cemmap

centre for microdata methods and practice

Control variables, discrete instruments, and identification of structural functions

Whitney Newey
Sami Stouli

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP55/18

Control Variables, Discrete Instruments, and Identification of Structural Functions

Whitney Newey* and Sami Stouli†

September 14, 2018.

Abstract

Control variables provide an important means of controlling for endogeneity in econometric models with nonseparable and/or multidimensional heterogeneity. We allow for discrete instruments, giving identification results under a variety of restrictions on the way the endogenous variable and the control variables affect the outcome. We consider many structural objects of interest, such as average or quantile treatment effects. We illustrate our results with an empirical application to Engel curve estimation.

KEYWORDS: Control variables, discrete instruments, structural functions, endogeneity, partially parametric, nonseparable models, identification.

JEL CLASSIFICATION: C14, C31, C35

*Department of Economics, MIT, wnewey@mit.edu.
†Department of Economics, University of Bristol, s.stouli@bristol.ac.uk

1 Introduction

Nonseparable and/or multidimensional heterogeneity is important. It is present in discrete choice models as in McFadden (1973) and Hausman and Wise (1978). Multidimensional heterogeneity in demand functions allows price and income elasticities to vary over individuals in unrestricted ways, e.g., Hausman and Newey (2016) and Kitamura and Stoye (2017). It allows general variation in production technologies. Treatment effects that vary across individuals require intercept and slope heterogeneity.

Endogeneity is often a problem in these models because we are interested in the effect of an observed choice, or treatment variable on an outcome. Control variables provide an important means of controlling for endogeneity with multidimensional heterogeneity. A control variable is an observed or estimable variable that makes heterogeneity and treatment independent when it is conditioned on. Observed covariates serve as control variables for treatment effects (Rosenbaum and Rubin, 1983). The conditional cumulative distribution function (CDF) of a choice variable given an instrument can serve as a control variable in economic models (Imbens and Newey, 2009).

Nonparametric identification of many objects of interest, such as average or quantile treatment effects, requires a full support condition, that the support of the control variable conditional on the treatment variable is equal to the marginal support of the control variable. This restriction is often not satisfied in practice; e.g., see Imbens and Newey (2009) for Engel curves. It cannot be satisfied when instruments are discrete. One approach to this problem is to focus on identified sets for objects of interest, as for quantile effect in Imbens and Newey (2009). Another approach is to consider restrictions on the model that allow for point identification. Florens et al. (2008)

did so by showing identification when the structural function is a polynomial in the endogenous variable and a measurable separability condition is satisfied. Torgovitsky (2015) and D'Haultfœuille and Février (2015) did so by showing identification for discrete instruments when the structural disturbance is a scalar.

In this paper we give identification results under a variety of restrictions on the way the treatment and control variables enter the control regression of the outcome of interest on the endogenous and control variables. The restrictions we consider generalise those of Florens et al. (2008) to allow for nonpolynomial functions of endogenous variables or control variables. We also take a different approach to identification, focusing here on conditional nonsingularity of second moment matrices instead of measurable separability.

A main benefit of our approach is that it allows for discrete instruments. We show that identification of average, distribution and quantile treatment effects requires that the instrument have at least as many points of support as there are known functions of the endogenous variable or the control variable that appear in the control regression. These results are obtained by viewing various control regression specifications as varying coefficient models.

These results provide an alternative approach to identifying objects of interest in nonseparable models with discrete instruments. Instead of restricting the dimension of the heterogeneity to obtain identification with discrete instruments we can allow for multidimensional heterogeneity but restrict the way the treatment or controls affect the outcome.

We illustrate our results using an empirical application to Engel curves estimation using British expenditure survey data. We find that estimates of average, distributional and quantile treatment effects of total expenditure on food and leisure expenditure are not very sensitive to discretisation of the income instruments. We find that as we "coarsen" the instrument by only using knowledge of income intervals the structural estimates do not change much until the instrument is very coarse. Thus, in this empirical example we find that one can obtain good structural estimates even with discrete instruments.

These results also generalise the identification conditions for the baseline parametric models considered by Chernozhukov et al. (2017). Identification conditions based on conditional nonsingularity as considered here are more general than identification conditions based on support conditions.

In Section 2 we introduce the parametric models we consider. In Section 3 we give basic identification results for parametric models where either the endogenous variable or the control variable affects the outcome linearly. In Section 4 we extend these identification results to general parametric models. Section 5 gives results for partially parametric models that allow for nonparametric components. Section 6 reports the results of an empirical application to Engel curve estimation.

2 Parametric Modelling of Control Regressions

Let Y denote an outcome variable of interest and X an endogenous treatment with supports denoted by $\mathcal{Y}$ and $\mathcal{X}$, respectively. For ε a structural disturbance vector of unknown dimension, a nonseparable control variable model takes the form

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

where X and ε are independent conditional on an observable or estimable control variable denoted V. Conditioning on the control variable allows to identify general features of the structural relationship between X and Y in model (2.1), such as those captured by the structural functions of Blundell and Powell (2003, 2004), and Imbens and Newey (2009). An important kind of model where X is independent of ε conditional on V is a structural triangular system where $X = h(Z, \eta)$ and $h(z, \eta)$ is one-to-one in η. If (ε, η) are jointly independent of Z then $V = F_{X|Z}(X \mid Z)$, the conditional CDF of X given Z, is a control variable in this model (Imbens and Newey, 2009).

Leading examples of structural functions are the average structural function, $\mu(x)$, the distribution structural function (DSF), $G(y, x)$, and the quantile structural function (QSF) $Q(p, x)$, given by

$$\begin{aligned} \mu(x) &:= \int g(x, \varepsilon) F_\varepsilon(d\varepsilon), \quad G(y, x) := \Pr(g(x, \varepsilon) \leq y), \\ Q(p, x) &:= p^{\text{th}} \text{ quantile of } g(x, \varepsilon), \end{aligned}$$

where x is fixed in these expressions. These structural functions may be identifiable from control regressions of Y on X and V, including the conditional mean $E[Y \mid X, V]$, CDF, $F_{Y|XV}(Y \mid X, V)$, and quantile function, $Q_{Y|XV}(U \mid X, V)$, of Y given (X, V). In particular, when the support $\mathcal{V}_x$ of V conditional on $X = x$ equals the marginal support $\mathcal{V}$ of V we have

$$\begin{aligned} \mu(x) &= \int_{\mathcal{V}} E[Y \mid X = x, V = v] F_V(dv), \quad G(y, x) = \int_{\mathcal{V}} F_{Y|XV}(y \mid x, v) F_V(dv), \\ Q(p, x) &= G^{\leftarrow}(p, x) := \inf\{y \in \mathbb{R} : G(y, x) \geq p\}; \end{aligned} \tag{2.2}$$

see Blundell and Powell (2003) and Imbens and Newey (2009).

The key condition for equation (2.2) is full support, that the support $\mathcal{V}_x$ of V conditional on $X = x$ equals the marginal support of V. Without full support the integrals would not be well defined because integration would be over a range of (x, v) values that are outside the joint support of (X, V). Having a full support for each x is equivalent to (X, V) having rectangular support. In the absence of a rectangular support, global identification of the structural functions at all x must rely on alternative conditions that identify $F_{Y|XV}(y \mid x, v)$ for all $(x, v) \in \mathcal{X} \times \mathcal{V}$ and not merely over the joint support $\mathcal{XV}$ of (X, V). An example of such conditions are functional form restrictions on the controlled regressions $F_{Y|XV}$ and $Q_{Y|XV}$ which thus constitute natural modelling targets in the context of nonseparable conditional independence models. Imbens and Newey (2009) did show that structural effects may be partially identified without the full support condition. Here we focus on achieving identification via restricting the form of control regressions.

We begin with parametric specifications that are linear combinations of a vector of known functions $w(X, V)$ having the kronecker product form $p(X) \otimes q(V)$, where $p(X)$ and $q(V)$ are vectors of transformations of X and V, respectively. Let Γ denote a strictly increasing continuous CDF, such as the Gaussian CDF Φ, with inverse function denoted Γ^{-1}. The control regression specifications we consider are

$$E[Y \mid X, V] = \beta_0'[p(X) \otimes q(V)], \quad F_{Y|XV}(y \mid X, V) = \Gamma(\beta(y)'[p(X) \otimes q(V)]),$$
$$Q_{Y|X,V}(u \mid X, V) = \beta(u)'[p(X) \otimes q(V)], \quad u \in (0, 1), \tag{2.3}$$

where the coefficients $\beta(y)$ and $\beta(u)$ are functions of y and u, respectively. When Y is discrete the conditional distribution specification can be thought of as a discrete choice model as in McFadden (1973). As usual the quantile and conditonal mean coefficients are related by $\beta_0 = \int_0^1 \beta(u)du$. Chernozhukov et al. (2017) gives examples of structural models that give rise to control regressions as in equation (2.3).

It is convenient in what follows to use a common notation for the conditional mean, distribution, and quantile control regressions. For $\mathcal{U} = (0,1)$ and an index set $\mathcal{T} = \{0\}$, $\mathcal{Y}$, or $\mathcal{U}$, we define the collection of functions indexed by $\tau \in \mathcal{T}$,

$$\varphi_\tau(x,v) = \begin{cases} E[Y \mid X = x, V = v] & \text{if } \mathcal{T}=\{0\} \\ \Gamma^{-1}\left(F_{Y|XV}(\tau \mid x, v)\right) & \text{if } \mathcal{T}=\mathcal{Y} \\ Q_{Y|XV}(\tau \mid x, v) & \text{if } \mathcal{T}=\mathcal{U} \end{cases}.$$

While the coefficients $y \mapsto \beta(y)$ and $u \mapsto \beta(u)$ in (2.3) are infinite-dimensional parameters, for each τ in $\mathcal{T}$ the three control regression specifications share the essentially parametric form

$$\varphi_\tau(X,V) = \beta_\tau' w(X,V), \quad w(X,V) := p(X) \otimes q(V),$$

where the coefficient β_τ is a finite-dimensional parameter vector. This interpretation motivates the following definition of a *parametric* class of conditional independence models.

Assumption 1. *(a) For the model in (2.1), there exists a control variable V such that X and ε are independent conditional on V. (b) For a specified set $\mathcal{T} = \{0\}$, $\mathcal{Y}$, or $\mathcal{U}$, and each $\tau \in \mathcal{T}$, the outcome Y conditional on (X,V) follows the model*

$$\varphi_\tau(X,V) = \beta_\tau' w(X,V), \quad w(X,V) := p(X) \otimes q(V). \tag{2.4}$$

Standard results such as those of Newey and McFadden (1994) imply that point identification of β_τ only requires positive definiteness of the second moment matrix $E[w(X,V)w(X,V)']$. Under this condition knowledge of the control regressions is achievable at all $(y, x, v) \in \mathcal{Y} \times \mathcal{X} \times \mathcal{V}$, and the structural functions are then point identified as functionals of $\varphi_\tau(X, V)$ without full support. The formulation of primitive conditions under which $E[w(X, V)w(X, V)']$ is positive definite thus provides a characterisation of the identifying power of parametric conditional independence models without the full support condition. Chernozhukov et al. (2017) gave simple sufficient conditions when the joint distribution of X and V has a continuous component. Here we generalize these results in a way that allows for the distribution of V given X (or X given V) to be discrete.

Our identification analysis will also apply to other interesting structural objects that do not require the rectangular support assumption for identification. For example, by independence of ε from X conditional on V,

$$\varphi_y(x, v) = F_{Y|XV}(y \mid x, v) = \int 1(g(x, \varepsilon) \leq y) F_{\varepsilon|V}(d\varepsilon \mid v),$$

and its inverse,

$$\varphi_u(x, v) = Q_{Y|XV}(u \mid x, v) = \inf \{y \in \mathbb{R} : \varphi_y(x, v) \geq u\},$$

are structural objects. For instance, when the treatment X is continuous,

$$\frac{\partial \varphi_u(x, v)}{\partial x} = E\left[\frac{\partial g(x, \varepsilon)}{\partial x} \mid V = v, g(x, \varepsilon) = \varphi_u(x, v)\right]$$

is an average derivative of the structural function with respect to x conditional on the control variable taking value v and the outcome taking value $\varphi_u(x, v)$, that is the Local Average Structural Derivative of Hoderlein and Mammen (2007) and defined as the Local Quantile Structural Function in Fernandez-Val et al. (2018). All these objects will be identified under our conditions.

We next give primitive conditions for identification in parametric conditional independence models. For triangular systems, we show that these conditions can be satisfied with discrete valued instrumental variables. Estimation and inference methods for control regression functions (2.4) and the corresponding structural functions in triangular systems are extensively analysed by Chernozhukov et al. (2017), and directly apply when V is observable.

Remark 1. An additional vector of exogenous covariates Z_1 can be incorporated straightforwardly in our models. Let $r(Z_1)$ be a vector of known transformations of Z_1, and define $w(X, Z_1, V) := p(X) \otimes r(Z_1) \otimes q(V)$ the augmented vector of regressors. The control regressions then take the form

$$\varphi_\tau(X, Z_1, V) = \beta_\tau' w(X, Z_1, V), \quad \tau \in \mathcal{T}.$$

Our identification analysis is not affected by the presence of additional covariates and for clarity of exposition we do not include them in the remaining of the paper. Chernozhukov et al. (2017) provide a detailed exposition of the models we consider in the presence of exogenous covariates. □

3 Identification in Baseline Parametric Models

In this Section we formulate conditions for positive definiteness of $E[w(X,V)w(X,V)']$ in the important particular case where one of the elements $q(V)$ or $p(X)$ of the vector of regressors $w(X,V)$ is restricted to its first two components. With either $q(V) = (1,V)'$ or $p(X) = (1,X)'$, each type of restriction defines a class of baseline parametric models. For triangular systems we show that a binary instrumental variable is sufficient for identification of the corresponding control regression and structural functions. These baseline specifications are thus of substantial interest for empirical practice, and are important instances of the general parametric framework we consider in Section 4.

3.1 Main Result

In the first class of baseline models we set $q(V) = (1, V)'$ and the corresponding vector of regressors in the control regression function $\varphi_\tau(X, V)$ is $w(X, V) = (p(X)', V p(X)')'$. We denote the cardinality of a set such as $\mathcal{X}$ and $\mathcal{V}_x$ by $|\mathcal{X}|$ and $|\mathcal{V}_x|$. The condition for identification can then be formulated in terms of the support of V conditional on X: letting

$$\mathcal{X}_V^o = \{x \in \mathcal{X} \, : \, |\mathcal{V}_x| \geq 2\} \, ,$$

a sufficient condition is that $E[1(X \in \widetilde{\mathcal{X}}) p(X) p(X)']$ be positive definite with $\widetilde{\mathcal{X}} \subseteq \mathcal{X}_V^o$. Under this condition $\mathcal{X}_V^o$ is a set with positive probability and V has positive variance conditional on $X = x$ for each x in that set.

Alternatively, with $p(X) = (1, X)'$, the vector of regressors in the control regression function $\varphi_\tau(X, V)$ that defines the second class of baseline models is $w(X, V) = (q(V)', X q(V)')'$. The condition for identification can then be formulated in terms of the support of X conditional on V: letting

$$\mathcal{V}_X^o = \{v \in \mathcal{V} \, : \, |\mathcal{X}_v| \geq 2\} \, ,$$

a sufficient condition is that $E[1(V \in \widetilde{\mathcal{V}}) q(V) q(V)']$ be positive definite with $\widetilde{\mathcal{V}} \subseteq \mathcal{V}_X^o$. Under this condition $\mathcal{V}_X^o$ is a set with positive probability and X has positive variance conditional on $V = v$ for each v in that set.

Let $C < \infty$ denote some generic positive constant whose value may vary from place to place.

Assumption 2. *(a) We have that $E[p(X)p(X)']$ exists, $\sup_{x\in\mathcal{X}} E[||q(V)||^2 \mid X = x] \leq C$ and, for some specified set $\widetilde{\mathcal{X}}$, $E[1(X \in \widetilde{\mathcal{X}})p(X)p(X)']$ is positive definite. (b) We have that $E[q(V)q(V)']$ exists, $\sup_{v\in\mathcal{V}} E[||p(X)||^2 \mid V = v] \leq C$, and, for some specified set $\widetilde{\mathcal{V}}$, $E[1(V \in \widetilde{\mathcal{V}})q(V)q(V)']$ is positive definite.*

The following theorem states our first main result. The proofs of all our formal results are given in Appendix A.

Theorem 1. *(i) Let $q(V) = (1, V)'$. If Assumption 2(a) holds with $\widetilde{\mathcal{X}} \subseteq \mathcal{X}_V^o$, then $E[w(X, V)w(X, V)']$ exists and is positive definite. (ii) Let $p(X) = (1, X)'$. If Assumption 2(b) holds with $\widetilde{\mathcal{V}} \subseteq \mathcal{V}_X^o$, then $E[w(V, X)w(V, X)']$ exists and is positive definite.*

The formulation of sufficient conditions for identification in terms of $\mathcal{X}_V^o$ and $\mathcal{V}_X^o$ emphasises the fact that the full support condition $\mathcal{V}_x = \mathcal{V}$ is not required for $E[w(V, X)w(V, X)']$ to be positive definite in the baseline specifications. Under Assumption 1, identification of the control regressions and structural functions then follows by Theorem 1 in Chernozhukov et al. (2017). We also note that identification does not depend on the dimension of the unrestricted element $p(X)$ or $q(V)$ entering the vector of regressors $w(X, V)$. Thus the baseline specifications allow for flexible modelling of either how X affects the control regression functions or how V affects the control regression functions. When $q(V) = (1, V)'$, complex features of the relationship between X and Y can also be incorporated in the specification of the structural functions.

An example illustrating the modelling trade-offs inherent to our baseline specifications is the random coefficient model

$$Y = g(X, \varepsilon) = \sum_{j=1}^{J} p_j(X)\varepsilon_j, \tag{3.1}$$

where the unobserved heterogeneity components ε_j, $j \in \{1, \dots, J\}$, satisfy the conditional independence property

$$\varepsilon_j = Q_{\varepsilon_j|XV}(U \mid X, V) = \sum_{k=1}^{K} \beta_{jk}(U)q_k(V), \quad U \mid X, V \sim U(0, 1), \tag{3.2}$$

and the control variable V is normalised to have mean zero. For the specification with $q(V) = (1, V)'$ and $K = 2$, for each $u \in \mathcal{U}$ the control conditional quantile function is

$$\begin{aligned} Q_{Y|XV}(u \mid X, V) &= \sum_{j=1}^{J} p_j(X)[\beta_{j1}(u) + \beta_{j2}(u)V] \\ &= \sum_{j=1}^{J} \beta_{j1}(u)p_j(X) + \sum_{j=1}^{J} \beta_{j2}(u)\{p_j(X)V\} = \beta_u'[p(X) \otimes q(V)], \end{aligned}$$

where $\beta_u = (\beta_{u1}', \beta_{u2}')'$, $\beta_{uk} = (\beta_{u1k}, \dots, \beta_{uJk})'$, $\beta_{ujk} := \beta_{jk}(u)$, $j \in \{1, \dots, J\}$, $k \in \{1, 2\}$, which has the form of (2.4) with $\mathcal{T} = \mathcal{U}$ and $\tau = u$ in Assumption 1. The corresponding control conditional mean function is

$$E[Y \mid X, V] = \int_0^1 Q_{Y|X,V}(u \mid X, V)du = \beta_0'[p(X) \otimes q(V)],$$

where $\beta_0 = (\beta_{01}', \beta_{02}')'$, $\beta_{0k} = (\beta_{01k}, \dots, \beta_{0Jk})'$, $\beta_{0jk} := \int_0^1 \beta_{jk}(u)du$, $j \in \{1, \dots, J\}$, $k \in \{1, 2\}$, which has the form of (2.4) with $\mathcal{T} = \{0\}$ and $\tau = 0$ in Assumption 1. Upon using that $\int_{\mathcal{V}} vF_V(dv) = 0$, the corresponding average structural function takes the form

$$\mu(x) = \int_{\mathcal{V}} E[Y \mid X = x, V = v]F_V(dv) = \beta_{01}'p(x).$$

Model (3.1)-(3.2) thus allows for flexible modelling of the relationship between the treatment X and the outcome Y in both the control regression and average structural functions, which are identified under the conditions of Theorem 2. Similarly, when $p(X) = (1, X)'$ and setting $J = 2$ in (3.1), complex features of the relationship between the source of endogeneity V and the outcome Y can be captured by the model specification, while the average structural function will be linear in X.

3.2 Identification in Triangular Systems

In triangular systems with control variable $V = F_{X|Z}(X \mid Z)$, the conditions given above for $E[w(X,V)w(X,V)']$ to be positive definite translate into primitive conditions in terms of $\mathcal{Z}_x$, the support of Z conditional on $X = x$. Letting

$$\mathcal{X}_Z^o = \{x \in \mathcal{X} : |\mathcal{Z}_x| \geq 2\},$$

the matrix $E[w(X,V)w(X,V)']$ will be positive definite if Assumption 2(a) holds for a set $\widetilde{\mathcal{X}} \subseteq \mathcal{X}_Z^o$ such that $F_{X|Z}(x \mid z) \neq F_{X|Z}(x \mid \tilde{z})$ for some $z, \tilde{z} \in \mathcal{Z}_x$ and all $x \in \mathcal{X}_Z^o$. For $v \mapsto Q_{X|Z}(v \mid Z)$ denoting the quantile function of X conditional on Z, the result also holds if Assumption 2(b) is satisfied for a set $\widetilde{\mathcal{V}} \subseteq (0,1)$ with positive probability such that $Q_{X|Z}(v \mid z) \neq Q_{X|Z}(v \mid \tilde{z})$ for some $z, \tilde{z} \in \mathcal{Z}$ and all $v \in \widetilde{\mathcal{V}}$.

Assumption 3. *(a) For some specified set $\widetilde{\mathcal{X}}$, we have $F_{X|Z}(x \mid z) \neq F_{X|Z}(x \mid \tilde{z})$ for some $z, \tilde{z} \in \mathcal{Z}_x$ and all $x \in \widetilde{\mathcal{X}}$. (b) For some specified set $\widetilde{\mathcal{V}}$, we have $Q_{X|Z}(v \mid z) \neq Q_{X|Z}(v \mid \tilde{z})$ for some $z, \tilde{z} \in \mathcal{Z}$ and all $v \in \widetilde{\mathcal{V}}$.*

Under this condition a discrete instrument, including binary, is then sufficient for our baseline parametric models to identify the structural functions.

Theorem 2. *Suppose that either the conditions of Theorem 1(i) and Assumption 3(a) hold with $\widetilde{\mathcal{X}} \subseteq \mathcal{X}_Z^o$, or the conditions of Theorem 1(ii) and Assumption 3(b) hold with $\widetilde{\mathcal{V}} \subseteq (0,1)$. If Assumption 1 holds with $\mathcal{T} = \mathcal{Y}$ or $\mathcal{U}$ then the average, distribution and quantile structural functions are identified. If Assumption 1 holds with $\mathcal{T} = \{0\}$ then the average structural function is identified.*

Theorem 2 demonstrates the relevance of the parametric specifications in a wide range of empirical settings, for instance triangular systems with a binary or discrete instrument and including a discrete or mixed continuous-discrete outcome.[1]

4 Generalisation

We generalise the results of the previous Section by expanding the set of regressors in the baseline specifications. In the more general case we consider here, both $p(X)$ and $q(V)$ are vectors of transformations of X and V, respectively. In practice these will typically consist of basis functions with good approximating properties such as splines, trigonometric or orthogonal polynomials (cf. Appendix D.2 for an illustration to parametric demand analysis with splines).

4.1 Identification in Parametric Models

One general condition for positive definiteness of $E[w(X,V)w(X,V)']$ is the existence of a set of values x of X with positive probability such that the smallest eigenvalue of $E[q(V)q(V)' \mid X = x]$ is bounded away from zero. An alternative general condition is the existence of a set of values v of V with positive probability such that the smallest

[1]For instance our baseline models can be used for the specification of parametric sample selection models with censored selection rule as considered in Fernandez-Val et al. (2018).

eigenvalue of $E[p(X)p(X)' \mid V = v]$ is bounded away from zero. This characterisation leads to natural sufficient conditions for $E\left[w(X,V)w(X,V)'\right]$ to be positive definite when the vectors $p(X)$ and $q(V)$ are unrestricted.

With $B > 0$ denoting some generic constant whose value may vary from place to place, let $\lambda_{\min}(x)$ denote the smallest eigenvalue of $E[q(V)q(V)' \mid X = x]$, and define

$$\mathcal{X}_V^* = \left\{x \in \mathcal{X} \, : \, \lambda_{\min}(x) \geq B > 0\right\}.$$

The smallest eigenvalue of $E[q(V)q(V)' \mid X = x]$ is then bounded away from zero uniformly over $x \in \mathcal{X}_V^*$, and a sufficient condition for identification is that Assumption 2(a) holds with $\widetilde{\mathcal{X}} \subseteq \mathcal{X}_V^*$. Alternatively, let $\lambda_{\min}(v)$ denote the smallest eigenvalue of $E[p(X)p(X)' \mid V = v]$, and define

$$\mathcal{V}_X^* = \left\{v \in \mathcal{V} \, : \, \lambda_{\min}(v) \geq B > 0\right\}.$$

The eigenvalues of $E[p(X)p(X)' \mid V = v]$ are then bounded away from zero uniformly over $v \in \mathcal{V}_X^*$, and a sufficient condition for identification is that Assumption 2(b) holds with $\widetilde{\mathcal{V}} \subseteq \mathcal{V}_X^*$.

Theorem 3. *For some $B > 0$, if either Assumption 2(a) holds with $\widetilde{\mathcal{X}} \subseteq \mathcal{X}_V^*$, or Assumption 2(b) holds with $\widetilde{\mathcal{V}} \subseteq \mathcal{V}_X^*$, then $E\left[w(X,V)w(X,V)'\right]$ exists and is positive definite*

Remark 2. For the baseline specifications, Proposition 2 in Appendix B shows that the conditions of Theorem 3 are satisfied by the conditions of Section 3. In the simple case $q(V) = (1, V)'$, if Assumption 2(a) holds with $\widetilde{\mathcal{X}} \subseteq \mathcal{X}_V^o$ then $\mathrm{Var}(V \mid X = x) \geq B > 0$ for each $x \in \mathcal{X}_V^o$, and Assumption 2(a) also holds with $\widetilde{\mathcal{X}} \subseteq \mathcal{X}_V^*$. In the simple case $p(X) = (1, X)'$, if Assumption 2(b) holds with $\widetilde{\mathcal{V}} \subseteq \mathcal{V}_X^o$ then $\mathrm{Var}(X \mid V = v) \geq B > 0$

for each $v \in \mathcal{V}_X^o$, and Assumption 2(b) also holds with $\widetilde{\mathcal{X}} \subseteq \mathcal{X}_V^*$.

4.2 Discussion

Theorem 3 gives a general identification result for models with regressors of a Kronecker product form $w(X,V) = p(X) \otimes q(V)$. Positive definiteness of the matrix $E[w(V,X)w(V,X)']$ is then a sufficient condition for uniqueness of the control regression functions with probability one. Thus the conditions of Theorem 3 are also sufficient for the models we consider to identify their corresponding structural functions.

Theorem 4. *Suppose the assumptions of Theorem 3 are satisfied. If Assumption 1 holds with $\mathcal{T} = \mathcal{Y}$ or $\mathcal{U}$ then the average, distribution and quantile structural functions are identified. If Assumption 1 holds with $\mathcal{T} = \{0\}$ then the average structural function is identified.*

The formulation of identification conditions in terms of the second conditional moment matrices of $p(X)$ and $q(V)$ is a considerable simplification relative to existing conditions in the literature. The assumptions of Theorems 3 and 4 are more primitive and easier to interpret than the dominance condition[2] proposed by Chernozhukov et al. (2017) for positive definiteness of $E[w(X,V)w(X,V)']$. These conditions are also weaker than the full support condition or the measurable separability condition of Florens et al. (2008), which require the control variable to have a continuous distribution conditional on X.

[2] Chernozhukov et al. (2017) assume that the joint probability distribution of X and V dominates a product probability measure $\mu(x) \times \rho(v)$ such that $E_\mu[p(X)p(X)']$ and $E_\rho[q(V)q(V)']$ are positive definite. This condition is sufficient for $E[w(X,V)w(X,V)']$ to be positive definite, but is difficult to interpret.

In a triangular system with control variable $V = F_{X|Z}(X \mid Z)$, our identification conditions can be satisfied by a discrete valued instrument, and admit an equivalent formulation in terms of the first stage model and the instrument Z. Letting $\widetilde{\lambda}_{\min}(x)$ denote the smallest eigenvalue of

$$E[q(F_{X|Z}(X \mid Z))q(F_{X|Z}(X \mid Z))' \mid X = x],$$

for $x \in \mathcal{X}$, define the corresponding set $\mathcal{X}_Z^* = \{x \in \mathcal{X} : \widetilde{\lambda}_{\min}(x) \geq B > 0\}$. Then $\widetilde{\lambda}_{\min}(x) = \lambda_{\min}(x)$ and $\mathcal{X}_Z^* = \mathcal{X}_V^*$. Thus Assumption 2(a) with $\widetilde{\mathcal{X}} \subseteq \mathcal{X}_Z^*$ is sufficient for identification by Theorem 3. Moreover, when Z is discrete a necessary condition for this to hold is that there is a set $\widetilde{\mathcal{X}}$ with positive probability such that, for each $x \in \widetilde{\mathcal{X}}$, the set $\mathcal{R}(x)$ of distinct values[3] of $F_{X|Z}(x \mid z)$ for $z \in \mathcal{Z}_x$ has cardinality greater than or equal to $K \equiv \dim(q(V))$.

Alternatively, letting $\widetilde{\lambda}_{\min}(v)$ denote the smallest eigenvalue of

$$E[p(Q_{X|Z}(v \mid Z))p(Q_{X|Z}(v \mid Z))'],$$

for $v \in (0,1)$, define the corresponding set $\mathcal{V}_Z^* = \{v \in (0,1) : \widetilde{\lambda}_{\min}(v) \geq B > 0\}$. Then, by independence of V from Z, $\widetilde{\lambda}_{\min}(v) = \lambda_{\min}(v)$ and $\mathcal{V}_Z^* = \mathcal{V}_X^*$. Thus Assumption 2(b) with $\widetilde{\mathcal{V}} \subseteq \mathcal{V}_Z^*$ is sufficient for identification by Theorem 3. Moreover, when Z is discrete with support $\mathcal{Z}_x = \{z_1, \ldots, z_{|\mathcal{Z}_x|}\}$ conditional on $X = x$, a necessary condition for this to hold is that there is a set $\widetilde{\mathcal{V}}$ with positive probability such that,

[3] Formally, for $x \in \mathcal{X}$, we define $\mathcal{R}(x) = \left\{F_{X|Z}(x \mid z_m)\right\}_{m \in \mathcal{M}(x)}$, where

$$\mathcal{M}(x) = \left\{m \in \{1, \ldots, |\mathcal{Z}_x|\} : F_{X|Z}(x \mid z_m) \neq F_{X|Z}(x \mid z_{m'}) \text{ for all } m' \in \{1, \ldots, |\mathcal{Z}_x|\}/\{m\}\right\}.$$

for each $v \in \widetilde{\mathcal{V}}$, the set $\mathcal{Q}(v)$ of distinct values[4] of $Q_{X|Z}(v \mid z)$ for $z \in \mathcal{Z}$ has cardinality greater than or equal to $J \equiv \dim(p(X))$.

Proposition 1. *Suppose that $|\mathcal{Z}| < \infty$. Then Assumptions 2(a) and 2(b) hold with $\widetilde{\mathcal{X}} \subseteq \mathcal{X}_Z^*$ and $\widetilde{\mathcal{V}} \subseteq \mathcal{V}_X^*$, respectively, only if there exist sets $\widetilde{\mathcal{X}} \subseteq \mathcal{X}$ and $\widetilde{\mathcal{V}} \subseteq (0,1)$ of positive probability such that $\inf_{x\in\widetilde{\mathcal{X}}} |\mathcal{R}(x)| \geq K$ and $\inf_{v\in\widetilde{\mathcal{V}}} |\mathcal{Q}(v)| \geq J$, respectively.*

Proposition 1 shows that the flexibility of parametric triangular systems is restricted by the cardinality of the set of instrumental values. Thus identification can be achieved in the presence of a binary instrument only when one of the two vectors $q(V)$ and $p(X)$ is of dimension two. More generally, identification in the class of models we consider cannot be achieved whenever $|\mathcal{Z}| < \min(J, K)$.

5 Partially Parametric Specifications

An important generalisation of the parametric specifications of the previous sections is one where either the relationship between X and Y or between V and Y is unspecified in the control regression functions. This gives rise to two classes of models with known functional form of either how X affects the control regression functions or how V affects the control regression functions, but not both. These models are special cases of functional coefficient regression models.

The first class of partially parametric models we consider is one where X is known to affect the control regression function $\varphi_\tau(X, V)$ only through a vector of known

[4]Formally, for $v \in (0,1)$, we define $\mathcal{Q}(v) = \left\{Q_{X|Z}(v \mid z_m)\right\}_{m\in\mathcal{N}(v)}$, where

$$\mathcal{N}(v) = \left\{m \in \{1,\ldots,|\mathcal{Z}|\} : Q_{X|Z}(v \mid z_m) \neq Q_{X|Z}(v \mid z_{m'}) \text{ for all } m' \in \{1,\ldots,|\mathcal{Z}|\}/\{m\}\right\}.$$

functions $p(X)$. We assume that

$$\varphi_\tau(X, V) = p(X)' q_\tau(V), \quad \tau \in \mathcal{T}, \tag{5.1}$$

where the vector of functions $q_\tau(V)$ is now unknown, rather than a linear combination of finitely many known transformations of V. This model is studied in Newey and Stouli (2018) and generalises the polynomial specifications of Florens et al. (2008) to allow $p(X)$ to be any functions of X rather than just powers of X. An example of a structural model that gives rise to control regression functions as in (5.1) is the random coefficient model (3.1), where the outcome $Y = \sum_{j=1}^{J} p_j(X)\varepsilon_j$ is continuous and the unobserved heterogeneity components ε_j satisfy the the conditional independence property

$$\varepsilon_j = Q_{\varepsilon_j \mid X,V}(U \mid X, V) = Q_{\varepsilon_j \mid V}(U \mid V), \; U \mid X, V \sim U(0, 1), \; j \in \{1, \ldots, J\}. \tag{5.2}$$

Letting $q_j(u, v) := Q_{\varepsilon_j \mid V}(u \mid v)$, $j \in \{1, \ldots, J\}$, and substituting representation (5.2) in the outcome equation, the control regression representation for Y conditional on (X, V) is:

$$Y = \sum_{j=1}^{J} p_j(X) Q_{\varepsilon_j \mid V}(U \mid V) = p(X)' q(U, V),$$

with $p(x) = (p_1(x), \ldots, p_J(x))'$ and $q(u, v) = (q_1(u, v), \ldots q_J(u, v))'$. The corresponding control conditional quantile function is:

$$Q_{Y \mid X,V}(u \mid X, V) = p(X)' q(u, V) =: p(X)' q_u(V), \quad u \in \mathcal{U},$$

which has the form of (5.1) with $\mathcal{T} = \mathcal{U}$ and $\tau = u$. The corresponding control conditional mean function is:

$$\begin{aligned}E[Y \mid X, V] &= \int_0^1 Q_{Y|X,V}(u \mid X, V) du \\ &= \int_0^1 p(X)' q(u, V) du = p(X)' \left\{ \int_0^1 q(u, V) du \right\} =: p(X)' q_0(V),\end{aligned}$$

which has the form of (5.1) with $\mathcal{T} = \{0\}$ and $\tau = 0$. Thus this is a model with known functional form of how X affects the control conditional mean and quantile functions.

The second class of partially parametric models we consider is one where V is known to affect the control regression function $\varphi_\tau(X, V)$ only through a vector of known functions $q(V)$. We assume that

$$\varphi_\tau(X, V) = p_\tau(X)' q(V), \quad \tau \in \mathcal{T}, \tag{5.3}$$

where the vector of functions $p_\tau(X)$ is now unknown, rather than just a linear combination of finitely many known transformations of X. When Y is continuous, an example of a structural model that gives rise to a control regression function as in (5.3) is the latent random coefficient model

$$\xi = \sum_{k=1}^{K} \varepsilon_k q_k(V), \quad \xi \mid X, V \sim \Gamma, \tag{5.4}$$

where the unobserved heterogeneity components ε_k satisfy the restrictions

$$\varepsilon_k = p_k(Y, X), \quad k \in \{1, \ldots, K\}, \tag{5.5}$$

with $y \mapsto p_k(y, x)$ strictly increasing, and the conditional independence property

$$F_{\varepsilon_k|V}(\varepsilon_k \mid V) = F_{\varepsilon_k|XV}(\varepsilon_k \mid X, V), \quad k \in \{1, \ldots, K\}.$$

Upon substituting expression (5.5) for ε_k in the latent variable equation (5.4), the control regression representation for ξ conditional on (X, V) is:

$$\xi = \sum_{k=1}^{K} p_k(Y, X) q_k(V) = p(Y, X)' q(V),$$

with $q(v) = (q_1(v), \ldots, q_K(v))$ and $p(y, x) = (p_1(y, x), \ldots p_K(y, x))'$. The corresponding control conditional CDF satisfies:

$$\Gamma^{-1}\left(F_{Y|XV}(y \mid X, V)\right) = p(y, X)' q(V) =: p_y(X)' q(V), \quad y \in \mathcal{Y},$$

which has the form of (5.3) with $\mathcal{T} = \mathcal{Y}$ and $\tau = y$. Thus this is a model with known functional form of how V affects the control conditional CDF.

Remark 3. Additional exogenous covariates Z_1 can be incorporated straightforwardly in these models through the known functional component of the control regression function $\varphi_\tau(X, V)$. With an exogenous vector of covariates Z_1, model (5.1) takes the form

$$\varphi_\tau(X, Z_1, V) = p(X, Z_1)' q_\tau(V),$$

where $p(X, Z_1)$ is a vector of known functions of (X, Z_1), and model (5.3) takes the form

$$\varphi_\tau(X, Z_1, V) = p_\tau(X)' q(Z_1, V),$$

where $q(Z_1, V)$ is a vector of known functions of (Z_1, V). □

The following assumption gathers the two classes of partially parametric specifications.

Assumption 4. *(a) For a specified set $\mathcal{T} = \{0\}$, $\mathcal{Y}$, or $\mathcal{U}$, and each $\tau \in \mathcal{T}$, the outcome Y conditional on (X, V) follows the model*

$$\varphi_\tau(X, V) = p(X)' q_\tau(V); \tag{5.6}$$

we have $E[Y^2] < \infty$ and $E\left[\|p(X)\|^2\right] < \infty$; and $E\left[p(X)\, p(X)' \mid V\right]$ exists and is nonsingular with probability one; or (b) for a specified set $\mathcal{T} = \{0\}$, $\mathcal{Y}$, or $\mathcal{U}$, and each $\tau \in \mathcal{T}$, the outcome Y conditional on (X, V) follows the model

$$\varphi_\tau(X, V) = q(V)' p_\tau(X); \tag{5.7}$$

we have $E[Y^2] < \infty$ and $E\left[\|q(V)\|^2\right] < \infty$; and $E\left[q(V)\, q(V)' \mid X\right]$ exists and is nonsingular with probability one.

The next result states our main identification result of this Section.

Theorem 5. *(i) If Assumption 4(a) holds then $q_\tau(V)$ is identified for each $\tau \in \mathcal{T}$. (ii) If Assumption 4(b) holds then $p_\tau(X)$ is identified for each $\tau \in \mathcal{T}$.*

We earlier discussed conditions for nonsingularity of $E\left[p(X)\, p(X)' \mid V\right]$ and $E\left[q(V)\, q(V)' \mid X\right]$. All those conditions are sufficient for identification of $q_\tau(V)$ and $p_\tau(X)$, including those that allow for discrete valued instrumental variables, under the important stricter condition that they hold on sets of V and X having probability one, respectively. We also note that identification of $q_\tau(V)$ and $p_\tau(X)$ means uniqueness on sets of V and X having probability one, respectively. Thus the structural functions corresponding to models (5.6) and (5.7) are identified. For example, in the first class of models the quantile and distribution structural functions will be identified as

$$Q(p,x) = G^{\leftarrow}(p,x), \quad G(y,x) = \int_{\mathcal{V}} \Gamma\left(p\left(x\right)' q_y\left(v\right)\right) F_V(dv),$$

since $p\left(X\right)$ and Γ are known functions and $q_y\left(V\right)$ is identified, and hence $\Gamma\left(p\left(x\right)' q_y\left(v\right)\right)$ also is.

Theorem 6. *Suppose Assumption 1(a) holds. If Assumption 4 holds with $\mathcal{T} = \mathcal{Y}$ or $\mathcal{U}$ then the average, distribution and quantile structural functions are identified. If Assumption 4 holds with $\mathcal{T} = \{0\}$ then the average structural function is identified.*

6 Empirical Application

In this Section we illustrate our identification results by estimating the QSF for a parametric triangular system for Engel curves. We focus on the structural relationship between household's total expenditure and household's demand for two goods: food and leisure. We take the outcome Y to be the expenditure share on either food or leisure, and X the logarithm of total expenditure. We use as an instrument a discretised version $\widetilde{Z}$ of the logarithm of gross earnings of the head of household Z^*. We also include an additional binary covariate Z_1 accounting for the presence of children in the household.

There is a large literature using nonseparable triangular systems for the identification and estimation of Engel curves (Imbens and Newey, 2003, Chernozhukov et al., 2015, Chernozhukov et al., 2017). We follow Chernozhukov et al. (2017) who consider estimation of structural functions for food and leisure using triangular parametric control regression specifications. For comparison purposes we use the same dataset, the 1995 U.K. Family Expenditure Survey. We restrict the sample to 1,655 married or cohabiting couples with two or fewer children, in which the head of the household is employed and between the ages of 20 and 55 years. For this sample we estimate the QSF for both goods using discrete instruments and then compare our results to those obtained with a continuous instrument by Chernozhukov et al. (2017).

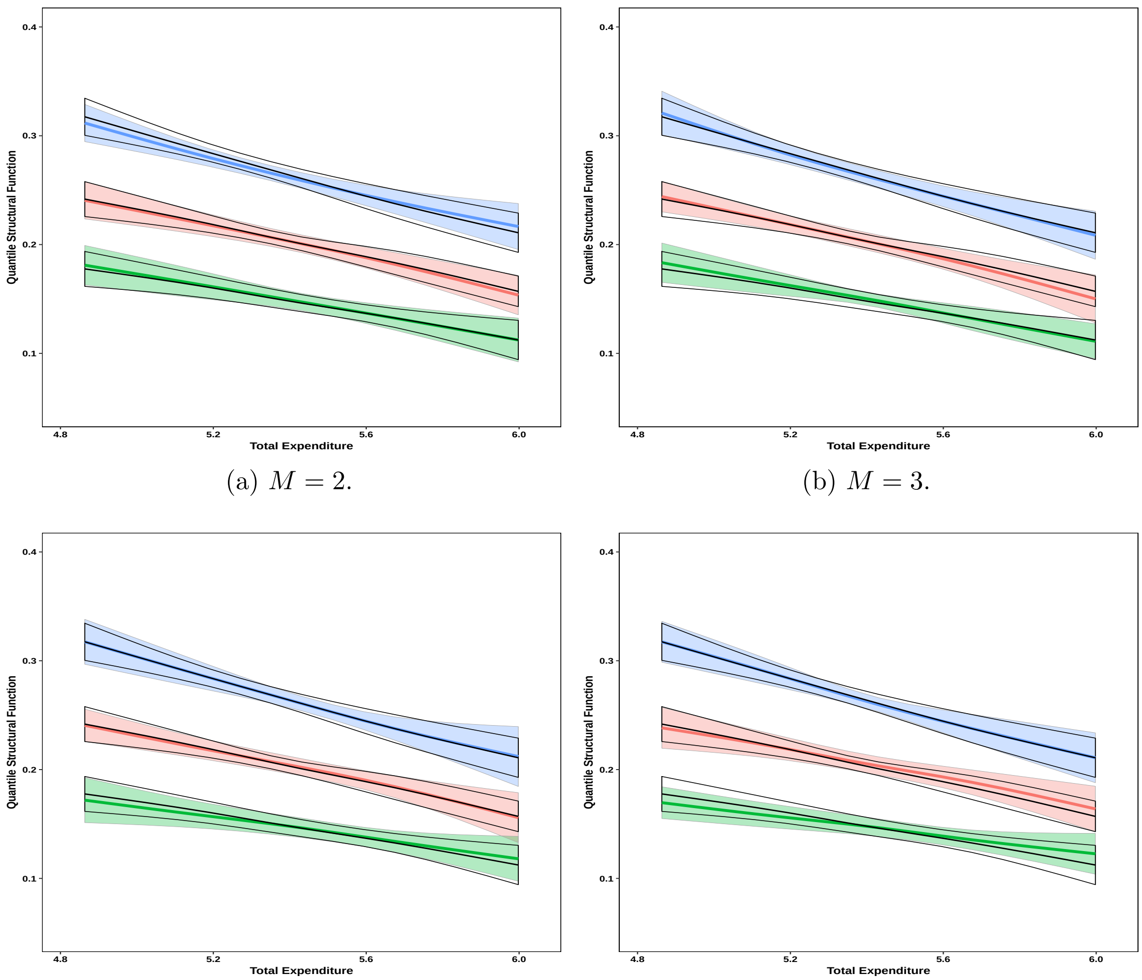
We consider the triangular system,

$$Y = Q_{Y|X,V}(U \mid X, V) = \beta(U)'[p(X) \otimes r(Z_1) \otimes q(V)], \quad U \mid X, Z_1, V \sim Unif(0,1)$$
$$X = Q_{X|Z}(V \mid Z) = \pi(V)'[s(\widetilde{Z}) \otimes r(Z_1)], \quad V \mid Z \sim Unif(0,1), \quad Z := (\widetilde{Z}, Z_1)',$$

where $s(\widetilde{Z}) = (1, \widetilde{Z})'$, $r(Z_1) = (1, Z_1)'$, $p(X) = (1, X)'$ and $q(V) = (1, \Phi^{-1}(V))'$. The corresponding QSFs are estimated by the quantile regression estimators of Chernozhukov et al. (2017), described in Appendix C. For our sample of $n = 1,655$ observations $\{(Y_i, X_i, Z_i)\}_{i=1}^n$, we construct two sets of four discrete valued instruments taking $M = 2, 3, 5$ and 15 values, respectively, and then estimate the QSFs using one instrument at a time. In the first set the instrument $\widetilde{Z}$ is uniformly distributed across its support (DESIGN 1). For $t_m = m/M$, $m \in \{0, 1, \ldots M\}$, let $\widehat{Q}_{Z^*}(t_m)$ denote the sample t_m quantile of Z^*. For $i \in \{1, \ldots, n\}$ and $m \in \{0, 1, \ldots M-1\}$ such that $Z_i^* \in [\widehat{Q}_{Z^*}(t_m), \widehat{Q}_{Z^*}(t_{m+1}))$, we define

$$\widetilde{Z}_i = \widehat{Q}_{Z^*}(t_m) + \frac{1}{2}\left[\widehat{Q}_{Z^*}(t_{m+1}) - \widehat{Q}_{Z^*}(t_m)\right].$$

For an observation i such that $Z_i^* = \max_{i \le n}(Z_i^*)$, we define $\widetilde{Z}_i = \widehat{Q}_{Z^*}(t_{M-1}) + \frac{1}{2}\left[\widehat{Q}_{Z^*}(t_M) - \widehat{Q}_{Z^*}(t_{M-1})\right]$. In the second set the instrument $\widetilde{Z}$ is discretised according to a non uniform distribution (DESIGN 2). Define the equispaced grid $\min_{i \le n}(Z_i^*) = \xi_0 < \xi_1 < \ldots < \xi_M = \max_{i \le n}(Z_i^*)$. For $i \in \{1, \ldots, n\}$ and $m \in \{0, \ldots, M-1\}$ such



(a) $M = 2$.

(b) $M = 3$.

(c) $M = 5$.

(d) $M = 15$.

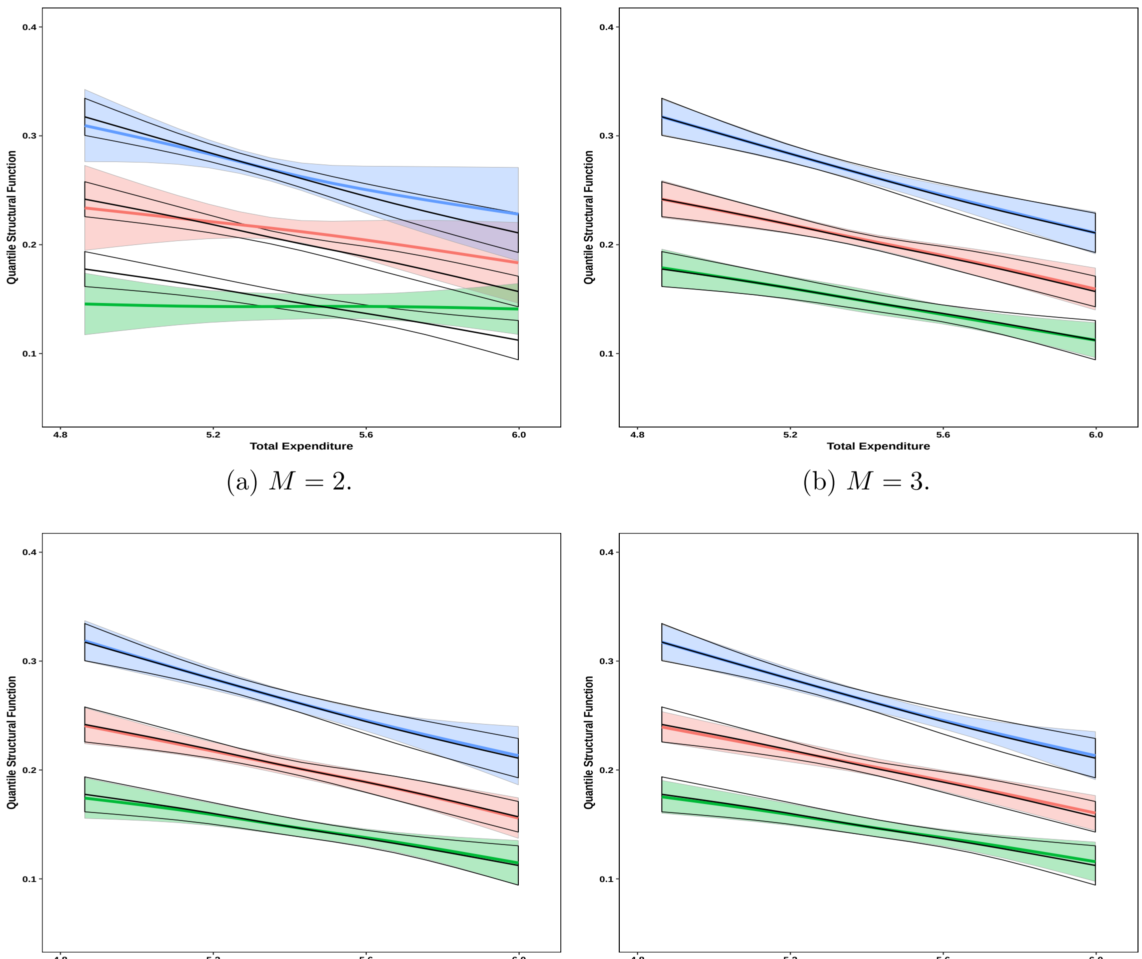
Figure 6.1: Design 1. QSF for food with discrete instrument $\widetilde{Z}$ (coloured) and with continuous instrument Z^* (black).

that $Z_i^* \in [\xi_m, \xi_{m+1})$ we define

$$\widetilde{Z}_i = \xi_m + \frac{1}{2}\left[\xi_{m+1} - \xi_m\right].$$

For an observation i such that $Z_i^* = \max_{i \leq n}(Z_i^*)$, we define $\widetilde{Z}_i = \xi_{M-1} + \frac{1}{2}\left[\xi_M - \xi_{M-1}\right]$.

Figures 6.1 and 6.2 show the 0.25, 0.5 and 0.75-QSFs for food estimated with each set of four instruments, respectively, as well as the corresponding benchmark QSFs



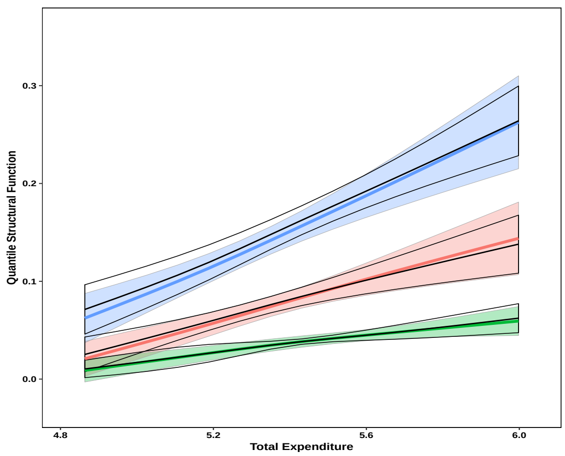
(a) $M = 2$. (b) $M = 3$.

(c) $M = 5$. (d) $M = 15$.

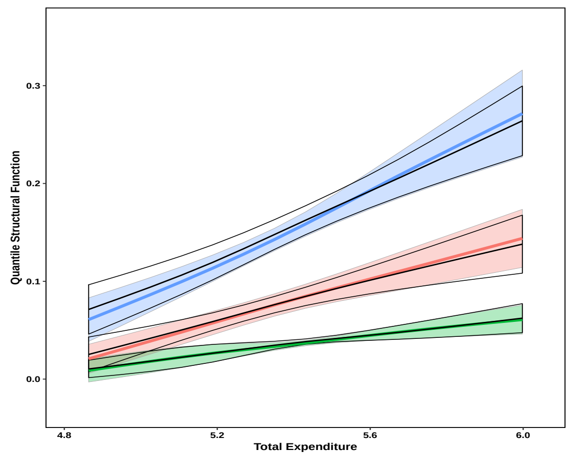
Figure 6.2: Design 2. QSF for food with discrete instrument $\widetilde{Z}$ (coloured) and with continuous instrument Z^* (black).

estimated using the original continuous instrument Z^*. Figures 6.3 and 6.4 show the corresponding QSFs for leisure. For comparison purposes the implementation is exactly as in Chernozhukov et al. (2017). We report weighted bootstrap 90%-confidence bands that are uniform over the support regions of the displayed QSFs,[5] constructed

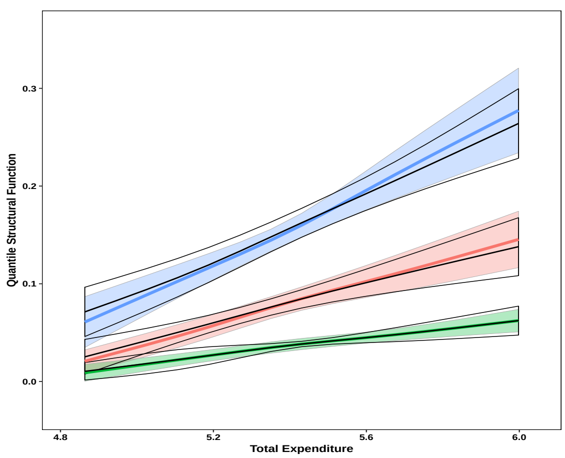
[5] All QSFs and uniform confidence bands are obtained over the region $[\widehat{Q}_X(0.1), \widehat{Q}_X(0.9)] \times \{0.25, 0.5, 0.75\}$, where the interval $[\widehat{Q}_X(0.1), \widehat{Q}_X(0.9)]$ is approximated by a grid of 5 points $\{\widehat{Q}_X(0.1), \widehat{Q}_X(0.3), \ldots, \widehat{Q}_X(0.9)\}$. For graphical representation the QSFs are then interpolated by splines over that interval.



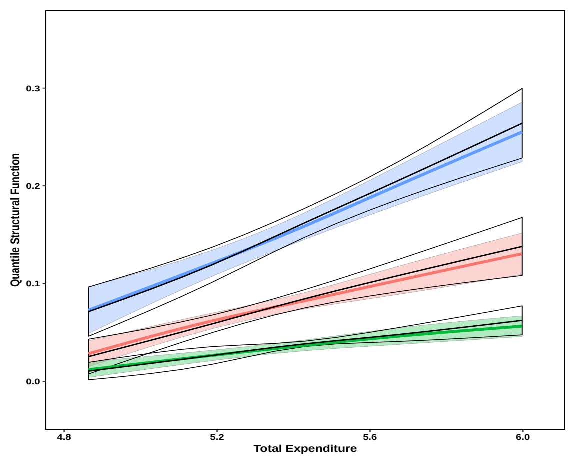
(a) $M = 2$.



(b) $M = 3$.



(c) $M = 5$.

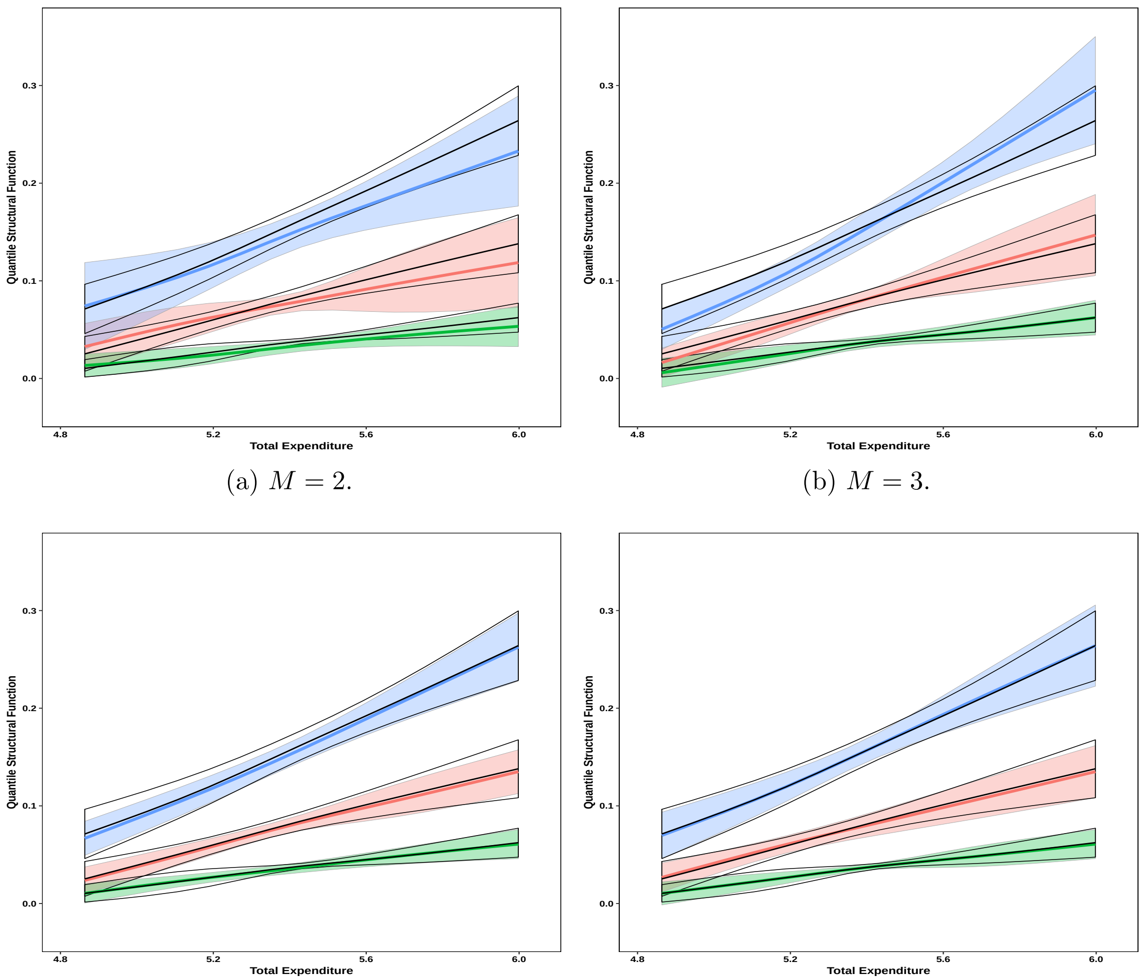


(d) $M = 15$.

Figure 6.3: Design 1. QSF for leisure with discrete instrument $\widetilde{Z}$ (coloured) and with continuous instrument Z^* (black).

with 199 bootstrap replications. Our empirical results show that both discretisation schemes deliver very similar QSF estimates and confidence bands that capture the main features of the benchmark QSFs estimated with a continuous instrument. The largest deviations from the benchmark QSFs occur for $M = 2$ and the non uniform Design 2, where the first value of $\widetilde{Z}$ is allocated to 6% of the observations only.

In Appendix D we show that our empirical findings also hold for the average and distribution structural functions. We also show that similar results hold for



(a) $M = 2$.

(b) $M = 3$.

(c) $M = 5$.

(d) $M = 15$.

Figure 6.4: Design 2. QSF for leisure with discrete instrument $\widetilde{Z}$ (coloured) and with continuous instrument Z^* (black).

nonlinear estimates of the QSF, when the vector $p(X)$ is augmented with spline transformations of X. Thus for this dataset the main features of Engel curves for food and leisure are well captured when estimation is performed with a discrete valued instrumental variable. Overall our empirical findings support our identification results and illustrate the use of discrete instruments for the estimation of structural functions in parametric triangular systems.

A Proof of Main Results

A.1 Proof of Theorem 1

Proof. Part (i). The proof builds on the proof of Lemma S3 in Spady and Stouli (2018). The matrix $E\left[w(X,V)w(X,V)'\right]$ is of the form

$$\begin{aligned} E\left[w(X,V)w(X,V)'\right] &= E\left[\{p(X)\otimes q(V)\}\{p(X)\otimes q(V)\}'\right] \\ &= E\left[\{p(X)p(X)'\}\otimes\{q(V)q(V)'\}\right] \\ &= E\begin{bmatrix} p(X)p(X)' & p(X)p(X)'V \\ p(X)p(X)'V & p(X)p(X)'V^2 \end{bmatrix}. \end{aligned}$$

Assumption 2(a) implies that $E[p(X)p(X)']$ is positive definite. Thus $E\left[w(X,V)w(X,V)'\right]$ is positive definite if and only if the Schur complement of $E[p(X)p(X)']$ in $E\left[w(X,V)w(X,V)'\right]$ is positive definite (Boyd and Vandenberghe, 2004, Appendix A.6), i.e. if and only if

$$\varUpsilon := E\left[p(X)p(X)'V^2\right] - E\left[p(X)p(X)'V\right]E\left[p(X)p(X)'\right]^{-1}E\left[p(X)p(X)'V\right]$$

satisfies $\det(\varUpsilon) > 0$.

With

$$\Xi = E\left[p(X)p(X)'V\right]E\left[p(X)p(X)'\right]^{-1},$$

we have that

$$\varUpsilon = E\left[\{p(X)V - \Xi p(X)\}\{p(X)V - \Xi p(X)\}'\right],$$

a finite positive definite matrix, if and only if for all $\lambda \neq 0$ there is no d such that $\Pr[\{\lambda' p(X)\}V = d'\{\Xi p(X)\}] > 0$; this is an application of the Cauchy-Schwarz inequality for matrices stated in Tripathi (1999).

For $\widetilde{\mathcal{X}} \subseteq \mathcal{X}_V^o$, positive definiteness of $E[1(X \in \widetilde{\mathcal{X}})p(X)p(X)']$ under Assumption 2(a) implies that for all $\lambda \neq 0$, $E[1(X \in \widetilde{\mathcal{X}})\{\lambda' p(X)\}^2] > 0$, which implies that for all $\lambda \neq 0$, the set $\{x \in \widetilde{\mathcal{X}} : \lambda' p(x) \neq 0\}$ has positive probability. By definition of $\mathcal{V}_x$ and the variance, we have that $\text{Var}(V \mid X = x) > 0$ for each $x \in \mathcal{X}_V^o$. Thus for all $\lambda \neq 0$, by Ξ being a constant matrix, there is no d such that $\Pr[\{\lambda' p(X)\}V = d'\{\Xi p(X)\}] > 0$, and $E[w(X,V)w(X,V)']$ is positive definite.

Part (ii). The proof is similar to Part (i). □

A.2 Proof of Theorem 2

Proof. Under the conditions of Theorem 1(i), $E[w(X,V)w(X,V)']$ exists and is positive definite. This follows from the proof of Theorem 1(i) upon substituting for $V = F_{X|Z}(X \mid Z)$ throughout, and using that Assumption 3(a) and the definitions of $\mathcal{Z}_x$ and the variance imply that $\text{Var}(F_{X|Z}(x \mid Z) \mid X = x) > 0$ for each $x \in \mathcal{X}_Z^o$. A similar argument shows that $E[w(X,V)w(X,V)']$ exists and is positive definite under the conditions of Theorem 1(ii) and Assumption 3(b). Identification of the structural functions then follows by Theorem 1 in Chernozhukov et al. (2017). □

A.3 Proof of Theorem 3

Proof. By iterated expectations, $E[w(X,V)w(X,V)']$ can be expressed as

$$E[w(X,V)w(X,V)'] = E[\{p(X)p(X)'\} \otimes E[q(V)q(V)' \mid X]].$$

We show that $E\left[w(X,V)w(X,V)'\right]$ is positive definite. By Assumption 2(a), there is a positive constant B such that

$$
\begin{aligned}
E\left[\{p(X)p(X)'\}\otimes E[q(V)q(V)' \mid X]\right] &\geq E\left[1(X\in\widetilde{\mathcal{X}})\{p(X)p(X)'\}\otimes\lambda_{\min}(X)I_K\right] \\
&\geq E\left[1(X\in\widetilde{\mathcal{X}})\{p(X)p(X)'\}\right]\otimes BI_K,
\end{aligned}
$$

where I_K is the $K\times K$ identity matrix, and the inequality means no less than in the usual partial ordering for positive semi-definite matrices. The conclusion then follows by the matrix following the last inequality being positive definite by Assumption 2(a).

Under Assumption 2(b) the proof is similar upon using that $E\left[w(X,V)w(X,V)'\right] = E\left[E[p(X)p(X)' \mid V]\otimes\{q(V)q(V)'\}\right]$. □

A.4 Proof of Theorem 4

Proof. By Theorem 3 the matrix $E\left[w(X,V)w(X,V)'\right]$ exists and is positive definite. The result then follows by Theorem 1 in Chernozhukov et al. (2017). □

A.5 Proof of Proposition 1

Proof. The proof builds on the proof of Proposition 1 in Newey and Stouli (2018). For $x\in\mathcal{X}_Z^*$, by definition of $\mathcal{Z}_x$ we have that $\Pr(Z=z_m \mid X=x)\geq\delta>0$ for $m\in\{1,\ldots,|\mathcal{Z}_x|\}$, and

$$
\begin{aligned}
E[q(V)q(V)' \mid X=x] = \sum_{m=1}^{|\mathcal{Z}_x|} \Big\{ &\left\{q(F_{X|Z}(x\mid z_m))q(F_{X|Z}(x\mid z_m))'\right\} \\
&\times\Pr(Z=z_m\mid X=x)\Big\},
\end{aligned}
$$

is a sum of $|\mathcal{R}(x)| \leq |\mathcal{Z}_x|$ rank one $K \times K$ distinct matrices which is singular if $|\mathcal{R}(x)| < K$. If there is no set $\widetilde{\mathcal{X}} \subseteq \mathcal{X}$ of positive probability such that $\inf_{x \in \widetilde{\mathcal{X}}} |\mathcal{R}(x)| \geq K$, the matrix $E[q(V)q(V)' \mid X]$ is then singular with probability one, and Assumption 2(a) cannot hold with $\widetilde{\mathcal{X}} \subseteq \mathcal{X}_Z^*$, by definition of $\mathcal{X}_Z^*$. Therefore Assumption 2(a) holds with $\widetilde{\mathcal{X}} \subseteq \mathcal{X}_Z^*$ only if there is a set $\widetilde{\mathcal{X}}$ of positive probability such that $\inf_{x \in \widetilde{\mathcal{X}}} |\mathcal{R}(x)| \geq K$. A similar argument shows that $\mathcal{V}_Z^*$ has positive probability only if there is a set $\widetilde{\mathcal{V}} \subseteq (0, 1)$ of positive probability such that $\inf_{v \in \widetilde{\mathcal{V}}} |\mathcal{Q}(v)| \geq J$. □

A.6 Proof of Theorem 5

Proof. The result follows from the proof of Theorem 1 in Newey and Stouli (2018). □

A.7 Proof of Theorem 6

Proof. The proof builds on the proof of identification of the average structural function in Theorem 2 of Newey and Stouli (2018). Under Assumption 4(a), $q_\tau(V)$ is identified for each $\tau \in \mathcal{T}$ by Theorem 5. This implies that, for $\mathcal{T} = \mathcal{Y}$, the conditional CDF $F_{Y|XV}(y \mid X, V) = \Gamma(p(X)'q_y(V))$ is unique with probability one for each $y \in \mathcal{Y}$, since $p(X)$ and Γ are known functions. The structural functions are then identified by (2.2) in the main text. For $\mathcal{T} = \mathcal{U}$, when Y is continuous the conditional quantile function $Q_{Y|XV}(u \mid X, V) = p(X)'q_u(V)$ is unique with probability one for each $u \in \mathcal{U}$. Since $y \mapsto F_{Y|XV}(y \mid XV)$ is the inverse function of $u \mapsto Q_{Y|XV}(u \mid X, V)$, the structural functions are also identified by (2.2) in the main text. □

B Formal Statement of Remark 2

Proposition 2. *(i) Let $q(V) = (1, V)'$. If Assumption 2(a) holds with $\widetilde{\mathcal{X}} \subseteq \mathcal{X}_V^o$ then it also holds with $\widetilde{\mathcal{X}} \subseteq \mathcal{X}_V^*$. (ii) Let $p(X) = (1, X)'$. If Assumption 2(a) holds with $\widetilde{\mathcal{V}} \subseteq \mathcal{V}_X^o$ then it also holds with $\widetilde{\mathcal{V}} \subseteq \mathcal{V}_X^*$.*

Proof. Part (i). Each $x \in \mathcal{X}_V^o$ satisfies $|\mathcal{V}_X| \geq 2$, which by the definitions of $\mathcal{V}_X$ and the variance implies that $\mathrm{Var}(V \mid X = x) \geq B > 0$. For $q(V) = (1, V)'$, the smallest eigenvalue of $E[q(V)q(V)' \mid X = x]$ is then bounded away from zero for each $x \in \mathcal{X}_V^o$, by Lemma 1 below. Therefore each $x \in \mathcal{X}_V^o$ also satisfies $x \in \mathcal{X}_V^*$, so that $\mathcal{X}_V^o \subseteq \mathcal{X}_V^*$. The result follows.

Part (ii). The proof is similar to Part (i). □

Lemma 1. *For a set of random variables $\{X(t)\}_{t \in \overline{\mathcal{T}}}$ such that $E[X(t)^2] \leq C$ and Var$(X(t)) \geq B > 0$, the smallest eigenvalue of*

$$\Sigma(t) = E\left[\begin{pmatrix} 1 \\ X(t) \end{pmatrix} \begin{pmatrix} 1 & X(t) \end{pmatrix}\right]$$

is bounded away from zero.

Proof. $\det(\Sigma(t)) = \mathrm{Var}(X(t)) = \lambda_{\max}(t)\lambda_{\min}(t)$ where $\lambda_{\max}(t)$ and $\lambda_{\min}(t)$ are the largest and smallest eigenvalues of $\mathrm{Var}(X(t))$, respectively. Note that for all $t \in \overline{\mathcal{T}}$

$$\lambda_{\max}(t) = \sup_{\lambda:||\lambda||=1} \lambda'\Sigma(t)\lambda \leq ||\lambda||^2||\Sigma(t)|| \leq ||\Sigma(t)|| \leq \widetilde{C}$$

by $E[X(t)^2]$ bounded. Therefore

$$\lambda_{\min}(t) = \frac{\mathrm{Var}(X(t))}{\lambda_{\max}(t)} \geq \frac{\mathrm{Var}(X(t))}{\widetilde{C}} \geq \frac{B}{\widetilde{C}},$$

and the result follows. □

C Quantile Regression Estimation of Structural Functions

In this Section we give a simplified summary of the key steps in the implementation of the quantile regression-based estimators for structural functions proposed by Chernozhukov et al. (2017). A detailed description and implementation algorithms for estimation and the weighted bootstrap procedures are given in Chernozhukov et al. (2017).

The estimators implemented in the empirical application have three main stages. In the first stage, we estimate the control variable, $\{\hat{V}_i\}_{i=1}^n$. In the second stage, we estimate the CDF, $\hat{F}_{Y|XZ_1V}(y \mid x, z_1, v)$. In the third and final stage, estimators $\widehat{G}(y,x)$, $\widehat{Q}(\tau,x)$ and $\widehat{\mu}(x)$ of the distribution, quantile and average structural functions, respectively, are obtained.

[First stage.] Denoting the usual check function by $\rho_v(z) = (v - 1(z < 0))z$, the quantile regression estimator for $F_{X|Z}$ is, for $(x,z) \in \mathcal{XZ}$,

$$\begin{aligned}
\hat{F}_{X|Z}(x \mid z) &= \epsilon + \int_{\epsilon}^{1-\epsilon} 1\left\{\hat{\pi}(v)'[s(\widetilde{z}) \otimes r(z_1)] \leq x\right\} dv, \\
\hat{\pi}(v) &\in \arg\min_{\pi \in \mathbb{R}^{\dim(Z)}} \sum_{i=1}^{n} \rho_v(X_i - \pi'[s(\widetilde{Z}_i) \otimes r(Z_{1i})]),
\end{aligned}$$

for some small constant $\epsilon > 0$. The control function estimator is then $\hat{V}_i = \hat{F}_{X|Z}(X_i \mid Z_i)$, $i \in \{1,\ldots,n\}$.

[Second stage.] The quantile regression estimator for $F_{Y|XZ_1V}$ is, for $(y,x,z_1,v) \in$

$\mathcal{Y}\mathcal{X}\mathcal{Z}_1\mathcal{V}$,

$$\begin{aligned}\hat{F}_{Y|XZ_1V}(y \mid x, z_1, v) &= \epsilon + \int_{\epsilon}^{1-\epsilon} 1\left\{\hat{\beta}(u)'w(x, z_1, v) \leq y\right\} du, \\ \hat{\beta}(u) &\in \arg \min_{\beta \in \mathbb{R}^{\dim(W)}} \sum_{i=1}^{n} \rho_u(Y_i - \beta' w(X_i, Z_{1i}, \hat{V}_i)).\end{aligned}$$

[Third stage.] Given estimates $(\{\hat{V}_i\}_{i=1}^n, \hat{F}_{Y|XZ_1V})$, the estimator for the DSF takes the form

$$\hat{G}(y, x) = \frac{1}{n} \sum_{i=1}^{n} \hat{F}_{Y|XZ_1V}(y \mid x, Z_{1i}, \hat{V}_i).$$

Given the DSF estimate, the QSF estimator is then defined as

$$\widehat{Q}(p, x) = \int_{\mathcal{Y}^+} 1\{\widehat{G}(y, x) \leq p\} dy - \int_{\mathcal{Y}^-} 1\{\widehat{G}(y, x) \geq p\} dy,$$

and the ASF estimator as

$$\widehat{\mu}(x) = \int_{\mathcal{Y}^+} [1 - \widehat{G}(y, x)] \nu(dy) - \int_{\mathcal{Y}^-} \widehat{G}(y, x) \nu(dy).$$

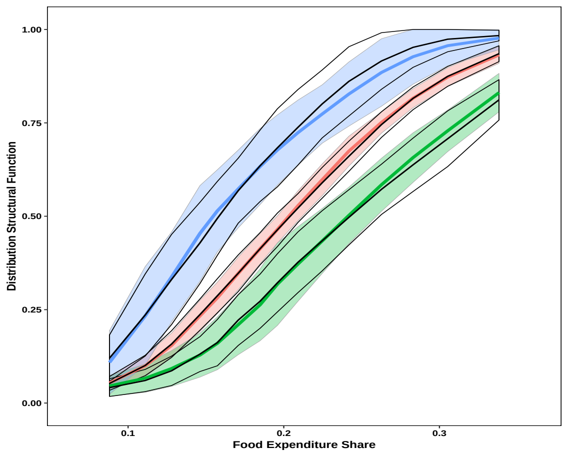
D Additional Results for the Empirical Application

In this Section we complement the empirical analysis of Section 6 by studying the robustness of the empirical findings for the QSF. We first report estimates for the DSF in Section D.1. We also estimated the average structural function for each good and each instrument and the results lead to similar conclusions to the QSF and DSF, and are thus omitted. We then report more flexible QSF estimates including spline transformations of the endogenous variable X. Overall, our robustness checks show that our empirical results are robust across structural functions and instrument specifications, and our additional results confirm our findings for the QSF discussed in the main text.

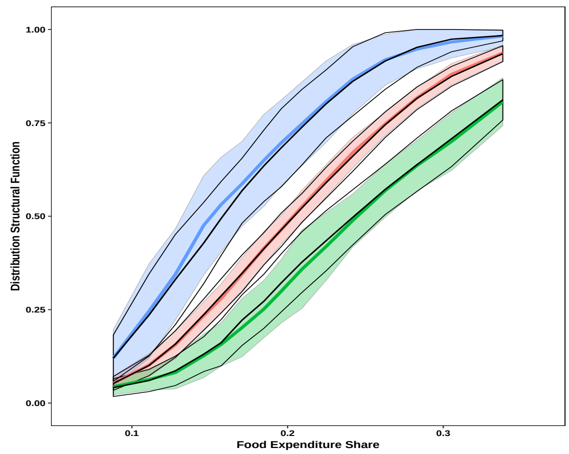
D.1 Distribution Structural Functions

To further check the robustness of our empirical findings we also estimated the DSF. For the DSF estimate $\widehat{G}(y,x)$ we give weighted bootstrap confidence bands uniform over the region $[\widehat{Q}_Y(0.1), \widehat{Q}_Y(0.9)] \times \{\widehat{Q}_X(0.1), \widehat{Q}_X(0.5), \widehat{Q}_X(0.9)\}$, constructed with 199 bootstrap replications. For the equispaced grid $0.1 = t_1 < \ldots < t_{15} = 0.9$, in our implementation the interval $[\widehat{Q}_Y(0.1), \widehat{Q}_Y(0.9)]$ is approximated by a grid of 15 points $\{\widehat{Q}_Y(t_1), \ldots, \widehat{Q}_Y(t_{15})\}$.

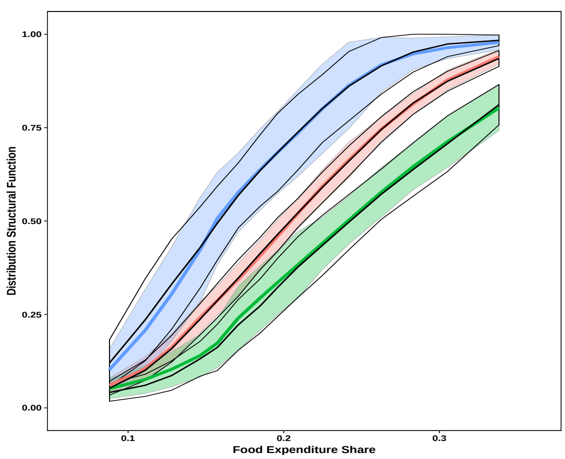
For each $x \in \{\widehat{Q}_X(0.1), \widehat{Q}_X(0.5), \widehat{Q}_X(0.9)\}$, Figures D.1 and D.2 show the corresponding three DSFs for food estimated with each set of four instruments, respectively, as well as the corresponding benchmark DSFs estimated using the original continuous instrument Z^*. Figures D.3 and D.4 show the corresponding DSFs for leisure. Similarly to the QSF estimates, our empirical results show that both discretisation schemes deliver very similar DSF estimates and confidence bands that capture the main features of the benchmark DSFs, and the largest deviations from the benchmark DSFs occur for $M = 2$ and the non uniform Design 2.



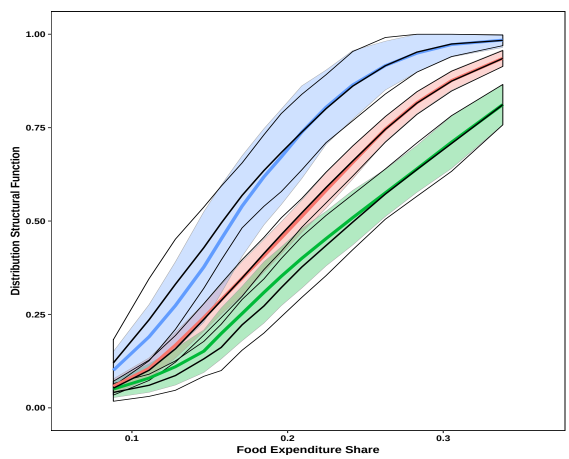
(a) $M = 2$.



(b) $M = 3$.

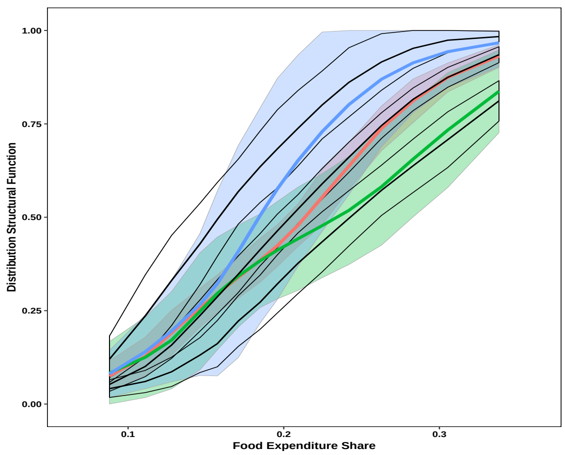


(c) $M = 5$.

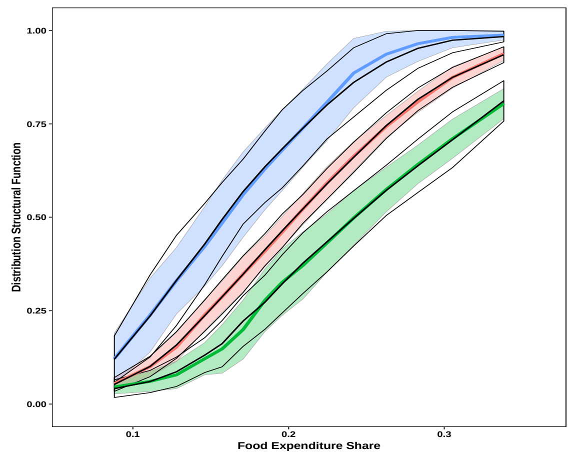


(d) $M = 15$.

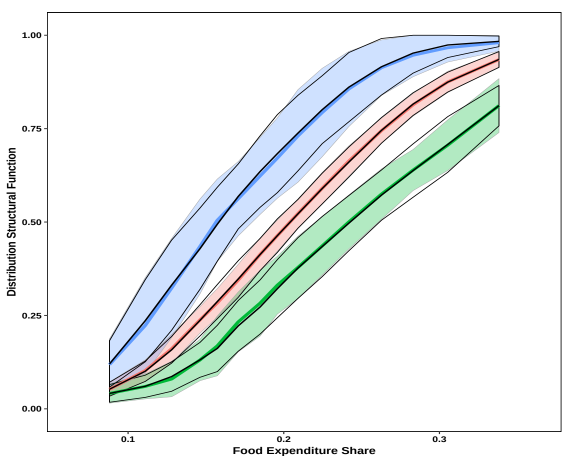
Figure D.1: Design 1. DSF for food with discrete instrument $\widetilde{Z}$ (coloured) and with continuous instrument Z^* (black).



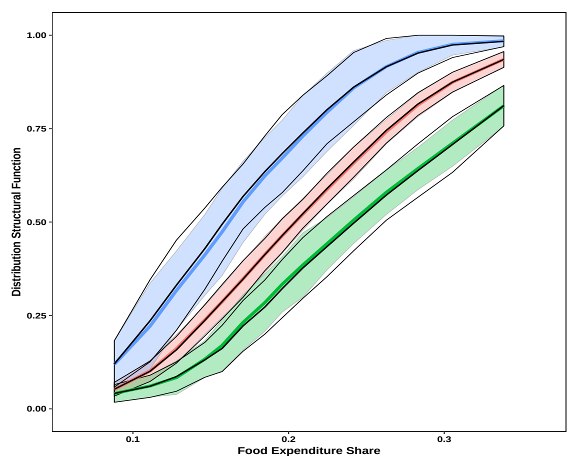
(a) $M = 2$.



(b) $M = 3$.

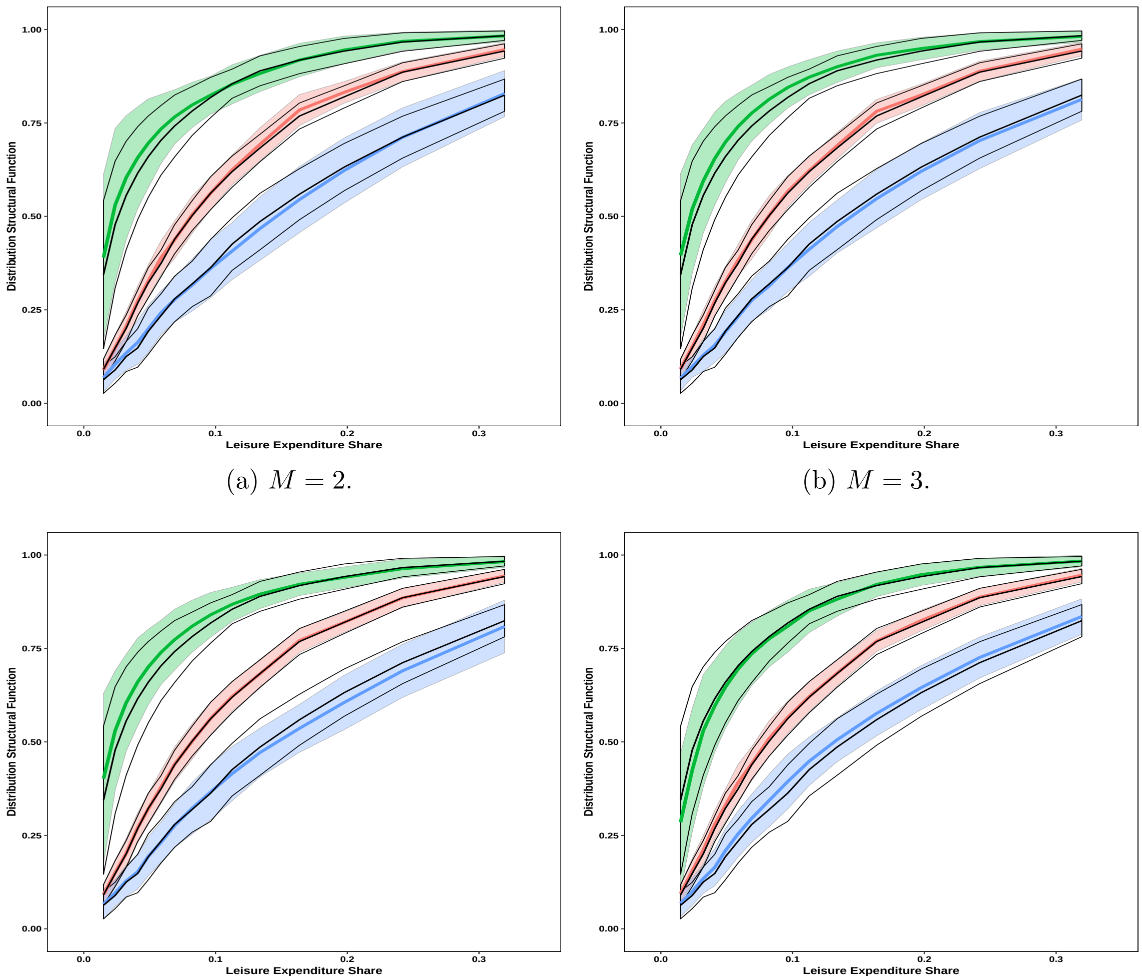


(c) $M = 5$.



(d) $M = 15$.

Figure D.2: Design 2. DSF for food with discrete instrument $\widetilde{Z}$ (coloured) and with continuous instrument Z^* (black).



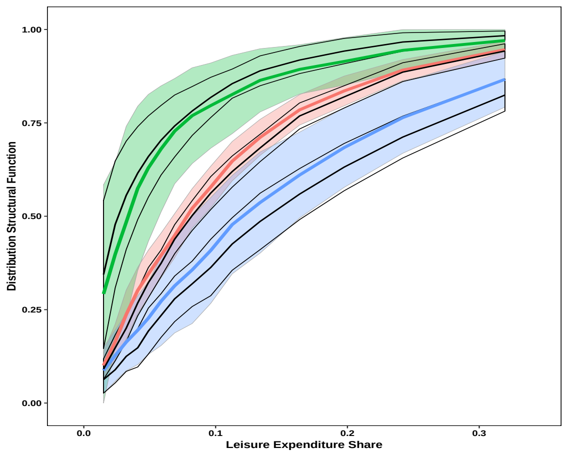
(a) $M = 2$.

(b) $M = 3$.

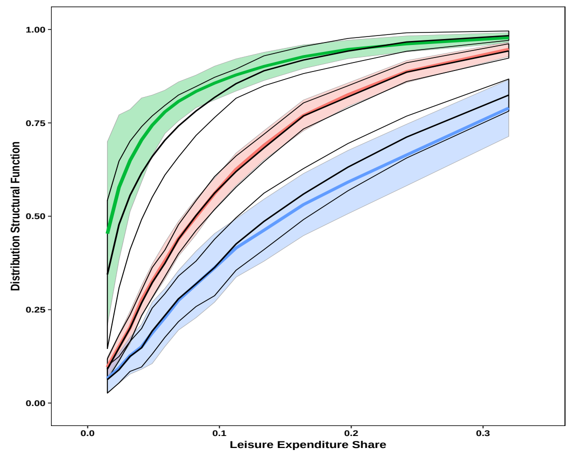
(c) $M = 5$.

(d) $M = 15$.

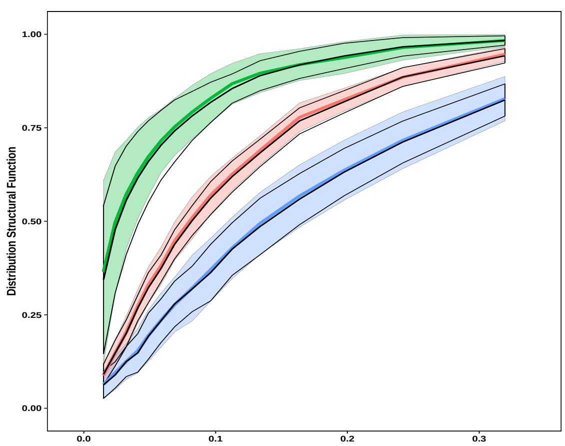
Figure D.3: Design 1. DSF for leisure with discrete instrument $\widetilde{Z}$ (coloured) and with continuous instrument Z^* (black).



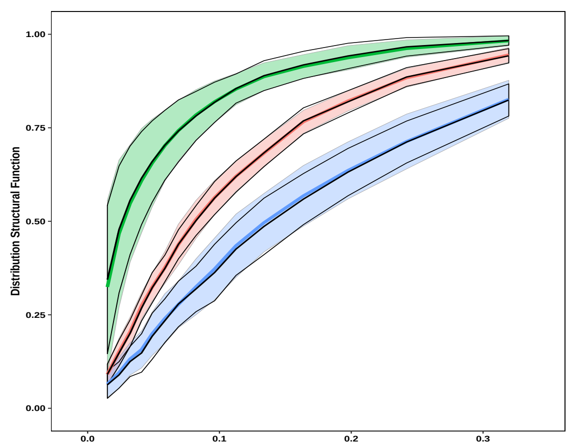
(a) $M = 2$.



(b) $M = 3$.



(c) $M = 5$.

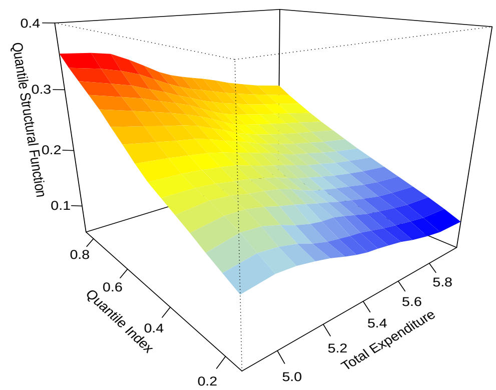


(d) $M = 15$.

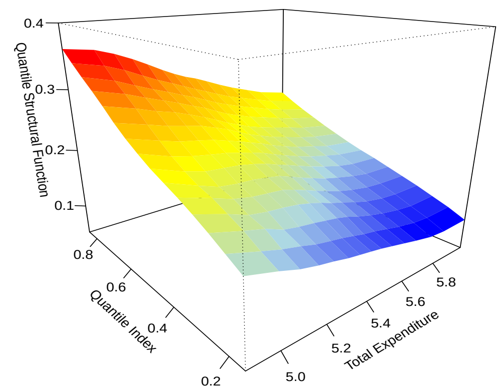
Figure D.4: Design 2. DSF for leisure with discrete instrument $\widetilde{Z}$ (coloured) and with continuous instrument Z^* (black).

D.2 Nonlinear Estimation

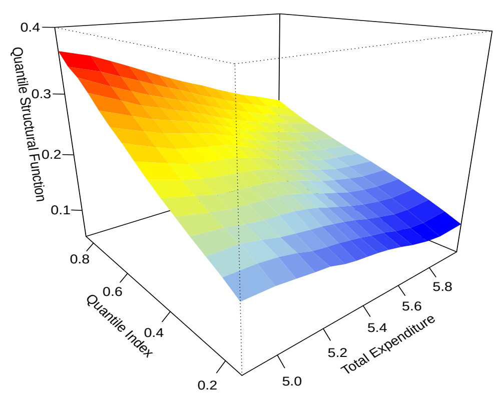
Our results in the main text show that identification is robust to increasing the number of terms in the specification of the regressor vector. Here we consider including spline transformations of X in the specification of $p(X)$ in order to account for potential nonlinearities in data. This increased flexibility allows for the estimation of nonlinear QSF without the need for a continuous instrument. We estimate the QSF for food and leisure by taking cubic B-splines transformations with 4 knots of log-total expenditure, with each set of four instruments, respectively. For the QSF estimate $\widehat{Q}(p, x)$ we give weighted bootstrap confidence bands uniform over the region $[0.15, 0.85] \times [\widehat{Q}_X(0.1), \widehat{Q}_X(0.9)]$, constructed with 199 bootstrap replications. Figures D.5 and D.6 show the QSFs for food for each set of instruments. Figures D.7 and D.8 show the QSFs for leisure for each set of instruments. These results show that our flexible estimates of the structural relationship between total expenditure and food and leisure shares are robust to instrument discretisation for both food and leisure.



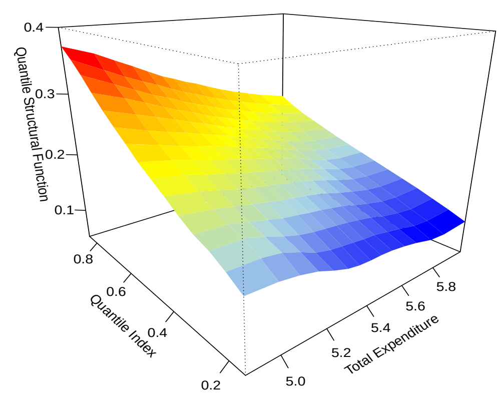
(a) $M = 2$.



(b) $M = 3$.

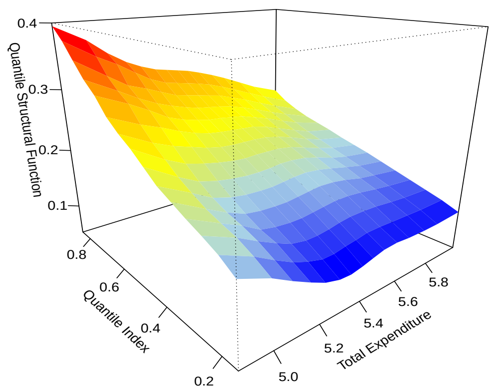


(c) $M = 5$.

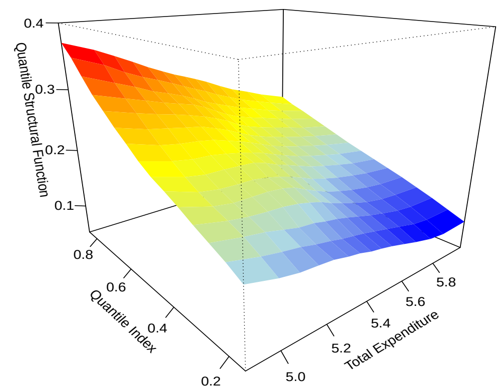


(d) $M = 15$.

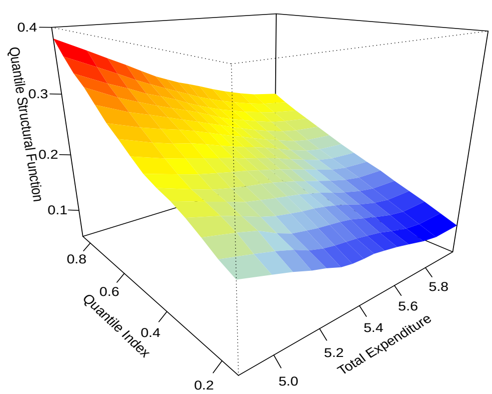
Figure D.5: Design 1. QSF for food.



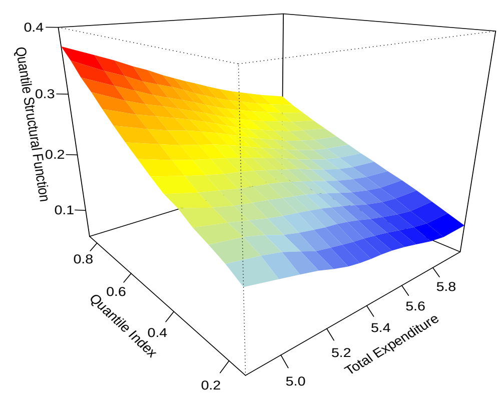
(a) $M = 2$.



(b) $M = 3$.

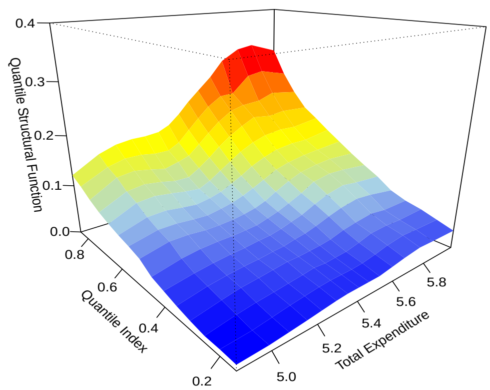


(c) $M = 5$.

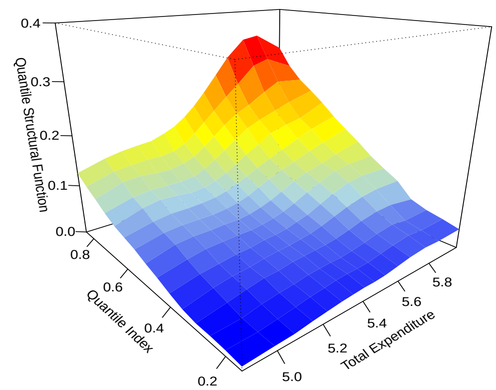


(d) $M = 15$.

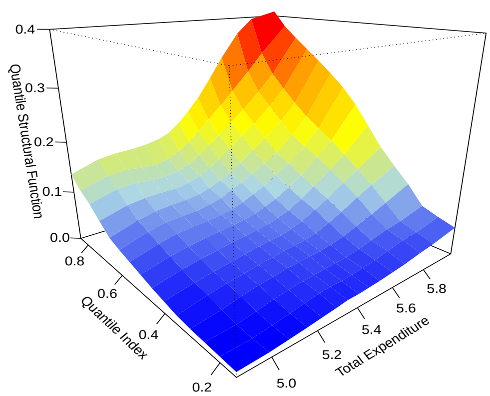
Figure D.6: DESIGN 2. QSF for food.



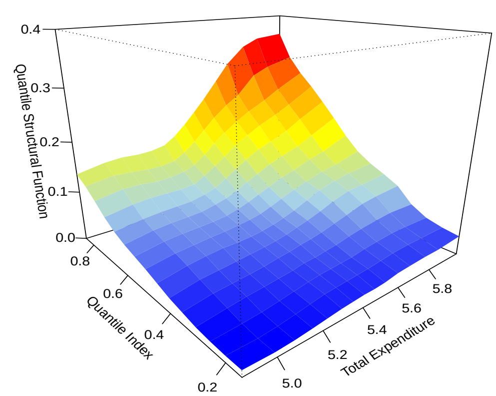
(a) $M = 2$.



(b) $M = 3$.

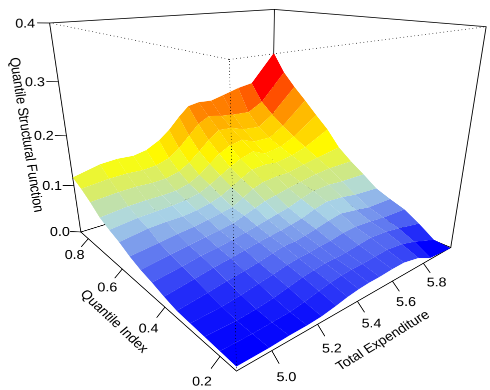


(c) $M = 5$.

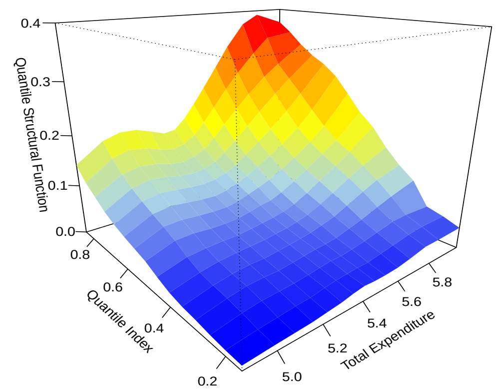


(d) $M = 15$.

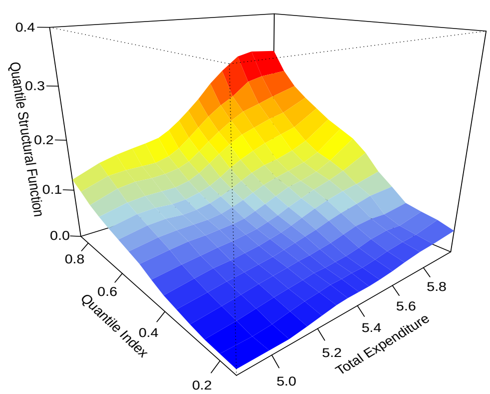
Figure D.7: Design 1. QSF for leisure.



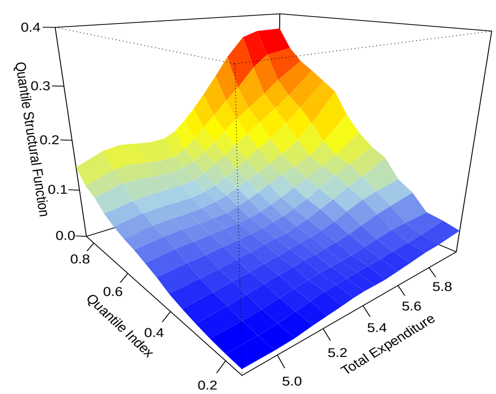
(a) $M = 2$.



(b) $M = 3$.



(c) $M = 5$.



(d) $M = 15$.

Figure D.8: Design 2. QSF for leisure.

References

[1] Boyd, S. P. and L. Vandenberghe, 2004, Convex optimization. Cambridge University Press, Cambridge.

[2] Blundell, R., and J. L. Powell, 2003, Endogeneity in nonparametric and semiparametric regression models. Econometric society monographs 36, pp. 312-357.

[3] Blundell, R., and J. L. Powell, 2004, Endogeneity in semiparametric binary response models. The Review of Economic Studies 71, pp. 655-679.

[4] Chernozhukov, V., Fernandez-Val, I., and A. Kowalski, 2015, Quantile regression with censoring and endogeneity. Journal of Econometrics 186, pp. 201-221.

[5] Chernozhukov, V., Fernandez-Val, I. Newey, W. K., Stouli, S. and F. Vella, 2017, Semiparametric estimation of structural functions in nonseparable triangular models. Arxiv.

[6] D'Haultfœuille, X. and P. Février, 2015, Identification of nonseparable triangular models with discrete instruments. Econometrica 83, pp. 1199-1210.

[7] Fernandez-Val, I. Van Vuuren, A. and F. Vella, 2017, Nonseparable sample selection models with censored selection rules. Arxiv.

[8] Florens, J. P., Heckman, J. J., Meghir, C. and E. Vytlacil, 2008, Identification of treatment effects using control functions in models with continuous, endogenous treatment and heterogeneous effects. Econometrica 76, pp. 1191-1206.

[9] Hausman, J. A. and W. K. Newey, 2016, Individual heterogeneity and average welfare. Econometrica 84, pp.1225-1248.

[10] Hausman, J. A. and D. Wise, 1978, A conditional probit model for qualitative choice: discrete decisions recognizing interdependence and heterogeneous preferences. Econometrica 46, pp. 403-426.

[11] Hoderlein, S. and E. Mammen, 2007, Identification of marginal effects in nonseparable models without monotonicity. Econometrica 75, pp. 1513-1518.

[12] Imbens, G. and W. K. Newey, 2009, Identification and estimation of triangular simultaneous equations models without additivity. Econometrica 77, pp. 1481-1512.

[13] Kitamura, Y. and J. Stoye, 2017, Nonparametric analysis of random utility models. Arxiv.

[14] McFadden, D., 1973, Conditional logit analysis of qualitative choice behavior, in: P. Zarambka (Ed.), Frontiers in econometrics. New York: Academic Press.

[15] Newey, W.K. and D. McFadden, 1994, Large sample estimation and hypothesis testing, in: Engle, R. and D. McFadden (Eds.), Handbook of econometrics. Elsevier, Berlin, pp. 2111-2245.

[16] Newey, W. K. and S. Stouli, 2018, Identification in partially parametric conditional independence models. Manuscript in preparation.

[17] Rosenbaum, P. R. and D. B. Rubin, 1983, The central role of the propensity score in observational studies for causal effects. Biometrika 70, pp.41-55.

[18] Spady, R. H. and S. Stouli, 2018, Dual regression. Biometrika 105, pp. 1-18.

[19] Torgovitsky, A., 2015, Identification of nonseparable models using instruments with small support. Econometrica 83, pp. 1185-1197.

[20] Tripathi, G., 1999, A matrix extension of the Cauchy-Schwarz inequality. Economics Letters 63, pp. 1-3.