag{5.18}$$

$$g_2(y, x, w, z, \theta) = d_v \left[\left(\frac{y - \lambda_1 \cdot d_v}{\lambda_2 \cdot d_v}\right)^2 - 1\right] \tag{5.19}$$

$$g_3(y, x, w, z, \theta) = z \left(\frac{x - \alpha_1 \cdot z}{\alpha_2 \cdot z}\right) \tag{5.20}$$

$$g_4(y, x, w, z, \theta) = z \left[\left(\frac{x - \alpha_1 \cdot z}{\alpha_2 \cdot z}\right)^2 - 1\right]. \tag{5.21}$$

Therefore, for the purpose of asymptotic analysis, the 2-step estimator of $Q_{Y|XWV}$ can be equivalently viewed as a stacked Method of Moments estimator with moments $g(y, x, w, z, \theta)$, solving the system

$$\frac{1}{n}\sum_{i=1}^{n} g(y_i, x_i, w_i, z_i, \hat{\theta}) = 0. \tag{5.22}$$

Under the assumption that the model is well specified and data is i.i.d, define G_o and S_o as

$$G_o = E\left[\left.\frac{\partial g}{\partial \theta'}\right|_{\theta_o}\right] \quad \text{and} \quad S_o = E\left[gg'|_{\theta_o}\right]. \tag{5.23}$$

Applying standard results for the Method of Moments (Newey and McFadden (1994)), there is

$$\sqrt{n}\left(\hat{\theta} - \theta_o\right) \xrightarrow{d} \mathcal{N}\left(0, G_o^{-1} S_o \left(G_o^{-1}\right)'\right). \tag{5.24}$$

A detailed characterization of the asymptotic variance-covariance matrix is given in Appendix D. This result shows that inference methods for dual regression estimates can be constructed from standard methods available for Method of Moments estimators. In particular, bootstrap methods can easily be applied to obtain standard errors.

5.4. **Structural Function.** Turning to the third step of the estimation procedure, for $v^* = .5$, plugging the expression for ε_i^* given in (5.6) in (5.4), there is

$$y_i = \beta_{10}^* + \beta_{11}^* x_i + \beta_{12}^* \cdot w_i + (\beta_{20}^* + \beta_{21}^* x_i + \beta_{22}^* \cdot w_i)\left[\gamma_{12}^* \eta_i + (1 + \gamma_{22}^* \eta_i) b_i\right]. \tag{5.25}$$

Given estimates of $\{\hat{\eta}_i, \hat{b}_i\}_{i=1}^n$ the $2 \times (dim(W) + 1) + 2$ vector of parameters $(\beta^*, \gamma)'$ obtains by solving

$$\min_{(\beta^*,\gamma^*)} \sum_{i=1}^{n} \left\{y_i - \left(\beta_{11}^* x_i + \beta_{12}^* w_i + (\beta_{21}^* x_i + \beta_{22}^* w_i)\left[\gamma_{12}^* + (1 + \gamma_{22}^* \hat{\eta}_i)\hat{b}_i\right]\right)\right\}^2. \tag{5.26}$$

Once the vector parameter β^* has been recovered and after rearranging terms, for $\tau \in (0, 1)$ the structural function is given by

$$\left(\beta_{11}^* + \beta_{21}^* F_{n\varepsilon^*}^{-1}(\tau)\right) x_i + \left(\beta_{12}^* + \beta_{22}^* F_{n\varepsilon^*}^{-1}(\tau)\right) \cdot w_i = \beta^*(\tau) \cdot d_i^*,$$

where $F_{n\varepsilon^*}$ denotes the empirical distribution of ε^*, $d_i^* = (x_i, w_i)$ and $\beta^*(\tau) = (\beta_{11}^* + \beta_{21}^* F_{n\varepsilon^*}^{-1}(\tau), (\beta_{12}^* + \beta_{22}^* F_{n\varepsilon^*}^{-1}(\tau))')'$ is the vector of quantile regression coefficients for $V = v^*$. Thus, $\beta^*(\tau) \cdot d^*$ is the structural conditional τ-quantile function of Y given X, W and $V = .5$.

5.5. **Local Quantile Structural Function Algorithm For Triangular Models.** The previous steps are summarized in the following algorithm for $v^* = .5$.

(1) *Estimate the control variable* $\{\hat{\eta}_i\}_{i=1}^n$ *by solving the dual regression optimization problem* (P.1)*.*

(2) *Estimate the conditional distribution function* $\{\hat{b}_i\}_{i=1}^n$ *by solving the dual regression optimization problem* (P.2)*, then for* $\hat{V} = F_{n\eta}(\hat{\eta})$*,* $\hat{Q}_{Y|XWV}(u|x, w, \hat{V})$ *is directly available.*

(3) For $v^* = .5$*, estimate simultaneously* $\hat{\beta}^*$ *and* $\hat{\gamma}$ *by solving*

$$\min_{(\beta^*,\gamma^*)} \sum_{i=1}^n \left\{ y_i - \left(\beta_{11}^* x_i + \beta_{12}^* w_i + (\beta_{21}^* x_i + \beta_{22}^* w_i) \left[\gamma_{12}^* + (1 + \gamma_{22}^* \hat{\eta}_i) \hat{b}_i \right] \right) \right\}^2. \tag{5.27}$$

An alternative is to solve directly the dual regression optimization problem (P) and proceed to Step 2. Indeed, Program (P) provides a one-step estimation procedure for $\hat{F}_{Y|XWV}(y|x, w, \hat{V})$. It is possible to formulate a one-step optimization problem for the structural function as well, which is left for future work. Also, note that taking transformations, such as splines or trigonometric basis functions, of x_i, w_i and $\hat{\eta}_i$ in (5.27) yields a very flexible yet particulary simple implementation of the approximation step.

6. **Empirical Application : Demand Estimation**

This Section illustrates how the estimation methodology can be applied to estimation of gasoline demand functions, where the price is treated as an endogenous regressor. This empirical application has been studied in Blundell, Horowitz, and Parey (2012), where they obtain well-behaved demand functions under appropriate shape constraints. They argue that non-decreasing estimated demand functions are merely an artifact of inappropriate estimation procedures and that, when imposing relevant constraints such as the Slutsky property, well-behaved demand estimates obtain. This view is reconsidered in the light of correcting for endogeneity.

All computational procedures are implemented in the software R (R-Development-Core-Team (2007)). Dual regression is implemented using Ipopt (Interior Point Optimizer), an open source software package for large-scale nonlinear optimization (Waechter and Biegler (2006)), and its R interface (Ypma (2011)). The first-stage quantile regression estimates shown in Figure 2 are obtained using the R package Quantreg (Koenker (2007)).

6.1. **Data and Empirical Specification.** The dataset is as in Blundell, Horowitz, and Parey (2012), where a detailed description can be found. The data are taken from the 2001 National Household Travel Survey (NHTS), conducted between March 2001 and May 2002.

The survey collects information on household characteristics, each household vehicle and on trips made during this time period. The households are from all geographic areas in the US. The total number of observations selected from the dataset is 4812 observations.

Here the main variables of interest are the outcome variable, Y, which will be annual gasoline consumption, gasoline prices per gallon, X, household income, R, and the instrumental variable, Z, the distance to the gulf of Mexico. Gasoline consumption is derived from odometer readings and estimates of the vehicle fuel economy (miles per gallon), and is aggregated over different vehicles owned by the household. Recorded prices are a weighted average of monthly prices, including taxes, in dollars per gallon, in the county where the household is located.

Price differences across local markets reflect proximity of supply, short-run shocks to supply, competition in the local market, and local differences in taxes and environmental programs. It is assumed that these factors are summarized through a scalar index captured by the control variable. Households report their annual income, before taxes, in 18 different ranges, and households income is set to the midpoint of the respective interval and households reporting an annual income over \$100,000 are assigned an income of \$120,000. The instrument is taken to be a distance measure (in 1000 km) from the source of supply in the Gulf of Mexico to the capital of the state in which the households is located. The starting point is a major oil platform located in the Gulf of Mexico.

The following empirical specification is considered for the structural equation:

$$y_i = \beta^*_{11}\tilde{x}_i + \beta^*_{12}r_i + \beta^*_{13}\tilde{x}_i r_i + (\beta^*_{21}\tilde{x}_i + \beta^*_{22}r_i + \beta^*_{23}\tilde{x}_i r_i)\varepsilon^*_i, \tag{6.1}$$

where y_i and r_i denote log gasoline demand and log income and $\tilde{x}_i$ is a vector of transformations of log prices, including an intercept. The transformations are taken to be cubic B-splines with 4 equally spaced knots. This augmented partially linear specification is flexible and allows for interactions between prices and income. Blundell, Horowitz, and Parey (2012) find that acccounting for dependence of the price elasticity on income matters in this dataset. The estimated structural parameter vector β^* is obtained conditionally on $V = .5$. Estimated demand functions under an alternative specification including demographic characteristics are shown in Appendix E. Following the description of the triangular location scale model given in Section 5.2, in the second step of the local QSF algorithm specification (5.15) serves as a basis for estimation, letting $d_i = (\tilde{x}_i, \tilde{x}_i r_i, \tilde{x}_i \eta_i, r_i \eta_i, \tilde{x}_i r_i \eta_i)$, for a total of $2 \times (4 \times dim(\tilde{X}) + 1)$ Lagrange multipliers λ obtained from solving for the 4812 parameters $\{b_i\}_{i=1}^n$ in problem (P2).

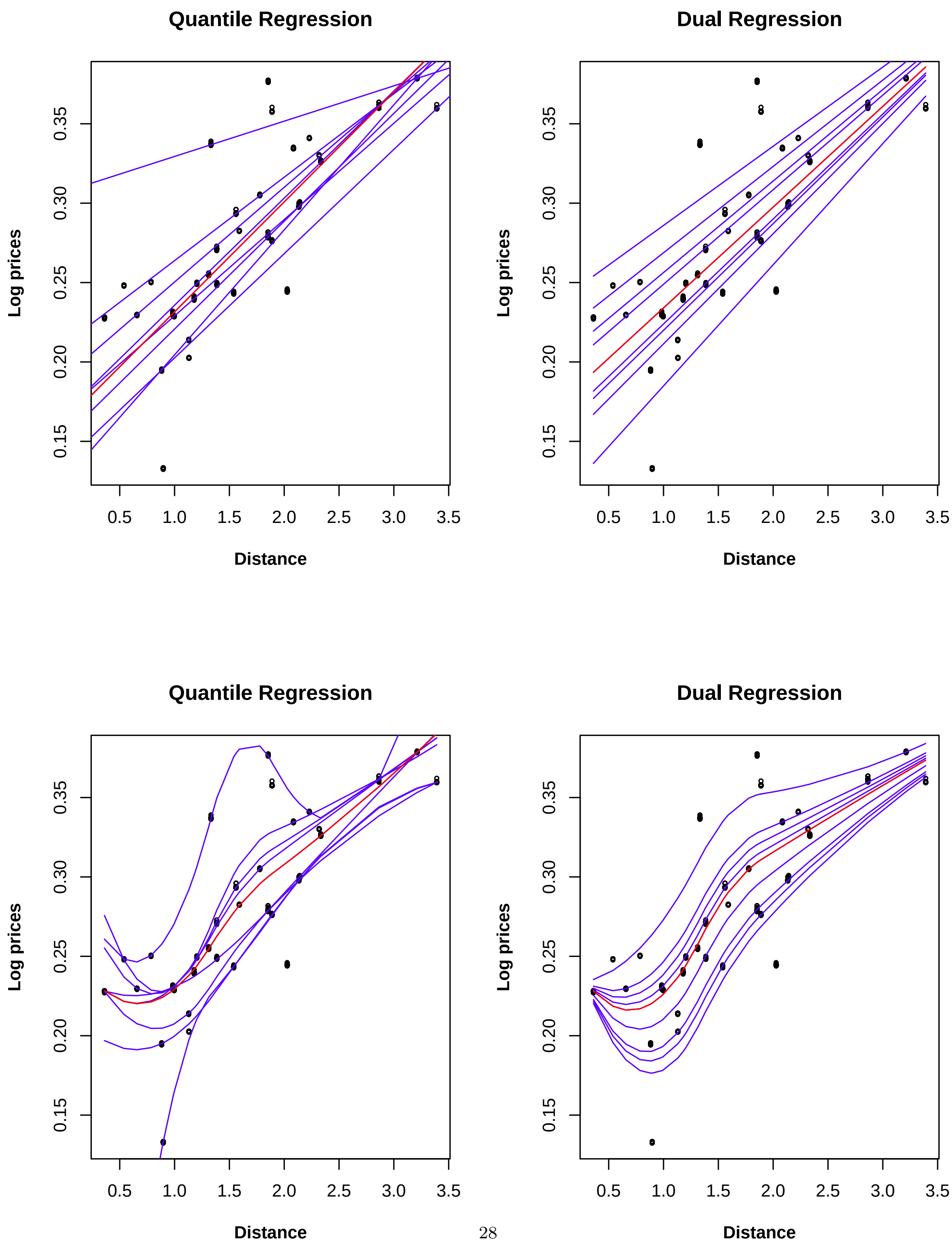


FIGURE 2. First stage: Linear (top) and nonlinear (bottom) quantile (left) and dual (right) regression of log prices on distance to the Gulf of Mexico. Conditional median (red) and conditional $\{\cdot 10, \cdot 15, \ldots, \cdot 90\}$ quantile functions (blue).

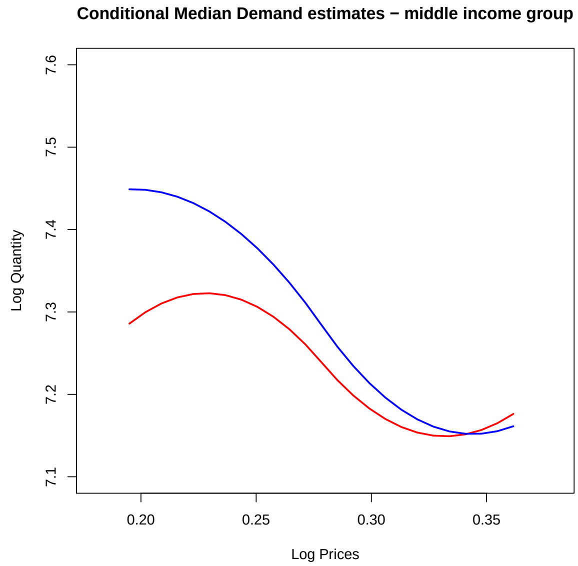


FIGURE 3. Statistical and structural conditional median demand for gasoline - Conditional on $V = .5$.

6.2. **First stage.** The distance measure is taken to be the distance to the respective state capital. Thus, an interesting feature of the data, is that the support of the instrument is effectively discrete with 15 points of support. In order to illustrate dual regression and its strengths, it is informative to give a close look at the relationship between gasoline price and the instrument. Fig. 2 shows estimates of conditional quantile functions of log prices of gasoline given the instrument, with both linear and nonlinear specifications, obtained by quantile and dual regression. It is striking that, given the nature of the data, the interpolative property of quantile regression in finite-samples is a drawback in the uncovering of the relationship of interest. Indeed, conditional quantile functions are seen to cross at several points in the quantile regression figures, whereas dual regression delivers well behaved estimates, not subject to crossing. The choice of a regression method is especially important when the control variable used in order to correct for endogeneity is the entire conditional distribution function of X given Z and not simply the first stage residuals from a mean regression as in more conventional control variable procedures.

6.3. **Estimation of demand functions.** Variations in price responsiveness of gasoline demand is investigated across three dimensions: (1) Across the income distribution, where the focus is on three income levels: a middle-income group at \$57,500, which corresponds to median income in the sample; a low-income group at \$42,500, which corresponds to the first quartile in the sample; a high-income group at \$72,500, which corresponds to the 59.-63.3th percentile in the sample[11]. (2) Across the price distribution, by using a flexible specification for the price variable. (3) Across unobserved heterogeneity in demand elasticity.

Figure 3 shows statistical and structural (i.e corrected for endogeneity) median demand curves conditional on $V = .5$. It is already apparent that correcting for endogeneity yields a better behaved demand curve suggesting that the unconstrained non-downward-sloping demand curves obtained by Blundell, Horowitz, and Parey (2012) are not merely an artefact of inappropriate estimation procedures but may also be a consequence of misspecification and bias originating in not accounting for endogeneity. Fig. 3 is in line with results obtained by Hausman and Newey (2012) who also find that correcting for endogeneity delivers downward-sloping demand estimates. These results raise questions about the virtue of imposing shape constraints coming from economic theory on data.

Figure 4 illustrates the results and plots the estimated conditional quantile functions of log demand given log prices. Fig. 4 plots the original and structural estimates of the quantile demand surfaces. The figure summarizes the structural relationship between price and demand for a grid of probability levels in the interval $[.3, .7]$, conditionally on $V = .5$. Estimates satisfy some basic smoothness requirements across the entire conditional quantile process. It is important to note that this feature does not typically characterize estimates of the conditional quantile process by quantile regression methods, as conditional quantile functions are then estimated sequentially and independently of each other.

Figure 4 shows that in this example, dual regression delivers monotone and well-behaved estimates, suggesting that imposing restrictions of the model and exploiting the specific structure of the stochastic element already delivers estimates that mostly do not exhibit upward sloping areas. This is apparent from the structural demand curves obtained after correcting for endogeneity which are largely free of upward sloping regions, without imposing further shape constraints, except for upper quantiles demand curves for the middle- and low- income groups. Again, this may rather be interpreted as a call for better specification of the model than for shape constraints. Correcting for endogeneity also yields demand curves with steeper slopes.

[11] The definition of income groups is as in Blundell, Horowitz, and Parey (2012).

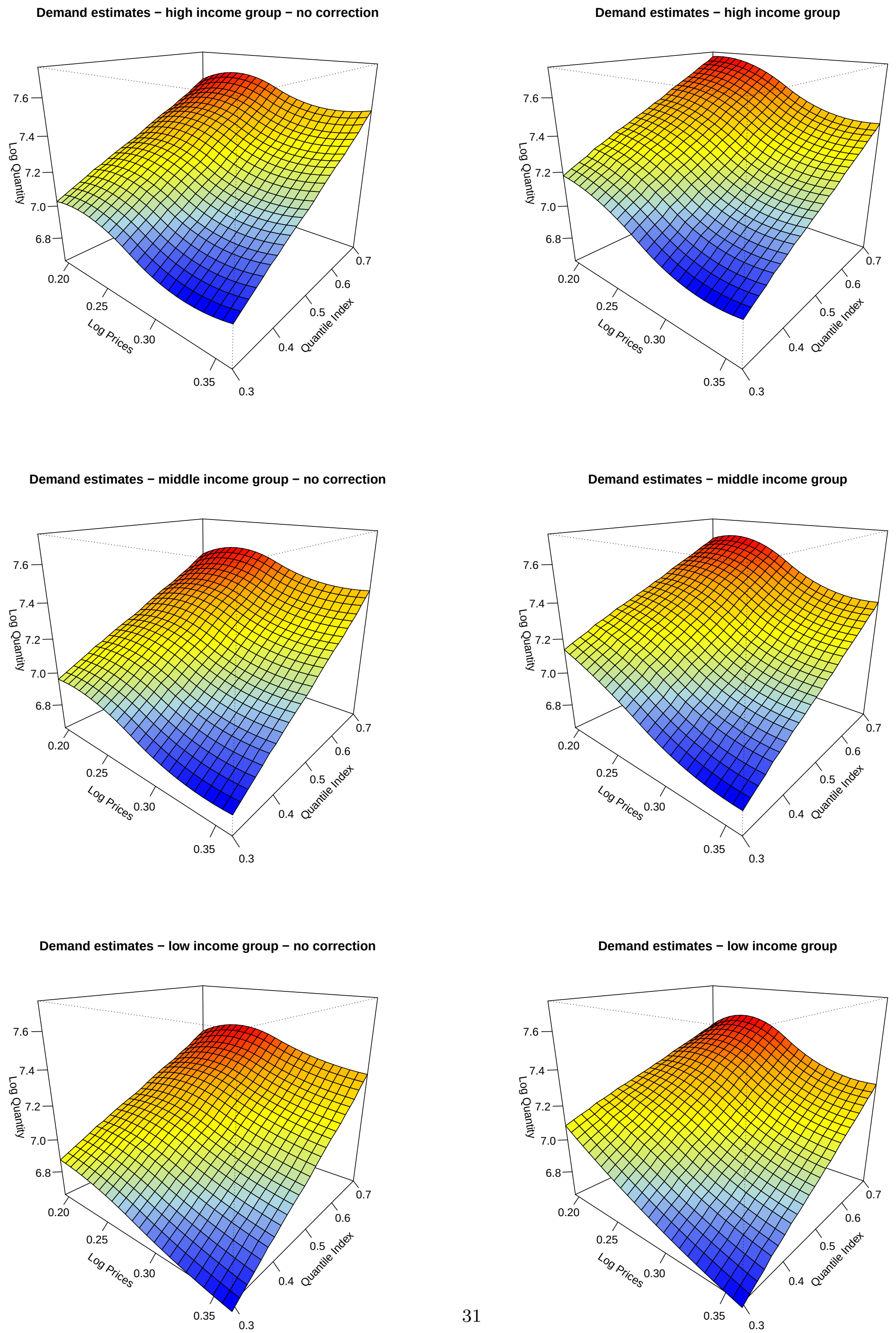


FIGURE 4. Statistical (Left) and structural (Right) demand functions by income groups - Conditional on $V = .5$.

7. **Concluding Remarks**

This paper proposes a new treatment of identification and estimation of structural functions in conditional independence models featuring a scalar disturbance. It is shown that information (i.e the identifying power of the model) is located at both a local level (information specific to the relationship between instruments and the endogenous regressor) and at the global level (where all restrictions of the model are exploited). After a reanalysis of local identification in nonseparable triangular models, local overidentification in the presence of nonmonotone instruments and the connection between Chesher (2003) and Imbens and Newey (2009) are established. At the global level, identification of the structural function conditional on a value of the control variable V as well as unconditionally are obtained by constructing a disturbance endowed with a specific structure determined by restrictions of the model, leading to global identification under very weak conditions.

A generic analog estimation strategy is proposed and an estimator for a semiparametric triangular location-scale model based on a new regression method is introduced. The applicability of the method and the estimator are illustrated by estimation of demand for gasoline in the United States using the dataset of Blundell, Horowitz, and Parey (2012). The empirical application shows that once corrected for endogeneity of prices estimated demand curves are mostly downward-sloping.

Of particular interest for future work is the connection between results in this paper and the latent variable modelling/measurement error literature. Also, revisiting identification in nonseparable triangular models with discrete X (see Chesher (2005); Jun, Pinkse, and Xu (2011) for recent work on this topic) in the light of the dual construction introduced appears as a natural next step.

The relevance of this contribution is also related to the fact that several results can be exploited to build specification tests in triangular models. The results in this paper also raise several questions regarding inference and efficient estimation in nonseparable models with endogenous regressors. Treatment of weakly identifying instrumental values constitutes an interesting research avenue as well in order to optimally combine the relative power of various instrumental values when the structural function is locally overidentified.

Appendix A. **Proofs for Section 2**

A.1. **Proof of Proposition 1.**

Proof. The result follows from the argument in the text. □

A.2. **Proof of Proposition 2.**

Proof. The result follows from the argument in the text. □

Appendix B. **Proofs for Section 3**

B.1. **Preliminaries.** Before proceding to the proof of Theorem 1, there is the following useful lemma.

Lemma 1. *For $v^* \in \mathcal{V}$ specified in Normalization 1 and ε^{ψ^*} as in Definition 2, $\varepsilon^{\psi^*} \perp\!\!\!\perp X|V = v^*$ implies that ε^{ψ^*} is uniformly distributed on $[0,1]$ conditional on $V = v^*$.*

Proof. The result follows upon noting that for $V = v^*$ and for all $\tau \in (0,1)$, $\varepsilon^{\psi^*} \perp\!\!\!\perp X|V = v^*$ implies that

$$F_{\varepsilon^{\psi^*}|V}(\tau|v^*) = F_{\varepsilon^{\psi^*}|XV}(\tau|x, v^*) = \tau, \tag{B.1}$$

since replacing ε^{ψ^*} by its definition $F_{\varepsilon^{\psi^*}|XV}(\tau|x, v^*)$ gives

$$\begin{aligned} P[\psi^*(F_{Y|XV}(Y|X, v^*), v^*) \leq \tau|X = x, V = v^*] &= P[F_{Y|XV}(Y|X, v^*) \leq \psi^{*-1}(\tau, v^*)|X = x, V = v^*] \\ &= P[F_{Y|XV}(Y|X, v^*) \leq \tau|X = x, V = v^*] && \text{(B.2)} \\ &= P[Y \leq Q_{Y|XV}(\tau|x, v^*)|X = x, V = v^*] && \text{(B.3)} \\ &= \tau, && \text{(B.4)} \end{aligned}$$

where the first equality is by strict monotonicity of $\psi^*(\tau, v^*)$ in τ, the second one by Definition 1 and the third by strict monotonicity of $F_{Y|XV}(y|x, v^*)$ in y for all $x \in \mathcal{X}$. □

B.2. **Proof of Theorem 1.**

Proof. **Step 1:** $\varepsilon^{\psi^*} \perp\!\!\!\perp X|V \;\Rightarrow$ (4.5) and $\psi^* = Q_{\varepsilon^*|V}$ and $h^* = H^*$.
Step 1A: $\varepsilon^{\psi^*} \perp\!\!\!\perp X|V \;\Rightarrow\; h^* = H^*$.
Local application of the conditional independence condition $\varepsilon^{\psi^*} \perp\!\!\!\perp X|V$ and $\psi^* \in \Psi^*$ imply that conditional on $V = v^*$, ε^{ψ^*} is endowed with the three properties of a random variable generated by the conditional distribution function of Y given X and $V = v^*$: conditional on $V = v^*$, ε^{ψ^*} is (i) independent of X, (ii) uniformly distributed on $[0,1]$ by Lemma 1, and (iii) the composite function $y \mapsto \psi^*(F_{Y|XV}(y|x, v^*), v^*)$ is strictly increasing in y for all $x \in \mathcal{X}$, since the composition of strictly increasing functions is strictly increasing. Therefore, for all $(y, x) \in \mathcal{Y} \times \mathcal{X}$,

$$\psi^*(F_{Y|XV}(y|x, v^*), v^*) = F_{Y|XV}(y|x, v^*). \tag{B.5}$$

Upon inverting $F_{Y|XV}(y|x, v^*)$ with respect to y in (B.5), there is

$$y = Q_{Y|XV}(\psi^*(F_{Y|XV}(y|x, v^*), v^*)|x, v^*) \equiv h^*(x, \psi^*(F_{Y|XV}(y|x, v^*), v^*)), \tag{B.6}$$

and there exists a function $h^*(x, \tau)$ strictly increasing in τ such that (4.5) holds at $V = v^*$. For $\mathcal{Y}^*$ denoting the conditional support of Y given X and v^*, the function $h^*(x, \tau)$ is defined as the restriction of the conditional quantile function of Y given X and V to $[0, 1] \times \mathcal{X} \times \{v^*\}$, i.e $h^* \equiv Q_{Y|XV}\big|_{[0,1]\times\mathcal{X}\times\{v^*\}} : [0, 1] \times \mathcal{X} \times \{v^*\} \to \mathcal{Y}^* = \mathcal{Y}$ by Condition 3(i). Since $Q_{Y|XV}(\tau|x, v^*)$ is strictly increasing in τ for all $x \in \mathcal{X}$, $h^*(x, \tau)$ is strictly increasing in τ for all $x \in \mathcal{X}$.
From (B.6), since h^* is strictly increasing in τ, its inverse $h^{*-1}(x, y)$ is well defined, and so there is

$$\psi^*(F_{Y|XV}(y|x, v^*), v^*) = h^{*-1}(x, y), \tag{B.7}$$

which by (B.5) yields

$$F_{Y|XV}(y|x, v^*) = h^{*-1}(x, y). \tag{B.8}$$

By Condition 2 and Normalization 1 there also is

$$F_{Y|XV}(y|x, v^*) = F_{\varepsilon^*|V}(H^{*-1}(x, y)|v^*) = H^{*-1}(x, y). \tag{B.9}$$

Therefore, from (B.8) and (B.9) and Condition 3(ii), conclude that $h^{*-1}(x, y) = H^{*-1}(x, y)$ for all $(y, x) \in \mathcal{Y} \times \mathcal{X}$.
Step 1B: $\varepsilon^{\psi^*} \perp\!\!\!\perp X|V$ and $h^* = H^* \ \Rightarrow\ $ (4.5) and $\psi^* = Q_{\varepsilon^*|V}$.
$\varepsilon^{\psi^*} \perp\!\!\!\perp X|V$ and $\psi^* \in \Psi^*$ imply that the random variable $U^{\psi^*} \equiv F_{\varepsilon^{\psi^*}|XV}(H^{*-1}(X, Y)|X, V)$ is endowed with the three properties of a random variable generated by the conditional distribution function of Y given X and V: U^{ψ^*} is (i) independent of X and V, (ii) uniformly distributed on $[0, 1]$ by construction, and (iii) the composite function $y \mapsto F_{\varepsilon^{\psi^*}|XV}(H^{*-1}(x, y)|x, v)$ is strictly increasing in y for all $x \in \mathcal{X}$ and $v \in \mathcal{V}$, since the composition of strictly increasing functions is strictly increasing. It follows that

$$F_{\varepsilon^{\psi^*}|XV}(H^{*-1}(x, y)|x, v) = F_{Y|XV}(y|x, v) = F_{\varepsilon^*|V}(H^{*-1}(x, y)|v), \tag{B.10}$$

where the second equality follows from Condition 2.
By Definition 2 of ε^{ψ^*}, $F_{\varepsilon^{\psi^*}|XV}(H^{*-1}(x, y)|x, v)$ is equal to

$$\begin{aligned}
P[\psi^*(F_{Y|XV}(Y|X, V), V) \le H^{*-1}(x, y)|X = x, V = v] &= P[F_{Y|XV}(Y|X, V) \le \psi^{*-1}(H^{*-1}(x, y), v)|X = x, V = v] \\
\text{(B.11)} \qquad &= P[Y \le Q_{Y|XV}(\psi^{*-1}(H^{*-1}(x, y), v)|x, v)|X = x, V = v] \\
\text{(B.12)} \qquad &= \psi^{*-1}(H^{*-1}(x, y), v),
\end{aligned}$$

where the first equality follows by the strict monotonicity property stated in Definition 1(i) and the conditioning on $V = v$, the second one by strict monotonicity of $F_{Y|XV}(y|x,v)$ in y, and the third one by $Q_{Y|XV}(\cdot|x,v)$ being the inverse function of $F_{Y|XV}(\cdot|x,v)$. Therefore, combining (B.10) and (B.12),

$$F_{\varepsilon^*|V}(H^{*-1}(x,y)|v) = \psi^{*-1}(H^{*-1}(x,y),v), \tag{B.13}$$

for all $(y,x,v) \in \mathcal{Y} \times \mathcal{X} \times \mathcal{V}$. By Condition 3(i), for each $v \in \mathcal{V}$, there is a value $x(v)$ of X such that

$$supp(H^{*-1}(x(v),Y)|X = x(v), V = v) = (0,1), \tag{B.14}$$

since Y has full support conditional on $X = x(v)$ and $V = v$. Therefore $F_{\varepsilon^*|V}(\tau|v) = \psi^{*-1}(\tau,v)$ for all τ and v in $(0,1) \times \mathcal{V}$, follows from (B.13) and (B.14), and $\psi^*(u,v) = Q_{\varepsilon^*|V}(u|v)$ for all u and v in $(0,1) \times \mathcal{V}$ as claimed.
Last, by Definition 2 of ε^{ψ^*} and the structural representations of Y given in Normalization 1 and Equation (4.1), for $\psi^* = Q_{\varepsilon^*|V}$ and $h^* = H^*$, there is

$$\varepsilon^{\psi^*} = \varepsilon^* = H^{*-1}(X,Y) = h^{*-1}(X,Y). \tag{B.15}$$

Therefore $y = h^*(x, \psi^*(F_{Y|XV}(y|x,v),v))$ for all $(y,x,v) \in \mathcal{Y} \times \mathcal{X} \times \mathcal{V}$ and the result follows.
Step 2: (4.5) and $\psi^* = Q_{\varepsilon^*|V}$ and $h^* = H^* \;\Rightarrow\; \varepsilon^{\psi^*} \perp\!\!\!\perp X|V$.
From the equivalent structural representation of Y given in (4.1) and Definition 2 of ε^{ψ^*}, for $\psi^* = Q_{\varepsilon^*|V}$ and $h^* = H^*$, there is

$$\varepsilon^{\psi^*} = Q_{\varepsilon^*|V}(F_{Y|XV}(Y|X,V),V) = H^{*-1}(X,Y) = \varepsilon^*, \tag{B.16}$$

and the result follows by Condition 2. □

B.3. Proof of Corollary 1.

Proof. Evaluating $y = h^*(x, \psi^*(F_{Y|XV}(y|x,v),v))$ at $V = v^*$, it follows from the steps in the proof of Theorem 1 that $\psi^*(u,v) = Q_{\varepsilon^*|V}(u|v)$ for all $(u,v) \in (0,1)\times\mathcal{V}$ and $h^*(x,\tau) = H^*(x,\tau)$ for all $x \in \mathcal{X}$ and $\tau \in (0,1)$. The converse is obvious. □

Appendix C. Illustration: Identification in a semiparametric location scale model

C.1. The Model.

The DGP considered is given by

$$y_i = \beta_{11}x_i + \beta_{12} \cdot w_i + (\beta_{21}x_i + \beta_{22} \cdot w_i)\varepsilon_i \tag{C.1}$$

$$x_i = \alpha_1 \cdot z_i + (\alpha_2 \cdot z_i)\eta_i \tag{C.2}$$

$$\varepsilon_i = \gamma_{11} + \gamma_{12}\eta_i + (\gamma_{21} + \gamma_{22}\eta_i)b_i, \tag{C.3}$$

where x_i is the endogenous regressor, w_i is a vector of additional regressors that includes an intercept and z_i is a vector of instrumental variables including an intercept, and potentially including additional covariates. η and b are disturbances normalized to have mean 0 and variance 1.

In this model the control function is given by $\eta_i = \frac{x_i - \alpha_1 \cdot z_i}{\alpha_2 \cdot z_i}$ where the α_1 and α_2 are identified from the moments conditions

$$\begin{aligned} Z^\top \left\{ \frac{x - \alpha_1 \cdot z}{\alpha_2 \cdot z} \right\} &= 0 \\ Z^\top \left[\left\{ \frac{x - \alpha_1 \cdot z}{\alpha_2 \cdot z} \right\}^2 - 1 \right] &= 0. \end{aligned} \tag{C.4}$$

Given knowledge of η, plugging the expression for ε_i given in (C.3) in (C.1), and rearranging terms, there is the following representation of the model which serves as a basis for estimation of $F_{Y|XWV}(Y|X,W,V)$:

$$y_i = \lambda_1 \cdot d_i + (\lambda_2 \cdot d_i)\, b_i \tag{C.5}$$

$$x_i = \alpha_1 \cdot z_i + (\alpha_2 \cdot z_i)\, \eta_i, \tag{C.6}$$

where $d_i = (x_i, w_i, x_i\eta_i, w_i\eta_i)$, and where the λ's are identified from the moment conditions

$$\begin{aligned} D^\top \left\{ \frac{y - \lambda_1 \cdot d}{\lambda_2 \cdot d} \right\} &= 0 \\ D^{*\top} \left[\left\{ \frac{y - \lambda_1 \cdot d}{\lambda_2 \cdot d} \right\}^2 - 1 \right] &= 0. \end{aligned} \tag{C.7}$$

C.2. **Discussion of the normalization condition: $v^* = .5$.** In a location-scale model, for $v^* = .5$, $\varepsilon^* = \frac{\varepsilon_i - \gamma_{11}}{\gamma_{21}}$, i.e

$$\varepsilon_i^* = \frac{\varepsilon_i - \gamma_{11}}{\gamma_{21}} \tag{C.8}$$

$$= \frac{\gamma_{12}}{\gamma_{21}}\eta_i + (1 + \frac{\gamma_{22}}{\gamma_{21}}\eta_i)b_i \tag{C.9}$$

$$\equiv \gamma_{11}^* + \gamma_{12}^*\eta_i + (\gamma_{21}^* + \gamma_{22}^*\eta_i)b_i, \tag{C.10}$$

with $\gamma_{11}^* = 0$, $\gamma_{12}^* = \frac{\gamma_{12}}{\gamma_{21}}$, $\gamma_{21}^* = 1$, $\gamma_{22}^* = \frac{\gamma_{22}}{\gamma_{21}}$. Since at $\eta = 0$

$$\varepsilon_i^* = \frac{\varepsilon_i - \gamma_{11}}{\gamma_{21}} = b_i, \tag{C.11}$$

the distributional normalization is simply that ε has mean 0 and variance 1 conditional on $\eta = 0$.

C.3. **Main steps.** From (C.11), there is

$$\varepsilon_i = \gamma_{11} + \gamma_{21}\varepsilon_i^* \tag{C.12}$$

Plugging the expression for ε_i given in (C.12) in (C.1), there is

$$\begin{aligned} y_i &= \beta_{11}x_i + \beta_{12}\cdot w_i + (\beta_{21}^* x_i + \beta_{22}\cdot w_i)\left[\gamma_{11} + \gamma_{21}\varepsilon_i^*\right] && \text{(C.13)} \\ &= (\beta_{11} + \beta_{21}\gamma_{11})x_i + (\beta_{12} + \beta_{22}\gamma_{11})\cdot w_i + ((\gamma_{21}\beta_{21})x_i + (\gamma_{21}\beta_{22})\cdot w_i)\varepsilon_i^* && \text{(C.14)} \\ &= \beta_{11}^* x_i + \beta_{12}^*\cdot w_i + (\beta_{21}^* x_i + \beta_{22}^*\cdot w_i)\varepsilon_i^*, && \text{(C.15)} \end{aligned}$$

with

$$\begin{aligned} \beta_{11}^* &= \beta_{11} + \beta_{21}\gamma_{11} && \text{(C.16)} \\ \beta_{12}^* &= \beta_{12} + \beta_{22}\gamma_{11} && \text{(C.17)} \\ \beta_{21}^* &= \gamma_{21}\beta_{21} && \text{(C.18)} \\ \beta_{22}^* &= \gamma_{21}\beta_{22}. && \text{(C.19)} \end{aligned}$$

Plugging the expression (C.10) for ε_i^* in (C.20), there is

$$y_i = \beta_{11}^* x_i + \beta_{12}^*\cdot w_i + (\beta_{21}^* x_i + \beta_{22}^*\cdot w_i)\left[\gamma_{12}^*\eta_i + (1+\gamma_{22}^*\eta_i)b_i\right]. \tag{C.20}$$

Given knowledge of $\{(\eta_i, b_i)\}_{i=1}^n$, the structural function is obtained by noting that one can solve for the $2\times(dim(W)+1)+2$ vector of parameters $(\beta^*, \gamma)'$ from the system of n equations

$$y_i - \left(\beta_{11}^* x_i + \beta_{12}^*\cdot w_i + (\beta_{21}^* x_i + \beta_{22}^*\cdot w_i)\left[\gamma_{12}^*\eta_i + (1+\gamma_{22}^*\eta_i)b_i\right]\right) = 0. \tag{C.21}$$

For observations i such that $\eta_i = 0$, (C.21) reduces to

$$\varepsilon_i^* = \frac{y_i - (\beta_{11}^* x_i + \beta_{12}^*\cdot w_i)}{\beta_{21}^* x_i + \beta_{22}^*\cdot w_i} = b_i, \tag{C.22}$$

so ε^* is known, and the $2\times dim(W)+2$ vector of structural parameters β^* is identified from the $2\times dim(W)+2$ moment conditions

$$\begin{aligned} D^{*\top}\left\{\frac{y-\beta_1^*\cdot d^*}{\beta_2^*\cdot d^*}\right\} &= 0 \\ D^{*\top}\left[\left\{\frac{y-\beta_1^*\cdot d^*}{\beta_2^*\cdot d^*}\right\}^2 - 1\right] &= 0, && \text{(C.23)} \end{aligned}$$

where $d_i^* = (x_i, w_i)$.

From knowledge of the betas

$$\beta_{11}^* x_i + \beta_{12}^*\cdot w_i + (\beta_{21}^* x_i + \beta_{22}^*\cdot w_i)F_{n\varepsilon^*}^{-1}(\tau) \equiv \beta(\tau)\cdot d_i^*,$$

and $\beta^*(\tau)\cdot d^*$ is the structural conditional τ-quantile function of Y given X and W and $V = .5$.

C.4. **Discussion of the normalization condition: general case.** ε^* is given by:

$$\varepsilon_i^* = \frac{\varepsilon_i - \left(\gamma_{11} + \gamma_{12} F_{n\eta}^{-1}(v^*)\right)}{\left(\gamma_{21} + \gamma_{22} F_{n\eta}^{-1}(v^*)\right)} \tag{C.24}$$

$$= \frac{(\gamma_{11} + \gamma_{12}\eta_i + (\gamma_{21} + \gamma_{22}\eta_i)b_i) - \left(\gamma_{11} + \gamma_{12} F_{n\eta}^{-1}(v^*)\right)}{\left(\gamma_{21} + \gamma_{22} F_{n\eta}^{-1}(v^*)\right)} \tag{C.25}$$

$$= \frac{\gamma_{12}}{\gamma_{21} + \gamma_{22} F_{n\eta}^{-1}(v^*)}(\eta_i - F_{n\eta}^{-1}(v^*)) + \left\{ \frac{\gamma_{21}}{\gamma_{21} + \gamma_{22} F_{n\eta}^{-1}(v^*)} + \frac{\gamma_{22}}{\gamma_{21} + \gamma_{22} F_{n\eta}^{-1}(v^*)}\eta_i \right\} b_i \tag{C.26}$$

$$\equiv \gamma_{11}^* + \gamma_{12}^*(\eta_i - F_{n\eta}^{-1}(v^*)) + (\gamma_{21}^* + \gamma_{22}^*\eta_i)b_i, \tag{C.27}$$

with $\gamma_{11}^* = 0$, $\gamma_{12}^* = \frac{\gamma_{12}}{\gamma_{21}+\gamma_{22}F_{n\eta}^{-1}(v^*)}$, $\gamma_{21}^* = \frac{\gamma_{21}}{\gamma_{21}+\gamma_{22}F_{n\eta}^{-1}(v^*)}$, $\gamma_{22}^* = \frac{\gamma_{22}}{\gamma_{21}+\gamma_{22}F_{n\eta}^{-1}(v^*)}$. Therefore, Normalization 1 can be imposed in a location-scale model by setting $\gamma_{11}^* = 0$ and centering appropriately the first occurence of η_i.

Appendix D. Asymptotics

The asymptotic distribution of the vector of parameters $\theta = (\alpha, \lambda)'$ obtains upon substituting $v(x_i, z_i; \alpha) = \frac{x_i - \alpha_1 \cdot z_i}{\alpha_2 \cdot z_i}$ for η_i, letting $d_{vi} = (x_i, w_i, x_i v(x_i, z_i; \alpha), w_i v(x_i, z_i; \alpha))$ and noting that the following $2 \times (dim(D) + dim(Z))$ moment conditions are available

$$D_v^\top \left\{ \frac{y - \lambda_1 \cdot d_v}{\lambda_2 \cdot d_v} \right\} = 0 \tag{D.1}$$

$$D_v^\top \left[\left\{ \frac{y - \lambda_1 \cdot d_v}{\lambda_2 \cdot d_v} \right\}^2 - 1 \right] = 0 \tag{D.2}$$

$$Z^\top \left\{ \frac{x - \alpha_1 \cdot z}{\alpha_2 \cdot z} \right\} = 0 \tag{D.3}$$

$$Z^\top \left[\left\{ \frac{x - \alpha_1 \cdot z}{\alpha_2 \cdot z} \right\}^2 - 1 \right] = 0, \tag{D.4}$$

corresponding to the vector of moments defined in the main text:

$$g(y, x, z, \theta) = \left(g_1(y, x, w, z, \theta)', g_2(y, x, w, z, \theta)', g_3(y, x, w, z, \theta)', g_4(y, x, w, z, \theta)'\right)'.$$

Therefore, the estimator in Step 2 of the estimation procedure can be viewed as a stacked Method of Moments estimator with moments $g(y, x, w, z, \theta)$, solving the system

$$\frac{1}{n}\sum_{i=1}^{n} g(y_i, x_i, w_i, z_i, \hat{\theta}) = 0. \tag{D.5}$$

Under the assumption that the model is well specified and data is i.i.d, define G_o and S_o as

$$G_o = E\left[\left. \frac{\partial g}{\partial \theta'} \right|_{\theta_o} \right] \quad \text{and} \quad S_o = E\left[gg'|_{\theta_o} \right]. \tag{D.6}$$

Applying standard results for the Method of Moments, there is

$$\sqrt{n}\left(\hat{\theta}-\theta_o\right) \xrightarrow{d} \mathcal{N}\left(0, G_o^{-1} S_o \left(G_o^{-1}\right)'\right). \tag{D.7}$$

Explicit characterization of the variance-covariance matrix follows from the following steps. Partitioning G_o, there is

$$G_o = \lim \frac{1}{n}\sum_{i=1}^{n} E \begin{bmatrix} \frac{\partial g_1(y_i,x_i,w_i,z_i,\theta)}{\partial \lambda_1'} & \frac{\partial g_1(y_i,x_i,w_i,z_i,\theta)}{\partial \lambda_2'} & \frac{\partial g_1(y_i,x_i,w_i,z_i,\theta)}{\partial \alpha_1'} & \frac{\partial g_1(y_i,x_i,w_i,z_i,\theta)}{\partial \alpha_2'} \\ \frac{\partial g_2(y_i,x_i,w_i,z_i,\theta)}{\partial \lambda_1'} & \frac{\partial g_2(y_i,x_i,w_i,z_i,\theta)}{\partial \lambda_2'} & \frac{\partial g_2(y_i,x_i,w_i,z_i,\theta)}{\partial \alpha_1'} & \frac{\partial g_2(y_i,x_i,w_i,z_i,\theta)}{\partial \alpha_2'} \\ \frac{\partial g_3(y_i,x_i,w_i,z_i,\theta)}{\partial \lambda_1'} & \frac{\partial g_3(y_i,x_i,w_i,z_i,\theta)}{\partial \lambda_2'} & \frac{\partial g_3(y_i,x_i,w_i,z_i,\theta)}{\partial \alpha_1'} & \frac{\partial g_3(y_i,x_i,w_i,z_i,\theta)}{\partial \alpha_2'} \\ \frac{\partial g_4(y_i,x_i,w_i,z_i,\theta)}{\partial \lambda_1'} & \frac{\partial g_4(y_i,x_i,w_i,z_i,\theta)}{\partial \lambda_2'} & \frac{\partial g_4(y_i,x_i,w_i,z_i,\theta)}{\partial \alpha_1'} & \frac{\partial g_4(y_i,x_i,w_i,z_i,\theta)}{\partial \alpha_2'} \end{bmatrix} \tag{D.8}$$

$$\equiv \begin{bmatrix} G_{11} & G_{12} & G_{13} & G_{14} \\ G_{21} & G_{22} & G_{23} & G_{24} \\ G_{31} & G_{32} & G_{33} & G_{34} \\ G_{41} & G_{42} & G_{43} & G_{44} \end{bmatrix}. \tag{D.9}$$

Some results needed.

$$\frac{\partial b_i}{\partial \lambda_1} = -\frac{d_{vi}}{(\lambda_2 \cdot d_{vi})} \tag{D.10}$$

$$\frac{\partial b_i}{\partial \lambda_2} = -\frac{d_{vi}}{\lambda_2 \cdot d_{vi}}\left(\frac{y_i - \lambda_1 \cdot d_{vi}}{\lambda_2 \cdot d_{vi}}\right) \tag{D.11}$$

$$\frac{\partial \eta_i}{\partial \alpha_1} = -\frac{z_i}{\alpha_2 \cdot z_i} \tag{D.12}$$

$$\frac{\partial \eta_i}{\partial \alpha_2} = -\frac{z_i}{\alpha_2 \cdot z_i}\left(\frac{x_i - \alpha_1 \cdot z_i}{\alpha_2 \cdot z_i}\right) \tag{D.13}$$

Also, for $j = 1, 2$, let $\tilde{\lambda}_j$ the $1 + dim(W)$ vector of components of λ_j corresponding to the interaction terms $xv(x, z; \alpha)$ and $wv(x, z; \alpha)$. Then for $j = 1, 2$,

$$\frac{\partial\left(y_i - \lambda_1 \cdot d_{vi}\right)}{\partial \alpha_j} = -\left(\tilde{\lambda}_1 \cdot \tilde{d}_{vi}\right)\frac{\partial \eta_i}{\partial \alpha_j} \tag{D.14}$$

$$\frac{\partial\left(\lambda_2 \cdot d_{vi}\right)}{\partial \alpha_j} = \left(\tilde{\lambda}_2 \cdot \tilde{d}_{vi}\right)\frac{\partial \eta_i}{\partial \alpha_j}, \tag{D.15}$$

so that

$$\frac{\partial b_i}{\partial \alpha_j} = \frac{\left(-\left(\tilde{\lambda}_1 \cdot \tilde{d}_{vi}\right)\frac{\partial \eta_i}{\partial \alpha_j}\right)(\lambda_2 \cdot d_{vi}) - \left(\left(\tilde{\lambda}_2 \cdot \tilde{d}_{vi}\right)\frac{\partial \eta_i}{\partial \alpha_j}\right)(y_i - \lambda_1 \cdot d_{vi})}{(\lambda_2 \cdot d_{vi})^2} \tag{D.16}$$

$$= \frac{-\left(\tilde{\lambda}_1 \cdot \tilde{d}_{vi}\right)\frac{\partial \eta_i}{\partial \alpha_j} - \left(\left(\tilde{\lambda}_2 \cdot \tilde{d}_{vi}\right)\frac{\partial \eta_i}{\partial \alpha_j}\right) b_i}{(\lambda_2 \cdot d_{vi})} \tag{D.17}$$

$$= -\left[\frac{\left(\tilde{\lambda}_1 + \tilde{\lambda}_2 b_i\right) \cdot \tilde{d}_{vi}}{\lambda_2 \cdot d_{vi}}\right]\frac{\partial \eta_i}{\partial \alpha_j}. \tag{D.18}$$

Last, there is the vector of derivatives

$$\frac{\partial d_{vi}}{\partial \alpha_j} = \left(0, 0, x_i \frac{\partial v(x_i, z_i; \alpha)}{\partial \alpha_j}, w_i \frac{\partial v(x_i, z_i; \alpha)}{\partial \alpha_j}\right)' \tag{D.19}$$

Computing the four blocks. ***Lower-left block:*** A first observation is that $G_{31}, G_{32}, G_{41}, G_{42}$ are all zeros as λ does not enter $g_3(y_i, x_i, w_i, z_i, \theta)$ and $g_4(y_i, x_i, w_i, z_i, \theta)$.

Upper-left block:

$$\begin{bmatrix} \frac{\partial g_1(y_i,x_i,w_i,z_i,\theta)}{\partial \lambda_1'} & \frac{\partial g_1(y_i,x_i,w_i,z_i,\theta)}{\partial \lambda_2'} \\ \frac{\partial g_2(y_i,x_i,w_i,z_i,\theta)}{\partial \lambda_1'} & \frac{\partial g_2(y_i,x_i,w_i,z_i,\theta)}{\partial \lambda_2'} \end{bmatrix} = \begin{bmatrix} -\frac{d_{vi}d_{vi}'}{\lambda_2 \cdot d_i} & -\frac{d_{vi}d_{vi}'}{\lambda_2 \cdot d_{vi}}\left(\frac{y_i - \lambda_1 \cdot d_{vi}}{\lambda_2 \cdot d_{vi}}\right) \\ -2\frac{d_{vi}d_{vi}'}{\lambda_2 \cdot d_{vi}}\left(\frac{y_i - \lambda_1 \cdot d_{vi}}{\lambda_2 \cdot d_{vi}}\right) & -2\frac{d_{vi}d_{vi}'}{\lambda_2 \cdot d_{vi}}\left(\frac{y_i - \lambda_1 \cdot d_{vi}}{\lambda_2 \cdot d_{vi}}\right)^2 \end{bmatrix}. \tag{D.20}$$

Upper-right block:

There is

$$\frac{\partial g_1(y_i, x_i, w_i, z_i, \theta)}{\partial \alpha_j'} = \frac{\partial d_{vi}}{\partial \alpha_j} b_i + d_{vi} \frac{\partial b_i}{\partial \alpha_j} \tag{D.21}$$

$$= \frac{\partial d_{vi}}{\partial \alpha_j} b_i - \left[\frac{\left(\tilde{\lambda}_1 + \tilde{\lambda}_2 b_i\right) \cdot \tilde{d}_{vi}}{\lambda_2 \cdot d_{vi}}\right] x_{vi} \frac{\partial \eta_i}{\partial \alpha_j} \tag{D.22}$$

$$\frac{\partial g_2(y_i, x_i, w_i, z_i, \theta)}{\partial \alpha_j'} = \frac{\partial d_{vi}}{\partial \alpha_j}\left(b_i^2 - 1\right) + 2 d_{vi} b_i \frac{\partial b_i}{\partial \alpha_j} \tag{D.23}$$

$$= \frac{\partial d_i}{\partial \alpha_j}\left(b_i^2 - 1\right) - 2\left[\frac{\left(\tilde{\lambda}_1 + \tilde{\lambda}_2 b_i\right) \cdot \tilde{d}_{vi}}{\lambda_2 \cdot d_{vi}}\right] d_{vi} \frac{\partial v(x_i, z_i; \alpha)}{\partial \alpha_j} b_i \tag{D.24}$$

so that

$$\frac{\partial g_1(y_i, x_i, w_i, z_i, \theta)}{\partial \alpha_1'} = \frac{\partial d_{vi}}{\partial \alpha_1} b_i + \left[\frac{\left(\tilde{\lambda}_1 + \tilde{\lambda}_2 b_i\right) \cdot \tilde{d}_{vi}}{\lambda_2 \cdot d_{vi}}\right] \frac{d_{vi} z_i'}{\alpha_2 \cdot z_i}$$

$$\frac{\partial g_1(y_i, x_i, w_i, z_i, \theta)}{\partial \alpha_2'} = \frac{\partial d_{vi}}{\partial \alpha_2} b_i + \left[\frac{\left(\tilde{\lambda}_1 + \tilde{\lambda}_2 b_i\right) \cdot \tilde{d}_{vi}}{\lambda_2 \cdot d_{vi}}\right] \frac{d_i z_i'}{\alpha_2 \cdot z_i} \eta_i$$

$$\frac{\partial g_2(y_i, x_i, w_i, z_i, \theta)}{\partial \alpha_1'} = \frac{\partial d_{vi}}{\partial \alpha_1}\left(b_i^2 - 1\right) + 2\left[\frac{\left(\tilde{\lambda}_1 + \tilde{\lambda}_2 b_i\right) \cdot \tilde{d}_{vi}}{\lambda_2 \cdot d_{vi}}\right] \frac{d_{vi} z_i'}{\alpha_2 \cdot z_i} b_i$$

$$\frac{\partial g_2(y_i, x_i^e, w_i, z_i, \theta)}{\partial \alpha_2'} = \frac{\partial d_{vi}}{\partial \alpha_2}\left(b_i^2 - 1\right) + 2\left[\frac{\left(\tilde{\lambda}_1 + \tilde{\lambda}_2 b_i\right) \cdot \tilde{d}_{vi}}{\lambda_2 \cdot d_{vi}}\right] \frac{d_{vi} z_i'}{\alpha_2 \cdot z_i} v(x_i, z_i; \alpha) b_i.$$

Lower-right block:

$$\begin{bmatrix} \frac{\partial g_3(y_i,x_i,w_i,z_i,\theta)}{\partial \alpha_1'} & \frac{\partial g_3(y_i,x_i,w_i,z_i,\theta)}{\partial \alpha_2'} \\ \frac{\partial g_4(y_i,x_i,w_i,z_i,\theta)}{\partial \alpha_1'} & \frac{\partial g_4(y_i,x_i,w_i,z_i,\theta)}{\partial \alpha_2'} \end{bmatrix} = \begin{bmatrix} -\frac{z_i z_i'}{\alpha_2 \cdot z_i} & -\frac{z_i z_i'}{\alpha_2 \cdot z_i}\left(\frac{x_i - \alpha_1 \cdot z_i}{\alpha_2 \cdot z_i}\right) \\ -2\frac{z_i z_i'}{\alpha_2 \cdot z_i}\left(\frac{x_i - \alpha_1 \cdot z_i}{\alpha_2 \cdot z_i}\right) & -2\frac{z_i z_i'}{\alpha_2 \cdot z_i}\left(\frac{x_i - \alpha_1 \cdot z_i}{\alpha_2 \cdot z_i}\right)^2 \end{bmatrix}. \tag{D.25}$$

Inverting G_o. Putting things together, there is:

$$G_o = \begin{bmatrix} G_{11} & G_{12} & G_{13} & G_{14} \\ G_{21} & G_{22} & G_{23} & G_{24} \\ 0 & 0 & G_{33} & G_{34} \\ 0 & 0 & G_{43} & G_{44} \end{bmatrix} \equiv \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}. \tag{D.26}$$

By application of the partitioned inverse formula, and since $\mathbf{C} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$, there is:

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{B}\mathbf{D}^{-1} \\ \mathbf{0} & \mathbf{D}^{-1} \end{bmatrix}. \tag{D.27}$$

Let's compute diagonal elements, $\mathbf{A}^{-1}$ and $\mathbf{D}^{-1}$, first. $\mathbf{A}$ can be simplified by noting that the off-diagonal element

$$G_{12} = E\left[-\frac{d_{vi}d'_{vi}}{\lambda_2 \cdot d_{vi}}\left(\frac{y_i - \lambda_1 \cdot d_{vi}}{\lambda_2 \cdot d_{vi}}\right)\right] = 0 \tag{D.28}$$

since $E[y_i|d_{vi}] = \lambda_1 \cdot d_{vi}$, and similarly for G_{21}. Therefore $\mathbf{A}^{-1}$ is simply given by

$$\mathbf{A}^{-1} = \begin{bmatrix} G_{11}^{-1} & 0 \\ 0 & G_{22}^{-1} \end{bmatrix}, \tag{D.29}$$

since $\mathbf{A}$ is diagonal. Similarly, $\mathbf{D}$ can be simplified by noting that the off-diagonal element

$$G_{34} = E\left[-\frac{z_i z'_i}{\alpha_2 \cdot z_i}\left(\frac{x_i - \alpha_1 \cdot z_i}{\alpha_2 \cdot z_i}\right)\right] = 0 \tag{D.30}$$

since $E[x_i|z_i] = \alpha_1 \cdot z_i$, and similarly for G_{43}. Therefore, $\mathbf{D}^{-1}$ is simply given by

$$\mathbf{D}^{-1} = \begin{bmatrix} G_{33}^{-1} & 0 \\ 0 & G_{44}^{-1} \end{bmatrix}. \tag{D.31}$$

Therefore, the off-diagonal element $-\mathbf{A}^{-1}\mathbf{B}\mathbf{D}^{-1}$ is given by

$$\begin{aligned} -\mathbf{A}^{-1}\mathbf{B}\mathbf{D}^{-1} &= -\begin{bmatrix} G_{11}^{-1} & 0 \\ 0 & G_{22}^{-1} \end{bmatrix}\begin{bmatrix} G_{13} & G_{14} \\ G_{23} & G_{24} \end{bmatrix}\begin{bmatrix} G_{33}^{-1} & 0 \\ 0 & G_{44}^{-1} \end{bmatrix} \\ &= -\begin{bmatrix} G_{11}^{-1}G_{13}G_{33}^{-1} & G_{11}^{-1}G_{14}G_{44}^{-1} \\ G_{22}^{-1}G_{23}G_{33}^{-1} & G_{22}^{-1}G_{24}G_{44}^{-1} \end{bmatrix}. \end{aligned} \tag{D.32}$$

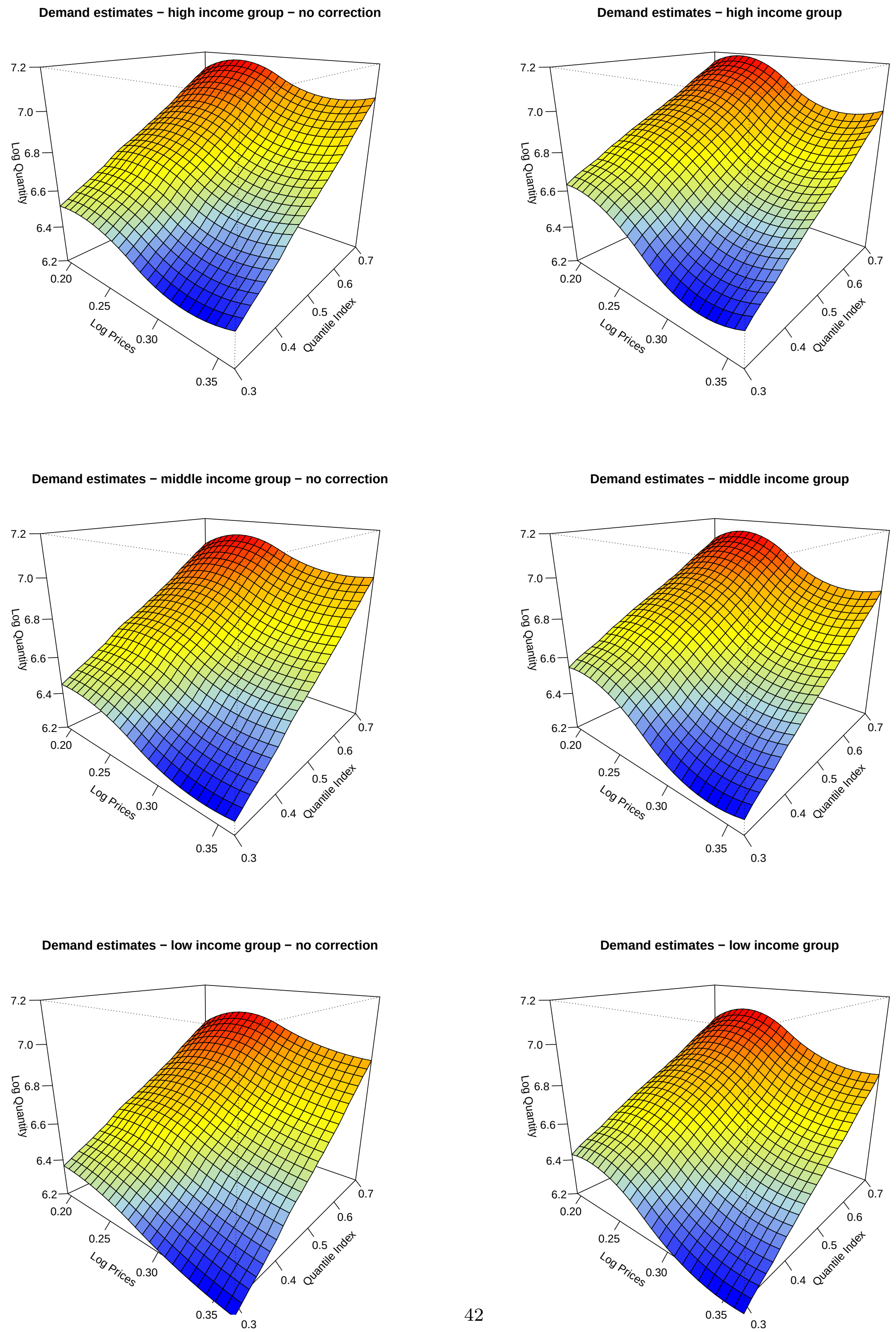


FIGURE 5. Statistical (Left) and structural (Right) demand functions by income groups - Conditional on $V = .5$.

Appendix E. **Additional Results for The Empirical Application**

Figure 5 shows estimated demand functions when the following empirical specification is considered for the structural equation:

$$y_i = \beta^*_{11}\tilde{x}_i + \beta^*_{12}r_i + \beta^*_{13}\tilde{x}_i r_i + \beta^*_{14}w_i + (\beta^*_{21}\tilde{x}_i + \beta^*_{22}r_i + \beta^*_{23}\tilde{x}_i r_i + \beta^*_{24}w_i)\varepsilon^*_i, \tag{E.1}$$

where, for observation i, r_i and y_i denote log income and log gasoline demand and $\tilde{x}_i$ is a vector of transformations of log prices. w_i is a vector of additional demographic characteristics of the household including: age of the household respondent, household size, and the number of drivers in the household (all measured in logs). w_i also includes an intercept. Number of employed household members is also included. As in the main text, transformations are taken to be cubic B-splines with 4 equally spaced knots.

References

Amemiya, T. (1982): "Two stage least absolute deviations estimators," *Econometrica*, 50(3), 689–711.

Belloni, A., C. V., and I. Fernandez-Val (2011): "Conditional Quantile Processes Based on Series or Many Regressors," .

Blundell, R., J. Horowitz, and M. Parey (2012): "Measuring the price responsiveness of gasoline demand: Economic shape restrictions and nonparametric demand estimation," *Quantitative Economics*, 3(1), 29–51.

Blundell, R., and J. Powell (2003): *Endogeneity in Nonparametric and Semiparametric Regression Models*. Cambridge University Press.

Chen, X. (2007): "Large sample sieve estimation of semi-nonparametric models," *Handbook of Econometrics*, 6, 5549–5632.

Chen, X., and D. Pouzo (2012): "Estimation of nonparametric conditional moment models with possibly nonsmooth generalized residuals," *Econometrica*, 80(1), 277–321.

Chernozhukov, V., F.-V. I., and A. Galichon (2010): "Quantile and Probability Curves Without Crossing," *Econometrica*, 78, 1093–1125.

Chernozhukov, V., I. Fernandez-Val, and A. Kowalski (2011): "Quantile Regression with Censoring and Endogeneity," *arXiv preprint arXiv:1104.4580*.

Chernozhukov, V., I. Fernandez-Val, and B. Melly (2009): "Inference on Counterfactual Distributions," *arXiv preprint arXiv:0904.0951*.

Chernozhukov, V., and C. Hansen (2005): "An IV Model of Quantile Treatment Effects," *Econometrica*, 73(1).

Chesher, A. (2001): “Exogenous Impact and Conditional Quantile Functions,” *CeMMAP*, CWP11/09.

——— (2003): “Identification in Nonseparable Models,” *Econometrica*, 71(5), 1405–1441.

——— (2005): “Nonparametric identification under discrete variation,” *Econometrica*, 73(5), 1525–1550.

——— (2007): “Instrumental values,” *Journal of Econometrics*, 139(1), 15–34.

——— (2010): “Instrumental variable models for discrete outcomes,” *Econometrica*, 78(2), 575–601.

Chesher, A., A. Rosen, and K. Smolinski (2011): “An instrumental variable model of multiple discrete choice,” *CeMMAP working papers*.

Chesher, A., and K. Smolinski (2012): “IV models of ordered choice,” *Journal of Econometrics*, 166(1), 33–48.

Cunha, F., J. Heckman, and S. Schennach (2010): “Estimating the technology of cognitive and noncognitive skill formation,” *Econometrica*, 78(3), 883–931.

d’Haultfoeuille, X., and P. Fevrier (2011): “Identification of Nonseparable Models with Endogeneity and Discrete Instruments,” .

Florens, J., J. Heckman, C. Meghir, and E. Vytlacil (2008): “Identification of Treatment Effects Using Control Functions in Models With Continuous, Endogenous Treatment and Heterogeneous Effects,” *Econometrica*, 76(5), 1191–1206.

Gagliardini, P., and O. Scaillet (2012): “Nonparametric Instrumental Variable Estimation of Structural Quantile Effects,” *Econometrica*, 80(4), 1533–1562.

Heckman, J., J. Stixrud, and S. Urzua (2006): “The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior,” *Journal of Labor Economics*, 24(3), 411–482.

Heckman, J., and E. Vytlacil (2007): “Econometric evaluation of social programs, part I: Causal models, structural models and econometric policy evaluation,” *Handbook of econometrics*, 6, 4779–4874.

Hoderlein, S., and E. Mammen (2007): “Identification of marginal effects in nonseparable models without monotonicity,” *Econometrica*, 75(5), 1513–1518.

Hoderlein, S., and A. Vanhems (2011): “Welfare analysis using nonseparable models,” Discussion paper, CeMMAP.

Horowitz, J., and S. Lee (2007): “Nonparametric instrumental variables estimation of a quantile regression model,” *Econometrica*, 75(4), 1191–1208.

Imbens, G. (2007): "Nonadditive models with endogenous regressors," *Econometric Society Monographs*, 43, 17.

Imbens, G., and J. Angrist (1994): "Identification and Estimation of Local Average Treatment effects," *Econometrica*, 61(2).

Imbens, G., and W. Newey (2009): "Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity," *Econometrica*, 77(5).

Jun, S. (2009): "Local structural quantile effects in a model with a nonseparable control variable," *Journal of Econometrics*, 151(1), 82–97.

Jun, S., J. Pinkse, and H. Xu (2011): "Tighter bounds in triangular systems," *Journal of Econometrics*, 161(2), 122–128.

Koenker, R. (2005): *Quantile Regression*. Cambridge University Press.

——— (2007): "quantreg: Quantile Regression, R package version 4.10," .

Lee, S. (2007): "Endogeneity in quantile regression models: A control function approach," *Journal of Econometrics*, 141(2), 1131–1158.

Li, Q., and J. Racine (2011): *Nonparametric econometrics: Theory and practice*. Princeton University Press.

Ma, L., and R. Koenker (2006): "Quantile regression methods for recursive structural equation models," *Journal of Econometrics*, 134, 471–506.

Matzkin, R. (2003): "Nonparametric estimation of nonadditive random functions," *Econometrica*, 71(5), 1339–1375.

Newey, W., and D. McFadden (1994): "Large sample estimation and hypothesis testing," *Handbook of econometrics*, 4, 2111–2245.

R-Development-Core-Team (2007): *R: A Language and Environment for Statistical Computing*R Foundation for Statistical Computing, Vienna, Austria.

Spady, R. (2006): "Identification and estimation of latent attitudes and their behavioral implications," Discussion paper, CeMMAP.

——— (2007): "Semiparametric methods for the measurement of latent attitudes and the estimation of their behavioural consequences," .

Spady, R., and S. Stouli (2012): "Dual Regression," *arXiv preprint arXiv:1210.6958*.

Su, C., and K. Judd (2012): "Constrained optimization approaches to estimation of structural models," *Econometrica*, 80(5), 2213–2230.

Torgovitsky, A. (2011): "Identification of nonseparable models with general instruments," Discussion paper, Working paper, Yale University.

WAECHTER, A., AND L. T. BIEGLER (2006): "On the implementation of a primaldual interior point filter line search algorithm for large-scale nonlinear programming," *Mathematical Programming*, 106(1), 25–57.

YPMA, J. (2011): "Introduction to ipoptr: an R interface to Ipopt," Discussion paper.